Optimal Central Redistribution with Decentralized Redistribution[‡]

Chunyang Fu
UCASS

Liutang Gong
Peking University

Wenjian Li
Zhejiang University

March 22, 2025

Abstract

This paper formalizes a framework to study optimal central income tax with strategic local governments who also redistribute with nonlinear income tax and compete tax with each other. It develops a sufficient-statistics approach to the multi-level contract design and identify a concise ABC-nest-ABC tax rule based on a few micro statistics. Our formula demonstrates precisely how the central government can engage local governments and achieve the optimal tax system using empirically observable data. The classic Diamond ABC is integrated into a novel Pigouvian ABC, comprising: (i) economic weight for each region, (ii) inter-government objective inconsistency, and (iii) the Diamond ABC. Unlike the canonical Mirrleesian tax, this optimal tax specifically addresses inconsistencies in inter-government objectives, including fiscal externality. A simple quantitative analysis to the U.S. indicates that decentralized redistribution is generally beneficial, particularly when supported by a well-designed central income tax. The welfare effect of local taxes is minor because the central income tax significantly crowds out local taxes while stabilizing the overall tax burden. From another perspective, local taxes address local needs without greatly impacting national welfare. The central income tax is essential for redistribution, even with free inter-region transfers. In contrast, implementing region-specific central income taxation would only slightly improve national welfare. (A quantitative analysis covering 51 jurisdictions in the U.S. is to be done)

Keywords. optimal income tax; objective inconsistency; multi-level government
JEL. H21;H71;H77.

[*]We thank Chong Liu, Mikhail Golosov, Shenxing Zhang, Jan Eeckhout, Xianwen Shi, Xiaoyong Cui, Guanrong Ma, Jinyou Li, Rong Li, Fucheng Wei, and the participants of the 2024 Asian Meeting of the Econometric Society and 7th China Public Finance Conference for their very useful comments. We also thank numerous colleagues and seminar audiences for their helpful comments. We gratefully acknowledge the support from the National Natural Science Foundation of China (Grant No. 72373133, 72003198), the National Social Science Foundation of China (Grant No. 19ZDA069), the WU Jiapei Award for Information Economics in 2016.

[†]Chunyang Fu, School of Economics, UCASS, fuchunyang@pku.edu.cn; Liutang Gong, Guanghua School of Management, Peking University, and Beijing Technology and Business University, ltgong@gsm.pku.edu.cn; Wenjian Li, School of Economics, Zhejiang University, wenjian_li@pku.edu.cn.

1 Introduction

The optimal Mirrleesian tax literature often assumes that there is a sole redistribution policymaker. However, in practice, multiple governments typically enforce redistribution policies in the same region. Nevertheless, organizations with varying degrees of centralization do exhibit significant differences in their redistribution policies, with more centralized governments generally having more centralized redistribution policies. This is exactually the case among the EEC, U.S., and China.[1] A crucial question thus arises: is decentralized redistribution desirable from a national perspective? And, how to design central redistribution with decentralized redistribution?

There are two difficulties in designing the central redistribution policy. Take the U.S. as an example. First, the diverse economic landscapes of U.S. states pose a significant challenge to the design of a fair and efficient federal income tax. Moreover, the U.S. Constitution mandates the uniformity of federal income tax.[2] Such a one-size-fits-all federal income tax system struggles to adequately address the inherent disparities. The central govenment has to balance the redistribution between regions and within regions. Second, central governments that want or are obliged to compress income gaps must take residents' incentive problems, as well as the local governments' reactions, into consideration. Empirical studies documents two important facts on how the behavior of local governments has changed the effectiveness of central redistribution. Higher central income tax rates lead to lower local income tax rates (see, e.g., Feldstein and Metcalf (1987); Goodspeed (2000)). Moreover, central tax reforms that would increase the progressivity of state effective tax rates result in decreased progressivity of statutory state income taxes (see, e.g., Scott and Triest (1993)).

Despite the global prevalence of multi-level governments, plausible empirical evidence on strategic responses of local government, and the challenge to the tax design for a union, normative research in this context is notably scarce. This paper thus studies optimal region-uniform central income tax with region-specific local income taxes designed by local governments. The underlying question here, which lies at the core of optimal taxation and fiscal federalism, is how to construct a central income tax that effectively leverages the strengths of both central and local governments while navigating their inherent conflicts. [3]

Bearing the above phenomenons and questions in mind, we formalize a multi-principal Mirrlees tax problem in a sequential game consisting of a central government, two asymmetric local governments, and a continuous heterogeneous taxpayer. Multi-level governments contract sequentially and non-cooperatively with the residents in their jurisdictions. At the beginning of the economy, the central government set its income tax schedule while considering the reactions of local governments and taxpayers. Local governments

[1] Many, if not most, countries and unions are governed by multi-level governments and all levels of government collect tax formally or informally. For example, in the United States, both federal and state governments levy income taxes. In the European context, both member states and the European Economic Community itself enact redistribution (such as the European Social Fund and the European Regional Development Fund). In China, although the local governments cannot formulate local income tax, they adjust the effective tax through a talent subsidy policy, a government-funded projects, and tax administration. These typical examples suggest that (i) despite differences in the degree of decentralization, multi-level redistribution is popular around the world; and (ii) typically, the more centralized the organization is, the more centralized its redistribution policies is (in the above cases, EEC is the most decentralized and China is the most centralized).

[2] Article I, Section 8, Clause 1 of the Constitution states that "all Duties, Imposts, and Excises shall be uniform throughout the United States." The U.S. courts have interpreted "uniformity" to mean geographical uniformity in how the tax is applied, not necessarily that everyone pays the same rate. This allows for a progressive tax system where rates vary based on income levels.

[3] Alexis de Tocqueville observed this nearly two centuries ago (see, e.g., Tocqueville (2003)): "The federal system was created with the intention of combining the different advantages that result from the magnitude and the littleness of nations."

then design their own income taxes, taking the central tax as given and anticipating taxpayer responses. Lastly, taxpayers, observing the tax system, act to maximize their utilities. Unlike the local governments, which consider only the local welfare, the central government is concerned with the national welfare. On the policy side, we follow the spirit of Mirrlees tax both to abstract the analysis from functional constraints on taxes and to capture the nonlinear reaction of local tax in the real economy. By doing so, we have, for the first time, unveiled the key principles of nonlinear tax design in a multi-level government system.

Our main contribution is developing a sufficient-statistics approach for the contract design in a multi-level principal-agent problem with both intensive-margin (labor supply) and extensive-margin (location choice) decisions. Follow the Mirrlees (1971) tradition, we allow all taxes to be arbitrarily non-linear.[4] Local income taxes are region-specific. In contrast, the central income tax is region-uniform, capturing political or information constraints of the central government. As Oates (1999) observed, central governments have been historically, and continue to be, prevented from implementing completely region-specific redistribution policies. Actually, the U.S. Constitution mandates the uniformity of federal income tax.[5] Another key feature of the current paper is the explicit objective inconsistency between governments. This inconsistency, often implicit and taken for granted, is the basis of the other issues. Without objective inconsistency, tax design reduces to a planning problem for a benevolent social planner.[6] Our paper highlights the impact of objective inconsistency and multi-level principal-agent problem on the tax design. Below, we elaborate on the baseline and extension components of this study, respectively.

In the benchmark model, we segregate the problem from labor mobility and highlight the role of objective inconsistency. We first analyze how local governments react to the central income tax, which determines the optimal local tax. Levying central income tax reduces local governments' benefit from motivating labor supply, which in turn increases the aggregate local income tax rate (i.e., the combined tax rate of the local and central income taxes). However, the aggregate tax rate is sticky in the sense that one unit increase of the central income tax rate crowds in less than one unit of aggregate tax rate. Therefore, although the aggregate tax rate increases, the local income tax rate is crowded out.

These observations align with empirical evidence. Previous empirical studies have demonstrated that increases in the level and progressivity of central income tax are often accompanied by decreases in the corresponding aspects of local income tax, partially counteracting the impact of central tax reforms (see, e.g., Feldstein and Metcalf (1987); Scott and Triest (1993); Goodspeed (2000)). Our theoretical framework echoes these empirical findings, demonstrating consistency in both the offsetting effect and the overall effects; and in both the level and progressivity of taxes.

Our primary finding is a statistic-based *ABC-nest-ABC* tax rule, which establishes our main findings. The rule integrates the classic Diamond *ABC* as the term *C* within a novel Pigouvian *ABC* formula. Specifically, the Pigouvian *A*, *B*, and *C* terms comprise: (*i*) the classic Diamond *ABC*, forming the baseline of

[4] We consider nonlinear taxes because (i) Empirical research has found that local taxes respond to central taxes in both level and progressive aspects. (ii) Reducing ad hoc policy constraints highlights the key of our tax design

[5] Article I, Section 8, Clause 1 of the Constitution states that "all Duties, Imposts, and Excises shall be uniform throughout the United States." The U.S. courts have interpreted "uniformity" to mean geographical uniformity. This allows for a progressive tax system where rates vary based on income levels.

[6] As is emphasized by Drazen (2018), externality and asymmetric information problems are relevant only with objective inconsistency.

the central tax rule; (*ii*) objective inconsistency, reflecting how inter-governmental conflicts cause deviations from the canonical Mirrleesian tax rule (the Diamond ABC); and (*iii*) region-specific economic weights, capturing the trade-off in redistribution across different regions.

The key insight of our optimal tax formula is its focus on addressing objective inconsistency, which is quite different from the canonical Mirrleesian tax. The central income tax is limited to remedy the objective inconsistency rather than forming incentives for the taxpayers. Therefore, the optimal central income tax rate is positive (negative) when the central government has a relatively higher (lower) obligation to redistribute. Second, there is a trade-off between intra-region redistributions. Since the central income tax does not vary by locality, the policy design balances redistributive effects within each locality, which is the core of the *ABC-nest-ABC* formula. Last, there is an interaction between the inter-region transfer and central income tax. We find that inter-region transfers help reduce inter-region objective inconsistency, which disappears in the benchmark with optimal transfers. If transfers are imperfect, the central income tax assumes the role of inter-region redistribution. In such cases, the optimal marginal tax rates at boundaries are not zero, challenging traditional findings.

In the model extension, we explore the influence of labor mobility and tax competition. Local governments lower tax rates to compete for tax revenue, while the central income tax serves as a Pigouvian tax to internalize the horizontal externalities.[7] This finding justifies the central income tax even when governments share the same preference. Solving the optimal central income tax in an economy with labor mobility and asymmetric jurisdictions is quite complex. While we employ the variational approach to address the problem in our main analysis, we also demonstrate the validity of the primal approach.

Our research demonstrates how to use a small number of sufficient statistics to design contracts in the context of multilayer principal-agent problem. First, we demonstrate the existence of general equilibrium elasticities (GEEs) that determine the reaction of labor income to the central income tax. These GEEs are elasticities of labor income with respect to the central income tax, taking the reaction of local income taxes into consideration. Second, we show how can these GEEs be determined by a small number of partial equilibrium elasticities (PEEs), i.e., labor income, population, an elasticities of labor income and populations. Interestingly, under common symmetry assumptions on elasticities, the *ABC-nest-ABC* tax rule still holds and our benchmark analysis is still explanatory. What's different is now the term B is modified to include the effect of fiscal externality due to the tax competition.

Positively, we did a simple quantitative analysis for the United States. Calibrating objectives of local governments to meet realistic income taxes, we find that decentralized redistribution is desirable when objective inconsistency is relatively small. Interestingly, local taxes have a minor effect on national welfare because the central income tax considerably crowds out local taxes while maintains a stable aggregate income tax. This finding illustrates why state income taxes are relatively low and flat in the U.S. From another perspective, the above finding indicates that local taxes address local needs without significantly affecting national welfare. Last, we find that the federal income tax is crucial even if inter-region transfer is perfect. In contrast, introducing a region-specific federal income tax would only mildly increase national welfare.

The study generates important policy recommendations. The basic principles of fiscal federalism as-

[7] Horizontal and vertical externalities refer to the effects between governments at the same and different levels, respectively.

sert that redistributive policies should be undertaken by the most central level of government.[8] In practice, multi-level redistribution is popular around the world; and generally, the more centralized the union is, the more centralized its redistributive policy system is.[9] Therefore, it is of general interest to ask to what extent redistribution should be decentralized; and, given the distribution of authority, how to design the central income tax. One can answer the first question only if the second one has been addressed. Our study suggests that local redistribution is generally beneficial, particularly when supported by a well-designed central income tax. Decentralized redistribution can enhance national welfare in certain situations and effectively address local issues. To prevent potential negative impacts on national welfare due to local actions, it is advisable to limit local redistribution within a suitable range. Furthermore, our findings indicate that region-specific central income taxes have a minimal impact on national welfare and, given the implementation costs, may not be an effective policy option. Last, our formula demonstrates precisely how the central government can engage local governments and achieve the optimal tax system using empirically observable data.

Literature. This paper makes two key contributions. First, it explores optimal Mirrleesian tax under tax overlap and tax competition, a novel study within the Mirrlees framework. Second, it addresses a gap in fiscal federalism research by developing a quantifiable model and a practical tool for fiscal institution design which is absent in previous work (see, e.g., Oates (2005)).[10]

To illustrate the contribution of this article in detail, we will discuss its connections to three relevant bodies of literature: (*i*) optimal taxation with multiple regions (jurisdictions), (*ii*) optimal taxation in multi-level governance contexts, and (*iii*) using empirically observable statistics to derive optimal taxation.

(i) The first body of literature is an emerging and influential field focused on redistribution in multi-region economies (see, e.g., Lehmann et al. (2014); Ales et al. (2015); Gomes et al. (2018); Eeckhout and Guner (2015); Huggett and Luo (2023)). For instance, Lehmann et al. (2014) examines optimal income taxation in a two-country model under Nash equilibrium, where each country, acting as a parallel principal, simultaneously designs its income tax while treating the other country's tax as fixed. In contrast, our model features a hierarchical structure in which one government acts as a leader while the others follow. This approach effectively captures the multi-level governance prevalent in various countries, which is often overlooked in optimal tax literature.[11]

Recently, Gaubert et al. (2021) investigated the desirability of inter-region transfers alongside region-uniform income redistribution, while Colas and Hutchinson (2021) explored the advantages of a progressive region-uniform income tax in conjunction with inter-region transfers. Both studies demonstrate that inter-region transfers and central income taxes cannot be substituted for each other. However, this literature primarily focuses on optimal taxation with parallel principals or a single principal. None of these works address optimal central redistribution in conjunction with strategic decentralized redistributions, a gap that

[8] There are also arguments for some decentralized redistribution policies. Please see the literature review for details.

[9] See footnote 1.

[10] As noted by Oates (2005), "while these models (second-generation fiscal federalism) offer a nuanced perspective on the balance between centralization and decentralization, they often neglect the specifics of fiscal institutional structures."A quantifiable fiscal federal model for policy design is essential but absent.

[11] This setup aligns with studies on a leader's optimal taxation in open economies. For example, tax reforms in large countries or unions, such as the United States and the European Union, may influence the tax policies of other nations. Consequently, the government of the country or union may design its tax policy as a leader rather than treating the taxes of other countries as given.

our paper aims to fill.

(ii) The second body of literature typically examines optimal policy within a multi-level principal-agent framework, addressing adverse selection or moral hazard issues in labor or credit markets (see, e.g., Golosov and Tsyvinski (2007); Chetty and Saez (2010); Scheuer (2013); Stantcheva (2014); Doligalski et al. (2023)). In contrast, our paper focuses on a different type of adverse selection—specifically, the interaction between local governments and their residents, which constitutes a multi-level government problem. Moreover, we account for asymmetric subprincipals (local governments) and practical policy constraints on the central principal. In this asymmetric environment, the central government must navigate inter- and intra-region redistribution while considering the strategic responses of local governments. The introduction of multi-order and asymmetric governance leads to significant changes in the dynamics at play. Our paper contributes to this literature with an approach to articulate optimal central income taxation by employing a small number of empirically observable statistics.

Research on optimal redistribution involving multi-level governments is relatively scarce. This paper is particularly related to Wildasin (1991) and Bordignon et al. (2001). Wildasin (1991) examines an economy characterized by exogenous labor supply, two types of individuals, and symmetric information between governments and residents, allowing for enforceable type-specific lump-sum taxes. Bordignon et al. (2001) highlights the role of asymmetric information among governments, yet their analysis excludes labor mobility and assumes individuals are identical within regions, thereby overlooking intra-region inequality and inconsistencies in regional objectives. Overall, Wildasin (1991), Bordignon et al. (2001), and the current paper complement each other by addressing three critical aspects of central tax design: externalities, imperfect information, and objective inconsistencies. Recently, Fajgelbaum and Gaubert (2020) studied optimal central policy in a general framework that incorporates local governments' reactions, focusing on spillovers from concentration and spatial sorting. While they analyze parameterized taxes in a model without endogenous labor supply, we investigate non-parametric income tax in the Mirrleesian tradition.

A core issue in fiscal federal literature is whether redistribution policy should be decentralized. A prevailing view suggests that due to adverse selection and pecuniary externalities, it is preferable for the central government to administer redistribution policies (see, e.g., Stigler (1957); Musgrave (1971); Oates (1972); Brown and Oates (1987); Kopczuk et al. (2005)). Conversely, arguments for decentralized redistribution policies highlight that local governments possess more accurate regional information and can implement targeted measures (see, e.g., Boadway and Wildasin (1984); Raff and Wilson (1997); Oates (1999)). Our study presents a structural fiscal federal model that facilitates the design of taxes tailored to different economies and evaluates the welfare implications of central versus decentralized redistribution quantitatively.

(iii) Lastly, this paper contributes to the literature on statistics-based optimal taxation (see, e.g., Piketty (1997); Diamond (1998); Saez (2001)) by examining a novel economic framework: a multi-level and asymmetric principal-agent model in which the leader is constrained by policy. Utilizing the variational approach, we derive tax incidences and present intuitive illustrations of optimal taxation. The variational approach has long been employed to solve for optimal nonlinear taxes (see, e.g., Mirrlees (1971); Piketty (1997); Christiansen (1984); Saez (2001)). Golosov et al. (2014) advocate for this method as a formal means of addressing optimal tax systems. Currently, the approach is widely recognized in the literature (see, e.g., Saez and

Stantcheva (2018); Park and Chang (2018); Sachs et al. (2020)). However, to the best of our knowledge, it has not yet been applied to a multi-level principal-agent problem with asymmetric contract competition and policy constraints. The statistics-based formula derived herein enhances the existing literature by illustrating how objective inconsistency as well as inter-region inequality affects optimal taxation. We also present the primal approach as an alternative method. The techniques developed in this paper may be extended to analyze other multi-level principal-agent scenarios. For instance, the optimal income tax with screening in the labor market (see, e.g., Stantcheva (2014)).

The structure of the remainder of this paper is organized as follows. We begin by outlining the model setup and the problems of taxpayers and local governments in Section 2. In Section 3, we derive solutions for taxpayers and local governments, and formally define the equilibrium. We then derive the tax incidence of central income tax within the equilibrium in Section 4 and present the optimal central income tax in Section 5. We introduce labor mobility in Section 6. Section 7 presents the numerical analysis. Section 8 concludes the paper. Appendixes A to E contain the proofs. In Appendix C, we address various issues, including tax base and optimal transfers, and present the primal approach.

2 Model

We consider a static two-region economy inhabited by a continuum of heterogeneous residents. Each region is governed by a distinct local government and a singular central authority. Our baseline model posits that individuals are immobile. In Section 6, we incorporate labor mobility to examine the effects of tax competition on optimal tax policies. While we focus on a two-region economy for clarity, the model can easily accommodate any finite number of regions.

Environment. Consider a country composed of two regions, denoted as $j \in \mathcal{J} \equiv \{1, 2\}$. The population consists of a continuum of residents indexed by skill level $\theta \in \Theta \equiv (0, \infty)$. We refer to a θ-skilled resident working in region j as a (θ, j) individual, using θ_j to specify the skill of these individuals. We normalize the total population of the economy to one, and the economy produces a homogeneous output, which serves as the numeraire.

Skill levels are distributed according to a cumulative distribution function (CDF) $F_\theta : \Theta \mapsto \mathbb{R}_+$. Let $f_\theta : \Theta \mapsto \mathbb{R}_+$ denote the probability density function (PDF) of θ. We denote $f_{\theta_j} : \Theta \mapsto \mathbb{R}_+$, $F_{\theta_j} : \Theta \mapsto \mathbb{R}_+$, and $N_j \in (0, 1)$ as the PDF, CDF, and total population of θ-skilled residents in region j, respectively. We have $f_\theta(\theta) = \sum_j N_j f_{\theta_j}(\theta)$ and $F_\theta(\theta) = \sum_j N_j F_{\theta_j}(\theta)$. Let $\underline{\theta}_j$ and $\overline{\theta}_j$ represent the lowest and highest skill levels in region j, respectively, with $\Theta_j \equiv (\underline{\theta}_j, \overline{\theta}_j)$.

As technical assumptions, we posit that: (*i*) $f_{\theta_j}(\cdot)$ is continuously differentiable on Θ; and (*ii*) $f_{\theta_j}(\cdot)$ is strictly positive on Θ_j and zero on Θ/Θ_j, for all $j \in \mathcal{J}$. Consequently, $f_\theta(\theta)$ is also continuously differentiable and positive on Θ.

Preference and Technology. Residents exhibit homogeneous preferences for consumption and leisure, characterized by quasi-linear consumption preferences and iso-elastic effort costs. These simplifying assumptions are common in the literature, as they facilitate the derivation of closed-form optimal tax formulas for a single government.[12] Let c_j and l_j denote the consumption and labor supply of a resident in region j,

[12] See Salanié (2003), chapter 4, p. 86.

respectively. The utility of an individual is expressed as $u(c_j, l_j) = c_j - \phi(l_j)$, where $\phi(\cdot)$ is a function that is twice continuously differentiable, strictly convex, strictly monotonically increasing, and satisfies the Inada conditions. The elasticity of effort costs, defined as $\varepsilon_j \equiv \frac{\phi'(l_j)}{l_j \phi''(l_j)}$, is assumed to be constant.[13] We denote the indirect utility of a (θ, j) agent as $V_j(\theta)$.

Production occurs via a linear production function based on labor inputs. The productivity of residents is determined by their skill levels and a region-specific productivity parameter $A_j \in (0, \infty)$. The income (output) of a (θ, j) resident is given by $y_j = A_j \theta l_j$.

Policy and Authority. In each jurisdiction, there is one local government. The j-region government redistribute income in region j by a local income tax $T_j : \mathbb{R}_+ \mapsto \mathbb{R}$ on y_j. The tax is levied to make redistribution and collect a transfer required by the central government. Denote $E_j \in \mathbb{R}$ as the exogenous per capita transfer from j-region government to the central government. Later in Section C.2 and C.3, we will consider endogenous inter-government transfers.

The central government, seeking to maximize the national welfare, redistributes income within the country by a central income tax $T : \mathbb{R}_+ \mapsto \mathbb{R}$ on y_j, $j \in \mathcal{J}$. The policy system $\mathcal{T} \equiv \left\{T, T_j; E_j\right\}_{j \in \mathcal{J}}$ thus is a combination of central policy system $\mathcal{T}^C \equiv \left\{T; E_j\right\}_{j \in \mathcal{J}}$ and regional policy system $\mathcal{T}^R = \left\{T_j\right\}_{j \in \mathcal{J}}$.

Both central and local income taxes are arbitrarily nonlinear in the tradition of Mirrlees (1971). However, the central income tax is restricted to be region-uniform. For any government, all the remaining fiscal revenue after the inter-government transfers are used as a lumpsum subsidy (tax, if negative) for the residents in their jurisdictions. For later use, we define the aggregate local income tax on y_j as a function $T_j^A : \mathbb{R}_+ \mapsto \mathbb{R}$ satisfying:

$$T_j^A\left(y_j\right) = T\left(y_j\right) + T_j\left(y_j\right). \tag{1}$$

Throughout this paper, we consider twice continuously differentiable tax functions, i.e., $T, T_j \in \mathcal{C}^2\left(\mathbb{R}_+, \mathbb{R}\right)$ and $j \in \mathcal{J}$.

Governments' Objectives. To characterize all Pareto constrained-efficient allocations, we employ a continuous probability density function (PDF) $\widetilde{f}_{\theta_j} : \Theta \mapsto \mathbb{R}_+$ to represent the Pareto weight schedule of the local government for (θ, j) individuals. The corresponding cumulative distribution function (CDF) is denoted as $\widetilde{F}_{\theta_j}(\theta) \equiv \int_0^\theta \widetilde{f}_{\theta_j}(x)\, dx$, representing the cumulative weight for residents in region j with skill levels lower than θ.[14] We also use continuous functions $\widetilde{f}_{\theta_j}^C : \Theta_j \mapsto \mathbb{R}_+$ to govern the central government's Pareto weights for (θ, j) individuals, ensuring that $\sum_j \left[N_j \int_\Theta \widetilde{f}_{\theta_j}^C(\theta)\, d\theta\right] = 1$. We require both $\widetilde{f}_{\theta_j}(\cdot)$ and $\widetilde{f}_{\theta_j}^C(\cdot)$ to be positive on Θ_j and zero on Θ / Θ_j.

Let $V_j(\theta)$ denote the individual utility of a (θ, j) agent. The objectives of the j-region government and the central government are expressed as:

$$W_j = \int_{\Theta_j} G(V_j(\theta)) \widetilde{f}_{\theta_j}(\theta)\, d\theta \quad \text{and} \quad W = \sum_j \left[N_j \int_{\Theta_j} G(V_j(\theta)) \widetilde{f}_{\theta_j}^C(\theta)\, d\theta\right].$$

[13] We employ a separable utility function to abstract from the effects of complementarity or substitutability between consumption and leisure on the incentive conditions. Readers interested in these effects can refer to Atkinson and Stiglitz (1976), Mirrlees (1976), and Christiansen (1984).

To emphasize the role of the authority system in tax design, we abstract from general equilibrium effects such as production spillovers. For those interested in the impact of spillovers, see Fajgelbaum and Gaubert (2020) for an example.

[14] See Saez and Stantcheva (2016) for generalized social welfare weights and Pareto weights.

In our benchmark model, we consider a linear social welfare function, $G(x) = x$, allowing $\widetilde{f}_j^C(\cdot)$ to reflect the governments' preferences for equality. Later in Section C, we will explore the implications of a concave welfare function.

The Pareto weight schedules of central and local governments can differ significantly. First, different governments may have varying preferences. Second, even if governments share the same preference for equality, their Pareto weights for a θ-skilled individual may diverge, as a high-skilled individual in region j could be classified as middle- or even low-skilled at the national level. Thus, although the total national social welfare weight sums to one, i.e., $\sum_j \left[N_j \int_{\Theta_j} \widetilde{f}_{\theta_j}^C(\theta)\, d\theta \right] = 1$, the local total social welfare weight $\int_{\Theta_j} \widetilde{f}_{\theta_j}^C(\theta)\, d\theta$ may not equal one. We will later introduce inter-region and intra-region social welfare weights to represent the government's preferences for inter- and intra-region equality.

Timeline. The economy is divided into three stages. At the outset, the central government enacts the central income tax. Subsequently, local governments implement regional income taxes simultaneously, treating the central income tax as a given. Finally, individuals choose their labor income to maximize their utilities, given the imposed taxes.

Formally, an individual of type (θ, j) maximizes their utility under the budget constraint, observing all tax policies:

$$\begin{aligned} V_j(\theta) = \max_{y_j} c_j - \phi\left(\frac{y_j}{A_j\theta}\right) \\ \text{s.t. } c_j \leq y_j - T_j^A\left(y_j\right). \end{aligned} \tag{P-1}$$

Treating the central income tax as given and anticipating the responses of residents, the government of region j designs local income taxes to maximize local social welfare, $j \in \mathcal{J}$:

$$\begin{aligned} \max_{T_j} \int_{\Theta_j} G\left(V_j(\theta)\right) \widetilde{f}_{\theta_j}(\theta)\, d\theta \\ \text{s.t. } E_j \leq \int_{\Theta_j} T_j\left(A_j\theta l_j(\theta)\right) f_{\theta_j}(\theta) d\theta \\ y_j(\theta) \in \arg\max_{y_j} \{\text{P-1}\}, \forall \theta \in \Theta. \end{aligned} \tag{P-2}$$

Considering the strategic reactions of local governments and individuals, the central government's problem involves choosing the central income tax $T(y)$ to maximize social welfare, subject to budget con-

straints:[15]

$$\max_{T} \sum_{j} N_j \int_{\Theta_j} G\left(V_j(\theta)\right) \widetilde{f}^{C}_{\theta_j}(\theta)\, d\theta \tag{P-3}$$

$$\text{s.t. } 0 \leq \sum_{j} N_j \left[\int_{\Theta_j} T\left(A_j \theta l_j(\theta)\right) f_{\theta_j}(\theta)\, d\theta_j\right] + \sum_{j} N_j E_j,$$

$$T_j \in \arg\max_{T_j} \{\text{P-2}\}, \forall j \in \mathcal{J},$$

$$y_j(\theta) \in \arg\max_{y_j} \{\text{P-1}\}, \forall \theta \in \Theta, j \in \mathcal{J}.$$

Equilibrium. We provide a formal definition of equilibrium after solving for the problems of all agents. Initially, we present an informal definition: given the central policies, the equilibrium allocation and price system ensure that the resulting allocation maximizes individuals' utilities while adhering to their budget constraints, and it also maximizes the objectives of local governments under their respective budget constraints. Furthermore, all markets must clear under the established price system, and the central government's budget constraint must be satisfied. This condition, when considered alongside the budget constraints of individuals and local governments, implies that the social resource constraint is also fulfilled.

The central government aims to select a central income tax that yields a constrained optimal equilibrium. To facilitate analysis, we assume that the first-order conditions for both individuals' and local governments' problems are both necessary and sufficient and there exist an unique solutions for all agents' problems.[16] Finally, we impose the common monotonicity assumptions:[17]

Assumption 1 *$y_j(\theta)$ and $c_j(\theta)$ are continuously differentiable, strictly positive, and strictly increasing in θ, for all $\theta \in \Theta_j$ and $j \in \mathcal{J}$.*

Under Assumption 1, we construct the CDF of labor income by $F_j\left(y_j(\theta)\right) = F_{\theta_j}(\theta)$ and PDF of labor income by $f_j\left(y_j\right) = F_j'\left(y_j\right)$. For later use, we also set income-based Pareto weights by $\widetilde{F}_j\left(y_j(\theta)\right) = \widetilde{F}_{\theta_j}(\theta)$ and $\widetilde{f}_j\left(y_j\right) = \widetilde{F}_j'\left(y_j\right)$; $\widetilde{F}^{C}_j\left(y_j(\theta)\right) = \widetilde{F}^{C}_{\theta_j}(\theta)$ and $\widetilde{f}^{C}_j\left(y_j\right) = \widetilde{F}^{C\prime}_j\left(y_j\right)$. We assume that $\widetilde{F}_{\theta_j}(\theta) \geq F_{\theta_j}(\theta)$ and $\widetilde{F}^{C}_{\theta_j}(\theta) \geq F_{\theta_j}(\theta)$, which means that governments have a weak preference for equality.

3 Solution

In this section, we solve for the MGTE. Note that local governments' solutions are parts of the equilibrium.

3.1 Individual's Solution

To solve the individual's problem P-1, we derive the following first-order conditions:

$$1 - T_j^{A\prime}(y_j(\theta)) = \frac{1}{A_j\theta}\phi'\left(\frac{y_j(\theta)}{A_j\theta}\right), \quad \forall \theta \in \Theta, j \in \mathcal{J}. \tag{2}$$

We denote by $y_j(\theta)$ the argmax of the above problem, and let $c_j(\theta; T_j^A)$ and $V_j(\theta; T_j^A)$ represent the resulting consumption and indirect utility, respectively. Following the literature (see, e.g., Jacquet et al. (2013)), we

[15] Recall that we consider $G(V_j) = V_j$ in our benchmark model.

[16] Additionally, we assume that the tax policies are twice continuously differentiable. When the central tax is restricted to be twice continuously differentiable, the local tax response to the central tax is also twice continuously differentiable.

[17] The Spence-Mirrlees condition implies that labor income and consumption be non-decreasing in wages. For simplicity, we assume that $y_j(\theta)$ and $c_j(\theta)$ increase strictly with θ, thereby excluding cases with mass points.

assume that the first-order conditions of individuals' problems are both necessary and sufficient, which holds true provided that the aggregate regional income taxes are not excessively concave.

As established in the literature on tax incidence (see, e.g., Saez (2001) and Golosov et al. (2014)), a first-order Taylor expansion of both sides of (2) around the baseline yields the change of labor income of θ_j-skiled individual ($\widehat{y}_j(\theta_j)$) in terms of the change of marginal tax rates ($\widehat{T}_j^{A\prime}(y_j(\theta_j))$). We formally define the tax reform and following changes of individual action in Appendix A.1, where we show:

Lemma 1 *For any $j \in \mathcal{J}$ and $\theta_j \in \Theta_j$,*

$$\frac{\widehat{y}_j(\theta_j)}{y_j(\theta_j)} = -\varepsilon_j^A(y_j(\theta_j))\frac{\widehat{T}_j^{A\prime}(y_j(\theta_j))}{1-T_j^{A\prime}(y_j(\theta_j))}, \tag{3}$$

where

$$\varepsilon_j^A(y_j(\theta)) = \frac{1}{\frac{1}{\varepsilon_j} + p_j^A(y_j(\theta))} \tag{4}$$

is the elasticity of labor income, and $p_j^A(y_j) = -\frac{d\ln[1-T^{A\prime}(y_j)]}{d\ln y_j}$ represents the progressivity of the aggregate income tax, with $\theta \in \Theta_j$, $j \in \mathcal{J}$.

Proof. See Appendix A.2. ■

The elasticity ε_j^A incorporates all micro factors determining the responsiveness of labor income to the net aggregate tax retention rate ($1 - T_j^{A\prime}$): the concavity of the utility function and the progressivity of the aggregate income tax. According to the taxpayer's problem, regardless of how local taxes respond to the central income tax, equation (3) must hold. It turns out that the elasticity ε_j^A, reflecting the responsiveness of labor income to $1 - T_j^{A\prime}$, also governs the response of labor income to the local income tax when given the central income tax.

3.2 Local Government's Solution

The local government's problem P-2 is analogous to the planner's problem in the canonical Mirrleesian tax framework (see, e.g., Piketty (1997), Diamond (1998), and Saez (2001)), with the exception of the introduction of a central income tax. We denote the solution to the local government's problem P-2 as $T_j^\star(\cdot)$.

Local Income Tax. The derivation of the regional income taxes is presented in Appendix A.1. Below, we provide the rule for local income tax:[18]

Lemma 2 *Set $\underline{y}_j = y_j\left(\underline{\theta}_j\right)$ and $\overline{y}_j = y_j\left(\overline{\theta}_j\right)$. For any $y_j \in \left(\underline{y}_j, \overline{y}_j\right)$, optimal local income tax satisfies:*

$$\frac{T_j^{\star\prime}\left(y_j\right)}{1-T'\left(y_j\right)-T_j^{\star\prime}\left(y_j\right)} = \frac{\widetilde{F}_j\left(y_j\right)-F_j\left(y_j\right)}{1-F_j\left(y_j\right)}\frac{1-F_j\left(y_j\right)}{y_j f_j\left(y_j\right)}\frac{1}{\varepsilon_j^A\left(y_j\right)}. \tag{5}$$

Proof. See Appendix A.3. ■

Formula (5) illustrates how local governments strategically respond to the central income tax. The central income tax is incorporated into the ABC tax formula (see, e.g., Piketty (1997), Diamond (1998), and Saez

[18] To maintain a consistent framework, we employ the variational approach to derive both regional and central income taxes. Alternatively, the mechanism design theory and primal approach could be utilized to address the governments' problems. Proposition 3, as an extension of Proposition 2, demonstrates that the optimal taxations obtained from these two approaches are consistent.

(2001)). Rearranging (5) provides an intuitive illustration of optimal taxation:

$$0 = \overbrace{\frac{\varepsilon_j^A(y_j)}{1-T'(y_j)-T_j^{\star\prime}(y_j)} T_j^{\star\prime}(y_j) y_j f_j(y_j)}^{\text{Local revenue effect}} \overbrace{- \frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\left[1-F_j(y_j)\right]}^{\text{Redistribution effect}}. \tag{6}$$

The local revenue effect and the redistribution effect correspond to the "elasticity effect" and "mechanical effect net of welfare loss" respectively, as described in Saez (2001). Notably, the percentage change in the retention rate is now represented by $\frac{1}{1-T'(y_j)-T_j^{\star\prime}(y_j)}$ instead of $\frac{1}{1-T_j^{\star\prime}(y_j)}$.[19]

The percentage change in the after-tax retention rate induced by any regional tax reform becomes more pronounced with the introduction of a central income tax, indicating that the local revenue effect is more sensitive to regional tax reforms. Specifically, the regional revenue losses resulting from individuals' responses to an increase in regional tax rates become larger, while the gains from individuals' reactions to a decrease in local tax rates diminish. This shift, in turn, leads to a decrease in the optimal local income tax. More clearly, we can observe the following transformation of (5):

$$T_j^{\star\prime}\left(y_j\right) = \frac{\frac{\widetilde{F}_j\left(y_j\right)-F_j\left(y_j\right)}{1-F_j\left(y_j\right)} \frac{1-F_j\left(y_j\right)}{y_j f_j\left(y_j\right)} \frac{1}{\varepsilon_j^A\left(y_j\right)}}{1+\frac{\widetilde{F}_j\left(y_j\right)-F_j\left(y_j\right)}{1-F_j\left(y_j\right)} \frac{1-F_j\left(y_j\right)}{y_j f_j\left(y_j\right)} \frac{1}{\varepsilon_j^A\left(y_j\right)}} \left[1-T'\left(y_j\right)\right]. \tag{7}$$

Note that the optimal response of local government indicates that the increase in the local tax rate resulting from an additional unit increase in the central tax rate is less than 1. This is evident from the multiplier of $1-T'(y_j)$ on the right side of (7). To elucidate this point, we present the result in terms of parameters in Corollary 1. Before doing so, we will introduce the following useful concepts.

Mirrleesian Tax. In the context of Mirrleesian tax literature, it is well known that in the absence of a central income tax, the optimal local income tax rate on $y_j(\theta)$ is $\tau_j^\star(\theta)$; conversely, without a local income tax, the optimal region-specific central income tax rate on $y_j(\theta)$ is represented as $\tau_j^{C,\star}(\theta)$:

$$\tau_j^\star\left(\theta\right) \equiv \frac{\frac{\widetilde{F}_{\theta_j}(\theta)-F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta}\left(1+\frac{1}{\varepsilon_j}\right)}{1+\frac{\widetilde{F}_{\theta_j}(\theta)-F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta}\left(1+\frac{1}{\varepsilon_j}\right)} \quad \text{and} \quad \tau_j^{C,\star}(\theta) \equiv \frac{\frac{\widetilde{F}^C_{\theta_j}(\theta)-F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta}\left(1+\frac{1}{\varepsilon_j}\right)}{1+\frac{\widetilde{F}^C_{\theta_j}(\theta)-F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta}\left(1+\frac{1}{\varepsilon_j}\right)}. \tag{8}$$

Note that we have assumed that $\widetilde{F}_{\theta_j}\left(\theta\right) \geq F_{\theta_j}\left(\theta\right)$ and $\widetilde{F}^C_{\theta_j}\left(\theta\right) \geq F_{\theta_j}\left(\theta\right)$. This assumption guarantees $\tau_j^\star \geq 0$ and $\tau_j^{C,\star} \geq 0$. Besides, $\tau_j^\star$ and $\tau_j^{C,\star}$ are less than one. Intuitively, there is no need to enforce a tax rate higher than one (see, e.g., Mirrlees (1971)). We will return to the above canonical Mirrleesian tax frequently in the following analysis.

[19] As a simple illustration following Saez (2001), suppose the marginal local income tax rates decrease by an amount $d\tau$ for incomes between y and $y+\triangle y$, where $d\tau$ is assumed to be second order compared to $\triangle y$. This tax reform increases the labor supply of each targeted taxpayer (whose income falls within $(y, y+\triangle y)$) by $\frac{d\tau}{1-T'(y_j(\theta))-T_j^{\star\prime}(y_j(\theta))}\varepsilon_j^A(y_j(\theta))$, which when multiplied by $y_j(\theta)f_{y_j}(y_j(\theta))T_j^{\star\prime}(y_j(\theta))\triangle y$, yields the increased tax revenue. Meanwhile, the tax burden for all taxpayers with income exceeding $y+\triangle y$ is reduced by $\triangle y d\tau$. The redistribution benefit of such a tax reform can be quantified as $-\frac{\widetilde{F}_{y_j}(y_j(\theta))-F_{y_j}(y_j(\theta))}{1-F_{y_j}(y_j(\theta))}\left[1-F_{y_j}(y_j(\theta))\right]\triangle y d\tau$ in terms of government revenue.

Crowd-out and Crowd-in Effects. With the notations established, we first find that:

Proposition 1 *For any $\theta \in \Theta_j$, and $j \in \mathcal{J}$, (i) the reaction of the local income tax to the central income tax satisfies*

$$T_j^{\star\prime}(y_j(\theta)) = \tau_j^\star(\theta)\left[1 - T'(y_j(\theta))\right]; \tag{9}$$

(ii) the reaction of the aggregate labor income tax to the central income tax satisfies

$$1 - T_j^{A\prime}(y_j(\theta)) = \left[1 - \tau_j^\star(\theta)\right]\left[1 - T'(y_j(\theta))\right]. \tag{10}$$

Proof. See Appendix A.3.[20] ■

Proposition 1 show that the marginal tax rate of local income tax for a specific agent decreases with the increase of marginal central income tax rate faced by the agent, while the marginal tax rate of aggregate income tax increases with the increase of marginal central income tax rate.

Remarkably, the canonical Mirrleesian tax rate $\tau_j^\star(\theta)$ captures the sensitivity of the reactions of local and aggregate income tax rates to the central income tax rates. Two important aspects are worth noting here: the crowd-out effect of the central income tax on local income tax and the crowd-in effect of the central income tax on aggregate income tax. First, an increase in $T'(y_j(\theta))$ leads to a decrease in $T_j^{\star\prime}(y_j(\theta))$. Second, although $T_j^{\star\prime}(y_j(\theta))$ decreases with respect to $T'(y_j(\theta))$, $T_j^{A\prime}(y_j(\theta))$ ultimately increases with respect to $T'(y_j(\theta))$ (see equation (10)), because a unit increase in $T'(y_j(\theta))$ induces a decrease of less than one unit in $T_j^{\star\prime}(y_j(\theta))$.[21] These two effects illustrate the significant impact of the central income tax on local and aggregate income taxes.

Comparison to the Canonical Mirrleesian Tax. To understand the aforementioned findings, we compare the local income tax to the canonical Mirrleesian tax. To facilitate further analysis, consider the case where $T'\left(y_j\left(\theta\right)\right) \geq 0$, which holds when the central government has a greater obligation to redistribute than regional governments. Proposition 1 directly implies the following results:[22]

1. When $T'\left(y_j\left(\theta\right)\right) \geq 0$, we have $T_j^{\star\prime}\left(y_j\left(\theta\right)\right) \leq \tau_j^\star\left(\theta\right)$ and $T_j^{A\prime}(y_j\left(\theta\right)) \geq \tau_j^\star\left(\theta\right)$;
2. When $T'\left(y_j\left(\theta\right)\right) \leq 0$, we have $T_j^{\star\prime}\left(y_j\left(\theta\right)\right) \geq \tau_j^\star\left(\theta\right)$ and $T_j^{A\prime}(y_j\left(\theta\right)) \leq \tau_j^\star\left(\theta\right)$.

The above results indicate that local governments will choose a lower local tax rate and a higher aggregate tax rate than it would in the absence of the central income tax. The reasoning is that, with a central income tax, a portion of tax revenues is allocated to the central government. Consequently, local governments undervalue the revenue benefits associated with reducing local income taxes. To clarify, the total revenue effect $\frac{\varepsilon_j^A\left(y_j\right)}{1-T'\left(y_j\right)-T_j^{\star\prime}\left(y_j\right)}\left[T_j^{\star\prime}\left(y_j\right)+T'\left(y_j\right)\right]y_j f_j\left(y_j\right)$ is larger than the local revenue term in equation (6), i.e., $\frac{\varepsilon_j^A(y_j)}{1-T'(y_j)-T_j^{\star\prime}(y_j)}T_j^{\star\prime}(y_j)y_j f_j(y_j)$. As a result, the local income tax rate chosen by the local government is

[20] The proof for Proposition 1 is included in the proof for Lemma 2.

[21] To see this, transform equation (10) to $T_j^{A\prime}(y_j(\theta)) = T'(y_j(\theta))\left[1-\tau_j^\star(\theta)\right] + \tau_j^\star(\theta)$. The term $1-\tau_j^\star(\theta) \leq 1$ captures the sensitivity of the aggregate income tax rate to the central income tax rate.

[22] Notice that $0 \leq \tau_j^\star\left(\theta\right) < 1$. Equation (10), combined with $T'\left(y_j\left(\theta\right)\right) \geq 0$, implies that $T_j^{A\prime}(y_j\left(\theta\right)) \geq \tau_j^\star\left(\theta\right)$. While equation (10) implies that $T_j^{A\prime}(y_j\left(\theta\right)) \geq \tau_j^\star\left(\theta\right)$, equation (9) indicates that $T_j^{\star\prime}\left(y_j\left(\theta\right)\right) \leq \tau_j^\star\left(\theta\right)$.

higher than a scenario without the central income tax. From another perspective, local governments overlook the greater total revenue when reducing the aggregate income tax rate, resulting in a higher aggregate income tax rate.

An interesting question arises: What prevents a local government from implementing its preferred income tax schedule? Given the availability of nonlinear income tax options, the local government could theoretically override any federal tax policy and substitute it with its preferred alternative. However, when weighing the revenue and redistribution effects induced by the local income tax, it is optimal for local governments to deviate from $\tau_j^\star(\theta)$. Consequently, the central government can influence the aggregate income tax by inserting "wedges" into the local government's decision-making process. Proposition 1 show precisely how can the central government adjust the total tax rate upwards and downwards.

Relation to Empirical Findings. There are plausible empirical findings on the strategic response of local governments to the central policy (see, e.g., Feldstein and Metcalf (1987), Scott and Triest (1993), Goodspeed (2000), Parchet (2019)). Our theoretical findings echo empirical findings on the strategic response of the local income tax to the central one. Goodspeed (2000) find that higher national income tax rates lead to lower local income tax rates. Scott and Triest (1993) consider a federal income tax reform, which would cause an increase in the progressivity of effective state tax. They find a decrease in the progressivity of statutory state income tax which offset the influence of federal tax reform. However, when state tax payments net of the federal subsidy are analyzed, state tax progressivity increases.[23]

3.3 Equilibrium

Throughout this paper, we will consider the following *Multi-level Governance Tax Equilibrium* (MGTE), where we refer to the allocation set $\mathcal{A} \equiv \left\{l_j(\theta_j), c_j(\theta_j)\right\}_{\theta_j \in \Theta, j \in \mathcal{J}}$ as a combination of consumption schedules $c_j : \Theta \mapsto \mathbb{R}_+$, and labor supply schedules $l_j : \Theta \mapsto \mathbb{R}_+$. In a MGTE, the local governments' behaviors have to be determined as a part of the equilibrium.

Definition 1 *Given a central policy system $\mathcal{T}^C$, a Multi-level Governance Tax Equilibrium (MGTE) is a combination of the central policy system $\mathcal{T}^C$ and the resulting allocation $\mathcal{A}$ and regional policy system $\mathcal{T}^R$, such that given the central government's policies $\mathcal{T}^C$, the local governments' policies $\mathcal{T}^R$ solve local governments' problem P-2 subject to the local budget constraint and individuals' rational reaction, and the allocation $\mathcal{A}$ solve residents' problem P-1 subject to the budget constraint under the policy system $\mathcal{T}$; and the final goods market is cleared:*

$$\sum_{j \in \mathcal{J}} Q_j = \sum_{j \in \mathcal{J}} \left[N_j \int_\Theta c_j(\theta) f_j(\theta) d\theta \right]. \tag{11}$$

By Walras's law, the central government's budget constraint is satisfied when the commodity market-clearing condition (resource constraint) and the local governments' budget constraints are satisfied. As social welfare increases in the agents' utilities, the budget constraints must be binding in the equilibrium. Since we have assumed that the first-order conditions are both necessary and sufficient, equations (10),

[23] Feldstein and Metcalf (1987) examines how federal deductibility of state and local taxes influences state and local fiscal behavior. They find that in states where the federal deductibility makes personal taxes cheaper for taxpayers, these states tend to rely more on these deductible personal taxes. Goodspeed (2000) use a sample of 13 OECD countries for the period 1975–1984. Scott and Triest (1993) examines the extent to which states adjusted the progressivity of state tax in response to federal tax changes during the 1980s of U.S.

(2) together with individuals' and local governments' budget constraints and commodity market-clearing conditions (11) determine the MGTE allocation.

4 Model Analysis: Tax Incidence

In this section, we derive the elasticity of labor income to the central income tax. We then combine the elasticities of labor income to the central and aggregate income tax to derive the reaction of local income tax to the central income tax.

Tax Perturbation. Consider a central tax perturbation $\widehat{T} \in \mathcal{C}^2(\mathbb{R}_+, \mathbb{R})$ so that the perturbed central income tax is $T + \mu\widehat{T}$, where $\mu \in \mathbb{R}_+$. Denote by $y_j(\theta; T)$ and $y_j\left(\theta; T + \mu\widehat{T}\right)$ the labor income under the initial tax and the tax after reform. The marginal response of labor income to such a perturbation at $T(\cdot)$ in the direction $\widehat{T}(\cdot)$ is then given by: $\widehat{y}_j\left(\theta; T, \widehat{T}\right) \equiv \lim_{\mu\to 0} \frac{y_j\left(\theta; T+\mu\widehat{T}\right) - y_j(\theta;T)}{\mu}$. Note that the change in the central income tax will alter the local income tax and the change of labor income takes this indirect effect into consideration.

Denote by $T_j^\star(y; T)$ and $T_j^\star\left(y; T + \mu\widehat{T}\right)$ the local income tax under the initial central income tax and after the tax reform. We assume that for any $y \in \mathbb{R}_+$, $T_j^\star(y; \cdot)$ and $T_j^{\star\prime}(y; \cdot)$ change continuously in response to infinitesimal perturbations of the initial central income tax.[24] We then derive the change in regional governments' behaviors following this perturbation as $\mu \to 0$. To do this, we first define the reform of local income tax and marginal tax rate of local income tax for a specific income level $y \in \mathbb{R}_+$ of region $j \in \mathcal{J}$ as:[25]

$$\widehat{T}_j^\star\left(y; T, \widehat{T}\right) \equiv \lim_{\mu\to 0} \frac{T_j^\star\left(y; T + \mu\widehat{T}\right) - T_j^\star(y; T)}{\mu} \quad \text{and} \quad \widehat{T}_j^{\star\prime}\left(y; T, \widehat{T}\right) \equiv \lim_{\mu\to 0} \frac{T_j^{\star\prime}\left(y; T + \mu\widehat{T}\right) - T_j^{\star\prime}(y; T)}{\mu}. \tag{12}$$

In particular, the change of lumpsum part is $\widehat{T}_j^\star\left(\underline{y}_j; T, \widehat{T}\right) = \lim_{\mu\to 0} \frac{T_j^\star\left(\underline{y}_j; T+\mu\widehat{T}\right) - T_j^\star\left(\underline{y}_j; T\right)}{\mu}$, where $\underline{y}_j = 0$ because $\underline{\theta}_j = 0$. For later use, denote

$$\widehat{T}_j^A\left(y; T, \widehat{T}\right) = \widehat{T} + \widehat{T}_j^\star\left(y; T, \widehat{T}\right) \tag{13}$$

as the reform of the aggregate income tax after the reform of the central income tax.

Reaction of Labor Income. Taking the first-order Taylor expansions of the equilibrium conditions (see Appendix A.4) at the baseline tax system (i.e., $\mu \to 0$) provides differential equations describing the first-order changes in allocations and tax policies. With these differential equations, we can derive the marginal changes in labor income:

Lemma 3 *Solving the local government's problem gives:*

$$\frac{\widehat{y}_j(\theta_j)}{y_j(\theta_j)} = -\varepsilon_j^C\left(y_j(\theta)\right) \frac{\widehat{T}'\left(y_j(\theta_j)\right)}{1 - T'\left(y_j(\theta_j)\right)}, \tag{14}$$

[24] Formally, we assume that for any $y \in \mathbb{R}_+$, initial central income tax $T(\cdot)$ and central income tax reform $\widehat{T}(\cdot)$, $T_j^\star(y; \cdot)$ and $T_j^{\star\prime}(y; \cdot)$ are locally Lipschitz continuous. That is, there exist $\overline{\mu} > 0$ and $M > 0$ such that $0 < \mu < \overline{\mu}$ implies $\left|T_j^\star(y; T + \mu\widehat{T}) - T_j^\star(y; T)\right| \leq M \times \mu$ and $\left|T_j^{\star\prime}(y; T + \mu\widehat{T}) - T_j^{\star\prime}(y; T)\right| \leq M \times \mu$.

[25] In particular, the change of lump-sum term as $\widehat{T}_k\left(\underline{y}_k\right) = \lim_{\mu\to 0} \frac{T_k\left(\underline{y}_k; T+\mu\widehat{T}\right) - T_k\left(\underline{y}_k; T\right)}{\mu}$.

where

$$\varepsilon_j^C\left(y_j(\theta)\right)=\frac{1}{\frac{1}{\varepsilon_j}+p^C\left(y_j(\theta)\right)} \tag{15}$$

is another elasticity of labor income and $p^C\left(y_j\right)=-\frac{d\ln\left[1-T'\left(y_j\right)\right]}{d\ln y_j}$ *is the progressivity of central income taxes,* $\theta\in\Theta_j, j\in\mathcal{J}$.

Proof. See Appendix A.4. ■

Equation (14) reveals a key insight: in a multi-level governance structure, the sensitivity of labor income to changes in the net-central-tax retention rate is given by $\varepsilon_j^C\left(y_j(\theta)\right)=\frac{1}{\frac{1}{\varepsilon_j}+p^C\left(y_j(\theta)\right)}$. This expression incorporates the induced changes in local income tax resulting from the central tax reform. For comparison, in a centralized setting, the corresponding elasticity is $\varepsilon_j^A\left(y_j(\theta)\right)=\frac{1}{\frac{1}{\varepsilon_j}+p^A\left(y_j(\theta)\right)}$.

Labor income elasticity $\varepsilon_j^A\left(y_j(\theta)\right)$ is familiar due to the optimal tax literature. Labor income elasticity $\varepsilon_j^C\left(y_j(\theta)\right)$ takes into the reaction of local income tax to the central income tax. Interestingly, equation (10), i.e., $1-T_j^{A\prime}(y_j(\theta))=\left[1-\tau_j^\star(\theta)\right]\left[1-T'(y_j(\theta))\right]$, suggest that the presence of local government does not alter the elasticity of labor income, i.e., $\varepsilon_j^C\left(y_j(\theta)\right)=\varepsilon_j^A\left(y_j(\theta)\right)$.[26] However, this equivalence generally do not hold when $\tau_j^\star(\theta)$, is endogenous. We will revisit this point in Section 6.

Reaction of Local Income Tax. By Lemma 1 and 3, we have

$$\frac{\widehat{y}_j\left(\theta_j\right)}{y_j\left(\theta_j\right)}=\varepsilon_j^A\left(y_j(\theta)\right)\frac{\widehat{T}\left(y_j(\theta)\right)+\widehat{T}_j^\star\left(y_j(\theta)\right)}{1-T_j^{A\prime}\left(y_j(\theta)\right)}=\varepsilon_j^C\left(y_j(\theta)\right)\frac{\widehat{T}_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j(\theta)\right)},$$

Using $\varepsilon_j^C\left(y_j(\theta)\right)=\varepsilon_j^A\left(y_j(\theta)\right)$ and equation (10), the above equation delivers the reaction of local income tax to the central income tax:

$$\widehat{T}_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)=-\widehat{T}'\left(y_j\left(\theta_j\right)\right)\tau_j^\star\left(\theta_j\right),\forall\theta\in\Theta_j, j\in\mathcal{J}. \tag{16}$$

Equation (16) shows the sensitivity of local income tax to central income tax reform is determined by the multiplier $\tau_j^\star\left(\theta_j\right)\in[0,1)$, which is a composite of income distribution, government preferences, and labor elasticity. This simplicity of the expression of $\widehat{T}_j^{\star\prime}\left(y_j\right)$ comes from the exogeneity of Pareto weight $\widetilde{f}_{\theta_j}(\theta)$. Despite its simplicity, it reflects the immediate factors that determine the behavior of local governments.

5 Optimal Centralized Redistribution

This section is attributed to solving problem P-3. As we have studied the incidence of central income tax, we can deliver the optimal central income tax formula by welfare analysis. In Appendix C.2, we also use the primal approach to derive the optimal central income, which verifies our main results below.

[26] To elaborate, recall that $p^C\left(y_j(\theta)\right)=-\frac{d\ln\left[1-T'\left(y_j\right)\right]}{d\ln y_j}$ and $p_j^A\left(y_j(\theta)\right)=-\frac{d\ln\left[1-T^{A\prime}\left(y_j\right)\right]}{d\ln y_j}$. Equation (10) implies $\frac{d\ln\left[1-T'\left(y_j\right)\right]}{d\ln y_j}=\frac{d\ln\left[1-T^{A\prime}\left(y_j\right)\right]}{d\ln y_j}$, leading to $p^C\left(y_j(\theta)\right)=p_j^A\left(y_j(\theta)\right)$. Applying the definitions of labor income elasticities then yields $\varepsilon_j^C\left(y_j(\theta)\right)=\varepsilon_j^A\left(y_j(\theta)\right)$.

5.1 Optimal Central Income Tax

Balanced Elementary Tax Reform. To derive the optimal central income tax, we consider one specific form of tax reform: the balanced elementary central income tax reform. An elementary central income tax reform is a balanced elementary central income tax reform if and only if the central government budget constraint holds after the tax reform. Formally, we consider

$$\widehat{T}(y) = 1_{\{y \geq y^*\}} + \widehat{r}(y^*), \tag{17}$$

in which $\widehat{r}(y^*)$ is a lump-sum tax guaranteeing the balance of the central government's budget. Denote the optimal central income tax as $T^\star : \mathbb{R}_+ \mapsto \mathbb{R}$. If the initial central income tax is optimal (i.e., the initial central income tax is $T^\star(\cdot)$), any tax reform $\left\{\widehat{T}^\star(\cdot; y^*)\right\}_{y^* \in \mathbb{R}_+}$ results in $\widehat{W} = 0$.

Use $\widetilde{\theta}_j(y) : \mathbb{R}_+ \mapsto \Theta$ to denote the ability of an agent who earns y in region j, and define $\theta_j^* \equiv \widetilde{\theta}_j(y^*)$. Set $\underline{y}_j = y_j(\underline{\theta}_j)$ and $\overline{y}_j = y_j(\overline{\theta}_j)$. For later use, set the lowest and highest income as $\underline{y} = \min\left(\underline{y}_j, \underline{y}_{-j}\right)$ and $\overline{y} = \max\left(\overline{y}_j, \overline{y}_{-j}\right)$, respectively. Without loss of generality, we consider the following case as an illustration:

$$\min\left\{y(\underline{\theta}_j)\right\}_{j\in\mathcal{J}} = \underline{y}_1,\ \max\left\{y_j(\underline{\theta}_j)\right\}_{j\in\mathcal{J}} = \underline{y}_2,\ \min\left\{y_j(\overline{\theta}_j)\right\}_{j\in\mathcal{J}} = \overline{y}_1, \text{and } \max\left\{y_j(\overline{\theta}_j)\right\}_{j\in\mathcal{J}} = \overline{y}_2.$$

The assumption above indicates that region 2 is richer than region 1. In particular, all individuals with income $[\underline{y}_1, \underline{y}_2)$ are from region 1, while all individuals with income $(\overline{y}_1, \overline{y}_2]$ are from region 2.

Inter- and Intra-region Welfare Weights. To capture the government's preference for inter- and intra-region equality, we define

$$\rho_j \equiv \frac{\int_{\Theta_j} G'\left(V_j(\theta)\right) \widetilde{f}_{\theta_j}^C(\theta)\, d\theta}{\sum_j N_j \left[\int_{\Theta_j} G'\left(V_j(\theta)\right) \widetilde{f}_{\theta_j}^C(\theta)\, d\theta\right]} \quad \text{and} \quad \widetilde{F}_{\theta_j}^C(\theta_j) \equiv \frac{\int_{\underline{\theta}_j}^{\theta_j} G'\left(V_j(x)\right) \widetilde{f}_{\theta_j}^C(x)\, dx}{\int_{\Theta_j} G'\left(V_j(\theta)\right) \widetilde{f}_{\theta_j}^C(\theta)\, d\theta} \tag{18}$$

as the inter-region social welfare weight for region j and intra-region social welfare weights for individuals with skill lower than θ_j. ρ_j is a normalized multiplier of fiscal expenditure. Note that $\widetilde{F}_{\theta_j}^C : \Theta \mapsto \mathbb{R}_+$ is a CDF and ρ_j satisfies $\sum_j N_j \rho_j = 1$. For later use, set $\widetilde{F}_j^C\left(y_j(\theta)\right) = \widetilde{F}_\theta^C(\theta_j)$ and $\widetilde{f}_j^C(y_j) = \widetilde{F}_j^{C\prime}(y_j)$.

Optimal Tax. In the optimum, any marginal tax reform induces a zero first-order effect on the national welfare, which delivers a necessary condition for the optimal central income tax.

Proposition 2 *(i) For any $y \in \cap_{j=1}^2 \left(\underline{y}_j, \overline{y}_j\right)$ and $j \in \mathcal{J}$, the optimal central income tax rates satisfy the following equations:*

$$T^{\star\prime}(y) = \sum_j \frac{N_j f_j(y)\, y \varepsilon_j^C(y)}{\sum_j N_j f_j(y)\, y \varepsilon_j^C(y)} \frac{\rho_j\left[\widetilde{F}_j^C(y) - \widetilde{F}_j(y)\right] + (1-\rho_j)\left[1 - F_j(y)\right]\left[\frac{\varepsilon_j^A(y)}{\varepsilon_j^C(y)} + \frac{\widetilde{F}_j(y) - F_j(y)}{y f_j(y) \varepsilon_j^C(y)}\right]}{\widetilde{F}_j(y) - F_j(y)} T_j^{\star\prime}(y), \tag{19}$$

where $\rho_j\left[\widetilde{F}_j(y) - \widetilde{F}_j(y)\right]$ and $(1-\rho_j)\left[1 - F_j(y)\right]\left[\frac{\varepsilon_j^A(y)}{\varepsilon_j^C(y)} + \frac{\widetilde{F}_j(y) - F_j(y)}{y f_j(y) \varepsilon_j^C(y)}\right]$ capture the intra-region and inter-region

objective inconsistencies, respectively. (ii) In particular, when $\rho_j = 1$ we have:

$$T^{\star\prime}(y) = \sum_j \frac{N_j f_j(y)\, y\varepsilon_j^C(y)}{\sum_j N_j f_j(y)\, y\varepsilon_j^C(y)} \frac{\widetilde{F}_j^C(y) - \widetilde{F}_j(y)}{\widetilde{F}_j(y) - F_j(y)} T_j^{\star\prime}(y). \tag{20}$$

Proof. See Appendix B.3. ■

Equations (19) and (20) show the relationship between the optimal central income tax and local income tax. When the central government is restricted to a region-uniform income tax, it can draw support from the local governments to implement a region-specific income redistribution policy, in which the central income tax is used to adjust the "objectives" of local governments, so that the resulting aggregate income tax rates are socially preferred solutions. To do this, it should enforce a positive tax rate when the local governments have lower obligations for redistribution.

In Appendix B.3, we derive the optimal central income taxation for each of the following intervals: $[\underline{y}_1, \underline{y}_2)$, $[\underline{y}_2, \overline{y}_1]$ and $(\overline{y}_1, \overline{y}_2]$. When a certain income level only appears in one region, the economic weight of that region is 1, and the weight of another region is zero. Therefore, the above rules still hold, but there is no longer a need to write in a weighted form.

The ABC-nest-ABC Formula. The combination of Lemma 2 and 2 establishes our main finding. It provides statistics-based optimal central income tax rates in terms of social welfare weights, elasticities, and income distributions:

Theorem 1 *(i) For any $y \in \left(\underline{y}, \overline{y}\right)$, we have*

$$\frac{T^{\star\prime}(y)}{1 - T^{\star\prime}(y)} = \sum_{j \in \mathcal{J}} A_j(y)\, B_j(y)\, C_j(y), \tag{21}$$

where

$$\begin{aligned}
A_j(y) &= \frac{\overbrace{N_j f_j(y)\, y\varepsilon_j^C(y)}^{\text{Economic weight}}}{\sum_j N_j f_j(y)\, y\varepsilon_j^C(y)}, \\
B_j(y) &= \frac{\overbrace{\rho_j \left[\widetilde{F}_j^C(y) - \widetilde{F}_j(y)\right] + (1-\rho_j)\left[1 - F_j(y)\right]\left[\frac{\varepsilon_j^A(y)}{\varepsilon_j^C(y)} + \frac{\widetilde{F}_j(y) - F_j(y)}{y f_j(y) \varepsilon_j^C(y)}\right]}^{\text{Objective inconsistency}}}{\widetilde{F}_j(y) - F_j(y)}, \\
C_j(y) &= \frac{\overbrace{\frac{\widetilde{F}_j(y) - F_j(y)}{1 - F_j(y)} \frac{1 - F_j(y)}{f_j(y) y} \frac{1}{\varepsilon_j^A(y)}}^{\text{Mirrleesian tax: Diamond ABC}}}{1 + \frac{\widetilde{F}_j(y) - F_j(y)}{1 - F_j(y)} \frac{1 - F_j(y)}{f_j(y) y} \frac{1}{\varepsilon_j^A(y)}}.
\end{aligned} \tag{22}$$

(ii) When $\rho_j = 1$ (inter-region objective is consistent), only intra-region objective inconsistency $\left[\widetilde{F}_j^C(y) - \widetilde{F}_j(y)\right]$ matters for the tax formula.

Proof. See Appendix B.4. ■

Theorem 1 makes a significant contribution to the literature on statistics-based optimal taxation (e.g., Piketty (1997); Diamond (1998); Saez (2001)) by examining it within the framework of a multi-level authority structure. Importantly, the formula relies on social welfare weights and commonly available statistics, including income distributions, labor elasticities, and expenditure multipliers (ρ_j). As previously demonstrated, labor elasticities can be expressed in terms of observable micro-elasticities, providing practical proxies even when $\varepsilon_j^A(y)$ or $\varepsilon_j^C(y)$ are not directly observable.

Theorem 1 addresses the challenge of redistribution facing objective inconsistencies and policy constraint. The optimal tax formula can be decomposed into three key components:

(*i*) Mirrleesian Term: This term represents the optimal local income tax rate in the absence of a central income tax. It serves as the baseline upon which the optimal central income tax is built.

(*ii*) Economic Weights: These weights capture the trade-offs inherent in redistributing resources across regions. A larger economic weight assigned to a region signifies a higher priority given to redistribution within that region. [27]

(*iii*) Objective Inconsistency Term: This component highlights how the central government incorporates objective inconsistencies between its own and the regions' preferences when designing the central income tax. Notably, when both inter-region and intra-regional objectives are aligned ($\rho_j = 1$ and $\widetilde{F}_j^C(y) = \widetilde{F}_j(y)$), this term, $B_j(y)$, becomes zero.

As indicated by Lemma 2, the central government possesses the ability to influence the aggregate income tax rate by adjusting the central income tax. However, such adjustments impact both regions, potentially leading to undesirable outcomes. The optimal tax formula presented in Theorem 1 effectively demonstrates how the central government can achieve constrained optimal region-specific aggregate income tax schedules by carefully balancing intra- and inter-region redistribution goals.

In conclusion, the ABC-nest-ABC formula elucidates how the central tax design: (*i*) balances the dual roles of intra- and inter-regional redistribution; (*ii*) optimizes the trade-off between redistribution across different regions; and (*iii*) leverages local redistributive efforts to achieve national redistribution goals, in the presence of objective inconsistency between governments.

A Special Case: $\rho_j = 1$. To see how the central government deals with the objective inconsistency, we consider a specific case with $\rho_j = 1$. In this case, the formula reduces to:

$$\frac{T^{\star\prime}(y)}{1-T^{\star\prime}(y)} = \sum_{j\in\mathcal{J}} \overbrace{\frac{N_j f_j(y)\, y\varepsilon_j^C(y)}{\sum_j N_j f_j(y)\, y\varepsilon_j^C(y)}}^{A_j(y)} \overbrace{\frac{\widetilde{F}_j^C(y) - \widetilde{F}_j(y)}{\widetilde{F}_j(y) - F_j(y)}}^{B_j(y)} \overbrace{\frac{\frac{\widetilde{F}_j(y)-F_j(y)}{yf_j(y)}\frac{1}{\varepsilon_j^A(y)}}{1+\frac{\widetilde{F}_j(y)-F_j(y)}{yf_j(y)}\frac{1}{\varepsilon_j^A(y)}}}^{C_j(y)}.$$

Two points merit mention here. First, under the optimal inter-region transfer scheme, ρ_j equals 1, which serves as a nice baseline (see Appendix C.2). This implies that inter-region objective inconsistencies can be effectively addressed through appropriate transfers. We will come back to this point later. Second, we

[27] In our setup with linear labor supply elasticity, ε_j^C (and consequently $\varepsilon_j^C(y)$) is independent of j. This allows for a simplification of the economic weights from $\frac{\varepsilon_j^C(y) f_j(y) N_j}{\sum_j \varepsilon_j^C(y) f_j(y) N_j}$ to $\frac{N_j f_j(y)}{\sum_j N_j f_j(y)}$. However, we retain $\varepsilon_j^C(y)$ within the economic weight expression to maintain its explicit representation.

observe that $C_j\left(y_j(\theta)\right)$ coincides with $\tau_j^\star(\theta)$. This equivalence will be revisited in the context of the model extension.

Here we illustrate how objective inconsistency affects optimal central income tax and the resulting aggregate income tax with the above case. By Theorem 1, we obtain the following results:

Corollary 1 *Consider $\rho_j = 1$. Under the optimal central income taxation:*

(i) If for any $j \in \mathcal{J}$, the governments have the same preference for $\left(\underline{\theta}_j, \widetilde{\theta}_j(y)\right)$-residents, i.e., $\widetilde{F}_{\theta_j}\left(\widetilde{\theta}_j(y)\right) = \widetilde{F}^C_{\theta_j}\left(\widetilde{\theta}_j(y)\right)$, then $T^{\star\prime}(y) = 0$;

(ii) If for any $j \in \mathcal{J}$, the local governments have lower preference for $\left(\underline{\theta}_j, \widetilde{\theta}_j(y)\right)$-residents, i.e., $\widetilde{F}_{\theta_j}\left(\widetilde{\theta}_j(y)\right) \leq \widetilde{F}^C_{\theta_j}\left(\widetilde{\theta}_j(y)\right)$, then $T^{\star\prime}(y) \geq 0$ and $T_j^{A\star\prime}(y) \geq \tau_j^\star\left(\widetilde{\theta}_j(y)\right)$, and vice versa.

Proof. The results above can be derived directly from Theorem 1 and Proposition 1. ■

Parts (*i*) of Corollary 1 shows that the central income tax is limited to correct tax rate deviation caused by objective inconsistency. Parts (*ii*) of Corollary 1 suggest how to deal with objective inconsistency. If the central government has a higher preference for $\left(\underline{\theta}_j, \widetilde{\theta}_j(y)\right)$-residents than the local governments, it should enforce non-negative marginal central income tax rates. Otherwise, it enforces non-positive marginal tax rates. As a result, when the central government has a higher preference or obligation for equality, the aggregate income tax rate is higher than the aggregate income tax rate in a decentralized economy without a central income tax, i.e., $T_j^{A\prime}(y) \geq \tau_j^\star\left(\widetilde{\theta}_j(y)\right)$.

5.2 Objective Inconsistency

Why do central income tax rates in practice tend to be positive and more progressive than the local taxes? Using Theorem 1, we proceed to show how objective inconsistency can contribute to the observed patterns of the tax system in the real economy.

Objective inconsistency results from heterogeneous preference (or obligation to redistribute) and differences in residents of each jurisdiction. First, the central government generally has a higher obligation to redistribute. As an example pointed out by Wellisch (2000), the German central government is obliged to compress inter-region income gaps. [28]. Second, the objectives of local and central governments differ due to the different demographics of residents. For instance, a high-income individual in a poor region might be considered middle- or low-income in a wealthy region, leading the central government to weigh their needs differently than the local government.

Objective inconsistency manifests in two forms: *intra-region* and *inter-region inconsistency*, captured by the inconsistency in intra- and inter-region social welfare weights (see equations (18)), respectively.

Intra-region Inconsistency. A single-region case is a useful starting point for understanding the effect of intra-region Inconsistency. Equivalently, one can consider an economy with symmetric regions. In both cases, the problem is abstracted from the inter-region inconsistency. We have the following result:

[28] See, e.g., Article 106 in combination with Article 72 of the German Constitution

Corollary 2 *Consider $\rho_j = 1$, $j \in \mathcal{J}$. When there is a single region, for any $\theta \in \Theta$, the optimal taxation satisfies*

$$T^{\star\prime}\left(y\left(\theta\right)\right) = \frac{\tau^{C,\star}(\theta) - \tau^{\star}\left(\theta\right)}{1 - \tau^{\star}\left(\theta\right)} \quad \text{and} \quad T^{A\prime}\left(y\left(\theta\right)\right) = \tau^{C,\star}(\theta), \tag{23}$$

which means we have the following results:

(i) Suppose that $\widetilde{F}^{C}_{\theta_j}\left(\theta\right) > F_{\theta_j}\left(\theta\right)$. For any $\theta \in \Theta$ we have the following results: $T^{\star\prime}_{j}\left(y\left(\theta\right)\right) < \tau^{\star}\left(\theta\right)$ if $\widetilde{F}^{C}_{\theta_j}\left(\theta\right) > \widetilde{F}_{\theta_j}\left(\theta\right)$; $T^{\star\prime}\left(y\left(\theta\right)\right) > T^{\star\prime}_{j}\left(y\left(\theta\right)\right)$ if and only if $\widetilde{F}^{C}_{\theta_j}\left(\theta\right) - \widetilde{F}_{\theta_j}\left(\theta\right) > \widetilde{F}_{\theta_j}\left(\theta\right) - F_{\theta_j}\left(\theta\right)$.

(ii) Suppose that $\frac{1-\tau^{\star}(\theta)}{1-\tau^{C,\star}(\theta)} > \tau^{\star}\left(\theta\right)$ and $\tau^{\star\prime}\left(\theta\right) < 0$.[29] Then the central income tax is more progressive at $y(\theta)$ than the local income tax in the sense that $p^{C}\left(y(\theta)\right) > p_j\left(y(\theta)\right)$.

Proof. See Appendix B.5. ■

Corollary 2 shows how preference inconsistency shapes central income tax. It suggests the central government's stronger desire for equality as a reason for a relatively progressive central income tax. First, only if the central government has a higher preference for equality or a stronger obligation to compress income gaps ($\widetilde{F}^{C}_{\theta_j}\left(\theta\right) > \widetilde{F}_{\theta_j}\left(\theta\right)$) will it enforce a positive marginal income tax rate. Second, one has $T'\left(y\right) > T'_j\left(y\right)$ when $\widetilde{F}^{C}_{\theta_j}\left(\theta\right) - \widetilde{F}_{\theta_j}\left(\theta\right) > \widetilde{F}_{\theta_j}\left(\theta\right) - F_{\theta_j}\left(\theta\right)$. From another perspective, the local government's low incentive to pursue equality may result in a more progressive central income tax. To see this, part two of Corollary 2 shows $p^{C}\left(y(\theta)\right) > p_j\left(y(\theta)\right)$ when $\tau^{\star\prime}\left(\theta\right) < 0$.

Formula (23) suggests that when the central government does not face region-uniform policy constraint (since there an unique region), the second-best allocation is achievable. The local government is only relevant for determining the decomposition of the aggregate income tax between central and local income taxes.

Inter-region Inconsistency. Even if the intra-region welfare weight is consistent, central income tax is desirable when there is inter-region inconsistency. For example, due to regional inequality, inter-region welfare weight is inconsistent, i.e., $\rho_j \neq \rho_{-j}$, unless the inter-region transfer is free and optimal (see Section C for detail).

Corollary 3 *We have the following results with intra-region consistency, i.e., $\widetilde{F}^{C}_{\theta_j}\left(\theta\right) = \widetilde{F}_{\theta_j}\left(\theta\right)$ for any $\theta \in \Theta$: when $\rho_j = 1$, one has $T^{\star\prime}\left(y\right) = 0$. Otherwise, $T^{\star\prime}\left(y\right)$ is not equal to zero for some $y \in \left(\underline{y}, \overline{y}\right)$.*

Proof. The results above can be derived directly from Theorem 1.[30] ■

Corollary 3 shows that Mirrlees (1971)'s finding of the optimal income tax rate on the boundary does not hold in an economy with multi-level governance and inter-region inequality. Intuitively, when inter-region transfer is not free, the central income tax bears the role of inter-region transfer. In this case, even if the intra-region welfare weights are consistent, a central income tax should be used to redistribute between regions.

Corollaries 1 to 3 suggest inter-region inequality, lack of inter-region transfer, and the central government's high obligation to promote equality as elements for the relatively high level and progressivity of the central income tax in the real economy. Section 6 presents labor mobility as another element.

[29] A sufficient condition for $\frac{1-\tau^{\star}(\theta)}{1-\tau^{C,\star}(\theta)} > \tau^{\star}\left(\theta\right)$ is $\tau^{C,\star}(\theta) \geq \tau^{\star}\left(\theta\right)$ or equivalently $\widetilde{F}^{C}_{j}\left(\theta\right) \geq \widetilde{F}_{j}\left(\theta\right)$.

[30] Note that if $\rho_j \geq 1$, one has $\rho_{-j} \leq 1$, where the equality holds if $\rho_j = 1$ (see equation (18)).

6 Model Extension

This section explores the impact of labor mobility and tax competition on policy design. The most important issue here is internalizing the horizontal revenue externalities due to tax competition. While formally deriving the optimal central taxation scheme under tax competition between local governments presents considerable analytical challenges, we establish a general framework and methodology to address this problem. Subsequently, we apply this framework to analyze both symmetric and asymmetric economies. Notably, under some common assumptions, the ABC formula derived in our benchmark model remains applicable even in the presence of tax competition. Introducing fiscal externality is analogous to endogenizing social welfare weights. Since our benchmark model elucidates how optimal taxation responds to changes in social welfare weights, it is still explanatory. The difference is that policy design now needs to take the endogeneity of social welfare weights into consideration, which results in a modified B.

Preference and Population. Remember that $V_j(\theta) = c_j(\theta) - \phi(l_j(\theta))$ is the *individual utility*. In the following analysis, we assume that $\varepsilon_j = \varepsilon \in \mathbb{R}_+$ is independent of j. Following Allen et al. (2019), the *gross utility* and the "population" of (θ, j) residents, are functions of $\theta, V_j(\theta), V_{-j}(\theta)$. We set $U_j(\theta) = U\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$ and $f_j^p(\theta) = f_j^p\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$ as the gross utility and population. Moreover, $U\left(\theta, V_j(\theta), V_{-j}(\theta)\right) = U\left(\theta, V_{-j}(\theta), V_j(\theta)\right)$, so that $U_j(\theta) = U_{-j}(\theta)$.[31] Underlying this assumption is the labor mobility between regions that results in the unity of gross utilities between regions for the same skills. We therefore omit the subscript of $U_j(\cdot)$ in the following analysis.

Set $\underline{\theta} = \min\{\underline{\theta}_j\}_{j\in\mathcal{J}}$ and $\overline{\theta} = \max\left\{\overline{\theta}_j\right\}_{j\in\mathcal{J}}$. For simplicity, throughout model extension, we assume that $\Theta_1 = \Theta_2 = \Theta = \left(\underline{\theta}, \overline{\theta}\right)$ with $0 \le \underline{\theta}, \overline{\theta} \le \infty$, $\overline{y}_j = \overline{y}_{-j} = \overline{y} \le \infty$, and $\underline{y}_j = \underline{y}_{-j} = \underline{y} < \overline{y}$. Set $\mathcal{Y} \equiv \left(\underline{y}, \overline{y}\right)$. Denote $F_j^p(\theta) \equiv \int_{\underline{\theta}}^{\theta} f_j^p(x)dx$ as the total population of residents in j region with skills lower than θ. $N_j = \int_{\underline{\theta}}^{\overline{\theta}} f_j^p(\theta)\, d\theta$ is the total population of region j. PDF and CDF of θ-skilled residents in region j are $f_{\theta_j}(\theta) = \frac{f_j^p(\theta)}{N_j}$ and $F_{\theta_j}(\theta) = \frac{F_j^p(\theta)}{N_j}$.

Policy and Governments' Objectives. The policies of local governments are the same as that of the benchmark model. The central government now has an additional policy — inter-government transfer E_j, which is the part of regional tax revenue belonging to the central government. All the remaining regional taxes are fed back to the regional residents as a lumpsum transfer (tax, if negative). The regional and central governments' budget constraints are the same as before.

As in our benchmark model, we consider welfare functions that are the weighted average of gross utilities. However, the gross utility $U(\theta)$ is now a nonlinear function of the individual utility $V_j(\theta)$. The central

[31] In particular, Allen et al. (2019) considers $U_j(\theta) = \left[\frac{f_j^p(\theta)}{f(\theta)}\right]^{-\kappa} V_j(\theta)$, where $\kappa \neq 0$ captures the extent of population externality. Solution of extensive problem thus determines the equilibrium distribution of populations $f_j^p(\theta)$ as a function of $\{V_j(\theta)\}_{j\in\mathcal{J}}$, which satisfies the following laborer supply condition: $f_j^p(\theta) = f(\theta)\frac{V_j(\theta)^{\frac{1}{\kappa}}}{\sum_j V_j(\theta)^{\frac{1}{\kappa}}}, \forall\theta \in \Theta, j \in \mathcal{J}$. Alternatively, one can take Eaton and Kortum (2002)'s approach to model the preference for location and labor mobility. As demonstrated by Allen et al. (2019), there is an equivalence between and the setups of Allen et al. (2019) and Eaton and Kortum (2002) in modeling gravity. Thus, the population externality can be treated as a general approach for modeling location choice. Our results do not rely on the specific form of $U_j(\theta)$.

government's objective is maximizing:

$$\int_{\Theta} U\left(\theta, V_j(\theta), V_{-j}(\theta)\right) \widetilde{f}_{\theta}^{C}(\theta)\, d\theta. \tag{24}$$

There is a complementarity or substitutability between the individual utilities of the same skilled individuals working in different cities. This complementarity or substitutability relation constructs an endogenous objective of local governments. The local government j treat $\left\{V_{-j}(\theta)\right\}_{\theta\in\Theta}$ as given and maximize:

$$\int_{\Theta} U\left(\theta, V_j(\theta), V_{-j}(\theta)\right) \widetilde{f}_{\theta_j}(\theta)\, d\theta. \tag{25}$$

Equilibrium. We now revise the definition of MGTE and define the *Symmetric Multi-level Governance Tax Equilibrium* (SMGTE).

Definition 2 *(i) Given a central policy system $\mathcal{T}^C = \left\{T, E_j\right\}_{j\in\mathcal{J}}$, a MGTE with labor mobility is a combination of the central policy system $\mathcal{T}^C$ and the resulting allocation $\mathcal{A} \equiv \left\{l_j(\theta), c_j(\theta), f_j^p(\theta)\right\}_{\theta\in\Theta, j\in\mathcal{J}}$ and regional policy system $\mathcal{T}^R$, such that given the central government's policies $\mathcal{T}^C$ and local government $-j$'s policy, the local government j's policy maximizes the local governments' objectives (25) under the regional budget constraint and individuals' rational reaction and labor maximize their gross utilities under given central and regional policies, which means for any $\theta_j \in \Theta$ and $j \in \mathcal{J}$ allocation $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right)\right\} \in \mathcal{A}$ solve (θ, j) resident's problem P-1, and the population $f_j^p(\theta) \in \mathcal{A}$ satisfies the laborer supply condition $f_j^p(\theta) = f_j^p\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$; and the final good market is cleared, i.e., (11) is satisfied. (ii) A SMGTE is a MGTE where A_j, $\widetilde{f}_{\theta_j}(\theta)$ and $\widetilde{f}_{\theta}^{C}(\theta)$ are independent of $j \in \mathcal{J}$, such that the allocation $\mathcal{A}$ and local governments' policies $\mathcal{T}^R$ are also independent of j.*

Throughout this sector, we consider interior equilibrium, where for any $\theta \in \Theta$ and $j \in \mathcal{J}$, $f_j^p(\theta) > 0$ whenever $f_{\theta}(\theta) > 0$. Under an SMGTE, $f_j^p(\theta) = \frac{f_{\theta}(\theta)}{2}$ is exogenous. The central government's problem is choosing a symmetric allocation $\{c(\theta), l(\theta)\}_{\theta\in\Theta}$ that maximize national welfare under the resource constraints and rational reactions of agents. We first solve for the optimal reaction of regional governments in an arbitrary MGTE. We then solve for the optimal central income tax of SMGTE.

6.1 Decentralized Redistribution

Partial Equilibrium Elasticities. We find that the optimal decentralized and centralized redistribution can be expressed in terms of a small number of partial equilibrium elasticities (PEEs) and social welfare weights. These PEEs are the labor income elasticities $\varepsilon_j^A\left(y_j(\theta)\right) = \frac{1}{\frac{1}{\varepsilon_j} + p_j^A\left(y_j(\theta)\right)}$ and $\varepsilon_j^C\left(y_j(\theta)\right) = \frac{1}{\frac{1}{\varepsilon_j} + p^C\left(y_j(\theta)\right)}$, labor income distributions, populations, and the following semi-elasticity of population $m_{jk}(\theta)$. Remember that we have assumed that $\varepsilon_j = \varepsilon \in \mathbb{R}_+$. Therefore, $\varepsilon_j^C(\cdot)$ and $\varepsilon_j^A(\cdot)$ are independent of j.

To capture the effect of individual utility on population mobility, we define the semi-elasticity of the population with respect to individual utility (semi-elasticity of the population for short) as:

$$m_{jk}(\theta) \equiv \frac{\partial \ln f_j\left(\theta; V_j(\theta), V_{-j}(\theta)\right)}{\partial V_k(\theta)}, \forall \theta \in \Theta, j, k \in \mathcal{J}. \tag{26}$$

All other elements in the statistics-based formula can be expressed in terms of the micro elasticities and social welfare weights.

Revenue Effects. Under quasi-linear utility, $m_{jk}(\theta)$ is also the semi-elasticity of the population with respect to consumption. To capture the revenue effects of local income taxes, we further define

$$r_{jk}(\theta) \equiv m_{jk}(\theta)\left[T_j\left(y_j(\theta)\right) - E_j\right] \quad \text{and} \quad \bar{r}_{jk}(\theta) \equiv \int_{\theta}^{\bar{\theta}} r_{jk}(x) \frac{f_{\theta_j}(x)}{1 - F_{\theta_j}(\theta)} dx, \tag{27}$$

for any $\theta \in \Theta, j, k \in \mathcal{J}$.

$r_{jk}(\theta)$ is the marginal revenue effect on region j induced by one extra unit of lumpsum subsidy to (θ, k) residents. Correspondingly, $\bar{r}_{jk}(\theta)$ is the weighted revenue effect. Marginal revenue effect $r_{jk}(\theta)$ is positive if and only if the net tax revenue of the local government from (θ, j) resident $T_j\left(y_j(\theta)\right) - E_j$ is positive.

Welfare Weights. Last, we define the marginal and weighted welfare weight of local government for any $\theta \in \Theta, j \in \mathcal{J}$:

$$g_j(\theta) \equiv \frac{1}{\lambda_j} \frac{\partial U(\theta)}{\partial V_j(\theta)} \frac{\widetilde{f}_{\theta_j}(\theta)}{N_j f_{\theta_j}(\theta)} \quad \text{and} \quad \bar{g}_j(\theta) \equiv \frac{\int_{\theta}^{\infty} g_j(x) f_{\theta_j}(x) dx}{1 - F_{\theta_j}(\theta)}. \tag{28}$$

λ_j is the shadow value of local government revenue.[32]

Notations. It is a bit of a symbol abuse, but for the sake of simplicity in expression, we set notations in terms of labor income: $g_j\left(y_j(\theta)\right) = g_j(\theta), \bar{g}_j\left(y_j(\theta)\right) = \bar{g}_j(\theta), r_{jk}\left(y_j(\theta)\right) = r_{jk}(\theta), m_{jk}\left(y_j(\theta)\right) = m_{jk}(\theta)$, for $\theta \in \Theta, j, k \in \mathcal{J}$. Under Assumption 1, above notations are well defined.

Optimal Local Income Tax. Solving the planning problem of local governments, we observe the strategic reaction of local governments to the federal income tax:[33]

Proposition 3 *The optimal local income tax and the resulting aggregate income tax again satisfy equations (9) and (10), except that $\tau_j^{\star}(\theta)$ is generalized to be*

$$\tau_j^{\star}(\theta) \equiv \frac{\left[1 - \bar{g}_j(\theta) - \bar{r}_{jj}(\theta)\right] \frac{1 - F_{\theta_j}(\theta)}{\theta f_{\theta_j}(\theta)} \left(1 + \frac{1}{\varepsilon_j}\right)}{1 + \left[1 - \bar{g}_j(\theta) - \bar{r}_{jj}(\theta)\right] \frac{1 - F_{\theta_j}(\theta)}{\theta f_{\theta_j}(\theta)} \left(1 + \frac{1}{\varepsilon_j}\right)}, \forall \theta \in \Theta, j \in \mathcal{J}, \tag{29}$$

which is exactly the local income tax rate without a central income tax.

Proof. See Appendix D.1. ■

It can be seen that $1 - \tau_j^{\star}(\theta)$ again captures the sensitivity of local income tax to central income tax. Except that now the optimal local income tax takes the tax competition and horizontal fiscal effect into consideration. Proposition 3 can be treated as an extension of Lemma 2, where the weight for residents with skill higher than θ is modified by the horizontal fiscal effect $\bar{r}_{jj}(\theta)$.

Increasing the utility $V_j(\theta)$ attracts more individuals of type (θ, j), thereby boosting tax revenue (subsidy if negative) in that region. Raising the revenue effect increases the region's welfare weight associated with $V_j(\theta)$, akin to increasing the Pareto weight in the benchmark model. Therefore, the higher the fiscal effect, the lower the optimal local income tax.

[32] See equation (D4) for the expression of λ_j.
[33] See Appendix E.1 for details.

6.2 Central Redistribution

The key to solving the optimal central income tax is describing the feasible allocation under enforceable policies. We can resort to Lemma 6 to describe the implementable allocations, except now that the population depends on the individual utility, and the policy constraint is modified. Besides, unlike the benchmark problem, we now allow the central government to manipulate the inter-government transfer by choosing E_j.

Due to the policy constraint, the asymmetric problem is fairly complicated. In particular, the federal income tax will change the social welfare weights, revenue effects, and regional income distributions, which in turn relax or tighten the policy constraints. We demonstrate the implementability of the constrained optimal allocation in Lemma 16 (see Appendix E.2).

6.2.1 Symmetric Example

Following Lemma 16, we can use Lagrange multipliers to solve for the constrained optimal allocation. However, the general problem is extremely complicated. It is out of our capacity to provide a transparent solution to such a complex problem. The symmetric case, where the uniform policy constraint is loosed (see, e.g., Corollary 5 in Appendix E.2), provides a useful section. Solving for the optimal central income tax of SMGTE:

Proposition 4 *Optimal central income taxation of SMGTE satisfies:*

$$\frac{T'\left(y\left(\theta\right)\right)}{1-T'\left(y\left(\theta\right)\right)}=\frac{\left[\overline{g}_j\left(\theta\right)+\overline{r}_{jj}\left(\theta\right)-\overline{g}\left(\theta\right)\right]\frac{1-F_\theta(\theta)}{\theta f(\theta)}\left(1+\frac{1}{\varepsilon}\right)}{1+\left[1-\overline{g}_j\left(\theta\right)-\overline{r}_{jj}\left(\theta\right)\right]\frac{1-F_\theta(\theta)}{\theta f(\theta)}\left(1+\frac{1}{\varepsilon}\right)},\forall\theta\in\Theta, \tag{30}$$

where $\overline{g}\left(\theta\right)$ is a weighted social welfare weights of the central government. $\overline{g}_j\left(\theta\right)$ and $\overline{r}_{jj}\left(\theta\right)$ are independent of j, $\forall j\in\mathcal{J}$.

Proof. See Appendix D.2. ■

Proposition 4 is a generation of Proposition 2 (see equation (20)). It highlights the key implication of this paper. The role of central income tax is to correct objective inconsistency. Therefore, the optimal central income tax depends on the gap between the Pareto weights (i.e., the gap between $\overline{g}_j\left(\theta\right)+\overline{r}_{jj}\left(\theta\right)$ and $\overline{g}\left(\theta\right)$). Introducing labor mobility modifies local government's weight for individual utilities and thus the optimal federal income tax. Proposition 3 and 4 show that while $T_j'\left(y\left(\theta\right)\right)$ decreases in $\overline{r}_{jj}\left(\theta\right)$, $T'\left(y\left(\theta\right)\right)$ increases in $\overline{r}_{jj}\left(\theta\right)$. Intuitively, to internalize the horizontal externality, the central government should increase the aggregate regional tax rate on $y\left(\theta\right)$ when $\overline{r}_{jj}\left(\theta\right)$ is positive.

As an illustration of the effect of revenue effect, consider the maximin welfare function. The importance of this maximin welfare function comes both from its simplicity and duality to the problem of maximizing revenue under a rational constraint.[34] When all governments maximize the utility of the lowest-skilled residents, $\overline{g}_j\left(\theta\right)=\overline{g}\left(\theta\right)$.[35] In this case, the sign of $T'\left(y\left(\theta\right)\right)$ is generally determined by the sign of $\overline{r}_{jj}\left(\theta\right)$. Therefore, the optimal federal income tax rate on the taxpayers is positive when $\overline{r}_{jj}\left(\theta\right)>0$. Otherwise, it is

[34]Moreover, the revenue-maximizing problem is analogous to the firm's profit-maximizing problem. The analysis under a maximin welfare function thus also sheds light on how to design tax with intra-firm income distribution.

[35]See Boadway and Jacquet (2008) and Lehmann et al. (2014) for details about the maximin welfare function and optimal taxation under this welfare function.

negative. Since high-income individuals pay more taxes, the revenue effect may result in a less progressive local income tax and a more progressive central income tax. Competition for tax revenue also adds to the explanations of why the federal income tax is higher than the state income taxes on average.

6.2.2 General Solution

Generalized Welfare Weights. To solve the optimal taxation with the variational approach, we first generalize some notations presented in our benchmark model to incorporate the effect of labor mobility and competition for taxes.

First, as analogs to inter- and intra-region welfare weights in the benchmark (see equation (18)), we define the inter- and intra-region welfare weights of local and central governments in our model extension:

$$\rho_j^C \equiv \frac{\int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial U(x)}{\partial V_j(x)} \widetilde{f}_\theta^C(x)\, dx}{\sum_j \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}_\theta^C(x) \frac{\partial U(x)}{\partial V_j(x)} dx},\ \widetilde{F}_j\left(y_j(\theta)\right) \equiv \frac{\int_{\underline{\theta}}^{\theta} \frac{\partial U(x)}{\partial V_j(x)} \widetilde{f}_j(x)\, dx}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial U(x)}{\partial V_j(x)} \widetilde{f}_j(x)\, dx}, \text{ and } \widetilde{F}_j^C\left(y_j(\theta)\right) \equiv \frac{\int_{\underline{\theta}}^{\theta} \frac{\partial U(x)}{\partial V_j(x)} \widetilde{f}^C(x)\, dx}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial U(x)}{\partial V_j(x)} \widetilde{f}^C(x)\, dx},$$

for $\theta \in \Theta, j \in \mathcal{J}$. For later use, we also set $\rho^C \equiv \sum_j \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}_\theta^C(x) \frac{\partial U(x)}{\partial V_j(x)} dx$.

To capture the influence of revenue effect on social welfare weight, we set

$$F_j^R\left(y_j(\theta)\right) \equiv \frac{\int_{\underline{\theta}}^{\theta} \left[1 - r_{jj}\left(y_j(x)\right)\right] f_j(x)\, dx}{\int_{\underline{\theta}}^{\overline{\theta}} \left[1 - r_{jj}\left(y_j(x)\right)\right] f_j(x)\, dx}, \forall \theta \in \Theta, j \in \mathcal{J}.$$

By definitions, one has $\frac{1-F_j^R\left(y_j(\theta)\right)}{1-F_\theta(\theta)} = \frac{1-\overline{r}_{jj}\left(y_j(\theta)\right)}{1-\overline{r}_{jj}}$, where $\overline{r}_{jj} \equiv \overline{r}_{jj}\left(y_j(\underline{\theta})\right)$.[36] Therefore, $1 - F_j^R(\cdot)$ is just another expression of one minus the weighted revenue effect $1 - \overline{r}_{jj}(\cdot)$. Under above definitions, one has another expression of $\tau_j^\star(\theta)$. $\tau_j^\star(\theta) = C_j\left(y_j(\theta)\right)$ and

$$C_j(y) \equiv \frac{\left(1 - \overline{r}_{jj}\right) \frac{\widetilde{F}_j(y) - F_j^R(y)}{y f_j(y)} \frac{1}{\varepsilon_j^A(y)}}{1 + \left(1 - \overline{r}_{jj}\right) \frac{\widetilde{F}_j(y) - F_j^R(y)}{y f_j(y)} \frac{1}{\varepsilon_j^A(y)}}, \forall y \in \mathcal{Y}. \tag{31}$$

Revenue Effect in General Equilibrium. Set $y_j^k(y) \equiv y_j(\theta)|_{y_k(\theta)=y}$ as the labor income of (θ, j) when $y_k(\theta) = y$, and define a general-equilibrium-weighted-average revenue effect for any $\theta \in \Theta$, and $j, k, n \in \mathcal{J}$:

$$\begin{aligned} \overline{r}_{kj}^G(\theta) &= \overline{r}_{kj}(\theta) + \sum_n \frac{N_n}{\sum_n N_n} \int_\theta^{\overline{\theta}} \left[T\left(y_n(x)\right) + E_n\right] m_{nj}(x) \frac{f_n(x)}{1 - F_k(\theta)} dx, \\ \overline{r}_{kj}^G(\underline{\theta}) &= \overline{r}_{kj}(\underline{\theta}) + \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_n(\theta)\right) + E_n\right] f_n(\theta) m_{nj}(\theta)\, d\theta. \end{aligned} \tag{32}$$

The revenue effect $\overline{r}_{kk}$ only takes the revenue effect of transfer to region k on region k. While the GRE takes revenue effect $\overline{r}_{kk}^C$ of the transfer on all regions into consideration. Last set $\left[\eta_{kn}^G\right] \equiv \left[\mathbf{I} - \left[\overline{r}_{kj}^G\right]\right]^{-1}$ as the

[36] By definition, $F_j^R\left(y_j(\theta)\right) = \frac{[1-F_\theta(\theta)] - \int_\theta^{\overline{\theta}} r_{kk}(x) f_j(x) dx}{1 - \int_{\underline{\theta}}^{\overline{\theta}} r_{kk}(x) f_j(x) dx}$. Notice that $\overline{r}_{kk}(\theta) = \frac{\int_\theta^{\overline{\theta}} r_{kk}(x) f_j(x) dx}{1 - F_\theta(\theta)}$ and $\overline{r}_{kk}(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} r_{kk}(x) f_j(x)\, dx$. We have $1 - F_j^R\left(y_j(\theta)\right) = [1 - F_\theta(\theta)] \frac{1-\overline{r}_{kk}(\theta)}{1-\overline{r}_{kk}(\underline{\theta})}$.

inverse matrix of $\mathbf{I} - \left[\bar{r}_{kj}^{G}\right]$. Then we define

$$\rho_j \equiv \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \Big/ \frac{N_j}{\sum_j N_j}$$

as an inter-region welfare weight modified by the revenue effect, $j \in \mathcal{J}$. Again, we have $\sum_j N_j \rho_j = 1$. Under $\sum_j N_j \rho_j = 1$, the symmetricity of ρ_j implies $\rho_j = 1$.

General Equilibrium Elasticity of Labor Income. Solving the individual's and local government's problems, we have the following results in the equilibrium:

Proposition 5 *Consider a tax reform without changes in inter-region transfers. Under regularity conditions provided in Assumption 2, there exist own and cross elasticities $\varepsilon_j^C : \mathcal{Y} \mapsto \mathbb{R}$, $\varepsilon_j^{Cross} : \mathcal{Y}^2 \mapsto \mathbb{R}$ such that*

$$\frac{\widehat{y}_j}{y_j} = -\varepsilon_j^C\left(y_j\right) \frac{\widehat{T}'\left(y_j\right)}{1 - T'\left(y_j\right)} - \int_{\underline{y}}^{\overline{y}} \varepsilon_j^{Cross}\left(y_j, y\right) \frac{\widehat{T}'\left(y\right)}{1 - T'\left(y\right)} dy, \forall j \in \mathcal{J}, y_j \in \mathcal{Y}.$$

The own elasticity turns out consistent with the elasticity in our benchmark, i.e., $\varepsilon_j^C\left(y_j\right) = \frac{1}{\frac{1}{\varepsilon_j} + T''\left(y_j\right) y_j / \left[1 - T'\left(y_j\right)\right]}$.

Proof. See Appendix D.3. ■

The own elasticity still satisfies (15). What's new is the cross elasticity $\varepsilon_j^{Cross}\left(y_j\left(\theta'\right), y\left(\theta\right)\right)$. The cross elasticity captures the influence of change in marginal tax rate $T'\left(y\left(\theta\right)\right)$ on the labor income of agent (θ', j). It turns out to be zero in our benchmark model but generally not zero in the model extension. This is because the change in tax rate at one income point will affect the tax rate choices at other income points by altering the fiscal revenue effect (see $\bar{r}_{jj}\left(\theta\right)$ in equation (30)), which is zero in the benchmark. As long as we know how labor income reacts to tax reform, the following is the same as before, except that now the agent's utility includes spillover effects. $\varepsilon_j^{Cross}\left(y_j, y\right)$ together with $\varepsilon_j^C\left(y_j\right)$ construct the general rquilibrium rlasticity (GEEs) that determine $\widehat{y}_j$.

Welfare Analysis. In the optimum, any marginal tax reform results in $\widehat{W} = 0$, where

$$\widehat{W} = \sum_{j \in \mathcal{J}} \int_{\Theta} \widetilde{f}^C\left(\theta\right) \frac{\partial U\left(\theta; U_j\left(\theta\right), U_{-j}\left(\theta\right)\right)}{\partial U_j\left(\theta\right)} \widehat{U}_j\left(\theta\right) d\theta$$

and $\widehat{V}_j\left(\theta\right)$ still satisfies (B4), i.e., $\widehat{V}_j\left(\theta\right) = -\widehat{T}\left(y_j\left(\theta\right)\right) - \widehat{T}_j\left(y_j\left(\theta\right)\right)$. Therefore, we have Lemma 4:

Lemma 4 *Given fiscal multipliers ρ_j^C, the change in national changes in the aggregate income taxes determine welfare:*

$$-\frac{\widehat{W}}{\rho^C} = \sum_{j \in \mathcal{J}} \rho_j^C \times \widehat{T}_j^A\left(\underline{y}\right) + \sum_j \int_{\underline{y}}^{\overline{y}} \rho_j^C \left[1 - \widetilde{F}_j^C\left(y\right)\right] \widehat{T}_j^{A\prime}\left(y\right) dy.$$

Proof. See Appendix D.4. ■

The change in social welfare is determined by changes in aggregate income taxes and how changes in taxes affect the social welfare, later of which is captured by the social welfare weights. Intuitively, one unit increase in $\widehat{T}_k^A\left(\underline{y}\right)$ increases the social welfare by ρ_k^C unit. The welfare effect of the marginal increase in $\widehat{T}_k^{A\prime}\left(y\right)$ is captured by $\rho_k^C\left[1 - \widetilde{F}_k^C\left(y\right)\right]$.

Optimal Central Income Tax. When the initial tax rate is at the optimal level, its marginal change will not have a first-order effect on social welfare. By searching for this central income tax, we can solve the conditions that the optimal central income tax needs to meet. It turns out that under some common conditions, the optimal tax formula is still ABC-nest-ABC. Set $y_m^j(y) = y_m\left(\widetilde{\theta}_j(y)\right)$, where $\widetilde{\theta}_j(y)$ is the skill of an agent whose income is y. We have the following result:

Proposition 6 *Suppose ρ_j and $\varepsilon_j^{Cross}(y', y)$ are independent of j, $\forall j \in \mathcal{J}$ and $y, y' \in \mathcal{Y}$.*[37] *The optimal central income tax satisfies:*

$$\frac{T'(y)}{1-T'(y)} = \sum_{j\in\mathcal{J}} A_j(y) B_j(y) C_j(y), \tag{33}$$

where $A_j(\cdot)$ is still given by (22), and $C_j(\cdot)$ is given by (31). Last, $B_j(y)$ is given by:

$$B_j(y) \equiv \frac{\left[\begin{array}{c}(1-\overline{r}_{jj})\left[1-\widetilde{F}_j(y)\right]-\left(1-\overline{r}_{jj}^G\right)\left[1-\widetilde{F}_j^C(y)\right]+\\ \overline{r}_{jj}(y)\left[1-F_j(y)\right]-\sum_m \frac{N_m}{N_j}\overline{r}_{mj}^G\left(y_m^j(y)\right)\left[1-F_m\left(y_m^j(y)\right)\right]\end{array}\right]}{(1-\overline{r}_{jj})\left[\widetilde{F}_j(y)-F_j^R(y)\right]}, \forall j\in\mathcal{J}, y\in\mathcal{Y}. \tag{34}$$

Proof. See Appendix D.11. ■

The difference between the optimal tax formula in the benchmark and the one above comes from the variation of $B_j(y)$, which now incorporate the inter-government inconsistency of objectives due to the fiscal externality. When there is no labor mobility and $\rho_j = 1$, we will return to the benchmark result:

$$B_j(y) = \frac{\widetilde{F}_j^C(y) - \widetilde{F}_j(y)}{\widetilde{F}_j(y) - F_j(y)}, \forall y \in \mathcal{Y}.$$

Although $B_j(y)$ becomes more complicated, the underlying principle of central tax design does not. The central income tax is used to remedy the objective inconsistency between governments. Local governments do not consider the negative effects of their tax reduction on the fiscal revenue of other regions, while the central government do. This difference in perspectives is reflected in the inclusion of the fiscal effect term within the objective inconsistency term.

When ρ_j and $\varepsilon_j^{Cross}(y', y)$ are not symmetric, the result is much more involved. In particular, one additional term $\widetilde{B}_j : \mathcal{Y} \mapsto \mathbb{R}_+$ appears in the optimal tax formula (see Theorem 3 in Appendix D.7). Nevertheless, we demonstrate that micro elasticities defined in Section 6.1 are still sufficient (see Proposition 7 in Appendix D.9).

7 Numerical Analysis

Our theoretical analysis reveals that the impact of regional income taxes on national income redistribution depends on regional characteristics and governmental preferences. When regions are homogenous and governmental preferences are inconsistent, fully centralized redistribution is optimal, unless inter-regional transfers are restricted. Conversely, when regional disparities exist alongside inconsistent governmental preferences, decentralized redistribution becomes more effective for intra-regional redistribution, making

[37] In this case $\rho_j = 1$. In Appendix D.12, we show that under some assumptions on symmetricity of elasticitie, $\rho_j = 1$ is optimal.

central income tax redundant. However, the desirability of decentralized redistribution becomes less clear when regional inequality coincides with misaligned objectives.

This section investigates the optimality of decentralized redistribution within the United States. We conduct a simplified numerical analysis of our benchmark model with endogenous social welfare weights and transfer(see Section C.3). As stated in theoretical research, although the benchmark model does not incorporate the externalities caused by tax competition directly, this impact can be partly captured by selecting reasonable Pareto weights.

Furthermore, we demonstrate how objective inconsistencies and regional disparities collectively shape the optimal configuration of federal and state income taxes. We calibrate our analysis using the U.S. wage distribution, dividing it into two representative regional distributions (consider the East and West Coasts for example) and exploring scenarios with varying degrees of preference inconsistency and regional inequality. Our findings indicate that decentralized redistribution consistently exerts a minimal impact on national welfare, while central redistribution emerges as a far more influential factor.

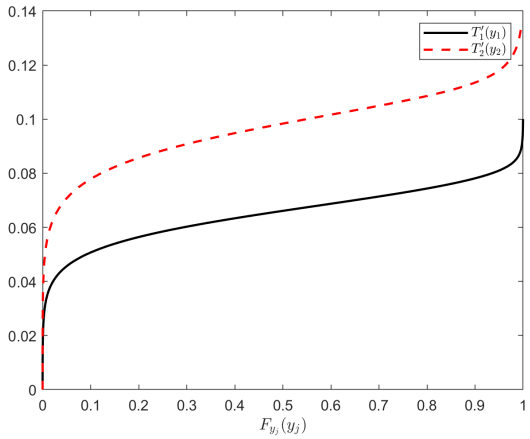
Environment. To calibrate $f_\theta(\theta)$, we should first make some assumptions about consumer preference and initial policies. We assume that the utility is $c - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$ with $\varepsilon = 0.33$ (see Chetty (2012)). Without loss of generality, we set the regional technology A_j at one and use the distribution of skills (i.e., $F_{\theta_j}(\theta)$) to govern the distributions of wages. The population $N_1 = N_2 = 0.5$. The labor income schedule is assumed to have a constant rate of progressivity. Following Heathcote et al. (2017), the original income tax on income y is $y - \frac{1-\varsigma}{1-\phi} y^{1-\phi}$. $\phi = 0.151$ is the rate of progressivity, and $\varsigma = -3$ governs the marginal tax rate in the lower bound. All tax revenue is transferred to individuals in the form of lumpsum transfers.

To get as close to the real $\widetilde{F}_{\theta_j}(\theta)$ as possible, we set state income tax rates basing on the real state income taxes in the United States. Specifically, we assume that the regional income tax in region 2 is progressive with a top marginal tax rate (MTR) of 0.133, which equals the highest marginal state income tax rate in the U.S. The regional income tax in Region 1 is also progressive with a benchmark top MTR of 10%. We considers different top MTR of region 1 (i.e., 5%, 10% and 13.3%), underlying which is the different preferences of local governments. Then, we can investigate how local government's preferences affect optimal taxation. In the benchmark case, the initial regional income tax rates and corresponding Pareto weights of local governments are given in Figure 1a and 1d.

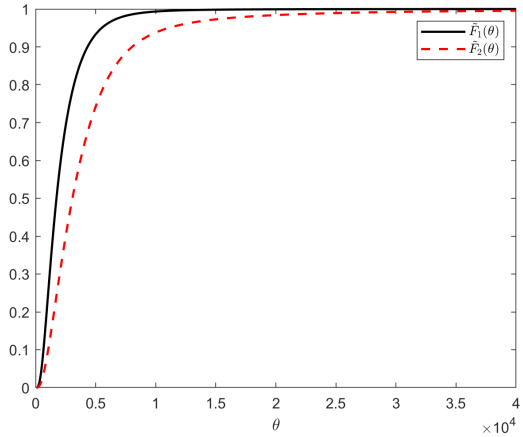
To capture the central government's desire to redistribute, we assume that the central government's objective is $\sum N_j \int_{\Theta_j} G\left(V_j(\theta)\right) f_{\theta_j}(\theta)\, d\theta$ and consider a concave social welfare function: $G(V) = \frac{V^{1-k^C}}{1-k^C}$. $k^C = 0$ corresponds to utilitarian preferences. As $k^C \to \infty$, the welfare weights approach that of a Rawlsian preference. We provide the optimal tax rates for $k^C = 2$, which is among the common values in the literature (in Sachs et al. (2020), $k^C \in [1, 3]$).

Construction of Regional Income Distributions. The income distribution is estimated according to the statistical results presented in Saez and Stantcheva (2018). They use IRS tax data for income distributions in 2007.[38] As in Saez (2001), we infer the distribution of θ and labor income from the income distribution and individual's first-order condition. We then construct two regions that are consistent with the calibrated

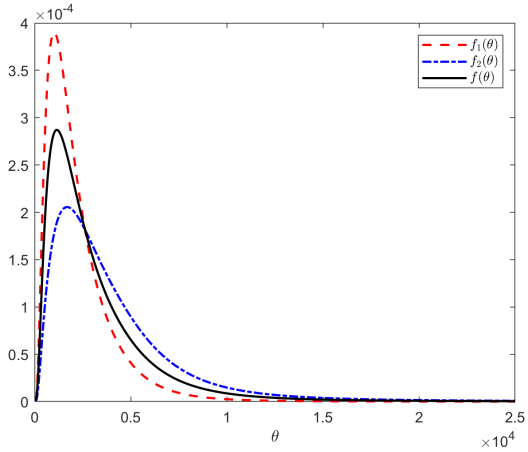
[38] As put by Saez and Stantcheva (2018): "*We choose 2007 as this is the most recent year of publicly available micro-level U.S. tax data available before the Great Recession.*"



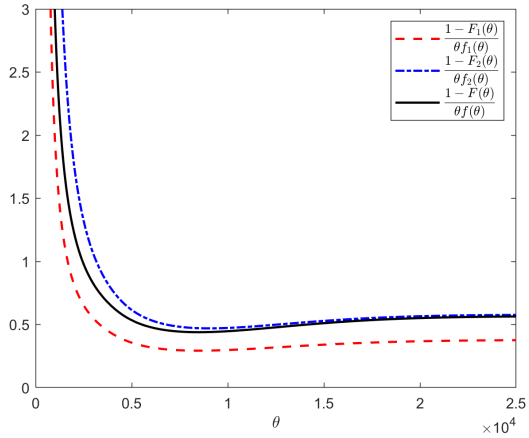
(a) Initial regional income tax rates



(b) The Pareto weights of local governments calculated by the initial regional income tax rates



(c) Regional and national wage distributions

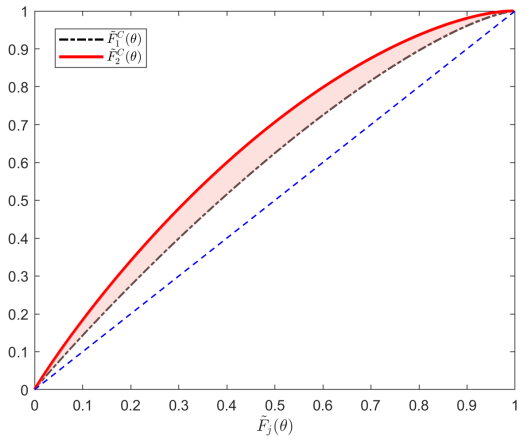


(d) Hazard ratios of wage distributions

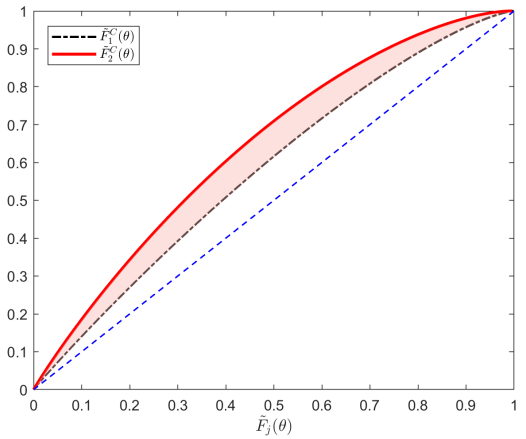
Figure 1: Pareto Weights and Wage Distributions

$f_\theta(\theta)$. Suppose region 1 is poorer than region 2. We construct the distribution of region 1 an 2 by $F_{\theta_1}(\theta) = 1 - [1 - F_\theta(\theta)]^h$ and $F_{\theta_2}(\theta) = \frac{F_\theta(\theta) - N_1 F_{\theta_1}(\theta)}{N_2}$. In the benchmark case, $h = 2$. Under the initial income taxation, the average income of Region 2 is 2.2 times of that of region 1. As a comparison, we also consider cases with $h = 1.2$ and $h = 3$. The distributions of wages and corresponding hazard ratios in the benchmark case are displayed in Figure 1c and 1d.

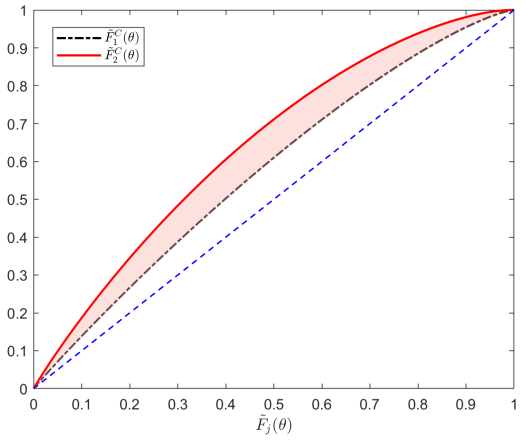
Given the federal and state income tax in the U.S., $\widetilde{F}_{\theta_j}(\theta)$ is calculated by the value of state income tax and regional governments' optimal reactions to the federal income tax (see, e.g., Lemma 2). We compare the Pareto weights of central and regional governments in Figure 2. Both the red and black lines are concave, and the red line is more concave than the black line. These results suggest that the central government has a higher desire for redistribution than the local governments, especially for region 2. Second, the decrease in the top marginal tax rate of region 1 reflects the decrease in the local government's desire for redistribution. Therefore, the preference inconsistency between the central and the local government of region 1 increases. As a reflection, the area of pink becomes larger. However, this increase in preference consistency does not necessarily result in welfare loss (see the below part on social welfare).



(a) $\widetilde{F}_j^C(\theta)$ versus $\widetilde{F}_j(\theta)$ when original Top MTR of region 1 equals 5%



(b) $\widetilde{F}_j^C(\theta)$ versus $\widetilde{F}_j(\theta)$ when original Top MTR of region 1 equals 10%



(c) $\widetilde{F}_j^C(\theta)$ versus $\widetilde{F}_j(\theta)$ when original Top MTR of region 1 equals 13.3%

Figure 2: Pareto Weights of Central and Local Governments

Optimal Tax System. Our algorithm uses the optimal tax formulas and an updating rule to identify a fixed-point optimal tax schedule (see Appendix B for details). We will compare the regional income tax T_j, federal income tax T, and aggregate regional income tax T_j^A under four different authority systems:

(a) only federal income tax (OF);

(b) only regional income tax (OR);

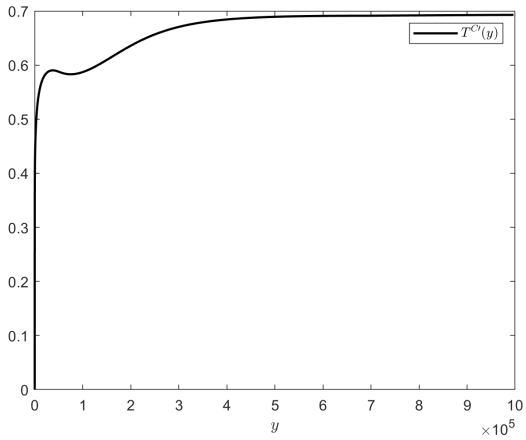
(c) both federal income tax and regional income taxes (F&R);

(d) region-specific federal income tax (RSF).

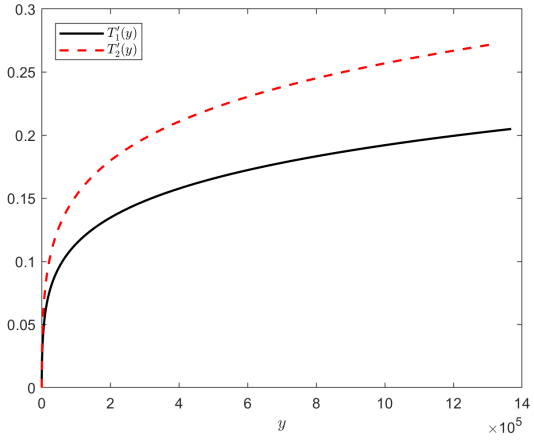
To isolate the welfare impact of income taxes, we allow for inter-regional transfers (E_j) while setting all income tax functions ($T(\underline{y})$ and $T_j(\underline{y}_j)$) to zero. Instead, we use t^C and t_j to represent lump-sum transfers from the central and local governments to their respective residents. Since only the difference between inter-regional transfers matters, we set $E_1 = 0$ and use E_2 to govern these transfers. Budget balance is maintained by ensuring that the sum of taxes and transfers to residents and governments equals zero.

Figure 3 illustrates the optimal taxation scheme under our benchmark parameters (i.e., $h = 1.5$ and a top marginal tax rate of 10% in region 1). Comparing Figures 3a, 3c, and 3d reveals how the federal income tax changes when transitioning from the OF system to the F&R system.

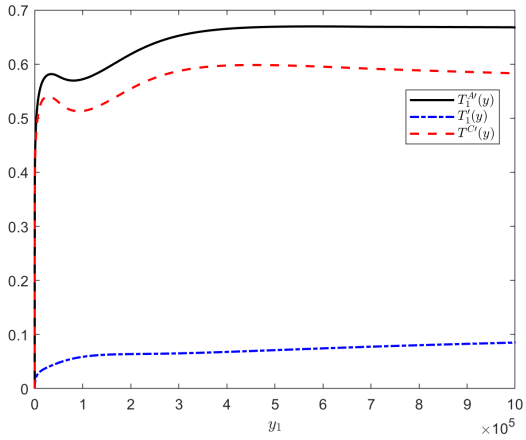
Figures 3c and 3d demonstrate how to optimally design federal income tax within an economy characterized by both inter- and intra-regional inequality. These figures illustrate the composition of optimal aggregate income taxes. Firstly, the central government implements a positive federal income tax due to its higher preference for redistribution (see Figure 2). Secondly, the aggregate income tax rate in the wealthier region ($T_2^{A\prime}(y)$) surpasses that of the poorer region ($T_1^{A\prime}(y)$) due to the higher regional income tax levied by region 2. This higher regional income tax in region 2 enables the central government to better approximate the optimal region-specific federal income tax structure.



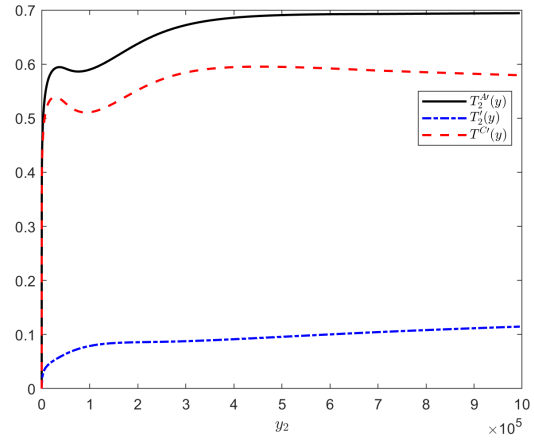
(a) Optimal $T^{C\prime}(y)$ under OF



(b) Optimal $T_j'(y)$ under OR



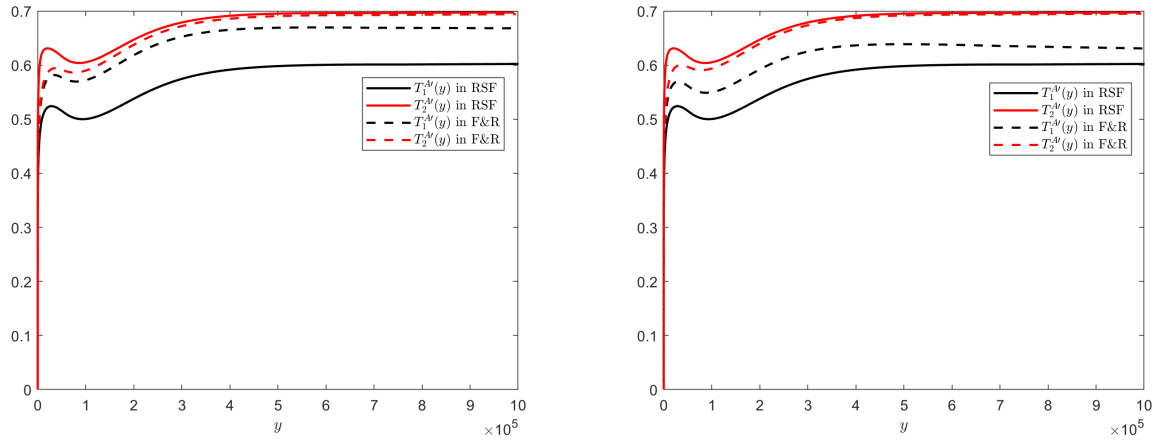
(c) Optimal $T_1'(y)$ and $T_1^{A\prime}(y)$ under F&R



(d) Optimal $T_2'(y)$ and $T_2^{A\prime}(y)$ under F&R

Figure 3: Optimal Taxation

Comparing Figure 3b to 3c and 3d shows the way regional income tax schedules change when the tax system switches from F&R to OR. It can be seen that although the regional income taxes under OR are much higher than those under F&R, they are substantially lower than the aggregate income tax under F&R. This finding already implies the limitation of regional income taxation as an instrument for redistribution. See the following welfare analysis for details.



(a) Optimal $T_j^{A\prime}(y)$ when Top MTR of region 1 equals 13.3%

(b) Optimal $T_j^{A\prime}(y)$ when Top MTR of region 1 equals 5%

Figure 4: Comparing Optimal Aggregate Tax Rates under RSF and R&F

Figures 4a and 4b compare aggregate income tax rates under a RSF system and a F&R system. Interestingly, the aggregate income tax rates for both regions under F&R fall within the range observed under RSF. This occurs because the central government faces a trade-off between inter- and intra-regional redistribution. While it might desire a higher tax rate in region 2, it refrains from doing so to avoid excessively increasing the tax rate in region 1. Switching from F&R to RSF eliminates this trade-off, allowing for an increase in $T_2^{A\prime}(y)$ and a decrease in $T_1^{A\prime}(y)$.

Social Welfare. Recall that national welfare is represented by $N_j \int_{\Theta_j} G\left(V_j(\theta)\right) f_{\theta_j}(\theta) d\theta$. Under our benchmark parameters, transitioning from OF to F&R yields a welfare gain equivalent to \$22.6 of per capita consumption. However, this benefit diminishes as the top marginal tax rate in region 1 increases, becoming a welfare loss of \$1.3 of per capita consumption when the top marginal tax rate reaches 13.3%. This occurs because a lower top marginal tax rate in region 1 leads to lower regional income taxes, bringing the aggregate income tax rates closer to those observed under RSF (see Figures 4a and 4b). This alignment stems from the central government's desire for a higher aggregate income tax rate in region 2 compared to region 1, which is facilitated by a relatively lower regional income tax rate in region 1.

These findings suggest that introducing decentralized redistribution can either enhance or diminish national welfare. However, in both cases, the impact of regional redistribution on national redistribution and welfare remains limited.

Similarly, switching from F&R to RSF generates modest welfare gains, increasing with higher values of h. For instance, when $h = 2$ (where the average income in region 2 is 3.4 times that of region 1), the welfare gain amounts to \$67.8 of per capita consumption. In contrast, transitioning OR to F&R yields significantly larger welfare gains, reaching \$8026.5 of per capita consumption in our benchmark case.

These results underscore that while regional income taxes have a minor influence on national welfare, the central income tax plays a crucial role, even when inter-regional transfers are freely available.

8 Conclusion

The diverse economic landscapes of U.S. states pose a significant challenge to the design of a fair and efficient federal income tax system. However, the U.S. Constitution mandates the uniformity of federal income tax. Such a one-size-fits-all federal income tax system struggles to adequately address the inherent disparities. This paper investigates how to construct a tax system that effectively leverages the strengths of both central and local governments while navigating their inherent conflicts. The global prevalence of multi-level governments and plausible empirical evidence on strategic local tax responses to central taxes underscore the need for normative research on optimal redistribution in multi-level governance contexts.

Our study show that the optimal central income tax is limited to correcting the objective inconsistency due to preference and object inconsistency as well as fiscal externality. Decentralized redistribution is desirable when local redistribution helps the central government mimic the optimal region-specific income taxation. Therefore, it is an empirical question whether the central government should nullify the redistributive effect of some regional income taxes through tax rebates. The numerical analysis demonstrates that, under a well-designed central income tax, local income taxes in the United States is generally desirable. Welfare effect of local taxes is small because the central income tax significantly crowds out local taxes while stabilizing the overall tax burden. From another perspective, local taxes address local needs without greatly impacting national welfare. The central income tax is essential for redistribution, even with free inter-region transfers. In contrast, implementing region-specific central income taxation would only slightly improve national welfare.

References

ALES, L., M. KURNAZ, AND C. SLEET (2015): "Technical Change, Wage Inequality, and Taxes," *The American Economic Review*, 105, 3061–3101.

ALLEN, T., C. ARKOLAKIS, AND Y. TAKAHASHI (2019): "Universal Gravity," *Journal of Political Economy*, 128, 393–433.

ATKINSON, A. B. AND J. E. STIGLITZ (1976): "The design of tax structure: direct versus indirect taxation," *Journal of Public Economics*, 6, 55–75.

BOADWAY, R. AND L. JACQUET (2008): "Optimal marginal and average income taxation under maximin," *Journal of Economic Theory*, 143, 425–441.

BOADWAY, R. W. AND D. E. WILDASIN (1984): *Public Sector Economics*, Little, Brown.

BORDIGNON, M., P. MANASSE, AND G. TABELLINI (2001): "Optimal Regional Redistribution under Asymmetric Information," *American Economic Review*, 91, 709–723.

BROWN, C. C. AND W. E. OATES (1987): "Assistance to the poor in a federal system," *Journal of Public Economics*, 32, 307–330.

CHETTY, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80, 969–1018.

CHETTY, R. AND E. SAEZ (2010): "Optimal Taxation and Social Insurance with Endogenous Private Insurance," *American Economic Journal: Economic Policy*, 2, 85–114.

CHRISTIANSEN, V. (1984): "Which commodity taxes should supplement the income tax?" *Journal of Public Economics*, 24, 195–220.

COLAS, M. AND K. HUTCHINSON (2021): "Heterogeneous Workers and Federal Income Taxes in a Spatial Equilibrium," *American Economic Journal: Economic Policy*, 13, 100–134.

DIAMOND, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

DOLIGALSKI, P., A. NDIAYE, AND N. WERQUIN (2023): "Redistribution with Performance Pay," *Journal of Political Economy Macroeconomics*, 1, 371–402, publisher: The University of Chicago Press.

DRAZEN, A. (2018): *Political Economy in Macroeconomics*, Princeton University Press.

EATON, J. AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

EECKHOUT, J. AND N. GUNER (2015): "Optimal Spatial Taxation: Are Big Cities Too Small?" SSRN Scholarly Paper ID 2554426, Social Science Research Network, Rochester, NY.

FAJGELBAUM, P. D. AND C. GAUBERT (2020): "Optimal Spatial Policies, Geography, and Sorting," *The Quarterly Journal of Economics*, 135, 959–1036.

FELDSTEIN, M. S. AND G. E. METCALF (1987): "The Effect of Federal Tax Deductibility on State and Local Taxes and Spending," *Journal of Political Economy*, 95, 710–736, publisher: The University of Chicago Press.

FINDEISEN, S. AND D. SACHS (2017): "Redistribution and insurance with simple tax instruments," *Journal of Public Economics*, 146, 58–78.

GAUBERT, C., P. M. KLINE, AND D. YAGAN (2021): "Place-Based Redistribution," Working Paper 28337, National Bureau of Economic Research.

GOLOSOV, M. AND A. TSYVINSKI (2007): "Optimal Taxation with Endogenous Insurance Markets," *The Quarterly Journal of Economics*, 122, 487–534.

GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2014): "A Variational Approach to the Analysis of Tax Systems," Tech. Rep. w20780, National Bureau of Economic Research.

——— (2016): "Chapter 10 - Recursive Contracts and Endogenously Incomplete Markets," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 725–841.

GOMES, R., J.-M. LOZACHMEUR, AND A. PAVAN (2018): "Differential Taxation and Occupational Choice," *Review of Economic Studies*, 85, 511–557.

GOODSPEED, T. J. (2000): "Tax structure in a federation," *Journal of Public Economics*, 75, 493–506.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2017): "Optimal Tax Progressivity: An Analytical Framework*," *The Quarterly Journal of Economics*, 132, 1693–1754.

HUGGETT, M. AND W. LUO (2023): "Optimal income taxation: An urban economics perspective," *Review of Economic Dynamics*, 51, 847–866.

JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2013): "Optimal redistributive taxation with both extensive and intensive responses," *Journal of Economic Theory*, 148, 1770–1805.

KAPLOW, L. (1996a): "Fiscal Federalism and the Deductibility of State and Local Taxes under the Federal Income Tax," *Virginia Law Review*, 82, 413–492.

——— (1996b): "Fiscal Federalism and the Deductibility of State and Local Taxes under the Federal Income Tax," *Virginia Law Review*, 82, 413–492, publisher: Virginia Law Review.

KOPCZUK, W., J. SLEMROD, AND S. YITZHAKI (2005): "The limitations of decentralized world redistribution: An optimal taxation approach," *European Economic Review*, 49, 1051–1079.

LEHMANN, E., L. SIMULA, AND A. TRANNOY (2014): "Tax me if you can! Optimal Nonlinear Income Tax Between Competing Governments," *The Quarterly Journal of Economics*, 129, 1995–2030.

LUENBERGER, D. G. (1997): *Optimization by Vector Space Methods*, New York: Wiley-Interscience, 1969th edition ed.

MANKIW, N. G., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

——— (1976): "Optimal tax theory: A synthesis," *Journal of Public Economics*, 6, 327–358.

MUSGRAVE, R. A. (1971): "Economics of Fiscal Federalism," *Nebraska Journal of Economics and Business*, 10, 3–13.

OATES, W. E. (1972): *Fiscal Federalism*, Edward Elgar Publishing.

——— (1999): "An Essay on Fiscal Federalism," *Journal of Economic Literature*, 37, 1120–1149.

——— (2005): "Toward A Second-Generation Theory of Fiscal Federalism," *International Tax and Public Finance*, 12, 349–373.

PARCHET, R. (2019): "Are Local Tax Rates Strategic Complements or Strategic Substitutes?" *American Economic Journal: Economic Policy*, 11, 189–224.

PARK, Y. AND Y. CHANG (2018): "Optimal Taxation with Private Insurance," Tech. Rep. no117, Institute of Economic Research, Seoul National University.

PIKETTY, T. (1997): "La redistribution fiscale face au chômage," *Revue française d'économie*, 12, 157–201.

RAFF, H. AND J. D. WILSON (1997): "Income Redistribution with Well-Informed Local Governments," *International Tax and Public Finance*, 4, 407–427.

SACHS, D., A. TSYVINSKI, AND N. WERQUIN (2020): "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *Econometrica*, 88, 469–493.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *The Review of Economic Studies*, 68, 205–229.

SAEZ, E. AND S. STANTCHEVA (2016): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," *The American Economic Review*, 106, 24–45.

——— (2018): "A simpler theory of optimal capital taxation," *Journal of Public Economics*, 162, 120–142.

SALANIÉ, B. (2003): *The economics of taxation*, The MIT Press.

SCHEUER, F. (2013): "Adverse selection in credit markets and regressive profit taxation," *Journal of Economic Theory*, 148, 1333–1360.

SCOTT, C. E. AND R. K. TRIEST (1993): "THE RELATIONSHIP BETWEEN FEDERAL AND STATE INDIVIDUAL INCOME TAX PROGRESSIVITY," *National Tax Journal*, publisher: The University of Chicago Press.

STANTCHEVA, S. (2014): "Optimal Income Taxation with Adverse Selection in the Labour Market," *The Review of Economic Studies*, 81, 1296–1329.

STARK, K. J. (2003): "Fiscal Federalism and Tax Progressivity: Should the Federal Income Tax Encourage State and Local Redistribution," *UCLA Law Review*, 51, 1389.

STIGLER, G. (1957): "The Tenable Range of Functions of Local Government," .

TOCQUEVILLE, A. D. (2003): *Democracy in America*, Regnery Publishing.

WELLISCH, D. (2000): *Theory of Public Finance in a Federal State*, Cambridge University Press.

WILDASIN, D. E. (1991): "Income Redistribution in a Common Labor Market," *The American Economic Review*, 81, 757–774.

ZEMYAN, S. M. (2012): *The Classical Theory of Integral Equations: A Concise Treatment*, Springer Science & Business Media, google-Books-ID: _PolpKk9Ii4C.

APPENDIX

A Decentralized Redistribution

A.1 Local Tax Incidence

In this subsection, given an initial tax system $\mathcal{T}$, we derive the first-order effects of a local tax perturbation with a given central income tax.

Local Tax Reform. Given an initial central income tax, consider a regional tax form on T_j with direction $\widehat{T}_j \in \mathcal{C}^2(\mathbb{R}_+, \mathbb{R})$ and extent $\mu_j \in \mathbb{R}_+$. We denote the labor supply, income, gross utility, and social welfare after such a local income tax form as $T_j\left(y_j; T_j + \mu_j \widehat{T}_j\right)$, $l_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)$, $y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)$, $V_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)$ and $W_j\left(T_j + \mu_j \widehat{T}_j\right)$. To make the expressions concise, we sometimes omit μ_j when $\mu_j \to 0$. For any μ_j, we assume that the above functions are continuously differentiable on μ_j.

When the magnitude of the tax change is small enough, the tax reform's first order effect on labor supply can be given formally by the Gateaux derivative of the labor supply functional $T_j \mapsto l_j\left(\theta_j; T_j\right)$ in the direction $\widehat{T}_j$:[39]

$$\widehat{l}_j\left(\theta_j\right) \equiv \lim_{\mu_j \to 0} \frac{1}{\mu_j}\left[l_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right) - l_j\left(\theta_j; T_j\right)\right]. \tag{A1}$$

Analogously, we can define the changes in (Gateaux derivatives of) individual's income, utility, and social welfare, and denote them as $\widehat{y}_j\left(\theta_j\right)$, $\widehat{V}_j\left(\theta_j\right)$, and $\widehat{W}_j$, respectively.

We call a regional tax reform a balanced regional tax reform if it maintains the local government's budget constraint. The following equations describe the economy under balanced tax reform:

$$\frac{\phi'\left(\frac{y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)}{A_j \theta_j}\right)}{A_j \theta_j} = 1 - T_j^{A\prime}\left(y_j\left(\theta_j; \mu_j\right); \mu_j\right), \forall \theta \in \Theta_j, \tag{A2}$$

and

$$E_j = \int_{\Theta_j}\left[T_j\left(y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)\right) + \mu_j \widehat{T}_j\left(y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)\right)\right] f_{\theta_j}\left(\theta_j\right) d\theta_j. \tag{A3}$$

Labor Income Effect. A first-order Taylor expansion of both sides of (A2) around the baseline labor supply delivers the Gateaux derivative $\widehat{y}_j\left(\theta_j\right)$ as a function of labor elasticity and tax reform, i.e., equation (3) in Lemma 1.

The expression of $\frac{\widehat{y}_j\left(\theta_j\right)}{y_j\left(\theta_j\right)}$ above is equivalent to the traditional expression if $T_j^{A\prime}\left(y_j\left(\theta_j\right)\right) = T_j'\left(y_j\left(\theta_j\right)\right)$. With the introduction of central income tax, the percentage change in the retention rate moves from $-\frac{\widehat{T}_j'\left(y_j\left(\theta_j\right)\right)}{1 - T_j'\left(y_j\left(\theta_j\right)\right)}$ to $-\frac{\widehat{T}_j'\left(y_j\left(\theta_j\right)\right)}{1 - T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}$ and the original expression changes into (3).

Utility and Social Welfare Effect. A first-order Taylor expansion of the perturbed gross utility, $V_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right) = y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right) - T_j^A\left(y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right); T_j + \mu_j \widehat{T}_j\right) - \phi'\left(\frac{y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right)}{A_j \theta_j}\right)$, around the baseline gross utility

[39] See Luenberger (1997) for details about the Gateaux derivative.

$V_j\left(\theta_j; T_j\right)$ yields:

$$\widehat{V}_j\left(\theta_j\right) = \left[1 - T_j^{A\prime}\left(y_j\left(\theta_j\right)\right) - \phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}\right]\widehat{y}_j\left(\theta_j\right) - \widehat{T}_j\left(y_j\left(\theta_j\right)\right). \tag{A4}$$

The individual's first order condition (2) implies that the first term on the right side of (A4) is zero, so that we obtain:

$$\widehat{V}_j\left(\theta_j\right) = -\widehat{T}_j\left(y_j\left(\theta_j\right)\right). \tag{A5}$$

Intuitively, taking the derivative of $V_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)$ with respect to μ_j, and using the envelop theory, we have $\widehat{V}_j\left(\theta_j\right) = -\widehat{T}_j\left(y_j\left(\theta_j\right)\right)$ when $\mu_j \to 0$. This is because under the agent's optimal reaction, the marginal tax reform has no first-order effect on the gross utility through its effort on the choice of agent. Thus, the change in the gross utility equals the change in the tax burden under the initial labor income. Last, the change in social welfare is a summation of $\widehat{V}_j\left(\theta\right)$, i.e., $\widehat{W}_j = \int_{\Theta_j}\widehat{V}_j\left(\theta_j\right)\widetilde{f}_j\left(\theta_j\right)d\theta_j$.

A.2 Proof of Lemma 1

Using the first-order condition under the baseline tax system is

$$1 - T'\left(y_j\left(\theta_j\right)\right) - T_j'\left(y_j\left(\theta_j\right)\right) = \phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}.$$

The perturbed individual first-order condition of individual θ_j is:

$$\begin{aligned} & 1 - T'\left(y_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)\right) - T_j'\left(y_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)\right) - \mu_j\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)\right) \\ = \; & \phi'\left(\frac{y_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}, \end{aligned}$$

where $\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)\right) = \widehat{T}_j'\left(y_j\left(\theta_j; T_j + \mu_j\widehat{T}_j\right)\right)$ because the local government treats the central income tax as given.

By the first-order Taylor expansion, one has:

$$-T''\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right) - T_j''\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right) - \widehat{T}_j^{A\prime}\left(y_j\left(\theta_j\right)\right) = \phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}\frac{1}{A_j\theta_j}\widehat{y}_j\left(\theta_j\right),$$

where $\widehat{y}_j\left(\theta\right)$ is a Gateaux derivative.

Thus, the change in labor income is:

$$\widehat{y}_j\left(\theta_j\right) = -\frac{\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}{\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}\frac{1}{A_j\theta_j} + T''\left(y_j\left(\theta_j\right)\right) + T_j''\left(y_j\left(\theta_j\right)\right)}.$$

Dividing both the numerator and denominator on the right side of the equation above by $\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j y_j\left(\theta_j\right)}$

or $\frac{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}{y_j\left(\theta_j\right)}$,[40] we have:

$$\begin{aligned}
\widehat{y}_j\left(\theta_j\right) &= -\frac{\frac{y_j\left(\theta_j\right)\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}}{\frac{\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}}{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}+\frac{y_j\left(\theta_j\right)T_j^{A\prime\prime}\left(y_j\left(\theta_j\right)\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}} \\
&= -\frac{\frac{y_j\left(\theta_j\right)\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}}{\frac{1}{\varepsilon_j}+p_j^A\left(y_j(\theta_j)\right)} \\
&= -\varepsilon_j^A\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\frac{\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)},
\end{aligned} \tag{A6}$$

where the second and third equations are derived from definitions of elasticity.

Remember that in the local government's problem, $\widehat{T}_j^{A\prime}(\cdot)=\widehat{T}_j'(\cdot)$. The last equation of (A6) implies Lemma 1.■

A.3 Proof of Lemma 2

Balanced Elementary Tax Reform. To derive the first-order necessary condition for the optimal local income tax rate, we consider a balanced elementary tax reform. A balanced elementary tax reform is constructed by an elementary tax reform on the marginal tax rates and the accompanying tax reform on lumpsum tax, which guarantees a balanced budget.[41] Formally, an elementary tax reform is represented by the following step function:

$$\widehat{T}_j\left(y_j;y^*\right)=1_{\left\{y_j\geq y^*\right\}}+\widehat{r}_j\left(y^*\right), \tag{A7}$$

in which $\widehat{r}_j\left(y^*\right)$ is a lumpsum tax used to guarantee a balanced budget. Under the tax reform above, we have

$$\widehat{T}_j'\left(y_j;y^*\right)=\delta\left(y_j-y^*\right), \tag{A8}$$

in which $\delta(\cdot)$ is a Dirac function.[42] The tax reform increases the tax on $y_j<y^*$ by $\widehat{r}_j\left(y^*\right)$ and $y_j\geq y^*$ by $1+\widehat{r}_j\left(y^*\right)$. For simplicity, we omit y^* in $\widehat{T}_j\left(y_j;y^*\right)$ and $\widehat{T}_j'\left(y_j;y^*\right)$ unless it leads to misunderstanding.

Denote the optimal local income tax given the central income tax by $T_j^\star:\mathbb{R}_+\mapsto\mathbb{R}$. If the initial local income tax is optimal (i.e., the initial local income tax is $T_j^\star(\cdot)$), any tax reform $\left\{\widehat{T}_j\left(y_j;y^*\right)\right\}_{y^*\in R}$, results in $\widehat{W}_j=0$. This is the key to solving for the optimal tax formula, with which one can obtain a necessary condition for optimal local income tax.

---

[40] Note that the agent's first order condition implies $\phi_j'\left(\frac{y_j(\theta)}{\theta}\right)\frac{1}{\theta y_j(\theta)}=\frac{1-T_j^{A\prime}\left(y_j(\theta)\right)}{y_j(\theta)}$.

[41] Saez and Stantcheva (2016) also considers a balanced tax reform. The elementary tax reform follows from Sachs et al. (2020).

[42] Although $\widehat{T}_j(y)$ is not a differentiable function, we can focus on it without losing generality because the elementary tax reforms can be approached with certain differentiable tax reforms (see, e.g., Gomes et al. (2018)). To see this, define $\widehat{H}_j\left(y;\delta\right)$ as below with parameter $\delta>0$: $\widehat{H}_j'\left(y;\delta\right)=\begin{cases}\frac{3}{2\delta}\left(\frac{y-y^*}{\delta}\right)^2, y^*-\delta\leq y\leq y^*+\delta;\\ 0, \text{otherwise.}\end{cases}$ and $\widehat{H}_j\left(y;\delta\right)=\int_{\underline{y}}^{y}\widehat{H}_j'\left(s;\delta\right)ds+\widehat{r}_j$. Then we have $\lim_{\delta>0,\delta\to0}\widehat{H}_j\left(y_j;\delta\right)=\widehat{T}_j\left(y_j\right)=1_{\{y\geq y^*\}}+\widehat{r}_j$, and $\lim_{\delta>0,\delta\to0}\widehat{H}_j'\left(y_j;\delta\right)=\widehat{T}_j'\left(y_j\right)=\delta\left(y_j-y^*\right)$, which characterize the elementary tax reform we considered.

Local Budget Balance. A first-order Taylor expansion of local government's perturbed budget constraint (A3) around the baseline budget constraint, $\int_{\Theta_j} T_j\left(y_j\left(\theta_j;T_j\right)\right) f_{\theta_j}\left(\theta_j\right) d\theta_j = E_j$, yields (A9):

$$0 = \int_{\Theta_j} \left[T_j'\left(y_j\left(\theta_j\right)\right) \widehat{y}_j\left(\theta_j\right) + \widehat{T}_j\left(y_j\left(\theta_j\right)\right)\right] f_{\theta_j}\left(\theta_j\right) d\theta_j. \tag{A9}$$

Local Social Welfare. A first-order Taylor expansion of the perturbed social welfare,

$$W_j\left(T_j + \mu_j \widehat{T}_j\right) = \int_{\Theta_j} V_j\left(y_j\left(\theta_j; T_j + \mu_j \widehat{T}_j\right); T_j + \mu_j \widehat{T}_j\right) \widetilde{f}_j\left(\theta_j\right) d\theta_j,$$

around the baseline social welfare $W_j\left(T_j\right)$ yields:

$$\widehat{W}_j = \int_{\Theta_j} \widehat{V}_j\left(\theta\right) \widetilde{f}_j\left(\theta_j\right) d\theta_j = -\int_{\Theta_j} \widehat{T}_j\left(y_j\left(\theta_j\right)\right) \widetilde{f}_j\left(\theta_j\right) d\theta_j. \tag{A10}$$

where the second equation is derived by the Envelope theorem (see, e.g., (A5)).

Optimal Local Income Tax Rate. Now, we consider the balanced elementary tax reform. For later use, define $\theta_j\left(y_j\right) : \mathbb{R}_+ \mapsto \mathbb{R}_+$ as the skill of agents who earn y_j in region j. We have $y_j\left(\theta_j\left(y\right)\right) = y$. $\theta_j\left(y_j\right)$ is well defined under our monotonicity assumption on $y_j\left(\theta_j\right)$, and we have $\theta_j'\left(y_j\right) > 0$. In particular, set $\theta_j^* \equiv \theta_j\left(y_j^*\right)$.

Under the balanced elementary tax reform, for any y^*, we have $\widehat{T}_j\left(y_j; y^*\right) = 1_{\left\{y_j \geq y^*\right\}} + \widehat{r}_j\left(y^*\right)$ and $\widehat{T}_j'\left(y_j; y^*\right) = \delta\left(y_j - y^*\right)$ (see, e.g., (A7) and (A8)). Substitute $\widehat{T}_j'\left(y_j\right)$ in (A6) with $\delta\left(y_j - y^*\right)$:

$$\widehat{y}_j\left(\theta_j\right) = -\varepsilon_j^r\left(y_j\left(\theta_j\right)\right) y_j\left(\theta_j\right) \frac{\delta\left(y_j\left(\theta_j\right) - y^*\right)}{1 - T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}, \forall \theta_j \in \Theta_j. \tag{A11}$$

Substitute $\widehat{y}_j\left(\theta\right)$ and $\widehat{T}_j\left(y_j\left(\theta\right)\right)$ in (A9) with (A11) and (A8) respectively:

$$\int_{\Theta_j} \left[-\frac{\theta_j T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)}{\left(1 + \frac{1}{\varepsilon_j}\right)\left[1 - T'\left(y_j\left(\theta_j\right)\right) - T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)\right]} \delta\left(y\left(\theta_j\right) - y_j\left(\theta_j^*\right)\right) + 1_{\left\{y_j \geq y_j\left(\theta_j^*\right)\right\}} + \widehat{r}_j\left(y^*\right)\right] f_{\theta_j}\left(\theta_j\right) d\theta_j = 0.$$

Solve for $\widehat{r}_j\left(y^*\right)$ with the equation above:

$$\widehat{r}_j\left(y^*\right) = \frac{f_j\left(\theta_j^*\right) \theta_j^* T_j^{\star\prime}\left(y_j^*\right)}{\left(1 + \frac{1}{\varepsilon_j}\right)\left[1 - T'\left(y_j^*\right) - T_j^{\star\prime}\left(y_j^*\right)\right]} - \left[1 - F_j\left(\theta_j^*\right)\right].$$

The above equation holds for any balanced elementary tax reform. Substitute $\widehat{r}_j\left(y^*\right)$ in (A7) with the above equation:

$$\widehat{T}_j\left(y_j\right) = 1_{\left\{y_j \geq y^*\right\}} - 1 - F_j\left(\theta_j^*\right) + \frac{\theta_j^* f_j\left(\theta_j^*\right) T_j^{\star\prime}\left(y^*\right)}{\left(1 + \frac{1}{\varepsilon_j}\right)\left[1 - T'\left(y^*\right) - T_j^{\star\prime}\left(y^*\right)\right]}. \tag{A12}$$

Substitute $\widehat{T}_j\left(y_j\right)$ in (A10) with (A12):

$$\widehat{W}_j=\int_{\theta_j\geq\theta_j^*}\left[f_j\left(\theta_j\right)-\widetilde{f}_j\left(\theta_j\right)\right]d\theta_j+\frac{\theta_j^*f_j\left(\theta_j^*\right)T_j^{\star\prime}\left(y_j\left(\theta_j^*\right)\right)}{\left(1+\frac{1}{\varepsilon_j}\right)\left[1-T'\left(y_j\left(\theta_j^*\right)\right)-T_j^{\star\prime}\left(y_j\left(\theta_j^*\right)\right)\right]}.$$

The equation above holds for any y^* and $\theta_j^*=\theta_j\left(y^*\right)$. At the optimum, any balanced elementary tax reform has no first-order effect on the regional welfare (i.e., $\widehat{W}_j=0$), which gives:

$$\frac{T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta_j\right)\right)-T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)}=\frac{\widetilde{F}_{\theta_j}\left(\theta_j\right)-F_{\theta_j}\left(\theta_j\right)}{\theta_jf_j\left(\theta_j\right)}\left(1+\frac{1}{\varepsilon_j}\right),\forall\theta_j\in\Theta_j,j\in\mathcal{J}. \tag{A13}$$

Using the notation $\tau_j^\star\left(\theta\right)$, we can rewrite the equation above as (9). Combination of $T_j^{A\prime}(y_j)=T'\left(y_j\right)+T_j'\left(y_j\right)$ and (9) gives (10).

Notice that:

$$\frac{T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta_j\right)\right)-T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)}=\frac{\widetilde{F}_j\left(y_j\left(\theta_j\right)\right)-F_j\left(y_j\left(\theta_j\right)\right)}{1-F_j\left(y_j\left(\theta_j\right)\right)}\frac{1-F_j\left(y_j\left(\theta_j\right)\right)}{y_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)}\frac{y_j\left(\theta_j\right)}{y_j'\left(\theta_j\right)\theta_j}\left(1+\frac{1}{\varepsilon_j}\right) \tag{A14}$$

and (see the following equation (A16)):

$$\frac{1}{\frac{1}{\varepsilon_j}+1}\frac{y_j'\left(\theta_j\right)\theta_j}{y_j\left(\theta_j\right)}=\varepsilon_j^A\left(y_j\left(\theta_j\right)\right). \tag{A15}$$

Last, combination of equations (A14) and (A15) gives (5).

To see why (A15) holds, take the derivative of both sides of individual's first-order condition, i.e., $1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)=\frac{1}{A_j\theta_j}\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)$, with respect to θ_j:

$$-T_j^{A\prime\prime}\left(y_j\left(\theta_j\right)\right)y_j'\left(\theta_j\right)=-\frac{1}{A_j\theta_j^2}\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)+\frac{1}{A_j\theta_j}\phi''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\left[\frac{y_j'\left(\theta_j\right)}{A_j\theta_j}-\frac{y_j\left(\theta_j\right)}{A_j\theta_j^2}\right].$$

Solve the above equation of (A16) for $\frac{y_j'\left(\theta_j\right)\theta_j}{y_j}$:

$$\begin{aligned}\frac{y_j'\left(\theta_j\right)\theta_j}{y_j\left(\theta_j\right)} &= \frac{\frac{\phi''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}}{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}+1}{\frac{T_j^{A\prime\prime}\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}+\frac{\phi''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}}{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}} \\ &= \frac{1+\frac{1}{\varepsilon_j}}{p_j^A\left(y_j(\theta_j)\right)+\frac{1}{\varepsilon_j}}.\end{aligned} \tag{A16}$$

By the definition of $p_j^A\left(y_j(\theta_j)\right)$ and $\varepsilon_j^A\left(y_j\left(\theta_j\right)\right)$, the above equation implies (A15).■

A.4 Proof of Lemma 3

Equilibrium. To derive the incidence of labor, one should take the reaction of local income tax to the central income tax into consideration. In the equilibrium, the reaction of labor income is determined by both changes in the central and local income tax. The following equation system describes the MGTE after the tax reform:

$$T_j^{\star\prime}\left(y_j\left(\theta;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)=\left[\begin{array}{c}1-T'\left(y_j\left(\theta;T+\mu\widehat{T}\right)\right)\\-\mu\widehat{T}'\left(y_j\left(\theta;T+\mu\widehat{T}\right)\right)\end{array}\right]\tau_j^{\star}\left(\theta\right), \tag{A17}$$

and

$$\int_{\Theta_j}T_j^{\star}\left(y_j\left(\theta;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)f_j\left(\theta_j\right)d\theta_j=E_j, \tag{A18}$$

and

$$\left[1-T_j^{\star\prime}\left(y_j\left(\theta;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)\right]\left[1-\tau_j^{\star}\left(\theta\right)\right]=\frac{\phi'\left(\frac{y_j\left(\theta;T+\mu\widehat{T}\right)}{A_j\theta_j}\right)}{A_j\theta_j}, \tag{A19}$$

and

$$\sum_j N_j\int_{\Theta_j}\left[\begin{array}{c}T\left(y_j\left(\theta;T+\mu\widehat{T}\right)\right)+\\\mu\widehat{T}\left(y_j\left(\theta;T+\mu\widehat{T}\right)\right)\end{array}\right]f_j\left(\theta_j\right)d\theta_j+\sum_j N_jE_j=0,\forall\theta\in\Theta_j,j\in\mathcal{J}. \tag{A20}$$

The optimal reactions of local governments imply the first streams of equations. The budget balance of the local government requires the second equation. The third stream of equations describes the taxpayers' behaviors. Last, the budget balance of the central government requires the fourth equation.

Tax Incidence of Labor Income. Lemma 2 immediately implies:

$$\begin{aligned}&T_j^{\star\prime}\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right);T_j+\mu_j\widehat{T}_j\right)\\=\;&\left[1-T'\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right)\right)-\mu\widehat{T}'\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right)\right)\right]\tau_j^{\star}\left(\theta_j\right).\end{aligned} \tag{A21}$$

By definition $\widehat{T}_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)=\lim_{\mu\to0}\frac{T_j^{\star\prime}\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right);T_j+\mu_j\widehat{T}_j\right)-T_j^{\star\prime}\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right);T_j\right)}{\mu}$ and the above equation, we have (16), i.e., $\widehat{T}_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)=-\widehat{T}'\left(y_j\left(\theta_j\right)\right)\tau_j^{\star}\left(\theta_j\right)$.

The perturbed individual first-order condition of individual θ is expressed as:

$$\left[1-T'\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right)\right)-\mu\widehat{T}'\left(y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right)\right)\right]\left[1-\tau_j^{\star}\left(\theta_j\right)\right]=\phi'\left(\frac{y_j\left(\theta_j;T_j+\mu_j\widehat{T}_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j},$$

where the left side is the retention rate taking the local government's reaction into consideration. Using the first-order condition in the baseline tax system, i.e., $\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]\left[1-\tau_j^{\star}\left(\theta_j\right)\right]=\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}$, we reach a first order change of labor income by first-order Taylor expansion and the defined Gateaux derivative $\widehat{y}_j\left(\theta\right)$:

$$\left[-T''\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)-\widehat{T}'\left(y_j\left(\theta_j\right)\right)\right]\left[1-\tau_j^{\star}\left(\theta_j\right)\right]=\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{\widehat{y}_j\left(\theta_j\right)}{A_j\theta_j}\frac{1}{\theta_j}.$$

Equivalently:

$$\widehat{y}_j\left(\theta_j\right) = -\frac{\left[1-\tau_j^{\star}\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{\left[1-\tau_j^{\star}\left(\theta_j\right)\right]T''\left(y_j\left(\theta_j\right)\right)+\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}\frac{1}{\theta_j}}. \tag{A22}$$

Dividing both the numerator and denominator on the right side of the equation above by $\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j y_j\left(\theta_j\right)}$ or $\frac{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}{y_j\left(\theta_j\right)}$,[43] we have:

$$\begin{aligned}
\widehat{y}_j\left(\theta_j\right) &= -\frac{\frac{y_j\left(\theta_j\right)\left[1-\tau_j^{\star}\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}}{\frac{\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}}{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}+\frac{y_j\left(\theta_j\right)\left[1-\tau_j^{\star}\left(\theta_j\right)\right]T''\left(y_j\left(\theta_j\right)\right)}{1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right)}} \\
&= -\frac{\frac{y_j(\theta)\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta_j\right)\right)}}{\frac{\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}}{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}+\frac{y_j(\theta)T''\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j(\theta)\right)}} \\
&= -\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\frac{\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta_j\right)\right)}.
\end{aligned} \tag{A23}$$

The second equation of (A23) is derived by (10), and the third equation is derived by definitions of elasticities, i.e., $p^{C}\left(y_j(\theta_j)\right) = \frac{y_j\left(\theta_j\right)T''\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta_j\right)\right)}$, $\frac{1}{\varepsilon_j} = \frac{\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}}{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}$ and $\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right) = \frac{1}{\frac{1}{\varepsilon_j}+p_j^{C}\left(y_j(\theta_j)\right)}$. The last equation in (A23) gives equation (14) of Lemma 3.■

B Centralized Redistribution

B.1 Central Tax Incidence

Tax Burden. In the above analysis, we derive the reaction of local income tax. We can further derive the change of tax burden:

Lemma 5 *For any $j \in \mathcal{J}$ and $\theta \in \Theta_j$, we have*

$$\widehat{T}_j^{\star}\left(y_j\left(\theta\right)\right) = \widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}\right)\right) + \int_{\underline{\theta}}^{\theta}\widehat{T}_j^{\star\prime}\left(y_j\left(x\right)\right)dy_j\left(x\right). \tag{B1}$$

and

$$0 = \int_{\Theta_j}T_j^{\star\prime}\left(y_j\left(\theta\right)\right)\widehat{y}_j\left(\theta_j\right)f_j\left(\theta_j\right)d\theta_j + \int_{\Theta_j}\widehat{T}_j^{\star}\left(y_j\left(\theta\right)\right)f_j\left(\theta_j\right)d\theta_j, \tag{B2}$$

where $\frac{\widehat{y}_j(\theta)}{y_j(\theta)} = -\varepsilon_j^{C}\left(y_j\left(\theta\right)\right)\frac{\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j(\theta)\right)}$.

Proof. See Appendix B.2. ■

[43] Note that the agent's first order condition implies $\phi_j'\left(\frac{y_j(\theta)}{\theta}\right)\frac{1}{\theta y_j(\theta)} = \frac{1-T_j^{A\prime}\left(y_j(\theta)\right)}{y_j(\theta)}$.

Equation (B1) shows that the change in tax burden at a specific income point is equal to the change in the marginal tax rate of income below that income point and the change in tax burden at the lowest income point.

Equation (B2) is derived by a first-order Taylor expansion of both sides of the local government's budget constraint (A18). It says that the overall tax changes come from two parts: Given individual income, the tax burden caused by changes in the tax function, and the tax changes caused by the impact of tax reform on individual income. The above two equations together determine $\widehat{T}_j^{\star}(\cdot)$.

Utility and Welfare. A first-order Taylor expansion of the perturbed gross utility $V_j\left(\theta_j; T+\mu\widehat{T}\right)$ around the baseline gross utility, $V_j(\theta; T)$, yields:

$$\widehat{V}_j(\theta_j) = \left[1 - T'\left(y_j(\theta_j)\right) - T_j^{\star\prime}\left(y_j(\theta_j)\right) - \frac{\phi'\left(l_j(\theta_j)\right)}{A_j\theta_j}\right]\widehat{y}_j(\theta_j) - \left[\widehat{T}\left(y_j(\theta_j)\right) + \widehat{T}_j^{\star}\left(y_j(\theta_j)\right)\right], \tag{B3}$$

where the first-order conditions imply $\phi'\left(\frac{y_j(\theta_j)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j} = T^{A\prime}\left(y_j(\theta_j)\right)$. Therefore,

$$\widehat{V}_j(\theta_j) = -\widehat{T}\left(y_j(\theta_j)\right) - \widehat{T}_j^{\star}\left(y_j(\theta_j)\right). \tag{B4}$$

The change in gross utility equals the sum of the adjustment in the central tax and the resulting change in the local tax. Last, the national welfare change is simply a weighted average of changes in gross utilities:

$$\widehat{W} = \sum\nolimits_j \left[N_j \int_{\Theta_j} \widehat{V}_j(\theta)\, \widetilde{f}_{\theta_j}^{C}(\theta)\, d\theta_j\right]. \tag{B5}$$

B.2 Proof of Lemma 5

First, the change of individual tax burden satisfies:

$$\begin{aligned}
& \lim_{\mu\to 0} \frac{T_j^{\star}\left(y_j\left(\theta_j; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star}\left(y_j(\theta_j; T); T\right)}{\mu} \\
= & \lim_{\mu\to 0} \frac{T_j^{\star}\left(y_j\left(\theta_j; T+\mu\widehat{T}\right); T\right) - T_j^{\star}\left(y_j(\theta_j; T); T\right)}{\mu} \\
& + \lim_{\mu\to 0} \frac{T_j^{\star}\left(y_j\left(\theta_j; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star}\left(y_j\left(\theta_j; T+\mu\widehat{T}\right); T\right)}{\mu} \\
= & T_j^{\star\prime}\left(y_j(\theta_j)\right)\widehat{y}_j(\theta_j) + \widehat{T}_j^{\star}\left(y_j(\theta_j)\right).
\end{aligned}$$

Similarly, we have:

$$\lim_{\mu\to 0} \frac{T_j^{\star\prime}\left(y_j\left(\theta_j; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star\prime}\left(y_j(\theta_j; T); T\right)}{\mu} = \widehat{T}_j^{\star\prime}\left(y_j(\theta_j)\right) + T_j^{\star\prime\prime}\left(y_j(\theta_j)\right)\widehat{y}_j(\theta_j).$$

Besides,

$$T_j^\star\left(y_j\left(\theta_j+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j\left(y_j\left(\underline{\theta}_j+\mu\widehat{T}\right);T+\mu\widehat{T}\right)=\int_{\underline{\theta}_j}^{\theta_j}\frac{\partial T_j^\star\left(y_j\left(x+\mu\widehat{T}\right);T+\mu\widehat{T}\right)}{\partial y_j}dy_j\left(x,T+\mu\widehat{T}\right), \quad \text{(B6)}$$

where

$$T_j^\star\left(y_j\left(\theta_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)=T_j^\star\left(y_j\left(\theta_j;T\right)\right)+\left[T_j^\star\left(y_j\left(\theta_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^\star\left(y_j\left(\theta_j;T\right)\right)\right], \quad \text{(B7)}$$

and

$$T_j^\star\left(y_j\left(\underline{\theta}_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)=T_j^\star\left(y_j\left(\underline{\theta}_j;T\right)\right)+\left[T_j^\star\left(y_j\left(\underline{\theta}_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^\star\left(y_j\left(\underline{\theta}_j;T\right)\right)\right]. \quad \text{(B8)}$$

Similarly, we have:

$$T_j^\star\left(y_j\left(\theta_j\right)\right)-T_j^\star\left(y_j\left(\underline{\theta}_j\right)\right)=\int_{\underline{\theta}_j}^{\theta_j}\frac{\partial T_j^\star\left(y_j\left(x\right)\right)}{\partial y_j}dy_j\left(x\right). \quad \text{(B9)}$$

Combination of (B6) and (B9) delivers:

$$\begin{aligned}&\int_{\underline{\theta}_j}^{\theta_j}\frac{\partial T_j^\star\left(y_j\left(x+\mu\widehat{T}\right);T+\mu\widehat{T}\right)}{\partial y_j}dy_j\left(x,T+\mu\widehat{T}\right)-\int_{\underline{\theta}_j}^{\theta_j}\frac{\partial T_j^\star\left(y_j\left(x\right)\right)}{\partial y_j}dy_j\left(x\right)\\ =\;&\left[T_j^\star\left(y_j\left(\theta_j+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^\star\left(y_j\left(\underline{\theta}_j+\mu\widehat{T}\right);T+\mu\widehat{T}\right)\right]-\left[T_j^\star\left(y_j\left(\theta_j\right)\right)-T_j^\star\left(y_j\left(\underline{\theta}_j\right)\right)\right].\end{aligned}$$

Rearrange the above equation:

$$\begin{aligned}&\int_{\underline{\theta}_j}^{\theta_j}\frac{\frac{\partial T_j^\star\left(y_j\left(x+\mu\widehat{T}\right);T+\mu\widehat{T}\right)}{\partial y_j}-\frac{\partial T_j^\star\left(y_j\left(x\right)\right)}{\partial y_j}}{\mu}dy_j\left(x,T+\mu\widehat{T}\right)+\int_{\underline{\theta}_j}^{\theta_j}\frac{\partial T_j^\star\left(y_j\left(x\right)\right)}{\partial y_j}d\left[\frac{y_j\left(x,T+\mu\widehat{T}\right)-y_j\left(x\right)}{\mu}\right]\\ =\;&\frac{T_j^\star\left(y_j\left(\theta_j+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^\star\left(y_j\left(\theta_j\right)\right)}{\mu}-\frac{T_j^\star\left(y_j\left(\underline{\theta}_j+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^\star\left(y_j\left(\underline{\theta}_j\right)\right)}{\mu}.\end{aligned}$$

As $\mu\to 0$, we have:

$$\begin{aligned}&\int_{\underline{\theta}_j}^{\theta_j}\left[\widehat{T}_j^{\star\prime}\left(y_j\left(x\right)\right)+T_j^{\star\prime\prime}\left(y_j\left(x\right)\right)\widehat{y}_j\left(x\right)\right]dy_j\left(x\right)+\int_{\underline{\theta}_j}^{\theta_j}T_j^{\star\prime}\left(y_j\left(x\right)\right)d\widehat{y}_j\left(x\right) \quad \text{(B10)}\\ =\;&\widehat{T}_j^\star\left(y_j\left(\theta_j\right)\right)+T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)-\widehat{T}_j^\star\left(y_j\left(\underline{\theta}_j\right)\right)-T_j^{\star\prime}\left(y_j\left(\underline{\theta}_j\right)\right)\widehat{y}_j\left(\underline{\theta}_j\right).\end{aligned}$$

Notice that

$$\begin{aligned}&T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)-T_j^{\star\prime}\left(y_j\left(\underline{\theta}_j\right)\right)\widehat{y}_j\left(\underline{\theta}_j\right) \quad \text{(B11)}\\ =\;&\int_{\underline{\theta}_j}^{\theta_j}d\left[T_j^{\star\prime}\left(y_j\left(x\right)\right)\widehat{y}_j\left(x\right)\right]\\ =\;&\int_{\underline{\theta}_j}^{\theta_j}\widehat{y}_j\left(x\right)dT_j^{\star\prime}\left(y_j\left(x\right)\right)+\int_{\underline{\theta}_j}^{\theta_j}T_j^{\star\prime}\left(y_j\left(x\right)\right)d\widehat{y}_j\left(x\right)\\ =\;&\int_{\underline{\theta}_j}^{\theta_j}T_j^{\star\prime\prime}\left(y_j\left(x\right)\right)\widehat{y}_j\left(x\right)dy_j\left(x\right)+\int_{\underline{\theta}_j}^{\theta_j}T_j^{\star\prime}\left(y_j\left(x\right)\right)d\widehat{y}_j\left(x\right).\end{aligned}$$

Combination of (B10) and (B11) delivers:

$$\widehat{T}_j^{\star}\left(y_j\left(\theta_j\right)\right)-\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)=\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}_j^{\star\prime}\left(y_j\left(x\right)\right)dy_j\left(x\right).$$

The local government's budget constraint (A18), i.e., $\int_{\Theta_j}T_j^{\star}\left(y_j\left(\theta_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)f_j\left(\theta_j\right)d\theta_j=E_j$, implies:

$$\int_{\Theta_j}\frac{T_j^{\star}\left(y_j\left(\theta_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\theta_j;T\right);T\right)}{\mu}f_j\left(\theta_j\right)d\theta_j=0$$

Therefore, we have:

$$\begin{aligned}0 &= \int_{\Theta_j}\frac{T_j^{\star}\left(y_j\left(\theta_j;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\theta_j;T\right);T+\mu\widehat{T}\right)}{\mu}f_j\left(\theta_j\right)d\theta_j\\ &+\int_{\Theta_j}\frac{T_j^{\star}\left(y_j\left(\theta_j;T\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\theta_j;T\right);T\right)}{\mu}f_j\left(\theta_j\right)d\theta_j\end{aligned}$$

When $\mu\to 0$, the above equation implies (B2).■

B.3 Proof of Proposition 2

Incidence of Labor Income. Equation (A23) already implies Lemma 3. For later use, we derive alternative expressions of $\widehat{y}_j\left(\theta\right)$. Taking differentiation of an individual's first-order condition with respect to the skill we have

$$\begin{aligned}&T''\left(y_j\left(\theta_j\right)\right)\left[1-\tau_j^{\star}\left(\theta_j\right)\right]y_j'\left(\theta_j\right)+\phi_j''\left(\frac{y_j\left(\theta_j\right)}{\theta_j}\right)\frac{1}{A_j\theta_j}\frac{1}{A_j\theta_j}y_j'\left(\theta_j\right)\\ =\;&\phi'\left(\frac{y_j\left(\theta_j\right)}{\theta_j}\right)\frac{1}{A_j\theta_j}\frac{1}{A_j\theta_j}+\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\frac{1}{A_j\theta_j}\frac{1}{\theta_j}+\\ &\frac{d\ln\left[1-\tau_j^{\star}\left(\theta_j\right)\right]}{d\ln\theta_j}\frac{\left[1-\tau_j^{\star}\left(\theta_j\right)\right]\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]}{\theta_j}.\end{aligned}$$

Utilizing $\frac{\phi_j''\left(\frac{y_j}{A_j\theta}\right)\frac{y_j}{A_j\theta}}{\phi'\left(\frac{y_j}{A_j\theta}\right)}=\frac{1}{\varepsilon_j}$, the above equation can be simplified as:

$$\begin{aligned}&\frac{\left[1+\frac{1}{\varepsilon_j}+\frac{d\ln\left[1-\tau_j^{\star}\left(\theta_j\right)\right]}{d\ln\theta_j}\right]\left[1-\tau_j^{\star}\left(\theta_j\right)\right]\left[1-T'\left(y_j\left(\theta\right)\right)\right]}{\frac{dy_j\left(\theta_j\right)}{d\theta_j}\theta_j}\\ =\;&\left[1-\tau_j^{\star}\left(\theta_j\right)\right]T''\left(y_j\left(\theta_j\right)\right)+\phi_j''\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\frac{1}{A_j\theta_j}\frac{1}{\theta_j}.\end{aligned}\tag{B12}$$

Utilize the above equation to substitute the denominator of (A22):

$$\widehat{y}_j\left(\theta_j\right)=-\frac{\widehat{T}'\left(y_j\left(\theta_j\right)\right)y_j'\left(\theta_j\right)\theta_j}{\left[1+\frac{1}{\varepsilon_j}+\frac{d\ln\left[1-\tau_j^{\star}\left(\theta_j\right)\right]}{d\ln\theta_j}\right]\left[1-T'\left(y_j\left(\theta\right)\right)\right]}. \tag{B13}$$

Central Budget Balance. A first-order Taylor expansion of the perturbed budget constraint of central government,

$$\sum_j N_j\int_{\Theta_j}\left[T\left(y_j\left(\theta_j;\mu\right)\right)+\mu\widehat{T}\left(y_j\left(\theta_j;\mu\right)\right)\right]f_j\left(\theta_j\right)d\theta_j+\sum_j N_jE_j=0,$$

around the baseline budget constraint,

$$\sum_j N_j\int_{\Theta_j}T\left(y_j\left(\theta_j\right)\right)f_j\left(\theta_j\right)d\theta_j+\sum_j N_jE_j=0,$$

yields:

$$\sum_j N_j\int_{\Theta_j}\left[T'\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)+\widehat{T}\left(y_j\left(\theta_j\right)\right)\right]f_j\left(\theta_j\right)d\theta_j=0.$$

By identity equation $\widehat{T}\left(y_j\left(\theta_j\right)\right)=\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right)+\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}'\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)$, the above equation can be transformed into:

$$\sum_j N_j\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right)+\sum_j N_j\int_{\Theta_j}\left[T'\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)+\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}'\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)\right]f_j\left(\theta_j\right)d\theta_j=0.$$

Remember that $\underline{y}=\min_j y_j\left(\underline{\theta}_j\right)$. One has $\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right)=\widehat{T}\left(\underline{y}\right)+\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)}\widehat{T}'\left(y_j\right)dy_j$. Therefore, the above equation gives

$$\begin{aligned}0 &= \sum_j N_j\widehat{T}\left(\underline{y}\right)+\sum_j N_j\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)}\widehat{T}'\left(y_j\right)dy_j+\sum_j N_j\int_{\Theta_j}T'\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)f_j\left(\theta_j\right)d\theta_j\\ &\quad+\sum_j N_j\int_{\Theta_j}\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}'\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)f_j\left(\theta_j\right)d\theta_j.\end{aligned}$$

Using the Fubini theorem, and $f_j\left(\theta_j\right)=f_j\left(y_j\left(\theta_j\right)\right)y_j'\left(\theta_j\right)$, the above equation is equivalent to:

$$\begin{aligned}0 &= \sum_j N_j\widehat{T}\left(\underline{y}\right)+\sum_j N_j\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)}\widehat{T}'\left(y_j\right)dy_j+\sum_j N_j\int_{\Theta_j}T'\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\ &\quad+\sum_j N_j\int_{\Theta_j}\left[1-F_j\left(y_j\left(\theta_j\right)\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right).\end{aligned}$$

Remember that $\widehat{y}_j\left(\theta_j\right)=-\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\frac{\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta\right)\right)}$. The above equation can be transformed into:

$$
\begin{aligned}
0 &= \sum_j N_j \widehat{T}\left(\underline{y}\right) + \sum_j N_j \int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)} \widehat{T}'\left(y_j\right) dy_j + \sum_j N_j \int_{\Theta_j} \left[1 - F_j\left(y_j\left(\theta_j\right)\right)\right] \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right) \\
&\quad - \sum_j N_j \int_{\Theta_j} f_j\left(y_j\left(\theta_j\right)\right) y_j\left(\theta_j\right) \varepsilon_j^C\left(y_j\left(\theta_j\right)\right) \frac{T'\left(y_j\left(\theta_j\right)\right)}{1 - T'\left(y_j\left(\theta\right)\right)} \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right),
\end{aligned}
$$

which means

$$
\begin{aligned}
\widehat{T}\left(\underline{y}\right) &= -\sum_j \frac{N_j}{\sum_j N_j} \int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)} \widehat{T}'\left(y_j\right) dy_j - \sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} \left[1 - F_j\left(\theta_j\right)\right] \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right) \\
&\quad + \sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} f_j\left(y_j\left(\theta_j\right)\right) y_j\left(\theta_j\right) \varepsilon_j^C\left(y_j\left(\theta_j\right)\right) \frac{T'\left(y_j\left(\theta_j\right)\right)}{1 - T'\left(y_j\left(\theta\right)\right)} \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right).
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right) &= \left[\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)} \widehat{T}'\left(y_j\right) dy_j - \sum_j \frac{N_j}{\sum_j N_j} \int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)} \widehat{T}'\left(y_j\right) dy_j\right] \\
&\quad - \sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} \left[1 - F_j\left(\theta_j\right)\right] \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right) \\
&\quad + \sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} f_j\left(y_j\left(\theta_j\right)\right) y_j\left(\theta_j\right) \varepsilon_j^C\left(y_j\left(\theta_j\right)\right) \frac{T'\left(y_j\left(\theta_j\right)\right)}{1 - T'\left(y_j\left(\theta\right)\right)} \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right).
\end{aligned} \tag{B14}
$$

Finally, we have

$$
\begin{aligned}
\widehat{T}\left(y_j\left(\theta_j\right)\right) &= \int_{\underline{\theta}_j}^{\theta_j} d\widehat{T}\left(y_j\left(x_j\right)\right) + \left[\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)} \widehat{T}'\left(y_j\right) dy_j - \sum_j \frac{N_j}{\sum_j N_j} \int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)} \widehat{T}'\left(y_j\right) dy_j\right] \\
&\quad - \sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} \left[1 - F_j\left(\theta_j\right)\right] \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right) \\
&\quad + \sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} f_j\left(y_j\left(\theta_j\right)\right) y_j\left(\theta_j\right) \varepsilon_j^C\left(y_j\left(\theta_j\right)\right) \frac{T'\left(y_j\left(\theta_j\right)\right)}{1 - T'\left(y_j\left(\theta\right)\right)} \widehat{T}'\left(y_j\left(\theta_j\right)\right) dy_j\left(\theta_j\right).
\end{aligned} \tag{B15}
$$

In the following analysis, we derive the optimal central income tax rate for each income interval.

Optimal Reaction of Local Governments. A first-order Taylor expansion of the perturbed budget constraint of local government,

$$
E_j = \int_{\Theta_j} T_j^\star\left(y_j\left(\theta_j; T + \mu\widehat{T}\right); T + \mu\widehat{T}\right) f_j\left(\theta_j\right) d\theta_j,
$$

around the baseline budget constraint,

$$
E_j = \int_{\Theta_j} T_j^\star\left(y_j\left(\theta_j; T\right); T\right) f_j\left(\theta_j\right) d\theta_j,
$$

On the other hand, the budget constraint implies:

$$
\int_{\Theta_j} \widehat{T}_j^\star\left(y_j\left(\theta_j\right)\right) f_j\left(\theta_j\right) d\theta_j + \int_{\Theta_j} T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right) \widehat{y}_j\left(\theta_j\right) f_j\left(\theta_j\right) d\theta_j = 0
$$

Remember that $\widehat{T}_j^{\star}\left(y_j\left(\theta_j\right)\right)=\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)+\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}_j^{\star\prime}\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)$. The above equation can be transformed into:

$$\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)+\int_{\Theta_j}\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}_j^{\star\prime}\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)f_j\left(\theta_j\right)d\theta_j+\int_{\Theta_j}T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)f_j\left(\theta_j\right)d\theta_j=0$$

Using the Fubini theorem, and $f_j\left(\theta_j\right)=f_j\left(y_j\left(\theta_j\right)\right)y_j'\left(\theta_j\right)$, we have:

$$\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)+\int_{\Theta_j}\left[1-F_j\left(y_j\left(\theta_j\right)\right)\right]\widehat{T}_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)+\int_{\Theta_j}T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)=0$$

By $T_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)=\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]\tau_j^{\star}\left(\theta_j\right)$, $\widehat{T}_j^{\star\prime}\left(y_j\left(\theta_j\right)\right)=-\widehat{T}'\left(y_j\left(\theta_j\right)\right)\tau_j^{\star}\left(\theta_j\right)$, and $\widehat{y}_j\left(\theta_j\right)=-\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\frac{\widehat{T}'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta\right)\right)}$, the above equation can be transformed into:

$$\begin{aligned}
&\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)\\
=&\int_{\Theta_j}\left[1-F_j\left(y_j\left(\theta_j\right)\right)\right]\tau_j^{\star}\left(\theta_j\right)d\widehat{T}\left(y_j\left(\theta_j\right)\right)-\int_{\Theta_j}\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]\tau_j^{\star}\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)\widehat{y}_j\left(\theta_j\right)dy_j\left(\theta_j\right)\\
=&\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\tau_j^{\star}\left(\theta_j\right)d\widehat{T}\left(y_j\left(\theta_j\right)\right)+\int_{\Theta_j}\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)\tau_j^{\star}\left(\theta_j\right)\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right).
\end{aligned}$$

Combination of the above equation and (B14) gives:

$$\begin{aligned}
&\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)+\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right) \qquad\qquad (B16)\\
=&\left[\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)}\widehat{T}'\left(y_j\right)dy_j-\sum_j\frac{N_j}{\sum_j N_j}\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)}\widehat{T}'\left(y_j\right)dy_j\right]\\
&+\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\tau_j^{\star}\left(\theta_j\right)d\widehat{T}\left(y_j\left(\theta_j\right)\right)-\sum_j\frac{N_j}{\sum_j N_j}\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&+\int_{\Theta_j}\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)\tau_j^{\star}\left(\theta_j\right)\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&+\sum_j\frac{N_j}{\sum_j N_j}\int_{\Theta_j}f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)\frac{T'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta\right)\right)}\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right).
\end{aligned}$$

Rearrange the above equation:

$$\begin{aligned}
&\widehat{T}_j^{\star}\left(y_j\left(\theta_j\right)\right)\\
=&\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\tau_j^{\star}\left(\theta_j\right)d\widehat{T}\left(y_j\left(\theta_j\right)\right)-\int_{\Theta_j}\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]\tau_j^{\star}\left(\theta_j\right)\widehat{y}_j\left(\theta_j\right)f_j\left(\theta_j\right)d\theta_j\\
&+\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}_j^{\star\prime}\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)\\
=&\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\tau_j^{\star}\left(\theta_j\right)d\widehat{T}\left(y_j\left(\theta_j\right)\right)-\int_{\Theta_j}\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]\tau_j^{\star}\left(\theta_j\right)\widehat{y}_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&-\int_{\underline{\theta}_j}^{\theta_j}\widehat{T}'\left(y_j\left(x_j\right)\right)\tau_j^{\star}\left(x_j\right)dy_j\left(x_j\right).
\end{aligned}$$

Substitute $\widehat{y}_j(\theta_j)$ in the above equation by $\widehat{y}_j(\theta_j) = -\varepsilon_j^C(y_j(\theta_j))\, y_j(\theta_j)\, \frac{\widehat{T}'(y_j(\theta_j))}{1-T'(y_j(\theta))}$:

$$\begin{aligned} & \widehat{T}_j^\star(y_j(\theta_j)) \\ = & \int_{\Theta_j}[1-F_j(\theta_j)]\,\tau_j^\star(\theta_j)\,d\widehat{T}(y_j(\theta_j)) - \int_{\underline{\theta}_j}^{\theta_j}\tau_j^\star(x_j)\,d\widehat{T}(y_j(x_j)) \\ & + \int_{\Theta_j}\tau_j^\star(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,y_j(\theta_j)\,f_j(y_j(\theta_j))\,d\widehat{T}(y_j(\theta_j)). \end{aligned} \tag{B17}$$

Remember that:

$$\begin{aligned} \tau_j^\star(\theta_j) &= \frac{\frac{\widetilde{F}_{\theta_j}(\theta_j)-F_{\theta_j}(\theta_j)}{f_{\theta_j}(\theta_j)\theta_j}\left(1+\frac{1}{\varepsilon_j}\right)}{1+\frac{\widetilde{F}_{\theta_j}(\theta_j)-F_{\theta_j}(\theta_j)}{f_{\theta_j}(\theta_j)\theta_j}\left(1+\frac{1}{\varepsilon_j}\right)} = \frac{\frac{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}{f_j(y_j(\theta_j))y_j(\theta_j)}\left(1+\frac{1}{\varepsilon_j}\right)\frac{y_j(\theta_j)}{y_j'(\theta_j)\theta_j}}{1+\frac{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}{f_j(y_j(\theta_j))y_j(\theta_j)}\left(1+\frac{1}{\varepsilon_j}\right)\frac{y_j(\theta_j)}{y_j'(\theta_j)\theta_j}} \\ &= \frac{\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}{1+\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}\Big|_{y_j=y_j(\theta_j)} = C_j(y_j(\theta_j)), \end{aligned} \tag{B18}$$

where we use $\frac{d\ln y_j(\theta_j)}{d\ln\theta_j} = \left(\frac{1}{\varepsilon_j}+1\right)\varepsilon_j^A(y_j(\theta_j))$ to derive the third equation. Therefore, we can substitute $\tau_j^\star(x_j)$ in (B17) by $C_j(y_j(\theta_j))$:

$$\begin{aligned} \widehat{T}_j^\star(y_j(\theta_j)) &= \int_{\Theta_j}[1-F_j(\theta_j)]\,C_j(y_j(\theta_j))\,d\widehat{T}(y_j(\theta_j)) - \int_{\underline{\theta}_j}^{\theta_j}C_j(y_j(x_j))\,d\widehat{T}(y_j(x_j)) \\ &\quad + \int_{\Theta_j}\varepsilon_j^C(y_j(\theta_j))\,y_j(\theta_j)\,f_j(y_j(\theta_j))\,C_j(y_j(\theta_j))\,d\widehat{T}(y_j(\theta_j)). \end{aligned} \tag{B19}$$

Utility and Social Welfare. Substitute $\widehat{T}_j^\star(y_j(\theta_j))$ and $\widehat{T}(y_j(\theta_j))$ in (B4) by (B1) and the identity equation $\widehat{T}(y_j(\theta_j)) = T\left(y_j\left(\underline{\theta}_j\right)\right) + \int_{\underline{\theta}_j}^{\theta_j}\widehat{T}'(y_j(x_j))\,dy_j(x_j)$:

$$\begin{aligned} \widehat{V}_j(\theta_j) &= -\widehat{T}(y_j(\theta_j)) - \widehat{T}_j^\star(y_j(\theta_j)) \\ &= -\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right) - \widehat{T}_j^\star\left(y_j\left(\underline{\theta}_j\right)\right) - \int_{\underline{\theta}_j}^{\theta_j}\widehat{T}'(y_j(x_j))\,dy_j(x_j) - \int_{\underline{\theta}_j}^{\theta_j}\widehat{T}_j^{\star\prime}(y_j(x_j))\,dy_j(x_j) \end{aligned} \tag{B20}$$

Use (16) and $\tau_j^\star(\theta_j) = C_j(y_j(\theta_j))$ to replace $\widehat{T}_j^{\star\prime}(y_j(x_j))$ in the above equation:

$$\widehat{V}_j(\theta_j) = -\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right) - \widehat{T}_j^\star\left(y_j\left(\underline{\theta}_j\right)\right) - \int_{\underline{\theta}_j}^{\theta_j}[1-C_j(y_j(x_j))]\,\widehat{T}'(y_j(x_j))\,dy_j(x_j)$$

Similar to the welfare effect of the local tax reform, we have $\widehat{W} = \sum_j N_j\left[\int_{\Theta_j}\widehat{V}_j(\theta_j)\,\widetilde{f}_j^C(\theta_j)\,d\theta_j\right]$. Substitute

$\widehat{V}_j\left(\theta_j\right)$ in $\widehat{W}$ with the above equation:

$$\begin{aligned}\widehat{W} &= -\sum_j N_j\rho_j\left[\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right)+\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)\right]\\ &\quad -\sum_j N_j\int_{\Theta_j}\int_{\underline{\theta}_j}^{\theta_j}\left[1-C_j\left(y_j\left(x_j\right)\right)\right]\widehat{T}'\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)\widetilde{f}_j^{C}\left(\theta_j\right)d\theta_j\end{aligned}$$

Using the Fubini theorem, and we know that $\sum_j N_j\rho_j=\sum_j N_j=1$ the equation above can be transformed into:

$$\begin{aligned}\widehat{W} &= -\sum_j \frac{N_j}{\sum_j N_j}\rho_j\left[\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right)+\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)\right]\\ &\quad -\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\left[1-C_j\left(y_j\left(\theta_j\right)\right)\right]\left[1-\widetilde{F}_j^{C}\left(y_j\left(\theta_j\right)\right)\right]d\widehat{T}\left(y_j\left(\theta_j\right)\right).\end{aligned} \tag{B21}$$

Multiply both sides of equation (B16) by $N_j\rho_j$ and sum across j:

$$\begin{aligned}&\sum_j N_j\rho_j\left[\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}_j\right)\right)+\widehat{T}\left(y_j\left(\underline{\theta}_j\right)\right)\right]\\ =\ &-\sum_j \frac{N_j}{\sum_j N_j}\left(1-\rho_j\right)\int_{\underline{y}}^{y_j\left(\underline{\theta}_j\right)}\widehat{T}'\left(y_j\right)dy_j\\ &+\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]C_j\left(y_j\left(\theta_j\right)\right)d\widehat{T}\left(y_j\left(\theta_j\right)\right)\\ &-\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\ &-\sum_j \frac{N_j\left(1-\rho_j\right)}{\sum_j N_j}\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\ &+\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)C_j\left(y_j\left(\theta_j\right)\right)\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\ &+\sum_j \frac{N_j}{\sum_j N_j}\int_{\Theta_j}f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^{C}\left(y_j\left(\theta_j\right)\right)\frac{T'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta\right)\right)}\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right).\end{aligned}$$

Rearrange the above equation:

$$
\begin{aligned}
&\sum_j N_j \rho_j \left[\widehat{T}_j^{\star} \left(y_j \left(\underline{\theta}_j \right) \right) + \widehat{T} \left(y_j \left(\underline{\theta}_j \right) \right) \right] \\
= \; & -\sum_j \frac{N_j}{\sum_j N_j} \left(1 - \rho_j \right) \int_{\underline{y}}^{y_j(\underline{\theta}_j)} \widehat{T}' \left(y_j \right) dy_j \\
& -\sum_j \frac{N_j}{\sum_j N_j} \left(1 - \rho_j \right) \int_{\Theta_j} \left[1 - F_j \left(\theta_j \right) \right] \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& -\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \left[1 - F_j \left(\theta_j \right) \right] \left[1 - C_j \left(y_j \left(\theta_j \right) \right) \right] d\widehat{T} \left(y_j \left(\theta_j \right) \right) \\
& +\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \varepsilon_j^C \left(y_j \left(\theta_j \right) \right) y_j \left(\theta_j \right) f_j \left(y_j \left(\theta_j \right) \right) C_j \left(y_j \left(\theta_j \right) \right) \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& +\sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} f_j \left(y_j \left(\theta_j \right) \right) y_j \left(\theta_j \right) \varepsilon_j^C \left(y_j \left(\theta_j \right) \right) \frac{T' \left(y_j \left(\theta_j \right) \right)}{1 - T' \left(y_j \left(\theta \right) \right)} \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right)
\end{aligned}
$$

Replacing $\sum_j N_j \rho_j \left[T_j^{\star} \left(y_j \left(\underline{\theta}_j \right) \right) + T \left(y_j \left(\underline{\theta}_j \right) \right) \right]$ in (B21) by the above equation:

$$
\begin{aligned}
\widehat{W} \; = \; & \sum_j \frac{N_j}{\sum_j N_j} \left(1 - \rho_j \right) \int_{\underline{y}}^{y_j(\underline{\theta}_j)} \widehat{T}' \left(y_j \right) dy_j \\
& +\sum_j \frac{N_j}{\sum_j N_j} \left(1 - \rho_j \right) \int_{\Theta_j} \left[1 - F_j \left(\theta_j \right) \right] \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& +\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \left[1 - F_j \left(\theta_j \right) \right] \left[1 - C_j \left(y_j \left(\theta_j \right) \right) \right] d\widehat{T} \left(y_j \left(\theta_j \right) \right) \\
& -\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \left[1 - C_j \left(y_j \left(\theta_j \right) \right) \right] \left[1 - \widetilde{F}_j^C \left(y_j \left(\theta_j \right) \right) \right] d\widehat{T} \left(y_j \left(\theta_j \right) \right) \\
& -\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \varepsilon_j^C \left(y_j \left(\theta_j \right) \right) y_j \left(\theta_j \right) f_j \left(y_j \left(\theta_j \right) \right) C_j \left(y_j \left(\theta_j \right) \right) \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& -\sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} f_j \left(y_j \left(\theta_j \right) \right) y_j \left(\theta_j \right) \varepsilon_j^C \left(y_j \left(\theta_j \right) \right) \frac{T' \left(y_j \left(\theta_j \right) \right)}{1 - T' \left(y_j \left(\theta \right) \right)} \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right)
\end{aligned}
$$

Rearrange the above equation:

$$
\begin{aligned}
\widehat{W} \; = \; & \sum_j \frac{N_j}{\sum_j N_j} \left(1 - \rho_j \right) \int_{\underline{y}}^{y_j(\underline{\theta}_j)} \widehat{T}' \left(y_j \right) dy_j \\
& +\sum_j \frac{N_j}{\sum_j N_j} \left(1 - \rho_j \right) \int_{\Theta_j} \left[1 - F_j \left(\theta_j \right) \right] \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& +\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \left[\widetilde{F}_j^C \left(y_j \left(\theta_j \right) \right) - F_j \left(y_j \left(\theta_j \right) \right) \right] \left[1 - C_j \left(y_j \left(\theta_j \right) \right) \right] \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& -\sum_j \frac{N_j}{\sum_j N_j} \rho_j \int_{\Theta_j} \varepsilon_j^C \left(y_j \left(\theta_j \right) \right) y_j \left(\theta_j \right) f_j \left(y_j \left(\theta_j \right) \right) C_j \left(y_j \left(\theta_j \right) \right) \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) \\
& -\sum_j \frac{N_j}{\sum_j N_j} \int_{\Theta_j} f_j \left(y_j \left(\theta_j \right) \right) y_j \left(\theta_j \right) \varepsilon_j^C \left(y_j \left(\theta_j \right) \right) \frac{T' \left(y_j \left(\theta_j \right) \right)}{1 - T' \left(y_j \left(\theta \right) \right)} \widehat{T}' \left(y_j \left(\theta_j \right) \right) dy_j \left(\theta_j \right) .
\end{aligned} \tag{B22}
$$

Notice that $-\frac{\widehat{y}_j(\theta_j)}{y_j(\theta_j)} = \varepsilon_j^A\left(y_j\left(\theta_j\right)\right)\frac{\widehat{T}_j^{A\prime}\left(y_j(\theta_j)\right)}{1-T_j^{A\prime}\left(y_j(\theta_j)\right)} = \varepsilon_j^C\left(y_j\left(\theta_j\right)\right)\frac{\widehat{T}'\left(y_j(\theta_j)\right)}{1-T'\left(y_j(\theta_j)\right)}$ (see Lemmas 1 and 3), $\widehat{T}_j^{A\prime}\left(y_j\left(\theta_j\right)\right) = \left[1-C_j\left(y_j\left(\theta_j\right)\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)$ (see equations (16) and (13)), and $1-T_j^{A\prime}\left(y_j\left(\theta_j\right)\right) = \left[1-C_j\left(y_j\left(\theta_j\right)\right)\right]\left[1-T'\left(y_j\left(\theta_j\right)\right)\right]$ (see Proposition 1). We have $\frac{\widehat{T}'\left(y_j(\theta_j)\right)}{1-T'\left(y_j(\theta_j)\right)} = \frac{\widehat{T}_j^{A\prime}\left(y_j(\theta_j)\right)}{1-T_j^{A\prime}\left(y_j(\theta_j)\right)}$ and $\frac{\varepsilon_j^C\left(y_j(\theta_j)\right)}{\varepsilon_j^A\left(y_j(\theta_j)\right)} = 1$. Therefore, equation (B22) can be rewritten as:

$$\begin{aligned}
\widehat{W} &= \sum_j \frac{N_j}{\sum_j N_j}\left(1-\rho_j\right)\int_{\underline{y}}^{y_j(\underline{\theta}_j)}\widehat{T}'\left(y_j\right)dy_j \qquad \text{(B23)}\\
&+\sum_j \frac{N_j}{\sum_j N_j}\left(1-\rho_j\right)\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&+\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\left[\widetilde{F}_j^C\left(y_j\left(\theta_j\right)\right)-F_j\left(y_j\left(\theta_j\right)\right)\right]\frac{\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)}{\varepsilon_j^A\left(y_j\left(\theta_j\right)\right)}\left[1-C_j\left(y_j\left(\theta_j\right)\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&-\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)C_j\left(y_j\left(\theta_j\right)\right)\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&-\sum_j \frac{N_j}{\sum_j N_j}\int_{\Theta_j}f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)\frac{T'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta\right)\right)}\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)
\end{aligned}$$

Notice that (see equation (B18)):

$$\begin{aligned}
1-C_j\left(y_j\left(\theta_j\right)\right) &= \frac{1}{1+\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}\Big|_{y_j=y_j(\theta_j)}\\
&= \frac{\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}{1+\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}\Big|_{y_j=y_j(\theta_j)}\\
&\quad\times\frac{f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)}{\widetilde{F}_j\left(y_j\left(\theta_j\right)\right)-F_j\left(y_j\left(\theta_j\right)\right)}\frac{\varepsilon_j^A\left(y_j\left(\theta_j\right)\right)}{\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)}\\
&= f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)\frac{1}{\widetilde{F}_j\left(y_j\left(\theta_j\right)\right)-F_j\left(y_j\left(\theta_j\right)\right)}\frac{\varepsilon_j^A\left(y_j\left(\theta_j\right)\right)}{\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)}C_j\left(y_j\left(\theta_j\right)\right).
\end{aligned}$$

Replacing $\left[1-C_j\left(y_j\left(\theta_j\right)\right)\right]$ in the right side of (B23), we have

$$\begin{aligned}
\widehat{W} &= \sum_j \frac{N_j}{\sum_j N_j}\left(1-\rho_j\right)\int_{\underline{y}}^{y_j(\underline{\theta}_j)}\widehat{T}'\left(y_j\right)dy_j+\sum_j \frac{N_j}{\sum_j N_j}\left(1-\rho_j\right)\int_{\Theta_j}\left[1-F_j\left(\theta_j\right)\right]\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&+\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)\frac{\widetilde{F}_j^C\left(y_j\left(\theta_j\right)\right)-F_j\left(y_j\left(\theta_j\right)\right)}{\widetilde{F}_j\left(y_j\left(\theta_j\right)\right)-F_j\left(y_j\left(\theta_j\right)\right)}C_j\left(y_j\left(\theta_j\right)\right)\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&-\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j}\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)f_j\left(y_j\left(\theta_j\right)\right)C_j\left(y_j\left(\theta_j\right)\right)\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right)\\
&-\sum_j \frac{N_j}{\sum_j N_j}\int_{\Theta_j}f_j\left(y_j\left(\theta_j\right)\right)y_j\left(\theta_j\right)\varepsilon_j^C\left(y_j\left(\theta_j\right)\right)\frac{T'\left(y_j\left(\theta_j\right)\right)}{1-T'\left(y_j\left(\theta\right)\right)}\widehat{T}'\left(y_j\left(\theta_j\right)\right)dy_j\left(\theta_j\right).
\end{aligned}$$

Collect the terms on the right side of the above equation:

$$\begin{aligned}\widehat{W} &= \sum_j \frac{N_j}{\sum_j N_j}(1-\rho_j)\int_{\underline{y}}^{y_j(\underline{\theta}_j)} \widehat{T}'(y_j)\,dy_j \\ &\quad -\sum_j \frac{N_j}{\sum_j N_j}\int_{\Theta_j} f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,\frac{T'(y_j(\theta_j))}{1-T'(y_j(\theta))}\widehat{T}'(y_j(\theta_j))\,dy_j(\theta_j) \\ &\quad +\sum_j \frac{N_j}{\sum_j N_j}\rho_j\int_{\Theta_j} f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,\frac{\widetilde{F}_j^C(y_j(\theta_j))-\widetilde{F}_j(y_j(\theta_j))}{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}C_j(y_j(\theta_j))\,\widehat{T}'(y_j(\theta_j))\,dy_j(\theta_j) \\ &\quad +\sum_j \frac{N_j}{\sum_j N_j}(1-\rho_j)\int_{\Theta_j}[1-F_j(\theta_j)]\,\widehat{T}'(y_j(\theta_j))\,dy_j(\theta_j)\end{aligned} \tag{B24}$$

Last, notice that

$$\begin{aligned}1 &= C_j(y_j(\theta_j))\frac{1}{C_j(y_j(\theta_j))} = C_j(y_j(\theta_j))\left[\frac{1+\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}{\frac{\widetilde{F}_j(y_j)-F_j(y_j)}{1-F_j(y_j)}\frac{1-F_j(y_j)}{f_j(y_j)y_j}\frac{1}{\varepsilon_j^A(y_j)}}\Big|_{y_j=y_j(\theta_j)}\right] \\ &= C_j(y_j(\theta_j))\left[1+\frac{f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^A(y_j(\theta_j))}{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}\right] \\ &= f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,\frac{\frac{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}{f_j(y_j(\theta_j))y_j(\theta_j)}\frac{1}{\varepsilon_j^C(y_j(\theta_j))}+\frac{\varepsilon_j^A(y_j(\theta_j))}{\varepsilon_j^C(y_j(\theta_j))}}{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}C_j(y_j(\theta_j)).\end{aligned} \tag{B25}$$

The last term on the right side of (B24) satisfies

$$\begin{aligned}&\sum_j \frac{N_j}{\sum_j N_j}(1-\rho_j)\int_{\Theta_j}[1-F_j(\theta_j)]\,\widehat{T}'(y_j(\theta_j))\,dy_j(\theta_j) \\ &= \sum_j \frac{N_j}{\sum_j N_j}(1-\rho_j)\int_{\Theta_j}\left[\begin{array}{c} f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,[1-F_j(\theta_j)]\times \\ \frac{\frac{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}{f_j(y_j(\theta_j))y_j(\theta_j)}\frac{1}{\varepsilon_j^C(y_j(\theta_j))}+\frac{\varepsilon_j^A(y_j(\theta_j))}{\varepsilon_j^C(y_j(\theta_j))}}{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}C_j(y_j(\theta_j)) \end{array}\right] d\widehat{T}(y_j(\theta_j)).\end{aligned}$$

Then, it is easy to see that equation (B24) can be rewritten as

$$\begin{aligned}\widehat{W} &= \sum_j \frac{N_j}{\sum_j N_j}(1-\rho_j)\int_{\underline{y}}^{y_j(\underline{\theta}_j)} d\widehat{T}(y_j) \\ &\quad -\sum_j \frac{N_j}{\sum_j N_j}\int_{\Theta_j} f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,\frac{T'(y_j(\theta_j))}{1-T'(y_j(\theta))}d\widehat{T}(y_j(\theta_j)) \\ &\quad +\sum_j \frac{N_j}{\sum_j N_j}\int_{\Theta_j}\left\{\left[\begin{array}{c} f_j(y_j(\theta_j))\,y_j(\theta_j)\,\varepsilon_j^C(y_j(\theta_j))\,C_j(y_j(\theta_j))\times \\ (1-\rho_j)\frac{[1-F_j(\theta_j)]\left[\frac{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))}{f_j(y_j(\theta_j))y_j(\theta_j)}\frac{1}{\varepsilon_j^C(y_j(\theta_j))}+\frac{\varepsilon_j^A(y_j(\theta_j))}{\varepsilon_j^C(y_j(\theta_j))}\right]}{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))} \\ +\rho_j\frac{\left[\widetilde{F}_j^C(y_j(\theta_j))-\widetilde{F}_j(y_j(\theta_j))\right]}{\widetilde{F}_j(y_j(\theta_j))-F_j(y_j(\theta_j))} \end{array}\right]\right\} d\widehat{T}(y_j(\theta_j)).\end{aligned} \tag{B26}$$

Optimal Tax Formula. Now, we consider the elementary balanced tax reform. Substitute $\widehat{T}\left(y_j\left(\theta_j\right)\right)$ in the above equation by $1_{\{y\geq y^*\}}+\widehat{r}\left(y^*\right)$ (remember that we have set $\theta_j^*=\widetilde{\theta}_j\left(y^*\right)$ as the skill of an individual whose income is y^*). Without loss of generality, we consider that $\underline{y}_1\leq\underline{y}_2$, and $\overline{y}_1\leq\overline{y}_2$, we can find the optimal rule by segment.

First consider a tax reform with $y^*\in\left(\underline{y}_2,\overline{y}_1\right)$. At the optimum, the marginal tax reform has a zero first order effect on the welfare ($\widehat{W}=0$). By equation (B26), we have:

$$\begin{aligned}&\frac{T'\left(y^*\right)}{1-T'\left(y^*\right)}\sum_j\frac{N_jf_j\left(y^*\right)y^*\varepsilon_j^C\left(y^*\right)}{\sum_jN_j}\\=&\sum_j\frac{N_jf_j\left(y^*\right)y^*\varepsilon_j^C\left(y^*\right)}{\sum_jN_j}\left\{\begin{array}{c}\left(1-\rho_j\right)\frac{\left[1-F_j(y^*)\right]\left[\frac{\widetilde{F}_j(y^*)-F_j(y^*)}{f_j(y^*)y^*}\frac{1}{\varepsilon_j^C(y^*)}+\frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}\right]}{\widetilde{F}_j(y^*)-F_j(y^*)}\\+\rho_j\frac{\left[\widetilde{F}_j^C(y^*)-\widetilde{F}_j(y^*)\right]}{\widetilde{F}_j(y^*)-F_j(y^*)}\end{array}\right\}C_j\left(y^*\right).\end{aligned}$$

Rearrange the above equation:

$$\frac{T'\left(y^*\right)}{1-T'\left(y^*\right)}=\sum_j\frac{N_jf_j\left(y^*\right)y^*\varepsilon_j^C\left(y^*\right)}{\sum_jN_jf_j\left(y^*\right)y^*\varepsilon_j^C\left(y^*\right)}\left\{\begin{array}{c}\rho_j\frac{\left[\widetilde{F}_j^C(y^*)-\widetilde{F}_j(y^*)\right]}{\widetilde{F}_j(y^*)-F_j(y^*)}+\\\left(1-\rho_j\right)\frac{\left[1-F_j(y^*)\right]\left[\frac{\widetilde{F}_j(y^*)-F_j(y^*)}{f_j(y^*)y^*}\frac{1}{\varepsilon_j^C(y^*)}+\frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}\right]}{\widetilde{F}_j(y^*)-F_j(y^*)}\end{array}\right\}C_j\left(y^*\right).$$

Notice that $T_j^{\star\prime}\left(y^*\right)=\left[1-T'\left(y^*\right)\right]C_j\left(y^*\right)$. We have

$$T'\left(y^*\right)=\sum_j\frac{N_jf_j\left(y^*\right)y^*\varepsilon_j^C\left(y^*\right)}{\sum_jN_jf_j\left(y^*\right)y^*\varepsilon_j^C\left(y^*\right)}\left\{\begin{array}{c}\rho_j\frac{\left[\widetilde{F}_j^C(y^*)-\widetilde{F}_j(y^*)\right]}{\widetilde{F}_j(y^*)-F_j(y^*)}+\\\left(1-\rho_j\right)\frac{\left[1-F_j(y^*)\right]\left[\frac{\widetilde{F}_j(y^*)-F_j(y^*)}{f_j(y^*)y^*}\frac{1}{\varepsilon_j^C(y^*)}+\frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}\right]}{\widetilde{F}_j(y^*)-F_j(y^*)}\end{array}\right\}T_j^{\star\prime}\left(y^*\right).\tag{B27}$$

Second, consider that $y^*\in\left[\underline{y}_1,\underline{y}_2\right]$. At the optimum, the marginal tax reform has a zero first order effect

on the welfare ($\widehat{W}=0$). By equation (B27), we have

$$\begin{aligned}
\frac{T'(y^*)}{1-T'(y^*)} &= \frac{N_1(1-\rho_1)}{N_1 f_1(y^*)\, y^* \varepsilon_1^C(y^*)} + \left\{ \begin{array}{c} \rho_1 \frac{\left[\widetilde{F}_1^C(y^*)-\widetilde{F}_1(y^*)\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} + \\ (1-\rho_1)\frac{[1-F_1(y^*)]\left[\frac{\widetilde{F}_1(y^*)-F_j(y^*)}{f_1(y^*)y^*}\frac{1}{\varepsilon_1^C(y^*)}+\frac{\varepsilon_1^A(y^*)}{\varepsilon_1^C(y^*)}\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} \end{array} \right\} C_1(y^*) \\
&= (1-\rho_1)\frac{N_1 f_1(y^*)\, y^* \varepsilon_1^C(y^*)}{N_1 f_1(y^*)\, y^* \varepsilon_1^C(y^*)} \frac{\frac{\widetilde{F}_1(y^*)-F_j(y^*)}{f_1(y^*)y^*}\frac{1}{\varepsilon_1^C(y^*)}+\frac{\varepsilon_1^A(y^*)}{\varepsilon_1^C(y^*)}}{\widetilde{F}_1(y^*)-F_1(y^*)} C_1(y^*) + \\
&\quad \left\{ \rho_1 \frac{\left[\widetilde{F}_1^C(y^*)-\widetilde{F}_1(y^*)\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} + (1-\rho_1)\frac{[1-F_1(y^*)]\left[\frac{\widetilde{F}_1(y^*)-F_j(y^*)}{f_1(y^*)y^*}\frac{1}{\varepsilon_1^C(y^*)}+\frac{\varepsilon_1^A(y^*)}{\varepsilon_1^C(y^*)}\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} \right\} C_1(y^*) \\
&= \left\{ \begin{array}{c} \rho_1 \frac{\left[\widetilde{F}_1^C(y^*)-\widetilde{F}_1(y^*)\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} + (1-\rho_1)\frac{[1-F_1(y^*)]\left[\frac{\widetilde{F}_1(y^*)-F_j(y^*)}{f_1(y^*)y^*}\frac{1}{\varepsilon_1^C(y^*)}+\frac{\varepsilon_1^A(y^*)}{\varepsilon_1^C(y^*)}\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} \\ + (1-\rho_1)\frac{\left[\frac{\widetilde{F}_1(y^*)-F_j(y^*)}{f_1(y^*)y^*}\frac{1}{\varepsilon_1^C(y^*)}+\frac{\varepsilon_1^A(y^*)}{\varepsilon_1^C(y^*)}\right]}{\widetilde{F}_1(y^*)-F_1(y^*)} \end{array} \right\} C_1(y^*).
\end{aligned}$$

where the second equation is derived by (B25).

Similarly, when $y^* \in [\overline{y}_1, \overline{y}_2]$, we have:

$$\frac{T'(y^*)}{1-T'(y^*)} = \left\{ \begin{array}{c} \rho_2 \frac{\left[\widetilde{F}_2^C(y^*)-\widetilde{F}_2(y^*)\right]}{\widetilde{F}_2(y^*)-F_2(y^*)} + (1-\rho_2)\frac{[1-F_2(y^*)]\left[\frac{\widetilde{F}_2(y^*)-F_j(y^*)}{f_2(y^*)y^*}\frac{1}{\varepsilon_2^C(y^*)}+\frac{\varepsilon_2^A(y^*)}{\varepsilon_2^C(y^*)}\right]}{\widetilde{F}_2(y^*)-F_2(y^*)} \\ + (1-\rho_2)\frac{\left[\frac{\widetilde{F}_2(y^*)-F_j(y^*)}{f_2(y^*)y^*}\frac{1}{\varepsilon_2^C(y^*)}+\frac{\varepsilon_2^A(y^*)}{\varepsilon_2^C(y^*)}\right]}{\widetilde{F}_2(y^*)-F_2(y^*)} \end{array} \right\} C_2(y^*).$$

In particular, for any $y^* \in \left[\underline{y}_1, \underline{y}_2\right] \cup [\overline{y}_1, \overline{y}_2]$:

$$T^{\star\prime}(y^*) = \left\{ \begin{array}{c} \rho_j \frac{\left[\widetilde{F}_j^C(y^*)-\widetilde{F}_j(y^*)\right]}{\widetilde{F}_j(y^*)-F_j(y^*)} + (1-\rho_j)\frac{\left[1-F_j(y^*)\right]\left[\frac{\widetilde{F}_j(y^*)-F_j(y^*)}{f_j(y^*)y^*}\frac{1}{\varepsilon_j^C(y^*)}+\frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}\right]}{\widetilde{F}_j(y^*)-F_j(y^*)} \\ + (1-\rho_j)\frac{\frac{\widetilde{F}_j(y^*)-F_j(y^*)}{f_j(y^*)y^*}\frac{1}{\varepsilon_j^C(y^*)}+\frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}}{\widetilde{F}_j(y^*)-F_j(y^*)} \end{array} \right\} T_j^{\star\prime}(y^*), \tag{B28}$$

where $j=1$ if $y^* \in \left[\underline{y}_1, \underline{y}_2\right]$ and $j=2$ if $y^* \in [\overline{y}_1, \overline{y}_2]$. In conclusion, we have Proposition 2.■

B.4 Proof of Theorem 1

Combination of (B27) and (B28) gives:

$$T^{\star\prime}(y^*) = \sum_j \frac{N_j f_j(y^*)\, y^* \varepsilon_j^C(y^*)}{\sum_j N_j f_j(y^*)\, y^* \varepsilon_j^C(y^*)} \left\{ \begin{array}{c} \rho_j \frac{\left[\widetilde{F}_j^C(y^*)-\widetilde{F}_j(y^*)\right]}{\widetilde{F}_j(y^*)-F_j(y^*)} + \\ (1-\rho_j)\frac{\left[1-F_j(y^*)\right]\left[\frac{\widetilde{F}_j(y^*)-F_j(y^*)}{f_j(y^*)y^*}\frac{1}{\varepsilon_j^C(y^*)}+\frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}\right]}{\widetilde{F}_j(y^*)-F_j(y^*)} \end{array} \right\} T_j^{\star\prime}(y^*), \forall y^* \in \left(\underline{y}_1, \overline{y}_2\right). \tag{B29}$$

This equation can be derived under the Case 2 of Section B.3 if we do not eliminate $\sum_j N_j(\rho_j - 1)$ by $\sum_j N_j(\rho_j - 1) = 0$.

Notice that $T_j^{\star\prime}(y^*) = [1 - T'(y^*)]\, C_j(y^*)$. Equation (B29) implies that for any $y^* \in \left(\underline{y}_1, \overline{y}_2\right)$:

$$\frac{T^{\star\prime}(y)}{1 - T^{\star\prime}(y)} = \sum_j \frac{N_j f_j(y^*)\, y^* \varepsilon_j^C(y^*)}{\sum_j N_j f_j(y^*)\, y^* \varepsilon_j^C(y^*)} \left\{ \begin{array}{c} \rho_j \frac{\left[\widetilde{F}_j^C(y^*) - \widetilde{F}_j(y^*)\right]}{\widetilde{F}_j(y^*) - F_j(y^*)} + \\ (1-\rho_j) \frac{\left[1 - F_j(y^*)\right]\left[\frac{\widetilde{F}_j(y^*) - F_j(y^*)}{f_j(y^*) y^*} \frac{1}{\varepsilon_j^C(y^*)} + \frac{\varepsilon_j^A(y^*)}{\varepsilon_j^C(y^*)}\right]}{\widetilde{F}_j(y^*) - F_j(y^*)} \end{array} \right\} C_j(y^*).$$

■

B.5 Proof of Corollary 2

(i) Firstly, substituting $T_j^{\star\prime}(y)$ in (20) by (9) we have:

$$\frac{T^{\star\prime}(y(\theta))}{1 - T^{\star\prime}(y(\theta))} = \frac{\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)} \tau^\star(\theta),$$

or equivalently,

$$T^{\star\prime}(y(\theta)) = \frac{\frac{\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)} \tau^\star(\theta)}{1 + \frac{\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)} \tau^\star(\theta)} = \frac{\frac{\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta} \left(1 + \frac{1}{\varepsilon_j}\right)}{1 + \frac{\widetilde{F}_{\theta_j}^C(\theta) - F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta} \left(1 + \frac{1}{\varepsilon_j}\right)} = \frac{\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}^C(\theta) - F_{\theta_j}(\theta)} \tau^{C,\star}(\theta). \tag{B30}$$

The second and third equations are derived from the definitions of $\tau^\star(\theta)$ and $\tau^{C,\star}(\theta)$ (see, e.g., (8)).

Substituting $1 - T^{\star\prime}(y(\theta))$ in the above equation by (10), we immediately have $T^{A\prime}(y(\theta)) = \tau^{C,\star}(\theta)$. Using $T^{A\prime}(y(\theta)) = T^{\star\prime}(y(\theta)) + T_j^{\star\prime}(y(\theta))$ and the above equation, we have:

$$\begin{aligned} T_j^{\star\prime}(y(\theta)) &= \frac{\widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}^C(\theta) - F_{\theta_j}(\theta)} \tau^{C,\star}(\theta) = \frac{\frac{\widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta} \left(1 + \frac{1}{\varepsilon_j}\right)}{1 + \frac{\widetilde{F}_{\theta_j}^C(\theta) - F_{\theta_j}(\theta)}{f_{\theta_j}(\theta)\theta} \left(1 + \frac{1}{\varepsilon_j}\right)} \\ &= \tau^\star(\theta) \frac{1 - \tau^{C,\star}(\theta)}{1 - \tau^\star(\theta)}, \end{aligned} \tag{B31}$$

where the second and third equations are derived by the definitions of $\tau^\star(\theta)$ and $\tau^{C,\star}(\theta)$. Combination of (B30) and (B31) delivers $T^{\star\prime}(y(\theta)) = \frac{\tau^{C,\star}(\theta) - \tau^\star(\theta)}{1 - \tau^\star(\theta)}$. In conclusion, we have (23).

(ii) Suppose that $\widetilde{F}_{\theta_j}^C(\theta) > F_{\theta_j}(\theta)$ such that $\tau^{C,\star}(\theta) > 0$. (B30) implies $T^{\star\prime}(y(\theta)) \geq 0$ if and only if $\widetilde{F}_{\theta_j}^C(\theta) \geq \widetilde{F}_{\theta_j}(\theta)$, and the equation holds only if $\widetilde{F}_{\theta_j}^C(\theta) = \widetilde{F}_{\theta_j}(\theta)$. (B31) suggests that $T_j^{\star\prime}(y(\theta)) \geq 0$ if and only if $\widetilde{F}_{\theta_j}(\theta) \geq F_{\theta_j}(\theta)$, and the equation holds only if $\widetilde{F}_{\theta_j}(\theta) = F_{\theta_j}(\theta)$. Moreover, if $\widetilde{F}_{\theta_j}^C(\theta) > \widetilde{F}_{\theta_j}(\theta)$ such that $\tau^{C,\star}(\theta) > \tau^\star(\theta)$, we have $T_j^{\star\prime}(y(\theta)) < \tau^\star(\theta)$. Last, combination of (B30) and (B31) suggests:

$$T^{\star\prime}(y(\theta)) - T_j^{\star\prime}(y(\theta)) = \left[\frac{\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}^C(\theta) - F_{\theta_j}(\theta)} - \frac{\widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)}{\widetilde{F}_{\theta_j}^C(\theta) - F_{\theta_j}(\theta)}\right] \tau^{C,\star}(\theta),$$

which means $T^{\star\prime}(y(\theta)) \geq T_j^{\star\prime}(y(\theta))$ if and only if $\widetilde{F}_{\theta_j}^C(\theta) - \widetilde{F}_{\theta_j}(\theta) \geq \widetilde{F}_{\theta_j}(\theta) - F_{\theta_j}(\theta)$.

(iii) Notice that $1-T^{\star\prime}\left(y\left(\theta\right)\right)=\frac{1-\tau^{C,\star}(\theta)}{1-\tau^{\star}(\theta)}$ and $1-T_j'\left(y_j\left(\theta\right)\right)=1-\frac{\left[1-\tau^{C,\star}(\theta)\right]\tau^{\star}(\theta)}{1-\tau^{\star}(\theta)}$, we have

$$\begin{aligned} p^C\left(y_j(\theta)\right)-p_j\left(y_j(\theta)\right) &= \frac{1}{\frac{d\ln y(\theta)}{d\theta}}\left[\frac{d\ln\left[1-T_j'\left(y\left(\theta\right)\right)\right]}{d\theta}-\frac{d\ln\left[1-T'\left(y\left(\theta\right)\right)\right]}{d\theta}\right] \\ &= \frac{1}{\frac{d\ln y(\theta)}{d\theta}}\left[\frac{d\ln\frac{1-\left[1-\tau^{C,\star}(\theta)\right]\frac{\tau^{\star}(\theta)}{1-\tau^{\star}(\theta)}}{\frac{1-\tau^{C,\star}(\theta)}{1-\tau^{\star}(\theta)}}}{d\theta}\right], \end{aligned}$$

where

$$\begin{aligned} \frac{d\ln\frac{1-\left[1-\tau^{C,\star}(\theta)\right]\frac{\tau^{\star}(\theta)}{1-\tau^{\star}(\theta)}}{\frac{1-\tau^{C,\star}(\theta)}{1-\tau^{\star}(\theta)}}}{d\theta} &= \frac{d\ln\left[\frac{1-\tau^{\star}(\theta)}{1-\tau^{C,\star}(\theta)}-\tau^{\star}\left(\theta\right)\right]}{d\theta} \\ &= \frac{\frac{-\tau^{\star\prime}(\theta)}{1-\tau^{C,\star}(\theta)}+\frac{[1-\tau^{\star}(\theta)]\tau^{C,\star\prime}(\theta)}{[1-\tau^{C,\star}(\theta)]^2}-\tau^{\star\prime}\left(\theta\right)}{\frac{1-\tau^{\star}(\theta)}{1-\tau^{C,\star}(\theta)}-\tau^{\star}\left(\theta\right)} \\ &= \frac{-\tau^{\star\prime}\left(\theta\right)}{\frac{1-\tau^{\star}(\theta)}{1-\tau^{C,\star}(\theta)}-\tau^{\star}\left(\theta\right)}. \end{aligned}$$

In conclusion, we have $p^C\left(y_j(\theta)\right)-p_j\left(y_j(\theta)\right)>0$ if $\frac{1-\tau^{\star}(\theta)}{1-\tau^{C,\star}(\theta)}>\tau^{\star}\left(\theta\right)$ and $\tau^{\star\prime}\left(\theta\right)<0$.■

B Numerical Analysis

Data. According to the findings of Saez and Stantcheva (2018),[44] we assume that the income distribution obeys the lognormal distribution at first. Starting from 50,000, it transforms from a lognormal distribution to a Pareto distribution. The transformation ends at 800,000. After 800,000, labor income obeys the Pareto distribution with a Pareto parameter equal to 1.4. Denote y_m as the average labor income of all individuals with labor income higher than y. To estimate the labor income distribution $F_y(y)$, we choose the Pareto mode, mean, and variance of lognormal distribution so that (i) $\frac{y_m}{y_m-y}$ is equal to the corresponding value of the sample at $y=80,0000$ and 200,000, respectively; and (ii) the average of y coincides with the average incomes in the sample. Finally, we assume that individuals' annual labor incomes are higher than 7,000 and normalize the labor income distribution. This treatment follows Mankiw et al. (2009), who assume that taxpayers earn not less than \$3.5 per hour.

Algorithm. We use equation (C31) with $\rho_j=1$ to derive the optimal federal income tax rates by fixed-point approach. To do this, we first substitute (9) in (C31) by $T_j^{\star\prime}\left(y^*\right)$ and substitute $\tau_j^{\star}\left(\theta\right)$ by (8):

$$\frac{T^{\star\prime}\left(y^*\right)}{1-T^{\star\prime}\left(y^*\right)}=\sum_j\left[\frac{N_j\varepsilon_j^C\left(y^*\right)y^*f_j\left(y^*\right)}{\sum_j N_j\varepsilon_j^C\left(y^*\right)y^*f_j\left(y^*\right)}\frac{\widetilde{F}_j^C\left(y^*\right)-\widetilde{F}_j\left(y^*\right)}{\widetilde{F}_j\left(y^*\right)-F_j\left(y^*\right)}\tau_j^{\star}\left(\theta_j\left(y^*\right)\right)\right]. \tag{B32}$$

Provided allocations, we can derive schedules of federal and state income taxes by (B32), as well as inter-region transfers that make $\rho_j=1$. Then, we can derive a new set of allocations by solving individuals'

[44] See, e.g., Figure 4 in Saez and Stantcheva (2018).

problems with the given policies. Repeating the above steps until the policy converges, we obtain the federal income tax that is potentially optimal. For details about this fixed-point approach, see Mankiw et al. (2009).

C Discussions and Robustness

This section discusses relevant issues that have not been considered in our benchmark analysis. (*i*) Is the choice of the tax base of central income tax relevant to the policy design? (*ii*) What's the optimal transfer and optimal tax under the optimal transfer? (*iii*) Can we extend our method to derive the optimal policy in a more general problem? For example, in an economy with labor mobility and endogenous social welfare weight.

C.1 Tax Base

In our benchmark model, we consider a central income tax on the before-regional-tax income. What if the central income tax is levied on the after-regional-tax income?[45] Denote a central income tax on after-regional-tax income as $T^{\#}$. The individual's problem can be reformulated as:

$$\begin{aligned} & \max_{l_j} c_j - \phi'\left(l_j\right) \\ \text{s.t. } c_j \quad = \quad & A_j\theta l_j - T_j\left(\theta l_j\right) - T^{\#}\left(\theta l_j - T_j\left(A_j\theta l_j\right)\right). \end{aligned}$$

Solving the above problem yields:

$$\left[1 - T_j'\left(y_j\left(\theta\right)\right)\right]\left[1 - T^{\#\prime}\left(y_j\left(\theta\right) - T_j'\left(y_j\left(\theta\right)\right)\right)\right] = \frac{1}{A_j\theta}\phi'\left(\frac{y_j\left(\theta\right)}{A_j\theta}\right).$$

It is easy to see that when there is only one region, an allocation implemented by $T\left(\cdot\right)$, can also be implemented by $T^{\#}\left(\cdot\right)$. One can construct $T^{\#}\left(\cdot\right)$ by:

$$\left[1 - T_j'\left(y_j\left(\theta\right)\right)\right]\left[1 - T^{\#\prime}\left(y_j\left(\theta\right) - T_j'\left(y_j\left(\theta\right)\right)\right)\right] = 1 - T_j'\left(y_j\left(\theta\right)\right) - T'\left(y_j\left(\theta\right)\right).$$

However, this is generally not the case when there are multiple regions. To see this, remember that when local tax is not deductible before the federal tax, $\frac{\phi'\left(l_j\left(\theta_j\right)\right)}{A_j\theta_j}\frac{1}{1-\tau_j^{\star}\left(\theta_j\right)} = \frac{\phi'\left(l_{-j}\left(\theta_{-j}\right)\right)}{A_{-j}\theta_{-j}}\frac{1}{1-\tau_{-j}^{\star}\left(\theta_{-j}\right)}$ as long as $y_{-j}\left(\theta_{-j}\right) = y_j\left(\theta_j\right)$. However, this is generally not the case when local tax is deductible before the federal tax.

The deductibility of state and local taxes carries significant implications for both revenue generation and tax equity (see, e.g., Kaplow (1996b)). However, the question of whether such taxes should be deductible from federal taxable income remains a subject of ongoing debate (see, e.g., Stark (2003)).

This debate gained prominence during the Reagan administration's 1984 tax reform proposal, which sought to eliminate state and local tax deductibility. As Feldstein and Metcalf (1987) recount, state and local governments vehemently opposed the proposal, arguing that deductibility was crucial for maintaining public support for essential services and that its elimination would result in unfair double taxation. While acknowledging these concerns, the Treasury Department countered that deductibility creates distortions in

[45] See Kaplow (1996a) and Stark (2003) for some facts about the deductibility of state and local taxes under federal income tax.

resource allocation and primarily benefits high-income taxpayers, leading to substantial federal revenue losses.

The deductibility debate intersects with fundamental principles of fiscal federalism. While these principles generally advocate for centralizing redistributive policies, our study suggests that local redistribution can offer distinct advantages.[46] Ultimately, the impact of deductibility on national welfare is a complex issue that warrants further investigation.

C.2 Endogenous Transfer

In our benchmark model, we deliberately decouple the problem of tax design from the determination of optimal intergovernmental transfers. It is important to note that endogenizing E_j would not alter our core findings, which emphasize the crucial role of objective inconsistency, income distributions, and labor supply elasticities in shaping optimal central redistribution. This robustness stems from the fact that the optimal local and central income tax rates in our benchmark model are independent of E_j.

However, the interplay between regions, mediated through intergovernmental transfers, indirectly influences optimal taxation when social welfare weights are endogenous. To create a nontrivial optimal transfer problem, we need to introduce either a strictly concave utility function or a nonlinear social welfare function in which social welfare weights depend on the transfers themselves. Otherwise, the solution becomes trivial: either there is no need for inter-regional transfers, or the central government would transfer all income from one region to the other.

Let's consider a scenario where the central government's social welfare weight is endogenous. In this case, the central government aims to maximize national welfare by choosing both the tax function, $T(\cdot)$, and the inter-regional transfer, E_j. By redistributing income through lump-sum transfers, the central government directly impacts income distributions, thereby influencing the social welfare weights. A noteworthy finding that emerges is the central government's indifference between regions under the optimal transfer scheme, implying that $\rho_j = \rho_{-j} = 1$. A detailed analysis using a planning problem with a Bergson-Samuelson social welfare function is provided in Section C.3.

C.3 Primal Approach

Mechanism design has difficulty in dealing with region-uniform policy constraints. The policy constraint requires individuals of the same income to pay the same amount of tax. This income-based constraint can not be directly expressed in terms of exogenous skills because income is endogenous. This section shows how to use the primal approach to derive the optimal taxation. In addition to providing an alternative approach, as an extension, we allow E_j to be free and consider a nonlinear Bergson-Samuelson social welfare function.

For simplicity, we assume that $\Theta_1 = \Theta_2 = \Theta = (0, \infty)$, such that $\underline{y}_j = \underline{y}_{-j}$ and $\overline{y}_j, \overline{y}_{-j} \to \infty$. Denote $\theta_j(\theta) : \Theta_{-j} \mapsto \Theta_j$ as the skill of worker who earns $y_{-j}(\theta)$ in region $-j$. Under Assumption 1, $\theta_j : \Theta_{-j} \mapsto \Theta_j$ is a one-to-one mapping with $\theta_j\left(\underline{\theta}_{-j}\right) = \underline{\theta}_j$ and $\theta_j\left(\overline{\theta}_{-j}\right) = \overline{\theta}_j$.

[46] Stark (2003) provides a nuanced analysis of this conflict, highlighting how the federal deduction for state and local taxes (SALT) deviates from core fiscal federalism principles. He argues that the current SALT deduction disproportionately favors itemizing taxpayers, progressive income and property taxes over regressive sales taxes, and high-income earners. Consequently, he proposes a more aligned approach: a more progressive federal tax structure coupled with less progressive state and local tax systems.

The central government's problem is choosing $\left\{y_j\left(\theta_j\right), c_j\left(\theta_j\right)\right\}_{\theta_j\in\Theta_j, j\in\mathcal{J}}$ to maximize:

$$\sum N_j \int_{\Theta_j} G\left(V_j\left(\theta\right)\right) \widetilde{f}^C_{\theta_j}\left(\theta\right) d\theta$$

subject to the incentive constraints:

$$V_j'\left(\theta\right) = \frac{1}{\theta}\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right), \forall j \in \mathcal{J}, \theta \in \Theta_j, \tag{C1}$$

the resource constraint:

$$\sum_j N_j \int_{\Theta_j} y_j\left(\theta_j\right) f_j(\theta) d\theta = \sum_j N_j \left[\int_{\Theta_j}\left[V_j\left(\theta\right) + \phi\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)\right] f_j(\theta)\delta_j\left(\theta\right) d\theta\right], \tag{C2}$$

and the policy constraints. That is for any θ_j and θ_{-j} that satisfy $y_j\left(\theta_j\right) = y_{-j}\left(\theta_{-j}\right)$, $T'\left(y_j\left(\theta_j\right)\right) = T'\left(y_{-j}\left(\theta_{-j}\right)\right)$. According to (2) and (10), the policy constraint is equivalent to

$$\frac{\phi'\left(\frac{y_j\left(\theta_j\right)}{A_j\theta_j}\right)}{A_j\theta_j}\frac{1}{1-\tau_j^\star\left(\theta_j\right)} = \frac{\phi'\left(\frac{y_{-j}\left(\theta_{-j}\right)}{A_{-j}\theta_{-j}}\right)}{A_{-j}\theta_{-j}}\frac{1}{1-\tau_{-j}^\star\left(\theta_{-j}\right)}, \tag{C3}$$

for any θ_j and θ_{-j} satisfying $y_j\left(\theta_j\right) = y_{-j}\left(\theta_{-j}\right)$.

The Bergson-Samuelson social welfare function $G\left(\cdot\right)$ is continuously differentiable, concave, and increasing in its argument. We take $\widetilde{f}_j^C\left(\cdot\right)$ as a PDF and use the region-uniform social welfare function $G\left(\cdot\right)$ to govern the central government's preference for inter-region equality. The general result about the implementability is delivered in Lemma 6.

Lemma 6 *Suppose that the FOCs of the individuals and the regional governments are both necessary and sufficient. The allocation of MGTE $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right)\right\}_{\theta_j\in\Theta_j, j\in\mathcal{J}}$ satisfies the following conditions jointly:*

(i) Incentive constraints (C1), where $y_j\left(\theta_j\right) = A_j\theta_j l_j\left(\theta_j\right)$, and $V_j\left(\theta_j\right) = c_j\left(\theta_j\right) - \phi\left(l_j\left(\theta_j\right)\right)$;

(ii) Policy constraints (C3), for any θ_j and θ_{-j} satisfying $y_j\left(\theta_j\right) = y_{-j}\left(\theta_{-j}\right)$;

(iii) Market clearing condition (C2).

Conversely, when the allocation $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right)\right\}_{\theta_j\in\Theta_j, j\in\mathcal{J}}$ satisfies the properties 1–3 above, there exists a central policy system $\mathcal{T}^C$ such that the allocation can be implemented by $\mathcal{T}^C$ at some regional income taxes $\mathcal{T}^R$ in a MGTE.

Proof. See Appendix C.1. ■

Following Lemma 6, we use Lagrange multipliers to solve for the constrained optimal allocation. To facilitate the Lagrange multiplier method, we should express the non-parametric policy constraint (C3) in terms of parametrics. It turns out to be easier to control $\left\{y_j\left(\theta_j\right), V_j\left(\theta_j\right)\right\}_{\theta_j\in\Theta_j, j\in\mathcal{J}}$ instead of $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right)\right\}_{\theta_j\in\Theta_j, j\in\mathcal{J}}$. Notice that $\theta_2\left(\theta\right)$ is determined by $y_1\left(\cdot\right)$ and $y_2\left(\cdot\right)$. Given $y_1\left(\cdot\right)$, we can replace $y_2\left(\cdot\right)$ by $\theta_2\left(\cdot\right)$, which deliv-

ers an alternative form of the policy constraint:

$$\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta}\frac{1}{1-\tau_1^\star(\theta)}=\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{1-\tau_2^\star(\theta_2(\theta))},\forall\theta\in\Theta_1. \tag{C4}$$

Notice that we have expressed the policy constraint in terms of $\theta \in \Theta_1$. We can also express the incentive constraint and resource constraints in Θ_1 space. The planner's problem can be described as choosing $\left\{y_1(\theta), \theta_2(\theta), V_j(\theta)\right\}_{\theta\in\Theta_1, j\in\mathcal{J}}$ to maximize the objective under the incentive and resource constraints.

The optimal tax formulas derived from the variational and primal approaches are consistent (see Appendix C.2). Interestingly, as is mentioned in Section C.2, ρ_j equals ρ_{-j} at the optimum. A more striking finding is:

Corollary 4 *When the central government chooses inter-region transfers freely, and the intra-region social welfare weights are consistent, the optimal federal income tax rates are zero.*

Proof. See Appendix C.2. ■

The reason behind this interesting finding is straightforward. When intra-region welfare weights are consistent, the only use of the federal income tax is mimicking inter-region transfer, which is better to be performed by the inter-region transfer.

Corollary 4 has two policy implications. For one thing, inter-region transfer reduces central income tax's use. For another, even if there is inter-region transfer, central income tax is generally desirable because of intra-region inconsistency (either due to preference or object inconsistency).

The above primal approach can also be used to study other multi-level principal-agent models. For example, the optimal income tax with screening in the labor market (see, e.g., Stantcheva (2014)). It will be interesting to see the optimal tax rate when the government attempts to maximize the lowest-skilled individuals' utility while firms maximize their profit under rational constraints. In this case, the firm's problem is the dual problem of government's problem. As firms deal with the incentive problem, the optimal income tax should differ greatly from the traditional optimal income tax.

In Appendix E, we also show how can the use of the primal approach be extended to an economy with labor mobility. In the following model extension, we study optimal central income tax with the variational approach.

C Proofs to Discussion and Robustness

C.1 Proof of Lemma 6

We first show that a MGTE allocation $\left\{l_j(\theta_j), c_j(\theta_j)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ must satisfy parts 1–3. First, by the definition of MGTE, (C2) must be satisfied. Second, by the definition of MGTE, individuals maximize their utilities, which results in (2) and (C1). Third, by the definition of MGTE, local governments maximize their objectives, which results in (10). (10) together with (2) gives (C3).

Next, suppose that we are given an allocation $\left\{l_j(\theta_j), c_j(\theta_j)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ satisfying parts 1–3. We now construct $\mathcal{T}^C$ and $\mathcal{T}^R$, which together with $\left\{l_j(\theta_j), c_j(\theta_j)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ constructs a MGTE. We first construct $\mathcal{T}^C$ and $\mathcal{T}^R$ with the given allocation $\left\{l_j(\theta_j), c_j(\theta_j)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$. We then show that the constructed policy system $\mathcal{T}^C$ and $\mathcal{T}^R$ together with $\left\{l_j(\theta_j), c_j(\theta_j)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ do construct a MGTE.

First, given the allocation $\{l_j(\theta_j), c_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ we construct $\{y_j(\theta_j), V_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ by $y_j(\theta_j) = A_j\theta_j l_j(\theta_j)$ and $V_j(\theta_j) = c_j(\theta_j) - \phi(l_j(\theta_j))$. In addition, we construct $\theta_2(\theta)$ by $y_2(\theta_2(\theta)) = y_1(\theta)$, $\forall\theta\in\Theta$. The aggregate regional tax rates $T_j^{A\prime}(\cdot)$ are constructed by (2):

$$T_j^{A\prime}(y_j(\theta)) = 1 - \frac{\phi'\left(\frac{y_j(\theta)}{A_j\theta}\right)}{A_j\theta}.$$

Then, we can construct $T'(y_j)$ by (10).

To fix the income taxes, we should solve for the boundary values of taxes. We use agents' budget constraints to fix the income taxes. We first construct $T_j(\cdot)$ and $T(\cdot)$ by

$$T_j(y_j(\theta)) = T_j(y_j(\underline{\theta})) - \int_{y_j(\underline{\theta})}^{y_j(\theta)} T_j'(y)\,dy$$

and

$$T(y_j(\theta)) = T(y_j(\underline{\theta})) - \int_{y_j(\underline{\theta})}^{y_j(\theta)} T'(y)\,dy$$

We then substitute $T_j(y_j(\theta))$ and $T(y_j(\theta))$ in the following equation by the above equations:

$$y_j(\theta) - T_j(y_j(\theta)) - T(y_j(\theta)) - c_j(\theta) = 0. \tag{C5}$$

We want to show that there exists $\{T_j(y_j(\underline{\theta}))\}_{j\in\mathcal{J}}$ and $T(y_j(\underline{\theta}))$ such that given $\{y_j(\theta_j), V_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$, equation (C5) is satisfied for any $\theta\in\Theta$. To be consistent with the $\underline{\theta}$-type agent's budget constraint, $\{T_j(y_j(\underline{\theta}))\}_{j\in\mathcal{J}}$ and $T(y_j(\underline{\theta}))$ must satisfy:

$$y_j(\underline{\theta}) - T_j(y_j(\underline{\theta})) - T(y_j(\underline{\theta})) - c_j(\underline{\theta}) = 0. \tag{C6}$$

We should show that any $\{T_j(y_j(\underline{\theta}))\}_{j\in\mathcal{J}}$ and $T(y_j(\underline{\theta}))$ satisfying (C6) are also consistent with other agents' budget constraints. This is equivalent to saying that:

$$y_j'(\theta)\left[1 - T_j'(y(\theta)) - T'(y(\theta))\right] - c_j'(\theta) = 0.$$

Substituting $1 - T_j'(y(\theta)) - T'(y(\theta))$ with (2), the above equation is equivalent to:

$$c_j'(\theta) - \frac{\phi'(l_j(\theta))}{A_j\theta}y_j'(\theta) = 0.$$

The above equations hold because we have:

$$\begin{aligned} V_j'(\theta) &= \frac{\phi'(l_j(\theta))\,l_j(\theta)}{A_j\theta} \\ &= c_j'(\theta) - \frac{\phi'(l_j(\theta))}{A_j\theta}y_j'(\theta) + \frac{\phi'(l_j(\theta))\,l_j(\theta)}{A_j\theta}. \end{aligned} \tag{C7}$$

The first equation of (C7) is the incentive compatible condition, and the second equation is derived by the definition of $V_j(\theta)$. We have shown that any $\left\{T_j\left(y_j(\underline{\theta})\right)\right\}_{j\in\mathcal{J}}$ and $T\left(y_j(\underline{\theta})\right)$ satisfying the equations (C6) are consistent with all agents' budget constraints. To fix $\left\{T_j\left(y_j(\underline{\theta})\right)\right\}_{j\in\mathcal{J}}$ and $T\left(y_j(\underline{\theta})\right)$ we need one more equation in addition to (C6). This is the resource constraint or, equivalently, the government's budget constraint:

$$\begin{aligned} 0 &= \sum_j N_j \int_\Theta \left[T_j\left(y_j(\theta)\right) + T\left(y_j(\theta)\right)\right] f_{\theta_j}(\theta)\, d\theta \\ &= \sum_j N_j \int_\Theta \left[T_j\left(y_j(\underline{\theta})\right) + \int_{y_j(\underline{\theta})}^{y_j(\theta)} T_j'(y)\, dy + T\left(y_j(\underline{\theta})\right) + \int_{y_j(\underline{\theta})}^{y_j(\theta)} T'(y)\, dy\right] f_{\theta_j}(\theta)\, d\theta \\ &= \sum_j N_j \int_\Theta \left[\int_{y_j(\underline{\theta})}^{y_j(\theta)} T_j'(y)\, dy + \int_{y_j(\underline{\theta})}^{y_j(\theta)} T'(y)\, dy\right] f_{\theta_j}(\theta)\, d\theta + \sum_j N_j T_j\left(y_j(\underline{\theta})\right) + T\left(y_j(\underline{\theta})\right). \end{aligned} \tag{C8}$$

Any $\left\{T_j\left(y_j(\underline{\theta})\right)\right\}_{j\in\mathcal{J}}$ and $T\left(y_j(\underline{\theta})\right)$ that solves the equation system (C6) and (C8) are consistent with the allocation.

We now construct $\left\{E_j\right\}_{j\in\mathcal{J}}$ by regional governments' budget constraints, which implies

$$N_j E_j = N_j \left[\int_\Theta y_j(\theta) f_j(\theta)\, d\theta - \int_\Theta c_j(\theta) f_j(\theta)\, d\theta - \int_\Theta T\left(y_j(\theta)\right) f_j(\theta)\, d\theta\right], \forall j \in \mathcal{J}.$$

In conclusion, given the allocation, we can construct a unique combination of federal income tax and intergovernment transfers that are consistent with the allocation in the equilibrium.

Last, we show that the allocation $\mathcal{A}$ satisfying parts 1–3 and the constructed policy system $\mathcal{T}$ construct an MGTE. First, the allocation satisfies the incentive conditions (C1), which are both necessary and sufficient for agents' optimal choice given the preconditions of Lemma 6. Second, under the constructed E_j, the constructed regional income taxes $\mathcal{T}^R$ satisfies (10), local governments' budget constraint and the incentive conditions (C1). Therefore, regional income taxes are the optimal solution to local governments' problems. Third, the goods market clear condition (C2) are satisfied. Lastly, agents' budget constraints are embedded in gross utilities' definitions and income taxes' construction. In conclusion, the constructed policy system $\mathcal{T}$ together with the given allocation $\mathcal{A}$ constructs an MGTE.■

C.2 Using Primal Approach to Derive Optimal Taxation

Define $\widetilde{V}_j(\theta) = V_j\left(\theta_j(\theta)\right)$ and introduce

$$\delta_j(\theta) = \theta_j'(\theta) \tag{C9}$$

as a control variable. $\theta_j(\theta)$ is a state variable. We have the following incentive conditions and national resource constraint:

$$\widetilde{V}_j'(\theta) = \frac{\frac{y_1(\theta)}{A_j\theta_j(\theta)}\phi'\left(\frac{y_1(\theta)}{A_j\theta_j(\theta)}\right)}{\theta_j(\theta)}\delta_j(\theta), \forall j \in \mathcal{J}, \theta \in \Theta, \tag{C10}$$

and

$$\sum_j N_j \int y_1(\theta) f_j(\theta_j(\theta))\delta_j(\theta)\, d\theta = \sum_j N_j \left[\int \left[\widetilde{V}_j(\theta) + \phi\left(\frac{y_1(\theta)}{\theta_j(\theta)}\right)\right] f_j(\theta_j(\theta))\delta_j(\theta)\, d\theta\right]. \tag{C11}$$

Formally, the central government's problem is:

$$\max_{\{\theta_2(\theta),V_2(\theta),V_1(\theta),\delta_2(\theta),y_1(\theta)\}_{\theta\in\Theta}} N_1 \int G\left(\widetilde{V}_1(\theta)\right) \widetilde{f}_1^C(\theta)\, d\theta + N_2 \int G\left(\widetilde{V}_2(\theta)\right) \widetilde{f}_2^C(\theta_2(\theta))\, \delta_2(\theta)\, d\theta$$

subject to (C10), (C11), (C9) and (C4). $G(\cdot)$ is continuously differentiable, concave, and increasing in its argument. In the following analysis, we assume that $\widetilde{f}_j^C(\cdot)$ is a PDF and use $G(\cdot)$ to govern the central government's preference for inter-region equality.

For later use, we derive the following term:

$$\begin{aligned}
\frac{\partial\left[\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{1-\tau_2^\star(\theta_2(\theta))}\right]}{\partial\theta_2(\theta)} &= \left[-\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)^2} - \frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{y_1(\theta)}{A_2\theta_2(\theta)^2}\right]\frac{1}{1-\tau_2^\star(\theta_2(\theta))} \\
&\quad + \frac{\tau_2^{\star\prime}(\theta_2(\theta))}{[1-\tau_2^\star(\theta_2(\theta))]^2}\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)} \\
&= \left[-\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)^2} - \frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{y_1(\theta)}{A_2\theta_2(\theta)^2}\right]\frac{1}{1-\tau_2^\star(\theta_2(\theta))} \\
&\quad + \frac{T_j^{A\prime\prime}(y(\theta))}{[1-\tau_2^\star(\theta_2(\theta))]\theta_2'(\theta)}y'(\theta) - \frac{T''(y_2(\theta))y_2'(\theta)}{\theta_2'(\theta)} \\
&= \left[-\frac{\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}}{1-\tau_2^\star(\theta_2(\theta))} - T''(y_1(\theta))\right]\frac{y_1'(\theta)}{\theta_2'(\theta)}.
\end{aligned} \tag{C12}$$

The second equation of the above equations is derived by the first-order condition of local government's problem (i.e., equation (10), $T_2^{A\prime}(y_1(\theta)) = T'(y_1(\theta))[1-\tau_2^\star(\theta_2(\theta))] + \tau_2^\star(\theta_2(\theta))$), which gives:

$$\frac{\tau_2^{\star\prime}(\theta_2(\theta))}{[1-\tau_2^\star(\theta_2(\theta))]^2} = \frac{\frac{T_2^{A\prime\prime}(y_1(\theta))}{[1-\tau_2^\star(\theta_2(\theta))]\theta_2'(\theta)}y_1'(\theta) - \frac{T''(y_1(\theta))y_1'(\theta)}{\theta_2'(\theta)}}{1-T_2^{A\prime}(y(\theta))}.$$

The third equation of (C12) is derived by the first-order condition of individual's problem (i.e., equation (2), $\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)} = 1 - T_2^{A\prime}(y_1(\theta))$), which gives:

$$T_2^{A\prime\prime}(y_1(\theta)) = \frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\left[\frac{y_1(\theta)}{A_2\theta_2(\theta)^2}\frac{\theta_2'(\theta)}{y_1'(\theta)} - \frac{1}{A_2\theta_2(\theta)}\right] + \frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2[\theta_2(\theta)]^2}\frac{\theta_2'(\theta)}{y_1'(\theta)}.$$

We now use Lagrange multipliers to solve the central government's problem.[47] The Lagrange function

[47] See Luenberger (1997) for details about the Lagrangian techniques. See Mirrlees (1976), Golosov et al. (2016), Findeisen and Sachs (2017) for the application to the field of public economics.

is set as:

$$\begin{aligned}\mathcal{L} &= N_1 \int G\left(\widetilde{V}_1(\theta)\right) \widetilde{f}_1^C(\theta)\, d\theta_1 + N_2 \int G\left(\widetilde{V}_2(\theta)\right) \widetilde{f}_2^C(\theta_2(\theta))\, \delta_2(\theta)\, d\theta_1 \\ &+ \int q_1(\theta) \left[\frac{\frac{y_1(\theta)}{A_1\theta}\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{\theta} - \widetilde{V}_1'(\theta)\right] d\theta + \int q_2(\theta) \left[\frac{\frac{y_1(\theta)}{A_2\theta_2(\theta)}\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\delta_2(\theta) - \widetilde{V}_2'(\theta)\right] d\theta \\ &+ \int \mu(\theta) \left[\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta}\frac{1}{1-\tau_1^\star(\theta)} - \frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{1-\tau_2^\star(\theta_2(\theta))}\right] d\theta \\ &+ \int \chi(\theta)\left[\delta_2(\theta) - \theta_2'(\theta)\right] d\theta \\ &+\lambda \left[\begin{array}{c} N_1 \int y_1(\theta) f_1(\theta)\, d\theta + N_2 \int y_1(\theta) f_2(\theta_2(\theta))\, \delta_2(\theta)\, d\theta \\ -N_1 \int \left[\widetilde{V}_1(\theta) + \phi\left(\frac{y_1(\theta)}{A_1\theta}\right)\right] f_1(\theta)\, d\theta - N_2 \int \left[\widetilde{V}_2(\theta) + \phi\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\right] f_2(\theta_2(\theta))\, \delta_2(\theta)\, d\theta \end{array}\right].\end{aligned}$$

Taking partially integrals yields:

$$-\int q_1(\theta)\widetilde{V}_1'(\theta)dy = \widetilde{V}_1(\underline{\theta})q_1(\underline{\theta}) - \widetilde{V}_1(\overline{\theta})q_1(\overline{\theta}) + \int q_1'(\theta)\widetilde{V}_1(\theta)d\theta$$

and

$$-\int q_2(\theta)\widetilde{V}_2'(\theta)dy = \widetilde{V}_2(\underline{\theta})q_2(\underline{\theta}) - \widetilde{V}_2(\overline{\theta})q_2(\overline{\theta}) + \int q_2'(\theta)\widetilde{V}_2(\theta)d\theta.$$

The derivatives with respect to the endpoint conditions yield the following boundary conditions:

$$q_j(\underline{\theta}) = q_j(\overline{\theta}) = 0, \forall j \in \mathcal{J}. \tag{C13}$$

Substituting the above conditions into the Lagrangian yields the following first-order conditions for any $\theta \in \Theta$:

$$\frac{\partial \mathcal{L}}{\partial y_1(\theta)} = \left\{\begin{array}{c} q_1(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{\theta}\frac{1}{A_1\theta}\left(1+\frac{1}{\varepsilon}\right) + q_2(\theta)\left[\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{\delta_2(\theta)}{A_2\theta_2(\theta)}\left(1+\frac{1}{\varepsilon}\right)\right] \\ +\mu(\theta)\left[\frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)}{[A_1\theta]^2}\frac{1}{1-\tau_1^\star(\theta)} - \frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{[A_2\theta_2(\theta)]^2}\frac{1}{1-\tau_2^\star(\theta_2(\theta))}\right] \\ +\lambda N_1 f_1(\theta)\left[1 - \frac{1}{A_1\theta}\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)\right] + \lambda N_2 f_2(\theta_2(\theta))\,\delta_2(\theta)\left[1 - \frac{1}{A_2\theta_2(\theta)}\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\right] \end{array}\right\} = 0, \tag{C14}$$

and

$$\frac{\partial \mathcal{L}}{\partial \theta_2(\theta)} = \left\{\begin{array}{c} N_2\widetilde{V}_2(\theta)\,\widetilde{f}_2^{C\prime}(\theta_2(\theta))\,\delta_2(\theta) - q_2(\theta_2)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{y_1(\theta)}{A_2\theta_2(\theta)^2}\left[2+\frac{1}{\varepsilon}\right]\delta(\theta) \\ +\lambda N_2\left[\begin{array}{c} y_1(\theta) f_2'(\theta_2(\theta))\,\delta_2(\theta) - \left[\widetilde{V}_2(\theta) + \phi\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\right] f_2'(\theta_2(\theta))\,\delta_2(\theta) \\ +f_2(\theta_2(\theta))\,\delta_2(\theta)\,\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)^2} \end{array}\right] \\ +\chi'(\theta) + \mu(\theta)\left[\left[\frac{\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}}{1-\tau_2^\star(\theta_2(\theta))} + T''(y_1(\theta))\right]\frac{y_1'(\theta)}{\theta_2'(\theta)}\right] \end{array}\right\} = 0, \tag{C15}$$

and

$$\frac{\partial \mathcal{L}}{\partial \delta_2(\theta)} = \begin{bmatrix} N_2 G\left(\widetilde{V}_2(\theta)\right) \widetilde{f}_2^{\,C}\left(\theta_2(\theta)\right) + \chi(\theta) + q_2(\theta) \frac{\frac{y_1(\theta)}{A_2\theta_2(\theta)} \phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)} \\ +\lambda N_2 f_2\left(\theta_2(\theta)\right)\left[y_1(\theta) - c_2\left(\theta_2(\theta)\right)\right] \end{bmatrix} = 0, \tag{C16}$$

and

$$\frac{\partial \mathcal{L}}{\partial \widetilde{V}_1(\theta)} = q_1'(\theta) + N_1 \widetilde{f}_1^{C}(\theta) G'\left(\widetilde{V}_1(\theta)\right) - N_1 \lambda f(\theta) = 0, \tag{C17}$$

and

$$\frac{\partial \mathcal{L}}{\partial \widetilde{V}_2(\theta)} = q_2'(\theta) + N_2 G'\left(\widetilde{V}_2(\theta)\right) \widetilde{f}_2^{C}\left(\theta_2(\theta)\right) \delta_2(\theta) - N_2 \lambda f_2\left(\theta_2(\theta)\right) \delta_2(\theta) = 0, \tag{C18}$$

where we substitute $\widetilde{V}_j(\theta) + \phi\left(\frac{y_j(\theta)}{\theta_j(\theta)}\right)$ by $c_j\left(\theta_j(\theta)\right)$.

According to (C17), (C18) and the boundary conditions we have:

$$-\frac{q_j'(\theta)}{\lambda} = \left[\frac{\widetilde{f}_j^{C}\left(\theta_j(\theta)\right)}{\lambda f_j\left(\theta_j(\theta)\right)} G'\left(\widetilde{V}_j(\theta)\right) - 1\right] N_j f_j\left(\theta_j(\theta)\right) \delta_j(\theta)$$

and

$$\begin{aligned} \frac{q_j(\theta)}{\lambda} &= \int_\theta^\infty \left[\frac{\widetilde{f}_j^{C}\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right)}{\lambda f_j\left(\theta_j(\theta)\right)} - 1\right] N_j f_j\left(\theta_j(\theta)\right) \delta_j(\theta)\, d\theta \\ &= N_j\left[1 - F_j\left(\theta_j(\theta)\right)\right]\left[\overline{g}_j\left(\theta_j(\theta)\right) - 1\right], \end{aligned} \tag{C19}$$

where

$$\overline{g}_j\left(\theta_j(\theta)\right) = \frac{\int_\theta^\infty \left[\frac{\widetilde{f}_j^{C}\left(\theta_j(x)\right) G'\left(\widetilde{V}_j(x)\right)}{\lambda f_j\left(\theta_j(x)\right)}\right] f_j\left(\theta_j(x)\right) \delta_j(x)\, dx}{1 - F_j\left(\theta_j(\theta)\right)} \tag{C20}$$

and

$$\lambda = \frac{\int_\Theta \widetilde{f}_j^{C}\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right) \delta_j(\theta)\, d\theta}{\int_\Theta f_j\left(\theta_j(\theta)\right) \delta_j(\theta)\, d\theta}, \forall j \in \mathcal{J}. \tag{C21}$$

For later use, define $\overline{g}_{y_j} : \mathcal{Y} \mapsto \mathbb{R}_+$ by:

$$\overline{g}_{y_j}\left(y_j\left(\theta_j(\theta)\right)\right) = \overline{g}_j\left(\theta_j(\theta)\right), \forall \theta \in \Theta_1, j \in \mathcal{J}. \tag{C22}$$

Note that the right side of (C21) must be independent of j. This condition inherently determines the optimal inter-region transfer. Notice that $\int_\Theta f_j\left(\theta_j(\theta)\right) \delta_j(\theta)\, d\theta = 1$. Equation (C21) implies

$$\lambda = \sum_j N_j \int_\Theta \widetilde{f}_j^{C}\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right) \delta_j(\theta)\, d\theta = \int_\Theta \widetilde{f}_j^{C}\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right) \delta_j(\theta)\, d\theta, j \in \mathcal{J}.$$

One can see that the inter-region weight $\rho_j = \frac{\int \widetilde{f}_j^{C}\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right) \delta_j(\theta) d\theta}{\sum_j N_j \int_\Theta \widetilde{f}_j^{C}\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right) \delta_j(\theta) d\theta} = 1$. Therefore, under the optimal inter-region transfer, the central government is indifferent between the regions.

Using (C20) to (C22) we can derive the following equation by the definition of ρ_j and $\widetilde{F}_j(\cdot)$ (see, e.g.,

(18)):

$$
\begin{aligned}
& \overline{g}_{y_j}\left(y_j\left(\theta_j(\theta)\right)\right)\left[1-F_j\left(y_j\left(\theta_j(\theta)\right)\right)\right] \qquad \text{(C23)}\\
= & \frac{\int_\Theta \widetilde{f}_j^C\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right)\delta_j(\theta)\,d\theta}{\sum_j N_j \int_\Theta \widetilde{f}_j^C\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right)\delta_j(\theta)\,d\theta}\int_\theta^\infty G'\left(\widetilde{V}_j(x)\right)\frac{\widetilde{f}_j^C\left(\theta_j(x)\right)\delta_j(x)}{\int_\Theta \widetilde{f}_j^C\left(\theta_j(\theta)\right) G'\left(\widetilde{V}_j(\theta)\right)\delta_j(\theta)\,d\theta}dx\\
= & \rho_j\left[1-\widetilde{F}_j^C\left(y_j\left(\theta_j(\theta)\right)\right)\right],
\end{aligned}
$$

where $\rho_j = 1$.

Taking derivative of the both sides of (C16) with respect to θ, we have:

$$
\begin{aligned}
-\chi'(\theta) = & N_2\widetilde{V}_2'(\theta)\,\widetilde{f}_2^C\left(\theta_2(\theta)\right) + N_2\widetilde{V}_2(\theta)\,\widetilde{f}_2^{C\prime}\left(\theta_2(\theta)\right)\delta_2(\theta) + q_2'(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)}}{\theta_2(\theta)}\\
& + q_2(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{y_1'(\theta)}{A_2\theta_2(\theta)}\left(1+\frac{1}{\varepsilon}\right)\\
& - q_2(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{y_1(\theta)}{A_2\theta_2(\theta)^2}\left(2+\frac{1}{\varepsilon}\right)\delta(\theta)\\
& + \lambda N_2\left[\begin{array}{c} f_2\left(\theta_2(\theta)\right)y_1'(\theta) + f_2'\left(\theta_2(\theta)\right)\delta_2(\theta)\left[y_1(\theta)-c_2\left(\theta_2(\theta)\right)\right] - \\ f_2\left(\theta_2(\theta)\right)\left[\widetilde{V}_2'(\theta) + \phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\left[\frac{y_1'(\theta)}{A_2\theta_2(\theta)} - \frac{y_1(\theta)\delta_2(\theta)}{A_2\theta_2(\theta)^2}\right]\right]\end{array}\right].
\end{aligned}
$$

Substitute $\chi'(\theta)$ in (C15) by the above equation and reorganizing the equation:

$$
\left\{\begin{array}{c}
-N_2\widetilde{V}_2'(\theta)\left[\widetilde{f}_2^C\left(\theta_2(\theta)\right) - \lambda f_2\left(\theta_2(\theta)\right)\right] - q_2'(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)}}{\theta_2(\theta)}\\
-q_2(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{y_1'(\theta)}{A_2\theta_2(\theta)}\left(1+\frac{1}{\varepsilon}\right)\\
-\lambda N_2\left[y'(\theta) - \phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1'(\theta)}{A_2\theta_2(\theta)}\right]f_2\left(\theta_2(\theta)\right)\\
+\mu(\theta)\left[\left[\frac{\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}}{1-\tau_2^\star\left(\theta_2(\theta)\right)} + T''\left(y(\theta)\right)\right]\frac{y_1'(\theta)}{\theta_2'(\theta)}\right]
\end{array}\right\} = 0. \qquad \text{(C24)}
$$

Substitute $q_2'(\theta)$ and $\widetilde{V}_2'(\theta)$ in (C24) by (C18) and (C10), respectively:

$$
\begin{aligned}
& \left[\begin{array}{c}
-\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)}}{\theta_2(\theta)}\left[G'\left(\widetilde{V}_2(\theta)\right)-1\right]N_2\widetilde{f}_2^C\left(\theta_2(\theta)\right)\delta_2(\theta)\frac{\theta_2'(\theta)}{y_1'(\theta)}\\
+q_2(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}\left(1+\frac{1}{\varepsilon}\right)\theta_2'(\theta)\\
+\lambda N_2 f_2\left(\theta_2(\theta)\right)\left[1-\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{1}{A_2\theta_2(\theta)}\right]\theta_2'(\theta)
\end{array}\right] \qquad \text{(C25)}\\
= & \mu(\theta)\left[T''\left(y_1(\theta)\right) + \frac{\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}}{1-\tau_2^\star\left(\theta_2(\theta)\right)}\right].
\end{aligned}
$$

Add two sides of the above equation to the two sides of (C14) and rearrange:

$$\begin{array}{rl} & \left[\begin{array}{c} \frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)}}{\theta_2(\theta)}\left[G'\left(\widetilde{V}_2(\theta)\right)-1\right]N_2\widetilde{f}_2^C(\theta_2(\theta))\,\delta_2(\theta)\,\frac{\theta_2'(\theta)}{y_1'(\theta)}+ \\ q_1(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{\theta}\frac{1}{A_1\theta}\left(1+\frac{1}{\varepsilon}\right)+\lambda N_1 f_1(\theta)\left[1-\frac{1}{A_1\theta}\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)\right] \end{array}\right] \\ = & -\mu(\theta)\left[\frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)}{(A_1\theta)^2}\frac{1}{1-\tau_1^\star(\theta)}+T''(y_1(\theta))\right]. \end{array} \tag{C26}$$

Substitute $\mu(\theta)$ in (C14) by (C25) (alternatively, one can substitute $\mu(\theta)$ by (C26)):

$$\begin{array}{rl} & q_1(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{\theta}\frac{1}{A_1\theta}\left(1+\frac{1}{\varepsilon}\right)+q_2(\theta)\left[\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{\delta_2(\theta)}{A_2\theta_2(\theta)}\left(1+\frac{1}{\varepsilon}\right)\right] \\ & +\lambda N_1 f_1(\theta)\left[1-\frac{1}{A_1\theta}\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)\right]+\lambda N_2 f_2(\theta_2(\theta))\,\delta_2(\theta)\left[1-\frac{1}{A_2\theta_2(\theta)}\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\right] \\ = & -\frac{\left[\frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)}{(A_1\theta)^2}-\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{(A_1\theta)^2}\right]\left[\begin{array}{c} -\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)}}{\theta_2(\theta)}\left[G'\left(\widetilde{V}_2(\theta)\right)-1\right]N_2\widetilde{f}_2^C(\theta_2(\theta))\,\frac{\delta_2(\theta)}{y_1'(\theta)}+ \\ q_2(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\frac{\theta_2(\theta)}{A_2\theta_2(\theta)}}\frac{1+\varepsilon}{\varepsilon}+\lambda N_2 f_2(\theta_2(\theta))\left[1-\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\right] \end{array}\right]\theta_2'(\theta)}{T''(y_1(\theta))+\frac{\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}}{1-\tau_2^\star(\theta_2(\theta))}}. \end{array} \tag{C27}$$

Notice that

$$\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{[A_2\theta_2(\theta)]^2}\frac{1}{1-\tau_2^\star(\theta_2(\theta))}=\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{y_1(\theta)}{A_2\theta_2(\theta)}}{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{y_1(\theta)A_2\theta_2(\theta)}\frac{1}{1-\tau_2^\star(\theta_2(\theta))}$$

and

$$\frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)}{(A_1\theta)^2}\frac{1}{1-\tau_1^\star(\theta)}=\frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)\frac{y_1(\theta)}{A_1\theta}}{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{y_1(\theta)A_1\theta}\frac{1}{1-\tau_1^\star(\theta)}.$$

We have $\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{[A_2\theta_2(\theta)]^2}\frac{1}{1-\tau_2^\star(\theta_2(\theta))}=\frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)}{(A_1\theta)^2}\frac{1}{1-\tau_1^\star(\theta)}$. Therefore, (C27) implies:

$$\left[\begin{array}{c} q_1(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{\theta}\frac{1}{A_1\theta}\left(1+\frac{1}{\varepsilon}\right)+\lambda N_1 f_1(\theta)\left[1-\frac{1}{A_1\theta}\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)\right] \\ q_2(\theta)\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{\theta_2(\theta)}\frac{1}{A_2\theta_2(\theta)}\left(1+\frac{1}{\varepsilon}\right)\theta_2'(\theta) \\ +\lambda N_2 f_2(\theta_2(\theta))\left[1-\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\frac{1}{A_2\theta_2(\theta)}\right]\theta_2'(\theta) \end{array}\right]=0. \tag{C28}$$

Equivalently, we have:

$$0 = \left[\frac{q_2(\theta)}{\lambda N_2 f_2(\theta_2(\theta))} \frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)^2} \left(1+\frac{1}{\varepsilon}\right) + \left[1 - \frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\right]\right] \lambda N_2 f_2(\theta_2(\theta)) \theta_2'(\theta)$$
$$+ \left[\frac{q_1(\theta)}{\lambda N_1 f_1(\theta)} \frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta^2} \left(1+\frac{1}{\varepsilon}\right) + \left[1 - \frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta}\right]\right] \lambda N_1 f_1(\theta_1).$$

When $\theta = \widetilde{\theta}_1(y)$, we have $\theta_2(\theta) = \widetilde{\theta}_2(y)$ and $\theta_2\left(\widetilde{\theta}_1(y)\right) = \widetilde{\theta}_2(y)$. Therefore, $\theta_2'\left(\widetilde{\theta}_1(y)\right)\widetilde{\theta}_1'(y) = \widetilde{\theta}_2'(y)$ and $\theta_2'(\theta) = \frac{\widetilde{\theta}_2'(y)}{\widetilde{\theta}_1'(y)}\theta_2'(\theta) = \theta_2'\left(\widetilde{\theta}_1(y)\right)$. Substitute $\frac{q_j(\theta)}{\lambda}$ by (C19) and use $\widetilde{\theta}_j(y)$:

$$0 = \left[\frac{\left[1-F_{y_2}(y)\right]\left[\overline{g}_{y_2}(y)-1\right]}{f_2\left(\widetilde{\theta}_2(y)\right)} \frac{\phi'\left(\frac{y}{A_2\widetilde{\theta}_2(y)}\right)}{A_2\widetilde{\theta}_2(y)^2} \left(1+\frac{1}{\varepsilon}\right) + \left[1 - \frac{\phi'\left(\frac{y}{A_2\widetilde{\theta}_2(y)}\right)}{A_2\widetilde{\theta}_2(y)}\right]\right] N_2 f_2\left(\widetilde{\theta}_2(y)\right) \frac{\theta_2'(y)}{\theta_1'(y)} \quad \text{(C29)}$$
$$+ \left[\frac{\left[1-F_{y_1}(y)\right]\left[\overline{g}_{y_1}(y)-1\right]}{f_1\left(\widetilde{\theta}_1(y)\right)} \frac{\phi'\left(\frac{y}{A_1\widetilde{\theta}_1(y)}\right)}{A_1\widetilde{\theta}_1(y)^2} \left(1+\frac{1}{\varepsilon}\right) + \left[1 - \frac{\phi'\left(\frac{y}{A_1\widetilde{\theta}_1(y)}\right)}{A_1\widetilde{\theta}_1(y)}\right]\right] N_1 f_1\left(\widetilde{\theta}_1(y)\right).$$

Substitute $\frac{1}{f_j\left(\widetilde{\theta}_j(y)\right)\widetilde{\theta}_j(y)} \frac{\phi'\left(\frac{y}{A_j\widetilde{\theta}_j(y)}\right)}{A_j\widetilde{\theta}_j(y)} \left(1+\frac{1}{\varepsilon}\right)$, $1 - \frac{\phi'\left(\frac{y}{A_j\widetilde{\theta}_j(y)}\right)}{A_j\widetilde{\theta}_j(y)}$ and $f_j\left(\widetilde{\theta}_j(y)\right)\widetilde{\theta}_j'(y)$ by $\frac{T_j'(y)}{\widetilde{F}_j(y)-F_j(y)}$ (see e.g., (A13)), $T_j'(y) + T'(y)$ and $f_j(y)$, respectively:

$$0 = \left[\frac{\left[1-F_{y_1}(y)\right]\left[\overline{g}_{y_1}(y)-1\right]}{\widetilde{F}_{y_1}(y) - F_{y_1}(y)} T_1'(y) + T_1'(y) + T'(y) \right] N_1 f_{y_1}(y)$$
$$+ N_2 f_{y_2}(y) \left[\frac{\left[1-F_{y_2}(y)\right]\left[\overline{g}_{y_2}(y)-1\right]}{\widetilde{F}_{y_2}(y) - F_{y_2}(y)} T_2'(y) + T_2'(y) + T'(y) \right].$$

Equivalently, we have:

$$T'(y) = \sum\nolimits_{j\in\mathcal{J}} \left[\frac{N_j f_j(y)}{\sum_{j\in\mathcal{J}} N_j f_j(y)} \frac{\left[1-\widetilde{F}_j(y)\right] - \left[1-F_j(y)\right]\overline{g}_{y_j}(y)}{\widetilde{F}_j(y) - F_j(y)} T_j'(y) \right]. \quad \text{(C30)}$$

Substitute $\overline{g}_{y_j}(y)$ in (C30) by (C23) gives:

$$T'(y) = \sum\nolimits_{j\in\mathcal{J}} \left[\frac{N_j f_j(y)}{\sum_{j\in\mathcal{J}} N_j f_j(y)} \frac{\left[1-\widetilde{F}_j(y)\right] - \rho_j\left[1-\widetilde{F}_j^C(y)\right]}{\widetilde{F}_j(y) - F_j(y)} T_j'(y) \right]. \quad \text{(C31)}$$

In particular, when $G(V_j) = V_j$ and $\int_{\Theta_j} \widetilde{f}_{\theta_j}^C(\theta)\, d\theta = 1$, we have $\left[1-F_j(\theta_j(\theta))\right]\left[\overline{g}_j(\theta_j(\theta)) - 1\right] =$

$F_j(y) - \frac{\widetilde{F}_j^C(y)}{\lambda}$ and $\lambda = 1$. In this case,

$$T'(y) = \sum\nolimits_{j \in \mathcal{J}} \left[\frac{N_j f_j(y)}{\sum_{j \in \mathcal{J}} N_j f_j(y)} \frac{\widetilde{F}_j^C(y) - \widetilde{F}_j(y)}{\widetilde{F}_j(y) - F_j(y)} T_j'(y) \right].$$

One thing worth noting here is that due to $\frac{\phi''\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{[A_2\theta_2(\theta)]^2} \frac{1}{1-\tau_2^\star(\theta_2(\theta))} = \frac{\phi''\left(\frac{y_1(\theta)}{A_1\theta}\right)}{(A_1\theta)^2} \frac{1}{1-\tau_1^\star(\theta)}$, the choice of $y_1(\theta)$ has no influence on the intensity of policy constraint. This is not true if the utility cost of working in different regions are different or ϕ is not constant elasticity with respect to labor supply.

ONLINE APPENDIX

D Model Extension

D.1 Proof of Proposition 3

We use Lagrange multipliers to solve the local government's problem. The authors also use the variational method to solve this problem and obtained the same results. The relevant results can be provided upon request.

The Lagrange function is set as:

$$\begin{aligned}\mathcal{L} &= \int_{\Theta} U(\theta)\,\widetilde{f}_j(\theta)\,d\theta + \int_{\Theta} q_j(\theta)\left[\frac{l_j(\theta)\,\phi'\left(l_j(\theta)\right)}{\theta} - V_j'(\theta)\right]d\theta \\ &\quad +\lambda_j\left[\begin{array}{c}\int_{\Theta} A_j\theta l_j(\theta)\,f_j^p(\theta)\,d\theta - \int_{\Theta}\left[V_j(\theta)+\phi\left(l_j(\theta)\right)\right]f_j^p(\theta)\,d\theta \\ -\int_{\Theta} T\left(A_j\theta l_j(\theta)\right)f_j^p(\theta)\,d\theta_j - \int_{\Theta} E_j f_j^p(\theta)\,d\theta\end{array}\right].\end{aligned}$$

Taking partially integrals yields:

$$-\int_{\Theta} q(\theta)V_j'(\theta)d\theta = V_j(\underline{\theta})q_j(\underline{\theta}) - V_j(\bar{\theta})q_j(\bar{\theta}) + \int_{\Theta} q_j'(\theta)V_j(\theta)d\theta.$$

The derivatives with respect to the endpoint conditions yield the following boundary conditions:

$$q_j(\underline{\theta}) = q_j(\bar{\theta}) = 0. \tag{D1}$$

Substituting the above conditions into the Lagrangian yields the following first-order conditions for any $\theta \in \Theta$:

$$\frac{\partial\mathcal{L}}{\partial l_j(\theta)} = q_j(\theta)\frac{\phi'\left(l_j(\theta)\right)}{\theta}\left(1+\frac{1}{\varepsilon_j}\right) + \lambda_j f_j^p(\theta)\,A_j\theta\left[1-\frac{\phi'\left(l_j(\theta)\right)}{A_j\theta} - T'\left(A_j\theta l_j(\theta)\right)\right] = 0, \tag{D2}$$

and

$$\frac{\partial\mathcal{L}}{\partial V_j(\theta)} = q_j'(\theta) + \frac{\partial U(\theta)}{\partial V_j(\theta)}\widetilde{f}_j(\theta) + \lambda_j\left[m_j^j(\theta)\left[T_j\left(y_j(\theta)\right) - E_j\right]f_j^p(\theta) - f_j^p(\theta)\right] = 0, \tag{D3}$$

where we substitute $A_j\theta l_j(\theta) - \left[V_j(\theta)+\phi\left(l_j(\theta)\right)\right] - T\left(y_j(\theta)\right)$ by $T_j\left(y_j(\theta)\right)$. Taking integration of both sides of the above equation gives

$$\lambda_j = \frac{\int_{\Theta}\frac{\partial U(\theta)}{\partial V_j(\theta)}\widetilde{f}_j(\theta)\,d\theta}{1-\int_{\Theta} m_j^j(\theta)\left[T_j\left(y_j(\theta)\right)-E_j\right]f_{\theta_j}(\theta)\,d\theta}\frac{1}{N_j}. \tag{D4}$$

We define

$$g_j(\theta) \equiv \frac{1}{\lambda_j}\frac{\partial U(\theta)}{\partial V_j(\theta)}\frac{\widetilde{f}_{\theta_j}(\theta)}{N_j f_{\theta_j}(\theta)} \quad \text{and} \quad \overline{g}_j(\theta) \equiv \frac{\int_{\theta}^{\infty} g_j(x)\,f_{\theta_j}(x)\,dx}{\left[1-F_{\theta_j}(\theta)\right]} \tag{D5}$$

as local government's welfare weight for $V_j(\theta)$, and weighted welfare weight for $\left\{V_j(x)\right\}_{x\geq\theta, x\in\Theta}$, respectively. λ_j is the shadow value of local government revenue.

According to (D3) and the boundary conditions we have

$$\begin{aligned} -\frac{q_j'(\theta)}{\lambda_j} &= \left[\frac{1}{\lambda_j}\frac{\partial U(\theta)}{\partial V_j(\theta)}\frac{\widetilde{f}_j(\theta)}{f_j^p(\theta)} - 1\right] f_j^p(\theta) + \varepsilon_j(\theta)\left[T_j\left(y_j(\theta)\right) - E_j\right] f_j^p(\theta) \\ &= \left[g_j(\theta) + r_{jj}(\theta) - 1\right] f_{\theta_j}(\theta) N_j \end{aligned}$$

and

$$\begin{aligned} \frac{q_j(\theta)}{\lambda_j} &= \int_\theta^\infty \left[\frac{1}{\lambda_j}\frac{\partial U(\theta)}{\partial V_j(\theta)}\frac{\widetilde{f}_j(\theta)}{f_{\theta_j}(\theta)} - 1\right] f_j^p(\theta)\, d\theta + \int_\theta^\infty m_j^j(x)\left[T_j\left(y_j(x)\right) - E_j\right] f_j^p(x)\, dx \\ &= N_j\left[1 - F_{\theta_j}(\theta)\right]\left[\overline{g}_j(\theta) + \overline{r}_{jj}(\theta) - 1\right]. \end{aligned} \tag{D6}$$

Rearranging (D2) we have:

$$\frac{1 - \frac{\phi'\left(l_j(\theta)\right)}{A_j\theta} - T'\left(A_j\theta l_j(\theta)\right)}{\frac{\phi'\left(l_j(\theta)\right)}{A_j\theta}} = -\frac{q_j(\theta)}{\theta\lambda_j f_j^p(\theta)}\left(1 + \frac{1}{\varepsilon_j}\right).$$

Substituting $\frac{1}{A_j\theta}\phi'\left(\frac{y}{A_j\theta}\right)$ in (D2) by $1 - T_j^{A\prime}\left(y_j(\theta)\right)$ and $\frac{q_j(\theta)}{\lambda_j}$ by (D6), we have:

$$\frac{T_j'\left(y_j(\theta)\right)}{1 - T_j^{A\prime}\left(y_j(\theta)\right)} = \left[1 - \overline{r}_{jj}(\theta) - \overline{g}_j(\theta)\right]\frac{1 - F_{\theta_j}(\theta)}{\theta f_j(\theta)}\left(1 + \frac{1}{\varepsilon_j}\right). \tag{D7}$$

In conclusion, we have Proposition 3.■

D.2 Proof of Proposition 4

We use Lagrange multipliers to solve the Planner's optimization problem. The The Lagrange function is set as

$$\begin{aligned} \mathcal{L} &= \int_\Theta U(\theta)\widetilde{f}^C(\theta)\, d\theta + \int_\Theta q(\theta)\left[\frac{l(\theta)\phi'(l(\theta))}{\theta} - V'(\theta)\right] d\theta \\ &\quad + \lambda\left[\int_\Theta A\theta l(\theta) f_\theta(\theta)\, d\theta - \int_\Theta \left[V(\theta) + \phi(l(\theta))\right] f_\theta(\theta)\, d\theta\right] \end{aligned}$$

Taking partial integrals yields

$$-\int_\Theta q(\theta)V'(\theta)d\theta = V(\underline{\theta})q(\underline{\theta}) - V(\bar{\theta})q(\bar{\theta}) + \int_\Theta q'(\theta)V(\theta)d\theta.$$

The derivatives with respect to the endpoint conditions yield the following boundary conditions:

$$q(\underline{\theta}) = q(\bar{\theta}) = 0. \tag{D8}$$

Substituting the above conditions into the Lagrangian yields the following first-order conditions for any

$\theta \in \Theta$:

$$\frac{\partial \mathcal{L}}{\partial l(\theta)} = q(\theta) \frac{\phi'(l(\theta))}{\theta} \left[1 + \frac{1}{\varepsilon}\right] + \lambda f(\theta) A\theta \left[1 - \frac{\phi'(l(\theta))}{A\theta}\right] = 0, \tag{D9}$$

and

$$\frac{\partial \mathcal{L}}{\partial V(\theta)} = q'(\theta) + \left[\frac{\partial U(\theta)}{\partial V_j(\theta)} + \frac{\partial U(\theta)}{\partial V_{-j}(\theta)}\right] |_{V_j(\theta)=V_{-j}(\theta)=V(\theta)} \widetilde{f}^C(\theta) - \lambda f(\theta) = 0. \tag{D10}$$

Taking integration of both sides of the above equation gives:

$$\lambda = \int_\Theta \left[\frac{\partial U(\theta)}{\partial V_j(\theta)} + \frac{\partial U(\theta)}{\partial V_{-j}(\theta)}\right] |_{V_j(\theta)=V_{-j}(\theta)=V(\theta)} \widetilde{f}^C(\theta) \, d\theta. \tag{D11}$$

According to (D10) and the boundary conditions (D8) we have

$$\begin{aligned} -\frac{q'(\theta)}{\lambda} &= \left[\frac{\left[\frac{\partial U(\theta)}{\partial V_j(\theta)} + \frac{\partial U(\theta)}{\partial V_{-j}(\theta)}\right] |_{V_j(\theta)=V_{-j}(\theta)=V(\theta)}}{\lambda} \frac{\widetilde{f}^C(\theta)}{f_\theta(\theta)} - 1\right] f_\theta(\theta) \\ &= [g(\theta) - 1] f_\theta(\theta) \end{aligned}$$

and

$$\begin{aligned} \frac{q(\theta)}{\lambda} &= \int_\theta^\infty \left[\frac{\left[\frac{\partial U(\theta)}{\partial V_j(\theta)} + \frac{\partial U(\theta)}{\partial V_{-j}(\theta)}\right] |_{V_j(\theta)=V_{-j}(\theta)=V(\theta)}}{\lambda} \frac{\widetilde{f}^C(\theta)}{f_\theta(\theta)} - 1\right] f_\theta(\theta) \, d\theta \\ &= [1 - F_\theta(\theta)] [\overline{g}(\theta) - 1], \end{aligned} \tag{D12}$$

where we define

$$g(\theta) \equiv \sum_j \frac{\frac{\partial U(\theta)}{\partial V_j(\theta)} \widetilde{f}^C(\theta)}{\lambda f(\theta)} |_{V_j(\theta)=V_{-j}(\theta)=V(\theta)} \quad \text{and} \quad \overline{g}(\theta) \equiv \frac{\int_\theta^\infty g(x) f(x) \, dx}{1 - F_\theta(\theta)} \tag{D13}$$

as the central government's marginal and weighted welfare weights for $V(\theta)$ and $\{V(x)\}_{x \geq \theta, x \in \Theta}$, respectively. λ is the shadow value of central government revenue.[48]

Rearranging (D9) we have:

$$\frac{1 - \frac{\phi'(l(\theta))}{A\theta} - T'(A\theta l(\theta))}{\frac{\phi'(l(\theta))}{A\theta}} = -\frac{q(\theta)}{\lambda f(\theta) \theta} \left(1 + \frac{1}{\varepsilon}\right). \tag{D14}$$

Substitute $\frac{1}{A\theta}\phi'(l(\theta))$ in the above equation by $1 - T^{A\prime}(y(\theta))$ and $\frac{q(\theta)}{\lambda}$ by (D12):

$$\frac{T^{A\prime}(y(\theta))}{1 - T^{A\prime}(y(\theta))} = [1 - \overline{g}(\theta)] \frac{1 - F_\theta(\theta)}{\theta f(\theta)} \left(1 + \frac{1}{\varepsilon}\right).$$

Equivalently,

$$\frac{1}{1 - T^{A\prime}(y(\theta))} = 1 + [1 - \overline{g}(\theta)] \frac{1 - F_\theta(\theta)}{\theta f(\theta)} \left(1 + \frac{1}{\varepsilon}\right).$$

[48] See (D11) for the expression of λ.

Substituting $T^{A\prime}\left(y\left(\theta\right)\right)$ in the above equation by $T^{A\prime}(y\left(\theta\right)) = 1 - \frac{1-T'(y(\theta))}{1+\left[1-\overline{g}_j(\theta)-\overline{r}_{jj}(\theta)\right]\frac{1-F_{\theta_j}(\theta)}{\theta f_j(\theta)}\left(1+\frac{1}{\varepsilon}\right)}$ (see Proposition 3), we have Proposition 4.■

D.3 Proof of Proposition 5

The proof is split into five parts. We first show how local income tax changes with the tax reform given $\frac{\widehat{y}_k(\theta_k)}{y_k(\theta_k)}$ and $\frac{\widehat{\tau}_k^\star(\theta_k)}{1-\tau_k^\star(\theta_k)}$ in Section D.3.1. We then show what's $\frac{\widehat{\tau}_k^\star(\theta_k)}{1-\tau_k^\star(\theta_k)}$ under the tax reform considered. We will show how E_j changes with the tax reform in Section D.3.3. We show how $\widehat{T}\left(y(\underline{\theta})\right) + \widehat{T}_j\left(y_j(\underline{\theta})\right)$ changes with the tax reform in section D.3.4. Last, a combination of the above results gives the result in Proposition 5.

D.3.1 Result 1: Reaction of Local Income Tax

In the first step, we want to derive the following results:

$$\frac{\widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)} = -\left[\frac{1}{\varepsilon_k} + \frac{T''\left(y_k\left(\theta_k\right)\right)y_k\left(\theta_k\right)}{1-T'\left(y_k\left(\theta_k\right)\right)}\right]\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)} + \frac{\widehat{\tau}_k^\star\left(\theta_k\right)}{1-\tau_k^\star\left(\theta_k\right)}. \tag{D15}$$

Alternatively,

$$\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)} = -\frac{1}{\frac{1}{\varepsilon_k} + \frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}}\frac{\widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)} + \frac{1}{\frac{1}{\varepsilon_k} + \frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}}\frac{\widehat{\tau}_k^\star\left(\theta_k\right)}{1-\tau_k^\star\left(\theta_k\right)}. \tag{D16}$$

Proof. To do this, notice that the local income tax before the tax reform is:

$$1-T'\left(y_k\left(\theta_k\right)\right) = \frac{1}{1-\tau_k^\star\left(\theta_k\right)}\left[1-T'\left(y_k\left(\theta_k\right)\right) - T_k^{\star\prime}\left(y_k\left(\theta_k\right)\right)\right] \tag{D17}$$

The local income tax after the tax reform is:

$$\begin{aligned} & 1-T'\left(y_k\left(\theta_k;T+\mu\widehat{T}\right)\right) - \mu\widehat{T}'\left(y_k\left(\theta_k;T+\mu\widehat{T}\right)\right) \\ = & \frac{\left[\begin{array}{c} 1-T'\left(y_k\left(\theta_k;T+\mu\widehat{T}\right)\right) - \mu\widehat{T}'\left(y_k\left(\theta_k;T+\mu\widehat{T}\right)\right) \\ -T_k^{\star\prime}\left(y_k\left(\theta_k;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right) \end{array}\right]}{1-\tau_k^\star\left(\theta_k;\mu\right)}. \end{aligned} \tag{D18}$$

Subtract the items on both sides of the equation (D18) from the items on both sides of equation (D17), and take $\mu \to 0$ to obtain

$$\begin{aligned} & -\left[T''\left(y_k\left(\theta_k\right)\right)\widehat{y}_k\left(\theta_k\right) + \widehat{T}'\left(y_k\left(\theta_k\right)\right)\right] \\ = & -\frac{1-T'\left(y_k\left(\theta_k\right)\right)}{1-T_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}\left\{T_k^{A\prime\prime}\left(y_k\left(\theta_k\right)\right)\widehat{y}_k\left(\theta_k\right) + \widehat{T}_k^{A\prime}\left(y_k\left(\theta_k\right)\right)\right\} - \frac{\widehat{\tau}_k^\star\left(\theta_k\right)}{1-\tau_k^\star\left(\theta_k\right)}\left[1-T'\left(y_k\left(\theta_k\right)\right)\right]. \end{aligned}$$

Rearrange the above equation:

$$-\frac{T''\left(y_k\left(\theta_k\right)\right)\widehat{y}_k\left(\theta_k\right) + \widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)} = -\frac{T_k^{A\prime\prime}\left(y_k\left(\theta_k\right)\right)\widehat{y}_k\left(\theta_k\right) + \widehat{T}_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}{1-T_k^{A\prime}\left(y_k\left(\theta_k\right)\right)} - \frac{\widehat{\tau}_k^\star\left(\theta_k\right)}{1-\tau_k^\star\left(\theta_k\right)}.$$

Equivalently:

$$\begin{aligned}
&\frac{T''\left(y_k\left(\theta_k\right)\right)y_k\left(\theta_k\right)}{1-T'\left(y_k\left(\theta_k\right)\right)}\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)}+\frac{\widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)}\\
= &\frac{T_k^{A\prime\prime}\left(y_k\left(\theta_k\right)\right)y_k\left(\theta_k\right)}{1-T_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)}+\frac{\widehat{T}_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}{1-T_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}+\frac{\widehat{\tau}_k^{\star}\left(\theta_k\right)}{1-\tau_k^{\star}\left(\theta_k\right)}.
\end{aligned}$$

Use $\frac{\widehat{y}_k(\theta_k)}{y_k(\theta_k)} = -\varepsilon_k^A\left(y_k\left(\theta_k\right)\right)\frac{\widehat{T}_k^{A\prime}(y_k(\theta_k))}{1-T_k^{A\prime}(y_k(\theta_k))} = -\frac{1}{\frac{1}{\varepsilon}+\frac{T_k^{A\prime\prime}(y_k(\theta_k))y_k(\theta_k)}{1-T_k^{A\prime}(y_k(\theta_k))}}\frac{\widehat{T}_k^{A\prime}(y_k(\theta_k))}{1-T_k^{A\prime}(y_k(\theta_k))}$ to substitute $\frac{\widehat{T}_k^{A\prime}(y_k(\theta_k))}{1-T_k^{A\prime}(y_k(\theta_k))}$ in the above equation:

$$\begin{aligned}
&\frac{T''\left(y_k\left(\theta_k\right)\right)y_k\left(\theta_k\right)}{1-T'\left(y_k\left(\theta_k\right)\right)}\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)}+\frac{\widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)}\\
= &\frac{T_k^{A\prime\prime}\left(y_k\left(\theta_k\right)\right)y_k\left(\theta_k\right)}{1-T_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)}-\left[\frac{1}{\varepsilon}+\frac{T_k^{A\prime\prime}\left(y_k\left(\theta_k\right)\right)y_k\left(\theta_k\right)}{1-T_k^{A\prime}\left(y_k\left(\theta_k\right)\right)}\right]\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)}+\frac{\widehat{\tau}_k^{\star}\left(\theta_k\right)}{1-\tau_k^{\star}\left(\theta_k\right)}\\
= &-\frac{1}{\varepsilon}\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)}+\frac{\widehat{\tau}_k^{\star}\left(\theta_k\right)}{1-\tau_k^{\star}\left(\theta_k\right)}.
\end{aligned}$$

Rearranging the above equation gives equations (D15) and (D16).■

D.3.2 Result 2: How $\tau_j^{\star}(\theta)$ Changes with the Reform

Set the following notations:

$$e_{kn}\left(y\right) = -\frac{\frac{[1-\overline{g}_k(y)-\overline{r}_{kk}(y)][1-F_k(y)]}{f_k(y)y\varepsilon_k^A(y)}}{\frac{[1-\overline{g}_k(y)-\overline{r}_{kk}(y)][1-F_k(y)]}{f_k(y)y\varepsilon_k^A(y)}+1}\frac{m_{kn}\left(y\right)}{\frac{1}{\varepsilon_k}+\frac{T''(y)y}{1-T'(y)}}, \tag{D19}$$

$$e_{kk}^{own}\left(y,y'\right) = -\frac{\frac{1-F_k(y)}{f_k(y)y\varepsilon_k^A(y)}\left[f_k\left(y'\right)1_{y'\geq y}-\frac{\overline{g}_k(y)}{\overline{g}_k\left(\underline{y}_k\right)}f_k\left(y'\right)\right]}{\frac{[1-\overline{g}_k(y)-\overline{r}_{kk}(y)][1-F_k(y)]}{f_k(y)y\varepsilon_k^A(y)}+1}\frac{m_{kk}\left(y'\right)}{\frac{1}{\varepsilon_k}+\frac{T''(y)y}{1-T'(y)}},$$

$$\begin{aligned}
e_{kn}^{cross}\left(y,y'\right) = &-\frac{\frac{y_k^n(y')\varepsilon_k^A\left(y_k^n(y')\right)\left(\frac{1}{\varepsilon_k}+1\right)}{y'\varepsilon_n^A(y')\left(\frac{1}{\varepsilon_n}+1\right)}\frac{[1-F_k\left(y_k^n(y)\right)]}{f_k\left(y_k^n(y)\right)y_k^n(y)\varepsilon_k^A\left(y_k^n(y)\right)}}{\frac{[1-\overline{g}_k\left(y_k^n(y)\right)-\overline{r}_{kk}\left(y_k^n(y)\right)][1-F_k\left(y_k^n(y)\right)]}{f_k\left(y_k^n(y)\right)y_k^n(y)\varepsilon_k^A\left(y_k^n(y)\right)}+1}\times\\
&\frac{\left\{\begin{array}{c}\left[\frac{f_k\left(y_k^n(y')\right)}{1-F_k\left(y_k^n(y)\right)}1_{y_k^n(y')\geq y_k^n(y)}-\frac{\overline{g}_k\left(y_k^n(y)\right)}{\overline{g}_k\left(\underline{y}_k\right)}f_k\left(y_k^n\left(y'\right)\right)\right]g_k\left(y_k^n\left(y'\right)\right)\mu_{kn}\left(y_k^n\left(y'\right)\right)+\\ \left[\begin{array}{c}\frac{\overline{g}_k\left(\underline{y}_k\right)1_{y_k^n(y')\geq y_k^n(y)}-\overline{g}_k\left(y_k^n(y)\right)}{\overline{g}_k\left(\underline{y}_k\right)}\left[r_{kk}\left(y_k^n\left(y'\right)\right)\varepsilon_{kn}^k\left(y_k^n\left(y'\right)\right)-1\right]\\ -F_k\left(y_k^n\left(y\right)\right)\overline{r}_{kk}\left(y_k^n\left(y\right)\right)\end{array}\right]m_{kn}\left(y_k^n\left(y'\right)\right)f_k\left(y_k^n\left(y'\right)\right)\end{array}\right\}}{\frac{1}{\varepsilon_k}+\frac{T''\left(y_k^n(y)\right)y_k^n(y)}{1-T'\left(y_k^n(y)\right)}},
\end{aligned}$$

where we know that can define the function $\widetilde{\theta}_j\left(y\right)$ as a inverse function of $y=y_j\left(\theta\right)$, which means that

$$y=y_j\left(\theta\right)\Leftrightarrow\theta=\theta_j\left(y\right)$$

Then we define $y_m^j\left(y\right)=y_m\left(\theta_j\left(y\right)\right)$, similar we have $y_k^n\left(y'\right)=y_k\left(\theta_n\left(y'\right)\right)$. Define that the semi-elasticity

of marginal utility $\mu_{kn}(\theta)$ as below:

$$\mu_{kn}(\theta) \equiv \frac{\frac{\partial^2 U}{\partial V_k \partial V_n}(\theta)}{\frac{\partial U}{\partial V_k}(\theta)}$$

We want to show that:

$$\begin{aligned}
-\frac{\widehat{\tau}_k^{\star}(\theta_k)}{1-\tau_k^{\star}(\theta_k)} &= -\sum_n e_{kn}\left(y_k(\theta_k)\right)\widehat{T}_n^{A}\left(y_n(\theta_k)\right) \\
&+\int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k(\theta_k), y_k(x_k)\right)\left[\widehat{T}_k^{A}\left(y_k(x_k)\right)-\widehat{T}\left(y_k(x_k)\right)-\widehat{E}_k\right]dy_k(x_k) \\
&+\int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k(\theta_k), y_k(x_k)\right)T_k^{\star\prime}\left(y_k(x_k)\right)y_k(x_k)\frac{\widehat{y}_k(x_k)}{y_k(x_k)}dy_k(x_k) \\
&-\sum_n\int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}\left(y_k(\theta_k), y_n(x_k)\right)\widehat{T}_n^{A}\left(y_n(x_k)\right)dy_n(x_k).
\end{aligned} \tag{D20}$$

Substitute $\frac{\widehat{\tau}_k^{\star}(\theta_k)}{1-\tau_k^{\star}(\theta_k)}$ in equation (D15), we have:

$$\begin{aligned}
\frac{\widehat{y}_k(\theta_k)}{y_k(\theta_k)} &= -\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}}\frac{\widehat{T}'\left(y_k(\theta_k)\right)}{1-T'\left(y_k(\theta_k)\right)}-\sum_n e_{kn}\left(y_k(\theta_k)\right)\widehat{T}_n^{A}\left(y_n(\theta_k)\right) \\
&+\int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k(\theta_k), y_k(x_k)\right)\left[\widehat{T}_k^{A}\left(y_k(x_k)\right)-\widehat{T}\left(y_k(x_k)\right)-\widehat{E}_k\right]dy_k(x_k) \\
&+\int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k(\theta_k), y_k(x_k)\right)T_k^{\star\prime}\left(y_k(x_k)\right)y_k(x_k)\frac{\widehat{y}_k(x_k)}{y_k(x_k)}dy_k(x_k) \\
&-\sum_n\int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}\left(y_k(\theta_k), y_n(x_k)\right)\widehat{T}_n^{A}\left(y_n(x_k)\right)dy_n(x_k).
\end{aligned} \tag{D21}$$

Alternatively, in terms of labor income, we have:

$$\begin{aligned}
\frac{\widehat{y}_k}{y_k} &= -\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k)y_k}{1-T'(y_k)}}\frac{\widehat{T}'\left(y_k\right)}{1-T'\left(y_k\right)}-\sum_n e_{kn}\left(y_k\right)\widehat{T}_n^{A}\left(y_n^k\left(y_k\right)\right) \\
&-\int_{\underline{y}_k}^{\overline{y}_k} e_{kk}^{own}\left(y_k, y_k'\right)\left[\widehat{E}_k+\widehat{T}\left(y_k'\right)-\widehat{T}_k^{A}\left(y_k'\right)\right]dy_k' \\
&-\sum_n\int_{\underline{y}_n}^{\overline{y}_n} e_{kn}^{cross}\left(y_k, y_n'\right)\widehat{T}_n^{A}\left(y_n'\right)dy_n'+\int_{\underline{y}_k}^{\overline{y}_k} e_{kk}^{own}\left(y_k, y_k'\right)T_k^{\star\prime}\left(y_k'\right)y_k'\frac{\widehat{y}_k'}{y_k'}dy_k'.
\end{aligned}$$

Proof. To see above results, remember that we have set $\tau_k^{\star}(\theta_k)$ in equation (29) in the main body as:

$$\frac{\tau_k^{\star}(\theta_k)}{1-\tau_k^{\star}(\theta_k)} = \left(1+\frac{1}{\varepsilon_k}\right)\frac{\left[1-\overline{g}_k\left(y_k(\theta_k)\right)-\overline{r}_{kk}\left(y_k(\theta_k)\right)\right]\left[1-F_k\left(y_k(\theta_k)\right)\right]}{f_k(\theta_k)\theta_k}.$$

Rearrange the above equation gives:

$$\begin{aligned}\tau_k^\star(\theta_k) &= \frac{\left(1+\frac{1}{\varepsilon_k}\right)\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\frac{d\ln y_k(\theta_k)}{d\ln\theta_k}}}{\left(1+\frac{1}{\varepsilon_k}\right)\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\frac{d\ln y_k(\theta_k)}{d\ln\theta_k}}+1} \\ &= \frac{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1},\end{aligned}$$

where the second equation is derived by $\frac{d\ln y_k(\theta_k)}{d\ln\theta_k}=\varepsilon_k^A\left(y_k\left(\theta_k\right)\right)\left(\frac{1}{\varepsilon_k}+1\right)$.

Remember that $\tau_k^\star(\theta_k)=C_k\left(y_k\left(\theta_k\right)\right)$:

$$C_k\left(y_k\left(\theta_k\right)\right)=\frac{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1}. \tag{D22}$$

Step 1. To derive $\widehat{\tau}_k^\star(\theta_k)$, we first express $\widehat{\tau}_k^\star(\theta_k)$ in terms of the change in revenue effect $\widehat{\overline{r}}_{kk}\left(y_k\left(\theta_k\right)\right)$, change in social welfare weight $\widehat{\overline{g}}_k\left(y_k\left(\theta_k\right)\right)$, and changes in labor income distributions, i.e., $\widehat{f}_{\theta_k}\left(\theta_k\right),\widehat{F}_{\theta_k}\left(\theta_k\right)$.

First, by definition of $\tau_k^\star(\theta_k)$, we have:

$$\begin{aligned}&\frac{\frac{\tau_k^\star\left(\theta_k;T+\mu\widehat{T}\right)-\tau_k^\star(\theta_k)}{\mu}}{\left[1-\tau_k^\star\left(\theta_k\right)\right]\left[1-\tau_k^\star\left(\theta_k;T+\mu\widehat{T}\right)\right]}=\frac{\frac{\tau_k^\star\left(\theta_k;T+\mu\widehat{T}\right)}{1-\tau_k^\star\left(\theta_k;T+\mu\widehat{T}\right)}-\frac{\tau_k^\star(\theta_k)}{1-\tau_k^\star(\theta_k)}}{\mu} \\ =&\left(1+\frac{1}{\varepsilon_k}\right)\frac{\left\{\begin{array}{c}\left[1-\overline{r}_{kk}\left(\theta_k;T+\mu\widehat{T}\right)-\overline{g}_k\left(\theta_k;T+\mu\widehat{T}\right)\right]\frac{1-F_k\left(\theta_k;T+\mu\widehat{T}\right)}{f_k\left(\theta_k;T+\mu\widehat{T}\right)\theta_k} \\ -\left[1-\overline{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)\right]\frac{1-F_k(\theta_k)}{f_k(\theta_k)\theta_k}\end{array}\right\}}{\mu},\end{aligned}$$

Take the derivative of both sides of the above equation with respect to μ:

$$\begin{aligned}\frac{\widehat{\tau}_k^\star\left(\theta_k\right)}{1-\tau_k^\star\left(\theta_k\right)}\frac{1}{\tau_k^\star\left(\theta_k\right)} &= \frac{\frac{[1-\overline{g}_k(\theta_k)-\widehat{\overline{r}_{kk}(\theta_k)}][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\overline{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}} \\ &= -\left[\frac{\widehat{\overline{r}}_{kk}\left(\theta_k\right)+\widehat{\overline{g}}_k\left(\theta_k\right)}{1-\overline{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)}+\frac{\widehat{f}_k\left(\theta_k\right)}{f_k\left(\theta_k\right)}+\frac{\widehat{F}_k\left(\theta_k\right)}{1-F_k\left(\theta_k\right)}\right],\end{aligned} \tag{D23}$$

where, by definition,

$$\widehat{\tau}_k^\star\left(\theta_k\right)\equiv\lim_{\mu\to0}\frac{\tau_k^\star\left(\theta_k;T+\mu\widehat{T}\right)-\tau_k^\star\left(\theta_k\right)}{\mu}.$$

Step 2. In this step, we express $\widehat{\tau}_k^\star\left(\theta_k\right)$ in terms of $\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right),\widehat{f}_k\left(x_k\right),\widehat{m}_{kk}\left(x_k\right),\widehat{T}_k^\star\left(y_k\left(x_k\right)\right),\widehat{y}_k\left(x_k\right)$. By the definition of $g_k\left(x_k\right)$:

$$g_k\left(x_k\right)=\frac{\frac{\partial U}{\partial V_k}\left(x_k\right)\frac{\widetilde{f}_k(x_k)}{f_k(x_k)}}{\int_{\underline{\theta}_k}^{\overline{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\frac{\widetilde{f}_k(x_k)}{f_k(x_k)}f_k\left(x_k\right)dx_k}\int_{\underline{\theta}_k}^{\overline{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)dx_k$$

We have the following expression for $\overline{g}_k(\theta_k) \equiv \int_{\theta_k}^{\overline{\theta}_k} \frac{g_k(x_k) f_k(x_k)}{1-F_k(\theta_k)} dx_k$ and $\overline{r}_{kk}(\theta_k) \equiv \int_{\theta_k}^{\overline{\theta}_k} r_{kk}(x) \frac{f_k(x)}{1-F_k(\theta_k)} dx$:

$$\begin{aligned}
\overline{g}_k(\theta_k) &= \frac{\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\partial U}{\partial V_k}(x_k) \frac{\widetilde{f}_k(x_k)}{f_k(x_k)} \frac{f_k(x_k) 1_{x_k \geq \theta_k}}{1-F_k(\theta_k)} dx_k}{\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\partial U}{\partial V_k}(x_k) \frac{\widetilde{f}_k(x_k)}{f_k(x_k)} f_k(x_k) dx_k} \times \int_{\underline{\theta}_k}^{\overline{\theta}_k} [1 - r_{kk}(x_k)] f_k(x_k) dx_k, \qquad \text{(D24)} \\
\overline{r}_{kk}(\theta_k) &= \int_{\underline{\theta}_k}^{\overline{\theta}_k} r_{kk}(x_k) \frac{f_k(x_k) 1_{x_k \geq \theta_k}}{1 - F_k(\theta_k)} dx_k.
\end{aligned}$$

In particular, one has $\overline{g}_k(\underline{\theta}) = 1 - \overline{r}_{kk}(\underline{\theta})$. Combination of definitions of $g_k(\theta_k), \overline{g}_k(\theta_k), \overline{r}_{kk}(\theta_k)$ gives:

$$\frac{\frac{\partial U(\theta_k)}{\partial V_k(\theta_k)} \widetilde{f}_k(\theta_k)}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial U(x)}{\partial V_k(x)} \widetilde{f}_k(x) dx} = \frac{g_k(\theta_k) f_k(\theta_k)}{\int_{\underline{\theta}_k}^{\overline{\theta}_k} [1 - r_{kk}(x_k)] f_k(x_k) dx_k} = \frac{g_k(\theta_k) f_k(\theta_k)}{1 - \overline{r}_{kk}(\underline{\theta})}, \qquad \text{(D25)}$$

and

$$\frac{\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\partial U}{\partial V_k}(x_k) \widetilde{f}_k(x_k) 1_{x_k \geq \theta_k}}{\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\partial U}{\partial V_k}(x_k) \widetilde{f}_k(x_k) dx_k} dx_k = \overline{g}_k(\theta_k) \frac{1 - F_k(\theta_k)}{\int_{\underline{\theta}_k}^{\overline{\theta}_k} [1 - r_{kk}(x_k)] f_k(x_k) dx_k} = \overline{g}_k(\theta_k) \frac{1 - F_k(\theta_k)}{1 - \overline{r}_{kk}(\underline{\theta})}. \qquad \text{(D26)}$$

Substitute $\overline{g}_k(\theta_k)$ and $\overline{r}_{kk}(\theta_k)$ in equation (29) by (D24):

$$\begin{aligned}
\ln \frac{\tau_k^\star(\theta_k)}{1 - \tau_k^\star(\theta_k)} &= \ln\left(1 + \frac{1}{\varepsilon_k}\right) + \ln \frac{[1 - \overline{g}_k(\theta_k) - \overline{r}_{kk}(\theta_k)][1 - F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k} \qquad \text{(D27)} \\
&= \ln\left(1 + \frac{1}{\varepsilon_k}\right) + \ln \left[\begin{array}{c} \frac{\int_{\underline{\theta}_k}^{\overline{\theta}_k} [1 - r_{kk}(x_k)] f_k(x_k) 1_{x_k \geq \theta_k} dx_k}{\int_{\underline{\theta}_k}^{\overline{\theta}_k} [1 - r_{kk}(x_k)] f_k(x_k) dx_k} \\ - \frac{\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\partial U}{\partial V_k}(x_k) \widetilde{f}_k(x_k) 1_{x_k \geq \theta_k} dx_k}{\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\partial U}{\partial V_k}(x_k) \widetilde{f}_k(x_k) dx_k} \end{array} \right] \\
&\quad + \ln \int_{\underline{\theta}_k}^{\overline{\theta}_k} [1 - r_{kk}(x_k)] f_k(x_k) dx_k - \ln f_k(\theta_k) \theta_k.
\end{aligned}$$

Taking the derivative of both sides of the above equation with respect to μ gives:

$$\begin{aligned}
&\frac{\frac{[1-\overline{g}_k(\theta_k)-\widehat{\overline{r}_{kk}(\theta_k)}][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\overline{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}} \\
&= \frac{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}-\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}} \frac{\left(\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}\right)\widehat{\ }}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}} \\
&\quad - \frac{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}-\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}} \frac{\left(\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}\right)\widehat{\ }}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}} \\
&\quad + \frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-\widehat{r_{kk}(x_k)}\right]f_k(x_k)dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}-\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)}.
\end{aligned} \tag{D28}$$

Basing on the definitions of $g_k(\theta_k)$, $\overline{g}_k(\theta_k)$, and $\overline{r}_{kk}(\theta_k)$, and using (D24) to (D26), we have:

$$\begin{aligned}
\frac{1-\overline{r}_{kk}(\theta_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} &= \frac{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}-\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}}, \\
\frac{\overline{g}_k(\theta_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} &= \frac{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}-\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}}.
\end{aligned} \tag{D29}$$

Substitute corresponding terms in equation (D28) by equations (D29):

$$\begin{aligned}
\frac{\frac{[1-\overline{g}_k(\theta_k)-\widehat{\overline{r}_{kk}(\theta_k)}][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\overline{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}} &= \frac{1-\overline{r}_{kk}(\theta_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} \frac{\left(\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}\right)\widehat{\ }}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}} \\
&\quad - \frac{\overline{g}_k(\theta_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} \frac{\left(\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}\right)\widehat{\ }}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}} \\
&\quad + \frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-\widehat{r_{kk}(x_k)}\right]f_k(x_k)dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}[1-r_{kk}(x_k)]f_k(x_k)dx_k}-\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)}.
\end{aligned} \tag{D30}$$

The right side of equation (D30) is constructed by the following terms:

$$\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)dx_k},\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)dx_k},$$
$$\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k},\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}.$$

Further calculations reveal the expressions of the four items listed above:

$$\begin{aligned}&\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)dx_k} \\ =&\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)dx_k}\left\{\frac{\widehat{f}_k\left(x_k\right)}{f_k\left(x_k\right)}-\frac{r_{kk}\left(x_k\right)}{1-r_{kk}\left(x_k\right)}\frac{\widehat{r}_{kk}\left(x_k\right)}{r_{kk}\left(x_k\right)}\right\}dx_k \\ =&\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{1-r_{kk}\left(x_k\right)}{1-\bar{r}_{kk}\left(\underline{\theta}\right)}f_k\left(x_k\right)\left\{\frac{\widehat{f}_k\left(x_k\right)}{f_k\left(x_k\right)}-\frac{r_{kk}\left(x_k\right)}{1-r_{kk}\left(x_k\right)}\frac{\widehat{r}_{kk}\left(x_k\right)}{r_{kk}\left(x_k\right)}\right\}dx_k,\end{aligned} \tag{D31}$$

and

$$\begin{aligned}\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)dx_k}&=\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)dx_k}\frac{\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)}{\frac{\partial U}{\partial V_k}\left(x_k\right)}dx_k \\ &=\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{g_k\left(x_k\right)}{1-\bar{r}_{kk}\left(\underline{\theta}\right)}f_k\left(x_k\right)\frac{\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)}{\frac{\partial U}{\partial V_k}\left(x_k\right)}dx_k,\end{aligned} \tag{D32}$$

and

$$\begin{aligned}\frac{\widehat{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}&=\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}\left(x_k\right)\widetilde{f}_k\left(x_k\right)dx_k}\frac{1}{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\partial U}{\partial V_k}(x_k)\widetilde{f}_k(x_k)dx_k}}1_{x_k\geq\theta_k}\frac{\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)}{\frac{\partial U}{\partial V_k}\left(x_k\right)}dx_k \\ &=\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{g_k\left(x_k\right)}{\overline{g}_k\left(\theta_k\right)}\frac{f_k\left(x_k\right)}{1-F_k\left(\theta_k\right)}1_{x_k\geq\theta_k}\frac{\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)}{\frac{\partial U}{\partial V_k}\left(x_k\right)}dx_k\end{aligned} \tag{D33}$$

and

$$\begin{aligned}
&\widehat{\frac{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}} \\
= &\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)1_{x_k\geq\theta_k}}{\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left[1-r_{kk}\left(x_k\right)\right]f_k\left(x_k\right)1_{x_k\geq\theta_k}dx_k}\left\{\frac{\widehat{f}_k\left(x_k\right)}{f_k\left(x_k\right)}-\frac{r_{kk}\left(x_k\right)}{1-r_{kk}\left(x_k\right)}\frac{\widehat{r}_{kk}\left(x_k\right)}{r_{kk}\left(x_k\right)}\right\}dx_k \\
= &\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{1-r_{kk}\left(x_k\right)}{1-\bar{r}_{kk}\left(\theta_k\right)}f_k\left(x_k\right)1_{x_k\geq\theta_k}\left\{\frac{\widehat{f}_k\left(x_k\right)}{f_k\left(x_k\right)}-\frac{r_{kk}\left(x_k\right)}{1-r_{kk}\left(x_k\right)}\frac{\widehat{r}_{kk}\left(x_k\right)}{r_{kk}\left(x_k\right)}\right\}dx_k.
\end{aligned} \tag{D34}$$

Substitute (D31) to (D34) into (D29) and use $\overline{g}_k\left(\underline{\theta}\right)=1-\bar{r}_{kk}\left(\underline{\theta}\right)$:

$$\begin{aligned}
&\frac{\widehat{[1-\overline{g}_k(\theta_k)-\bar{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\bar{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}=\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{1-r_{kk}\left(x_k\right)}{1-\bar{r}_{kk}\left(\underline{\theta}\right)}f_k\left(x_k\right)\left\{\frac{\widehat{f}_k\left(x_k\right)}{f_k\left(x_k\right)}-\frac{r_{kk}\left(x_k\right)}{1-r_{kk}\left(x_k\right)}\frac{\widehat{r}_{kk}\left(x_k\right)}{r_{kk}\left(x_k\right)}\right\}dx_k-\frac{\widehat{f}_k\left(\theta_k\right)}{f_k\left(\theta_k\right)} \\
&+\frac{1-\bar{r}_{kk}\left(\theta_k\right)}{1-\bar{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)}\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left\{\begin{array}{c}\frac{1-r_{kk}(x_k)}{1-\bar{r}_{kk}(\theta_k)}f_k\left(x_k\right)1_{x_k\geq\theta_k} \\ -\frac{1-r_{kk}(x_k)}{1-\bar{r}_{kk}(\underline{\theta})}f_k\left(x_k\right)\end{array}\right\}\left\{\frac{\widehat{f}_k\left(x_k\right)}{f_k\left(x_k\right)}-\frac{r_{kk}\left(x_k\right)}{1-r_{kk}\left(x_k\right)}\frac{\widehat{r}_{kk}\left(x_k\right)}{r_{kk}\left(x_k\right)}\right\}dx_k \\
&-\frac{\overline{g}_k\left(\theta_k\right)}{1-\bar{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)}\left[\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{g_k\left(x_k\right)}{\overline{g}_k\left(\theta_k\right)}\frac{f_k\left(x_k\right)}{1-F_k\left(\theta_k\right)}1_{x_k\geq\theta_k}\frac{\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)}{\frac{\partial U}{\partial V_k}\left(x_k\right)}dx_k-\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{g_k\left(x_k\right)}{\overline{g}_k\left(\underline{\theta}\right)}f_k\left(x_k\right)\frac{\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)}{\frac{\partial U}{\partial V_k}\left(x_k\right)}dx_k\right]
\end{aligned} \tag{D35}$$

To derive $\widehat{r}_{kk}\left(x_k\right)$ in the above equation, remember that $r_{kk}\left(x_k\right)\equiv\left[T_k^{\star}\left(y_k\left(x_k\right)\right)-E_k\right]m_{kk}\left(x_k\right)$, and

$$r_{kk}\left(x_k;T+\mu\widehat{T}\right)=\left[T_k^{\star}\left(y_k\left(x_k;T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-E_k-\mu\widehat{E}_k\right]m_{kk}\left(x_k;T+\mu\widehat{T}\right).$$

Taking the derivative of both sides of the above equation with respect to μ gives:

$$\widehat{r}_{kk}\left(x_k\right)=\left[T_k^{\star}\left(y_k\left(x_k\right)\right)-E_k\right]\widehat{m}_{kk}\left(x_k\right)+\left[\widehat{T}_k^{\star}\left(y_k\left(x_k\right)\right)-\widehat{E}_k+T_k^{\star\prime}\left(y_k\left(x_k\right)\right)\widehat{y}_k\left(x_k\right)\right]m_{kk}\left(x_k\right).$$

Using $1-\frac{1-\bar{r}_{kk}(\theta_k)}{1-\bar{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)}=-\frac{\overline{g}_k(\theta_k)}{1-\bar{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)}$ and the above expression of $\widehat{r}_{kk}\left(x_k\right)$ to rewrite (D35):

$$\begin{aligned}
&\frac{\widehat{[1-\overline{g}_k(\theta_k)-\bar{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\bar{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}=-\frac{\widehat{f}_k\left(\theta_k\right)}{f_k\left(\theta_k\right)}+\int_{\underline{\theta}_k}^{\bar{\theta}_k}\left\{\begin{array}{c}\frac{\left[f_k(x_k)1_{x_k\geq\theta_k}-\frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k(x_k)\right][1-r_{kk}(x_k)]}{1-\bar{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)}\frac{\widehat{f}_k(x_k)}{f_k(x_k)} \\ -\frac{\left[\frac{f_k(x_k)}{1-F_k(\theta_k)}1_{x_k\geq\theta_k}-\frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k(x_k)\right]g_k(x_k)}{1-\bar{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)}\frac{\widehat{\frac{\partial U}{\partial V_k}}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)}\end{array}\right\}dx_k \\
&-\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\left[f_k\left(x_k\right)1_{x_k\geq\theta_k}-\frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k\left(x_k\right)\right]r_{kk}\left(x_k\right)}{1-\bar{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)}\frac{\widehat{m}_{kk}\left(x_k\right)}{m_{kk}\left(x_k\right)}dx_k \\
&-\int_{\underline{\theta}_k}^{\bar{\theta}_k}\frac{\left[f_k\left(x_k\right)1_{x_k\geq\theta_k}-\frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k\left(x_k\right)\right]m_{kk}\left(x_k\right)}{1-\bar{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)}\left[\widehat{T}_k^{\star}\left(y_k\left(x_k\right)\right)-\widehat{E}_k+T_k^{\star\prime}\left(y_k\left(x_k\right)\right)\widehat{y}_k\left(x_k\right)\right]dx_k.
\end{aligned} \tag{D36}$$

Step 3. In this step, we derive expressions of $\widehat{\frac{\partial U}{\partial V_k}}\left(x_k\right)$, $\widehat{f}_k\left(x_k\right)$, $\widehat{m}_{kk}\left(x_k\right)$. By lemmas in Appendix E.2, one

has

$$\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)} = -\left[\sum_n m_{kn}(\theta_k)\,\widehat{T}_n^A(y_n(\theta_k)) - \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k)\,f_k(x_k)\,\widehat{T}_n^A(y_n(x_k))\,dx_k\right],$$

and

$$\frac{\widehat{\frac{\partial U}{\partial V_k}}(\theta_k)}{\frac{\partial U}{\partial V_k}(\theta_k)} = -\sum_n \mu_{kn}(\theta_k)\,\widehat{T}_n^A(y_n(\theta_k)); \frac{\widehat{m}_{kk}(\theta_k)}{m_{kk}(\theta_k)} = -\sum_n \left[\varepsilon_{kn}^k(\theta_k) - 1\right] m_{kn}(\theta_k)\,\widehat{T}_n^A(y_n(\theta_k)). \tag{D37}$$

By $\frac{d\ln y_n(x_k)}{d\ln x_k} = \varepsilon_n^A(y_n(x_k))\left(\frac{1}{\varepsilon_n}+1\right)$, the expression of $\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)}$ can be expressed in terms of labor income:

$$\begin{aligned}
&\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)} \\
=\ & -\sum_n m_{kn}(y_k(\theta_k))\,\widehat{T}_n^A(y_n(\theta_k)) - \\
& \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\bar{\theta}_k)} \frac{m_{kn}(y_k(x_k))\,f_k(y_k(x_k))\,y_k(x_k)}{y_n(x_k)}\widehat{V}_n(y_n(x_k))\,\frac{d\ln y_k(x_k)/d\ln x_k}{d\ln y_n(x_k)/d\ln x_k}dy_n(x_k) \\
=\ & -\sum_n m_{kn}(y_k(\theta_k))\,\widehat{T}_n^A(y_n(\theta_k)) - \\
& \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\bar{\theta}_k)} m_{kn}(y_k(x_k))\,\frac{\varepsilon_k^A(y_k(x_k))\,\frac{1+\varepsilon_k}{\varepsilon_k}}{\varepsilon_n^A(y_n(x_k))\,\frac{1+\varepsilon_n}{\varepsilon_n}}\frac{f_k(y_k(x_k))\,y_k(x_k)}{y_n(x_k)}\widehat{V}_n(y_n(x_k))\,dy_n(x_k) \\
=\ & -\sum_n m_{kn}(y_k(\theta_k))\,\widehat{T}_n^A(y_n(\theta_k)) + \\
& \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\bar{\theta}_k)} m_{kn}(y_k(x_k))\,\frac{\varepsilon_k^A(y_k(x_k))\,\frac{1+\varepsilon_k}{\varepsilon_k}}{\varepsilon_n^A(y_n(x_k))\,\frac{1+\varepsilon_n}{\varepsilon_n}}\frac{f_k(y_k(x_k))\,y_k(x_k)}{y_n(x_k)}\widehat{T}_n^A(y_n(x_k))\,dy_n(x_k).
\end{aligned} \tag{D38}$$

Step 4. Substitute $\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)}$ and $\frac{\widehat{\frac{\partial U}{\partial V_k}}(\theta_k)}{\frac{\partial U}{\partial V_k}(\theta_k)}$ in equation (D36) by (D37) and (D38):

$$\begin{aligned}
&\frac{\frac{[1-\overline{g}_k(\theta_k)-\widehat{\overline{r}_{kk}(\theta_k)}][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\overline{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}} \\
=\ & \sum_n m_{kn}(\theta_k)\,\widehat{T}_n^A(y_n(\theta_k)) - \int_{\underline{\theta}_k}^{\bar{\theta}_k} \frac{\left[f_k(x_k)\,1_{x_k\geq\theta_k} - \frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k(x_k)\right] m_{kk}(x_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} \begin{bmatrix} \widehat{T}_k^\star(y_k(x_k)) - \widehat{E}_k \\ +T_k^{\star\prime}(y_k(x_k))\,\widehat{y}_k(x_k) \end{bmatrix} dx_k \\
& + \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} \left\{ \begin{array}{c} \frac{\left[\frac{f_k(x_k)}{1-F_k(\theta_k)}1_{x_k\geq\theta_k} - \frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k(x_k)\right] g_k(x_k)\mu_{kn}(x_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} + \\ \{[1-\overline{r}_{kk}(\theta_k)]\,[1-F_k(\theta_k)] - \overline{g}_k(\theta_k)\}\,m_{kn}(x_k)\,f_k(x_k) \\ -\{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)\}\,m_{kn}(x_k)\,f_k(x_k) \\ -\left[f_k(x_k)\,1_{x_k\geq\theta_k} - \frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k(x_k)\right][1-r_{kk}(x_k)]\,m_{kn}(x_k) \\ \frac{+\left[f_k(x_k)\,1_{x_k\geq\theta_k} - \frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}f_k(x_k)\right] r_{kk}(x_k)\,[\varepsilon_{kn}^k(y_k(x_k))-1]\,m_{kn}(x_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)} \end{array} \right\} \widehat{T}_n^A(y_n(x_k))\,dx_k.
\end{aligned} \tag{D39}$$

Using definitions of $\varepsilon_{kn}^k, r_{kk}, \overline{r}_{kk}, \overline{g}_k$, and (D24) to (D26), we can simplify the right side of the above equa-

tion by the following equation:

$$\begin{aligned}
&\left\{\left[1-\overline{r}_{kk}\left(\theta_k\right)\right]\left[1-F_k\left(\theta_k\right)\right]-\overline{g}_k\left(\theta_k\right)\right\} m_{kn}\left(x_k\right) f_k\left(x_k\right)-\left[1-\overline{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)\right] m_{kn}\left(x_k\right) f_k\left(x_k\right) \quad \text{(D40)}\\
&-\left[f_k\left(x_k\right) 1_{x_k \geq \theta_k}-\frac{\overline{g}_k\left(\theta_k\right)}{\overline{g}_k(\underline{\theta})} f_k\left(x_k\right)\right]\left[1-r_{kk}\left(x_k\right)\right] m_{kn}\left(x_k\right)\\
&+\left[f_k\left(x_k\right) 1_{x_k \geq \theta_k}-\frac{\overline{g}_k\left(\theta_k\right)}{\overline{g}_k(\underline{\theta})} f_k\left(x_k\right)\right] r_{kk}\left(x_k\right)\left[\varepsilon_{kn}^k\left(x_k\right)-1\right] m_{kn}\left(x_k\right)\\
=\ &\left\{\frac{\overline{g}_k(\underline{\theta}) 1_{x_k \geq \theta_k}-\overline{g}_k\left(\theta_k\right)}{\overline{g}_k(\underline{\theta})}\left[r_{kk}\left(x_k\right) \varepsilon_{kn}^k\left(x_k\right)-1\right]-F_k\left(\theta_k\right) \overline{r}_{kk}\left(\theta_k\right)\right\} m_{kn}\left(x_k\right) f_k\left(x_k\right).
\end{aligned}$$

Substitute the corresponding term in the right side of (D39) by (D40):

$$\begin{aligned}
&\frac{\frac{[1-\overline{g}_k(\theta_k)-\widehat{\overline{r}_{kk}(\theta_k)}][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\overline{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}=\sum_n m_{kn}\left(\theta_k\right) \widehat{T}_n^A\left(y_n\left(\theta_k\right)\right) \quad \text{(D41)}\\
&-\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\left[f_k\left(x_k\right) 1_{x_k \geq \theta_k}-\frac{\overline{g}_k\left(\theta_k\right)}{\overline{g}_k(\underline{\theta})} f_k\left(x_k\right)\right] m_{kk}\left(x_k\right)}{1-\overline{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)}\left[\widehat{T}_k^A\left(y_k\left(x_k\right)\right)-\widehat{T}\left(y_k\left(x_k\right)\right)-\widehat{E}_k\right] dx_k\\
&-\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\left[f_k\left(x_k\right) 1_{x_k \geq \theta_k}-\frac{\overline{g}_k\left(\theta_k\right)}{\overline{g}_k(\underline{\theta})} f_k\left(x_k\right)\right] m_{kk}\left(x_k\right) T_k^{\star\prime}\left(y_k\left(x_k\right)\right) y_k\left(x_k\right)}{1-\overline{r}_{kk}\left(\theta_k\right)-\overline{g}_k\left(\theta_k\right)} \frac{\widehat{y}_k\left(x_k\right)}{y_k\left(x_k\right)} dx_k\\
&+\sum_n \int_{\underline{\theta}_k}^{\overline{\theta}_k}\left\{\begin{array}{c}\frac{\left[\frac{f_k(x_k)}{1-F_k(\theta_k)} 1_{x_k \geq \theta_k}-\frac{\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})} f_k(x_k)\right] g_k(x_k) \mu_{kn}(x_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)}+\\ \frac{\left[\frac{\overline{g}_k(\underline{\theta}) 1_{x_k \geq \theta_k}-\overline{g}_k(\theta_k)}{\overline{g}_k(\underline{\theta})}\left(r_{kk}(x_k) \varepsilon_{kn}^k(x_k)-1\right)-F_k(\theta_k) \overline{r}_{kk}(\theta_k)\right] m_{kn}(x_k) f_k(x_k)}{1-\overline{r}_{kk}(\theta_k)-\overline{g}_k(\theta_k)}\end{array}\right\} \widehat{T}_n^A\left(y_n\left(x_k\right)\right) dx_k.
\end{aligned}$$

Substitute (D41) into (D23):

$$
\begin{aligned}
&-\frac{\widehat{\tau}_k^{\star}(\theta_k)}{1-\tau_k^{\star}(\theta_k)} \\
=\;& \frac{\left[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))\right]\frac{[1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1}\left(\frac{\frac{[1-\overline{g}_k(\theta_k)-\widehat{\overline{r}_{kk}(\theta_k)}][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}{\frac{[1-\overline{g}_k(\theta_k)-\overline{r}_{kk}(\theta_k)][1-F_k(y_k(\theta_k))]}{f_k(\theta_k)\theta_k}}\right) \\
=\;& \sum_n \frac{\left[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))\right]\frac{[1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1} m_{kn}(y_k(\theta_k))\,\widehat{T}_n^A(y_n(\theta_k)) \\
&-\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\frac{[1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}\begin{bmatrix} f_k(y_k(x_k))1_{y_k(x_k)\geq y_k(\theta_k)} \\ -\frac{\overline{g}_k(y_k(\theta_k))}{\overline{g}_k(y_k(\underline{\theta}_k))}f_k(y_k(x_k)) \end{bmatrix} m_{kk}(y_k(x_k))}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1}\begin{bmatrix} \widehat{T}_k^A(y_k(x_k))- \\ \widehat{T}(y_k(x_k))-\widehat{E}_k \end{bmatrix} dy_k(x_k) \\
&-\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\frac{[1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}\begin{bmatrix} f_k(y_k(x_k))1_{y_k(x_k)\geq y_k(\theta_k)} \\ -\frac{\overline{g}_k(y_k(\theta_k))}{\overline{g}_k(y_k(\underline{\theta}_k))}f_k(y_k(x_k)) \end{bmatrix} m_{kk}(y_k(x_k))\,T_k^{\star\prime}(y_k(x_k))\,y_k(x_k)}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1}\frac{\widehat{y}_k(x_k)}{y_k(x_k)}dy_k(x_k) \\
&+\sum_n\int_{\underline{\theta}_k}^{\overline{\theta}_k} \frac{\frac{[1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}}{\frac{[1-\overline{g}_k(y_k(\theta_k))-\overline{r}_{kk}(y_k(\theta_k))][1-F_k(y_k(\theta_k))]}{f_k(y_k(\theta_k))y_k(\theta_k)\varepsilon_k^A(y_k(\theta_k))}+1}\frac{y_k(x_k)\,\varepsilon_k^A(y_k(x_k))\left(\frac{1}{\varepsilon_k}+1\right)}{y_n(x_k)\,\varepsilon_n^A(y_n(x_k))\left(\frac{1}{\varepsilon_n}+1\right)}\widehat{T}_n^A(y_n(x_k)) \\
&\times\left\{\begin{array}{c} \left[\frac{f_k(y_k(x_k))}{1-F_k(\theta_k)}1_{y_k(x_k)\geq y_k(\theta_k)}-\frac{\overline{g}_k(y_k(\theta_k))}{\overline{g}_k(y_k(\underline{\theta}_k))}f_k(y_k(x_k))\right]g_k(y_k(x_k))\,\mu_{kn}(y_k(x_k))+ \\ \begin{bmatrix} \frac{\overline{g}_k(y_k(\underline{\theta}_k))1_{y_k(x_k)\geq y_k(\theta_k)}-\overline{g}_k(y_k(\theta_k))}{\overline{g}_k(y_k(\underline{\theta}_k))}\left[r_{kk}(y_k(x_k))\,\varepsilon_{kn}^k(y_k(x_k))-1\right] \\ -F_k(y_k(\theta_k))\,\overline{r}_{kk}(y_k(\theta_k)) \end{bmatrix} m_{kn}(y_k(x_k))\,f_k(y_k(x_k)) \end{array}\right\} dy_n(x_k).
\end{aligned}
$$

Using the definitions of $e_{kn}(y)$, $e_{kn}^{cross}(y,y')$, $e_{kk}^{own}(y,y')$ to simplify the above equation, gives equation (D20). Last, combination of (D20) and (D16) gives (D21).■

D.3.3 Result 3: Local Government's Budget Constraint

Then the local government's Budget Constraint can be presented as below:

$$
\begin{aligned}
&\widehat{T}_k^A(y_k(\underline{\theta}))-\sum_j \overline{r}_{kj}(\underline{\theta})\,\widehat{T}_j^A(y_j(\underline{\theta})) \qquad \text{(D42)} \\
=\;& \left[\widehat{E}_k+\widehat{T}(y_k(\underline{\theta}))+\int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}[1-F_k(\theta)]\,\widehat{T}'(y_k(\theta))\,dy_k(\theta)\right] \\
&-\left\{\int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}[1-F_k(\theta)]\,\widehat{T}_k^{A\prime}(y_k(\theta))\,dy_k(\theta)-\sum_j\int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})}[1-F_k(\theta)]\,\overline{r}_{kj}(\theta)\,\widehat{T}_j^{A\prime}(y_j(\theta))\,dy_j(\theta)\right\} \\
&-\int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k(y_k(\theta))\,y_k(\theta)\,C_k(y_k(\theta))\left[1-T'(y_k(\theta))\right]\frac{\widehat{y}_k(\theta)}{y_k(\theta)}dy_k(\theta).
\end{aligned}
$$

Proof. The budget constraints of local governments before the central tax reform is:

$$
N_k E_k=\int_{\underline{\theta}}^{\overline{\theta}} T_k(y_k(\theta))\,f_k^p(\theta;U_k(\theta),U_{-k}(\theta))\,d\theta. \qquad \text{(D43)}
$$

The budget constraint after the tax reform is:

$$\begin{aligned} & N_k\left(T+\mu\widehat{T}\right)E_k\left(T+\mu\widehat{T}\right) \\ = & \int_{\underline{\theta}}^{\overline{\theta}} T_k\left(y_k(\theta;T+\mu\widehat{T});T+\mu\widehat{T}\right) f_k^p\left(\theta;V_k\left(\theta;T+\mu\widehat{T}\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)d\theta, \end{aligned} \tag{D44}$$

where

$$N_k=\int_{\underline{\theta}}^{\overline{\theta}} f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta\right)\right)d\theta, \tag{D45}$$

and

$$N_k\left(T+\mu\widehat{T}\right)=\int_{\underline{\theta}}^{\overline{\theta}} f_k^p\left(\theta;V_k\left(\theta;T+\mu\widehat{T}\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)d\theta. \tag{D46}$$

Subtract the items on both sides of equation (D43) from the items on the corresponding sides of equation (D44) to obtain the following equation:

$$\begin{aligned} & \frac{N_k\left(T+\mu\widehat{T}\right)E_k\left(T+\mu\widehat{T}\right)-E_kN_k}{\mu} \\ = & \frac{N_k\left(T+\mu\widehat{T}\right)-N_k}{\mu}E_k\left(T+\mu\widehat{T}\right)+\frac{E_k\left(T+\mu\widehat{T}\right)-E_k}{\mu}N_k \\ = & \int_{\underline{\theta}}^{\overline{\theta}}\frac{T_k\left(y_k(\theta;T+\mu\widehat{T});T+\mu\widehat{T}\right)-T_k\left(y_k(\theta)\right)}{\mu}f_k^p\left(\theta;V_k\left(\theta;T+\mu\widehat{T}\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)d\theta \\ & +\int_{\underline{\theta}}^{\overline{\theta}} T_k\left(y_k(\theta)\right)\frac{f_k^p\left(\theta;V_k\left(\theta;T+\mu\widehat{T}\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)-f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)}{\mu}d\theta \\ & +\int_{\underline{\theta}}^{\overline{\theta}} T_k\left(y_k(\theta)\right)\frac{f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)-f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta\right)\right)}{\mu}d\theta. \end{aligned} \tag{D47}$$

Subtract the items on both sides of equation (D45) from the items on the corresponding sides of equation (D46) to obtain the following equation:

$$\begin{aligned} & \frac{N_k\left(T+\mu\widehat{T}\right)-N_k}{\mu} \\ = & \int_{\underline{\theta}}^{\overline{\theta}}\frac{f_k^p\left(\theta;V_k\left(\theta;T+\mu\widehat{T}\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)-f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)}{\mu}d\theta \\ & +\int_{\underline{\theta}}^{\overline{\theta}}\frac{f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta;T+\mu\widehat{T}\right)\right)-f_k^p\left(\theta;V_k\left(\theta\right),V_{-k}\left(\theta\right)\right)}{\mu}d\theta. \end{aligned} \tag{D48}$$

3001

When $\mu \to 0$, equations (D47) and (D48) are reduced to:

$$\begin{aligned}\widehat{N}_k E_k + \widehat{E}_k N_k &= \int_{\underline{\theta}}^{\overline{\theta}} \left[\widehat{T}_k\left(y_k(\theta)\right) + T_k'\left(y_k(\theta)\right)\widehat{y}_k(\theta)\right] f_k^p\left(\theta; V_k\left(\theta\right), V_{-k}\left(\theta\right)\right) d\theta \\ &\quad + \sum_j \int_{\underline{\theta}}^{\overline{\theta}} T_k\left(y_k(\theta)\right) \frac{\partial f_k^p\left(\theta; V_k\left(\theta\right), V_{-k}\left(\theta\right)\right)}{\partial V_j\left(\theta\right)} \widehat{V}_j\left(\theta\right) d\theta\end{aligned} \tag{D49}$$

and

$$\widehat{N}_k = \sum_j \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial f_k^p\left(\theta; V_k\left(\theta\right), V_{-k}\left(\theta\right)\right)}{\partial V_j\left(\theta\right)} \widehat{V}_j\left(\theta\right) d\theta. \tag{D50}$$

Substitute $\widehat{N}_k$ in (D49) by (D50):

$$\begin{aligned}\widehat{E}_k N_k &= \int_{\underline{\theta}}^{\overline{\theta}} \left[\widehat{T}_k\left(y_k(\theta)\right) + T_k'\left(y_k(\theta)\right)\widehat{y}_k(\theta)\right] f_k^p\left(\theta; V_k\left(\theta\right), V_{-k}\left(\theta\right)\right) d\theta \\ &\quad + \sum_j \int_{\underline{\theta}}^{\overline{\theta}} \left[T_k\left(y_k(\theta)\right) - E_k\right] \frac{\partial f_k^p\left(\theta; V_k\left(\theta\right), V_{-k}\left(\theta\right)\right)}{\partial V_j\left(\theta\right)} \widehat{V}_j\left(\theta\right) d\theta\end{aligned}$$

By the definition of $r_{kj}\left(\theta\right)$, the above equation is equivalent to:

$$\widehat{E}_k = \int_{\underline{\theta}}^{\overline{\theta}} \left[\widehat{T}_k\left(y_k(\theta)\right) + T_k'\left(y_k(\theta)\right)\widehat{y}_k(\theta)\right] f_k\left(\theta\right) d\theta + \sum_j \int_{\underline{\theta}}^{\overline{\theta}} f_k\left(\theta\right) r_{kj}\left(\theta\right) \widehat{V}_j\left(\theta\right) d\theta$$

Using $d\left[1 - F_k\left(\theta\right)\right] = -f_k\left(\theta\right) d\theta$ and Lemma 13, i.e., $\widehat{T}_j^{\star}\left(y_j\left(\theta\right)\right) = \widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}\right)\right) + \int_{\underline{\theta}}^{\theta} \widehat{T}_j^{\star\prime}\left(y_j\left(x\right)\right) dy_j\left(x\right)$, the above equation can be rewritten as:

$$\begin{aligned}\widehat{E}_k &= \int_{\underline{\theta}}^{\overline{\theta}} \left[\widehat{T}_k^{\star}\left(y_k\left(\theta\right)\right) + T_k^{\star\prime}\left(y_k(\theta)\right)\widehat{y}_k(\theta)\right] f_k\left(\theta\right) d\theta + \sum_j \int_{\underline{\theta}}^{\overline{\theta}} f_k\left(\theta\right) r_{kj}\left(\theta\right) \widehat{V}_j\left(\theta\right) d\theta \\ &= -\int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}_k^{\star}\left(y_k\left(\theta\right)\right) d\left[1 - F_k\left(\theta\right)\right] + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} \frac{f_k\left(\theta\right)\theta}{\frac{d\ln y_k(\theta)}{d\ln\theta}} T_k^{\star\prime}\left(y_k(\theta)\right) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) \\ &\quad - \sum_j \int_{\underline{\theta}}^{\overline{\theta}} f_k\left(\theta\right) r_{kj}\left(\theta\right) \widehat{T}_j^{A}\left(y_j\left(\theta\right)\right) d\theta \\ &= -\int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}_k^{\star}\left(y_k\left(\theta\right)\right) d\left[1 - F_k\left(\theta\right)\right] + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k(\theta)\right) y_k(\theta) T_k^{\star\prime}\left(y_k(\theta)\right) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) \\ &\quad - \sum_j \int_{\underline{\theta}}^{\overline{\theta}} f_k\left(\theta\right) r_{kj}\left(\theta\right) \widehat{T}_j^{A}\left(y_j\left(\theta\right)\right) d\theta.\end{aligned}$$

By the definition of $\overline{r}_{jk}\left(\theta\right)$ and $d\left\{\left[1 - F_k\left(\theta\right)\right]\overline{r}_{kj}\left(\theta\right)\right\} = -f_k\left(\theta\right) r_{kj}\left(\theta\right) d\theta$, the above equation is equivalent to:

$$\begin{aligned}\widehat{E}_k &= -\int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}_k^{\star}\left(y_k\left(\theta\right)\right) d\left[1 - F_k\left(\theta\right)\right] + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k(\theta)\right) y_k(\theta) T_k^{\star\prime}\left(y_k(\theta)\right) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) \\ &\quad + \sum_j \int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}_j^{A}\left(y_j\left(\theta\right)\right) d\left\{\left[1 - F_k\left(\theta\right)\right]\overline{r}_{kj}\left(\theta\right)\right\}.\end{aligned}$$

Using integration by part, the above equation is transformed to:

$$\begin{aligned}\widehat{E}_k &= \widehat{T}_k^{\star}\left(y_k\left(\underline{\theta}\right)\right) - \sum_j \overline{r}_{kj}\left(\underline{\theta}\right)\widehat{T}_j^{A}\left(y_j\left(\underline{\theta}\right)\right) + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\widehat{T}_k^{\star\prime}\left(y_k\left(\theta\right)\right)dy_k\left(\theta\right) \\ &\quad - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\overline{r}_{kj}\left(\theta\right)d\widehat{T}_j^{A}\left(y_j\left(\theta\right)\right) - \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k\left(\theta\right)\right)y_k\left(\theta\right)T_k^{\star\prime}\left(y_k(\theta)\right)\frac{\widehat{y}_k(\theta)}{y_k(\theta)}dy_k(\theta).\end{aligned}$$

Add $\widehat{T}\left(y_k\left(\underline{\theta}\right)\right) + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\widehat{T}'\left(y_k\left(\theta\right)\right)dy_k\left(\theta\right)$ in both sides of above equation to get the following expression:

$$\begin{aligned}&\widehat{E}_k + \widehat{T}\left(y_k\left(\underline{\theta}\right)\right) + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\widehat{T}'\left(y_k\left(\theta\right)\right)dy_k\left(\theta\right) \\ = \;& \widehat{T}_k^{A}\left(y_k\left(\underline{\theta}\right)\right) - \sum_j \overline{r}_{kj}\left(\underline{\theta}\right)\widehat{T}_j^{A}\left(y_j\left(\underline{\theta}\right)\right) + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\widehat{T}_k^{A\prime}\left(y_k\left(\theta\right)\right)dy_k\left(\theta\right) \\ &- \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\overline{r}_{kj}\left(\theta\right)d\widehat{T}_j^{A}\left(y_j\left(\theta\right)\right) - \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k\left(\theta\right)\right)y_k\left(\theta\right)T_k^{\star\prime}\left(y_k(\theta)\right)\frac{\widehat{y}_k(\theta)}{y_k(\theta)}dy_k(\theta).\end{aligned}$$

Rearrange the above equation:

$$\begin{aligned}&\widehat{T}_k^{A}\left(y_k\left(\underline{\theta}\right)\right) - \sum_j \overline{r}_{kj}\left(\underline{\theta}\right)\widehat{T}_j^{A}\left(y_j\left(\underline{\theta}\right)\right) \\ = \;& \left[\widehat{E}_k + \widehat{T}\left(y_k\left(\underline{\theta}\right)\right) + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\widehat{T}'\left(y_k\left(\theta\right)\right)dy_k\left(\theta\right)\right] \\ &- \left\{\begin{array}{c}\int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\widehat{T}_k^{A\prime}\left(y_k\left(\theta\right)\right)dy_k\left(\theta\right) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})}\left[1-F_k\left(\theta\right)\right]\overline{r}_{kj}\left(\theta\right)d\widehat{T}_j^{A}\left(y_j\left(\theta\right)\right) \\ + \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k\left(\theta\right)\right)y_k\left(\theta\right)T_k^{\star\prime}\left(y_k(\theta)\right)\frac{\widehat{y}_k(\theta)}{y_k(\theta)}dy_k(\theta)\end{array}\right\}.\end{aligned}$$

Notice that $T_k^{\star\prime}\left(y_k\left(\theta\right)\right) = C_k\left(y_k\left(\theta\right)\right)\left[1-T'\left(y_k\left(\theta\right)\right)\right]$, the above equation is equivalent to equation (D42).■

D.3.4 Result 4: Central Government's Budget Constraint

We set the following notations for later use**:**

$$\begin{aligned}r_{kj}^{G}\left(\theta\right) &\equiv r_{kj}\left(\theta\right) + \sum_n \frac{N_n}{\sum_n N_n}\frac{f_n\left(\theta\right)}{f_k\left(\theta\right)}r_{nj}\left(\theta\right); \\ \overline{r}_{kj}^{G}\left(\theta\right) &\equiv \int_{\theta}^{\overline{\theta}} r_{kj}^{C}\left(x\right)\frac{f_k\left(x\right)}{1-F_k\left(\theta\right)}dx\end{aligned}$$

Under the above definitions, we have $\overline{r}_{kj}^{G}\left(\theta\right) = \int_{\theta}^{\overline{\theta}} r_{kj}\left(x\right)\frac{f_k(x)}{1-F_k(\theta)}dx + \sum_n \frac{N_n}{\sum_n N_n}\int_{\theta}^{\overline{\theta}} r_{nj}\left(x\right)\frac{f_n(x)}{1-F_k(\theta)}dx$ and $\overline{r}_{kj}^{G}\left(\underline{\theta}\right) = \overline{r}_{kj}\left(\underline{\theta}\right) + \sum_n \frac{N_n}{\sum_n N_n}\int_{\underline{\theta}}^{\overline{\theta}} r_{nj}\left(\theta\right)f_n\left(\theta\right)d\theta$. Define $\left[\eta_{kj}^{G}\right]$ as the inverse matrix of $I - \left[\overline{r}_{kj}^{G}\left(\underline{\theta}\right)\right]$, i.e., $\left[\eta_{kj}^{G}\right] = \left[I - \left[\overline{r}_{kj}^{G}\left(\underline{\theta}\right)\right]\right]^{-1}$. Under the above definitions, we have the following result by the before- and after-tax-

reform budget constraints of the central government:

$$\begin{aligned}\widehat{T}\left(\underline{y}\right) &= \sum_j \widehat{T}_j^A\left(y_j\left(\underline{\theta}\right)\right) \times \left\{\sum_n \frac{N_n}{\sum_n N_n}\int_{\underline{\theta}}^{\overline{\theta}}\left[T\left(y_n(x)\right)+E_n\right] f_n\left(\theta\right) m_{nj}\left(\theta\right) d\theta\right\} \\ &- \sum_n \frac{N_n}{\sum_n N_n}\widehat{E}_n - \sum_n \frac{N_n}{\sum_n N_n}\int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n\left(\theta\right)\right] \\ &- \sum_n \frac{N_n}{\sum_n N_n}\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\left[1-F_n\left(\theta\right)\right]\widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \\ &+ \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})}\left\{\sum_n \frac{N_n}{\sum_n N_n}\int_{\theta}^{\overline{\theta}}\left[T\left(y_n(x)\right)+E_n\right] f_n\left(x\right) m_{nj}\left(x\right) dx\right\}\widehat{T}_j^{A\prime}\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) \\ &- \sum_n \frac{N_n}{\sum_n N_n}\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right)\frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta).\end{aligned} \tag{D51}$$

The above equation together with equation (D42) gives:

$$\begin{aligned}&\widehat{T}_n^A\left(\underline{y}_n\right) = \sum_j \eta_{nj}^G\left\{\widehat{E}_j + \int_{\underline{y}}^{\underline{y}_j} d\widehat{T}\left[y_j\right] - \sum_m \frac{N_m}{\sum_m N_m}\widehat{E}_m - \sum_m \frac{N_m}{\sum_m N_m}\int_{\underline{y}}^{\underline{y}_m} d\widehat{T}\left[y_m\right]\right\} \\ &+ \sum_j \int_{\underline{y}_j}^{\overline{y}_j}\left[\eta_{kj}^G - \sum_m \eta_{nm}^G \frac{N_j}{\sum_m N_m}\right]\left[1-F_j\left(y_j\right)\right]\left[1-T'\left(y_j\right)\right]\frac{\widehat{T}'\left(y_j\right)}{1-T'\left(y_j\right)} dy_j \\ &+ \sum_j \int_{\underline{y}_j}^{\overline{y}_j}\left\{\begin{array}{c}\left[\eta_{nj}^G\left[1-F_j\left(y_j\right)\right] - \sum_m \eta_{nm}^G\left[1-F_m\left(y_m^j\left(y_j\right)\right)\right]\overline{r}_{mj}^G\left(y_m^j\left(y_j\right)\right)\right]\frac{1-T_j^{A\prime}\left(y_j\right)}{\varepsilon_j^A\left(y_j\right)} \\ -\left[\eta_{nj}^G C_j\left(y_j\right) - \sum_m \eta_{nm}^G \frac{N_j}{\sum_n N_j}\frac{T'\left(y_j\right)}{1-T'\left(y_j\right)}\right] f_j\left(y_j\right) y_j\left[1-T'\left(y_j\right)\right]\end{array}\right\}\frac{\widehat{y}_j}{y_j} dy_j.\end{aligned} \tag{D52}$$

and

$$\begin{aligned}&\widehat{T}\left(\underline{y}_n\right) + \widehat{E}_n \\ &= \left\{\widehat{E}_n + \int_{\underline{y}}^{\underline{y}_n} d\widehat{T}\left[y_n\right] - \sum_m \frac{N_m}{\sum_m N_m}\widehat{E}_m - \sum_m \frac{N_m}{\sum_m N_m}\int_{\underline{y}}^{\underline{y}_m} d\widehat{T}\left[y_m\right]\right\} - \\ &\sum_j \int_{\underline{y}_j}^{\overline{y}_j}\frac{N_j\left[1-F_j\left(y_j\right)\right]\left[1-T'\left(y_j\right)\right]}{\sum_m N_m}\frac{\widehat{T}'\left(y_j\right)}{1-T'\left(y_j\right)} dy_j - \sum_j \int_{\underline{y}_j}^{\overline{y}_j}\frac{N_j f_j\left(y_j\right) y_j T'\left(y_j\right)}{\sum_m N_m}\frac{\widehat{y}_j}{y_j} dy_j.\end{aligned} \tag{D53}$$

Proof. We now prove that there are equations (D51) to (D53). The budget constraint of the central government before the tax reform is:

$$0 = \sum_j \int_{\underline{\theta}}^{\overline{\theta}} T\left(y_j(\theta)\right) f_j^p\left(\theta; V_j\left(\theta\right), V_{-j}\left(\theta\right)\right) d\theta + \sum_j N_j E_j. \tag{D54}$$

The budget constraint of the central government after the tax reform is:

$$\begin{aligned}0 &= \sum_j \int_{\underline{\theta}}^{\overline{\theta}}\left[T\left(y_j(\theta; T+\mu\widehat{T})\right) + \mu\widehat{T}\left(y_j\left(\theta, T+\mu\widehat{T}\right)\right)\right] f_j^p\left(\theta; V_j\left(\theta; T+\mu\widehat{T}\right), V_{-j}\left(\theta; T+\mu\widehat{T}\right)\right) d\theta_j \\ &+ \sum_j N_j\left(T+\mu\widehat{T}\right) E_j\left(T+\mu\widehat{T}\right),\end{aligned} \tag{D55}$$

where we denote by $N_j\left(T+\mu\widehat{T}\right)$ and $E_j\left(T+\mu\widehat{T}\right)$ the population and regional transfer after the tax reform.

Subtract the items on both sides of equation (D55) from the items on the corresponding sides of equation (D54) to obtain the following equation:

$$\begin{aligned}
0 &= \sum_j \int_{\underline{\theta}}^{\overline{\theta}} \left[\begin{array}{c} \frac{T\left(y_j(\theta;T+\mu\widehat{T})\right)+\mu\widehat{T}\left(y_j\left(\theta,T+\mu\widehat{T}\right)\right)-T\left(y_j(\theta)\right)}{\mu} \times \\ f_j^p\left(\theta;V_j\left(\theta;T+\mu\widehat{T}\right),V_{-j}\left(\theta;T+\mu\widehat{T}\right)\right) \end{array} \right] d\theta_j \\
&\quad + \sum_j \int_{\underline{\theta}}^{\overline{\theta}} T\left(y_j(\theta)\right) \frac{f_j^p\left(\theta;V_j\left(\theta;T+\mu\widehat{T}\right),V_{-j}\left(\theta;T+\mu\widehat{T}\right)\right)-f_j^p\left(\theta;V_j\left(\theta\right),V_{-j}\left(\theta\right)\right)}{\mu} d\theta_j \\
&\quad + \sum_j \frac{N_j\left(T+\mu\widehat{T}\right)E_j\left(T+\mu\widehat{T}\right)-N_jE_j}{\mu}.
\end{aligned}$$

When $\mu \to 0$, we have:

$$\begin{aligned}
0 &= \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} \left[T'\left(y_k(\theta)\right)\widehat{y}_k(\theta)+\widehat{T}\left(y_k(\theta)\right)+\widehat{E}_k\right] f_k\left(\theta\right) d\theta \\
&\quad - \sum_k \sum_j N_j \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_j(\theta)\right)+E_j\right] f_j\left(\theta\right) m_{jk}\left(\theta\right) \widehat{T}_k^A\left(y_k\left(\theta\right)\right) d\theta \\
&= \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} T'\left(y_k(\theta)\right) \frac{f_k\left(\theta\right)\theta}{\frac{d\ln y_k(\theta)}{d\ln\theta}} \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) + \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}\left(y_k(\theta)\right) f_k\left(\theta\right) d\theta \\
&\quad + \sum_k N_k\widehat{E}_k - \sum_k \sum_j N_j \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_j(\theta)\right)+E_j\right] f_j\left(\theta\right) m_{jk}\left(\theta\right) \widehat{T}_k^A\left(y_k\left(\theta\right)\right) d\theta \\
&\quad + \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} T'\left(y_k(\theta)\right) f_k\left(y_k(\theta)\right) y_k(\theta) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) + \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}\left(y_k(\theta)\right) f_k\left(\theta\right) d\theta \\
&\quad + \sum_k N_k\widehat{E}_k - \sum_k \sum_j N_j \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_j(\theta)\right)+E_j\right] f_j\left(\theta\right) m_{jk}\left(\theta\right) \widehat{T}_k^A\left(y_k\left(\theta\right)\right) d\theta
\end{aligned}$$

Using $d\left[1-F_k\left(\theta\right)\right] = -f_k\left(\theta\right)d\theta$, we can represent above equation as below:

$$\begin{aligned}
0 &= \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} T'\left(y_k(\theta)\right) f_k\left(y_k(\theta)\right) y_k(\theta) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) - \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}\left(y_k(\theta)\right) d\left[1-F_k\left(\theta\right)\right] \\
&\quad + \sum_k N_k\widehat{E}_k - \sum_k \sum_j N_j \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_j(\theta)\right)+E_j\right] f_j\left(\theta\right) m_{jk}\left(\theta\right) \widehat{T}_k^A\left(y_k\left(\theta\right)\right) d\theta
\end{aligned}$$

Using $\widehat{T}_k^A\left(y_k\left(\theta\right)\right) = \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{A\prime}\left(y_k\left(x\right)\right) dy_k\left(x\right)$ (see Lemma 14) to substitute $\widehat{T}_k^A\left(y_k\left(\theta\right)\right)$ in the

above equation:

$$\begin{aligned}
0 &= \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} T'\left(y_k(\theta)\right) f_k\left(y_k(\theta)\right) y_k(\theta) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) + \sum_k N_k \widehat{E}_k - \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} \widehat{T}\left(y_k(\theta)\right) d\left[1 - F_k(\theta)\right] \\
&\quad - \sum_k \sum_j N_j \left\{ \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_j(\theta)\right) + E_j\right] f_j(\theta) m_{jk}(\theta) d\theta \right\} \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) \\
&\quad - \sum_k \sum_j N_j \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} \widehat{T}\left[T\left(y_j(\theta)\right) + E_j\right] f_j(\theta) m_{jk}(\theta) \widehat{T}_k^{A\prime}\left(y_k(x)\right) dy_k(x) d\theta
\end{aligned}$$

Using the Fubini Theorem to change the order of integral and using integration by part, we have the following result:

$$\begin{aligned}
0 &= \sum_k N_k \widehat{T}\left(y_k(\underline{\theta})\right) - \sum_k \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) \times \left\{ \sum_j N_j \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_j(\theta)\right) + E_j\right] f_j(\theta) m_{jk}(\theta) d\theta \right\} \\
&\quad + \sum_k N_k \widehat{E}_k + \sum_k N_k \int_{\underline{\theta}}^{\overline{\theta}} \left[1 - F_k(\theta)\right] d\widehat{T}\left(y_k(\theta)\right) \\
&\quad - \sum_k \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} \left\{ \sum_j N_j \int_{\theta}^{\overline{\theta}} \left[T\left(y_j(x)\right) + E_j\right] f_j(x) m_{jk}(x) dx \right\} \widehat{T}_k^{A\prime}\left(y_k(\theta)\right) dy_k(\theta) \\
&\quad + \sum_k N_k \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k(\theta)\right) y_k(\theta) T'\left(y_k(\theta)\right) \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta).
\end{aligned}$$

Notice that $\sum_k N_k = 1$ and $\widehat{T}\left(y_k\left(\underline{\theta}\right)\right) = \widehat{T}\left(\underline{y}\right) + \int_{\underline{y}}^{y_k(\underline{\theta})} d\widehat{T}\left(y_k(\theta)\right)$. We have equation (D51) by the above equation.

We can use $\widehat{T}\left(y_k\left(\underline{\theta}\right)\right) = \widehat{T}\left(\underline{y}\right) + \int_{\underline{y}}^{y_k(\underline{\theta})} d\widehat{T}\left(y_k(\theta)\right)$ (see Lemma 13) to replace $\widehat{T}\left(y_k(\underline{\theta})\right)$ in equation (D42), and use (D51) to replace $\widehat{T}\left(\underline{y}\right)$ in (D42), to derive the following equation:

$$\begin{aligned}
&\widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) - \sum_j \widehat{T}_j^A\left(y_j\left(\underline{\theta}\right)\right) \times \left\{ \overline{r}_{kj}\left(\underline{\theta}\right) + \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_n(\theta)\right) + E_n\right] f_n(\theta) m_{nj}(\theta) d\theta \right\} \qquad \text{(D56)} \\
= &- \left\{ \begin{array}{c} \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} \left[1 - F_k(\theta)\right] \widehat{T}_k^{A\prime}\left(y_k(\theta)\right) dy_k(\theta) - \\ \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left\{ \left[1 - F_k(\theta)\right] \overline{r}_{kj}(\theta) + \sum_n \frac{N_n}{\sum_n N_n} \int_{\theta}^{\overline{\theta}} \left[T\left(y_n(x)\right) + E_n\right] f_n(x) m_{nj}(x) dx \right\} \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta) \end{array} \right\} \\
&+ \left\{ \widehat{E}_k - \sum_n \frac{N_n}{\sum_n N_n} \widehat{E}_n \right\} + \left\{ \int_{\underline{y}}^{y_k(\underline{\theta})} d\widehat{T}\left[y_k(\theta)\right] - \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{\underline{y}_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right] \right\} \\
&+ \left\{ \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} \left[1 - F_k(\theta)\right] \widehat{T}'\left(y_k(\theta)\right) dy_k(\theta) - \sum_n \frac{N_n}{\sum_n N_n} \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left[1 - F_n(\theta)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \right\} \\
&- \left[\begin{array}{c} \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k(\theta)\right) y_k(\theta) C_k\left(y_k(\theta)\right) \left[1 - T'\left(y_k(\theta)\right)\right] \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) \\ + \sum_n \frac{N_n}{\sum_n N_n} \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \end{array} \right].
\end{aligned}$$

Remember that by definitions of $\overline{r}_{kj}(\theta)$ and $\overline{r}^{G}_{kj}(\theta)$ (see equation (32)):

$$\begin{aligned}
\overline{r}^{G}_{kj}(\theta) &= \overline{r}_{kj}(\theta) + \sum_n \frac{N_n}{\sum_n N_n} \int_{\theta}^{\overline{\theta}} \left[T\left(y_n(x)\right) + E_n\right] m_{nj}(x) \frac{f_n(x)}{1 - F_k(\theta)} dx \\
\overline{r}^{G}_{kj}(\underline{\theta}) &= \overline{r}_{kj}(\underline{\theta}) + \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{\theta}}^{\overline{\theta}} \left[T\left(y_n(\theta)\right) + E_n\right] f_n(\theta) m_{nj}(\theta) d\theta.
\end{aligned}$$

Therefore, equation (D56) can be simplified as:

$$\begin{aligned}
&\widehat{T}^{A}_{k}\left(y_k\left(\underline{\theta}\right)\right) - \sum_j \overline{r}^{G}_{kj}\left(\underline{\theta}\right) \times \widehat{T}^{A}_{j}\left(y_j\left(\underline{\theta}\right)\right) \qquad \text{(D57)} \\
= &- \left\{ \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} \left[1 - F_k(\theta)\right] \widehat{T}^{A\prime}_{k}\left(y_k(\theta)\right) dy_k(\theta) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left[1 - F_k(\theta)\right] \overline{r}^{G}_{kj}(\theta) \widehat{T}^{A\prime}_{j}\left(y_j(\theta)\right) dy_j(\theta) \right\} \\
&+ \left\{ \widehat{E}_k - \sum_n \frac{N_n}{\sum_n N_n} \widehat{E}_n \right\} + \left\{ \int_{\underline{y}}^{y_k(\theta)} d\widehat{T}\left[y_k(\theta)\right] - \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{\underline{y}_n(\theta)} d\widehat{T}\left[y_n(\theta)\right] \right\} \\
&+ \left\{ \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} \left[1 - F_k(\theta)\right] \widehat{T}'\left(y_k(\theta)\right) dy_k(\theta) - \sum_n \frac{N_n}{\sum_n N_n} \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left[1 - F_n(\theta)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \right\} \\
&- \begin{bmatrix} \int_{y_k(\underline{\theta})}^{y_k(\overline{\theta})} f_k\left(y_k(\theta)\right) y_k(\theta) C_k\left(y_k(\theta)\right) \left[1 - T'\left(y_k(\theta)\right)\right] \frac{\widehat{y}_k(\theta)}{y_k(\theta)} dy_k(\theta) \\ + \sum_n \frac{N_n}{\sum_n N_n} \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \end{bmatrix}.
\end{aligned}$$

Using $\left[\eta^{G}_{kj}\right] = \left[I - \left[\overline{r}^{G}_{kj}\left(\underline{\theta}\right)\right]\right]^{-1}$ and solve equation (D57) for $\widehat{T}^{A}_{k}\left(y_k\left(\underline{\theta}\right)\right)$ gives:

$$\begin{aligned}
&\widehat{T}^{A}_{k}\left(y_k\left(\underline{\theta}\right)\right) = \sum_j \eta^{G}_{kj} \left\{ \widehat{E}_j - \sum_n \frac{N_n}{\sum_n N_n} \widehat{E}_n \right\} + \sum_j \eta^{G}_{kj} \left\{ \begin{array}{c} \int_{\underline{y}}^{y_j(\underline{\theta})} d\widehat{T}\left[y_j(\theta)\right] - \\ \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right] \end{array} \right\} \qquad \text{(D58)} \\
&+ \left\{ \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta^{G}_{kj} \left[1 - F_j(\theta)\right] \widehat{T}'\left(y_j(\theta)\right) dy_j(\theta) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left(\sum_m \eta^{G}_{km} \right) \frac{N_j}{\sum_m N_m} \left[1 - F_j(\theta)\right] \widehat{T}'\left(y_j(\theta)\right) dy_j(\theta) \right\} \\
&- \left\{ \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta^{G}_{kj} \left[1 - F_j(\theta)\right] \widehat{T}^{A\prime}_{j}\left(y_j(\theta)\right) dy_j(\theta) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_m \eta^{G}_{km} \left[1 - F_m(\theta)\right] \overline{r}^{G}_{mj}(\theta) \widehat{T}^{A\prime}_{j}\left(y_j(\theta)\right) dy_j(\theta) \right\} \\
&- \begin{bmatrix} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta^{G}_{kj} f_j\left(y_j(\theta)\right) y_j(\theta) C_j\left(y_j(\theta)\right) \left[1 - T'\left(y_j(\theta)\right)\right] \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta) + \\ \left(\sum_m \eta^{G}_{km}\right) \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{N_j}{\sum_m N_m} f_j\left(y_j(\theta)\right) y_j(\theta) \frac{T'\left(y_j(\theta)\right)}{1 - T'\left(y_j(\theta)\right)} \left[1 - T'\left(y_j(\theta)\right)\right] \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta) \end{bmatrix}.
\end{aligned}$$

Notice that we have the following two equations

$$\begin{aligned}
&\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta^{G}_{kj} \left[1 - F_j(\theta)\right] \widehat{T}'\left(y_j(\theta)\right) dy_j(\theta) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left(\sum_m \eta^{G}_{km} \right) \frac{N_j \left[1 - F_j(\theta)\right] \widehat{T}'\left(y_j(\theta)\right)}{\sum_m N_m} dy_j(\theta) \qquad \text{(D59)} \\
= &\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left[\eta^{G}_{nj} - \sum_m \eta^{G}_{nm} \frac{N_j}{\sum_m N_m} \right] \left[1 - F_j\left(y_j(\theta)\right)\right] \left[1 - T'\left(y_j(\theta)\right)\right] \frac{\widehat{T}'\left(y_j(\theta)\right)}{1 - T'\left(y_j(\theta)\right)} dy_j(\theta),
\end{aligned}$$

and

$$-\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta_{kj}^G \left[1-F_j(\theta)\right] \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_m \eta_{km}^G \left[1-F_m(\theta)\right] \overline{r}_{mj}^G(\theta) \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta) \tag{D60}$$

$$= \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left[\eta_{nj}^G \left[1-F_j\left(y_j(\theta)\right)\right] - \sum_m \eta_{nm}^G \left[1-F_m\left(y_m(\theta)\right)\right] \overline{r}_{mj}^G\left(y_m(\theta)\right)\right] \frac{1-T_j^{A\prime}\left(y_j(\theta)\right)}{\varepsilon_j^A\left(y_j(\theta)\right)} \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta),$$

where we have used $\widehat{T}_j^{A\prime}\left(y_j(\theta)\right) = -\frac{1-T_j^{A\prime}\left(y_j(\theta)\right)}{\varepsilon_j^A\left(y_j(\theta)\right)} \frac{\widehat{y}_j(\theta)}{y_j(\theta)}$ in equation (D60).

Using (D60) to substitute a term in the right side of (D58):

$$-\left\{\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta_{kj}^G \left[1-F_j(\theta)\right] \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta) - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_m \eta_{km}^G \left[1-F_m(\theta)\right] \overline{r}_{mj}^G(\theta) \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta)\right\} \tag{D61}$$

$$-\begin{bmatrix} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \eta_{kj}^G f_j\left(y_j(\theta)\right) y_j(\theta) C_j\left(y_j(\theta)\right) \left[1-T'\left(y_j(\theta)\right)\right] \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta) \\ + \left(\sum_m \eta_{km}^G\right) \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{N_j}{\sum_m N_m} f_j\left(y_j(\theta)\right) y_j(\theta) \frac{T'\left(y_j(\theta)\right)}{1-T'\left(y_j(\theta)\right)} \left[1-T'\left(y_j(\theta)\right)\right] \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta) \end{bmatrix}$$

$$= \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \begin{Bmatrix} \left[\eta_{nj}^G \left[1-F_j\left(y_j(\theta)\right)\right] - \sum_m \eta_{nm}^G \left[1-F_m\left(y_m(\theta)\right)\right] \overline{r}_{mj}^G\left(y_m(\theta)\right)\right] \frac{1-T_j^{A\prime}\left(y_j(\theta)\right)}{\varepsilon_j^A\left(y_j(\theta)\right)} \\ -\left[\eta_{nj}^G C_j\left(y_j(\theta)\right) - \sum_m \eta_{nm}^G \frac{N_j}{\sum_m N_m} \frac{T'\left(y_j(\theta)\right)}{1-T'\left(y_j(\theta)\right)}\right] f_j\left(y_j(\theta)\right) y_j(\theta) \left[1-T'\left(y_j(\theta)\right)\right] \end{Bmatrix} \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta)$$

Substitute terms in the right side of (D58) by (D61) and (D59) gives:

$$\widehat{T}_n^A\left(y_n(\underline{\theta})\right) = \sum_j \eta_{nj}^G \left\{\widehat{E}_j - \sum_m \frac{N_m}{\sum_m N_m} \widehat{E}_m\right\} \tag{D62}$$

$$+\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \left[\eta_{nj}^G - \sum_m \eta_{nm}^G \frac{N_j}{\sum_m N_m}\right] \left[1-F_j\left(y_j(\theta)\right)\right] \left[1-T'\left(y_j(\theta)\right)\right] \frac{\widehat{T}'\left(y_j(\theta)\right)}{1-T'\left(y_j(\theta)\right)} dy_j(\theta)$$

$$+\sum_j \eta_{nj}^G \left\{\int_{\underline{y}}^{y_j(\theta)} d\widehat{T}\left[y_j(\theta)\right] - \sum_m \frac{N_m}{\sum_m N_m} \int_{\underline{y}}^{y_m(\theta)} d\widehat{T}\left[y_m(\theta)\right]\right\}$$

$$+\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \begin{Bmatrix} \left[\eta_{nj}^G \left[1-F_j\left(y_j(\theta)\right)\right] - \sum_m \eta_{nm}^G \left[1-F_m\left(y_m(\theta)\right)\right] \overline{r}_{mj}^G\left(y_m(\theta)\right)\right] \frac{1-T_j^{A\prime}\left(y_j(\theta)\right)}{\varepsilon_j^A\left(y_j(\theta)\right)} \\ -\left[\eta_{nj}^G C_j\left(y_j(\theta)\right) - \sum_m \eta_{nm}^G \frac{N_j}{\sum_m N_m} \frac{T'\left(y_j(\theta)\right)}{1-T'\left(y_j(\theta)\right)}\right] f_j\left(y_j(\theta)\right) y_j(\theta) \left[1-T'\left(y_j(\theta)\right)\right] \end{Bmatrix} \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta).$$

Rearrange (D62) gives equation (D52). Using $\widehat{T}\left(y_k\left(\underline{\theta}_k\right)\right) = \widehat{T}\left(\underline{y}\right) + \int_{\underline{y}}^{\underline{y}_k(\theta)} d\widehat{T}\left(y_k(\theta)\right)$ (see Lemma 13) and equation (D51), eqaution (D52) can be transformed to be equation (D53).■

D.3.5 Result 5: Elasticity of Labor Income

We set notations and conditions for the following analysis. Set the following notations:

$$
\begin{aligned}
e_{kj}\left(y,y'\right) &= e_{kj}\left(y\right)\frac{1-T_j^{A\prime}\left(y'\right)}{\varepsilon_j^A\left(y'\right)}, e_{kk}^{C,own}\left(y_k,y_k'\right)=\int_{y_k'}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k''\right)dy_k\left(s_k\right)\left[1-T'\left(y_k'\right)\right], \qquad \text{(D63)}\\
e_{kj}^{C,cross}\left(y_k,y_j\right) &= \left\{\begin{array}{c}\sum_n\left[e_{kn}\left(y_k\right)+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}\left(y_k,y_n'\right)dy_n'\right]\left[\eta_{nj}^G-\sum_m\eta_{nm}^G\frac{N_j}{\sum_m N_m}\right]\\ -\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k'\right)dy_k'\right]\left[\eta_{kj}^G+\frac{N_j}{\sum_m N_m}\left(1-\sum_m\eta_{km}^G\right)\right]\end{array}\right\}\left[1-F_j\left(y_j\right)\right]\left[1-T'\left(y_j\right)\right],\\
e_{kk}^{E,own}\left(y_k\right) &= \int_{\underline{y}}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k'\right)dy_k', e_{kj}^{E,cross}\left(y_k\right)=\left\{\eta_{kj}^G\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k'\right)dy_k'-\sum_n\left[e_{kn}\left(y_k\right)+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}\left(y_k,y_n'\right)dy_n'\right]\eta_{nj}^G\right\},\\
e_{kk}^{y,own}\left(y_k,y_k'\right) &= e_{kk}^{own}\left(y_k,y_k'\right)T_k^{\star\prime}\left(y_k'\right)y_k'-\frac{1-T_k^{A\prime}\left(y_k'\right)}{\varepsilon_k^A\left(y_k'\right)}\int_{y_k'}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k''\right)dy_k'',\\
e_{kj}^{y,cross}\left(y_k,y_j\right) &= \left[\begin{array}{c}\int_{y_j}^{\overline{y}} e_{kj}^{cross}\left(y_k,y_j''\right)dy_j''\frac{1-T_j^{A\prime}(y_j)}{\varepsilon_j^A(y_j)}+\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k'\right)dy_k'\right]\left\{\begin{array}{c}\eta_{kj}^G\left[1-F_j\left(y_j\right)\right]-\\ \sum_m\eta_{km}^G\left[1-F_m\left(y_m^j\left(y_j\right)\right)\right]\overline{r}_{mj}^G\left(y_m^j\left(y_j\right)\right)\end{array}\right\}\frac{1-T_j^{A\prime}(y_j)}{\varepsilon_j^A(y_j)}\\ -\sum_n\left[e_{kn}\left(y_k\right)+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}\left(y_k,y_n'\right)dy_n'\right]\left\{\eta_{nj}^G\left[1-F_j\left(y_j\right)\right]-\sum_m\eta_{nm}^G\left[1-F_m\left(y_m^j\left(y_j\right)\right)\right]\overline{r}_{mj}^G\left(y_m^j\left(y_j\right)\right)\right\}\frac{1-T_j^{A\prime}(y_j)}{\varepsilon_j^A(y_j)}\\ -\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}\left(y_k,y_k'\right)dy_k'\right]\left[\eta_{kj}^G C_j\left(y_j\right)-\left(1+\sum_m\eta_{km}^G\right)\frac{N_j}{\sum_n N_j}\frac{T'(y_j))}{1-T'(y_j)}\right]f_j\left(y_j\right))y_j\left[1-T'\left(y_j\right)\right]\\ +\sum_n\left[e_{kn}\left(y_k\right)+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}\left(y_k,y_n'\right)dy_n'\right]\left[\eta_{kj}^G C_j\left(y_j\right)-\sum_m\eta_{nm}^G\frac{N_j}{\sum_n N_j}\frac{T'(y_j))}{1-T'(y_j)}\right]f_j\left(y_j\right))y_j\left[1-T'\left(y_j\right)\right]\end{array}\right].
\end{aligned}
$$

The following assumption is made to facilitate the resolvent used to solve the integral equation. For any $j\in\mathcal{J}$ and $\theta\in\Theta_j$, set

$$
\begin{aligned}
c_e^y \equiv & \left\{\sum_{j=1}^K\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kj}\left(y_k,y_j'\right)\right]^2\left|\frac{\varepsilon_j^A\left(y_j'\right)\left(\frac{1}{\varepsilon_j}+1\right)}{\varepsilon_k^A\left(y_k\right)\left(\frac{1}{\varepsilon_k}+1\right)}\right|dy_j'dy_k\right\}^{\frac{1}{2}}\\
&+\left\{\sum_{j=1}^K\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kj}^{y,cross}\left(y_k,y_j'\right)\right]^2\left|\frac{\varepsilon_j^A\left(y_j'\right)\left(\frac{1}{\varepsilon_j}+1\right)}{\varepsilon_k^A\left(y_k\right)\left(\frac{1}{\varepsilon_k}+1\right)}\right|dy_j'dy_k\right\}^{\frac{1}{2}}\\
&+\left\{\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kk}^{y,own}\left(y_k,y_k'\right)\right]^2\left|\frac{\varepsilon_k^A\left(y_k'\right)}{\varepsilon_k^A\left(y_k\right)}\right|dy_k'dy_k\right\}^{\frac{1}{2}}.
\end{aligned}
$$

Assumption 2 $c_e^y<1$ *and*

$$\sum_{k=1}^{K}\int_{\underline{y}}^{\overline{y}}\left|1-T'\left(y_k\right)\right|^2+\left|1-T_k^{A\prime}\left(y_k\right)\right|^2+\left|\frac{1}{1-T'\left(y_k\right)}\right|^2+\left|\frac{1}{1-T_k^{A\prime}\left(y_k\right)}\right|^2 dy_k<+\infty,$$

$$\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\underline{y}}^{\overline{y}}\left|\overline{r}_{kj}^{G}\left(y_k\right)\right|^2 dy_k<+\infty,$$

$$\sum_{k=1}^{K}\int_{\underline{y}}^{\overline{y}}\left|\varepsilon_k^{A}\left(y_k\right)\right|^2+\left|\frac{1}{\varepsilon_k^{A}\left(y_k\right)}\right|^2 dy_k<+\infty,$$

$$\sum_{k=1}^{K}\int_{\underline{y}}^{\overline{y}}\left|\varepsilon_k^{C}\left(y_k\right)\right|^2+\left|\frac{1}{\varepsilon_k^{C}\left(y_k\right)}\right|^2 dy_k<+\infty,$$

$$\sum_{k=1}^{K}\int_{\underline{y}}^{\overline{y}}\left|C_k\left(y_k\right)\right|^2 dy_k<+\infty,$$

$$\sum_{k=1}^{K}\int_{\underline{y}}^{\overline{y}}\left|1-F_k\left(y_k\right)\right|^2+\left|f_k\left(y_k\right)y_k\right|^2 dy_k<+\infty.$$

We can make the following resolvent when Assumption 2 is satisfied. When Assumption 2 is satisfied, define resolvent $\left\{\overline{h}_{kj}\left(y_k,y_j'\right)\right\}$ and $\left\{h_{kj}\left(y_k,y_j'\right)\right\}$:

$$\overline{h}_{kj}\left(y_k,y_j'\right)\equiv e_{kj}\left(y_k,y_j'\right)+\sum_{m=1}^{K}\int_{y_m^j\left(y_j'\right)}^{y_m^k\left(y_k\right)}e_{km}\left(y_k,y_m''\right)\overline{h}_{mj}\left(y_m'',y_j'\right)y_m'', \tag{D64}$$

and

$$\begin{aligned}h_{kj}\left(y_k,y_j'\right) \equiv\ & e_{kk}^{y,own}\left(y_k,y_j'\right)1_{k=j}+e_{kj}^{y,cross}\left(y_k,y_j'\right)+\int_{y_k''\left(y_j'\right)}^{\overline{y}}e_{kk}^{y,own}\left(y_k,y_k''\right)\overline{h}_{kj}\left(y_k'',y_j'\right)y_k'' \\ &+\sum_{m=1}^{K}\int_{y_m''\left(y_j'\right)}^{\overline{y}}e_{km}^{y,cross}\left(y_k,y_m''\right)\overline{h}_{mj}\left(y_m'',y_j'\right)y_m''+\sum_{m=1}^{K}\int_{\underline{y}}^{y_m^k\left(y_k\right)}e_{km}\left(y_k,y_m''\right)h_{mj}\left(y_m'',y_j'\right)dy_m'' \\ &+\int_{\underline{y}}^{\overline{y}}e_{kk}^{y,own}\left(y_k,y_k''\right)h_{kj}\left(y_k'',y_j'\right)y_k''+\sum_{m=1}^{K}\int_{\underline{y}}^{\overline{y}}e_{km}^{y,cross}\left(y_k,y_m''\right)h_{mj}\left(y_m'',y_j'\right)y_m''.\end{aligned} \tag{D65}$$

Using the above resolvent operator, we can solve for the $\frac{\widehat{y}_k}{y_k}$ in terms of the following elasticities (see the following Lemma 7)

$$\varepsilon_k^{C}\left(y_k\right)\equiv\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k)y_k}{1-T'(y_k)}}, \tag{D66}$$

and

$$\begin{aligned}&\varepsilon_{kj}^{Cross}\left(y_k,y_j'\right) \\ \equiv\ & \overline{h}_{kj}\left(y_k,y_j'\right)\varepsilon_j^{C}\left(y_j'\right)1_{y_j'\leq y_j^k\left(y_k\right)}+h_{kj}\left(y_k,y_j'\right)\varepsilon_j^{C}\left(y_j'\right)+e_{kk}^{C,own}\left(y_k,y_k'\right)1_{j=k}+e_{kj}^{C,cross}\left(y_k,y_j'\right) \\ &+\int_{\underline{y}}^{y_j^k\left(y_k\right)}\overline{h}_{kj}\left(y_k,y_j''\right)e_{jj}^{C,own}\left(y_j'',y_j'\right)dy_j''+\int_{\underline{y}}^{\overline{y}}h_{kj}\left(y_k,y_j''\right)e_{jj}^{C,own}\left(y_j'',y_j'\right)dy_j'' \\ &+\sum_{m}\int_{\underline{y}}^{y_m^k\left(y_k\right)}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{C,own}\left(y_m'',y_j'\right)dy_m''+\sum_{m}\int_{\underline{y}}^{\overline{y}}h_{km}\left(y_k,y_m''\right)e_{mj}^{C,own}\left(y_m'',y_j'\right)dy_m'',\end{aligned} \tag{D67}$$

and

$$\begin{aligned}
&\varepsilon_{kj}^{E}\left(y_k\right) \\
\equiv\ & e_{kk}^{E,own}\left(y_k\right)1_{j=k}+e_{kj}^{E,cross}\left(y_k\right)+\int_{\underline{y}}^{y_j^k(y_k)}\overline{h}_{kj}\left(y_k,y_j'\right)e_{jj}^{E,own}\left(y_j'\right)dy_j'+\int_{\underline{y}}^{\overline{y}}h_{kj}\left(y_k,y_j'\right)e_{jj}^{E,own}\left(y_j'\right)dy_j' \\
&+\sum_m\int_{\underline{y}}^{y_m^k(y_k)}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{E,own}\left(y_m'',y_j'\right)dy_m''+\sum_m\int_{\underline{y}}^{\overline{y}}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{E,own}\left(y_m'',y_j'\right)dy_m'' \\
&-\frac{N_j}{\sum_m N_m}\left\{\begin{array}{c}e_{kk}^{E,own}\left(y_k\right)+\sum_j e_{kj}^{E,cross}\left(y_k\right)+\sum_j\int_{\underline{y}}^{y_j^k(y_k)}\overline{h}_{kj}\left(y_k,y_j'\right)e_{jj}^{E,own}\left(y_j'\right)dy_j' \\ +\sum_j\int_{\underline{y}}^{\overline{y}}h_{kj}\left(y_k,y_j'\right)e_{jj}^{E,own}\left(y_j'\right)dy_j' \\ +\sum_j\sum_m\int_{\underline{y}}^{y_m^k(y_k)}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{E,own}\left(y_m'',y_j'\right)dy_m'' \\ +\sum_j\sum_m\int_{\underline{y}}^{\overline{y}}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{E,own}\left(y_m'',y_j'\right)dy_m''\end{array}\right\}.
\end{aligned} \tag{D68}$$

For simplification, we also set

$$\varepsilon_k^{Cross}\left(y_k,y'\right)\equiv\sum_{j\in\mathcal{J}}\varepsilon_{kj}^{Cross}\left(y_k,y'\right). \tag{D69}$$

We now derive the change in labor incomes:

Lemma 7 *We have the following result regarding the change of labor income*

$$\begin{aligned}
\frac{\widehat{y}_k}{y_k} =\ & \sum_j\int_{\underline{y}}^{y_j^k(y_k)}e_{kj}\left(y_k,y_j'\right)\frac{\widehat{y}_j'}{y_j'}dy_j'+\int_{\underline{y}}^{\overline{y}}e_{kk}^{y,own}\left(y_k,y_k'\right)\frac{\widehat{y}_k'}{y_k'}dy_k' \\
&+\sum_j\int_{\underline{y}}^{\overline{y}}e_{kj}^{y,cross}\left(y_k,y_j\right)\frac{\widehat{y}_j'}{y_j'}dy_j'-\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k)y_k}{1-T'(y_k)}}\frac{\widehat{T}'\left(y_k\right)}{1-T'\left(y_k\right)} \\
&-\int_{\underline{y}}^{\overline{y}}e_{kk}^{C,own}\left(y_k,y_k'\right)\frac{\widehat{T}'\left(y_k'\right)}{1-T'\left(y_k'\right)}dy_k'-\sum_j\int_{\underline{y}}^{\overline{y}}e_{kj}^{C,cross}\left(y_k,y_j\right)\frac{\widehat{T}'\left(y_j\right)}{1-T'\left(y_j\right)}dy_j \\
&-e_{kk}^{E,own}\left(y_k\right)\left[\widehat{E}_k-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right]-\sum_j e_{kj}^{E,cross}\left(y_k\right)\left[\widehat{E}_j-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right].
\end{aligned} \tag{D70}$$

Proof. See the following proof. ■

Proof of Lemma 7. First, use $\widehat{T}_k^A\left(y_k\left(x_k\right)\right) = \widehat{T}_k^A\left(\underline{y}_k\right) + \int_{y_k(\underline{\theta})}^{y_k(x_k)} \widehat{T}_k^{A\prime}\left(y_k\left(x_k'\right)\right) dy_k\left(x_k'\right)$ and $\widehat{T}_k\left(y_k\left(x_k\right)\right) = \widehat{T}_k\left(\underline{y}_k\right) + \int_{y_k(\underline{\theta})}^{y_k(x_k)} \widehat{T}_k'\left(y_k\left(x_k'\right)\right) dy_k\left(x_k'\right)$ to rewrite (D21) as:

$$\begin{aligned}
\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)} &= -\frac{1}{\frac{1}{\varepsilon_k} + \frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}} \frac{\widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)} \\
&\quad - \sum_n e_{kn}\left(y_k\left(\theta_k\right)\right)\left[\widehat{T}_n^A\left(\underline{y}_n\right) + \int_{y_k(\underline{\theta})}^{y_n(\theta_k)} \widehat{T}_n^{A\prime}\left(y_n\left(x_k'\right)\right) dy_n\left(x_k'\right)\right] \\
&\quad + \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k\left(\theta_k\right), y_k\left(x_k\right)\right) \left[\begin{array}{l} \widehat{T}_k^A\left(\underline{y}_k\right) + \int_{y_k(\underline{\theta})}^{y_k(x_k)} \widehat{T}_k^{A\prime}\left(y_k\left(x_k'\right)\right) dy_k\left(x_k'\right) - \\ \widehat{T}_k\left(\underline{y}_k\right) - \int_{y_k(\underline{\theta})}^{y_k(x_k)} \widehat{T}_k'\left(y_k\left(x_k'\right)\right) dy_k\left(x_k'\right) - \widehat{E}_k \end{array}\right] dy_k\left(x_k\right) \\
&\quad + \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k\left(\theta_k\right), y_k\left(x_k\right)\right) T_k^{\star\prime}\left(y_k\left(x_k\right)\right) y_k\left(x_k\right) \frac{\widehat{y}_k\left(x_k\right)}{y_k\left(x_k\right)} dy_k\left(x_k\right) \\
&\quad - \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}\left(y_k\left(\theta_k\right), y_n\left(x_k\right)\right)\left[\widehat{T}_n^A\left(\underline{y}_n\right) + \int_{y_k(\underline{\theta})}^{y_n(\theta_k)} \widehat{T}_n^{A\prime}\left(y_n\left(x_k'\right)\right) dy_n\left(x_k'\right)\right] dy_n\left(x_k\right).
\end{aligned}$$

Use the Fubini Theorem to change the order of integral:

$$\begin{aligned}
&\frac{\widehat{y}_k\left(\theta_k\right)}{y_k\left(\theta_k\right)} = -\frac{1}{\frac{1}{\varepsilon_k} + \frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}} \frac{\widehat{T}'\left(y_k\left(\theta_k\right)\right)}{1-T'\left(y_k\left(\theta_k\right)\right)} \\
&- \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} \left[\int_{y_k(x_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k\left(\theta_k\right), y_k\left(s_k\right)\right)\left[1 - T'\left(y_k\left(x_k\right)\right)\right] dy_k\left(s_k\right)\right] \frac{\widehat{T}'\left(y_k\left(x_k\right)\right)}{1-T'\left(y_k\left(x_k\right)\right)} dy_k\left(x_k\right) \\
&+ \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k\left(\theta_k\right), y_k\left(x_k\right)\right) T_k^{\star\prime}\left(y_k\left(x_k\right)\right) y_k\left(x_k\right) \frac{\widehat{y}_k\left(x_k\right)}{y_k\left(x_k\right)} dy_k\left(x_k\right) \\
&+ \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} \left[\int_{y_k(x_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k\left(\theta_k\right), y_k\left(s_k\right)\right) dy_k\left(s_k\right)\right] \widehat{T}_k^{A\prime}\left(y_k\left(x_k\right)\right) dy_k\left(x_k\right) \\
&- \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\theta_k)} e_{kn}\left(y_k\left(\theta_k\right)\right) \widehat{T}_n^{A\prime}\left(y_n\left(x_k\right)\right) dy_n\left(x_k\right) \\
&- \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} \left[\int_{y_n(x_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}\left(y_k\left(\theta_k\right), y_n\left(s_k\right)\right) dy_n\left(s_k\right)\right] \widehat{T}_n^{A\prime}\left(y_n\left(x_k\right)\right) dy_n\left(x_k\right) \\
&- \sum_n \left[e_{kn}\left(y_k\left(\theta_k\right)\right) + \int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}\left(y_k\left(\theta_k\right), y_n\left(x_k\right)\right) dy_n\left(x_k\right)\right] \widehat{T}_n^A\left(y_n\left(\underline{\theta}_k\right)\right) \\
&+ \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}\left(y_k\left(\theta_k\right), y_k\left(x_k\right)\right) dy_k\left(x_k\right)\left[\widehat{T}_k^A\left(y_k\left(\underline{\theta}_k\right)\right) - \widehat{T}\left(y_k\left(\underline{\theta}_k\right)\right) - \widehat{E}_k\right].
\end{aligned}$$

Notice that $\widehat{T}_n^{A\prime}\left(y_n\left(x_k\right)\right) = -\frac{1-T_n^{A\prime}(y_n(x_k))}{\varepsilon_n^A(y_n(x_k))} \frac{\widehat{y}_n(x_k)}{y_n(x_k)}$, we can substitute $\widehat{T}_n^{A\prime}\left(y_n\left(x_k\right)\right)$ by $\frac{\widehat{y}_n(x_k)}{y_n(x_k)}$. Then, the above

equation is transformed into

$$\begin{aligned}
\frac{\widehat{y}_k(\theta_k)}{y_k(\theta_k)} &= -\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}}\frac{\widehat{T}'(y_k(\theta_k))}{1-T'(y_k(\theta_k))} \\
&- \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} \left[\int_{y_k(x_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}(y_k(\theta_k), y_k(s_k))\left[1-T'(y_k(x_k))\right] dy_k(s_k)\right] \frac{\widehat{T}'(y_k(x_k))}{1-T'(y_k(x_k))} dy_k(x_k) \\
&- \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} \left[\begin{array}{c} \frac{1-T_k^{A\prime}(y_k(x_k))}{\varepsilon_k^A(y_k(x_k))}\int_{y_k(x_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}(y_k(\theta_k), y_k(s_k))\, dy_k(s_k) \\ -e_{kk}^{own}(y_k(\theta_k), y_k(x_k))\, T_k^{\star\prime}(y_k(x_k))\, y_k(x_k) \end{array}\right] \frac{\widehat{y}_k(x_k)}{y_k(x_k)} dy_k(x_k) \\
&+ \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\theta_k)} e_{kn}(y_k(\theta_k)) \frac{1-T_n^{A\prime}(y_n(x_k))}{\varepsilon_n^A(y_n(x_k))}\frac{\widehat{y}_n(x_k)}{y_n(x_k)} dy_n(x_k) \\
&+ \sum_n \int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} \left[\frac{1-T_n^{A\prime}(y_n(x_k))}{\varepsilon_n^A(y_n(x_k))}\int_{y_n(x_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}(y_k(\theta_k), y_n(s_k))\, dy_n(s_k)\right] \frac{\widehat{y}_n(x_k)}{y_n(x_k)} dy_n(x_k) \\
&- \sum_n \left[e_{kn}(y_k(\theta_k)) + \int_{y_n(\underline{\theta}_k)}^{y_n(\overline{\theta}_k)} e_{kn}^{cross}(y_k(\theta_k), y_n(x_k))\, dy_n(x_k)\right] \widehat{T}_n^A(y_n(\underline{\theta}_k)) \\
&- \int_{y_k(\underline{\theta}_k)}^{y_k(\overline{\theta}_k)} e_{kk}^{own}(y_k(\theta_k), y_k(x_k))\, dy_k(x_k)\left[\widehat{T}(y_k(\underline{\theta}_k)) + \widehat{E}_k - \widehat{T}_k^A(y_k(\underline{\theta}_k))\right].
\end{aligned} \tag{D71}$$

Second, combination of (D62) and (D53) gives:

$$\begin{aligned}
&\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k), y_k')\, dy_k' \left[\widehat{T}(y_k(\underline{\theta})) + \widehat{E}_k - \widehat{T}_k^A(y_k(\underline{\theta}))\right] + \sum_n \left[e_{kn}(y_k(\theta_k)) + \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k), y_n')\, dy_n'\right] \widehat{T}_n^A(y_n(\underline{\theta})) \qquad \text{(D72)} \\
= \;& \int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k), y_k')\, dy_k' \left\{\widehat{E}_k - \sum_m \frac{N_m}{\sum_m N_m}\widehat{E}_m\right\} \\
&+ \sum_j \left\{\begin{array}{c} \sum_n \left[e_{kn}(y_k(\theta_k)) + \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k), y_n')\, dy_n'\right] \eta_{nj}^G \\ -\eta_{kj}^G \int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k), y_k')\, dy_k' \end{array}\right\} \left[\widehat{E}_j - \sum_m \frac{N_m}{\sum_m N_m}\widehat{E}_m\right] \\
&- \sum_j \int_{\underline{y}}^{\overline{y}} \left\{\begin{array}{c} \left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k), y_k')\, dy_k'\right]\left[\eta_{kj}^G + \frac{N_j}{\sum_m N_m}(1-\sum_m \eta_{km}^G)\right] \\ -\sum_n \left[e_{kn}(y_k(\theta_k)) + \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k), y_n')\, dy_n'\right]\left[\eta_{nj}^G - \sum_m \eta_{nm}^G \frac{N_j}{\sum_m N_m}\right] \end{array}\right\} \\
&\times \left[1-F_j\left(y_j\left(x_j\right)\right)\right]\left[1-T'\left(y_j\left(x_j\right)\right)\right] \frac{\widehat{T}'\left(y_j\left(x_j\right)\right)}{1-T'\left(y_j\left(x_j\right)\right)} dy_j\left(x_j\right) - \\
&\sum_j \int_{\underline{y}}^{\overline{y}} \left\{\begin{array}{c} \left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k), y_k')\, dy_k'\right] \left[\begin{array}{c} \eta_{kj}^G\left[1-F_j\left(y_j\left(x_j\right)\right)\right] - \\ \sum_m \eta_{km}^G\left[1-F_m\left(y_m\left(x_j\right)\right)\right]\bar{r}_{mj}^G\left(y_m\left(x_j\right)\right) \end{array}\right] \frac{1-T_j^{A\prime}(y_j(x_j))}{\varepsilon_j^A(y_j(x_j))} - \\ \sum_n \left[\begin{array}{c} e_{kn}(y_k(\theta_k)) + \\ \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k), y_n')\, dy_n' \end{array}\right] \left[\begin{array}{c} \eta_{nj}^G\left[1-F_j\left(y_j\left(x_j\right)\right)\right] - \\ \sum_m \eta_{nm}^G\left[1-F_m\left(y_m\left(x_j\right)\right)\right]\bar{r}_{mj}^G\left(y_m\left(x_j\right)\right) \end{array}\right] \frac{1-T_j^{A\prime}(y_j(x_j))}{\varepsilon_j^A(y_j(x_j))} \\ -\left\{\begin{array}{c} \left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k), y_k')\, dy_k'\right] \left[\begin{array}{c} \eta_{kj}^G C_j\left(y_j\left(x_j\right)\right) - \\ \frac{(1+\sum_m \eta_{km}^G)N_j}{\sum_n N_j}\frac{T'(y_j(x_j))}{1-T'(y_j(x_j))} \end{array}\right] \\ -\sum_n \left[\begin{array}{c} e_{kn}(y_k(\theta_k)) + \\ \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k), y_n')\, dy_n' \end{array}\right] \left[\begin{array}{c} \eta_{nj}^G C_j\left(y_j\left(x_j\right)\right) - \\ \sum_m \frac{\eta_{nm}^G N_j}{\sum_n N_j}\frac{T'(y_j(x_j))}{1-T'(y_j(x_j))} \end{array}\right] \end{array}\right\} \times \\ f_j\left(y_j(x_j)\right) y_j(x_j)\left[1-T'\left(y_j\left(x_j\right)\right)\right] \end{array}\right\} \frac{\widehat{y}_j(x_j)}{y_j(x_j)} dy_j\left(x_j\right).
\end{aligned}$$

Use (D72) to substitute the left side of equation (D72) in (D71):

$$
\begin{aligned}
&\frac{\widehat{y}_k(\theta_k)}{y_k(\theta_k)} = -\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k(\theta_k))y_k(\theta_k)}{1-T'(y_k(\theta_k))}}\frac{\widehat{T}'(y_k(\theta_k))}{1-T'(y_k(\theta_k))} \\
&-\int_{\underline{y}}^{\overline{y}}\left[\int_{y_k(x_k)}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k(s_k))\left[1-T'(y_k(x_k))\right]dy_k(s_k)\right]\frac{\widehat{T}'(y_k(x_k))}{1-T'(y_k(x_k))}dy_k(x_k) - \\
&\sum_j\int_{\underline{y}}^{\overline{y}}\left\{\begin{array}{c}\sum_n\left[\begin{array}{c}e_{kn}(y_k(\theta_k))+\\ \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k),y_n')\,dy_n'\end{array}\right]\left[\eta_{nj}^G-\frac{N_j}{\sum_m N_m}\sum_m\eta_{nm}^G\right]\left[1-F_j\left(y_j\left(x_j\right)\right)\right]\left[1-T'\left(y_j\left(x_j\right)\right)\right]\\ -\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]\left[1-F_j\left(y_j\left(x_j\right)\right)\right]\left[1-T'\left(y_j\left(x_j\right)\right)\right]\\ +\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]\left[\eta_{kj}^G-\frac{N_j}{\sum_m N_m}\sum_m\eta_{km}^G\right]\left[1-F_j\left(y_j\left(x_j\right)\right)\right]\left[1-T'\left(y_j\left(x_j\right)\right)\right]\end{array}\right\}\frac{\widehat{T}'\left(y_j\left(x_j\right)\right)dy_j\left(x_j\right)}{1-T'\left(y_j\left(x_j\right)\right)} \\
&-\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]\left\{\widehat{E}_k-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right\} \\
&-\sum_j\left\{\sum_n\left[e_{kn}(y_k(\theta_k))+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k),y_n')\,dy_n'\right]\eta_{nj}^G-\eta_{kj}^G\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right\}\times\left\{\widehat{E}_j-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right\} \\
&+\sum_j\int_{\underline{y}}^{y_j(\theta_k)} e_{kj}(y_k(\theta_k))\frac{1-T_j^{A\prime}\left(y_j\left(x_j\right)\right)}{\varepsilon_j^A\left(y_j\left(x_j\right)\right)}\frac{\widehat{y}_j(x_j)}{y_j(x_j)}dy_j\left(x_j\right) \\
&+\int_{\underline{y}}^{\overline{y}}\left[e_{kk}^{own}(y_k(\theta_k),y_k(x_k))\,T_k^{\star\prime}(y_k(x_k))\,y_k(x_k)-\frac{1-T_k^{A\prime}(y_k(x_k))}{\varepsilon_k^A(y_k(x_k))}\int_{y_k(x_k)}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k(s_k))\,dy_k(s_k)\right]\frac{\widehat{y}_k(x_k)}{y_k(x_k)}dy_k(x_k) \\
&+\sum_j\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\int_{y_j(x_j)}^{\overline{y}} e_{kj}^{cross}\left(y_k(\theta_k),y_j\left(s_j\right)\right)dy_j\left(s_j\right)\frac{1-T_j^{A\prime}(y_j(x_j))}{\varepsilon_j^A(y_j(x_j))}+\\ \left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]\left\{\begin{array}{c}\eta_{kj}^G\left[1-F_j\left(y_j\left(x_j\right)\right)\right]-\\ \sum_m\eta_{km}^G\left[1-F_m\left(y_m\left(x_j\right)\right)\right]\bar{r}_{mj}^G\left(y_m\left(x_j\right)\right)\end{array}\right\}\frac{1-T_j^{A\prime}(y_j(x_j))}{\varepsilon_j^A(y_j(x_j))}-\\ \sum_n\left[\begin{array}{c}e_{kn}(y_k(\theta_k))+\\ \int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k),y_n')\,dy_n'\end{array}\right]\left\{\begin{array}{c}\eta_{nj}^G\left[1-F_j\left(y_j\left(x_j\right)\right)\right]-\\ \sum_m\eta_{nm}^G\left[1-F_m\left(y_m\left(x_j\right)\right)\right]\bar{r}_{mj}^G\left(y_m\left(x_j\right)\right)\end{array}\right\}\frac{1-T_j^{A\prime}(y_j(x_j))}{\varepsilon_j^A(y_j(x_j))}+\\ \frac{N_j}{\sum_n N_j}\left\{\begin{array}{c}\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]+\\ \sum_m\eta_{km}^G\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]-\\ \sum_m\sum_n\left[e_{kn}(y_k(\theta_k))+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k),y_n')\,dy_n'\right]\eta_{nm}^G\end{array}\right\}f_j\left(y_j(x_j)\right)y_j(x_j)T'\left(y_j(x_j)\right)-\\ \left\{\begin{array}{c}\left[\int_{\underline{y}}^{\overline{y}} e_{kk}^{own}(y_k(\theta_k),y_k')\,dy_k'\right]\eta_{kj}^G-\\ \sum_n\left[e_{kn}(y_k(\theta_k))+\int_{\underline{y}}^{\overline{y}} e_{kn}^{cross}(y_k(\theta_k),y_n')\,dy_n'\right]\eta_{nj}^G\end{array}\right\}C_j\left(y_j\left(x_j\right)\right)f_j\left(y_j(x_j)\right)y_j(x_j)\left[1-T'\left(y_j\left(x_j\right)\right)\right]\end{array}\right]\frac{\widehat{y}_j(x_j)}{y_j(x_j)}dy_j\left(x_j\right).
\end{aligned}
$$

By definitions of (D63), we can write the above equation as:

$$
\begin{aligned}
\frac{\widehat{y}_k}{y_k} &= -\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k)y_k}{1-T'(y_k)}}\frac{\widehat{T}'(y_k)}{1-T'(y_k)} \\
&\quad -\int_{\underline{y}}^{\overline{y}} e_{kk}^{C,own}\left(y_k,y_k'\right)\frac{\widehat{T}'\left(y_k'\right)}{1-T'\left(y_k'\right)}dy_k'-\sum_j\int_{\underline{y}}^{\overline{y}} e_{kj}^{C,cross}\left(y_k,y_j\right)\frac{\widehat{T}'\left(y_j\right)}{1-T'\left(y_j\right)}dy_j \\
&\quad -e_{kk}^{E,own}(y_k)\left\{\widehat{E}_k-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right\}-\sum_j e_{kj}^{E,cross}(y_k)\left\{\widehat{E}_j-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right\} \\
&\quad +\sum_j\int_{\underline{y}}^{y_j^k(y_k)} e_{kj}\left(y_k,y_j'\right)\frac{\widehat{y}_j'}{y_j'}dy_j'+\int_{\underline{y}}^{\overline{y}} e_{kk}^{y,own}\left(y_k,y_k'\right)\frac{\widehat{y}_k'}{y_k'}dy_k'+\sum_j\int_{\underline{y}}^{\overline{y}} e_{kj}^{y,cross}\left(y_k,y_j\right)\frac{\widehat{y}_j'}{y_j'}dy_j'.
\end{aligned}
$$

■

We now derive Proposition 5 by showing that when Assumption 2 holds, there is a unique solution to

equation (D70). Moreover, there exist own and cross elasticities $\varepsilon_j^C : \mathcal{Y} \mapsto \mathbb{R}$ and $\varepsilon_j^{Cross} : \mathcal{Y}^2 \mapsto \mathbb{R}$ and $\varepsilon_{jk}^E : \mathcal{Y} \mapsto \mathbb{R}$ such that

$$\frac{\widehat{y}_j(\theta)}{y_j(\theta)} = -\varepsilon_j^C\left(y_j(\theta)\right)\frac{\widehat{T}'\left(y_j(\theta)\right)}{1-T'\left(y_j(\theta)\right)} - \int_{\underline{y}}^{\overline{y}} \varepsilon_j^{Cross}\left(y_j(\theta), y(x)\right)\frac{\widehat{T}'\left(y(x)\right)}{1-T'\left(y(x)\right)}dy(x) - \sum_{k\in\mathcal{J}} \varepsilon_{jk}^E\left(y_j\right)\widehat{E}_k, \quad \text{(D73)}$$

where $\varepsilon_j^C\left(y_j(\theta)\right) = \frac{1}{\frac{1}{\varepsilon_j} + \frac{T''\left(y_j(\theta)\right)y_j(\theta)}{1-T'\left(y_j(\theta)\right)}}$.

Proof. **Step 1**. Set $\overrightarrow{x} \equiv \left[\begin{array}{ccc} x_1 & \cdots & x_K \end{array}\right]^T$ as a K dimension vector function, such that for $k = 1, \ldots, K$,

$$\begin{aligned} x_k(\theta) \equiv & -\frac{1}{\frac{1}{\varepsilon_k} + \frac{T''(y_k(\theta))y_k(\theta)}{1-T'(y_k(\theta))}}\frac{\widehat{T}'\left(y_k(\theta)\right)}{1-T'\left(y_k(\theta)\right)} \\ & -\int_{\underline{\theta}}^{\overline{\theta}} e_{kk}^{C,own}\left(y_k(\theta), y_k(\theta')\right)\frac{\widehat{T}'\left(y_k(\theta')\right)}{1-T'\left(y_k(\theta')\right)}\varepsilon_k^A\left(y_k(\theta')\right)\left(\frac{1}{\varepsilon_k}+1\right)d\theta' \\ & -\sum_j \int_{\underline{\theta}}^{\overline{\theta}} e_{kj}^{C,cross}\left(y_k(\theta), y_j(\theta')\right)\varepsilon_j^A\left(y_j(\theta')\right)\left(\frac{1}{\varepsilon_j}+1\right)\frac{\widehat{T}'\left(y_j(\theta')\right)}{1-T'\left(y_j(\theta')\right)}d\theta' \\ & -e_{kk}^{E,own}\left(y_k(\theta)\right)\left\{\widehat{E}_k - \sum_m \frac{N_m}{\sum_m N_m}\widehat{E}_m\right\} - \sum_j e_{kj}^{E,cross}\left(y_k(\theta)\right)\left\{\widehat{E}_j - \sum_m \frac{N_m}{\sum_m N_m}\widehat{E}_m\right\}. \end{aligned} \quad \text{(D74)}$$

Set a $K \times K$ dimension vector function $\overrightarrow{\overline{\gamma}} \equiv \left[\begin{array}{ccc} \overline{\gamma}_{11} & \cdots & \overline{\gamma}_{K\times K} \end{array}\right]^T$, $\overrightarrow{\gamma} \equiv \left[\begin{array}{ccc} \gamma_{11} & \cdots & \gamma_{K\times K} \end{array}\right]^T$ such that for any $k, j = 1, \ldots, K$:

$$\begin{aligned} \overline{\gamma}_{kj}(\theta, \theta') &\equiv e_{kj}\left(y_k(\theta), y_j(\theta')\right)\varepsilon_j^A\left(y_j(\theta')\right)\left(\frac{1}{\varepsilon_j}+1\right); \\ \gamma_{kj}(\theta, \theta') &\equiv \left[e_{kk}^{y,own}\left(y_k(\theta), y_j(\theta')\right)1_{k=j} + e_{kj}^{y,cross}\left(y_k(\theta), y_j(\theta')\right)\right]\varepsilon_j^A\left(y_j(\theta')\right)\left(\frac{1}{\varepsilon_j}+1\right). \end{aligned}$$

Define a linear operator $T\left(\overrightarrow{f}\right) = \left[\begin{array}{ccc} T_1\left(\overrightarrow{f}\right) & \cdots & T_K\left(\overrightarrow{f}\right) \end{array}\right]^T$ where for any $k = 1, ..., K$:

$$T_k\left(\overrightarrow{f}\right)(\theta) \equiv \sum_{j=1}^K \int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta, \theta') f_j(\theta')\, d\theta' + \sum_{j=1}^K \int_{\Theta} \gamma_{kj}(\theta, \theta') f_j(\theta')\, d\theta';$$

Notice that equation (D70) is equivalent to the following equation (D75):

$$T\left(\overrightarrow{f}\right) + \overrightarrow{x} = \overrightarrow{f}. \quad \text{(D75)}$$

If there exists a unique solution to the problem (D70), then there exist $\left[\begin{array}{ccc} \frac{\widehat{y}_k(\theta)}{y_k(\theta)} & \cdots & \frac{\widehat{y}_k(\theta)}{y_k(\theta)} \end{array}\right]^T$ as a solution to the above equation, and it is unique.

Step 2. To use Lemma 10 to Lemma 12 to solve the above problem, we should show that $\overrightarrow{x} \in L_2^K(\Theta)$ and $\overrightarrow{\overline{\gamma}}, \overrightarrow{\gamma} \in L_2^{K\times K}(\Theta \times \Theta)$. To see this, notice that under Condition 2 the following functions belong to

$L_2(\Theta)$:

$$e_{kj}^{E,cross}\left(y_k(\theta)\right), e_{kk}^{E,own}\left(y_k(\theta)\right), \frac{1}{1-T'\left(y_j(\theta)\right)}, \varepsilon_j^A\left(y_j(\theta)\right), \frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k(\theta))y_k(\theta)}{1-T'(y_k(\theta))}}.$$

Second, the following functions belong to $L_2(\Theta\times\Theta)$:

$$e_{kj}^{C,cross}\left(y_k(\theta), y_j(\theta')\right), e_{kk}^{C,own}\left(y_k(\theta), y_k(\theta')\right).$$

Using Holder inequality and triangle inequality for absolute value, one can see $x_k(\theta)$ consists of functions belonging to $L_2(\Theta)$. Then, use Minkowski inequality, one has $x_k(\theta)$ belong to $L_2(\Theta)$. Therefore $\overrightarrow{x}\in L_2^K(\Theta)$. Similarly, one can see that $\gamma_{kj}(\theta,\theta')$ and $\overline{\gamma}_{kj}(\theta,\theta')$ are consist of functions belong to $L_2(\Theta\times\Theta)$. Therefore, $\overrightarrow{\overline{\gamma}}, \overrightarrow{\gamma}\in L_2^{K\times K}(\Theta\times\Theta)$.

Step 3. To use Lemma 10 to Lemma 12 to solve the problem (D75), one also need to show that $\left\|\overrightarrow{\overline{\gamma}}\right\|_2^{K\times K}+\left\|\overrightarrow{\gamma}\right\|_2^{K\times K}<1$. To do this, first notice that

$$\begin{aligned}
&\left\|\overrightarrow{\overline{\gamma}}\right\|_2^{K\times K}+\left\|\overrightarrow{\gamma}\right\|_2^{K\times K} \qquad\qquad (D76)\\
=&\left\{\sum_{j=1}^K\sum_{k=1}^K\int_\Theta\int_\Theta\left[\overline{\gamma}_{kj}(\theta,\theta')\right]^2 d\theta' d\theta\right\}^{\frac{1}{2}}+\left\{\sum_{j=1}^K\sum_{k=1}^K\int_\Theta\int_\Theta\left[\gamma_{kj}(\theta,\theta')\right]^2 d\theta' d\theta\right\}^{\frac{1}{2}}\\
=&\left\{\sum_{j=1}^K\sum_{k=1}^K\int_\Theta\int_\Theta e_{kj}\left(y_k(\theta),y_j(\theta')\right)^2\left[\frac{dy_j(\theta')}{d\theta'}\right]^2 d\theta' d\theta\right\}^{\frac{1}{2}}\\
&+\left\{\begin{array}{c}\sum_{j=1}^K\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kj}^{y,cross}\left(y_k(\theta),y_j(\theta')\right)\right]^2\left[\frac{dy_j(\theta')}{d\theta'}\right]^2 d\theta' d\theta+\\ \sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kk}^{y,own}\left(y_k(\theta),y_k(\theta')\right)\right]^2\left[\frac{dy_k(\theta')}{d\theta'}\right]^2 d\theta' d\theta+\\ +2\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kk}^{y,own}\left(y_k(\theta),y_k(\theta')\right)e_{kk}^{y,cross}\left(y_k(\theta),y_k(\theta')\right)\right]\left[\frac{dy_k(\theta')}{d\theta'}\right]^2 d\theta' d\theta\end{array}\right\}^{\frac{1}{2}}.
\end{aligned}$$

Using Holder's inequality, we have:

$$\begin{aligned}
&2\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kk}^{y,own}\left(y_k(\theta),y_k(\theta')\right)e_{kk}^{y,cross}\left(y_k(\theta),y_k(\theta')\right)\right]\left[\frac{dy_k(\theta')}{d\theta'}\right]^2 d\theta' d\theta \qquad (D77)\\
\leq\ &2\sqrt{\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kk}^{y,own}\left(y_k(\theta),y_k(\theta')\right)\right]^2\left[\frac{dy_k(\theta')}{d\theta'}\right]^2 d\theta' d\theta}\\
&\times\sqrt{\sum_{k=1}^K\int_\Theta\int_\Theta\left[e_{kk}^{y,own}\left(y_k(\theta),y_k(\theta')\right)\right]^2\left[\frac{dy_k(\theta')}{d\theta'}\right]^2 d\theta' d\theta}
\end{aligned}$$

Therefore,

$$
\begin{aligned}
& \sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}^{y,cross}\left(y_k\left(\theta\right),y_j\left(\theta'\right)\right)\right]^2\left[\frac{dy_j\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta \\
& +\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta \\
& +2\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)e_{kk}^{y,cross}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta \\
\leq\ & \sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}^{y,cross}\left(y_k\left(\theta\right),y_j\left(\theta'\right)\right)\right]^2\left[\frac{dy_j\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta \\
& +\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta \\
& +2\sqrt{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta} \\
& \times\sqrt{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta}.
\end{aligned}
\tag{D78}
$$

Take the square root of both sides of equation (D78):

$$
\begin{aligned}
& \left\{\begin{array}{c}
\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}^{y,cross}\left(y_k\left(\theta\right),y_j\left(\theta'\right)\right)\right]^2\left[\frac{dy_j\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta+ \\
\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta+ \\
+2\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)e_{kk}^{y,cross}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta
\end{array}\right\}^{\frac{1}{2}} \\
\leq\ & \sqrt{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta} \\
& +\sqrt{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\left[\frac{dy_k\left(\theta'\right)}{d\theta'}\right]^2 d\theta' d\theta}.
\end{aligned}
$$

The above inequality together with equation (D76) implies:

$$
\begin{aligned}
\left\|\overrightarrow{\gamma}\right\|_2^{K\times K}+\left\|\overrightarrow{\gamma}\right\|_2^{K\times K} \leq\ & \left\{\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}e_{kj}\left(y_k\left(\theta\right),y_j\left(\theta'\right)\right)^2\frac{\frac{dy_j\left(\theta'\right)}{d\theta'}}{\frac{dy_k\left(\theta\right)}{d\theta}}dy_j\left(\theta'\right)dy_k\left(\theta\right)\right\}^{\frac{1}{2}} \\
& +\left\{\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}^{y,cross}\left(y_k\left(\theta\right),y_j\left(\theta'\right)\right)\right]^2\frac{\frac{dy_j\left(\theta'\right)}{d\theta'}}{\frac{dy_k\left(\theta\right)}{d\theta}}dy_j\left(\theta'\right)dy_k\left(\theta\right)\right\}^{\frac{1}{2}} \\
& +\left\{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right),y_k\left(\theta'\right)\right)\right]^2\frac{\frac{dy_k\left(\theta'\right)}{d\theta'}}{\frac{dy_k\left(\theta\right)}{d\theta}}dy_k\left(\theta'\right)dy_k\left(\theta\right)\right\}^{\frac{1}{2}}.
\end{aligned}
$$

Use $\frac{d\ln y_k(\theta_k)}{d\ln \theta_k} = \varepsilon_k^A\left(y_k\left(\theta_k\right)\right)\left(\frac{1}{\varepsilon_k}+1\right)$ to substitude $\frac{d\ln y_k(\theta_k)}{d\ln \theta_k}$ in the above equation:

$$\left\|\overrightarrow{\overrightarrow{\gamma}}\right\|_2^{K\times K} + \left\|\overrightarrow{\gamma}\right\|_2^{K\times K} \leq \left\{\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta} e_{kj}\left(y_k\left(\theta\right), y_j\left(\theta'\right)\right)^2 \left|\frac{\varepsilon_j^A\left(y_j\left(\theta_j'\right)\right)\left(\frac{1}{\varepsilon_j}+1\right)}{\varepsilon_k^A\left(y_k\left(\theta_k\right)\right)\left(\frac{1}{\varepsilon_k}+1\right)}\right| dy_j\left(\theta'\right) dy_k\left(\theta\right)\right\}^{\frac{1}{2}}$$
$$+\left\{\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}^{y,cross}\left(y_k\left(\theta\right), y_j\left(\theta'\right)\right)\right]^2 \left|\frac{\varepsilon_j^A\left(y_j\left(\theta_j'\right)\right)\left(\frac{1}{\varepsilon_j}+1\right)}{\varepsilon_k^A\left(y_k\left(\theta_k\right)\right)\left(\frac{1}{\varepsilon_k}+1\right)}\right| dy_j\left(\theta'\right) dy_k\left(\theta\right)\right\}^{\frac{1}{2}}$$
$$+\left\{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k\left(\theta\right), y_k\left(\theta'\right)\right)\right]^2 \left|\frac{\varepsilon_k^A\left(y_k\left(\theta_k'\right)\right)}{\varepsilon_k^A\left(y_k\left(\theta_k\right)\right)}\right| dy_k\left(\theta'\right) dy_k\left(\theta\right)\right\}^{\frac{1}{2}}$$

The rearrangement of the above equation gives:

$$\begin{aligned}\left\|\overrightarrow{\overrightarrow{\gamma}}\right\|_2^{K\times K} + \left\|\overrightarrow{\gamma}\right\|_2^{K\times K} \quad \leq \quad & \left\{\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}\left(y_k, y_j'\right)\right]^2 \left|\frac{\varepsilon_j^A\left(y_j'\right)\left(\frac{1}{\varepsilon_j}+1\right)}{\varepsilon_k^A\left(y_k\right)\left(\frac{1}{\varepsilon_k}+1\right)}\right| dy_j' dy_k\right\}^{\frac{1}{2}} \\ & + \left\{\sum_{j=1}^{K}\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kj}^{y,cross}\left(y_k, y_j'\right)\right]^2 \left|\frac{\varepsilon_j^A\left(y_j'\right)\left(\frac{1}{\varepsilon_j}+1\right)}{\varepsilon_k^A\left(y_k\right)\left(\frac{1}{\varepsilon_k}+1\right)}\right| dy_j' dy_k\right\}^{\frac{1}{2}} \\ & + \left\{\sum_{k=1}^{K}\int_{\Theta}\int_{\Theta}\left[e_{kk}^{y,own}\left(y_k, y_k'\right)\right]^2 \left|\frac{\varepsilon_k^A\left(y_k'\right)}{\varepsilon_k^A\left(y_k\right)}\right| dy_k' dy_k\right\}^{\frac{1}{2}}\end{aligned}$$

By the definition of c_e^y and Assumption 2, the above inequality gives:

$$\left\|\overrightarrow{\overrightarrow{\gamma}}\right\|_2^{K\times K} + \left\|\overrightarrow{\gamma}\right\|_2^{K\times K} \leq c_e^y < 1.$$

Step 4. A combination of Step 3 and Step 2 suggests that we can use Lemma 10 and Lemma 11 to solve problem (D75). In particular, it suggests that there is a unique solution to the problem (D75). Moreover, there exist resolvent $\left\{\overline{h}_{kj}\left(y_k, y_j'\right)\right\}$ and $\left\{h_{kj}\left(y_k, y_j'\right)\right\}$ to the problem:

$$\overline{h}_{kj}\left(y_k, y_j'\right) \equiv e_{kj}\left(y_k, y_j'\right) + \sum_{m=1}^{K}\int_{y_m^j\left(y_j'\right)}^{y_m^k\left(y_k\right)} e_{km}\left(y_k, y_m''\right)\overline{h}_{mj}\left(y_m'', y_j'\right) y_m'',$$

and

$$\begin{aligned}h_{kj}\left(y_k, y_j'\right) \quad \equiv \quad & e_{kk}^{y,own}\left(y_k, y_j'\right)1_{k=j} + e_{kj}^{y,cross}\left(y_k, y_j'\right) + \int_{y_k''\left(y_j'\right)}^{\overline{y}} e_{kk}^{y,own}\left(y_k, y_k''\right)\overline{h}_{kj}\left(y_k'', y_j'\right) y_k'' \\ & + \sum_{m=1}^{K}\int_{y_m''\left(y_j'\right)}^{\overline{y}} e_{km}^{y,cross}\left(y_k, y_m''\right)\overline{h}_{mj}\left(y_m'', y_j'\right) y_m'' + \sum_{m=1}^{K}\int_{\underline{y}}^{y_m^k\left(y_k\right)} e_{km}\left(y_k, y_m''\right) h_{mj}\left(y_m'', y_j'\right) dy_m'' \\ & + \int_{\underline{y}}^{\overline{y}} e_{kk}^{y,own}\left(y_k, y_k''\right) h_{kj}\left(y_k'', y_j'\right) y_k'' + \sum_{m=1}^{K}\int_{\underline{y}}^{\overline{y}} e_{km}^{y,cross}\left(y_k, y_m''\right) h_{mj}\left(y_m'', y_j'\right) y_m''.\end{aligned}$$

The above resolvent can be derived by taking the limit of sequences $\left\{\overline{h}_{kj}^{n}\left(y_k,y_j'\right)\right\}$ and $\left\{h_{kj}^{n}\left(y_k,y_j'\right)\right\}$. To see this, first we set the following sequence of $\overline{h}_{kj}^{n}$:

$$\begin{aligned}\overline{h}_{kj}^{0}\left(y_k,y_j'\right) &\equiv e_{kj}\left(y_k,y_j'\right),\\ \overline{h}_{kj}^{n+1}\left(y_k,y_j'\right) &\equiv e_{kj}\left(y_k,y_j'\right)+\sum_{m=1}^{K}\int_{y_m^j\left(y_j'\right)}^{y_m^k(y_k)} e_{km}\left(y_k,y_m''\right)\overline{h}_{mj}^{n}\left(y_m'',y_j'\right)y_m'',n\geq 0.\end{aligned}$$

We have $\overline{h}_{kj}\left(y_k,y_j'\right)=\lim_{n\to+\infty}\overline{h}_{kj}^{n}\left(y_k,y_j'\right)$. In addition, for any $n\geq 0$, we set the following sequence of $h_{kj}^{n}\left(y_k,y_j'\right)$:

$$\begin{aligned}h_{kj}^{0}\left(y_k,y_j'\right) &\equiv e_{kk}^{y,own}\left(y_k,y_j'\right)1_{k=j}+e_{kj}^{y,cross}\left(y_k,y_j'\right)+\int_{y_k''\left(y_j'\right)}^{\overline{y}} e_{kk}^{y,own}\left(y_k,y_k''\right)\overline{h}_{kj}\left(y_k'',y_j'\right)y_k''\\ &\quad+\sum_{m=1}^{K}\int_{y_m''\left(y_j'\right)}^{\overline{y}} e_{km}^{y,cross}\left(y_k,y_m''\right)\overline{h}_{mj}\left(y_m'',y_j'\right)y_m'',\\ h_{kj}^{n+1}\left(y_k,y_j'\right) &\equiv h_{kj}^{0}\left(y_k,y_j'\right)+\sum_{m=1}^{K}\int_{\underline{y}}^{y_m^k(y_k)} e_{km}\left(y_k,y_m''\right)h_{mj}^{n}\left(y_m'',y_j'\right)dy_m''\\ &\quad+\int_{\underline{y}}^{\overline{y}} e_{kk}^{y,own}\left(y_k,y_k''\right)h_{kj}^{n}\left(y_k'',y_j'\right)y_k''+\sum_{m=1}^{K}\int_{\underline{y}}^{\overline{y}} e_{km}^{y,cross}\left(y_k,y_m''\right)h_{mj}^{n}\left(y_m'',y_j'\right)y_m''.\end{aligned}$$

By Lemma 11, we have $h_{kj}\left(y_k,y_j'\right)=\lim_{n\to+\infty}h_{kj}^{n}\left(y_k,y_j'\right)$.

Step 5. Using Lemma 12 to solve problem (D75) with resolvent $\left\{\overline{h}_{kj}\left(y_k,y_j'\right)\right\}$ and $\left\{h_{kj}\left(y_k,y_j'\right)\right\}$ gives:

$$\frac{\widehat{y}_k(\theta)}{y_k(\theta)}=x_k(\theta)+\sum_j\int_{\underline{y}}^{y_j(\theta)}\overline{h}_{kj}\left(y_k(\theta),y_j(\theta')\right)x_j(\theta')\,dy_j(\theta')+\sum_j\int_{\underline{y}}^{\overline{y}}h_{kj}\left(y_k(\theta),y_j(\theta')\right)x_j(\theta')\,dy_j(\theta')$$

Substitute $x_k(\theta)$ in the above equation by (D74). Then, using the Fubini theorem to exchange the integration order and collecting the terms gives:

$$\begin{aligned}\frac{\widehat{y}_k}{y_k} &= -\frac{1}{\frac{1}{\varepsilon_k}+\frac{T''(y_k)y_k}{1-T'(y_k)}}\frac{\widehat{T}'(y_k)}{1-T'(y_k)}-\sum_j\int_{\underline{y}}^{y_j^k(y_k)}\frac{\overline{h}_{kj}\left(y_k,y_j'\right)}{\frac{1}{\varepsilon_j}+\frac{T''\left(y_j'\right)y_j(\theta)}{1-T'\left(y_j(\theta)\right)}}\frac{\widehat{T}'\left(y_j'\right)}{1-T'\left(y_j'\right)}dy_j'\\ &\quad-\sum_j\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c} e_{kk}^{C,own}\left(y_k,y_k'\right)1_{j=k}+e_{kj}^{C,cross}\left(y_k,y_j'\right)+\frac{h_{kj}\left(y_k,y_j'\right)}{\frac{1}{\varepsilon_j}+\frac{T''\left(y_j(\theta')\right)y_j(\theta')}{1-T'\left(y_j(\theta')\right)}}+\\ \int_{\underline{y}}^{y_j^k(y_k)}\overline{h}_{kj}\left(y_k,y_j''\right)e_{jj}^{C,own}\left(y_j'',y_j'\right)dy_j''+\int_{\underline{y}}^{\overline{y}}h_{kj}\left(y_k,y_j''\right)e_{jj}^{C,own}\left(y_j'',y_j'\right)dy_j''\\ +\sum_m\int_{\underline{y}}^{y_m^k(y_k)}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{C,own}\left(y_m'',y_j'\right)dy_m''\\ +\sum_m\int_{\underline{y}}^{\overline{y}}h_{km}\left(y_k,y_m''\right)e_{mj}^{C,own}\left(y_m'',y_j'\right)dy_m''\end{array}\right]\frac{\widehat{T}'\left(y_j'\right)}{1-T'\left(y_j'\right)}dy_j'\\ &\quad-\sum_j\left[\begin{array}{c} e_{kk}^{E,own}\left(y_k\right)1_{j=k}+e_{kj}^{E,cross}\left(y_k\right)\\ +\int_{\underline{y}}^{y_j^k(y_k)}\overline{h}_{kj}\left(y_k,y_j'\right)e_{jj}^{E,own}\left(y_j'\right)dy_j'+\int_{\underline{y}}^{\overline{y}}h_{kj}\left(y_k,y_j'\right)e_{jj}^{E,own}\left(y_j'\right)dy_j'\\ +\sum_m\int_{\underline{y}}^{y_m^k(y_k)}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{E,own}\left(y_m'',y_j'\right)dy_m''\\ +\sum_m\int_{\underline{y}}^{\overline{y}}\overline{h}_{km}\left(y_k,y_m''\right)e_{mj}^{E,own}\left(y_m'',y_j'\right)dy_m''\end{array}\right]\left[\widehat{E}_j-\sum_m\frac{N_m}{\sum_m N_m}\widehat{E}_m\right].\end{aligned}$$

By the definitions of $\varepsilon_k^C(y_k)$, $\varepsilon_{kj}^{Cross}\left(y_k, y_j'\right)$, $\varepsilon_j^{Cross}(y_k, y')$, and $\varepsilon_{kj}^E(y_k)$ given in equations (D66) to (D69), the above equation is equivalent to equation (D73). Since we have proved that there is a unique solution to the integral equation (D75), the equation (D73) gives one solution to the integral equation. The above equation is the unique solution to the equation (D75).■

D.4 Proof of Lemma 4

First, we want to prove that

$$\frac{\widehat{W}}{\rho^C} = -\left\{\sum_k \rho_k^C \times \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) + \sum_k \int_{\underline{\theta}}^{\overline{\theta}} \rho_k^C \left[1 - \widetilde{F}_k^C\left(y_k(x)\right)\right] \widehat{T}_k^{A\prime}\left(y_k(x)\right) dy_k(x)\right\}, \tag{D79}$$

where $\rho^C \equiv \sum_k \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(x) \frac{\partial U(x_k)}{\partial V_k(x_k)} dx_k$.

First, by Lemma 13 to 14, we have:

$$\begin{aligned} \widehat{U}(\theta) &= \sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \widehat{V}_k(\theta); \\ \widehat{V}_k(\theta) &= -\widehat{T}_k^A\left(y_k(\theta)\right) = -\left[\widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{A\prime}\left(y_k(x)\right) dy_k(x)\right] \end{aligned} \tag{D80}$$

By the defintion of $W \equiv \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(\theta) U(\theta) d\theta$ and $\widehat{W}$, we have:

$$\widehat{W} = \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(\theta) \widehat{U}(\theta) d\theta.$$

Substitute $\widehat{U}(\theta)$ in the above equation by equations (D80):

$$\widehat{W} = -\sum_k \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(x) \frac{\partial U(x)}{\partial V_k(x)} dx \times \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) - \sum_k \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} \widetilde{f}^C(\theta) \frac{\partial U(\theta)}{\partial V_k(\theta)} \widehat{T}_k^{A\prime}\left(y_k(x)\right) dy_k(x) d\theta.$$

Use the Fubini to exchange the order of the double integration in the right side of the above equation:

$$\widehat{W} = -\sum_k \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(x) \frac{\partial U(x)}{\partial V_k(x)} dx \times \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) - \sum_k \int_{\underline{\theta}}^{\overline{\theta}} \left[\int_{\theta}^{\overline{\theta}} \widetilde{f}^C(x) \frac{\partial U(x)}{\partial V_k(x)} dx\right] \widehat{T}_k^{A\prime}\left(y_k(\theta)\right) dy_k(\theta).$$

By definitions of ρ_k^C and $\widetilde{F}_k^C\left(y_k(\theta)\right)$:

$$\widehat{W} = -\left[\sum_k \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(x) \frac{\partial U(x)}{\partial V_k(x)} dx\right] \left[\sum_k \rho_k^C \times \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) + \sum_k \int_{\underline{\theta}}^{\overline{\theta}} \rho_k^C \left[1 - \widetilde{F}_k^C\left(y_k(\theta)\right)\right] \widehat{T}_k^{A\prime}\left(y_k(\theta)\right) dy_k(\theta)\right].$$

Rearrangment of the above equation gives (D80).■

D.5 Lemma 8

Define the following notations:

$$B_n^1\left(y_n\left(\theta\right)\right) \equiv \rho_n \frac{\left\{\begin{array}{c}\left[1-F_n\left(y_n\left(\theta\right)\right)\right]-\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]- \\ \sum_m \frac{\rho_m}{\rho_n}\frac{N_m}{N_n}\left[1-F_m\left(y_m\left(\theta\right)\right)\right]\overline{r}_{mn}^G\left(y_m\left(\theta\right)\right)-\frac{\rho_n^C}{\sum_k \rho_k^C \eta_{kn}^G}\left[1-\widetilde{F}_n^C\left(y_n\left(\theta\right)\right)\right]\end{array}\right\}}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]},$$

$$B_n^2\left(y_n\left(\theta\right)\right) \equiv (1-\rho_n)\frac{\left[1-F_n\left(y_n\left(\theta\right)\right)\right]\left[\frac{\varepsilon_n^A(y_n(\theta))}{\varepsilon_n^C(y_n(\theta))}+\left(1-\overline{r}_{nn}\right)\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^C(y_n(\theta))}\right]}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]},$$

and

$$\widetilde{B}_k\left(y_k\right) \equiv \rho_k \frac{\left\{\begin{array}{c}\left[1-\overline{r}_{kk}\right]\left[\frac{1-\widetilde{F}_k(y_k)}{1-F_k(y_k)}\right]+\overline{r}_{kk}\left(y_k\right)-\left[\frac{\rho_k^C}{\sum_m \rho_m^C \eta_{mk}^G}\right]\left[\frac{1-\widetilde{F}_k^C(y_k)}{[1-F_k(y_k)]}\right] \\ -\sum_m \frac{\rho_m N_m}{\rho_k N_n}\overline{r}_{mn}^G\left(y_m^k\left(y_k\right)\right)\frac{1-F_m\left(y_m^k(y_k)\right)}{1-F_k(y_k)}\end{array}\right\}}{\left[1-\overline{r}_{kk}\right]\left[\frac{1-F_k^R(y_k)-\left[1-\widetilde{F}_k(y_k)\right]}{1-F_k(y_k)}\right]} \tag{D81}$$

$$B_k\left(y_k\right) \equiv \widetilde{B}_k\left(y_k\right)+(1-\rho_k)\frac{\left[1-F_k\left(y_k\right)\right]\left[\frac{\varepsilon_k^A(y_k)}{\varepsilon_k^C(y_k)}+\left(1-\overline{r}_{kk}\right)\frac{\widetilde{F}_k(y_k)-F_k^R(y_k)}{y_k f_k(y_k)}\frac{1}{\varepsilon_k^C(y_k)}\right]}{\left[1-\overline{r}_{kk}\right]\left[\widetilde{F}_k\left(y_k\right)-F_k^R\left(y_k\right)\right]};$$

Using the above notations, we have the following results:

Lemma 8 *The marginal change of social welfare satisfies:*

$$\begin{aligned}
&-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k \rho_k^G \eta_{kj}^G\right]}=-\sum_n(1-\rho_n)\frac{N_n}{\sum_n N_n}\widehat{E}_n \qquad \text{(D82)}\\
&-\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\frac{N_n}{\sum_n N_n}f_n\left(y_n(\theta)\right)y_n(\theta)T'\left(y_n(\theta)\right)\frac{\widehat{y}_n(\theta)}{y_n(\theta)}dy_n(\theta)\\
&+\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\frac{N_n}{\sum_n N_n}y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)B_n^1\left(y_n\left(\theta\right)\right)C_n\left(y_n\left(\theta\right)\right)\left[1-T'\left(y_n\left(\theta\right)\right)\right]\frac{\widehat{y}_n(\theta)}{y_n(\theta)}dy_n\left(\theta\right)\\
&-\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\frac{N_n}{\sum_n N_n}\varepsilon_n^C\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)B_n^2\left(y_n\left(\theta\right)\right)C_n\left(y_n\left(\theta\right)\right)\widehat{T}'\left(y_n(\theta)\right)dy_n(\theta).
\end{aligned}$$

In particular, when Assumption 2 holds, we have:

$$\begin{aligned}
&-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k \rho_k^C \overline{\eta}_{kj}^G\right]} \qquad \text{(D83)}\\
=\;&-\sum_n\left\{\frac{N_n}{\sum_n N_n}(1-\rho_n)-\sum_m\int_{\underline{y}}^{\overline{y}}\Lambda\left(y\right)A_m\left(y\right)\left[\sum_n\varepsilon_{mn}^E\left(y\right)\right]\left\{T'\left(y\right)-\widetilde{B}_m\left(y\right)C_m\left(y\right)\left[1-T'\left(y\right)\right]\right\}dy\right\}\widehat{E}_n\\
&+\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\Lambda\left(y\right)\sum_n A_n\left(y\right)\left\{T'\left(y\right)-B_n\left(y\right)C_n\left(y\right)\left[1-T'\left(y\right)\right]\right\}\\ +\int_{\underline{y}}^{\overline{y}}\Lambda\left(y'\right)\sum_m A_m\left(y'\right)\left\{T'\left(y'\right)-\widetilde{B}_m\left(y'\right)C_m\left(y'\right)\left[1-T'\left(y'\right)\right]\right\}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'\end{array}\right]\frac{1}{1-T'\left(y\right)}d\widehat{T}\left(y\right),
\end{aligned}$$

where $\Lambda\left(y\right)=\sum_n\frac{N_n}{\sum_n N_n}f_n\left(y\right)y\varepsilon_n^C\left(y\right)$.

Proof. See the following proof. ■

D.6 Proof of Lemma 8

Step 1. We want to use (D62) to substitute $\sum_k \rho_k^C \times \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)$ in (D79). Before this, notice that by (D62), $\sum_k \rho_k^C \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)$ equals to:

$$\begin{aligned}
\sum_k \rho_k^C \quad \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right) = & \sum_j \sum_k \rho_k^C \eta_{kj}^G \left\{\widehat{E}_j - \sum_n \frac{N_n}{\sum_n N_n}\widehat{E}_n\right\} + \sum_j \sum_k \rho_k^C \eta_{kj}^G \left\{\begin{array}{c} \int_{\underline{y}}^{\underline{y}_j(\underline{\theta})} d\widehat{T}\left(y_j\left(\theta\right)\right) - \\ \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{\underline{y}_n(\underline{\theta})} d\widehat{T}\left(y_n\left(\theta\right)\right) \end{array}\right\} \\
& + \left\{\begin{array}{c} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_k \rho_k^C \eta_{kj}^G \left[1 - F_j\left(\theta\right)\right] \widehat{T}'\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) - \\ \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left(\sum_j \sum_k \rho_k^C \eta_{kj}^G\right) \frac{N_n}{\sum_n N_n} \left[1 - F_n\left(\theta\right)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \end{array}\right\} \\
& - \left\{\begin{array}{c} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_k \rho_k^C \eta_{kj}^C \left[1 - F_j\left(\theta\right)\right] \widehat{T}_j^{A\prime}\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) - \\ \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_n \sum_k \rho_k^C \eta_{kn}^G \left[1 - F_n\left(\theta\right)\right] \overline{r}_{nj}^G\left(\theta\right) \widehat{T}_j^{A\prime}\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) \end{array}\right\} \\
& - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_k \rho_k^C \eta_{kj}^C f_j\left(y_j\left(\theta\right)\right) y_j\left(\theta\right) C_j\left(y_j\left(\theta\right)\right) \left[1 - T'\left(y_j\left(\theta\right)\right)\right] \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta) \\
& + \left(\sum_m \sum_k \rho_k^C \eta_{km}^C\right) \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta).
\end{aligned}$$

Divide both sides of the above equation by $\sum_j \sum_k \rho_k^C \eta_{kj}^G$:

$$\begin{aligned}
\frac{\sum_k \rho_k^C \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} = & \sum_j \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left\{\widehat{E}_j - \sum_n \frac{N_n}{\sum_n N_n}\widehat{E}_n\right\} \\
& + \sum_j \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left\{\int_{\underline{y}}^{\underline{y}_j(\underline{\theta})} d\widehat{T}\left[y_j\left(\theta\right)\right] - \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{\underline{y}_n(\underline{\theta})} d\widehat{T}\left[y_n\left(\theta\right)\right]\right\} \\
& + \left\{\begin{array}{c} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left[1 - F_j\left(\theta\right)\right] \widehat{T}'\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) \\ - \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} \left[1 - F_n\left(\theta\right)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \end{array}\right\} \\
& - \left\{\begin{array}{c} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kj}^C}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left[1 - F_j\left(\theta\right)\right] \widehat{T}_j^{A\prime}\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) \\ - \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_n \frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left[1 - F_n\left(\theta\right)\right] \overline{r}_{nj}^G\left(\theta\right) \widehat{T}_j^{A\prime}\left(y_j\left(\theta\right)\right) dy_j\left(\theta\right) \end{array}\right\} \\
& - \left[\begin{array}{c} \sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kj}^C}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} f_j\left(y_j\left(\theta\right)\right) y_j\left(\theta\right) C_j\left(y_j\left(\theta\right)\right) \left[1 - T'\left(y_j\left(\theta\right)\right)\right] \frac{\widehat{y}_j(\theta)}{y_j(\theta)} dy_j(\theta) \\ + \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \end{array}\right].
\end{aligned} \tag{D84}$$

To simplify the above equation, we first rearrange terms on the right side of the above equation one by one:

$$\sum_j \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left\{\widehat{E}_j - \sum_n \frac{N_n}{\sum_n N_n}\widehat{E}_n\right\} = \sum_n \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n}\right) \widehat{E}_n,$$

and

$$\begin{aligned}
&\sum_j \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left\{ \int_{\underline{y}}^{\underline{y}_j(\underline{\theta})} d\widehat{T}\left[y_j(\theta)\right] - \sum_n \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{\underline{y}_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right] \right\} \\
= &\sum_n \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n} \right) \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right],
\end{aligned}$$

and

$$\begin{aligned}
&\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left[1 - F_j(\theta)\right] \widehat{T}'\left(y_j(\theta)\right) dy_j(\theta) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} \left[1 - F_n(\theta)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \\
= &\sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n} \right) \left[1 - F_n(\theta)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta),
\end{aligned}$$

and

$$\begin{aligned}
&\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kj}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left[1 - F_j(\theta)\right] \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta) - \\
&\sum_j \int_{y_j(\underline{\theta})}^{y_j(\overline{\theta})} \sum_n \frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \left[1 - F_n(\theta)\right] \bar{r}_{nj}^G(\theta) \widehat{T}_j^{A\prime}\left(y_j(\theta)\right) dy_j(\theta) \\
= &\sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left\{ \frac{\sum_k \rho_k^G \eta_{kn}^G}{\sum_n \sum_k \rho_k^G \eta_{kn}^G} \left[1 - F_n(\theta)\right] - \sum_m \frac{\sum_k \rho_k^G \eta_{km}^G}{\sum_m \sum_k \rho_k^G \eta_{km}^G} \left[1 - F(\theta)\right] \bar{r}_{mn}^G(\theta) \right\} \widehat{T}_n^{A\prime}\left(y_n(\theta)\right) dy_n(\theta).
\end{aligned}$$

Using the above equations to simplify the right side of (D84):

$$\begin{aligned}
&\frac{\sum_k \rho_k^C \widehat{T}_k^A\left(y_k(\underline{\theta})\right)}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} \\
= &\sum_n \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n} \right) \widehat{E}_n + \sum_n \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n} \right) \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right] \\
&+ \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n} \right) \left[1 - F_n(\theta)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left\{ \frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} \left[1 - F_n(\theta)\right] - \sum_m \frac{\sum_k \rho_k^C \eta_{km}^G}{\sum_m \sum_k \rho_k^C \eta_{km}^G} \left[1 - F(\theta)\right] \bar{r}_{mn}^G(\theta) \right\} \times \\
&\frac{\left[1 - T_n^{A\prime}\left(y_n(\theta)\right)\right]}{\varepsilon_n^A\left(y_n(\theta)\right)\left[1 - T'\left(y_n(\theta)\right)\right]} \left[1 - T'\left(y_n(\theta)\right)\right] \frac{\varepsilon_n^A\left(y_n(\theta)\right) \widehat{T}_n^{A\prime}\left(y_n(\theta)\right)}{1 - T_n^{A\prime}\left(y_n(\theta)\right)} dy_n(\theta) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} f_n\left(y_n(\theta)\right) y_n(\theta) C_n\left(y_n(\theta)\right) \left[1 - T'\left(y_n(\theta)\right)\right] \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta).
\end{aligned} \tag{D85}$$

Remember that $1-T'\left(y_n\left(\theta\right)\right)=\frac{1-T'\left(y_n(\theta)\right)-T_n^{\star\prime}\left(y_n(\theta)\right)}{1-\tau_n^{\star}(\theta_n)}$. We have

$$\begin{aligned}
& 1-\frac{T_n^{\star\prime}\left(y_n\left(\theta\right)\right)}{1-T'\left(y_n\left(\theta\right)\right)} \\
= \; & 1-\tau_n^{\star}\left(\theta_n\right)=1-\frac{\left[1-\overline{r}_{nn}\right]\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^A(y_n(\theta))}}{1+\left[1-\overline{r}_{nn}\right]\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^A(y_n(\theta))}} \\
= \; & \frac{\varepsilon_n^A\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]}\frac{\left[1-\overline{r}_{nn}\right]\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^A(y_n(\theta))}}{1+\left[1-\overline{r}_{nn}\right]\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^A(y_n(\theta))}} \\
= \; & \frac{y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]}C_n\left(y_n\left(\theta\right)\right),
\end{aligned}$$

where $C_n\left(y_n\left(\theta\right)\right)=\tau_n^{\star}\left(\theta_n\right)$. Therefore, we have

$$\frac{\left[1-T_n^{A\prime}\left(y_n\left(\theta\right)\right)\right]}{\varepsilon_n^A\left(y_n\left(\theta\right)\right)\left[1-T'\left(y_n\left(\theta\right)\right)\right]}=y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)\frac{C_n\left(y_n\left(\theta\right)\right)}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]}. \tag{D86}$$

Substitute $\frac{\left[1-T_n^{A\prime}(y_n(\theta))\right]}{\varepsilon_n^A(y_n(\theta))[1-T'(y_n(\theta))]}$ in (D85) by (D86):

$$\begin{aligned}
& \frac{\sum_k\rho_k^C\widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)}{\sum_j\sum_k\rho_k^G\eta_{kj}^G} \\
= \; & \sum_n\left(\frac{\sum_k\rho_k^C\eta_{kn}^G}{\sum_n\sum_k\rho_k^C\eta_{kn}^G}-\frac{N_n}{\sum_nN_n}\right)\widehat{E}_n+\sum_n\left(\frac{\sum_k\rho_k^C\eta_{kn}^G}{\sum_n\sum_k\rho_k^C\eta_{kn}^G}-\frac{N_n}{\sum_nN_n}\right)\times \\
& \int_{\underline{y}}^{y_n(\underline{\theta})}d\widehat{T}\left[y_n\left(\theta\right)\right]+\sum_n\int_{y_n(\underline{\theta})}^{y_n\left(\overline{\theta}\right)}\left(\frac{\sum_k\rho_k^C\eta_{kn}^G}{\sum_n\sum_k\rho_k^C\eta_{kn}^G}-\frac{N_n}{\sum_nN_n}\right)\left[1-F_n\left(\theta\right)\right]\widehat{T}'\left(y_n(\theta)\right)dy_n(\theta) \\
& -\sum_n\int_{y_n(\underline{\theta})}^{y_n\left(\overline{\theta}\right)}y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)\frac{\frac{\sum_k\rho_k^C\eta_{kn}^G}{\sum_n\sum_k\rho_k^C\eta_{kn}^G}\left[1-F_n\left(\theta\right)\right]-\sum_m\frac{\sum_k\rho_k^C\eta_{km}^G}{\sum_m\sum_k\rho_k^C\eta_{km}^G}\left[1-F\left(\theta\right)\right]\overline{r}_{mn}^G\left(\theta\right)}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]} \\
& \times C_n\left(y_n\left(\theta\right)\right)\left[1-T'\left(y_n\left(\theta\right)\right)\right]\frac{\varepsilon_n^A\left(y_n\left(\theta\right)\right)\widehat{T}_n^{A\prime}\left(y_n\left(\theta\right)\right)}{1-T_n^{A\prime}\left(y_n\left(\theta\right)\right)}dy_n\left(\theta\right) \\
& -\sum_n\int_{y_n(\underline{\theta})}^{y_n\left(\overline{\theta}\right)}\frac{\sum_k\rho_k^C\eta_{kn}^G}{\sum_n\sum_k\rho_k^C\eta_{kn}^G}f_n\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)C_n\left(y_n\left(\theta\right)\right)\left[1-T'\left(y_n\left(\theta\right)\right)\right]\frac{\widehat{y}_n(\theta)}{y_n(\theta)}dy_n(\theta) \\
& -\sum_n\int_{y_n(\underline{\theta})}^{y_n\left(\overline{\theta}\right)}\frac{N_n}{\sum_nN_n}f_n\left(y_n(\theta)\right)y_n(\theta)T'\left(y_n(\theta)\right)\frac{\widehat{y}_n(\theta)}{y_n(\theta)}dy_n(\theta).
\end{aligned}$$

Notice that $\frac{\widehat{y}_n(\theta)}{y_n(\theta)} = -\frac{\varepsilon_n^A(y_n(\theta))\widehat{T}_n^{A\prime}(y_n(\theta))}{1-T_n^{A\prime}(y_n(\theta))}$. The above equation is equivalent to:

$$
\begin{aligned}
&\frac{\sum_k \rho_k^C \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)}{\sum_j \sum_k \rho_k^G \eta_{kj}^G} \\
= \;& \sum_n \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n}\right) \widehat{E}_n + \sum_n \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n}\right) \\
&\times \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n\left(\theta\right)\right] + \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left(\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} - \frac{N_n}{\sum_n N_n}\right) \left[1 - F_n\left(\theta\right)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \\
&+ \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} y_n\left(\theta\right) f_n\left(y_n\left(\theta\right)\right) \frac{\frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G}\left[1 - F_n\left(\theta\right)\right] - \sum_m \frac{\sum_k \rho_k^C \eta_{km}^G}{\sum_m \sum_k \rho_k^C \eta_{km}^G}\left[1 - F\left(\theta\right)\right] \overline{r}_{mn}^G\left(\theta\right)}{\left[1 - \overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right) - F_n^R\left(y_n\left(\theta\right)\right)\right]} \\
&\times C_n\left(y_n\left(\theta\right)\right)\left[1 - T'\left(y_n\left(\theta\right)\right)\right] \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n\left(\theta\right) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} f_n\left(y_n\left(\theta\right)\right) y_n\left(\theta\right) C_n\left(y_n\left(\theta\right)\right)\left[1 - T'\left(y_n\left(\theta\right)\right)\right] \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta).
\end{aligned}
$$

By definition $\rho_n \equiv \frac{\sum_k \rho_k^C \eta_{kn}^G}{\sum_n \sum_k \rho_k^C \eta_{kn}^G} / \frac{N_n}{\sum_n N_n}$, we can rearrange the above equation to derive:

$$
\begin{aligned}
&\frac{\sum_k \rho_k^G \widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)}{\sum_j \sum_k \rho_k^G \eta_{kj}^G} \qquad\qquad \text{(D87)} \\
= \;& \sum_n \left(\rho_n - 1\right) \frac{N_n}{\sum_n N_n} \widehat{E}_n + \sum_n \left(\rho_n - 1\right) \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n\left(\theta\right)\right] \\
&+ \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left(\rho_n - 1\right) \frac{N_n}{\sum_n N_n}\left[1 - F_n\left(\theta\right)\right] \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta) \\
&- \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
&+ \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{f_n\left(y_n\left(\theta\right)\right) y_n\left(\theta\right) \rho_n N_n}{\sum_n N_n} \frac{\left\{\begin{array}{c}\left[1 - F_n\left(\theta\right)\right] - \left[1 - \overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right) - F_n^R\left(y_n\left(\theta\right)\right)\right] \\ - \sum_m \frac{\sum_k \rho_k^C \eta_{km}^G}{\sum_m \sum_k \rho_k^C \eta_{km}^G}\left[1 - F_m\left(\theta\right)\right] \overline{r}_{mn}^G\left(\theta\right)\end{array}\right\}}{\left[1 - \overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right) - F_n^R\left(y_n\left(\theta\right)\right)\right]} \\
&\times C_n\left(y_n\left(\theta\right)\right)\left[1 - T'\left(y_n\left(\theta\right)\right)\right] \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n\left(\theta\right).
\end{aligned}
$$

Remember that

$$
C_n\left(y_n\left(\theta\right)\right) = \frac{\left[1 - \overline{r}_{nn}\right] \frac{\widetilde{F}_n(y_n(\theta)) - F_n^R(y_n(\theta))}{y_n(\theta) f_n(y_n(\theta))} \frac{1}{\varepsilon_n^A(y_n(\theta))}}{1 + \left[1 - \overline{r}_{nn}\right] \frac{\widetilde{F}_n(y_n(\theta)) - F_n^R(y_n(\theta))}{y_n(\theta) f_n(y_n(\theta))} \frac{1}{\varepsilon_n^A(y_n(\theta))}}.
$$

Equivalently,

$$\begin{aligned} & \frac{\left[1-F_n\left(\theta\right)\right]}{\varepsilon_n^C\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)C_n\left(y_n\left(\theta\right)\right)} \\ = & \frac{\left[1-F_n\left(y_n\left(\theta\right)\right)\right]\left[\frac{\varepsilon_n^A(y_n(\theta))}{\varepsilon_n^C(y_n(\theta))}+\left(1-\overline{r}_{nn}\right)\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^C(y_n(\theta))}\right]}{\left[1-\overline{r}_{kk}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]}. \end{aligned} \tag{D88}$$

Therefore,

$$\begin{aligned} & \sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\left(\rho_n-1\right)\frac{N_n}{\sum_n N_n}\left[1-F_n\left(\theta\right)\right]\widehat{T}'\left(y_n(\theta)\right)dy_n(\theta) \\ = & -\sum_n\left(1-\rho_n\right)\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\frac{N_n}{\sum_n N_n}\frac{\varepsilon_n^C\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)\left[1-F_n\left(\theta\right)\right]}{\varepsilon_n^C\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)C_n\left(y_n\left(\theta\right)\right)}C_n\left(y_n\left(\theta\right)\right)\widehat{T}'\left(y_n(\theta)\right)dy_n(\theta) \\ = & -\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\frac{N_n}{\sum_n N_n}\varepsilon_n^C\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)\left(1-\rho_n\right)C_n\left(y_n\left(\theta\right)\right)\widehat{T}'\left(y_n(\theta)\right) \\ & \times\frac{\left[1-F_n\left(y_n\left(\theta\right)\right)\right]\left[\frac{\varepsilon_n^A(y_n(\theta))}{\varepsilon_n^C(y_n(\theta))}+\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1-\overline{r}_{nn}}{\varepsilon_n^C(y_n(\theta))}\right]}{\left[1-\overline{r}_{kk}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right]}dy_n(\theta), \end{aligned}$$

where the second equation uses equation (D88).

Then substitute $\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\left(\rho_n-1\right)\frac{N_n}{\sum_n N_n}\left[1-F_n\left(\theta\right)\right]\widehat{T}'\left(y_n(\theta)\right)dy_n(\theta)$ in equation (D87) by the above equation:

$$\begin{aligned} & \frac{\sum_k\rho_k^C\widehat{T}_k^A\left(y_k\left(\underline{\theta}\right)\right)}{\sum_j\sum_k\rho_k^C\eta_{kj}^G} \\ = & \sum_n\left(\rho_n-1\right)\frac{N_n}{\sum_n N_n}\widehat{E}_n+\sum_n\left(\rho_n-1\right)\frac{N_n}{\sum_n N_n}\int_{\underline{y}}^{y_n(\underline{\theta})}d\widehat{T}\left[y_n\left(\theta\right)\right] \\ & -\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\frac{N_n}{\sum_n N_n}f_n\left(y_n(\theta)\right)y_n(\theta)T'\left(y_n(\theta)\right)\frac{\widehat{y}_n(\theta)}{y_n(\theta)}dy_n(\theta) \\ & -\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\left\{\begin{array}{c}\frac{N_n}{\sum_n N_n}\varepsilon_n^C\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)f_n\left(y_n\left(\theta\right)\right)\left(1-\rho_n\right)C_n\left(y_n\left(\theta\right)\right)\widehat{T}'\left(y_n(\theta)\right) \\ \times\frac{\left[1-F_n(y_n(\theta))\right]\left[\frac{\varepsilon_n^A(y_n(\theta))}{\varepsilon_n^C(y_n(\theta))}+\left(1-\overline{r}_{nn}\right)\frac{\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))}{y_n(\theta)f_n(y_n(\theta))}\frac{1}{\varepsilon_n^C(y_n(\theta))}\right]}{\left[1-\overline{r}_{kk}\right]\left[\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))\right]}\end{array}\right\}dy_n(\theta) \\ & +\sum_n\int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})}\left\{\begin{array}{c}f_n\left(y_n\left(\theta\right)\right)y_n\left(\theta\right)\rho_n\frac{N_n}{\sum_n N_n}C_n\left(y_n\left(\theta\right)\right)\left[1-T'\left(y_n\left(\theta\right)\right)\right]\frac{\widehat{y}_n(\theta)}{y_n(\theta)} \\ \times\frac{\left\{\begin{array}{c}\left[1-F_n\left(\theta\right)\right]-\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n\left(\theta\right)\right)-F_n^R\left(y_n\left(\theta\right)\right)\right] \\ -\sum_m\frac{\sum_k\rho_k^C\eta_{km}^G}{\sum_m\sum_k\rho_k^C\eta_{km}^G}\left[1-F_m\left(\theta\right)\right]\overline{r}_{mn}^G\left(\theta\right)\end{array}\right\}}{\left[1-\overline{r}_{nn}\right]\left[\widetilde{F}_n(y_n(\theta))-F_n^R(y_n(\theta))\right]}\end{array}\right\}dy_n\left(\theta\right). \end{aligned} \tag{D89}$$

Rearrange eqaution (D79):

$$
\begin{aligned}
&-\frac{\widehat{W}}{\left[\sum_k \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(x)\, \frac{\partial U(x)}{\partial V_k(x)} dx\right]\left[\sum_n \sum_k \rho_k^C \eta_{kn}^G\right]} \\
= \quad & \frac{\sum_k \rho_k^C \widehat{T}_k^A\left(y_k(\underline{\theta})\right)}{\sum_j \sum_k \rho_k^C \eta_{kj}^G} + \frac{\sum_n \int_{\underline{\theta}}^{\overline{\theta}} \rho_n^C \left[1 - \widetilde{F}_n^C\left(y_n(\theta)\right)\right] \widehat{T}_n^{A\prime}\left(y_n(\theta)\right) dy_n(\theta)}{\sum_n \sum_k \rho_k^C \eta_{kn}^G}.
\end{aligned}
$$

Substitute $\frac{\sum_k \rho_k^C \widehat{T}_k^A(y_k(\underline{\theta}))}{\sum_j \sum_k \rho_k^C \eta_{kj}^G}$ in the above equation by (D89):

$$
\begin{aligned}
&-\frac{\widehat{W}}{\left[\sum_k \int_{\underline{\theta}}^{\overline{\theta}} \widetilde{f}^C(x)\, \frac{\partial U(x)}{\partial V_k(x)} dx\right]\left[\sum_n \sum_k \rho_k^C \eta_{kn}^G\right]} \qquad \text{(D90)} \\
= \quad & \sum_n (\rho_n - 1) \frac{N_n}{\sum_n N_n} \widehat{E}_n + \sum_n (\rho_n - 1) \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right] \\
& - \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
& - \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left\{ \begin{array}{c} \frac{N_n}{\sum_n N_n} \varepsilon_n^C\left(y_n(\theta)\right) y_n(\theta) f_n\left(y_n(\theta)\right)(1-\rho_n) C_n\left(y_n(\theta)\right) \widehat{T}'\left(y_n(\theta)\right) \\ \times \frac{[1-F_n(y_n(\theta))]\left[\frac{\varepsilon_n^A(y_n(\theta))}{\varepsilon_n^C(y_n(\theta))} + (1-\bar{r}_{nn}) \frac{\widetilde{F}_n(y_n(\theta)) - F_n^R(y_n(\theta))}{y_n(\theta) f_n(y_n(\theta))} \frac{1}{\varepsilon_n^C(y_n(\theta))}\right]}{[1-\bar{r}_{kk}]\left[\widetilde{F}_n(y_n(\theta)) - F_n^R(y_n(\theta))\right]} \end{array} \right\} dy_n(\theta) \\
& + \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \left\{ \begin{array}{c} \frac{N_n}{\sum_n N_n} y_n(\theta) f_n\left(y_n(\theta)\right) C_n\left(y_n(\theta)\right) \frac{[1-T'(y_n(\theta))]\widehat{y}_n(\theta)}{y_n(\theta)} \\ \times \frac{\left[\begin{array}{c} \rho_n\left[1 - F_n(\theta)\right] - \rho_n\left[1 - \bar{r}_{nn}\right]\left[\widetilde{F}_n\left(y_n(\theta)\right) - F_n^R\left(y_n(\theta)\right)\right] - \\ \sum_m \rho_m \frac{N_m}{N_n}\left[1 - F_m(\theta)\right] \bar{r}_{mn}^G(\theta) - \frac{\rho C}{\sum_k \rho_k^C \eta_{kn}^G} \rho_n \left[1 - \widetilde{F}_n^C\left(y_n(\theta)\right)\right] \end{array}\right]}{[1-\bar{r}_{nn}]\left[\widetilde{F}_n(y_n(\theta)) - F_n^R(y_n(\theta))\right]} \end{array} \right\} dy_n(\theta).
\end{aligned}
$$

Use the definition of $B_n^1\left(y_n(\theta)\right)$, $B_n^2\left(y_n(\theta)\right)$, and ρ^C to simplify the expression of (D90):

$$
\begin{aligned}
-\frac{\widehat{W}}{\rho^C\left[\sum_j \sum_k \rho_k^G \eta_{kj}^G\right]} &= -\sum_n (1-\rho_n) \frac{N_n}{\sum_n N_n} \widehat{E}_n - \sum_n (1-\rho_n) \frac{N_n}{\sum_n N_n} \int_{\underline{y}}^{y_n(\underline{\theta})} d\widehat{T}\left[y_n(\theta)\right] \qquad \text{(D91)} \\
& - \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
& + \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} y_n(\theta) f_n\left(y_n(\theta)\right) B_n^1\left(y_n(\theta)\right) C_n\left(y_n(\theta)\right)\left[1 - T'\left(y_n(\theta)\right)\right] \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
& - \sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} \varepsilon_n^C\left(y_n(\theta)\right) y_n(\theta) f_n\left(y_n(\theta)\right) B_n^2\left(y_n(\theta)\right) C_n\left(y_n(\theta)\right) \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta).
\end{aligned}
$$

Since we have assumed that $\underline{\theta}_k = \underline{\theta}$, $\overline{\theta}_k = \overline{\theta}$, $y_n(\underline{\theta}) = \underline{y}$, $\overline{y}_n(\overline{\theta}) = \overline{y}$. We can simplify the above equation:

$$
\begin{aligned}
&-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k \rho_k^G \eta_{kj}^G\right]} = -\sum_n (1-\rho_n)\frac{N_n}{\sum_n N_n}\widehat{E}_n \\
&-\sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} f_n\left(y_n(\theta)\right) y_n(\theta) T'\left(y_n(\theta)\right) \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
&+\sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} y_n(\theta) f_n\left(y_n(\theta)\right) B_n^1\left(y_n(\theta)\right) C_n\left(y_n(\theta)\right)\left[1-T'\left(y_n(\theta)\right)\right] \frac{\widehat{y}_n(\theta)}{y_n(\theta)} dy_n(\theta) \\
&-\sum_n \int_{y_n(\underline{\theta})}^{y_n(\overline{\theta})} \frac{N_n}{\sum_n N_n} \varepsilon_n^C\left(y_n(\theta)\right) y_n(\theta) f_n\left(y_n(\theta)\right) B_n^2\left(y_n(\theta)\right) C_n\left(y_n(\theta)\right) \widehat{T}'\left(y_n(\theta)\right) dy_n(\theta).
\end{aligned}
$$

Step2. When Condition 2 holds, by Propostion 5, we have:

$$
\frac{\widehat{y}_n(\theta)}{y_n(\theta)} = -\varepsilon_n^C\left(y_n(\theta)\right)\frac{\widehat{T}'\left(y_n(\theta)\right)}{1-T'\left(y_n(\theta)\right)} - \int_{y(\underline{\theta})}^{y(\overline{\theta})} \varepsilon_n^{Cross}\left(y_n(\theta), y(x)\right)\frac{\widehat{T}'\left(y(x)\right)}{1-T'\left(y(x)\right)} dy(x) - \sum_j \varepsilon_{nj}^E\left(y_n(\theta)\right)\widehat{E}_j,
$$

where $\varepsilon_n^C\left(y_n(\theta)\right) = \frac{1}{\frac{1}{\varepsilon_n} + \frac{T''(y_n(\theta))y_n(\theta)}{1-T'(y_n(\theta))}}$. Substitute $\frac{\widehat{y}_n(\theta)}{y_n(\theta)}$ in equation (D91) by the above equation:

$$
\begin{aligned}
&-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k \rho_k^C \overline{\eta}_{kj}^G\right]} = -\sum_n (1-\rho_n)\frac{N_n}{\sum_n N_n}\widehat{E}_n \\
&-\sum_n \int_{\underline{y}}^{\overline{y}} \frac{N_n}{\sum_n N_n} \varepsilon_n^C\left(y_n\right) y_n f_n\left(y_n\right) B_n^2\left(y_n\right) C_n\left(y_n\right) \widehat{T}'\left(y_n\right) dy_n \\
&-\sum_n \int_{\underline{y}}^{\overline{y}} \frac{N_n}{\sum_n N_n} f_n\left(y_n\right) y_n T'\left(y_n\right) \left\{ \begin{array}{c} -\varepsilon_n^C\left(y_n\right)\frac{\widehat{T}'(y_n)}{1-T^{*\prime}(y_n)} - \sum_m \varepsilon_{nm}^E\left(y_n(\theta)\right)\widehat{E}_m \\ -\int_{\underline{y}}^{\overline{y}} \varepsilon_n^{Cross}\left(y_n, y'\right)\frac{\widehat{T}'(y')}{1-T'(y')} dy' \end{array} \right\} dy_n \\
&+\sum_n \int_{\underline{y}}^{\overline{y}} \frac{N_n}{\sum_n N_n} y_n f_n\left(y_n\right) B_n^1\left(y_n\right) C_n\left(y_n\right)\left[1-T'\left(y_n\right)\right] \left\{ \begin{array}{c} -\varepsilon_n^C\left(y_n\right)\frac{\widehat{T}'(y_n)}{1-T^{*\prime}(y_n)} - \sum_m \varepsilon_{nm}^E\left(y_n\right)\widehat{E}_m \\ -\int_{\underline{y}}^{\overline{y}} \varepsilon_n^{Cross}\left(y_n, y'\right)\frac{\widehat{T}'(y')}{1-T'(y')} dy' \end{array} \right\} dy_n.
\end{aligned}
$$

Use the Fubini theorem to exchange the order of integration:

$$
\begin{aligned}
&-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k \rho_k^C \overline{\eta}_{kj}^G\right]} \\
= \;& -\sum_n \frac{N_n}{\sum_n N_n}\left\{(1-\rho_n) - \sum_m \int_{\underline{y}}^{\overline{y}} \frac{N_m}{N_n} y f_m(y)\,\varepsilon_{mn}^E(y)\left\{T'\left(y'\right) - B_m^1(y) C_m(y)\left[1-T'\left(y'\right)\right]\right\} dy\right\}\widehat{E}_n \\
&+\int_{\underline{y}}^{\overline{y}} \left[\begin{array}{c} \sum_n \frac{N_n}{\sum_n N_n} y f_n(y)\,\varepsilon_n^C(y)\left\{T'\left(y'\right) - \left[B_n^1(y) + B_n^2(y)\right] C_n(y)\left[1-T'\left(y'\right)\right]\right\} \\ +\sum_m \int_{\underline{y}}^{\overline{y}} \frac{N_m}{\sum_n N_n} f_m\left(y'_m\right) y'_m \left\{T'\left(y'_m\right) - B_m^1\left(y'_m\right) C_m\left(y'_m\right)\left[1-T'\left(y'_m\right)\right]\right\} \varepsilon_m^{Cross}\left(y'_m, y\right) dy'_m \end{array} \right] \frac{d\widehat{T}(y)}{1-T'(y)}.
\end{aligned}
$$

Use defintions of $\widetilde{B}_n(y)$ and $B_n(y)$:

$$-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k\rho_k^C\overline{\eta}_{kj}^G\right]}=-\sum_n\left\{\begin{array}{c}\frac{N_n}{\sum_n N_n}(1-\rho_n)-\\ \sum_m\int_{\underline{y}}^{\overline{y}}\frac{N_m}{\sum_n N_n}yf_m(y)\left[\sum_n\varepsilon_{mn}^E(y)\right]\left[T'y)-\widetilde{B}_m(y)C_m(y)\left[1-T'(y)\right]\right]dy\end{array}\right\}\widehat{E}_n$$
$$+\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\sum_n\frac{N_n}{\sum_n N_n}f_n(y)y\varepsilon_n^C(y)\left\{T'(y)-B_n(y)C_n(y)\left[1-T'(y)\right]\right\}\\ +\sum_m\int_{\underline{y}}^{\overline{y}}\frac{N_m}{\sum_n N_n}f_m(y')y'\left\{T'(y')-\widetilde{B}_m(y')C_m(y')\left[1-T'(y')\right]\right\}\times\varepsilon_m^{Cross}(y',y')dy'\end{array}\right]\frac{d\widehat{T}(y)}{\left[1-T'(y)\right]}.$$

Use $\Lambda(y)=\sum_n\frac{N_n}{\sum_n N_n}f_n(y)y\varepsilon_n^C(y)$ to simplify the above equation:

$$\begin{aligned}&-\frac{\widehat{W}}{\rho^C\left[\sum_j\sum_k\rho_k^C\overline{\eta}_{kj}^G\right]}\\ =\;&-\sum_n\left\{\begin{array}{c}\frac{N_n}{\sum_n N_n}(1-\rho_n)-\\ \sum_m\int_{\underline{y}}^{\overline{y}}\Lambda(y))A_m(y)\left[\sum_n\varepsilon_{mn}^E(y)\right]\left\{T'(y)-\widetilde{B}_m(y)C_m(y)\left[1-T'(y)\right]\right\}dy\end{array}\right\}\widehat{E}_n\\ &+\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\Lambda(y)\sum_n A_n(y)\left\{T'(y)-B_n(y)C_n(y)\left[1-T'(y)\right]\right\}+\\ \int_{\underline{y}}^{\overline{y}}\Lambda(y')\sum_m A_m(y')\left\{T'(y')-\widetilde{B}_m(y')C_m(y')\left[1-T'(y')\right]\right\}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'\end{array}\right]\frac{d\widehat{T}(y)}{1-T'(y)}.\end{aligned}\tag{D92}$$

The above equation is equivalent to (D79).■

D.7 Theorem 3: Integral Equation for the Optimal Taxation

For the optimal central income tax, we have the following general result.

Theorem 3 *When the regular Assumption 2 hold, for any $y\in\mathcal{Y}$, the optimal central income tax rate satisfies:*

$$\begin{aligned}T'(y)\;=\;&\frac{\sum_n A_n(y)B_n(y)C_n(y)}{\sum_n A_n(y)\left[1+B_n(y)C_n(y)\right]}\\ &+\int_{\underline{y}}^{\overline{y}}\frac{\Lambda(y')}{\Lambda(y)}\frac{\sum_n A_n(y')\widetilde{B}_n(y')C_n(y')}{\sum_n A_n(y)\left[1+B_n(y)C_n(y)\right]}\frac{\varepsilon_n^{Cross}(y',y)}{\varepsilon_n^C(y')}dy'\\ &-\int_{\underline{y}}^{\overline{y}}\frac{\Lambda(y')}{\Lambda(y)}\frac{\sum_n A_n(y')\left[1+\widetilde{B}_n(y')C_n(y')\right]}{\sum_n A_n(y)\left[1+B_n(y)C_n(y)\right]}\frac{\varepsilon_n^{Cross}(y',y)}{\varepsilon_n^C(y')}T'(y')\,dy',\end{aligned}\tag{D93}$$

where $\Lambda(y_n(\theta))=\sum_n\frac{N_n}{\sum_n N_n}f_n(y_n(\theta))y_n(\theta)\varepsilon_n^C(y_n(\theta))$, $A_j(y)=\frac{N_jf_j(y)y\varepsilon_j^C(y)}{\sum_j N_jf_j(y)y\varepsilon_j^C(y)}$, $B_j(y)$ is given in equation (D81), $C_j(y)$ in given in equation (31). Moreover, if the inter-region transfer is freely chosen by the central government, we have

$$0=\frac{N_n}{\sum_n N_n}(1-\rho_n)-\int_{\underline{y}}^{\overline{y}}\sum_m A_m(y')\left[\begin{array}{c}T'(y')-\\ \widetilde{B}_m(y')C_m(y')\left[1-T'(y')\right]\end{array}\right]\left[\Lambda(y')\sum_n\varepsilon_{mn}^E(y')\right]dy'.\tag{D94}$$

Proof. See the following proof. ■

D.8 Proof of Theorem 3

Step 1. When the central government can not manipulate E_n, and the central income tax is optimal, we have $\widehat{W}=0$ for any tax perturbation and $\widehat{E}_n=0$. Then according to (D83), we have:

$$0=\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\Lambda(y)\sum_n A_n(y)\left\{T^{\star\prime}(y')-B_n(y)C_n(y)\left[1-T^{\star\prime}(y)\right]\right\}\\+\int_{\underline{y}}^{\overline{y}}\Lambda(y')\sum_m A_m(y')\left\{T^{\star\prime}(y')-\widetilde{B}_m(y')C_m(y')\left[1-T^{\star\prime}(y')\right]\right\}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'\end{array}\right]\frac{d\widehat{T}(y)}{1-T^{\star\prime}(y)}.$$

Since for any $\widehat{T}(y)$ the above equation hold, we can consider the elementary tax reform so that $\left\{\widehat{T}(\widetilde{y})=1_{y=\widetilde{y}}\right\}_{\widetilde{y}\in[\underline{y},\overline{y}]}$. Then for any $\widetilde{y}\in\mathcal{Y}$:

$$0=\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\Lambda(y)\sum_n A_n(y)\left\{T^{\star\prime}(y')-B_n(y)C_n(y)\left[1-T^{\star\prime}(y)\right]\right\}\\+\int_{\underline{y}}^{\overline{y}}\Lambda(y')\sum_m A_m(y')\left\{T^{\star\prime}(y')-\widetilde{B}_m(y')C_m(y')\left[1-T^{\star\prime}(y')\right]\right\}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'\end{array}\right]\frac{d1_{y=\widetilde{y}}}{1-T^{\star\prime}(y)}.$$

We have the following result for any $\widetilde{y}\in\mathcal{Y}$:

$$\begin{aligned}0 &= \sum_n A_n(\widetilde{y})\left\{T'(\widetilde{y})-B_n(\widetilde{y})C_n(\widetilde{y})\left[1-T'(\widetilde{y})\right]\right\}+\\&\int_{\underline{y}}^{\overline{y}}\sum_m A_m(y')\left\{T'(y')-\widetilde{B}_m(y')C_m(y')\left[1-T'(y')\right]\right\}\frac{\Lambda(y')}{\Lambda(\widetilde{y})}\frac{\varepsilon_m^{Cross}(y',\widetilde{y})}{\varepsilon_m^C(y')}dy'\end{aligned}$$

Alternatively, for $y\in\mathcal{Y}$:

$$\begin{aligned}0 &= \sum_n A_n(y)\left\{T'(y)-B_n(y)C_n(y)\left[1-T'(y)\right]\right\}+\\&\int_{\underline{y}}^{\overline{y}}\sum_m A_m(y')\left\{T'(y')-\widetilde{B}_m(y')C_m(y')\left[1-T'(y')\right]\right\}\frac{\Lambda(y')}{\Lambda(y)}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'.\end{aligned}\tag{D95}$$

Rearranging (D95) gives (D93).

Step 2. If the central government can both choose central income tax and inter-region transfer, then one has optimal conditions for the central income tax and transfer. To derive the conditions, we first reform the central income tax while keeping the inter-region transfer unchanged. Then we again has the above equation (D93). On the other hand, since in the neighbor of optimal tax and transfer, for any marginal reforms $\left\{\widehat{T}(y),\widehat{E}_k\right\}$, one has $\widehat{W}=0$, we have the following equation by equation (D83):

$$\begin{aligned}0=&-\sum_n\left[\begin{array}{c}\frac{N_n}{\sum_n N_n}(1-\rho_n)-\\\sum_m\int_{\underline{y}}^{\overline{y}}\Lambda(y)A_m(y)\left[\sum_n\varepsilon_{mn}^E(y)\right]\left\{T^{\star\prime}(y)-\widetilde{B}_m(y)C_m(y)\left[1-T^{\star\prime}(y)\right]\right\}dy\end{array}\right]\widehat{E}_n\\&+\int_{\underline{y}}^{\overline{y}}\left[\begin{array}{c}\Lambda(y)\sum_n A_n(y)\left\{T^{\star\prime}(y')-B_n(y)C_n(y)\left[1-T^{\star\prime}(y)\right]\right\}+\\\int_{\underline{y}}^{\overline{y}}\Lambda(y')\sum_m A_m(y')\left\{T^{\star\prime}(y')-\widetilde{B}_m(y')C_m(y')\left[1-T^{\star\prime}(y')\right]\right\}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'\end{array}\right]\frac{d\widehat{T}(y)}{1-T^{\star\prime}(y)}.\end{aligned}$$

Since for any $y \in \mathcal{Y}$, (D95) is satisfied, the above equation can be simplified as:

$$0 = -\sum_n \left\{ \begin{array}{c} \frac{N_n}{\sum_n N_n}(1-\rho_n) - \\ \sum_m \int_{\underline{y}}^{\overline{y}} \Lambda(y) A_m(y) \left[\sum_n \varepsilon_{mn}^E(y)\right] \left[T^{\star\prime}(y) - \widetilde{B}_m(y) C_m(y) \left[1 - T^{\star\prime}(y)\right]\right] dy \end{array} \right\} \widehat{E}_n.$$

Notice that for any $\left\{\widehat{E}_n\right\}$ the above equation should be zero when the original tax and transfer are optimal. We have equation (D94).■

D.9 Proposition 7: Solution to the Integral Equation

In Proposition 7, we show the existence and uniqueness of the solution to the integral equation.

Assumption 4 *Functions* $\frac{\sum_n A_n(y)B_n(y)C_n(y)}{\sum_n A_n(y)[1+B_n(y)C_n(y)]}$, $\frac{\sum_n A_n(y')\widetilde{B}_n(y')C_n(y')}{\sum_n A_n(y)[1+B_n(y)C_n(y)]}$, *and* $\frac{\Lambda(y')}{\Lambda(y)}\frac{\varepsilon_n^{Cross}(y',y)}{\varepsilon_n^C(y')}$ *are square-integrable function in* $\mathcal{Y}$, *where* $\Lambda(y) = \sum_n \frac{N_n}{\sum_n N_n} f_n(y) y \varepsilon_n^C(y)$. *They satisfy the following inequality:*

$$\left[\int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{\overline{y}} \left\{ \frac{\Lambda(y')}{\Lambda(y)} \frac{\sum_n A_n(y') \left[1 + \widetilde{B}_n(y') C_n(y')\right]}{\sum_n A_n(y) \left[1 + B_n(y) C_n(y)\right]} \frac{\varepsilon_n^{Cross}(y',y)}{\varepsilon_n^C(y')} \right\}^2 dy'dy \right]^{\frac{1}{2}} < 1.$$

Under Assumption 4, we can well define $B_n^*(y)$:

$$B_n^*(y) = B_n(y) + \int_{\underline{y}}^{\overline{y}} \left\{ \begin{array}{c} \frac{A_n(y')C_n(y')}{A_n(y)C_n(y)} \left[\begin{array}{c} \frac{\Lambda(y')}{\Lambda(y)} \frac{\varepsilon_n^{Cross}(y',y)}{\varepsilon_n^C(y')} \\ + \int_{\underline{y}}^{\overline{y}} g(y,y'') \frac{\Lambda(y')}{\Lambda(y'')} \frac{\varepsilon_n^{Cross}(y',y'')}{\varepsilon_n^C(y')} \frac{\sum_n A_n(y)[1+B_n(y)C_n(y)]}{\sum_n A_n(y'')[1+B_n(y'')C_n(y'')]} dy'' \end{array} \right] \widetilde{B}_n(y') \\ + g(y,y') \frac{A_n(y')C_n(y')}{A_n(y)C_n(y)} \frac{\sum_n A_n(y)[1+B_n(y)C_n(y)]}{\sum_n A_n(y')[1+B_n(y')C_n(y')]} B_n(y') \end{array} \right\} dy', \tag{D96}$$

where $g(y,y')$ is the solution to the following integral equation

$$\begin{aligned} g(y,y') &= -\frac{\Lambda(y')}{\Lambda(y)} \frac{\sum_n A_n(y') \left[1 + \widetilde{B}_n(y') C_n(y')\right]}{\sum_n A_n(y) \left[1 + B_n(y) C_n(y)\right]} \frac{\varepsilon_n^{Cross}(y',y)}{\varepsilon_n^C(y')} - \\ &\quad \int_{\underline{y}}^{\overline{y}} \frac{\Lambda(y')}{\Lambda(y'')} \frac{\sum_n A_n(y') \left[1 + \widetilde{B}_n(y') C_n(y')\right]}{\sum_n A_n(y'') \left[1 + B_n(y'') C_n(y'')\right]} \frac{\varepsilon_n^{Cross}(y',y'')}{\varepsilon_n^C(y')} g(y'',y') \, dy''. \end{aligned}$$

Assumption 4 guarantees the existence and uniqueness of the solution to the above equation. Under the above assumptions and definitions, we have the following proposition:

Proposition 7 *When Assumption 2 holds, there exist unique optimal central income tax schedule, which satisfies*

$$\begin{aligned} T^{\star\prime}(y) &= \frac{\sum_n A_n(y) B_n(y) C_n(y)}{\sum_n A_n(y) \left[1 + B_n(y) C_n(y)\right]} \\ &\quad + \int_{\underline{y}}^{\overline{y}} \frac{\Lambda(y')}{\Lambda(y)} \frac{\sum_m \varepsilon_{nm}^{Cross}(y',y)}{\varepsilon_n^C(y')} \frac{\sum_n A_n(y') \widetilde{B}_n(y') C_n(y')}{\sum_n A_n(y) \left[1 + B_n(y) C_n(y)\right]} dy' \\ &\quad - \int_{\underline{y}}^{\overline{y}} \frac{\Lambda(y')}{\Lambda(y)} \frac{\sum_n A_n(y') \left[1 + \widetilde{B}_n(y') C_n(y')\right]}{\sum_n A_n(y) \left[1 + B_n(y) C_n(y)\right]} \frac{\sum_m \varepsilon_{nm}^{Cross}(y',y)}{\varepsilon_n^C(y')} T^{\star\prime}(y') \, dy'. \end{aligned} \tag{D97}$$

where $\Lambda(y) = \sum_n \frac{N_n}{\sum_n N_n} f_n(y) y\varepsilon_n^C(y)$. *Alternatively, we have:*

$$T^{\star\prime}(y) = \frac{\sum_n A_n(y) B_n^*(y) C_n(y)}{\sum_n A_n(y)[1 + B_n(y) C_n(y)]}. \tag{D98}$$

Proof. See the following proof. ■

D.10 Proof of Proposition 7

To shorten the expression, set $x(y)$ and $\gamma(y, y')$:

$$\begin{aligned} x(y) &= \frac{\sum_n A_n(y) B_n(y) C_n(y)}{\sum_n A_n(y)[1 + B_n(y) C_n(y)]} + \int_{\underline{y}}^{\overline{y}} \frac{\Lambda(y')}{\Lambda(y)} \frac{\varepsilon_n^{Cross}(y', y)}{\varepsilon_n^C(y')} \frac{\sum_n A_n(y') \widetilde{B}_n(y') C_n(y')}{\sum_n A_n(y)[1 + B_n(y) C_n(y)]} dy', \\ \gamma(y, y') &= -\frac{\Lambda(y')}{\Lambda(y)} \frac{\sum_n A_n(y') \left[1 + \widetilde{B}_n(y') C_n(y')\right]}{\sum_n A_n(y)[1 + B_n(y) C_n(y)]} \frac{\varepsilon_n^{Cross}(y', y)}{\varepsilon_n^C(y')}. \end{aligned}$$

Then the equation (D93) is equivalent to:

$$T'(y) = x(y) + \int_{\underline{y}}^{\overline{y}} \gamma(y, y') T'(y') dy'. \tag{D99}$$

When Assumption 4 holds, by Minkowski inequality and Holder inequality, we know $\gamma(y, y')$ and $x(y)$ are square integrable function. By Lemma 10, there is a unique solution to the above equation (D99). By Lemma 11, when Assumption 4 holds, there exist a unique resolvent $g(y, y')$ to the equation (D99):

$$g(y, y') = \gamma(y, y') + \int_{\underline{y}}^{\overline{y}} \gamma(y, y'') g(y'', y') dy''$$

Moreover, we can solve $g(y, y')$ by taking the limit of sequence $\{g^n(y, y')\}$, i.e., $g(y, y') = \lim_{n\to\infty} g^n(y, y')$:

$$g^{n+1}(y, y') = \gamma(y, y') + \int_{\underline{y}}^{\overline{y}} \gamma(y, y'') g^n(y'', y') dy''; n \geq 0, g^0(y, y') = \gamma(y, y').$$

By Lemma 12, when Assumption 4 holds, the solution of equation (D99) is

$$T'(y) = x(y) + \int_{\underline{y}}^{\overline{y}} g(y, y') x(y') dy'.$$

By definitions of $x(y)$ and $g(y, y')$:

$$T^{\star\prime}(y) = \frac{\sum_n A_n(y) \left[B_n(y) + \int_{\underline{y}}^{\overline{y}} \frac{A_n(y') C_n(y')}{A_n(y) C_n(y)} \left(\begin{array}{c} \frac{\Lambda(y')}{\Lambda(y)} \frac{\varepsilon_n^{Cross}(y', y)}{\varepsilon_n^C(y')} \widetilde{B}_n(y') + \\ \widetilde{B}_n(y') \int_{\underline{y}}^{\overline{y}} g(y, y'') \frac{\Lambda(y')}{\Lambda(y'')} \frac{\varepsilon_n^{Cross}(y', y'')}{\varepsilon_n^C(y')} \frac{\sum_n A_n(y)[1+B_n(y)C_n(y)]}{\sum_n A_n(y'')[1+B_n(y'')C_n(y'')]} dy'' \\ + g(y, y') \frac{\sum_n A_n(y)[1+B_n(y)C_n(y)]}{\sum_n A_n(y')[1+B_n(y')C_n(y')]} B_n(y') \end{array} \right) dy' \right] C_n(y)}{\sum_n A_n(y)[1 + B_n(y) C_n(y)]}$$

Use the definition of $B_n^*(y)$, the above equation is equivalent to (D98)

$$T^{\star\prime}(y) = \frac{\sum_n A_n(y) B_n^*(y) C_n(y)}{\sum_n A_n(y)[1 + B_n(y) C_n(y)]}.$$

■

D.11 Proof of Proposition 6: Optimal Tax When $\rho_n = 1$

By Proposition 7 when $\rho_n = 1$, one has for any $y_n(\theta)$ and n (see equation (D81)):

$$\widetilde{B}_n(y_n(\theta)) = B_n(y_n(\theta)).$$

Then equation (D97) can be written as:

$$\begin{aligned} 0 &= \sum_n A_n(y)\left\{T^{\star\prime}(y) - B_n(y)C_n(y)\left[1 - T^{\star\prime}(y)\right]\right\} + \\ &\int_{\underline{y}}^{\overline{y}} \sum_m A_m(y') \frac{\Lambda(y')}{\Lambda(y)} \frac{\varepsilon_m^{Cross}(y', y)}{\varepsilon_m^C(y')} \left\{T^{\star\prime}(y') - B_m(y')C_m(y')\left[1 - T^{\star\prime}(y')\right]\right\} dy'. \end{aligned}$$

Remember that ε_m^C is independent of m. In addition, if $\varepsilon_m^{Cross}(y', y)$ is independent of $m \in \mathcal{J}$ for $y', y \in \mathcal{Y}$, we have

$$\begin{aligned} 0 &= \sum_n A_n(y)\left\{T^{\star\prime}(y) - B_n(y)C_n(y)\left[1 - T^{\star\prime}(y)\right]\right\} + \\ &\int_{\underline{y}}^{\overline{y}} \frac{\varepsilon^{Cross}(y', y)}{\varepsilon^C(y')} \left\{\sum_m A_m(y')T^{\star\prime}(y') - \sum_m A_m(y')B_m(y')C_m(y')\left[1 - T^{\star\prime}(y')\right]\right\} dy'. \end{aligned} \tag{D100}$$

where we omit the subscript of ε_m^C and ε_m^{Cross} because they are indpendent of m. We have the following $T^{\star\prime}(y)$ as a solution to the above equation.

$$T^{\star\prime}(y) = \frac{\sum_n A_n(y)B_n(y)C_n(y)}{\sum_n A_n(y)\left[1 + B_n(y)C_n(y)\right]}$$

Moreover, since there is a unique solution to (D100), the above solution is the unique solution.■

D.12 Supplement to Proposition 6: When $\rho_n = 1$ is Optimal

When the central government can freely choose central income tax and inter-region transfers, the optimal policies satisfy

$$\rho_n = 1; \frac{T^{\star\prime}(y)}{1 - T^{\star\prime}(y)} = \sum_n A_n(y)B_n(y)C_n(y) \tag{D101}$$

when Assumptions 2 and 4 hold, and

$$\varepsilon_{m'n}^E(y) = \varepsilon_{mn}^E(y), \varepsilon_{m'}^{Cross}(y', y) = \varepsilon_m^{Cross}(y', y), \forall m', m, n \in \mathcal{J}, y', y \in \mathcal{Y}. \tag{D102}$$

To see the above result, we first prove that when the central government can freely choose central income tax and inter-region transfers, $\rho_n = 1$. By equation (D98) in Proposition 7, one has the following result when the central government can choose inter-region transfer freely:

$$\frac{N_n}{\sum_n N_n}(1 - \rho_n) = \int_{\underline{y}}^{\overline{y}} \sum_m A_m(y') \begin{Bmatrix} \left[1 + \widetilde{B}_m(y')C_m(y')\right]T^{\star\prime}(y) \\ -\widetilde{B}_m(y')C_m(y') \end{Bmatrix} \times \left[\Lambda(y')\sum_n \varepsilon_{mn}^E(y')\right] dy'. \tag{D103}$$

By equation (D97) in Proposition 7, the optimal tax rate satisfies:

$$0 = \left\{\sum_n A_n(y)\left[1+B_n(y)C_n(y)\right]T^{\star\prime}(y) - \sum_n A_n(y)B_n(y)C_n(y)\right\} \tag{D104}$$
$$+\int_{\underline{y}}^{\overline{y}}\left\{\sum_m A_m(y')\left[1+\widetilde{B}_m(y')C_m(y')\right]T^{\star\prime}(y) - \sum_m A_m(y')\widetilde{B}_m(y')C_m(y')\right\}\frac{\Lambda(y')}{\Lambda(y)}\frac{\varepsilon_m^{Cross}(y',y)}{\varepsilon_m^C(y')}dy'.$$

Proposition 7 has shown that there is a unique solution. Therefore, we need only to show that (D101) is a solution.

In the Proof of Proposition 6, we have shown that when $\rho_n = 1$ and for any $m', m \in \mathcal{J}$, $y', y \in \mathcal{Y}$, $\varepsilon_{m'}^{Cross}(y', y) = \varepsilon_m^{Cross}(y', y)$, (D104) is equivalent to:

$$\frac{T^{\star\prime}(y)}{1-T^{\star\prime}(y)} = \sum_n A_n(y)B_n(y)C_n(y).$$

Therefore, (D101) satisfies (D104). When for any $m', m, n \in \mathcal{J}$, $\varepsilon_{m'n}^E(y) = \varepsilon_{mn}^E(y)$, the left side of (D103) is equivalent to

$$\frac{N_n}{\sum_n N_n}(1-\rho_n) - \int_{\underline{y}}^{\overline{y}}\sum_m A_m(y')\left\{\left[1+\widetilde{B}_m(y')C_m(y')\right]T^{\star\prime}(y) - \widetilde{B}_m(y')C_m(y')\right\}\left[\Lambda(y')\sum_n \varepsilon_{mn}^E(y')\right]dy'$$
$$= \frac{N_n}{\sum_n N_n}(1-\rho_n) - \int_{\underline{y}}^{\overline{y}}\left\{\sum_m A_m(y')\left[1+\widetilde{B}_m(y')C_m(y')\right]T^{\star\prime}(y) - \sum_m A_m(y')\widetilde{B}_m(y')C_m(y')\right\}\left[\Lambda(y')\sum_n \varepsilon_n^E(y)(y')\right]dy',$$

where we save subscript m of ε_{mn}^E in the right side of the above equation because ε_{mn}^E is independent of m.

Under (D101) we have $\rho_n = 1$ and $\widetilde{B}_m(y') = B_m(y')$, which implies

$$0 = \frac{N_n}{\sum_n N_n}(1-\rho_n) - \int_{\underline{y}}^{\overline{y}}\left\{\sum_m A_m(y')\left[1+\widetilde{B}_m(y')C_m(y')\right]T^{\star\prime}(y) - \sum_m A_m(y')\widetilde{B}_m(y')C_m(y')\right\}\left[\Lambda(y')\sum_n \varepsilon_n^E(y)(y')\right]dy'.$$

Therefore, when (D102) holds, (D101) guarantees (D103). In conclusion, (D101) is the unique solution.■

E Supplements for Model Extension

E.1 Propositions about Integral Equations

We first provide the necessary definitions and then present the integral equation problem to be solved in Section E.1.2 (see equation (A.P)). Finally, we solve the integral equation problem.

E.1.1 Definition of Space and Operator

Definition 3 *The L^2-normed Vector Space. (i) Given that $\Theta \subseteq \mathbb{R}$ is a closed interval in $\mathbb{R}$ with a lower bound $\underline{\theta}$, $L_2(\Theta)$ is defined as the vector space consisting of all square-integrable functions on Θ (equivalence classes up to a Lebesgue measure zero set), which forms a normed vector space equipped with the 2-norm $\|\cdot\|_2$:*

$$L_2(\Theta) \equiv \left\{f | f:\Theta\to\mathbb{R} \text{ and } \int_\Theta |f(\theta)|^2 d\theta < \infty\right\}/\sim; \|f\|_2 \equiv \left[\int_\Theta |f(\theta)|^2 d\theta\right]^{1/2}. \tag{E1}$$

Here, the symbol "/~" indicates that this is a vector space composed of equivalence classes. Similarly, $L_2(\Theta\times\Theta)$ is defined as follows:

$$L_2(\Theta\times\Theta) \equiv \left\{h|h:\Theta\to\mathbb{R} \text{ and } \int_\Theta\int_\Theta \left|h(\theta,\theta')\right|^2 d\theta d\theta' < \infty\right\}/\tilde{}; \quad \text{(E2)}$$
$$\|h\|_2 \equiv \left[\int_\Theta\int_\Theta \left|h(\theta,\theta')\right|^2 d\theta d\theta'\right]^{1/2}.$$

(ii) The normed vector space $L_2^K(\Theta)$ is the vector space consisting of all K-dimensional vector functions that are square-integrable on Θ, which forms a normed vector space when equipped with the 2-norm $\|\bullet\|_2^K$. We define the K-dimensional vector function $\overrightarrow{f}:\Theta\to\mathbb{R}^K$ on Θ as follows:

$$\overrightarrow{f} = \left[\begin{array}{ccc} f_1 & \cdots & f_K \end{array}\right]^T : \Theta\to\mathbb{R}^K; \text{where } \forall\, k=1,\ldots,K,\ f_k:\Theta\to\mathbb{R} \text{ is a function on } \Theta.$$

where $L_2^K(\Theta)$ and $\|\bullet\|_2^K : L_2^K(\Theta)\to[0,+\infty)$ are defined as

$$L_2^K(\Theta) \equiv \left\{\overrightarrow{f} = \left[\begin{array}{ccc} f_1 & \cdots & f_K \end{array}\right]^T : \Theta\to\mathbb{R}^K | f_k\in L_2(\Theta), \forall\, k=1,\ldots,K\right\}/\tilde{}; \quad \text{(E3)}$$
$$\left\|\overrightarrow{f}\right\|_2^K \equiv \left[\sum_{k=1}^K \int_\Theta |f_k(\theta)|^2 d\theta\right]^{\frac{1}{2}}.$$

Similarly, the normed vector space $L_2^{K\times K}(\Theta\times\Theta)$ can be defined as the vector space consisting of $K\times K$-dimensional vector functions on $\Theta\times\Theta$, equipped with the 2-norm $\|\bullet\|_2^{K\times K}$. Define the $K\times K$-dimensional vector function $\overrightarrow{h}:\Theta\times\Theta\to\mathbb{R}^{K\times K}$ on $\Theta\times\Theta$ as follows:

$$\overrightarrow{h} = \left[\begin{array}{ccccccc} h_{11} & \cdots & h_{1K} & h_{21} & \cdots & h_{K\times K} \end{array}\right]^T : \Theta\times\Theta\to\mathbb{R}^{K\times K},$$

where for any $k,j=1,\ldots,K$, $h_{kj}:\Theta\times\Theta\to\mathbb{R}$ is a function on $\Theta\times\Theta$. Then define $L_2^{K\times K}(\Theta\times\Theta)$ and $\|\bullet\|_2^{K\times K}$: $L_2^{K\times K}(\Theta\times\Theta)\to[0,+\infty)$:

$$L_2^{K\times K}(\Theta\times\Theta) \equiv \left\{\begin{array}{c} \overrightarrow{h} = \left[\begin{array}{ccccccc} h_{11} & \cdots & h_{1K} & h_{21} & \cdots & h_{K\times K} \end{array}\right]^T : \\ \Theta\times\Theta\to\mathbb{R}^{K\times K} | h_{kj}\in L_2(\Theta\times\Theta), \forall\, k,j=1,\ldots,K \end{array}\right\}/\tilde{}; \quad \text{(E4)}$$
$$\left\|\overrightarrow{h}\right\|_2^{K\times K} \equiv \left[\sum_{k=1}^K\sum_{j=1}^K \int_\Theta\int_\Theta \left|h_{kj}(\theta,\theta')\right|^2 d\theta d\theta'\right]^{\frac{1}{2}}.$$

Note that $L_2(\Theta)$ and $L_2(\Theta\times\Theta)$ are complete spaces. Since the product space of a finite number of complete spaces remains complete, it follows that the metric spaces $L_2^K(\Theta)$ and $L_2^{K\times K}(\Theta\times\Theta)$ are complete spaces. To simplify the narrative, in the following appendix, we omit the equivalence class description in the text while retaining "/~" in the definition to indicate that equivalence classes are taken here.

Definition 4 *Bounded Mapping, Linear Operator, and Contraction Operator. Given that $\Theta\subseteq\mathbb{R}$ is a closed interval with a lower bound $\underline{\theta}$, and $L_2^K(\Theta)$ and $L_2^{K\times K}(\Theta\times\Theta)$ are complete normed vector spaces: (i) $T:L_2^K(\Theta)\to L_2^K(\Theta)$ is a bounded operator, if there exists a constant $M>0$ such that for any $\overrightarrow{f}\in L_2^K(\Theta)$, the following inequality holds:*

$$\left\|T\left(\overrightarrow{f}\right)\right\|_2^K \le M\left\|\overrightarrow{f}\right\|_2^K;$$

(ii) Similarly, the operator $\widetilde{T}: L_2^{K\times K}(\Theta\times\Theta)\to L_2^{K\times K}(\Theta\times\Theta)$ *is said to be bounded if there exists a constant* $M>0$ *such that for any* $\overrightarrow{h}\in L_2^{K\times K}(\Theta\times\Theta)$*, the following inequality holds:*

$$\left\|\widetilde{T}\left(\overrightarrow{h}\right)\right\|_2^{K\times K}\leq M\left\|\overrightarrow{h}\right\|_2^{K\times K};$$

(iii) $T: L_2^K(\Theta)\to L_2^K(\Theta)$ *is a linear operator, if for any* $a_1,a_2,a\in\mathbb{R}$ *and any* $\overrightarrow{f}_1,\overrightarrow{f}_2,\overrightarrow{f}\in L_2^K(\Theta)$*, one has:*

$$T\left(a_1\overrightarrow{f}_1+a_2\overrightarrow{f}_2\right)=a_1T\left(\overrightarrow{f}_1\right)+a_2T\left(\overrightarrow{f}_2\right);T\left(a\overrightarrow{f}\right)=aT\left(\overrightarrow{f}\right);$$

(iv) Similarly, $\widetilde{T}: L_2^{K\times K}(\Theta\times\Theta)\to L_2^{K\times K}(\Theta\times\Theta)$ *is a linear operator, if for any* $a_1,a_2,a\in\mathbb{R}$ *and any* $\overrightarrow{h}_1,\overrightarrow{h}_2,\overrightarrow{h}\in L_2^{K\times K}(\Theta\times\Theta)$*, one has:*

$$\widetilde{T}\left(a_1\overrightarrow{h}_1+a_2\overrightarrow{h}_2\right)=a_1\widetilde{T}\left(\overrightarrow{h}_1\right)+a_2\widetilde{T}\left(\overrightarrow{h}_2\right);\widetilde{T}\left(a\overrightarrow{h}\right)=a\widetilde{T}\left(\overrightarrow{h}\right);$$

(v) $T: L_2^K(\Theta)\to L_2^K(\Theta)$ *is a contraction operator, if there exist* $\delta\in[0,1)$ *such that for any* $\overrightarrow{f}_1,\overrightarrow{f}_2\in L_2^K(\Theta)$*, the following inequality holds:*

$$\left\|T\left(\overrightarrow{f}_1\right)-T\left(\overrightarrow{f}_2\right)\right\|_2^K\leq\delta\left\|\overrightarrow{f}_1-\overrightarrow{f}_2\right\|_2^K;$$

(vi) Similarly, $\widetilde{T}: L_2^{K\times K}(\Theta\times\Theta)\to L_2^{K\times K}(\Theta\times\Theta)$ *is a contraction operator, if there exist* $\delta\in[0,1)$ *such that for any* $\overrightarrow{h}_1,\overrightarrow{h}_2\in L_2^{K\times K}(\Theta\times\Theta)$*, the following inequality holds:*

$$\left\|\widetilde{T}\left(\overrightarrow{h}_1\right)-\widetilde{T}\left(\overrightarrow{h}_2\right)\right\|_2^{K\times K}\leq\delta\left\|\overrightarrow{h}_1-\overrightarrow{h}_2\right\|_2^{K\times K}.$$

E.1.2 The Problem to be Solved

Set vector function $\overrightarrow{\overline{\gamma}}\equiv\left[\begin{array}{ccc}\overline{\gamma}_{11} & \cdots & \overline{\gamma}_{KK}\end{array}\right]^T$, $\overrightarrow{\gamma}\equiv\left[\begin{array}{ccc}\gamma_{11} & \cdots & \gamma_{KK}\end{array}\right]^T\in L_2^{K\times K}(\Theta\times\Theta)$, where $\overline{\gamma}_{kj},\gamma_{kj}\in L_2(\Theta\times\Theta)$, $k,j=1,\ldots,K$. For any $\overrightarrow{f}\equiv\left[\begin{array}{ccc}f_1 & \cdots & f_K\end{array}\right]^T\in L_2^K(\Theta)$, define linear operator $T\left(\overrightarrow{f}\right)=\left[\begin{array}{ccc}T_1\left(\overrightarrow{f}\right) & \cdots & T_K\left(\overrightarrow{f}\right)\end{array}\right]^T$ as:

$$T_k\left(\overrightarrow{f}\right)(\theta)\equiv\sum_{j=1}^K\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta')\,d\theta'+\sum_{j=1}^K\int_{\Theta}\gamma_{kj}(\theta,\theta')f_j(\theta')\,d\theta',k=1,...,K;\tag{E5}$$

Under the above-defined linear operator, the equation

$$\overrightarrow{f}=T\left(\overrightarrow{f}\right)+\overrightarrow{x}\tag{A.P}$$

represent the following equation system.

$$f_k(\theta)=x_k(\theta)+\sum_{j=1}^K\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta')\,d\theta'+\sum_{j=1}^K\int_{\Theta}\gamma_{kj}(\theta,\theta')f_j(\theta')\,d\theta';k=1,...,K.$$

This is the system of equations to be solved below.

Lemma 9 *The operator $T\left(\overrightarrow{f}\right)$ is a well-defined bounded linear operator from $L_2^K(\Theta)$ to $L_2^K(\Theta)$.*

Proof. See the following proof. ■

E.1.3 Proof of Lemma 9

As linearity is obvious, to prove Lemma 9 , we only need to prove that for any $k = 1, ..., K$, $T_k\left(\overrightarrow{f}\right)$ is a well-defined (i.e., bounded) operator from $L_2^K(\Theta)$ to $L_2(\Theta)$. That is to say $\left\|T_k\left(\overrightarrow{f}\right)\right\|_2 < \infty$.

Step 1. Using the properties of absolute values, we have the following inequality

$$\begin{aligned}
\max\left\{\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta'), -\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta')\right\} &\leq \left|\overline{\gamma}_{kj}(\theta,\theta')\right||f_j(\theta')|, \\
\max\left\{\gamma_{kj}(\theta,\theta')f_j(\theta'), -\gamma_{kj}(\theta,\theta')f_j(\theta')\right\} &\leq \left|\gamma_{kj}(\theta,\theta')\right||f_j(\theta')|.
\end{aligned}$$

We can derive the following inequality between integrations by the above inequality:

$$\begin{aligned}
\left|\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta')\,d\theta'\right| &\leq \int_{\underline{\theta}}^{\theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta', \\
\left|\int_{\Theta}\gamma_{kj}(\theta,\theta')f_j(\theta')\,d\theta'\right| &\leq \int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta'.
\end{aligned}$$

Notice that $\int_{\underline{\theta}}^{\theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta' \leq \int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta$, one has the following inequalities by the above inequalities:

$$\begin{aligned}
\left|\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta')\,d\theta'\right| &\leq \int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta, \\
\left|\int_{\Theta}\gamma_{kj}(\theta,\theta')f_j(\theta')\,d\theta'\right| &\leq \int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta'.
\end{aligned} \tag{E6}$$

Step 2. Since $\overline{\gamma}_{kj}, \gamma_{kj}: \Theta\times\Theta \to \mathbb{R}$ is a square-integrable function, we have the following inequality:

$$\int_{\Theta}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta' d\theta < \infty, \int_{\Theta}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' d\theta < \infty.$$

Based on the definition of integrability, it can be inferred that $\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta' < \infty$ and $\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' < \infty$ is true almost everywhere in Θ. Using the Holder inequality, we have

$$\begin{aligned}
\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta &\leq \left[\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \times \left[\int_{\Theta}|f_j(\theta')|^2 d\theta'\right]^{\frac{1}{2}}, \\
\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right||f_j(\theta')|\,d\theta' &\leq \left[\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \times \left[\int_{\Theta}|f_j(\theta')|^2 d\theta'\right]^{\frac{1}{2}}.
\end{aligned} \tag{E7}$$

Combination of (E6) and (E7) suggests that the following inequality hold almost everywhere in Θ:

$$\begin{aligned}
\left|\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{kj}(\theta,\theta')f_j(\theta')\,d\theta'\right|^2 &\leq \left[\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right] \times \left[\int_{\Theta}|f_j(\theta')|^2 d\theta'\right], \\
\left|\int_{\Theta}\gamma_{kj}(\theta,\theta')f_j(\theta')\,d\theta'\right|^2 &\leq \left[\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right] \times \left[\int_{\Theta}|f_j(\theta')|^2 d\theta'\right].
\end{aligned} \tag{E8}$$

Notice that $\overrightarrow{f} \in L_2^K(\Theta)$, one has $\int_\Theta \left|f_j(\theta')\right|^2 d\theta' < \infty$. Therefore the following inequality holds almost everywhere in Θ:

$$\begin{aligned}
\left|\int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta,\theta') f_j(\theta') d\theta'\right| &\leq \left[\int_\Theta \left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \times \left[\int_\Theta \left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} < \infty, \\
\left|\int_{\Theta} \gamma_{kj}(\theta,\theta') f_j(\theta') d\theta'\right| &\leq \left[\int_\Theta \left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \times \left[\int_\Theta \left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} < \infty
\end{aligned}$$

When the above inequality holds, one can take integration of both sides of (E8), which gives:

$$\begin{aligned}
\int_\Theta \left[\int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta,\theta') f_j(\theta') d\theta'\right]^2 d\theta &\leq \int_\Theta \left[\int_\Theta \left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta' \times \int_\Theta \left|f_j(\theta')\right|^2 d\theta'\right] d\theta, \\
\int_\Theta \left[\int_{\Theta} \gamma_{kj}(\theta,\theta') f_j(\theta') d\theta'\right]^2 d\theta &\leq \int_\Theta \left[\int_\Theta \left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' \times \int_\Theta \left|f_j(\theta')\right|^2 d\theta'\right] d\theta.
\end{aligned}$$

Using the Fubini theorem, the above inequality can be transformed into:

$$\begin{aligned}
\int_\Theta \left[\int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta,\theta') f_j(\theta') d\theta'\right]^2 d\theta &\leq \int_\Theta \int_\Theta \left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta' d\theta \times \int_\Theta \left|f_j(\theta')\right|^2 d\theta' < \infty, \\
\int_\Theta \left[\int_{\Theta} \gamma_{kj}(\theta,\theta') f_j(\theta') d\theta'\right]^2 d\theta &\leq \int_\Theta \int_\Theta \left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' d\theta \times \int_\Theta \left|f_j(\theta')\right|^2 d\theta' < \infty.
\end{aligned} \tag{E9}$$

The above inequality suggests that $\int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta,\theta') f_j(\theta') d\theta'$ and $\int_\Theta \gamma_{kj}(\theta,\theta') f_j(\theta') d\theta'$ are square integrable functions in Θ.

Step 3. Rememeber that we have defined the following operator (see equation (E5)):

$$T_k\left(\overrightarrow{f}\right)(\theta) \equiv \sum_{j=1}^K \int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta,\theta') f_j(\theta') d\theta' + \sum_{j=1}^K \int_\Theta \gamma_{kj}(\theta,\theta') f_j(\theta') d\theta'.$$

For any $k = 1,..,K$, using the triangle inequality of absolute values, for any $\overrightarrow{f} \in L_2^K(\Theta)$ we have the following inequality:

$$\begin{aligned}
\left|T_k\left(\overrightarrow{f}\right)(\theta)\right| &= \left|\sum_{j=1}^K \int_{\underline{\theta}}^{\theta} \overline{\gamma}_{kj}(\theta,\theta') f_j(\theta') d\theta' + \sum_{j=1}^K \int_\Theta \gamma_{kj}(\theta,\theta') f_j(\theta') d\theta'\right| \\
&\leq \sum_{j=1}^K \int_{\underline{\theta}}^{\theta} \left|\overline{\gamma}_{kj}(\theta,\theta')\right| \left|f_j(\theta')\right| d\theta' + \sum_{j=1}^K \int_\Theta \left|_{kj}(\theta,\theta')\right| \left|f_j(\theta')\right| d\theta'.
\end{aligned} \tag{E10}$$

Using (E6) and (E7), we can rewrite (E10) as:

$$\begin{aligned}
\left|T_k\left(\overrightarrow{f}\right)(\theta)\right| \leq & \sum_{j=1}^K \left[\int_\Theta \left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \left[\int_\Theta \left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \\
& + \sum_{j=1}^K \left[\int_\Theta \left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \left[\int_\Theta \left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}.
\end{aligned}$$

Using Holder's inequality, the above inequality can be written as

$$\begin{aligned}\left|T_k\left(\overrightarrow{f}\right)(\theta)\right| &\leq \left[\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}\left[\sum_{j=1}^{K}\int_{\Theta}\left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \\ &\quad+\left[\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}\left[\sum_{j=1}^{K}\int_{\Theta}\left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}.\end{aligned}$$

Rearrange the above inequality:

$$\begin{aligned}\left\|T_k\left(\overrightarrow{f}\right)\right\|_2 &\leq \left\{\int_{\Theta}\left(\frac{\left[\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}}{+\left[\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}}\right)^2 d\theta\right\}^{\frac{1}{2}}\left[\sum_{j=1}^{K}\int_{\Theta}\left|f_j(\theta')\right|^2 d\theta'\right]^{\frac{1}{2}} \\ &= \left\{\int_{\Theta}\left(\frac{\left[\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}}{+\left[\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}}\right)^2 d\theta\right\}^{\frac{1}{2}}\left\|\overrightarrow{f}\right\|_2^K.\end{aligned}$$

Notice that $\left\|T\left(\overrightarrow{f}\right)\right\|_2^K=\left(\sum_{k=1}^{K}\left\|T_k\left(\overrightarrow{f}\right)\right\|_2^2\right)^{\frac{1}{2}}$. We have

$$\left\|T_k\left(\overrightarrow{f}\right)\right\|_2\leq\left\|T\left(\overrightarrow{f}\right)\right\|_2^K\leq\left\{\sum_{k=1}^{K}\int_{\Theta}\left(\frac{\left[\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}}{+\left[\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'\right]^{\frac{1}{2}}}\right)^2 d\theta\right\}^{\frac{1}{2}}\left\|\overrightarrow{f}\right\|_2^K. \tag{E11}$$

On the other hand, for any $k=1,..,K$, we have

$$\begin{aligned}&\sum_{k=1}^{K}\left(\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'}+\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'}\right)^2 \qquad \text{(E12)}\\ =\;&\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'+\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' \\ &+2\sum_{k=1}^{K}\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'}\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'}.\end{aligned}$$

Using Holder inequality for vector, we have

$$\begin{aligned}&\sum_{k=1}^{K}\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta'}\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta'} \\ \leq\;&\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta' d\theta}\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' d\theta}\end{aligned}$$

Combination of the above inequality and equation (E12) gives

$$\begin{aligned}
&\sum_{k=1}^{K}\left(\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'}+\sqrt{\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'}\right)^{2} \qquad \text{(E13)}\\
\leq\ &\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'+\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\left|\gamma_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'\\
&+2\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{\gamma}_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'd\theta}\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\gamma_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'd\theta}\\
=\ &\left(\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{\gamma}_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'd\theta}+\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\gamma_{kj}\left(\theta,\theta'\right)\right|^{2}d\theta'd\theta}\right)^{2}\\
=\ &\left(\left\|\overrightarrow{\overline{\gamma}}\right\|_{2}^{K\times K}+\|\overrightarrow{\gamma}\|_{2}^{K\times K}\right)^{2}.
\end{aligned}$$

Combination of (E11) and (E13) gives

$$\left\|T_{k}\left(\overrightarrow{f}\right)\right\|_{2}\leq\left\|T\left(\overrightarrow{f}\right)\right\|_{2}^{K}\leq\left(\left\|\overrightarrow{\overline{\gamma}}\right\|_{2}^{K\times K}+\|\overrightarrow{\gamma}\|_{2}^{K\times K}\right)\times\left\|\overrightarrow{f}\right\|_{2}^{K}. \qquad \text{(E14)}$$

Notice that $\overrightarrow{\overline{\gamma}},\overrightarrow{\gamma}\in L_{2}^{K\times K}\left(\Theta\times\Theta\right)$. One can see that $T\left(\overrightarrow{f}\right)$ and $T_{k}\left(\overrightarrow{f}\right)$ are well-defined bounded operators. The linear property of the above operator is obvious.■

E.1.4 Existence and Uniqueness of the Solution

The following Theorem 5 and Lemma 10 are used to prove the existence and uniqueness of the solution to the integral equations.

Theorem 5 *Banach Contraction Mapping. (i) if (a)* $T:L_{2}^{K}\left(\Theta\right)\rightarrow L_{2}^{K}\left(\Theta\right)$ *is linear bounded contraction mapping, and (b)* $\overrightarrow{x}\in L_{2}^{K}\left(\Theta\right)$*, we have the following results:*

(a) There is a unique solution $\overrightarrow{f}_{*}\in L_{2}^{K}\left(\Theta\right)$ *to the following equation:*

$$\overrightarrow{f}=T\left(\overrightarrow{f}\right)+\overrightarrow{x}$$

(b) For any $\overrightarrow{x}_{0}\in L_{2}^{K}\left(\Theta\right)$*, one can construct sequence* $\left\{\overrightarrow{f}_{n}\right\}_{n=0}^{+\infty}$ *as below, such that for any* $n\geq0$ *there is* $\overrightarrow{f}_{n}\in L_{2}^{K}\left(\Theta\right)$ *satisfying* $\lim_{n\rightarrow+\infty}\overrightarrow{f}_{n}=\overrightarrow{f}_{*}$*:*

$$\overrightarrow{f}_{n+1}=T\left(\overrightarrow{f}_{n}\right)+\overrightarrow{x},n\geq0,\overrightarrow{f}_{0}=\overrightarrow{x}_{0}$$

(ii) If (a) $\widetilde{T}:L_{2}^{K\times K}\left(\Theta\times\Theta\right)\rightarrow L_{2}^{K\times K}\left(\Theta\times\Theta\right)$*is linear bounded contraction mapping, and (b)* $\overrightarrow{y}\in L_{2}^{K\times K}\left(\Theta\times\Theta\right)$*, we have the following results:*

(a) There is a unique solution $\overrightarrow{h}_{*}\in L_{2}^{K\times K}\left(\Theta\times\Theta\right)$ *to the following equation:*

$$\overrightarrow{h}=\widetilde{T}\left(\overrightarrow{h}\right)+\overrightarrow{y}$$

(b) If for any $\overrightarrow{y}_0 \in L_2^{K\times K}(\Theta\times\Theta)$, *one can construct sequence* $\left\{\overrightarrow{h}_n\right\}_{n=0}^{+\infty}$ *as below, such that for any* $n\geq 0$, *there is* $\overrightarrow{h}_n \in L_2^{K\times K}(\Theta\times\Theta)$ *satisfying* $\lim_{n\to+\infty}\overrightarrow{h}_n = \overrightarrow{h}_*$:

$$\overrightarrow{h}_{n+1} = \widetilde{T}(h_n) + \overrightarrow{y}, n\geq 0, h_0 = \overrightarrow{y}_0$$

Proof. For the above theorem, see Theorem 6.2.1 and 6.2.2 in CH6 of Zemyan (2012). ■

Sachs et al. (2020) also use the above theorem to prove the existence and uniqueness of the solution to the integral equation. By the above theorem, we have the following Lemma 10.

Lemma 10 *If the following contraction condition holds:*

$$\left\|\overrightarrow{\overline{\gamma}}\right\|_2^{K\times K} + \left\|\overrightarrow{\gamma}\right\|_2^{K\times K} < 1; \tag{E15}$$

i.e.,

$$\sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_\Theta\int_\Theta\left|\overline{\gamma}_{kj}(\theta,\theta')\right|^2 d\theta' d\theta} + \sqrt{\sum_{k=1}^{K}\sum_{j=1}^{K}\int_\Theta\int_\Theta\left|\gamma_{kj}(\theta,\theta')\right|^2 d\theta' d\theta} < 1,$$

then for any $\overrightarrow{x}\in L_2^K(\Theta)$ *there is a unique solution* $\overrightarrow{f}_*\in L_2^K(\Theta)$ *to problem (A.P), i.e.,* $\overrightarrow{f}_* = T\left(\overrightarrow{f}_*\right) + \overrightarrow{x}$. *The uniqueness mentioned above means that, up to a set of measure zero, there exists a unique* $\overrightarrow{f}_*\in L_2^K(\Theta)$ *such that* $\overrightarrow{f}_* = T\left(\overrightarrow{f}_*\right) + \overrightarrow{x}$. *In particular, if* $\overrightarrow{f}$ *is continuous, the solution* $\overrightarrow{f}_*$ *exist and is unique.*

Proof. See the following proof. ■

E.1.5 Proof of Lemma 10

First, to prove that problem (A.P) is a well-defined problem, one only need to show that for any $\overrightarrow{f}\in L_2^K(\Theta)$, there is $T\left(\overrightarrow{f}\right)+\overrightarrow{x}\in L_2^K(\Theta)$. Notice that T is linear and bounded mapping, one has $T\left(\overrightarrow{f}\right)\in L_2^K(\Theta)$. Moreover, since $\overrightarrow{x}\in L_2^K(\Theta)$, the Minkowski inequality implies $T\left(\overrightarrow{f}\right)+\overrightarrow{x}\in L_2^K(\Theta)$3002

Second, we now prove that T is a contractive mapping. Use inequality (E14) and condition E15, one can see that there exist $\delta = \left\|\overrightarrow{\overline{\gamma}}\right\|_2^{K\times K} + \left\|\overrightarrow{\gamma}\right\|_2^{K\times K}\in[0,1)$, such that for any $\overrightarrow{f}\in L_2^K(\Theta)$ there is the following inequality:

$$\left\|T\left(\overrightarrow{f}\right)\right\|_2^K \leq \delta\left\|\overrightarrow{f}\right\|_2^K < \left\|\overrightarrow{f}\right\|_2^K$$

It can be seen that T is a contractive mapping. Then by Theorem 5, there exist a unqiue solution to problem (A.P). Moreover, for any $\overrightarrow{x}_0\in L_2^K(\Theta)$, one can construct sequence $\left\{\overrightarrow{f}_n\right\}_{n=0}^{+\infty}$ as below, so that for any $n\geq 0$, there is $\overrightarrow{f}_n\in L_2^K(\Theta)$ satisfying $\lim_{n\to+\infty}\overrightarrow{f}_n = \overrightarrow{f}_*$:

$$\overrightarrow{f}_{n+1} = T\left(\overrightarrow{f}_n\right) + \overrightarrow{x}, n\geq 0, \overrightarrow{f}_0 = \overrightarrow{x}_0.$$

In conclusion, we have Lemma 10.■

E.1.6 Resolvent of Problem

Considering the possibility of iterative convergence in solving the problem, it can be conjectured that there exists a vector function $\overrightarrow{\overline{\Gamma}} \equiv \left[\begin{array}{ccc}\overline{\Gamma}_{11} & \cdots & \overline{\Gamma}_{KK}\end{array}\right]^T$, $\overrightarrow{\Gamma} \equiv \left[\begin{array}{ccc}\Gamma_{11} & \cdots & \Gamma_{KK}\end{array}\right]^T \in L_2^{K\times K}(\Theta\times\Theta)$, such that the

solution $\overrightarrow{f}_* = \left[\begin{array}{ccc} f_1^* & \cdots & f_K^* \end{array} \right]^T$ satisfies the following representation form:

$$f_k^*(\theta) = x_k(\theta) + \sum_j \int_{\underline{\theta}}^{\theta} \overline{\Gamma}_{kj}(\theta, \theta') x_j(\theta') d\theta' + \sum_j \int_{\Theta} \Gamma_{kj}(\theta, \theta') x_j(\theta') d\theta', k = 1, .., K$$

Substituting it into the problem to be solved, i.e., problem (A.P), it can be found that if $\overrightarrow{\overline{\Gamma}}$ and $\overrightarrow{\Gamma}$ is the solution to problem (A.P), for any $k, j = 1, .., K$, $\overrightarrow{\overline{\Gamma}}$ and $\overrightarrow{\Gamma}$ should satisfy the following equations:

$$\begin{aligned} \overline{\Gamma}_{kj}(\theta, \theta') &= \overline{\gamma}_{kj}(\theta, \theta') + \sum_{m=1}^{K} \int_{\theta'}^{\theta} \overline{\gamma}_{km}(\theta, \theta'') \overline{\Gamma}_{mj}(\theta'', \theta') d\theta''; \\ \Gamma_{kj}(\theta, \theta') &= \left[\gamma_{kj}(\theta, \theta') + \sum_{m=1}^{K} \int_{\theta'' \geq \theta'} \gamma_{km}(\theta, \theta'') \overline{\Gamma}_{mj}(\theta'', \theta') d\theta'' \right] \\ &\quad + \sum_{m=1}^{K} \int_{\underline{\theta}}^{\theta} \overline{\gamma}_{km}(\theta, \theta'') \Gamma_{mj}(\theta'', \theta') d\theta'' + \sum_{m=1}^{K} \int_{\Theta} \gamma_{km}(\theta, \theta'') \Gamma_{mj}(\theta'', \theta') d\theta''. \end{aligned} \tag{E16}$$

To this end, we refer to the existence and uniqueness of the solution ($\overrightarrow{\overline{\Gamma}}$ and $\overrightarrow{\Gamma}$) to the above problem as the resolvent problem of (A.P).

We have the following result regarding the resolvent problem of (A.P).

Lemma 11 *When condition E15 holds, there exist a unique solution* $\overrightarrow{\overline{\Gamma}} \in L_2^{K \times K}(\Theta \times \Theta)$ *and* $\overrightarrow{\Gamma} \in L_2^{K \times K}(\Theta \times \Theta)$ *such that there is a solution to the resolvent problem (E16). Moreover the resolvent* $\left\{\overline{\Gamma}_{kj}(\theta, \theta')\right\}$ *and* $\left\{\Gamma_{jk}(\theta, \theta')\right\}$ *can be constructed by the following sequence. First construct a convergent sequence* $\left\{\overline{\Gamma}_{kj}^n(\theta, \theta')\right\}_{n=0}^{+\infty}$ *of the resolvent* $\left\{\overline{\Gamma}_{kj}(\theta, \theta')\right\}$*:*

$$\begin{aligned} \overline{\Gamma}_{kj}^0(\theta, \theta') &\equiv \overline{\gamma}_{kj}(\theta, \theta'), \\ \overline{\Gamma}_{kj}^{n+1}(\theta, \theta') &\equiv \overline{\gamma}_{kj}(\theta, \theta') + \sum_{m=1}^{K} \int_{\theta'}^{\theta} \overline{\gamma}_{km}(\theta, \theta'') \overline{\Gamma}_{mj}^n(\theta'', \theta') d\theta'', \forall n \geq 1. \end{aligned}$$

One has $\Gamma_{kj}(\theta, \theta') = \lim_{n \to \infty} \Gamma_{kj}^n(\theta, \theta')$*. Second, construct a convergent sequence* $\left\{\Gamma_{kj}^n(\theta, \theta')\right\}_{n=0}^{+\infty}$ *of the resolvent* $\left\{\Gamma_{kj}(\theta, \theta')\right\}$*:*

$$\begin{aligned} \Gamma_{kj}^0(\theta, \theta') &\equiv \gamma_{kj}(\theta, \theta') + \sum_{m=1}^{K} \int_{\theta'' \geq \theta'} \gamma_{km}(\theta, \theta'') \overline{\Gamma}_{mj}(\theta'', \theta') d\theta''; \\ \Gamma_{kj}^{n+1}(\theta, \theta') &\equiv \left[\gamma_{kj}(\theta, \theta') + \sum_{m=1}^{K} \int_{\theta'' \geq \theta'} \gamma_{km}(\theta, \theta'') \overline{\Gamma}_{mj}(\theta'', \theta') d\theta'' \right] \\ &\quad + \sum_{m=1}^{K} \int_{\underline{\theta}}^{\theta} \overline{\gamma}_{km}(\theta, \theta'') \Gamma_{mj}^n(\theta'', \theta') d\theta'' + \sum_{m=1}^{K} \int_{\Theta} \gamma_{km}(\theta, \theta'') \Gamma_{mj}^n(\theta'', \theta') d\theta''. \end{aligned}$$

One has $\overline{\Gamma}_{kj}(\theta, \theta') = \lim_{n \to \infty}^n \overline{\Gamma}_{kj}^n(\theta, \theta'), \forall n \geq 1$.

Proof. See the following proof. ■

E.1.7 Proof of Lemma 11

We first prove that there exist a unique $\overrightarrow{\overline{\Gamma}} \in L_2^{K\times K}(\Theta\times\Theta)$ such that

$$\overline{h}_{kj}(\theta,\theta') \equiv \overline{\gamma}_{kj}(\theta,\theta') + \sum_{m=1}^{K}\int_{\theta'}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\overline{h}_{mj}(\theta'',\theta')\,d\theta''; \tag{E17}$$

Then, use the property that $\overrightarrow{\overline{\Gamma}} \in L_2^{K\times K}(\Theta\times\Theta)$, we prove that there exist a unique $\overrightarrow{\Gamma} \in L_2^{K\times K}(\Theta\times\Theta)$ such that the following equation holds:

$$\begin{aligned} h_{kj}(\theta,\theta') \equiv{}& \left[\gamma_{kj}(\theta,\theta') + \sum_{m=1}^{K}\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta''\right] \\ &+ \sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,h_{mj}(\theta'',\theta')\,d\theta'' + \sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\,h_{mj}(\theta'',\theta')\,d\theta''. \end{aligned} \tag{E18}$$

Step 1. Prove that there exists a unique solution $\overrightarrow{\overline{\Gamma}} \in L_2^{K\times K}(\Theta\times\Theta)$ to the problem (E17).

First, for any $\overrightarrow{\overline{h}} \in L_2^{K\times K}(\Theta\times\Theta)$, define linear operator $\widetilde{T}\left(\overrightarrow{\overline{h}}\right) = \left[\begin{array}{ccc}\widetilde{T}_{11}\left(\overrightarrow{\overline{h}}\right) & \cdots & \widetilde{T}_{K\times K}\left(\overrightarrow{\overline{h}}\right)\end{array}\right]^T : L_2^{K\times K}(\Theta\times\Theta) \to L_2^{K\times K}(\Theta\times\Theta)$ such that for any $k,j = 1,..,K$:

$$\widetilde{T}_{kj}\left(\overrightarrow{\overline{h}}\right)(\theta,\theta') = \sum_{m=1}^{K}\int_{\theta'}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\overline{h}_{mj}(\theta'',\theta')\,d\theta'';$$

Then, we only need to prove that there exists a unique solution to the following equation:

$$\overrightarrow{\overline{h}} = \widetilde{T}\left(\overrightarrow{\overline{h}}\right) + \overrightarrow{\overline{\gamma}}$$

where $\overrightarrow{\overline{\gamma}} \in L_2^{K\times K}(\Theta\times\Theta)$.

Now, we only need to prove that $\widetilde{T}$ is a contraction mapping. Similar to the process of obtaining (E12) in the proof of Lemma 9 by utilizing (E6) and (E7), we can derive the following inequality:

$$\begin{aligned} &\int_{\Theta}\int_{\Theta}\left|\widetilde{T}_{kj}\left(\overrightarrow{\overline{h}}\right)(\theta,\theta')\right|^2 d\theta d\theta' \\ \leq{}& \int_{\Theta}\int_{\Theta}\left(\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^2 d\theta''\right]^{\frac{1}{2}}\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^2 d\theta''\right]^{\frac{1}{2}}\right)^2 d\theta d\theta' \\ ={}& \int_{\Theta}\int_{\Theta}\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^2 d\theta''\right]\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^2 d\theta''\right] d\theta d\theta' \end{aligned}$$

Summing over k and j yields:

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\widetilde{T}_{kj}\left(\overrightarrow{h}\right)(\theta,\theta')\right|^{2}d\theta d\theta' \\
\leq\ &\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^{2}d\theta''\right]\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^{2}d\theta''\right]d\theta d\theta'
\end{aligned}$$

With Fubini's theorem, the above inequality can be written in the following form:

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\widetilde{T}_{kj}\left(\overrightarrow{h}\right)(\theta,\theta')\right|^{2}d\theta d\theta' \\
\leq\ &\left\{\sum_{k=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^{2}d\theta''d\theta d\theta''\right\}\times\left\{\sum_{j=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^{2}d\theta''d\theta'\right\}
\end{aligned}$$

Using notations defined before, the above inequality can be expressed in the following form:

$$\left\|\widetilde{T}\left(\overrightarrow{h}\right)\right\|_{2}^{K\times K}\leq\left\|\overrightarrow{\gamma}\right\|_{2}^{K\times K}\left\|\overrightarrow{h}\right\|_{2}^{K\times K}<\left\|\overrightarrow{h}\right\|_{2}^{K\times K}$$

By Theorem 5, (a) there exist a unique solution in $L_{2}^{K\times K}(\Theta\times\Theta)$ to problem (E17). Set $\overrightarrow{\Gamma}$ as the solution; moreover (b) we can construct sequence $\left\{\overrightarrow{\Gamma}_{n}\right\}_{n=0}^{+\infty}$ such that $\lim_{n\to+\infty}\overrightarrow{\Gamma}_{n}=\overrightarrow{\Gamma}$:

$$\overrightarrow{\Gamma}_{n+1}=\widetilde{T}\left(\overrightarrow{\Gamma}_{n}\right)+\overrightarrow{\gamma},n\geq0;\overrightarrow{\Gamma}_{0}=\overrightarrow{\gamma}.$$

Specifically,

$$\begin{aligned}
\overline{\Gamma}_{kj}^{0}(\theta,\theta') &\equiv \overline{\gamma}_{kj}(\theta,\theta');\\
\overline{\Gamma}_{kj}^{n+1}(\theta,\theta') &\equiv \overline{\gamma}_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta'}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}^{n}(\theta'',\theta')\,d\theta'',n\geq0.
\end{aligned}$$

Step 2. We now prove that there exists a unque solution $\overrightarrow{\Gamma}\in L_{2}^{K\times K}(\Theta\times\Theta)$ to problem (E18)

In Step 1, we have proved that there exists a unique solution $\overrightarrow{\Gamma}\in L_{2}^{K\times K}(\Theta\times\Theta)$ to problem (E17). Now set $\overrightarrow{y}\equiv\left[\begin{array}{ccc}y_{11} & \cdots & y_{K\times K}\end{array}\right]^{T}$ such that for any $k,j=1,..,K$:

$$y_{kj}(\theta,\theta')\equiv\gamma_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta''$$

Similar to the process of obtaining (E8) in the proof of Lemma 9 by utilizing (E6) and (E7), we can use the fact that $\overrightarrow{\gamma}$, $\overrightarrow{\Gamma}\in L_{2}^{K\times K}(\Theta\times\Theta)$ to prove that $\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta''\in L_{2}^{K\times K}(\Theta\times\Theta)$.

Then use Minkowski inequaltiy, one has $y_{kj}\in L_{2}(\Theta\times\Theta)$. Therefore $\overrightarrow{y}\in L_{2}^{K\times K}(\Theta\times\Theta)$. Then, set an

operator $\overrightarrow{\overrightarrow{T}}\left(\overrightarrow{h}\right)=\left[\begin{array}{ccc}\overrightarrow{\overrightarrow{T}}_{11}\left(\overrightarrow{h}\right) & \cdots & \overrightarrow{\overrightarrow{T}}_{K\times K}\left(\overrightarrow{h}\right)\end{array}\right]^{T}: L_2^{K\times K}(\Theta\times\Theta)\to L_2^{K\times K}(\Theta\times\Theta)$ as below:

$$\overrightarrow{\overrightarrow{T}}_{kj}\left(\overrightarrow{h}\right)(\theta,\theta')=\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,h_{mj}(\theta'',\theta')\,d\theta''+\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\,h_{mj}(\theta'',\theta')\,d\theta'', k,j=1,..,K.$$

Then the original problem becomes proving that there exists a unique solution to the following equation:

$$\overrightarrow{h}=\overrightarrow{\overrightarrow{T}}\left(\overrightarrow{h}\right)+\overrightarrow{y},$$

where $\overrightarrow{y}\in L_2^{K\times K}(\Theta\times\Theta)$.

Now we only need to prove that $\overrightarrow{\overrightarrow{T}}$ is a contractive mapping. Similar to the process of obtaining (E12) in the proof of Lemma 9 by utilizing (E6) and (E7), we can derive the following inequality:

$$\begin{aligned}&\left|\overrightarrow{\overrightarrow{T}}\left(\overrightarrow{h}\right)(\theta,\theta')\right|\\ \leq\ &\sum_{m=1}^{K}\left[\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^2d\theta''\right]^{\frac{1}{2}}\left[\int_{\Theta}\left|h_{mj}(\theta'',\theta')\right|^2d\theta''\right]^{\frac{1}{2}}\\ &+\sum_{m=1}^{K}\left[\int_{\Theta}\left|\gamma_{km}(\theta,\theta'')\right|^2d\theta''\right]^{\frac{1}{2}}\left[\int_{\Theta}\left|h_{mj}(\theta'',\theta')\right|^2d\theta''\right]^{\frac{1}{2}}\\ \leq\ &\left\{\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^2d\theta''\right]^{\frac{1}{2}}+\left[\sum_{m=1}^{K}\int_{\Theta}\left|\gamma_{km}(\theta,\theta'')\right|^2d\theta''\right]^{\frac{1}{2}}\right\}\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^2d\theta''\right]^{\frac{1}{2}}.\end{aligned}$$

Similar to the process of obtaining (E11) and (E13) in the proof of Lemma 9, we can derive the following inequality by the above inequality:

$$\begin{aligned}&\sum_{k=1}^{K}\sum_{j=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overrightarrow{\overrightarrow{T}}\left(\overrightarrow{h}\right)(\theta,\theta')\right|^2d\theta d\theta'\\ \leq\ &\int_{\Theta}\int_{\Theta}\left\{\sum_{k=1}^{K}\left[\sum_{m=1}^{K}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^2d\theta''\right]^{\frac{1}{2}}+\left[\sum_{m=1}^{K}\int_{\Theta}\left|\gamma_{km}(\theta,\theta'')\right|^2d\theta''\right]^{\frac{1}{2}}\right\}^2d\theta'\\ &\times\left[\sum_{j=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^2d\theta''d\theta'\right]\\ \leq\ &\left\{\left[\sum_{k=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{\gamma}_{km}(\theta,\theta'')\right|^2d\theta''d\theta'\right]^{\frac{1}{2}}+\left[\sum_{k=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\int_{\Theta}\left|\gamma_{km}(\theta,\theta'')\right|^2d\theta''d\theta'\right]^{\frac{1}{2}}\right\}^2\\ &\times\left[\sum_{j=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\int_{\Theta}\left|\overline{h}_{mj}(\theta'',\theta')\right|^2d\theta''d\theta'\right].\end{aligned}$$

Using notations defined before, the above inequality can be expressed in the following form:

$$\left\|\overrightarrow{\overrightarrow{T}}\left(\overrightarrow{h}\right)\right\|_2^{K\times K}\leq\left(\left\|\overrightarrow{\overrightarrow{\gamma}}\right\|_2^{K\times K}+\left\|\overrightarrow{\overrightarrow{\gamma}}\right\|_2^{K\times K}\right)\left\|\overrightarrow{h}\right\|_2^{K\times K}<\left\|\overrightarrow{h}\right\|_2^{K\times K}$$

It can be seen that $\widetilde{\overrightarrow{T}}$ is a linear and bounded contractive mapping defined on $L_2^{K\times K}(\Theta\times\Theta)$. By Theorem 5, since $\overrightarrow{y}\in L_2^{K\times K}(\Theta\times\Theta)$, we have (a) there exist a unique solution in $L_2^{K\times K}(\Theta\times\Theta)$ to problem (E18). Set the solution as $\overrightarrow{\Gamma}$. Moreover, (b) one can construct sequence $\left\{\overrightarrow{\Gamma}_n\right\}_{n=0}^{+\infty}$ such that $\lim_{n\to+\infty}\overrightarrow{\Gamma}_n=\overrightarrow{\Gamma}$:

$$\overrightarrow{\Gamma}_0=\overrightarrow{y},\overrightarrow{\Gamma}_{n+1}=\widetilde{T}\left(\overrightarrow{\Gamma}_n\right)+\overrightarrow{y},n\geq 0.$$

i.e.,

$$\begin{aligned}\Gamma_{kj}^0(\theta,\theta') &\equiv \gamma_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta'',\\ \Gamma_{kj}^{n+1}(\theta,\theta') &\equiv \left[\gamma_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta''\right]\\ &\quad+\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\Gamma_{mj}^n(\theta'',\theta')\,d\theta''+\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\,\Gamma_{mj}^n(\theta'',\theta')\,d\theta'',\end{aligned}$$

where $n\geq 0$. The proof is completed.■

E.1.8 Solution of the Problem

Lemma 12 *The fixed point of operator T (see equation (E5)) $\overrightarrow{f}_*=\begin{bmatrix}f_1^* & \ldots & f_K^*\end{bmatrix}$ satisfies:*

$$f_k^*(\theta)=x_k(\theta)+\sum_j\int_{\underline{\theta}}^{\theta}\overline{\Gamma}_{kj}(\theta,\theta')\,x_j(\theta')\,d\theta'+\sum_j\int_{\Theta}\Gamma_{kj}(\theta,\theta')\,x_j(\theta')\,d\theta',k=1,...,K \tag{E19}$$

Proof. See the following proof. ■

E.1.9 Proof of Lemma 12

Since operator T is a contractive mapping, there exists a unique mapping. Therefore, we only need to prove that $\overrightarrow{f}_*$ satisfies $\overrightarrow{f}_*=T\left(\overrightarrow{f}_*\right)+\overrightarrow{x}$. In other words, we only need to prove that for any $k=1,..,K$:

$$f_k^*=T_k\left(\overrightarrow{f}_*\right)+x_k.$$

By the definition of T (see equation (E5)), we have the following equation:

$$\left[T_k\left(\overrightarrow{f}_*\right)+x_k\right](\theta)=x_k(\theta)+\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta')\,f_m^*(\theta')\,d\theta'+\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta')\,f_m^*(\theta')\,d\theta' \tag{E20}$$

Substitude $\overrightarrow{f}_*$ in the above equation by equation (E19):

$$\begin{aligned}\left[T_k\left(\overrightarrow{f}_*\right)+x_k\right](\theta) &= x_k(\theta)+\sum_{j=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{kj}(\theta,\theta')\,x_j(\theta')\,d\theta'+\sum_{j=1}^{K}\int_{\Theta}\gamma_{kj}(\theta,\theta')\,x_j(\theta)\,d\theta' \\ &+\sum_{j=1}^{K}\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\int_{\underline{\theta}}^{\theta}\overline{\Gamma}_{mj}(\theta'',\theta')\,x_j(\theta')\,d\theta'd\theta'' \\ &+\sum_{j=1}^{K}\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\int_{\Theta}\Gamma_{mj}(\theta'',\theta')\,x_j(\theta')\,d\theta'd\theta'' \\ &+\sum_{j=1}^{K}\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\int_{\underline{\theta}}^{\theta}\overline{\Gamma}_{mj}(\theta'',\theta')\,x_j(\theta')\,d\theta' \\ &+\sum_{j}\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\int_{\Theta}\Gamma_{mj}(\theta'',\theta')\,x_j(\theta')\,d\theta'd\theta''.\end{aligned}$$

Using integration by part, we have the following expression for $\left[T_k\left(\overrightarrow{f}_*\right)+x_k\right](\theta)$:

$$\begin{aligned}&\left[T_k\left(\overrightarrow{f}_*\right)+x_k\right](\theta) \qquad\qquad (E21)\\ =\;& x_k(\theta)+\sum_{j=1}^{K}\int_{\underline{\theta}}^{\theta}\left[\overline{\gamma}_{kj}(\theta,\theta')++\sum_{m=1}^{K}\int_{\theta'}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta''\right]x_j(\theta')\,d\theta' \\ &+\sum_{j=1}^{K}\int_{\Theta}\left[\begin{array}{c}\gamma_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta'' \\ +\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\Gamma_{mj}(\theta'',\theta')\,d\theta''+\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\,\Gamma_{mj}(\theta'',\theta')\,d\theta''\end{array}\right]x_j(\theta)\,d\theta'\end{aligned}$$

Remember that $\overrightarrow{\overline{\Gamma}}$ and $\overrightarrow{\Gamma}$ are solution to the resolvent (E16):

$$\begin{aligned}\overline{\Gamma}_{kj}(\theta,\theta') &\equiv \overline{\gamma}_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta'}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta'';\\ \Gamma_{kj}(\theta,\theta') &\equiv \left[\gamma_{kj}(\theta,\theta')+\sum_{m=1}^{K}\int_{\theta''\geq\theta'}\gamma_{km}(\theta,\theta'')\,\overline{\Gamma}_{mj}(\theta'',\theta')\,d\theta''\right] \\ &\quad+\sum_{m=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\gamma}_{km}(\theta,\theta'')\,\Gamma_{mj}(\theta'',\theta')\,d\theta''+\sum_{m=1}^{K}\int_{\Theta}\gamma_{km}(\theta,\theta'')\,\Gamma_{mj}(\theta'',\theta')\,d\theta''\end{aligned}$$

Therefore equation (E21) can be written as:

$$\left[T_k\left(\overrightarrow{f}_*\right)+x_k\right](\theta)=x_k(\theta)+\sum_{j=1}^{K}\int_{\underline{\theta}}^{\theta}\overline{\Gamma}_{kj}(\theta,\theta')\,x_j(\theta')\,d\theta'+\sum_{j=1}^{K}\int_{\Theta}\Gamma_{kj}(\theta,\theta')\,x_j(\theta)\,d\theta'$$

By the definition of $\overrightarrow{f}_*$, we have

$$f_k^*(\theta)=\left[T_k\left(\overrightarrow{f}_*\right)+x_k\right](\theta),k=1,...,K, \qquad (E22)$$

i.e.,

$$\overrightarrow{f}_*=T\left(\overrightarrow{f}_*\right)+\overrightarrow{x}.$$

Therefore function vector $\overrightarrow{f}_*$ is a solution to problem (A.P). Since the solution to problem (A.P) exist and is unique, the solution to the problem (A.P) must satisfy:

$$f_k^*(\theta) = x_k(\theta) + \sum_j \int_{\underline{\theta}}^{\theta} \overline{\Gamma}_{kj}(\theta, \theta') x_j(\theta') d\theta' + \sum_j \int_{\Theta} \Gamma_{kj}(\theta, \theta') x_j(\theta') d\theta', k = 1, ..., K.$$

In conclusion, we have Lemma 12.■

E.2 Supplementary Lemmas

E.2.1 Lemma 13

Lemma 13 *For any $j \in \mathcal{J}$ and $\theta \in \Theta_j$, we have*

$$\widehat{T}_j^{\star}(y_j(\theta)) = \widehat{T}_j^{\star}(y_j(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}_j^{\star\prime}(y_j(x)) dy_j(x). \tag{E23}$$

Proof. See the following proof. ■

E.2.2 Proof of Lemma 13

First, using $\lim_{\mu\to 0} y_j\left(\theta; T + \mu\widehat{T}\right) = y_j(\theta)$, we have

$$\begin{aligned}
&\lim_{\mu\to 0} \frac{T_j^{\star}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star}\left(y_j(\theta; T); T\right)}{\mu} \\
= &\lim_{\mu\to 0} \frac{T_j^{\star}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T\right) - T_j^{\star}\left(y_j(\theta; T); T\right)}{\mu} \\
&+ \lim_{\mu\to 0} \frac{T_j^{\star}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T\right)}{\mu} \\
= &T_j^{\star\prime}(y_j(\theta))\,\widehat{y}_j(\theta) + \widehat{T}_j^{\star}(y_j(\theta)).
\end{aligned}$$

On the other hand, we have

$$\lim_{\mu\to 0} \frac{T_j^{\star\prime}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star\prime}\left(y_j(\theta; T); T\right)}{\mu} = \widehat{T}_j^{\star\prime}(y_j(\theta)) + T_j^{\star\prime\prime}(y_j(\theta))\,\widehat{y}_j(\theta).$$

Besides,

$$\begin{aligned}
&T_j^{\star}\left(y_j\left(\theta+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star}\left(y_j\left(\underline{\theta}+\mu\widehat{T}\right); T+\mu\widehat{T}\right) \\
= &\int_{\underline{\theta}}^{\theta} \frac{\partial T_j^{\star}\left(y_j\left(x+\mu\widehat{T}\right); T+\mu\widehat{T}\right)}{\partial y_j} dy_j\left(x, T+\mu\widehat{T}\right),
\end{aligned} \tag{E24}$$

where

$$T_j^{\star}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) = T_j^{\star}(y_j(\theta; T)) + \left[T_j^{\star}\left(y_j\left(\theta; T+\mu\widehat{T}\right); T+\mu\widehat{T}\right) - T_j^{\star}(y_j(\theta; T))\right], \tag{E25}$$

and

$$T_j^{\star}\left(y_j\left(\underline{\theta};T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)=T_j^{\star}\left(y_j\left(\underline{\theta};T\right)\right)+\left[T_j^{\star}\left(y_j\left(\underline{\theta};T+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\underline{\theta};T\right)\right)\right]. \tag{E26}$$

Similarly, we have:

$$T_j^{\star}\left(y_j\left(\theta\right)\right)-T_j^{\star}\left(y_j\left(\underline{\theta}\right)\right)=\int_{\underline{\theta}}^{\theta}\frac{\partial T_j^{\star}\left(y_j\left(x\right)\right)}{\partial y_j}dy_j\left(x\right). \tag{E27}$$

Combination of (E24) and (E27) delivers

$$\begin{aligned}&\int_{\underline{\theta}}^{\theta}\frac{\partial T_j^{\star}\left(y_j\left(x+\mu\widehat{T}\right);T+\mu\widehat{T}\right)}{\partial y_j}dy_j\left(x,T+\mu\widehat{T}\right)-\int_{\underline{\theta}}^{\theta}\frac{\partial T_j^{\star}\left(y_j\left(x\right)\right)}{\partial y_j}dy_j\left(x\right)\\ =\;&\left[T_j^{\star}\left(y_j\left(\theta+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\underline{\theta}+\mu\widehat{T}\right);T+\mu\widehat{T}\right)\right]-\left[T_j^{\star}\left(y_j\left(\theta\right)\right)-T_j^{\star}\left(y_j\left(\underline{\theta}\right)\right)\right].\end{aligned}$$

Rearrange the above equation:

$$\begin{aligned}&\int_{\underline{\theta}}^{\theta}\frac{\frac{\partial T_j^{\star}\left(y_j\left(x+\mu\widehat{T}\right);T+\mu\widehat{T}\right)}{\partial y_j}-\frac{\partial T_j^{\star}\left(y_j\left(x\right)\right)}{\partial y_j}}{\mu}dy_j\left(x,T+\mu\widehat{T}\right)+\int_{\underline{\theta}}^{\theta}\frac{\partial T_j^{\star}\left(y_j\left(x\right)\right)}{\partial y_j}d\left[\frac{y_j\left(x,T+\mu\widehat{T}\right)-y_j\left(x\right)}{\mu}\right]\\ =\;&\frac{T_j^{\star}\left(y_j\left(\theta+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\theta\right)\right)}{\mu}-\frac{T_j^{\star}\left(y_j\left(\underline{\theta}+\mu\widehat{T}\right);T+\mu\widehat{T}\right)-T_j^{\star}\left(y_j\left(\underline{\theta}\right)\right)}{\mu}.\end{aligned}$$

As $\mu\to 0$, we have

$$\begin{aligned}&\int_{\underline{\theta}}^{\theta}\left[\widehat{T}_j^{\star\prime}\left(y_j\left(x\right)\right)+T_j^{\star\prime\prime}\left(y_j\left(x\right)\right)\widehat{y}_j\left(x\right)\right]dy_j\left(x\right)+\int_{\underline{\theta}}^{\theta}T_j^{\star\prime}\left(y_j\left(x\right)\right)d\widehat{y}_j\left(x\right)\\ =\;&\widehat{T}_j^{\star}\left(y_j\left(\theta\right)\right)+T_j^{\star\prime}\left(y_j\left(\theta\right)\right)\widehat{y}_j\left(\theta\right)-T_j^{\star\prime}\left(y_j\left(\underline{\theta}\right)\right)\widehat{y}_j\left(\underline{\theta}\right)-\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}\right)\right).\end{aligned}$$

Notice that

$$\begin{aligned}T_j^{\star\prime}\left(y_j\left(\theta\right)\right)\widehat{y}_j\left(\theta\right)-T_j^{\star\prime}\left(y_j\left(\underline{\theta}\right)\right)\widehat{y}_j\left(\underline{\theta}\right)&=\int_{\underline{\theta}}^{\theta}d\left[T_j^{\star\prime}\left(y_j\left(x\right)\right)\widehat{y}_j\left(x\right)\right]\\ &=\int_{\underline{\theta}}^{\theta}\widehat{y}_j\left(x\right)dT_j^{\star\prime}\left(y_j\left(x\right)\right)+\int_{\underline{\theta}}^{\theta}T_j^{\star\prime}\left(y_j\left(x\right)\right)d\widehat{y}_j\left(x\right)\\ &=\int_{\underline{\theta}}^{\theta}T_j^{\star\prime\prime}\left(y_j\left(x\right)\right)\widehat{y}_j\left(x\right)dy_j\left(x\right)+\int_{\underline{\theta}}^{\theta}T_j^{\star\prime}\left(y_j\left(x\right)\right)d\widehat{y}_j\left(x\right).\end{aligned} \tag{E28}$$

Combination of (B10) and (E28) delivers

$$\widehat{T}_j^{\star}\left(y_j\left(\theta\right)\right)-\widehat{T}_j^{\star}\left(y_j\left(\underline{\theta}\right)\right)=\int_{\underline{\theta}}^{\theta}\widehat{T}_j^{\star\prime}\left(y_j\left(x\right)\right)dy_j\left(x\right). \tag{E29}$$

■

E.2.3 Lemma 14

Lemma 14 *We have the following results regarding the change of utility with tax reform:*

$$\widehat{V}_k(\theta) = -\widehat{T}_k^A(y_k(\theta)) = -\left[\widehat{T}(y_k(\theta)) + \widehat{T}_k^\star(y_k(\theta))\right] = -\left[\widehat{T}_k^A(y_k(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{A\prime}(y_k(x))\, dy_k(x)\right] \tag{E30}$$

and

$$\begin{aligned}\widehat{U}(\theta) &= \sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \widehat{V}_k(\theta) = -\sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \widehat{T}_k^A(y_k(\theta)) \\ &= -\sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \left[\widehat{T}_k^A(y_k(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{A\prime}(y_k(x))\, dy_k(x)\right]\end{aligned} \tag{E31}$$

and

$$\frac{\widehat{\frac{\partial U}{\partial V_k}}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} = \sum_n \frac{\frac{\partial^2 U}{\partial V_k \partial V_n}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} \widehat{V}_n(x_k) = \sum_n \mu_{kn}(x_k)\, \widehat{V}_n(x_k) = -\sum_n \mu_{kn}(y_k(x_k))\, \widehat{T}_n^A(y_n(x_k)). \tag{E32}$$

Proof. See the following proof. ■

E.2.4 Proof of Lemma 14

Firstly, we have

$$\begin{aligned}\widehat{V}_k(\theta) &= -\widehat{T}_k^A(y_k(\theta)) = -\left[\widehat{T}(y_k(\theta)) + \widehat{T}_k^\star(y_k(\theta))\right] \\ &= -\left[\widehat{T}(y_k(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}'(y_k(x))\, dy_k(x) + \widehat{T}_k^\star(y_k(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{\star\prime}(y_k(x))\, dy_k(x)\right] \\ &= -\left[\widehat{T}_k^A(y_k(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{A\prime}(y_k(x))\, dy_k(x)\right],\end{aligned}$$

where we use equation (E29) to derive the third equation.

Remember that $U(\theta) = U(\theta; V_j(\theta), V_{-j}(\theta))$, $\frac{\partial U(\theta)}{\partial V_j(\theta)} = \frac{\partial U\left(\theta; V_j(\theta), V_{-j}(\theta)\right)}{\partial V_j(\theta)}$, and

$$\widehat{U}(\theta) \equiv \lim_{\mu \to 0} \frac{U\left[\theta; V_j\left(\theta; T + \mu\widehat{T}\right), V_{-j}\left(\theta; T + \mu\widehat{T}\right)\right] - U\left[\theta; V_j(\theta), V_{-j}(\theta)\right]}{\mu},$$

we have

$$\widehat{U}(\theta) = \sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \widehat{V}_k(\theta) = -\sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \widehat{T}_k^A(y_k(\theta)).$$

Use equation (E29) again:

$$\widehat{U}(\theta) = -\sum_k \frac{\partial U(\theta)}{\partial V_k(\theta)} \times \left[\widehat{T}_k^A(y_k(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \widehat{T}_k^{A\prime}(y_k(x))\, dy_k(x)\right]$$

By definitions of $\frac{\partial U}{\partial V_k}(x_k)$ and $\frac{\widehat{\partial U}}{\partial V_k}(x_k)$, we have

$$\frac{\frac{\widehat{\partial U}}{\partial V_k}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} \equiv \lim_{\mu \to 0} \frac{\frac{\partial U}{\partial V_k}\left(x_k; V_j\left(x_k; T+\mu\widehat{T}\right), V_{-j}\left(x_k; T+\mu\widehat{T}\right)\right) - \frac{\partial U}{\partial V_k}\left(x_k; V_j(x_k), V_{-j}(x_k)\right)}{\mu}$$

By the chain rule, we have:

$$\begin{aligned}
& \frac{\frac{\widehat{\partial U}}{\partial V_k}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} \\
= & \lim_{\mu \to 0} \frac{\begin{bmatrix} \frac{\partial U}{\partial V_k}\left(x_k; V_j\left(x_k; T+\mu\widehat{T}\right), V_{-j}\left(x_k; T+\mu\widehat{T}\right)\right) \\ -\frac{\partial U}{\partial V_k}\left(x_k; V_j(x_k), V_{-j}\left(x_k; T+\mu\widehat{T}\right)\right) \end{bmatrix}}{\frac{\partial U(x_k)}{\partial V_k}\left[V_j\left(x_k; T+\mu\widehat{T}\right) - V_j(x_k)\right]} \lim_{\mu \to 0} \frac{V_j\left(x_k; T+\mu\widehat{T}\right) - V_j(x_k)}{\mu} \\
& + \lim_{\mu \to 0} \frac{\begin{bmatrix} \frac{\partial U}{\partial V_k}\left(x_k; V_j(x_k), V_{-j}\left(x_k; T+\mu\widehat{T}\right)\right) \\ -\frac{\partial U}{\partial V_k}\left(x_k; V_j(x_k), V_{-j}(x_k)\right) \end{bmatrix}}{\frac{\partial U(x_k)}{\partial V_k}\left[V_{-j}\left(x_k; T+\mu\widehat{T}\right) - V_{-j}(x_k)\right]} \lim_{\mu \to 0} \frac{V_{-j}\left(x_k; T+\mu\widehat{T}\right) - V_{-j}(x_k)}{\mu} \\
= & \sum\nolimits_n \frac{\frac{\partial^2 U}{\partial V_k \partial V_n}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} \widehat{V}_n(x_k).
\end{aligned}$$

By $\mu_{kn}(x_k) \equiv \frac{\frac{\partial^2 U}{\partial V_k \partial V_n}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)}$, we have:

$$\frac{\frac{\widehat{\partial U}}{\partial V_k}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} = \sum\nolimits_n \frac{\frac{\partial^2 U}{\partial V_k \partial V_n}(x_k)}{\frac{\partial U}{\partial V_k}(x_k)} \widehat{V}_n(x_k) = \sum\nolimits_n \mu_{kn}(x_k) \widehat{V}_n(x_k) = -\sum\nolimits_n \mu_{kn}(y_k(x_k)) \widehat{T}_n^A(y_n(x_k)).$$

In conclusion, we have Lemma 14.■

E.2.5 Lemma 15

Lemma 15 *We have the following results regarding to the change of populations with tax reform:*

$$\begin{aligned}
\frac{\widehat{f}_k^p(\theta_k)}{f_k^p(\theta_k)} &= \sum_n m_{kn}(\theta_k) \widehat{V}_n(\theta_k) = -\sum_n m_{kn}(\theta_k) \widehat{T}_n^A(\theta_k), \\
\frac{\widehat{N}_k}{N_k} &= \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k) f_k(x_k) \widehat{V}_n(x_k) dx_k = -\sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k) f_k(x_k) \widehat{T}_n^A(x_k) dx_k, \\
\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)} &= \sum_n m_{kn}(\theta_k) \widehat{V}_n(\theta_k) - \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k) f_k(x_k) \widehat{V}_n(x_k) dx_k \\
&= -\left[\sum_n m_{kn}(\theta_k) \widehat{T}_n^A(\theta_k) - \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k) f_k(x_k) \widehat{T}_n^A(x_k) dx_k\right], \\
\frac{\widehat{m}_{ki}(\theta)}{m_{ki}(\theta)} &= \sum\nolimits_n \left[\varepsilon_{in}^k(\theta) - 1\right] m_{kn}(\theta) \widehat{V_n}(\theta) = -\sum\nolimits_n \left[\varepsilon_{in}^k(\theta) - 1\right] m_{kn}(\theta) \widehat{T}_n^A(\theta).
\end{aligned}$$

Proof. See the following proof. ■

E.2.6 Proof of Lemma 15

We first calculate the change of $f_k^p(\theta_k)$. By definition of $\widehat{f}_k^p(\theta_k)$ and $m_{kn}(\theta_k) \equiv \frac{\partial \ln f_k^p(\theta_k; V_k(\theta_k), V_{-k}(\theta_k))}{\partial V_n(\theta_k)}$, we have

$$\begin{aligned}\frac{\widehat{f}_k^p(\theta_k)}{f_k^p(\theta_k)} &\equiv \lim_{\mu\to 0} \frac{\frac{f_k^p\left(\theta_k; V_k\left(\theta_k; T+\mu\widehat{T}\right), V_{-k}\left(\theta_k; T+\mu\widehat{T}\right)\right) - f_k^p(\theta_k; V_k(\theta_k), V_{-k}(\theta_k))}{\mu}}{f_k^p(\theta_k; V_k(\theta_k), V_{-k}(\theta_k))} \\ &= \sum_n \frac{\partial \ln f_k^p(\theta_k; V_k(\theta_k), V_{-k}(\theta_k))}{\partial V_n(\theta_k)} \widehat{V}_n(\theta_k) \\ &= \sum_n m_{kn}(\theta_k)\, \widehat{V}_n(\theta_k);\end{aligned} \tag{E33}$$

By definition of N_k and $\widehat{N}_k$:

$$\begin{aligned}\frac{\widehat{N}_k}{N_k} &\equiv \lim_{\mu\to 0} \int_{\underline{\theta}_k}^{\bar{\theta}_k} \frac{\frac{f_k^p\left(x_k; V_k\left(x_k; T+\mu\widehat{T}\right), V_{-k}\left(x_k; T+\mu\widehat{T}\right)\right) - f_k^p(x_k; V_k(x_k), V_{-k}(x_k))}{\mu}}{f_k^p(x_k; V_k(x_k), V_{-k}(x_k))} \frac{f_k^p(x_k; V_k(x_k), V_{-k}(x_k))}{N_k} dx_k \\ &= \int_{\underline{\theta}_k}^{\bar{\theta}_k} \frac{f_k^p(x_k)}{N_k} \times \frac{\widehat{f}_k^p(x_k)}{f_k^p(x_k)} dx_k = \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k)\, f_k(x_k)\, \widehat{V}_n(x_k)\, dx_k.\end{aligned} \tag{E34}$$

By the definition of $f_k(\theta_k)$, we have:

$$\begin{aligned}\frac{\widehat{f}_k(\theta_k)}{f_k(\theta_k)} &\equiv \lim_{\mu\to 0} \frac{\frac{f_k^p\left(\theta_k; T+\mu\widehat{T}\right)}{N_k\left(T+\mu\widehat{T}\right)} - \frac{f_k^p(\theta_k)}{N_k}}{\mu} \frac{1}{f_k(\theta_k)} = \frac{\widehat{f_k^p}(\theta_k)}{f_k^p(\theta_k)} - \frac{\widehat{N}_k}{N_k} \\ &= \sum_n m_{kn}(\theta_k)\, \widehat{V}_n(\theta_k) - \sum_n \int_{\underline{\theta}_k}^{\bar{\theta}_k} m_{kn}(x_k)\, f_k(x_k)\, \widehat{V}_n(x_k)\, dx_k.\end{aligned} \tag{E35}$$

Last, combination of definitions of definitions of $m_{kk}(x_k)$ and $\varepsilon_{in}^k(x_k)$, and $m_{ki}(\theta) = \frac{\partial \ln\left[f_k^p/f\right]}{\partial V_i}$, we have (E36). To see this, remember that by the definition of $\varepsilon_{ik}^j(\theta)$, we have

$$\begin{aligned}\varepsilon_{ik}^j(\theta) &\equiv \frac{\frac{\partial^2\left[f_j^p/f\right]}{\partial V_i \partial V_k} \frac{f_j^p}{f}}{\frac{\partial\left[f_j^p/f\right]}{\partial V_i} \frac{\partial\left[f_j^p/f\right]}{\partial V_k}} = \frac{\frac{\partial \ln\left\{\frac{\frac{\partial\left[f_j^p/f\right]}{\partial V_i}}{f_j^p/f} \times f_j^p/f\right\}}{\partial V_k}}{\frac{\partial \ln\left[f_j^p/f\right]}{\partial V_k}} \\ &= \frac{\frac{\partial \ln \frac{\partial\left[f_j^p/f\right]}{\partial V_i}}{\partial V_k}}{\frac{\partial \ln\left[f_j^p/f\right]}{\partial V_k}} + \frac{\frac{\partial \ln\left[f_j^p/f\right]}{\partial V_k}}{\frac{\partial \ln\left[f_j^p/f\right]}{\partial V_k}} = \frac{\frac{\partial \ln m_{ji}(\theta)}{\partial V_k}}{m_{jk}(\theta)} + 1,\end{aligned}$$

which gives $\frac{\partial \ln m_{ji}(\theta)}{\partial V_k} = \left[\varepsilon_{ik}^j(\theta) - 1\right] m_{jk}(\theta)$. Equivalently,

$$\frac{\partial \ln m_{ki}(\theta)}{\partial V_n(\theta)} = \left[\varepsilon_{in}^k(\theta) - 1\right] m_{kn}(\theta).$$

Therefore, we have the following result:

$$\begin{aligned}\widehat{m}_{ki}(\theta) &= m_{ki}(\theta)\sum\nolimits_n \frac{\partial \ln m_{ki}(\theta)}{\partial V_n(\theta)}\widehat{V}_n(\theta) \\ &= m_{ki}(\theta)\sum\nolimits_n \left[\varepsilon_{in}^k(\theta)-1\right] m_{kn}(\theta)\widehat{V}_n(\theta) \\ &= -m_{ki}(\theta)\sum\nolimits_n \left[\varepsilon_{in}^k(\theta)-1\right] m_{kn}(\theta)\widehat{T}_n^A(\theta).\end{aligned} \tag{E36}$$

In conclusion, we have Lemma 15.■

E Primal Approach with Labor Mobility

This section presents the problems of the local and central governments and presents solutions. We give relevant proof in the next four sections.

E.1 Decentralized Redistribution

Local government's reaction to federal income tax can be treated in a number of different ways. In the heuristic argument that follows, j local government adopts a feasible direct, truthful mechanism, $\left\{c_j(\theta), y_j(\theta)\right\}_{\theta\in\Theta}$, to implement allocation rules of consumption and labor $\left\{c_j(\theta), l_j(\theta)\right\}_{\theta\in\Theta}$ that maximize regional welfare under regional resource constraint (local government's budget constraint). Under such a mechanism, an agent who reports θ should earn $y_j(\theta)$ units of labor income in region j, and will receive $c_j(\theta)$ units of consumption as a reward. It is worth noting that j local government takes the policies of other governments as given, which suggests that it also takes $\left\{V_{-j}(\theta)\right\}_{\theta\in\Theta}$ as given.

It turns out to be easier if we take $l_j(\theta)$ rather than $y_j(\theta)$. Further, we substitute $c_j(\theta)$ with $V_j(\theta)$ by $V_j(\theta) = c_j(\theta) - \phi\left(l_j(\theta)\right)$ as our control variables. The j local government must choose schedules $\left\{l_j(\theta), V_j(\theta)\right\}_{\theta\in\Theta}$ to maximize its objective such that the allocation satisfies incentive-compatible conditions:[49]

$$V_j'(\theta) = \frac{l_j(\theta)\phi'\left(l_j(\theta)\right)}{\theta}, \forall\theta\in\Theta, \tag{E37}$$

the laborer supply condition, and the local government budget constraint (equivalently, the regional resource constraint):

$$\int_\Theta A_j\theta l_j(\theta) f_j^p(\theta)\,d\theta - \int_\Theta \left[V_j(\theta)+\phi\left(l_j(\theta)\right)\right] f_j^p(\theta)\,d\theta - \int_\Theta T\left(A_j\theta l_j(\theta)\right) f_j^p(\theta)\,d\theta \geq E_j. \tag{E38}$$

Solving the above planning problem, we observe the optimal local income tax formula presented in Proposition 3.

[49] Individual report types to maximize individual utility $U_j(\theta) \equiv \max_{\theta'\in\Theta} c_j(\theta') - \phi\left(\frac{y_j(\theta')}{A_j\theta}\right)$. Denote the utility of a θ-type individual who reports θ' as $U_j(\theta'|\theta) = c_j(\theta') - \phi\left(\frac{y_j(\theta')}{A_j\theta}\right)$. The first-order necessary incentive-compatible conditions require $\frac{\partial U_j(\theta'|\theta)}{\partial\theta'}|_{\theta'=\theta} = 0$, $\forall\theta\in\Theta$. It is common in the optimal Mirrleesian tax literature that under our preference, the first-order incentive condition is not only necessary but also sufficient and is equivalent to (E37).

E.2 Centralized Redistribution

Policy Constraint. A region-uniform federal income tax should satisfy the following policy constraint: for any θ_j and θ_{-j} that satisfy $y_j\left(\theta_j\right)=y_{-j}\left(\theta_{-j}\right)$, $T'\left(y_j\left(\theta_j\right)\right)=T'\left(y_{-j}\left(\theta_{-j}\right)\right)$. According to Proposition 3 and (2), this is equivalent to say that for any θ_j and θ_{-j} satisfying $y_j\left(\theta_j\right)=y_{-j}\left(\theta_{-j}\right)$,

$$\begin{aligned}&\frac{\phi'\left(\frac{y_1(\theta_1)}{A_1\theta_1}\right)}{A_1\theta_1}\left[1+\left[1-\overline{g}_1\left(\theta_1\right)-\overline{r}_{11}\left(\theta_1\right)\right]\frac{N_1-F_1^p\left(\theta_1\right)}{\theta_1 f_1^p\left(\theta_1\right)}\left(1+\frac{1}{\varepsilon_1}\right)\right] \\ =\ &\frac{\phi'\left(\frac{y_2(\theta_2)}{A_2\theta_2}\right)}{A_2\theta_2}\left[1+\left[1-\overline{g}_2\left(\theta_2\right)-\overline{r}_{22}\left(\theta_2\right)\right]\frac{N_2-F_2^p\left(\theta_2\right)}{\theta_2 f_2^p\left(\theta_2\right)}\left(1+\frac{1}{\varepsilon_2}\right)\right].\end{aligned} \tag{E39}$$

The above policy constraint is non-parametric. There is a way to express the policy constraint in the space of parametric. Introduce $\theta_j\left(\theta_1\right)$ as the skill of worker who earns $y_1\left(\theta_1\right)$ in region j, $j\in\mathcal{J}$. Note that $\theta_1\left(\theta_1\right)=\theta_1$ and $\theta_2\left(\theta_1\right)$ is determined by $y_1\left(\cdot\right)$ and $y_2\left(\cdot\right)$. Under Assumption 1, $\theta_2:\Theta\mapsto\Theta$ is a one-to-one mapping. Thus, one can replace $y_2\left(\cdot\right)$ by $\theta_2\left(\cdot\right)$. Then we have an alternative form of policy constraint. For any $\theta\in\Theta$,

$$\begin{aligned}&\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta}\left[1+\left[1-\overline{g}_1\left(\theta\right)-\overline{r}_{11}\left(\theta\right)\right]\frac{N_1-F_1^p\left(\theta\right)}{\theta f_1^p\left(\theta\right)}\left(1+\frac{1}{\varepsilon_1}\right)\right] \\ =\ &\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2\left(\theta\right)}\left[1+\left[1-\overline{g}_2\left(\theta_2\left(\theta\right)\right)-\overline{r}_{22}\left(\theta_2\left(\theta\right)\right)\right]\frac{N_2-F_2^p\left(\theta_2\left(\theta\right)\right)}{\theta_2\left(\theta\right)f_2^p\left(\theta_2\left(\theta\right)\right)}\left(1+\frac{1}{\varepsilon_2}\right)\right].\end{aligned} \tag{E40}$$

Wedges. In traditional second-best solution, marginal distortions in agents' choices can be described with wedges. This is also the case in our economy. Since individuals have two possible choices (consumption and leisure), and local governments have one type of policy, there are two types of wedges. Tax wedge for individuals and local governments are captured by the aggregate local income tax rate $T_j^{A\prime}\left(\cdot\right)$ and federal income tax rate $T'\left(\cdot\right)$, respectively. To see the later, redefine $\tau_j^\star(\theta)$ as

$$\tau_j^\star(\theta)\equiv\frac{\left[1-\overline{g}_j\left(\theta\right)-\overline{r}_{jj}\left(\theta\right)\right]\frac{1-F_{\theta_j}(\theta)}{\theta f_j(\theta)}\left(1+\frac{1}{\varepsilon_j}\right)}{1+\left[1-\overline{g}_j\left(\theta\right)-\overline{r}_{jj}\left(\theta\right)\right]\frac{1-F_{\theta_j}(\theta)}{\theta f_j(\theta)}\left(1+\frac{1}{\varepsilon_j}\right)}, \tag{E41}$$

which, according to Proposition 3, is the optimal local income tax rate without federal income tax. Marginal federal income tax act as a wedge between $\frac{1}{A_j\theta_j}\phi'\left(\frac{y_j(\theta_j)}{A_j\theta_j}\right)$ and $1-\tau_j^\star(\theta)$. Combination of Proposition 3 and (2) gives:

$$T'\left(y_j\left(\theta\right)\right)=1-\frac{\frac{1}{A_j\theta}\phi'\left(\frac{y_j(\theta)}{A_j\theta}\right)}{1-\tau_j^\star(\theta)},\forall\theta\in\Theta,j\in\mathcal{J}. \tag{E42}$$

Once $T_j^{A\prime}\left(\cdot\right)$ and $T'\left(\cdot\right)$ are determined, we can derive $T_j'\left(\cdot\right)$ by $T_j'\left(\cdot\right)=T_j^{A\prime}\left(\cdot\right)-T'\left(\cdot\right)$.

Implementability. In addition to intensive and extensive incentive-compatible constraints and resource constraint, the implementable allocation must satisfy the policy constraints when federal income tax is restricted to be region-uniform.

Lemma 16 *Suppose that the FOCs of the individuals and the local governments are both necessary and sufficient. The allocation of MGTE with labor mobility,* $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right), f_j^p\left(\theta\right)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$*, satisfies the following conditions jointly:*

(i) Incentive constraints (C1) and laborer supply condition $f_j^p(\theta) = f_j^p\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$,[50] *where* $y_j\left(\theta_j\right) = A_j\theta_j l_j\left(\theta_j\right)$, $V_j\left(\theta_j\right) = c_j\left(\theta_j\right) - \phi\left(l_j\left(\theta_j\right)\right)$, $V_2\left(\theta_2\left(\theta\right)\right) = V_2\left(\underline{\theta}_2\right) + \int_{\underline{\theta}_2}^{\theta_2(\theta)} V_2'\left(x\right)dx$, $V_1\left(\theta_2\left(\theta\right)\right) = V_1\left(\underline{\theta}_1\right) + \int_{\underline{\theta}_1}^{\theta_2(\theta)} V_1'\left(x\right)dx$. $\theta_2\left(\theta\right)$ *is determined by* $y_2\left(\theta_2\left(\theta\right)\right) = y_1\left(\theta\right)$*, i.e.,* $y_1\left(\theta\right) = \int_{\underline{\theta}_2}^{\theta_2(\theta)} y_2'\left(x\right)dx$*;*

(ii) Policy constraints,

$$\begin{aligned} &\frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2\left(\theta\right)}\left[1+\left[1-\overline{g}_2\left(\theta_2\left(\theta\right)\right)-\overline{r}_{22}\left(\theta_2\left(\theta\right)\right)\right]\frac{1-F_2\left(\theta_2\left(\theta\right)\right)}{\theta_2\left(\theta\right)f_2\left(\theta_2\left(\theta\right)\right)}\frac{1+\varepsilon_2}{\varepsilon_2}\right] \\ = \;&\frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta}\left[1+\left[1-\overline{g}_1\left(\theta\right)-\overline{r}_{11}\left(\theta\right)\right]\frac{1-F_1\left(\theta\right)}{\theta f_1\left(\theta\right)}\frac{1+\varepsilon_1}{\varepsilon_1}\right], \forall\theta\in\Theta, \end{aligned} \tag{E43}$$

where terms in equations (E43) can be expressed as functionals of $\left\{V_j\left(\cdot\right), \theta_2\left(\cdot\right), \theta_2'\left(\cdot\right)\right\}_{j=1,2}$*;*

(iii) Goods market clearing conditions (C2).

Conversely, suppose the allocation $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right), f_j^p\left(\theta\right)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ *satisfies the properties 1–3. Then there exists a central policy system* $\mathcal{T}^C$ *such that the allocation can be implemented by* $\mathcal{T}^C$ *at some regional income taxes* $\mathcal{T}^R$ *in a MGTE.*

Proof. See Appendix E.3. ■

Symmetric Case. In a SMGTE, the policy constraint (E43) is naturally satisfied. This property of symmetric allocation helps us to simplify the problem. Moreover, under a symmetric economy, $f_{\theta_j}\left(\theta\right) = f_\theta\left(\theta\right)$ is exogenous. Thus, the central government's problem is reduced to the traditional one. Formally, the central government's planning problem is choosing a symmetric allocation $\{c(\theta), l(\theta)\}_{\theta\in\Theta}$ that maximize the national welfare $\int_\Theta U\left(\theta\right)\widetilde{f}^C\left(\theta\right)d\theta$ under the incentive constraints

$$V'\left(\theta\right) = \frac{l\left(\theta\right)\phi'\left(l\left(\theta\right)\right)}{\theta}, \forall\theta\in\Theta, \tag{E44}$$

and national resource constraint

$$\int_\Theta y\left(\theta\right)f_\theta\left(\theta\right)d\theta = \int_\Theta c\left(\theta\right)f_\theta\left(\theta\right)d\theta. \tag{E45}$$

Formally, we have the following Corollary 5:

Corollary 5 *Suppose that the FOCs of the individuals and the local governments are both necessary and sufficient. The allocation* $\left\{l_j\left(\theta_j\right), c_j\left(\theta_j\right)\right\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ *of a SMGTE satisfies the following conditions jointly:*

(i) Incentive conditions (E44);

(ii) Final market clearing condition (E45).

[50] Notice that $f_j^p(\theta_2\left(\theta\right)) = f_j^p\left(\theta_2\left(\theta\right), V_j(\theta_2\left(\theta\right)), V_{-j}(\theta_2\left(\theta\right))\right)$. To express $f_j^p(\theta_2\left(\theta\right))$, $V_2\left(\theta_2\left(\theta\right)\right)$ and $V_1\left(\theta_2\left(\theta\right)\right)$ are needed.

Conversely, suppose an allocation $\{l_j(\theta_j), c_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ satisfies the properties 1–2 above. Then there exists a central policy system $\mathcal{T}^C$ such that the allocation can be implemented by $\mathcal{T}$ at some regional income taxes $\mathcal{T}^R$ in a SMGTE.

Proof. See Appendix E.4. ■

Corollary 5 suggests that even if laborers are mobile, any incentive-compatible and feasible allocation can be implemented by the central government. Intuitively, the central government can get any symmetrical aggregate income tax by adjusting the federal income tax. Thus, any symmetric allocation that can be achieved in an economy without regional income taxes can also be achieved with regional income taxes.

E.3 Proof of Lemma 16

We first show that a MGTE allocation $\{l_j(\theta_j), c_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ must satisfy parts 1–3. First, by the definition of MGTE, (C2) must be satisfied. Second, by the definition of MGTE, individuals maximize their utilities, which results in (C1), $f_j^p(\theta) = f_j^p(\theta, V_j(\theta), V_{-j}(\theta))$, and (2). Third, by the definition of MGTE, local governments maximize their objectives, which results in Proposition 3. Proposition 3 together with (2) gives (E43).

Next, suppose that we are given an allocation $\{l_j(\theta_j), c_j(\theta_j), f_j^p(\theta)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ satisfying parts 1–3. We now construct $\mathcal{T}^C$ and $\mathcal{T}^R$, which together with $\{l_j(\theta_j), c_j(\theta_j), f_j^p(\theta)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ constructs a MGTE. We first construct $\mathcal{T}^C$ and $\mathcal{T}^R$ with the given allocation $\{l_j(\theta_j), c_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$. We then show that the constructed policy system $\mathcal{T}^C$ and $\mathcal{T}^R$ together with $\{l_j(\theta_j), c_j(\theta_j), f_j^p(\theta)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ do construct a MGTE.

First, given the allocation $\{l_j(\theta_j), c_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ we construct $\{y_j(\theta_j), V_j(\theta_j)\}_{\theta_j\in\Theta, j\in\mathcal{J}}$ by $y_j(\theta_j) = A_j\theta_j l_j(\theta_j)$ and $V_j(\theta_j) = c_j(\theta_j) - \phi(l_j(\theta_j))$. In addition, we construct $\theta_2(\theta)$ by $y_2(\theta_2(\theta)) = y_1(\theta)$, $\forall\theta\in\Theta$. The aggregate regional tax rates $T_j^{A\prime}(\cdot)$ are constructed by (2):

$$T_j^{A\prime}(y_j(\theta)) = 1 - \frac{\phi'\left(\frac{y_j(\theta)}{A_j\theta}\right)}{A_j\theta}.$$

We now construct $T'(y_j)$. First, we construct $\theta_2(\theta)$, $V_2(\theta_2(\theta))$ and $V_1(\theta_2(\theta))$ by $y_2(\theta_2(\theta)) = y_1(\theta)$, $V_2(\theta_2(\theta)) = V_2(\underline{\theta}_2) + \int_{\underline{\theta}_2}^{\theta_2(\theta)} V_2'(x)\,dx$ and $V_1(\theta_2(\theta)) = V_1(\underline{\theta}_1) + \int_{\underline{\theta}_1}^{\theta_2(\theta)} V_1'(x)\,dx$, respectively. Remember that, $f_j^p(\theta) = f_j^p(\theta, V_j(\theta), V_{-j}(\theta))$ and $f_j^p(\theta_2(\theta)) = f_j^p(\theta_2(\theta), V_j(\theta_2(\theta)), V_{-j}(\theta_2(\theta)))$. Therefore, $N_j = \int_\Theta f_j^p(\theta)\,d\theta$, $F_2^p(\theta_2(\theta)) = \int_{\underline{\theta}_2}^{\theta_2(\theta)} f_2^p(x)\,dx$, and $F_1^p(\theta) = \int_{\underline{\theta}_1}^{\theta} f_1^p(x)\,dx$ can be expressed as functionals of $\{V_j(\cdot), \theta_2(\cdot)\}_{j=1,2}$. By definitions (see, e.g., (C20)), $\overline{g}_2(\theta_2(\theta))$ and $\overline{g}_1(\theta)$ can also be expressed as functionals of $\{V_j(\cdot), \theta_2(\cdot)\}_{j=1,2}$. Now set $\widetilde{T}(\theta) \equiv T(y_1(\theta))$, $\widetilde{T}_2(\theta) \equiv T_2(y_1(\theta)) = y_1(\theta) - \left[V_2(\theta_2(\theta)) + \phi\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\right] - \widetilde{T}(\theta)$ and $\widetilde{T}_1(\theta) \equiv T_1(y_1(\theta)) = y_1(\theta) - \left[V_1(\theta) + \phi\left(\frac{y_1(\theta)}{A_1\theta}\right)\right] - \widetilde{T}(\theta)$. We have:

$$r_{11}(\theta) = m_{11}(\theta)\left[y_1(\theta) - \left[V_1(\theta) + \phi\left(\frac{y_1(\theta)}{A_1\theta}\right)\right] - \widetilde{T}(\theta) - E_1\right]$$

and

$$r_{22}(\theta_2(\theta)) = m_{22}(\theta_2(\theta))\left[y_1(\theta) - \left[V_2(\theta_2(\theta)) + \phi\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)\right] - \widetilde{T}(\theta) - E_2\right],$$

where m_{jj} is given by (26). In addition, $\bar{r}_{11}(\theta) = \frac{\int_\theta^\infty r_{11}(x) f_1^p(x) dx}{N_1 - \widehat{F}_1(\theta)}$ and $\bar{r}_{22}(\theta_2(\theta)) = \frac{\int_\theta^\infty r_{jj}\left(\widetilde{\theta}_2(y)\right) f_2^p\left(\widetilde{\theta}_2(y)\right) d\theta_2(y)}{N_2 - \widehat{F}_2(\theta_2(\theta))}$.

According to Proposition 3, we have:

$$\begin{aligned}
1 - T'(y_1(\theta)) &= 1 - \frac{\widetilde{T}'(\theta)}{y_1'(\theta)} \\
&= \frac{\phi'\left(\frac{y_1(\theta)}{A_1\theta}\right)}{A_1\theta}\left[1 + [1 - \overline{g}_1(\theta) - \bar{r}_{11}(\theta)]\frac{N_1 - F_1^p(\theta)}{\theta f_1^p(\theta)}\left(1 + \frac{1}{\varepsilon_1}\right)\right] \\
&= \frac{\phi'\left(\frac{y_1(\theta)}{A_2\theta_2(\theta)}\right)}{A_2\theta_2(\theta)}\left[1 + [1 - \overline{g}_2(\theta_2(\theta)) - \bar{r}_{22}(\theta_2(\theta))]\frac{N_2 - F_2^p(\theta_2(\theta))}{\theta_2(\theta) f_2^p(\theta_2(\theta))}\left(1 + \frac{1}{\varepsilon_2}\right)\right].
\end{aligned}$$

Therefore, one can derive $\widetilde{T}'(\cdot)$ as a functional of $\left\{V_j(\cdot), \theta_2(\cdot), \theta_2'(\cdot), y_1(\cdot), y_1'(\cdot)\right\}_{j=1,2}$ and $\widetilde{T}(\overline{\theta})$. Since $T'(y_1(\theta)) = \frac{\widetilde{T}'(\theta)}{y_1'(\theta)}$, given $\widetilde{T}(\overline{\theta})$, $T'(\cdot)$ is also a functional of $\left\{V_j(\cdot), \theta_2(\cdot), \theta_2'(\cdot), y_1(\cdot), y_1'(\cdot)\right\}_{j\in\mathcal{J}}$. Last, we can construct the local income tax rate by $T_j'(y_j) = T_j^{A\prime}(y_j) - T'(y_j)$.

To fix the policies, we should solve for the boundary values of taxes and $\left\{E_j\right\}_{j\in\mathcal{J}}$. This part follows the proof of Lemma 6 in Appendix C.1.

Last, we show that the allocation $\mathcal{A}$ satisfying parts 1–3 together with the constructed policy system $\mathcal{T}$ constructs an MGTE. First, the allocation satisfies the incentive conditions (C1) and $f_j^p(\theta) = f_j^p\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$, which are both necessary and sufficient for agents' optimal choice under our preconditions of Lemma 16. Second, under the constructed E_j, the constructed local income taxes $\mathcal{T}^R$ satisfies Proposition 3, local governments' budget constraint and the incentive conditions (C1) and $f_j^p(\theta) = f_j^p\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$. Therefore, regional income taxes are the optimal solution to local governments' problems. Third, the goods market clear condition (C2) are satisfied. Lastly, agents' budget constraints are embedded in the definitions of gross utilities and the construction of income taxes. In conclusion, the constructed policy system $\mathcal{T}$ together with the given allocation $\mathcal{A}$ constructs an MGTE.■

E.4 Proof of Corollary 5

The first part of Corollary 5 is obvious. To prove part 2 of Corollary 5 we only need to show that the policy constraints are naturally satisfied and the labor distribution must be symmetric such that given an allocation solving the planner's problem, we can construct an MGTE following the steps given in the proof of Lemma 16 and such an MGTE is an SMGTE. To see this, first note that under symmetric allocations, for any $\theta \in \Theta$, $V_j(\theta)$ is independent on j, thus according to labor market clear conditions $f_j^p(\theta) = f_j^p\left(\theta, V_j(\theta), V_{-j}(\theta)\right)$, the resulting distribution of labors must be symmetric. Second, as allocation is symmetric, $y_j(\theta)$ must equals $y_{-j}(\theta)$. Under Assumption 1, $y_j(\theta) = y_{-j}(\theta)$ implies $\widetilde{\theta}_j(y) = \theta_{-j}(y)$. Therefore, under symmetric allocation, the policy constraints are naturally satisfied. In conclusion, we have Corollary 5.■