Optimal Taxation in the Age of AI Uncertainty

Anastasios Karantounias* Radoslaw Paluszynski† Pei Cheng Yu‡

March 22, 2026

Preliminary (please do not circulate)

Abstract

This paper examines the impact of non-insurable technological shocks on labor markets and optimal tax policy. Using survey data, we document a substantial dispersion in artificial intelligence (AI) pessimism that is negatively correlated with worker income. Motivated by these findings, we build a dynamic Mirrlees income taxation model with ambiguity-averse workers. Workers form endogenous worst-case beliefs about the impact of new technologies, that vary by skill level. This endogenous heterogeneity in expectations enables the formulation of government policies that indirectly influence workers' subjective beliefs to enhance welfare. We calibrate the model to the empirical findings on AI expectations and solve for optimal tax schedules and AI investment. We find that uncertainty over AI motivates a more progressive labor income tax schedule, combined with a mostly flat savings subsidy of about 10%. The optimal policy reverses the sign of the empirical relationship between AI pessimism and worker income.

JEL Codes: D81; D82; E61; E62; H21.

Keywords: Optimal taxation; multiplier preferences; endogenous beliefs; redistribution; artificial intelligence.

*a.karantounias@surrey.ac.uk, University of Surrey.

†rpaluszynski@uh.edu, University of Houston.

‡pei-cheng.yu@unsw.edu.au, University of New South Wales.

1 Introduction

Throughout history, technology has repeatedly reshaped the labor market. The current wave, fueled by rapid advances in information technology, artificial intelligence (AI), and robotics, is unfolding at an exceptional pace and appears primed to transform how work is organized and rewarded. The next generation of technological advancements has the potential to amplify income disparities by reallocating earnings opportunities across the skill distribution in ways that are difficult to foresee. Designing policies that can prepare us for this uncertainty, rather than reacting after the fact, is both immediate and essential.

The urgency of this issue is well recognized in economics. A large body of work studies how automation shifts demand away from low-skilled or routine-intensive tasks (Acemoglu and Restrepo, 2018) and evaluates the best policy responses (Guerreiro et al., 2022; Costinot and Werning, 2023; Thuemmel, 2023). By contrast, far less is known about how the latest AI tools will reshape the labor force, especially for high-skilled occupations. For example, in medical diagnostics, AI systems can evaluate X-rays at or above the accuracy of many radiologists (Agarwal et al., 2025). It is uncertain whether this new technology will reduce the demand for radiologists, or increase their value by enabling them to focus on more critical tasks. It is similarly uncertain if AI is more likely to displace most of current low-skill workers, or raise their productivity by supplementing their labor with expert-level knowledge.

We start by investigating emerging evidence from a nationally representative panel survey on people's expectations of workplace impact of AI in the near future. We find a wide heterogeneity in the sentiments of respondents, with pessimists generally outnumbering optimists. Pessimism is also negatively correlated with household income, while optimists can be found across all income ranks.

These findings underscore the challenges of the impending AI revolution: its unpredictable and difficult to quantify effects across various skill levels. This stands in sharp contrast to much of the optimal taxation literature, which typically treats technological change as a predictable shift in production technology and derives policy prescriptions under the assumption that the direction and magnitude of skill bias are known. This paper seeks to bridge this gap by formulating policies tailored to the potentially unpredictable dynamics of

the labor market brought on by new technology. We characterize the optimal instruments that mitigate the short-run effects of new technology across the full spectrum of skills.

Our paper develops a quantitative two-period Mirrlees taxation model with permanent skills in which the impact of technological change in the second period is ambiguous rather than risky. In the standard framework, technological change is treated as well-known, quantifiable *risk*: the government and agents share a well-specified and correct probability distribution over aggregate technology shocks, and optimal policies balance redistribution and incentives under that known distribution. To capture the unpredictable impact of the impending technological advances on the entire skill distribution, we move beyond risk and consider Knightian uncertainty or *ambiguity*. Agents entertain *multiple* probability distributions and are *averse* to this ambiguity: they behave cautiously and form worst-case assessments about the likelihood of shocks. Their worst-case beliefs are tilted towards adverse outcomes. We represent this using the multiplier preferences of Hansen and Sargent (2001), a tractable specification that is widely used in macroeconomics.

Embedding ambiguity-averse agents with heterogeneous skills in a Mirrlees setting yields two novel features for tax design. First, worst-case beliefs depend on continuation utility, so they are endogenously determined by policy. Specifically, the government can manipulate the set of states that agents fear most by changing the allocation through its policies.[1] Second, worst-case beliefs are skill-dependent: high- and low-skill agents may be pessimistic about different technological outcomes. The government can exploit this heterogeneity in beliefs when designing income and savings taxes to relax the incentive compatibility constraints. We show how the optimal policy uses belief manipulation to raise efficiency and improve insurance against unpredictable technological change.

A key theoretical object that characterizes the optimal policies is a set of belief elasticities, which measure how sensitive the agents' subjective beliefs are to changes in allocations. With multiplier preferences, adjustments to consumption or output under a given shock not only change utilities but also the subjective probabilities attached to that and other shocks. We capture this via in-state belief elasticities (how a change in consumption or output under a

[1]Related works by Karantounias (2013, 2023) and Ferriere and Karantounias (2019) examine this belief-manipulation channel in representative-agent environments, focusing on efficiency and abstracting from redistribution.

given shock affects the perceived probability of that shock) and cross-state elasticities (how a change under one shock affects the perceived probabilities of other shocks). These belief elasticities summarize how policies reshape pessimism about technology across states and across the skill distribution, and they underpin our characterization of the optimal labor and savings wedges. As a result, and in contrast to the tax wedges in standard Mirrlees settings, the optimal labor income and savings wedges in our model are modified by an additional term composed of belief elasticities.

We solve the model numerically and bring it to the data by calibrating a positive version of it which we refer to as "current policies" model. In particular, we identify the deep parameters of the model that drive technological shocks and ambiguity aversion by targeting the empirical findings on the distribution of expectations of the workplace impact of AI from the survey data. This allows us to embed both the factual dispersion of pessimism, as well as its negative relationship with income, among the agents in our model. We then solve for optimal policies in this environment. We find that ambiguity regarding the impact of AI has a quantitatively significant influence on the optimal policy prescriptions. Specifically, the optimal labor income tax becomes more progressive, while the savings wedge is fairly uniform across income and negative, implying a savings subsidy of around 10%. Crucially, the optimal policy changes the sign of the relationship between pessimism and income, from negative as in the data (and our calibration) to positive. This implies that belief management can act as a tool of government redistribution in the presence of ambiguity around AI.

The model also predicts that, faced with ambiguity-averse agents, the planner should tax investments in AI to reduce its capability by about 30%. This result echoes many voices in the public debate that call for pausing and regulating the AI sector. While our model is too simple to tackle this question comprehensively, we conduct another exercise where the government chooses optimal policies without an ability to influence AI investments. In this scenario, the optimal labor wedge is mostly unaffected while the savings wedge shifts up aggressively, taxing savings and motivating agents to work harder in the future.

1.1 Literature Review

Our paper is related to two strands of the literature: (i) optimal policies with multiplier or general recursive preferences, and (ii) optimal policies with technological change, such as automation or AI.

Our paper contributes to the literature on policy design with multiplier preferences that focus on agents who are ambiguity averse. Existing studies, such as Karantounias (2013), Ferriere and Karantounias (2019), and Karantounias (2023), typically consider full-information models and explore belief manipulation as a novel policy channel. In these models, ambiguity aversion leads agents—and occasionally planners, as in Karantounias (2023)—to form endogenous worst-case beliefs. Policies can then adjust the payoffs of different states to manage these beliefs and improve welfare.

Our work is closely related to studies that incorporate multiplier preferences or general recursive preferences within models featuring asymmetric information. Farhi and Werning (2008) and Kapicka (2024) explore planning problems with recursive preferences in environments characterized by asymmetric information. Using multiplier preferencs, Bhandari (2015) characterizes the optimal dynamic risk-sharing arrangement between agents facing uncertainty regarding the endowment process. Miao and Rivera (2016) study a dynamic principal-agent framework where the principal has doubts about the output process. Bhandari et al. (2023) analyze a static Mirrlees model in which the planner is uncertain about the productivity distribution.[2] In contrast, our work investigates a planning problem where agents themselves face uncertainty.

Most closely related is the work of Lensman and Troshkin (2022), who consider a dynamic Mirrlees setting with uncertainty in the data-generating process for the aggregate shock. They establish general properties of the optimal policy, showing that it is periodically redesigned in response to new shocks. This periodic reform leads to simplified policies that are not fully contingent on future shocks and may not depend on past history. However, since Lensman and Troshkin (2022) consider a general framework that does not specify the procedure that generates the set of beliefs, they do not fully characterize the resulting tax

[2] Vairo (2024) examines a static Mirrlees setting where the government faces uncertainty regarding agents' production technology and employs a worst-case objective to determine the optimal robust policy.

policies. In contrast, by adopting multiplier preferences, we obtain a specific structured set of beliefs, which enables us to derive explicit optimal income and capital tax policies. Crucially, our approach allows us to demonstrate the precise economic mechanism that drives the optimal distortions.

In our paper, due to the multiplier preferences, consumption and labor are no longer additively separable, and utility is non-separable across states. The literature has extensively examined the impact of non-separable preferences on optimal tax design. Canonical static models by Mirrlees (1971) and Saez (2001) feature utility functions that are non-separable in consumption and labor. Recent contributions in macroeconomic public finance explore the effects of non-separable preferences in dynamic settings (Farhi and Werning, 2013; Golosov et al., 2016). Grochulski and Kocherlakota (2010) and Koehne and Kuhn (2015) study optimal taxes with preferences that are non-separable across time. Most closely related is the work of Hellwig (2021), who considers a general framework incorporating non-separable preferences in both static and dynamic environments. Since the non-separability in our paper stems from ambiguity aversion, we offer new economic interpretations of the optimal taxes through belief elasticities shaped by agents' endogenous expectations management.

Our paper is related to the literature on policies under technological change. Slavik and Yazici (2014) study optimal policy in a model with multiple types of capital and show that it is optimal to raise taxes on capital that is more complementary to high-skill agents in order to raise redistribution. In a similar spirit, Guerreiro et al. (2022) analyze a setting where the price of automation falls over time and find that a short-run robot tax can slow down automation to protect older workers who cannot retrain, while the tax is phased out in the long run to improve efficiency once new cohorts adapt. Thuemmel (2023) incorporates richer occupational heterogeneity and also finds that distorting the pace of automation can compress wage disparities. Costinot and Werning (2023) adopt a sufficient-statistics approach and conclude that optimal robot taxes are small and sensitive to marginal changes in the wage distribution. Some papers consider setups where new capital, such as robots and AI, cannot be directly taxed and examine how other policies should adjust in response to technological change (Loebbing, 2019; Kina et al., 2024). In contrast, our paper highlights the ambiguous nature of new technologies and formulates optimal policies around this uncertainty.

The remainder of the paper is organized as follows. Section 2 provides motivating survey evidence concerning attitudes towards artificial intelligence. Section 3 outlines our model, and Section 4 derives our main theoretical results. Section 5 presents our quantitative analysis. Proofs for the theoretical results in the paper are in Appendix A.

2 Facts About AI Expectations

In this section we use survey data to analyze workers' expectations about the impact on AI technologies on labor markets. We present two main findings. First, there is a considerable dispersion in the expectations. While about a half of respondents are skeptical about significant impacts, the other half tends to be dominated by pessimists who believe AI will have negative effects on job opportunities and pay. Second, such pessimism is negatively correlated with current income. While optimists who expect a positive impact of AI come from all ranks of the income distribution, the pessimists tend to be concentrated among lower income respondents.

2.1 Data Description

The main dataset we employ to study the attitudes towards AI is the Understanding America Study (UAS) project, a survey administered by the Center for Economic and Social Research (CESR) at the University of Southern California. The UAS is a nationally representative, probability-based panel consisting of approximately 15,000 U.S. residents and an additional 3,500-person oversample from California. It collects information about a range of socio-economic indicators that shape Americans' individual and family experiences over the life cycle. In particular, since 2023, it has included an annual wave of questions about the respondents' expectations of the impact of AI in the workplace which is the focus of our study.

Table 1 presents basic summary statistics from the data and contrasts them with analogous moments for the U.S. population. In particular, the first column shows the statistics based on the subsample of respondents who provided answers to the AI block of the survey, while the second column aggregates all participants of the latest wave. As the table shows,

Table 1: Summary Statistics

Variable	Respondents only	Whole sample (2025)	U.S. Population
Demographics			
Female (%)	51.06	51.29	50.50
Median age	42.00	47.00	39.10
65 and older (%)	9.14	22.89	18.00
Born in US (%)	86.42	88.24	86.10
Race/Ethnicity			
White (%)	71.78	72.82	74.80
Black (%)	13.04	13.19	13.70
Asian (%)	7.00	6.00	6.70
Other race (%)	7.58	7.38	4.80
Hispanic (%)	19.23	17.33	20.00
Education			
HS or higher (%)	94.55	92.10	89.40
Bachelor's or higher (%)	43.23	36.41	35.00
Socio-economic			
Median income ($000)	87.50	67.50	78.50
In labor force (%)	90.70	65.60	63.00
Disabled (%)	2.00	9.27	9.10
Household size (mean)	2.86	2.70	2.50
Married (%)	53.11	52.00	47.00
Geography			
Northeast (%)	17.42	17.28	17.28
Midwest (%)	20.23	20.45	20.45
South (%)	37.78	38.73	38.73
West (%)	24.51	23.50	23.50
Observations:	7015	10733	

Notes: Sample statistics calculated using sampling weights. "Respondents only" refers to the subsample of respondents who answered the questions from the Artificial Intelligence block of the survey in at least one wave. "Whole sample" calculates statistics from all respondents of the 2025 survey. U.S. Population data from 2024 Census Survey.

the survey is indeed highly representative of the wider U.S. population, matching the gender, racial and educational composition of the respondents well. A few discrepancies arise among the group who provided answers to the AI-related questions. Such respondents tend to be younger due to under-representation of retirement age people (9% compared with 18% in the population), as well as disabled individuals (2% versus 9%). On the other hand, over 90% of the respondents to these questions are in the labor force, compared with 63% in the

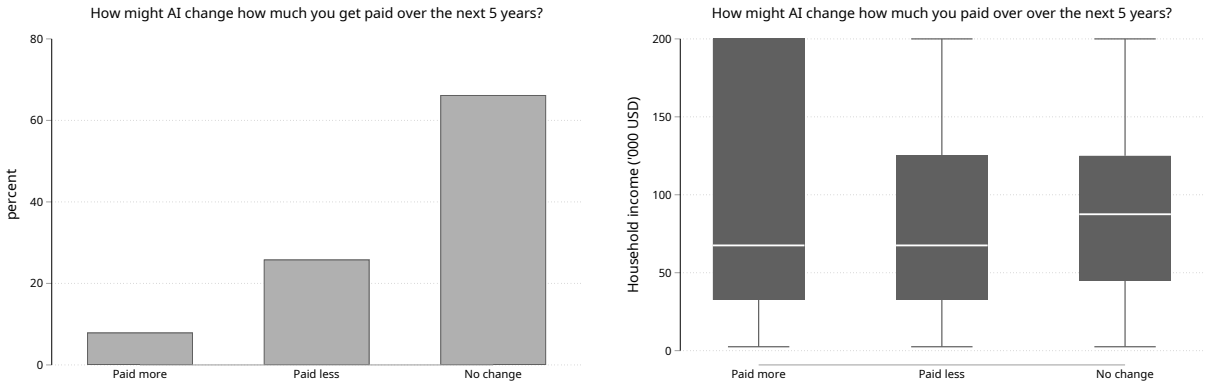
population. This suggests that such questions were most likely skipped by individual who are not active in the labor market. Because the questions are related to the impact of AI on the workplace, such selection bias seems natural and desirable.

2.2 Main Findings

We focus our attention on two main questions asked in the survey that are most related to the theory developed in the following section. First, question $wk005$ in the survey is categorical and reads as follows:

> Thinking of your current job, how do you think AI technologies might influence how much you get paid over the next 5 years?

Figure 1 summarizes the responses to this question. Panel (a) shows that over 60% of respondents believe there will be no change in their level of pay due to AI. Among the rest, about three quarters (30% of total) say they will get paid less while only one quarter (under 10% of all) say they will get paid more. Panel (b) plots the distribution of household income among each respondent group. The optimists, who are on the whole least numerous, span the entire range of income while the pessimists, but also the "agnostics", are concentrated among lower-income households. In Appendix B we test this relationship more systematically by running a multinomial logit regression of the three qualitative responses on household income



(a) Raw responses (b) Responses by income

Notes: Sample consists of each respondent's most recent answer. Sample size is 7015.

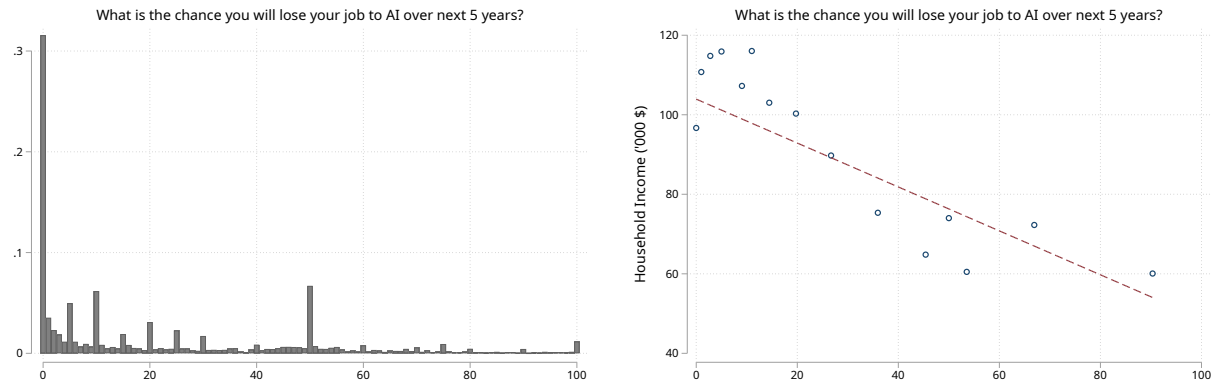
Figure 1: "How might AI change how much you get paid over the next 5 years?"

and a host of demographic controls. We find that higher income is a strong predictor of lower likelihood of being a pessimist, while there is no association of income with the likelihood of being an optimist.

The second part of the survey most relevant to our paper is question $wk007$ which is more quantitative in its nature and reads as follows:

> On a scale from 0 to 100, where 0 means absolutely no chance and 100 means absolutely certain, what are the chances that you will lose your job because it is replaced by AI in the next 5 years?

Figure 2 summarizes the responses to this question. Panel (a) shows that there is a wide dispersion of beliefs spanning the entire range from 0% to 100%. As is common in such surveys, the responses tend to bunch at "round numbers". About a third of the respondents are "agnostics", assigning zero probability to the prospect of losing their job to AI, while other popular responses include 5%, 10%, 20%, 50%, and 100%. Panel (b) shows the relationship of these sentiments with household income using a binned scatterplot. The graph uncovers a strong negative relationship between between the concern about future job loss and income level with a binned correlation of -0.9. In Appendix B we test this relationship more systematically by regressing the survey response on log household income while controlling for a host of demographic characteristics. We find that income is indeed the strongest predic-



(a) Raw responses

(b) Responses by income

Notes: Sample consists of each respondent's most recent answer. Sample size is 7015.

Figure 2: "What is the chance you will lose your job to AI over the next 5 years?"

tor of the perceived job loss probability and associates with it negatively. We will use this statistical relationship to calibrate our quantitative model in Section 5.

Analyzing the responses to questions $wk005$ and $wk007$ in the survey leaves us with two main takeaways. First, there is a wide dispersion in the expectations about the impact of AI technologies on future job outcomes. Agnostics, who are skeptical about any large impact are quantitatively a dominating group while pessimists are more numerous than optimists. Second, pessimism about the impact of AI is strongly correlated with income. These two facts will be crucial for the construction of our model with ambiguity aversion in Section 3. The actual distribution of the responses to the quantitative question $wk007$ will also be used to calibrate are model in Section 5.

2.3 Other Findings

In this section, we summarize the findings from other AI-related questions in the survey, and we discuss the time trends in expectations. While the two questions discussed in the previous section are most useful to our paper, it is important to emphasize that the entire AI block of the USC survey paints a consistent picture. We conclude with a reference to an alternative data source that further supports our main takeaways.

Appendix B summarizes the responses to other AI-related survey questions. Three additional categorical questions ask the respondents about AI's impact on the number of jobs like theirs in the nearest future, its impact on their enjoyment of their job, and the effect on how much they'll be able to accomplish at work. The additional numerical question asks the respondents to predict the fraction of their job tasks that could be performed by AI in the near future. The responses to all of these questions are highly consistent with the main takeaways from the previous section. While "agnostics" are most numerous, pessimist tend to dominate optimists. There is also a strong positive correlation between pessimism and household income.

Because there have already been three waves of the survey on AI expectations, we can also investigate the evolution of such attitudes over time. As documented in Appendix B, the results are mostly stable. If anything, there is a slight trend of increasing pessimism in years 2023-2025 but a quantitatively negligible one.

Finally, we note that there exists an alternative survey that confirms our findings.[3]. While we do not have access to the microdata from that survey, the results summary reveals a considerable dispersion in the attitudes towards AI in the workplace, with pessimists dominating the optimists in general, but especially so among lower income workers.

3 Model

We consider a two-period economy: $t = 0, 1$. It is populated by a continuum of agents indexed by their permanent skill θ, which is distributed according to density $f(\theta)$ and cdf $F(\theta)$ with support $\Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_+$. Agents privately learn their skills in $t = 0$.

In $t = 0$, an agent with skill θ supplying labor l produces output $y = \theta l$, i.e., the marginal productivity of labor in the first period is equal to the skill. An aggregate investment of ϕM, with $\phi \in (0, 1)$, generates machine capability M in the following period. In $t = 1$, machine M changes the marginal productivity of labor according to an aggregate uninsurable technological shock $\xi_i \in \Xi = \{\xi_0, \xi_1, \ldots, \xi_S\}$. Precisely, an agent with skill θ supplying labor l produces output $y = \omega(M, \theta, \xi_i) l$ in $t = 1$ when the technological shock is ξ_i. We assume that $\omega(M, \theta, \xi_i)$ is differentiable in all arguments and increasing in both M and θ. We can interpret M as the capability of new advanced machines, such as AI or robots, and the shock ξ_i affects the complementarity between the agents' skill θ and machine M. For both periods, the government only observes each agent's effective labor y, but not their marginal productivity of labor nor their effort l. Machine M and shock ξ_i are public information.

The labor market is competitive, so the wage is equal to the marginal productivity of labor, and output is the gross income. We use the Hicksian coefficient of complementarity between machine and skill to measure the effect of the technological shock. The Hicksian complementarity is defined as

$$\epsilon^H = \frac{\frac{\partial^2 \omega}{\partial \theta \partial M} \omega}{\frac{\partial \omega}{\partial \theta} \frac{\partial \omega}{\partial M}}.$$

There are two other crucial elasticities that describe the effects of the technological shock. Let $\epsilon^{\omega}_{\theta} = \frac{\partial \omega}{\partial \theta} \frac{\theta}{\omega}$ denote the wage elasticity of skill in $t = 1$, and $\epsilon^{\omega}_{M} = \frac{\partial \omega}{\partial M} \frac{M}{\omega}$ denote the wage

[3] "U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace, Pew Research Center, February 25, 2025

elasticity of machine in $t = 1$. Therefore, we have the following relationship:

$$\epsilon^H - 1 = \frac{1}{\epsilon_M^\omega} \frac{\partial \epsilon_\theta^\omega}{\partial M} \frac{M}{\epsilon_\theta^\omega} = \frac{1}{\epsilon_\theta^\omega} \frac{\partial \epsilon_M^\omega}{\partial \theta} \frac{\theta}{\epsilon_M^\omega}.$$

Hence, when $\epsilon^H > 1$, then $\partial \epsilon_\theta^\omega / \partial M > 0$. That is, the wage elasticity of skill ϵ_θ^ω increases as machine M rises. In essence, machine and skill are complements, the associated technological shock is skill-biased.

Utility is additively separable across time and between consumption and leisure. Specifically, agents have the following flow utility in t :

$$u(c_t) - h(l_t).$$

The utility of consumption $u(c)$ is twice differentiable, strictly increasing, and strictly concave in consumption $c : u', -u'' > 0$. The disutility of work $h(l)$ is twice differentiable, strictly increasing, and strictly convex in the labor supply $l : h', h'' > 0$ with $h(0) = 0$. Let $\beta < 1$ be the discount factor.

3.1 Ambiguity Aversion

The innovative feature of our paper is that there are doubts regarding the distribution of the technological shock ξ_i. Precisely, there is an estimate: ξ_i is distributed according to a *reference* probability mass function $p(\xi_i)$, i.e., $p(\xi_i) > 0$ for all $\xi_i \in \Xi$ and $\sum_{i=0}^{S} p(\xi_i) = 1$.

The reference distribution is a good approximation for the true probability measure, but both the government and the agents are concerned that the reference distribution is misspecified. To deal with the possibility of misspecification, a set of alternative distributions $\tilde{p}_{M,\theta}(\xi_i)$ that are statistically close to the reference distribution is entertained. Note that we allow the set of alternative distributions being considered to depend on M and differ across skills. We denote the likelihood ratio between the reference and the alternative distributions as $m_{M,\theta}(\xi_i) = \tilde{p}_{M,\theta}(\xi_i) / p(\xi_i)$. By construction, the alternative distribution has to sum up to unity: $\sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i) = 1$.

We use relative entropy or Kullback–Leibler divergence to measure the statistical prox-

imity between the alternative distribution $\tilde{p}_{M,\theta}(\xi_i)$ and the reference distribution $p(\xi_i)$:

$$D(\tilde{p}_{M,\theta}||p) = \sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i) \ln(m_{M,\theta}(\xi_i)).$$

The relative entropy $D(\tilde{p}_{M,\theta}||p)$ is non-negative and zero for identical distributions, i.e., $\tilde{p}_{M,\theta} = p$. The relative entropy also increases with the distinguishability between the two distributions.[4] While relative entropy is also defined for logarithms to different bases, the natural logarithm is most useful for our purposes.

To capture the agents' aversion towards uncertainty, we assume that they have multiplier preferences (Hansen and Sargent, 2001, 2008).[5] For a given truth-telling direct mechanism, which we detail later, let $Q(\theta) = \{c_0(\theta), c_1(\theta, \xi_i), y_0(\theta), y_1(\theta, \xi_i)\}_{\xi_i \in \Xi}$ denote the allocations that θ-skilled agents choose. The expected utility of these agents at $t = 1$ is

$$\min_{m_{M,\theta}(\xi_i)} \left\{ \sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i) \left[u(c_1(\theta, \xi_i)) - h\left(\frac{y_1(\theta, \xi_i)}{\omega(M, \theta, \xi_i)} \right) \right] + \eta D(\tilde{p}_{M,\theta}||p) \right\},$$

subject to $\sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i) = 1$. The first term in the objective function is the expected utility of the θ-skilled agents evaluated under the alternative distribution. The second term is a penalty that penalizes alternative distributions that are too distant from the reference distribution in terms of relative entropy. The parameter η represents the agents' confidence in the reference model. Higher confidence in the reference distribution corresponds to larger penalties. As $\eta \to \infty$, our model collapses to the standard expected utility model without ambiguity or uncertainty.

Solving the minimization problem above gives us

$$m_{M,\theta}(\xi_i) = \frac{\exp\left(-\frac{1}{\eta} \left[u(c_1(\theta, \xi_i)) - h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)} \right) \right] \right)}{\sum_{i=0}^{S} p(\xi_i) \exp\left(-\frac{1}{\eta} \left[u(c_1(\theta, \xi_i)) - h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)} \right) \right] \right)}.$$

Ambiguity-averse agents choose the alternative distribution to guard against misspecifica-

[4] The relative entropy can be thought of as measuring the 'distance' between two distributions, but it is technically a *divergence*. See Cover and Thomas (1991) for more information on relative entropy.

[5] Strzalecki (2011) provides an axiomatic foundation for the multiplier preferences.

tion by assigning higher probability than the reference on low-utility events and placing smaller weight than the reference on high-utility events. As a result, the agents' pessimism is endogenously determined by the utility in $t = 1$.

Substituting the optimal $m_{M,\theta}(\xi_i)$ into the utility function yields the following lifetime discounted expected utility for agents with skill θ :

$$U_0(\theta) = u(c_0(\theta)) - h\left(\frac{y_0(\theta)}{\theta}\right) + \beta \tilde{U}_1\left(\{c_1(\theta,\xi_i), y_1(\theta,\xi_i)\}_{\xi_i \in \Xi}; M, \theta\right), \tag{1}$$

where

$$\tilde{U}_1\left(\{c_1(\theta,\xi_i), y_1(\theta,\xi_i)\}_{\xi_i \in \Xi}; M, \theta\right) = -\eta \ln\left[\sum_{i=0}^{S} p(\xi_i) \exp\left(-\frac{1}{\eta}\left[u(c_1(\theta,\xi_i)) - h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\right]\right)\right].$$

Due to ambiguity aversion, consumption and leisure (labor supply) are not additively separable in $t = 1$. Equally important is the fact that $\tilde{U}_1$ is also non-separable across states. Despite its complexity, $\tilde{U}_1$ is strictly concave and strictly increasing in consumption and leisure.[6]

3.2 Belief Manipulation

With endogenous pessimism, policies not only determine allocations but can also influence beliefs—a key insight developed in Karantounias (2013, 2023) and Ferriere and Karantounias (2019). Therefore, a key ingredient in determining the optimal policies is the elasticity of the subjective likelihood with respect to allocations, which are defined as

$$\epsilon_c^m(M,\theta,\xi_i,\xi_k) = \frac{\partial m_{M,\theta}(\xi_i)}{\partial c_1(\theta,\xi_k)} \frac{c_1(\theta,\xi_k)}{m_{M,\theta}(\xi_i)} \text{ and } \epsilon_y^m(M,\theta,\xi_i,\xi_k) = \frac{\partial m_{M,\theta}(\xi_i)}{\partial y_1(\theta,\xi_k)} \frac{y_1(\theta,\xi_k)}{m_{M,\theta}(\xi_i)}.$$

[6]It is immediate that $\tilde{U}_i$ is increasing in consumption c_1 and decreasing in output y_1. To see why $\tilde{U}_1$ is strictly concave, note that $\exp\left(-\frac{1}{\eta}\left[u(c_1) - h\left(\frac{y_1}{\omega}\right)\right]\right)$ is strictly log-convex. The sum of strictly log-convex functions is still strictly log-convex (Boyd and Vandenberghe, 2004), so the strict concavity of $\tilde{U}_1$ follows.

We refer to these as *belief elasticities*. When $\xi_i = \xi_k$, they are *in-state belief elasticities*, and when $\xi_i \neq \xi_k$, they are *cross-state belief elasticities*. For notational convenience, define $\epsilon_q^m(M, \theta, \xi_i) \equiv \epsilon_q^m(M, \theta, \xi_i, \xi_i)$. The belief elasticities determine the sensitivity of the subjective likelihoods with respect to changes in consumption and output, i.e., the effectiveness of belief manipulation in the design of policies. When there is full confidence in the reference distribution, then the belief elasticities are perfectly inelastic for all skills and shocks.

The following lemma states that the in-state (cross-state) belief elasticity of consumption ϵ_c^m is negative (positive), while the in-state (cross-state) belief elasticity of output ϵ_y^m is positive (negative). It also demonstrates that when the allocation changes under shock ξ_i, there is a uniform cross-state impact on beliefs for all other shocks.

Lemma 1 *For any M, $\theta \in \Theta$, and $\xi_i, \xi_k \in \Xi$ with $\xi_i \neq \xi_k$,*

$$\epsilon_c^m(M, \theta, \xi_i) = -\frac{c_1(\theta, \xi_i)}{\eta} u'(c_1(\theta, \xi_i)) [1 - p(\xi_i) m_{M,\theta}(\xi_i)] \leq 0,$$

$$\epsilon_y^m(M, \theta, \xi_i) = \frac{y_1(\theta, \xi_i)}{\eta \omega(M, \theta, \xi_i)} h'\left(\frac{y_1(\theta, \xi_i)}{\omega(M, \theta, \xi_i)}\right) [1 - p(\xi_i) m_{M,\theta}(\xi_i)] \geq 0,$$

$$\epsilon_c^m(M, \theta, \xi_k, \xi_i) = \frac{c_1(\theta, \xi_i)}{\eta} u'(c_1(\theta, \xi_i)) p(\xi_i) m_{M,\theta}(\xi_i) \geq 0,$$

$$\epsilon_y^m(M, \theta, \xi_k, \xi_i) = -\frac{y_1(\theta, \xi_i)}{\eta \omega(M, \theta, \xi_i)} h'\left(\frac{y_1(\theta, \xi_i)}{\omega(M, \theta, \xi_i)}\right) p(\xi_i) m_{M,\theta}(\xi_i) \leq 0.$$

Lemma 1 allows us to connect the belief elasticities to well-known economic objects. It also introduces new economic effects. To see this, let

$$\tilde{U}_{1,q}(M, \theta, \xi_i) = \frac{\partial \tilde{U}_1(M, \theta)}{\partial q_1(\theta, \xi_i)} \text{ and } \tilde{U}_{1,qr}(M, \theta, \xi_i, \xi_k) = \frac{\partial \tilde{U}_{1,q}(M, \theta, \xi_i)}{\partial r_1(\theta, \xi_k)}$$

with partial derivatives of other functions expressed analogously. For notational simplicity, let $\tilde{U}_{1,qr}(M, \theta, \xi_i) \equiv \tilde{U}_{1,qr}(M, \theta, \xi_i, \xi_i)$.

From Lemma 1, the relative risk aversion of $\tilde{U}_1$ for skill θ and shock ξ_i is

$$-\frac{\tilde{U}_{1,cc}(M, \theta, \xi_i) c_1(\theta, \xi_i)}{\tilde{U}_{1,c}(M, \theta, \xi_i)} = \sigma_1(\theta, \xi_i) - \epsilon_c^m(M, \theta, \xi_i), \tag{2}$$

where σ_t is the relative risk aversion of u at t.[7] By Lemma 1, the in-state belief elasticity of consumption ϵ_c^m is negative, so (2) shows that ambiguity aversion adds more curvature to the utility $\tilde{U}_1$ than the baseline full-confidence case. Intuitively, in addition to the existing curvature in the consumption utility u (characterized by σ_1), a small increase in consumption under shock ξ_i also lowers the probability agents place on shock ξ_i, which decreases the marginal utility of consumption for $\tilde{U}_1$ in state ξ_i.

The degree of complementarity between consumption and labor is pinned down by ϵ_y^m :

$$\frac{\tilde{U}_{1,cy}\left(M,\theta,\xi_i\right)y_1\left(\theta,\xi_i\right)}{\tilde{U}_{1,c}\left(M,\theta,\xi_i\right)}=\epsilon_y^m\left(M,\theta,\xi_i\right). \tag{3}$$

By Lemma 1, the in-state belief elasticity of output ϵ_y^m is positive, so (3) shows that consumption and labor are complements (or, consumption and leisure are substitutes) for ambiguity averse agents.[8] This is because an increase in output y_1 under shock ξ_i raises the probability weight the agents place on shock ξ_i occurring, which correspondingly increases the marginal utility of consumption for $\tilde{U}_1$ in state ξ_i. In contrast, when agents are fully confident in the reference distribution, consumption and labor are additively separable.

Furthermore, since ambiguity aversion also induces non-separability across states, a change in consumption in one state affects the marginal utilities of consumption in all other states. We define the *cross-state complementarity in consumption* between technological shocks ξ_i and ξ_k as

$$\frac{\tilde{U}_{1,cc}\left(M,\theta,\xi_i,\xi_k\right)c_1\left(\theta,\xi_k\right)}{\tilde{U}_{1,c}\left(M,\theta,\xi_i\right)}.$$

Intuitively, an increase in consumption under shock ξ_k lowers the agent's belief that ξ_k will occur, thereby raising the belief assigned to other states—the probability of ξ_i. A higher subjective probability of ξ_i in turn increases the marginal utility of consumption under ξ_i.

Similarly, we can also define the *cross-state consumption-labor substitutability* between

[7] Precisely, $\sigma_0\left(\theta\right)=-\frac{u''(c_0(\theta))c_0(\theta)}{u'(c_0(\theta))}$ and $\sigma_1\left(\theta,\xi_i\right)=-\frac{u''(c_1(\theta,\xi_i))c_1(\theta,\xi_i)}{u'(c_1(\theta,\xi_i))}$.

[8] Note that $\tilde{U}_{1,cl}\left(M,\theta,\xi_i\right)l_1\left(\theta,\xi_i\right)=\tilde{U}_{1,cy}\left(M,\theta,\xi_i\right)y_1\left(\theta,\xi_i\right)$. Therefore, the complementarity between consumption and output is equivalent to the complementarity between consumption and labor.

technological shocks ξ_i and ξ_k as

$$-\frac{\tilde{U}_{1,cy}\left(M,\theta,\xi_i,\xi_k\right)y_1\left(\theta,\xi_k\right)}{\tilde{U}_{1,c}\left(M,\theta,\xi_i\right)}.$$

An output increase under shock ξ_k raises the agent's subjective belief that shock ξ_k would occur. Hence, the agent places less weight on other shocks such as ξ_i, which in turn lowers the marginal utility of consumption in state ξ_i. As a result, labor is substitutable with consumption of other states.

Given the mechanism, these cross-state interactions are related to the cross-state belief elasticities. By Lemma 1, we can show that the larger the cross-state belief elasticity of consumption, the stronger the complementarity of consumptions across states, while an analogous relationship holds between the cross-state belief elasticity of output and the cross-state consumption-labor substitutability. The following corollary summarizes these relationships.

Corollary 1 *For any* M, $\theta\in\Theta$, *and* $\xi_i,\xi_k\in\Xi$ *with* $\xi_i\neq\xi_k$,

$$\frac{\tilde{U}_{1,cc}\left(M,\theta,\xi_i,\xi_k\right)c_1\left(\theta,\xi_k\right)}{\tilde{U}_{1,c}\left(M,\theta,\xi_i\right)}=\epsilon_c^m\left(M,\theta,\xi_i,\xi_k\right),$$

and

$$\frac{\tilde{U}_{1,cy}\left(M,\theta,\xi_i,\xi_k\right)y_1\left(\theta,\xi_k\right)}{\tilde{U}_{1,c}\left(M,\theta,\xi_i\right)}=\epsilon_y^m\left(M,\theta,\xi_i,\xi_k\right).$$

Finally, since changes in machine investment can alter the subjective likelihoods. We define the following elasticity:

$$\epsilon_M^m\left(M,\theta,\xi_i\right)=\frac{\partial m_{M,\theta}\left(\xi_i\right)}{\partial M}\frac{M}{m_{M,\theta}\left(\xi_i\right)},$$

where ϵ_M^m is the elasticity of the subjective likelihood with respect to machine. It follows that ϵ_M^m can be expressed as

$$\epsilon_M^m\left(M,\theta,\xi_i\right)=-\sum_{k=0}^{S}\epsilon_y^m\left(M,\theta,\xi_i,\xi_k\right)\epsilon_M^{\omega}\left(M,\theta,\xi_k\right),$$

which will be useful in characterizing the optimal investment in M.

3.3 Planner's Problem

Due to information asymmetry, the optimal allocations must be incentive compatible. By the revelation principle, we can first analyze a direct truth-telling mechanism—agents report their private information truthfully to the government—to characterize constrained efficient allocations. Then, we can use the results of the direct mechanism as a blueprint to decentralize the optimum. Therefore, instead of choosing the tax functions, the direct mechanism specifies the allocations for all agents $\{Q(\theta)\}_{\theta\in\Theta}$. Also, the set of allocations $\{Q(\theta)\}_{\theta\in\Theta}$ must be incentive compatible: for any $\theta, \theta' \in \Theta$,

$$u(c_0(\theta)) - h\left(\frac{y_0(\theta)}{\theta}\right) + \beta\tilde{U}_1\left(\{c_1(\theta,\xi_i), y_1(\theta,\xi_i)\}_{\xi_i\in\Xi}; M, \theta\right)$$
$$\geq u(c_0(\theta')) - h\left(\frac{y_0(\theta')}{\theta}\right) + \beta\tilde{U}_1\left(\{c_1(\theta',\xi_i), y_1(\theta',\xi_i)\}_{\xi_i\in\Xi}; M, \theta\right). \quad (4)$$

The incentive constraints (4) ensure that agents report their skills truthfully at $t = 0$.

We adopt the first-order approach and replace the incentive constraints (4) with the following envelope condition: for any $\theta \in \Theta$,

$$\frac{\partial U_0(\theta)}{\partial\theta} = \frac{y_0(\theta)}{\theta^2}h'\left(\frac{y_0(\theta)}{\theta}\right) - \beta\sum_{i=0}^{S}\tilde{U}_{1,y}(M,\theta,\xi_i)\, y_1(\theta,\xi_i)\,\frac{\epsilon_\theta^\omega(M,\theta,\xi_i)}{\theta}. \quad (5)$$

The envelope condition (5) is a necessary condition for incentive compatibility. In standard models, the envelope condition along with monotonicity conditions (y_t increases with θ) are also sufficient for incentive compatibility. However, in our setting, the envelope and the monotonicity conditions are not enough to guarantee sufficiency: Even with monotonicity, ambiguity-averse agents may still want to misreport to decrease the probability weighting on low-utility states. Despite this, we follow the literature and solve a relaxed planner's problem where we ignore the monotonicity condition and replace the incentive constraints (4) with its necessary condition (5), and then check the solutions for incentive compatibility numerically.

We consider a utilitarian government with full commitment that shares the same doubts regarding the reference distribution as the agents. The government solves a relaxed problem

by choosing the allocations $\{Q(\theta)\}_{\theta\in\Theta}$ and machine M to maximize the utilitarian welfare

$$\int_{\Theta} U_0(\theta)\, dF(\theta)$$

subject to (5) and the resource constraints: for any $\xi_i \in \Xi$,

$$\int_{\Theta} \left\{ [y_0(\theta) - c_0(\theta)] + \frac{1}{R} [y_1(\theta, \xi_i) - c_1(\theta, \xi_i)] \right\} dF(\theta) \geq \phi M, \tag{6}$$

where R is the exogenous gross rate of return. For simplicity, we assume that $\beta R = 1$. Note here that the resource constraint (6) implies that the technological shock ξ_i is uninsurable.

3.4 Full-Information Benchmark

Before presenting our main results, it is helpful to first see how ambiguity aversion shapes the planner's problem in the absence of incentive constraints by examining a full-information environment—skill θ is publicly observable. Although allocations no longer need to satisfy incentive compatibility, ambiguity aversion directly shapes the allocations, leading to a full-information optimum that differs from the standard case.

Proposition 1 *The full-information optimum satisfies, for any $\theta \in \Theta$, (i) Belief-adjusted Euler equation: $u'(c_0(\theta)) = \sum_{i=0}^{S} p(\xi_i)\, m_{M,\theta}(\xi_i)\, u'(c_1(\theta, \xi_i))$, (ii) Full insurance in $t = 0$: $c_0(\theta) = c_0(\theta')$ for any $\theta' \in \Theta$, (iii) Efficient output: $\theta u'(c_0(\theta)) = h'(y_0(\theta)/\theta)$ and $\omega(M, \theta, \xi_i)\, u'(c_1(\theta, \xi_i)) = h'(y_1(\theta, \xi_i)/\omega(M, \theta, \xi_i))$ for any $\xi_i \in \Xi$, (iv) Belief-adjusted insurance in $t = 1$: $m_{M,\theta}(\xi_i)\, u'(c_1(\theta, \xi_i)) = m_{M,\theta'}(\xi_i)\, u'(c_1(\theta', \xi_i))$ for all $\xi_i \in \Xi$ and $\theta' \in \Theta$, (v) Efficient machine investment:*

$$\phi u'(c_0(\theta)) = \frac{1}{RM} \int_{\Theta} \sum_{i=0}^{S} p(\xi_i)\, m_{M,\theta}(\xi_i)\, y_1(\theta, \xi_i)\, u'(c_1(\theta, \xi_i))\, \epsilon_M^{\omega}(M, \theta, \xi_i)\, dF(\theta).$$

Proposition 1 shows that in the full-information efficient optimum, production is efficient in every period and agents are fully insured in $t = 0$. These outcomes mirror those under standard preferences. However, ambiguity aversion introduces new forces into the intertemporal condition and alters the full-insurance property in $t = 1$. Intuitively, because beliefs vary

with skills, the Euler equation accounts for this heterogeneity in probability assessments. These differing beliefs also redefine the notion of full insurance: If θ-skilled agents assign a higher likelihood to shock ξ_i than θ'-skilled agents, then allocating more consumption to θ-skilled agents when ξ_i occurs is welfare increasing. In the next section, we demonstrates how ambiguity aversion interacts with the incentive problem.

The final part of Proposition 1 characterizes the optimal machine investment in the full-information efficient optimum. The left-hand side of the condition is the marginal cost of investment in terms of utility. The right-hand side is the economy-wide marginal benefit of investment per unit of machine M. The marginal utility gain from machine via higher output is captured by $y_1(\theta, \xi_i) u'(c_1(\theta, \xi_i)) \epsilon_M^{\omega}(M, \theta, \xi_i)$, which is a standard effect.

4 Theoretical Properties

This section characterizes the labor and savings wedges and the optimal investment in technological advancement M, which provide the foundations for any policy decentralization.

The labor wedges characterize the optimal distortions in the marginal rate of substitution between consumption and leisure, which are defined as

$$\tau_0^w(\theta) = 1 - \frac{h'(l_0)}{\theta u'(c_0)} \text{ and } \tau_1^w(M, \theta, \xi_i) = 1 - \frac{h'(l_1)}{\omega(M, \theta, \xi_i) u'(c_1)}.$$

Since the equilibrium wage is equal to the marginal productivity of labor in a competitive labor market, if $\tau_t^w \neq 0$, then agents are not working at the efficient level. In particular, if $\tau^w > 0$ ($\tau^w < 0$), then there is an under-supply (over-supply) of labor given the market wage. In essence, the labor wedge provides us with the implicit tax on income.

The savings wedge captures the distortion in the intertemporal choice between consumption in $t = 0$ and $t = 1$. We define the savings wedge τ^s as

$$\tau^s = 1 - \frac{u'(c_0)}{\beta R \tilde{\mathbb{E}}_\theta [u'(c_1)]},$$

where $\tilde{\mathbb{E}}_\theta$ denotes the expectations operator with respect to θ-skilled agents' subjective beliefs. If $\tau^s > 0$ ($\tau^s < 0$), then savings is inefficiently low (high) from the agent's perspective.

The savings wedge provides us with the implicit tax on savings.

4.1 Labor Wedge

We begin by characterizing the labor distortions. Let

$$\gamma_0(\theta) = \frac{h'\left(\frac{y_0(\theta)}{\theta}\right)}{\frac{y_0(\theta)}{\theta}h''\left(\frac{y_0(\theta)}{\theta}\right)} \text{ and } \gamma_1(M,\theta,\xi_i) = \frac{h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}{\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}h''\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}$$

define the Frisch elasticity of labor supply in $t=0$ and $t=1$, respectively. Since there is no ambiguity regarding the state of the economy in $t=0$, the labor wedge in $t=0$ satisfies the standard Mirrlees-Diamond-Saez formula (Mirrlees, 1971; Diamond, 1998; Saez, 2001):

$$\frac{\tau_0^w(\theta)}{1-\tau_0^w(\theta)} = A_0(\theta) B_0(\theta) C_0(\theta),$$

where

$$A_0(\theta) = \frac{1-F(\theta)}{\theta f(\theta)},$$

$$B_0(\theta) = 1 + \frac{1}{\gamma_0(\theta)},$$

$$C_0(\theta) = \int_\theta^{\overline{\theta}} \exp\left(\int_\theta^x \sigma_0(z)\frac{c_{0,z}(z)}{c_0(z)}dz\right)\left[1 - \frac{u'(c_0(x))}{\sum_{i=0}^S \mu(\xi_i)}\right]\frac{f(x)}{1-F(\theta)}dx,$$

and $\mu(\xi_i)$ is the multiplier on the resource constraint (6), which is a function of ξ_i since the technological shock is uninsurable. The sum of the multiplier $\mu(\xi_i)$ is associated with the cost of increasing the average utility from consumption in $t=0$.[9]

The Mirrlees-Diamond-Saez formula captures the economic forces that determine the cost and benefit of distorting labor supply. To understand it, first note that $A_0(\theta)$ captures the balance between the reduction in total output, which is proportional to $\theta f(\theta)$ due to increased labor distortions for agents at this skill level, and the relaxation of incentive

[9] At the optimum, $\frac{1}{\sum_{i=0}^S \mu(\xi_i)} = \int_\Theta \frac{1}{u'(c_0(\theta))}dF(\theta)$. The inverse marginal utility represents the consumption cost of an incremental utility increase, so the cost of enhancing average utility at $t=1$ equals to $\frac{1}{\sum_{i=0}^S \mu(\xi_i)}$.

constraints for higher-skilled agents. Meanwhile, the increased labor distortion on θ-skilled agents lowers their labor supply, and $B_0(\theta)$ captures this decrease. Intuitively, it shows that the optimal labor distortion is inverse to the labor supply elasticity γ_0.

Finally, $C_0(\theta)$ reflects the government's motive to provide insurance against skill disparities. When the cost of elevating average utility is comparatively low against the cost of improving the utility for agents with skill x, that is, $\left[\sum_{i=0}^{S}\mu(\xi_i)\right]^{-1} < \left[u'(c_0(x))\right]^{-1}$, then the advantages of raising labor distortions for θ-skilled agents to ease the incentive constraints for higher productivity types ($x \geq \theta$) surpass the associated costs. Also, the exponential term in $C_0(\theta)$, which is greater than one, captures the income effect on higher-skilled agents due to the labor distortion on θ-skilled agents. Intuitively, increasing the labor distortion on θ-skilled agents relaxes the incentive constraints of higher-skilled agents and, through the resulting income effect, induces them to work more. When the coefficient of relative risk aversion σ_0 is higher, labor distortions raise the marginal utility of consumption more sharply, further relaxing those incentive constraints and magnifying the income effect. As a result, the labor wedge for θ-skilled agents escalates with the income effect on higher-skilled agents. The next proposition shows how ambiguity augments the labor wedge in $t=1$.

Theorem 1 *The labor wedge in $t=1$ satisfies, for any $\theta \in \Theta$, and $\xi_i \in \Xi$,*

$$\frac{\tau_1^w(M,\theta,\xi_i)}{1-\tau_1^w(M,\theta,\xi_i)} = A_1(M,\theta,\xi_i)\,B_1(M,\theta,\xi_i)\,C_1(M,\theta,\xi_i), \tag{7}$$

where

$$A_1(M,\theta,\xi_i) = A_0(\theta)\,\epsilon_\theta^{\omega}(M,\theta,\xi_i),$$

$$B_1(M,\theta,\xi_i) = 1 + \frac{1}{\gamma_1(M,\theta,\xi_i)},$$

$$C_1(M,\theta,\xi_i) = \int_\theta^{\bar\theta} \exp\left(\int_\theta^x \left[\sigma_1(z,\xi_i)\frac{c_{1,z}(z,\xi_i)}{c_1(z,\xi_i)} - \Psi(M,z,\xi_i)\right]dz\right)\frac{\left[1-\frac{\tilde U_{1,c}(M,x,\xi_i)}{\mu(\xi_i)}\right]}{1-F(\theta)}dF(x)$$

with $\Psi(M,\theta,\xi_i) = \sum_{k=0}^{S}\epsilon_c^m(M,\theta,\xi_i,\xi_k)\frac{c_{1,\theta}(\theta,\xi_k)}{c_1(\theta,\xi_k)} + \sum_{k=0}^{S}\epsilon_y^m(M,\theta,\xi_i,\xi_k)\frac{y_{1,\theta}(\theta,\xi_k)}{y_1(\theta,\xi_k)}$.

Before we explain Theorem 1, first consider the standard setting with full confidence in

the reference distribution ($\eta \to \infty$). Two important things to keep in mind for the standard setting. First, it is optimal for the government to smooth the cost of providing incentives, leading to a distorted labor supply in $t = 1$ that mirrors the explanation previously offered for τ_0^w even without dynamic private information. Second, if the technological shock can be fully insured against, then $c_0(\theta) = c_1(\theta, \xi_i)$ for any $\theta \in \Theta$ and $\xi_i \in \Xi$. However, since the technological shock affects the marginal utility for leisure in $t = 1$ (γ_1 may differ across shocks), the labor wedges are generally not equal across periods even with full insurance against the technology shock.

Theorem 1 shows the influence of the technological shocks on the labor wedge, which is captured by $A_1(M, \theta, \xi_i)$. The component $A_1(M, \theta, \xi_i)$ is not affected by ambiguity, but it highlights how the labor wedge increases with the wage elasticity of skill $\epsilon_\theta^\omega(M, \theta, \xi_i)$. When ϵ_θ^ω is large, small differences in skill lead to big changes in productivity. This raises the temptation for high-skilled agents to mimic downwards. As a result, the distortion in labor is increased to maintain incentive compatibility. In contrast, if $\epsilon_\theta^\omega = 0$ for all agents, then the marginal productivity is equalized, eliminating the incentive to distort the labor supply for any particular group of agents.

From Theorem 1, the only component affected by ambiguity aversion is $C_1(M, \theta, \xi_i)$. There are two main changes. First, since the beliefs are endogenous, they also influence the marginal utility of consumption: $\tilde{U}_{1,c}(M, \theta, \xi_i) = p(\xi_i) m_{M,\theta}(\xi_i) u'(c_1(\theta, \xi_i))$. Second, due to the non-separability across states and between consumption and labor, the labor distortion on θ-skilled agents introduces an additional economic force in the income effect on labor supply of higher-skilled agents. This novel effect enters through $\Psi(M, \theta, \xi_i)$, which is comprised of belief elasticities. In contrast, under full confidence ($\eta \to \infty$), we have $m_{M,\theta}(\xi_1) = 1$ and $\Psi(M, \theta, \xi_i) = 0$ for all $\theta \in \Theta$ and $\xi_i \in \Xi$. The rest of our discussion will focus on the new component $\Psi(M, \theta, \xi_i)$.

By (2), (3), and Corollary 1, the belief elasticities in $\Psi(M, \theta, \xi_i)$ can be expressed with familiar objects. In particular, the income effect on the labor supply of higher-skilled agents enters the labor wedge in a fashion similar to the standard Mirrlees-Diamond-Saez formula

with non-separable utility:

$$\sigma_1(\theta,\xi_i)\frac{c_{1,\theta}(\theta,\xi_i)}{c_1(\theta,\xi_i)}-\Psi(M,\theta,\xi_i)=\sum_{k=0}^{S}\left[-\frac{\tilde{U}_{1,cc}(M,\theta,\xi_i,\xi_k)\,c_1(\theta,\xi_k)}{\tilde{U}_{1,c}(M,\theta,\xi_i)}\frac{c_{1,\theta}(\theta,\xi_k)}{c_1(\theta,\xi_k)}\right.$$
$$\left.-\frac{\tilde{U}_{1,cy}(M,\theta,\xi_i,\xi_k)\,y_1(\theta,\xi_k)}{\tilde{U}_{1,c}(M,\theta,\xi_i)}\frac{y_{1,\theta}(\theta,\xi_k)}{y_1(\theta,\xi_k)}\right]. \quad (8)$$

The innovation of our work is that the income effect works through two channels, the *in-state channel* and the *cross-state channel*. The in-state channel considers the income effect within a given shock, while the cross-state channel takes into account the income effect emanating from other shocks.

We begin by examining how the in-state effect influences the labor wedge. This in-state channel enters much as in the standard Mirrlees-Diamond-Saez formula with utility that is non-separable in consumption and labor. Indeed, if we temporarily ignore the cross-state terms, (7) can be translated to the intratemporal component in Golosov et al. (2016). Despite the similarities, ambiguity aversion introduces a new interpretation of the income effect, which is composed of two distinct parts.

The first part shows how the curvature of $\tilde{U}_1$ with respect to consumption has a direct effect on labor supply. Intuitively, if the relative risk aversion of $\tilde{U}_1$ is large, then the marginal utility rises rapidly when consumption drops due to an increase in the labor wedge, which relaxes the incentive constraints of higher-skilled agents. From Lemma 1 and (2), the relative risk aversion of ambiguity averse agents is larger than the standard model with full confidence: In addition to the curvature of the consumption utility u, the endogenous beliefs also adjust with consumption in a way that strengthens the risk aversion of ambiguity averse agents. Therefore, the in-state belief elasticity of consumption $\epsilon_c^m(M,\theta,\xi_i)$ increases the labor distortion on θ-skilled agents to raise the income effect on labor supply of higher-skilled agents, which leads them to work more.

The second part illustrates how the complementarity between consumption and labor influences labor decisions. Intuitively, if consumption and labor are complements, then consumption and leisure are substitutes. As a result, when consumption falls due to a rise in

labor distortion, then agents would tend to respond by increasing leisure through a decrease in labor supply, which tightens the incentive constraints on higher-skilled agents. From (3), the complementarity between consumption and labor in our model is simply the in-state belief elasticity of output $\epsilon_y^m(M, \theta, \xi_i)$, which is positive (Lemma 1). Therefore, the in-state belief elasticity of output counteracts the in-state belief elasticity of consumption: The in-state belief elasticity of output $\epsilon_y^m(M, \theta, \xi_i)$ lowers the labor distortion on θ-skilled agents to limit the decrease in labor supply of higher-skilled agents.

We now consider how the novel cross-state channel influences labor distortions. It captures how the income effect on labor supply under shock ξ_i is altered by changes in allocation under other technological shocks through a shift in subjective beliefs. Consider changes in consumption and output under shock ξ_k. When consumption increases in state ξ_k, the subjective likelihood $m_{M,\theta}(\xi_k)$ falls, which increases the probability of ξ_i and thereby affects the marginal utility of consumption in ξ_i. Similarly, when output rises under ξ_k, it raises $m_{M,\theta}(\xi_k)$, reducing the weight assigned to ξ_i and lowering its marginal utility of consumption. These mechanisms generate cross-state complementarity in consumption (driven by changes in consumption under ξ_k) and cross-state substitutability between consumption and labor (driven by changes in output under ξ_k). Thus, following the logic for in-state belief elasticities, the cross-state belief elasticity of consumption $\epsilon_c^m(M, \theta, \xi_i, \xi_k)$ decreases the labor distortion on θ-skilled agents, while the cross-state belief elasticity of output $\epsilon_y^m(M, \theta, \xi_i, \xi_k)$ has the opposite effect. Therefore, the cross-state channel is similar to the in-state channel but with opposite signs.

The discussion above shows how the optimal labor distortion uses belief manipulation to enhance welfare, while illustrating its limitations. The in-state channel is counteracted by the cross-state channel, and within each channel, the belief elasticity of consumption is offset by the belief elasticity of output.

4.2 Savings Wedge

Since ambiguity aversion induces non-separability in consumption and leisure and across states, the Euler equation may not hold.[10] The following theorem demonstrates how ambiguity augments the intertemporal margin.

Theorem 2 *The constrained efficient allocation satisfies the ambiguity-tilted Euler equation: For any* $\theta \in \Theta$,

$$u'\left(c_0\left(\theta\right)\right) = \sum_{i=0}^{S} p\left(\xi_i\right) m_{M,\theta}\left(\xi_i\right) \Gamma\left(M,\theta,\xi_i\right) u'\left(c_1\left(\theta,\xi_i\right)\right), \tag{9}$$

where

$$\Gamma\left(M,\theta,\xi_i\right) = \left[1 - \frac{1}{B_1\left(M,\theta,\xi_i\right)} \frac{\tau_1^w\left(M,\theta,\xi_i\right)}{1-\tau_1^w\left(M,\theta,\xi_i\right)} \sum_{k=0}^{S} \epsilon_y^m\left(M,\theta,\xi_i,\xi_k\right) \frac{\epsilon_\theta^\omega\left(M,\theta,\xi_k\right)}{\epsilon_\theta^\omega\left(M,\theta,\xi_i\right)}\right]^{-1}. \tag{10}$$

Theorem 2 shows how the optimal intertemporal distortion is affected by ambiguity aversion and private information. We can see that if agents had full confidence in the reference distribution ($\eta \to \infty$), then $\Gamma\left(M,\theta,\xi_i\right) = m_{M,\theta}\left(\xi_i\right) = 1$ for all θ and ξ_i, so (9) becomes the standard Euler equation. Furthermore, without private information, the labor wedge is not distorted, so $\Gamma\left(M,\theta,\xi_i\right) = 1$ for all θ and ξ_i. Therefore, in a full-information environment, the Euler equation holds under subjective beliefs $p\left(\xi_i\right) m_{M,\theta}\left(\xi_i\right)$ for all θ :

$$u'\left(c_0\left(\theta\right)\right) = \sum_{i=0}^{S} p\left(\xi_i\right) m_{M,\theta}\left(\xi_i\right) u'\left(c_1\left(\theta,\xi_i\right)\right).$$

As a result, $\Gamma\left(M,\theta,\xi_i\right)$ represents the interaction between ambiguity and private information.

To understand the mechanism behind $\Gamma\left(M,\theta,\xi_i\right)$, let $\tau^k\left(M,\theta,\xi_i\right) = \Gamma\left(M,\theta,\xi_i\right) - 1$ denote the shock-dependent savings subsidy (when τ^k is positive) or tax (when τ^k is negative). The following theorem demonstrates how the belief elasticity of output determine whether savings is subsidized or taxed.

[10] It is well-known that the Euler equation does not hold when utility is non-separable in consumption and leisure (Farhi and Werning, 2013).

Theorem 3 *For any* $\theta \in \Theta$, *and* $\xi_i \in \Xi$,

$$\operatorname{sgn}\left[\tau^k\left(M,\theta,\xi_i\right)\right] = \operatorname{sgn}\left[\sum_{k=0}^{S} \epsilon_y^m\left(M,\theta,\xi_i,\xi_k\right)\epsilon_\theta^\omega\left(M,\theta,\xi_k\right)\right],$$

and $\tau^k\left(M,\theta,\xi_i\right) = 0$ *when* $\theta \in \left\{\underline{\theta},\overline{\theta}\right\}$.

Theorem 3 follows from the well-known fact that the optimal labor wedges are positive and the in-state belief elasticity of output is strictly positive while the cross-state ones are negative (Lemma 1). From (10), we can see that the savings of the least skilled and most skilled agents are not distorted due to the fact that their labor wedges are zero for both periods—the well-known no-distortion at the top and bottom result.

In our setting, the sign of τ^k depends on which effect is stronger: the in-state belief elasticity of output or the cross-state belief elasticity of output, both weighted by the wage elasticity of skill (which captures the effectiveness of incentivizing higher-skilled agents to work). By (3), the in-state belief elasticity of output measures the complementarity between consumption and output within the same shock ξ_i. A large in-state belief elasticity of output thus indicates that a savings subsidy in state ξ_i would increase consumption and thereby maintain the incentives to work in ξ_i. Conversely, if the cross-state belief elasticity of output dominates, then by Corollary 1, consumption under shock ξ_i is highly substitutable with output in other states. Therefore, providing a savings subsidy under shock ξ_i would reduce work incentives under those other shocks, so Theorem 3 recommends a tax on savings instead.

Another part of $\Gamma\left(M,\theta,\xi_i\right)$ is the dependence of the savings subsidy on the labor wedge and Frisch elasticity of labor supply. To understand how labor distortions influence intertemporal distortions, note that the marginal impact of the savings subsidy Γ is bigger for agents whose labor distortions are large. In essence, the savings subsidy substitutes the labor distortions with intertemporal distortions, which is why the subsidy is increasing in the labor wedge. Following a similar logic, since labor distortions are more costly for agents whose Frisch elasticity is high, the savings subsidy Γ also increases with the Frisch elasticity of labor supply γ_1 to help lower the labor wedge.

We can rewrite $\Gamma\left(M,\theta,\xi_i\right)$ and reinterpret it in terms of the labor wedge in $t=0$, the degree of complementarity between skill θ and consumption in $t=1$, and the income effect

on labor supply in $t = 1$ (8):

$$\Gamma(M, \theta, \xi_i) = 1 + \frac{1}{B_0(\theta)} \frac{\tau_0^w(\theta)}{1 - \tau_0^w(\theta)} \times \left[\left(-\frac{\theta \tilde{U}_{1,\theta c}(M, \theta, \xi_i)}{\tilde{U}_{1,c}(M, \theta, \xi_i)} \right) - \theta \left(\sigma_1(\theta, \xi_i) \frac{\frac{\partial c_1(\theta, \xi_i)}{\partial \theta}}{c_1(\theta, \xi_i)} - \Psi(M, \theta, \xi_i) \right) \right], \quad (11)$$

where the skill-consumption complementarity is defined as $\theta \tilde{U}_{1,\theta c}(M, \theta, \xi_i) / \tilde{U}_{1,c}(M, \theta, \xi_i)$. Furthermore, we can show that

$$-\frac{\theta \tilde{U}_{1,\theta c}(M, \theta, \xi_i)}{\tilde{U}_{1,c}(M, \theta, \xi_i)} = \theta \left[\sigma_1(\theta, \xi_i) \frac{\frac{\partial c_1(\theta, \xi_i)}{\partial \theta}}{c_1(\theta, \xi_i)} - \Psi(M, \theta, \xi_i) \right] + \sum_{k=0}^{S} \epsilon_y^m(M, \theta, \xi_i, \xi_k) \epsilon_\theta^\omega(M, \theta, \xi_k).$$

The details for deriving (11) and the skill-consumption complementarity are provided in the proof for Theorem 2 in Appendix A.

To understand (11), consider a fixed technological shock ξ_i. The effect of the labor wedge in $t = 0$ on Γ follows the same logic as the impact of τ_1^w on Γ in (10). The income effect on labor supply (8) influences the savings distortion due to the non-separability between consumption and labor as well as the non-separability across technological shocks. The explanation for how the income effect influences Γ mirrors the argument presented for the labor wedge in Theorem 1.

To understand how the skill-consumption complementarity affects savings distortions, we follow Hellwig (2021) in interpreting this complementarity as a measure of the government's *needs-based* redistributive motive.[11] If the skill-consumption complementarity is negative, then consumption needs decrease with skill—higher-skilled agents have a lower marginal utility of consumption. Since there is no uncertainty in $t = 0$, the complementarity between skill and consumption in $t = 0$ is zero for all agents. Therefore, higher-skilled agents prefer to frontload their consumption relative to low-skilled agents. Thus, by Theorem 3, it is optimal to subsidize savings to backload the consumption of low-skilled agents, which lowers

[11] Hellwig (2021) considers a dynamic Mirrlees taxation setup with non-separable preferences in consumption and labor, but not across states. Crucially, due to the uncertainty regarding the future, our model only has non-separable preferences in $t = 1$ and not in $t = 0$, while preferences are non-separable in every period in Hellwig (2021).

the incentives of higher-skilled agents to mimic them and relaxes the incentive constraints. The opposite holds if the skill-consumption complementarity is positive.

A novel way of implementing the savings distortion in Theorem 2 is to interpret $\Gamma(M, \theta, \xi_i)$ as the return of an asset. Note that the ambiguity-tilted Euler equation (9) can be expressed as the following asset pricing equation:

$$1 = \tilde{\mathbb{E}}\left[D(\theta, \xi_i)\, \Gamma(M, \theta, \xi_i)\right],$$

where $\tilde{\mathbb{E}}$ is the expectations operator under the alternative distribution $p(\xi_i)\, m_{M,\theta}(\xi_i)$ and $D(\theta, \xi_i) = u'(c_1(\theta, \xi_i)) / u'(c_0(\theta))$ is the stochastic discount factor for θ-skilled agents under shock ξ_i. Under this interpretation, θ-skilled agents can freely buy or sell an asset priced at 1 in $t = 0$ whose gross return is $\Gamma(M, \theta, \xi_i)$ in $t = 1$. Note that the asset pricing equation can be rewritten as

$$1 = \tilde{\mathbb{E}}\left[D(\theta, \xi_i)\right] \tilde{\mathbb{E}}\left[\Gamma(M, \theta, \xi_i)\right] + \tilde{\mathrm{cov}}\left(D(\theta, \xi_i), \Gamma(M, \theta, \xi_i)\right), \tag{12}$$

where $\tilde{\mathrm{cov}}$ is the covariance operator under the alternative distribution $p(\xi_i)\, m_{M,\theta}(\xi_i)$.

Returning to (10), we can implement Γ as an asset that depends on the agents' future marginal income tax τ_1^w. Note that, if there is no bunching, then a classical argument shows that the optimal allocation depends on skill θ through y_0. In essence, all allocations and policies can be rewritten as functions of the observed gross income y_0 instead of the unobservable skill θ. As a result, based on their income y_0 in $t = 0$, the government sells to each agent an income-dependent asset that returns $\Gamma(M, y_0, \xi_i)$ at a price of 1. Under this interpretation, Γ is the payoff to a bet on how the technology advancement affects each skill type. If the asset Γ covaries negatively with consumption c_1, then it acts as an insurance and helps agents smooth consumption. Later, in our quantitative exercise, we will examine (12) and see whether the asset return Γ covaries positively or negatively with the stochastic discount factor D.

4.3 Investment in Machines

This section characterizes optimal investment M when its effects on complementarity with labor are uncertain. The following theorem presents the main result of this subsection.

Theorem 4 *The optimal investment in $t = 0$ satisfies*

$$\int_{\Theta} \sum_{i=0}^{S} \frac{p(\xi_i) m_{M,\theta}(\xi_i)}{RMu'(c_0(\theta))} y_1(\theta, \xi_i) [1 - \tau_1^w(M, \theta, \xi_i)]$$

$$\times \Omega(M, \theta, \xi_i) u'(c_1(\theta, \xi_i)) \epsilon_M^{\omega}(M, \theta, \xi_i) dF(\theta) = \phi, \quad (13)$$

where

$$\Omega(M, \theta, \xi_i) = 1 - \frac{\epsilon_\theta^{\omega}(M, \theta, \xi_i)}{B_0(\theta)} \frac{\tau_0^w(\theta)}{1 - \tau_0^w(\theta)} \left[\left(\epsilon^H(M, \theta, \xi_i) - 1 \right) - B_1(M, \theta, \xi_i) + \frac{\epsilon_M^m(M, \theta, \xi_i)}{\epsilon_M^{\omega}(M, \theta, \xi_i)} \right].$$

From Theorem 4, in the absence of asymmetric information, the labor wedges in both periods are zero for all agents ($\tau_0^w = \tau_1^w = 0$), so (13) reverts back to the full-information efficient optimum of Proposition 1. Crucially, Theorem 4 shows that private information distorts investment through two channels.

The first channel works its way through the labor distortion in $t = 1$. Since agents only receive $1 - \tau_1^w$ fraction of the marginal utility gain from machine investment, the positive labor wedge naturally lowers the returns of investment M.

The second channel—the focus of the remainder of this subsection—is captured by Ω. It illustrates how changing aggregate investment M can help relax or tighten incentive constraints. Notably, the investment distortion increases with the labor wedge τ_0^w, the wage elasticity of skill ϵ_θ^{ω}, and the Frisch elasticity of labor supply in $t = 0$. Following the logic behind Γ in Theorem 2, the distortion in investment M can partially offset the distortions in labor supply. However, the direction of the investment distortion, whether optimal M should be increased or decreased, depends on the relative strength of the following three factors.

First, note that skill-biased technological shocks ($\epsilon^H(M, \theta, \xi_i) > 1$) tend to reduce invest-

ment M. When shocks are more likely to be skill-biased, increasing investment disproportionately raises the productivity of high-skilled relative to low-skilled individuals, thereby tightening the incentive constraints at $t = 0$. Thus, the presence of skill-biased shocks leads to a downward distortion in investment M; conversely, investment is distorted upwards for shocks that favor less-skilled agents ($\epsilon^H(M, \theta, \xi_i) < 1$).

Second, Ω is decreasing in the Frisch elasticity of labor supply in $t = 1$, implying that investment M increases when γ_1 is low. A higher level of M raises the productivity ω of all agents in $t = 1$, which in turn reduces the cost of screening productivity. To see why, consider a simple one-period example without technological shocks and two productivity types: $\theta \in \{\theta_H, \theta_L\}$, with $\theta_H > \theta_L$. At the optimum, the downward incentive compatibility constraint binds:

$$u(c_H) - h\left(\frac{y_H}{\omega(M, \theta_H)}\right) \geq u(c_L) - h\left(\frac{y_L}{\omega(M, \theta_H)}\right),$$

and let $\lambda(\theta_H) > 0$ denote the multiplier on this constraint. Then, a marginal increase in M relaxes the incentive constraint by

$$\lambda(\theta_H) \frac{\epsilon^{\omega}_M(M, \theta_H)}{M} \left[h'\left(\frac{y_H}{\omega(M, \theta_H)}\right) \frac{y_H}{\omega(M, \theta_H)} - h'\left(\frac{y_L}{\omega(M, \theta_H)}\right) \frac{y_L}{\omega(M, \theta_H)}\right] > 0.$$

Due to the convexity of h and the fact that θ_H-agents produce more output, a larger M lowers the disutility of working more for the truthful θ_H-agents than the ones who lied downwards. Thus, the temptation to misreport falls, which increases welfare. Importantly, this effect is larger when labor supply is inelastic, so optimal investment M rises as the Frisch elasticity of labor supply in $t = 1$ falls, i.e., as B_1 increases.

Finally, investment M can also play a role in manipulating beliefs regarding the technological shock, which is a novel economic force due to ambiguity. The government can choose investment M such that the subjective beliefs place smaller weights on shocks that tighten the incentive constraints. When $\epsilon^m_M(M, \theta, \xi_i)$—the elasticity of subjective probability for θ-skilled agents with respect to M—is positive, an increase in investment raises the subjective probability assigned to technological shock ξ_i. If the incentive constraint tightens when shock ξ_i is more likely, i.e., $\frac{\tau^w_0(\theta)}{1-\tau^w_0(\theta)} \epsilon^{\omega}_{\theta}(M, \theta, \xi_i)$ is large, then raising M would place greater weight

on ξ_i and magnify the cost of inducing effort in $t = 0$. As a result, when

$$\epsilon_M^m(M, \theta, \xi_i) \frac{\epsilon_\theta^\omega(M, \theta, \xi_i)}{B_0(\theta)} \frac{\tau_0^w(\theta)}{1 - \tau_0^w(\theta)}$$

is positive and large, Ω declines, reducing the marginal benefit of machine investment. In essence, optimal investment policy recommends lowering M to relax the incentive constraints by shifting weight away from the most distortionary shocks.

In the next section, we explore the quantitative properties of the optimal policy in a calibrated model, including under various specifications for the distribution of the technological shock.

5 Quantitative Analysis

In this section, we solve the model numerically and calibrate it to the U.S. data. We then present the optimal tax wedge analysis and conduct various experiments to understand the economic forces at play.

5.1 Calibration

We begin our discussion of the model calibration by assuming typical functional forms for agent preferences and set a collection of fairly standard parameter values (Table 2). These include the constant relative risk aversion utility function of the form $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ with the parameter $\sigma = 2$; as well as the labor disutility function of the form $h(\ell) = \frac{\ell^{1+\frac{1}{\kappa}}}{1+\frac{1}{\kappa}}$ with the constant Frisch elasticity of labor supply parameter set to 0.5. We also assume the discount

Table 2: Parameters set externally

Param.	**Meaning**	**Value**	**Source**
σ	Relative risk aversion	2	Standard value
κ	Frisch elasticity of labor	0.5	Standard value
β	Discount factor	0.8	15-year horizon at $r = 1.5\%$
γ	Skill-AI complementarity	2	Literature
M	AI capability	1778	Corresponds to $F(\theta) = 1 - 10^{-4}$
ϕ	Price of AI (in normative)	0.00021	$M = 1778$ w/o ambiguity in normative

factor $\beta = 0.8$ and consequently the interest rate $R = 1.25$. We think of a period here as 10-20 years; assuming a midpoint of 15 years yields an annual real return of 1.5%, roughly in line with what we have witnessed in the first quarter of the twenty-first century.

Next, we turn to the parameters that are calibrated internally in the model by targeting the results from Section 2, presented in Table 3.

Table 3: Calibrated parameters and targeted moments

Param.	**Meaning**	**Value**	**Moment**	**Model**	**Data**
Individual skill distribution					
μ_N	Mean of normal	5.24	Mean	6.00	5.62
σ_N	St. dev. of normal	3.65	St. deviation	12.33	12.45
μ_L	Mean of lognormal	0.79	Mode	4.39	4.39
σ_L	St. dev. of lognormal	0.88	10th percentile	0.99	0.96
α	Pareto shape	1.57	75/25 perc. ratio	3.18	3.19
θ^T	Threshold of Pareto	18.17	90/10 perc. ratio	11.34	11.06
ρ	Fraction of Pareto	0.47	Skewness	1.19	1.22
p_N	Probability of normal	0.51	Kurtosis	3.19	3.18
p_L	Probability of lognormal	0.21	Top 1% share	11.68%	11.44%
Technological shock and ambiguity					
η	Ambiguity aversion	0.15	Pessimism-income coeff.	-0.06	-0.06
α	Share of θ	0.14	$\text{Corr}(y_0, \frac{\mathbb{E}_p(y_1)}{y_0})$	0.00	0.00
σ_ξ	Scale parameter	0.71	Mean *pessimism*	0.21	0.19
β_ξ	Shape parameter	3.94	St. dev. *pessimism*	0.19	0.25
μ_ξ	Location parameter	0.71	$\mathbb{E}_p(\xi) - \gamma$	0.00	0.00

"Current policies" model To calibrate the distribution of individual skill θ, we use data on the distribution of earnings in the U.S. economy. However, we cannot do it directly because earnings in the model are a product of individual skill and the labor supply. For this reason, to infer the unobserved skills, we follow Paluszynski and Yu (2023) by constructing an auxiliary, positive model of the world which we call the "current policies" model. This model, while mimicking the main theoretical framework from Section 3, does not involve optimal policy and instead it admits a decentralized equilibrium with fairly realistic fiscal policy. By matching the predicted distribution of earning in this model to its data counterpart, we are able to infer the underlying distribution of individual skills.

In the "current policies" model, each agent with individual skill level θ solves the following

problem

$$\max_{c_0,\ell_0,\{c_1(\xi)\},\{\ell_1(\xi)\}} u(c_0) - h(\ell_0) + \beta(-\eta)\log\left[\sum_{i=1}^{S} p(\xi_i)\exp\left(-\frac{1}{\eta}\left[u\left(c_1(\xi_i)\right) - h\left(\ell_1(\xi_i)\right)\right]\right)\right]$$

subject to the budget constraints

$$c_0 + s_1 = \theta\ell_0 - T(\theta\ell_0)$$

$$c_1(\xi_i) = \omega(M,\theta,\xi_i)\ell_1 - T\left(\omega(M,\theta,\xi_i)\,\ell_1\right) + Rs_1\ \forall_{\xi_i}$$

The labor income tax is assumed to take the functional form $T(y_t) = y - \lambda(y_t)^{1-\tau}$ (Heathcote et al., 2017), with parameter values $\tau = 0.161$ and $\lambda = 0.839$ set following Heathcote and Tsujiyama (2021). Agents can save in the first period at the real interest rate R. For simplicity, we do not impose a balanced government budget constraint or model government debt and public goods.

Skill distribution For a given distribution of θ, the model predicts a distribution of first period earnings $y_0 = \theta\ell_0$. We assume that the underlying skill distribution is a mixture of three parametric distributions: normal, lognormal, and Pareto, and we pick all nine parameters of such a hybrid distribution to match the distribution of earnings in the U.S. economy. We target several moments of the empirical distribution of earnings obtained from Guvenen et al. (2022). This data is based on the 2015 US Census and features a high degree of granularity, including information about the tails of the distribution.

Second period productivity We assume the following functional form for the second period productivity

$$\omega(M,\theta,\xi) = \left[\alpha\theta^{1-\xi} + (1-\alpha)M^{1-\gamma}\right]^{1/(1-\gamma)} = \left[\alpha\left(\theta\cdot\theta^{\frac{\gamma-\xi}{1-\gamma}}\right)^{1-\gamma} + (1-\alpha)M^{1-\gamma}\right]^{1/(1-\gamma)} \quad (14)$$

Formula (14) assumes that individual skill θ and new technology M mix within a nested CES production structure. In particular, individual skill is additionally affected by the technological shock realization ξ. This assumption allows individual skill to exhibit its own

curvature (governed by ξ) before combining with capital in a CES aggregator (governed by γ). This specification is designed to capture the essence of the ambiguity over the AI revolution, namely how it will affect different groups of workers. Suppose $\gamma > 1$ meaning that individual skill and AI are generally complementary. If $\xi > \gamma$, then AI is complementary with high individual skills implying a skill-biased technological change. If $\xi < \gamma$ then it works in reverse, i.e. AI is more complementary with low-skilled workers.

Technological shock distribution We assume the AI technological shock in the second period ξ follows a *generalized normal distribution*. This distribution is characterized by three parameters – location (μ_ξ), scale (σ_ξ), and shape (β_ξ) – and features a probability density function of the form

$$f(\xi_i) = \frac{\beta_\xi}{2\sigma_\xi \Gamma(1/\beta_\xi)} e^{-(|\xi_i - \mu_\xi|/\sigma_\xi)}$$

When the shape $\beta_\xi = 2$, we have a normal distribution. The advantage of working with a generalized normal distribution is that it allows us to control the thickness of tails while preserving the general Gaussian shape of the PDF. For $\beta_\xi < 2$, the distribution features thick tails (leptokurtic), while for $\beta_\xi > 2$, the tails are thin (platykurtic). Controlling the tail behavior of a distribution of technological shocks appears particularly important in the context of AI and will prove a useful feature in our calibration and numerical exercises with the model.

AI quality While our main normative model features an optimal investment in AI quality (M), at the calibration stage we need to set it to a value most closely resembling our "current policies" world, i.e. a free market outcome. For the technological revolution to be meaningful in the model, we set it to a high number, corresponding to θ at 0.1 of the top percentile of the distribution of individual skill. Artificial intelligence is already performing specialized tasks at the level equivalent to the top 10^{-6} of human talent (math problems, chess, GO). With other tasks, it is believed to perform at the level of the top 10^{-2} of human ability (coding tasks, translation). Taking a stand somewhere in between, and noticing that our second period is still ahead in the future (next 10-20 years), we assume that M takes a value equivalent to the top 10^{-4} of the distribution of θ. Although it need not be that exact

number, the point is to set it high enough. For our parameterization of the distribution of θ, this value is 1778.[12]

Matching survey data To find the remaining parameters values, we engineer the behavior of agents in our model to mimic the results from the UAS survey discussed in Section 2. Specifically, we aim to replicate two main findings: 1) a broad dispersion of sentiments about the impact of AI on future jobs, and 2) a negative relationship between individual income and pessimism about such impact. For the purpose of this calibration, we define a measure of *pessimism* in our model as follows

$$\text{pessimism}_j = 1 - \frac{\mathbb{E}_m(y_{j,1})}{\mathbb{E}_p(y_{j,1})} \tag{15}$$

In formula (15), $\mathbb{E}_m(\cdot)$ denotes subjective expectations of an agent about their future income (evaluated in terms of the worst-case distribution), while $\mathbb{E}_p(\cdot)$ denotes objective expectations about future income (evaluated in terms of the reference distribution). Hence, *pessimism* close to zero implies that an agent's subjectively expected income is nearly the same as the objective one (i.e. agent is not biased in their evaluation of the distribution of outcomes). On the other hand, a value near one indicates near total pessimism meaning that an agent's subjective beliefs are so biased they do not expect to have much income in the second period at all.

Armed with this measure, we aim to match the distribution of *pessimism* with the responses to question $wk007$ in the UAS survey (Section 2) which asks respondents about the probability they may lose their jobs to AI in the near future. Matching these two indicators is convenient as both are numerical objects bounded between 0 and 100%. However, we must also acknowledge a major discrepancy between them in that the survey question is inherently about an extensive margin of the respondents' expectations (job loss), while the pessimism measure defined in (15) is about an intensive margin (magnitude of income loss). Hence, by matching these two we assume that agents' beliefs about the probability

[12]Notice that this assumption does not imply that AI can displace 99.9% of human workers. While it excels at some tasks, AI must still mix with human talent in order to yield an effective productivity in the second period. A realization of the technological shock ξ determines what type of workers it mixes most efficiently with.

of job loss (extensive margin) map directly to expected income losses (intensive margin). Specifically, an agent who assigns probability p to job loss treats this as equivalent to a deterministic p percent reduction in income. While not a perfect assumption (we would ideally calibrate to question $wk005$ in the survey, which asks about the expected pay change due to AI, if the responses to it were numeric and not categorical), we think it is reasonable as a first-order approximation. Job displacement represents the dominant source of labor income risk, meaning the extensive margin captures the majority of perceived income uncertainty. If anything, this assumption understates true risk by ignoring wage fluctuations conditional on employment and persistent scarring effects of job loss.

Parameter values and model fit Table 2 summarizes the parameter values we set externally or otherwise outside of the solution of the "current policies" model. Preference parameters are rather standard. The general AI-skill complementarity is set to 2, implying that individual skill and AI are in principle quite complementary (in line with findings of the vast literature on capital and labor complementarity), although the exact elasticity of substitution depends on realization of technological shock ξ. The AI capability is set to 1778 which corresponds to the top 0.1% of the distribution of individual skill θ.

Table 3 summarizes the parameter values obtained internally from the solution of the "current policies" model. The nine parameters that shape the distribution of individual skills are primarily informed by targeting nine moments of the empirical distribution of earnings. We target the usual central moments such as mean, standard deviation, skewness, kurtosis; but also the 10th percentile, mode, the top 1% share, and the 75/25 and 90/10 percentile ratios. As the table demonstrates, our hybrid distribution of θ is flexible enough to allow us to match all of these moments well.

Within the same calibration procedure, we also target the behavior of *pessimism* in the model. We target the mean and standard deviation of survey responses to the job loss probability question, as well as the regression coefficient of pessimism on log of household income. We also impose a hand-tying condition that the correlation of current income with future income growth (evaluated with the reference distribution) is zero. We impose this condition in order not to take a stand on the objective impact of the AI revolution (which

we, as modelers, cannot predict). Hence, we assume that under the veil of ignorance, AI can objectively be expected to have a neutral impact on income growth. Finally, we also impose a normalization assumption that the expected value of the technological shock is equal to γ, i.e. in expectation the productivity formula (14) is a CES function. Matching these moments (which our model does well) jointly informs the parameters of the AI shock and ambiguity aversion. The value of the ambiguity aversion parameter is 0.15, while the share of individual skill in the second period productivity formula is 0.14. It is also notable that the calibration requires us to set the shape parameter of the distribution of the AI shock to 4, which implies that the distribution has relatively thin tails.

Cost of AI capital (ϕ) The final step of our calibration exercise is to find the price parameter of M (AI capability), ϕ. While this value does not play an important role in our positive "current policies" world (as the government budget constraint need not balance out), it is crucial for the normative Mirrleesian model developed in Section 3 in which the planner decides on the optimal investment in such machines subject to an aggregate resource constraint.

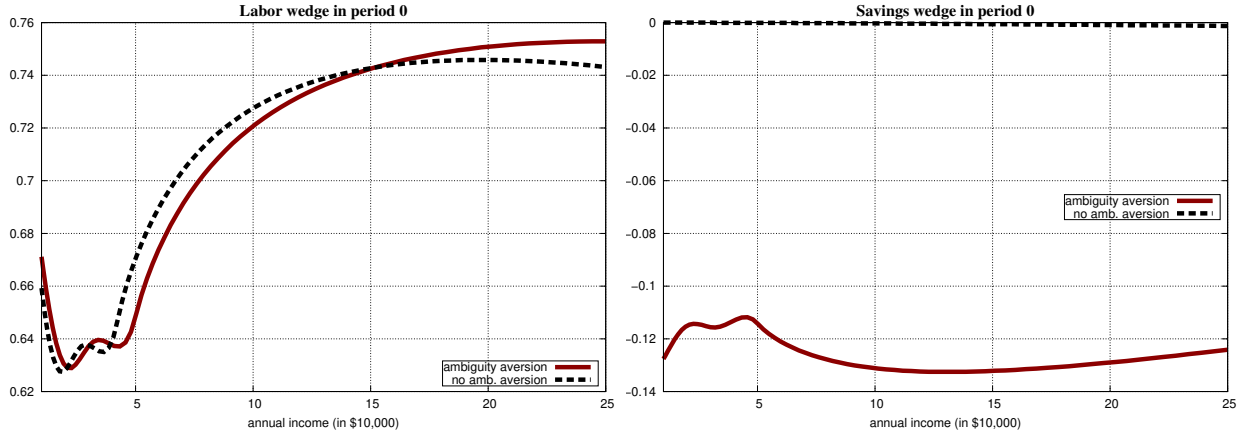
To calibrate this parameter, we use the variant of our optimal Mirrleesian model *without* ambiguity aversion and set it such that the optimal investment in M is equal to our assumption in the positive "current policies" model. By doing so, we implicitly assume that the factual AI quality corresponds to a world with optimal fiscal policy that does not take into account the fact that agents are ambiguity averse with respect to the looming technological change. Consequently, in the following section, we will show how optimal policy differs once the government does take this preference into account.

5.2 Optimal policies

In this section, we present the main results on the optimal fiscal policy in our model. Figure 3 shows the optimal wedges in period zero under two scenarios: ambiguity averse agents, as calibrated in the preceding section, and no ambiguity aversion whatsoever. The left panel shows the optimal labor distortion which generally takes the U shape, common for quantitative Mirrlees models. At the very bottom of the income distribution, agents face a

fairly high marginal tax rate because they are rather unproductive and the planner tries to prevent higher-ability agents from mimicking them. The optimal distortion then declines for the "middle class" before increasing sharply and stabilizing at a rather high level due to the presence of a Pareto tail in the distribution of skills. As is evident, the presence of ambiguity aversion calls for a minor reduction in the marginal tax rates for middle and upper-middle income households, while increasing them sharply for the highest earning individuals.

Optimal tax schedules The right panel of Figure 3 shows the optimal savings wedge in period zero. It is optimal for the government to subsidize savings of all agents, especially so at the bottom of the income range. The savings subsidy helps alleviate the agents' worries about the worst-case scenario and optimally amounts to just over 10%.



(a) Labor wedge

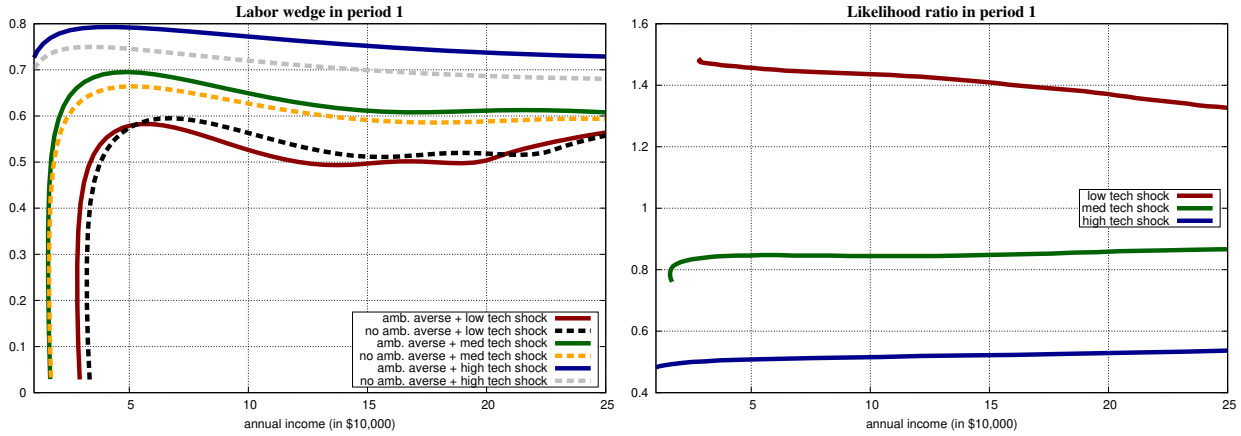
(b) Savings wedge (subjective)

Figure 3: Period 0 wedges

Figure 4(a) plots the optimal labor wedges in the second period (which also depend on the realization of the technology shock). For the low values of the shock, which favor low-skill workers, the optimal tax schedule has a usual U shape, shifted downwards. For the high realization, which favor high-skill workers, the optimal marginal income tax rate is mostly a decreasing function, incentivizing high-ability workers to put effort and discouraging them from misreporting their type. Relative to the standard preferences, ambiguity aversion reduces the tax schedule in the former case and shifts it upward in the latter.

Figure 4(b) plots the likelihood ratios associated with different realizations of the tech-

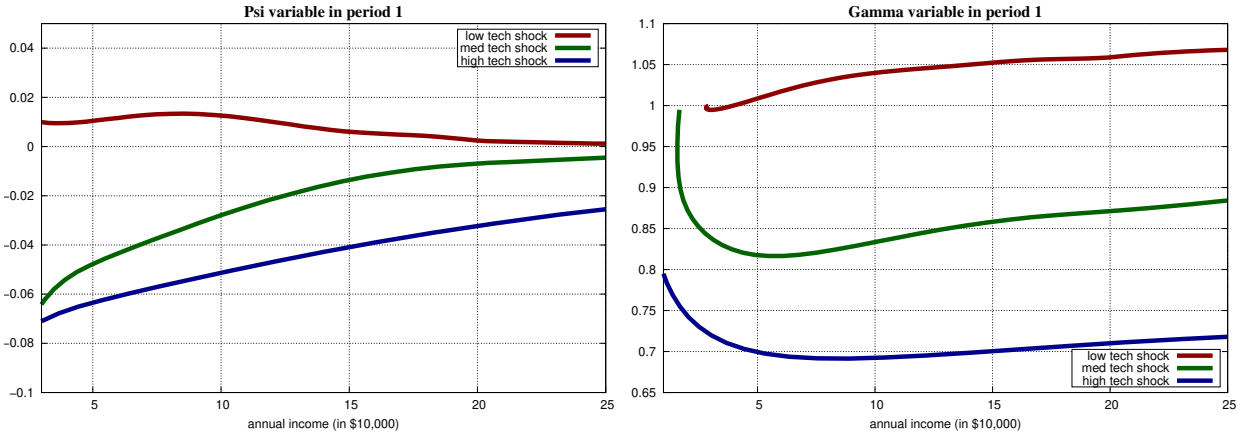
nology shocks. Low values generally focus more of the agents' concerns than high values, and the variance is especially pronounced for low income individuals. This pattern helps to understand ambiguity aversion's effect on the optimal income tax schedules. Government reduces distortions in the states that are subject to more concern, and they increase distortions in the states that agents are less anxious about.



(a) Labor wedge (b) Likelihood ratio

Figure 4: Period 1 wedges and beliefs

Figure 5 further illuminates these results by plotting two crucial objects defined in Section 4. As the key component of the second period labor wedge, the Ψ component is positive when the technology shock is low (leading to reduced distortions) and turns negative as



(a) Psi (b) Gamma

Figure 5: Gamma and Psi variables

the shock realization increases (higher distortions). The Γ component of the savings wedge similarly indicates a savings subsidy (when the shock is low) or a savings tax (when it is high). Because the likelihood ratio is much higher in the former case, it ultimately dominates, and the optimal wedge in the first period is negative (i.e. a savings subsidy).

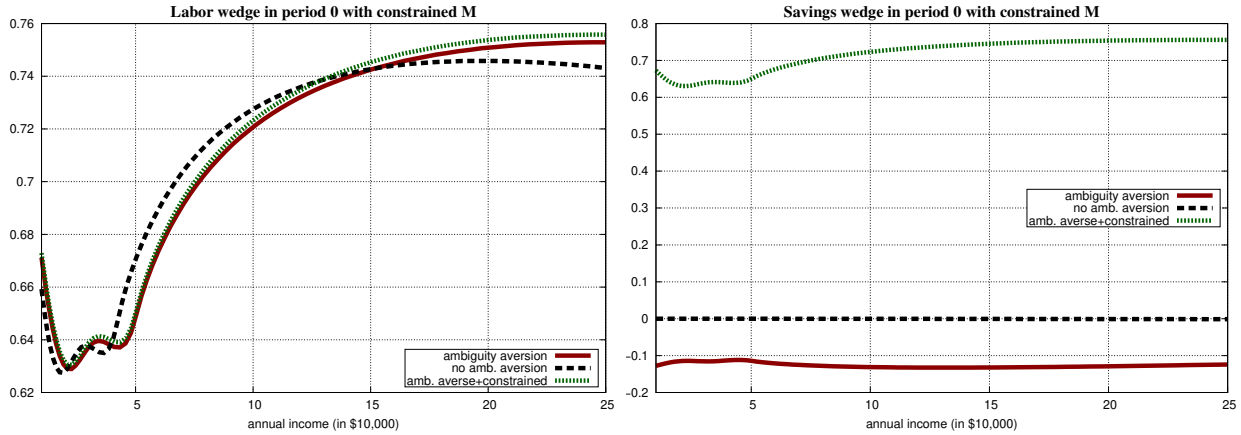
Optimal AI capability Beyond the agent allocations, our Mirrleesian model also predicts an optimal level of investment in the AI capability M. Table 4 shows basic comparative statics exercises to understand the forces behind this decision. The numbers in the table are normalized and equal to 100 for the baseline calibration with no ambiguity aversion (which was targeted to match what we believe to be a "factual" AI capability). We uncover three findings which align with simple intuition. First, holding everything else fixed, optimal AI capability is reduced by around 30% when agents are ambiguity-averse. This result reflects the quantitative power of ambiguity aversion in our setting, which arises from the calibration to survey data. In addition to balancing the efficiency-equity tradeoff, government policy also has a role in alleviating workers' uncertainty regarding the future impact of the new technology. Second, optimal AI is reduced as the tails of the technology shock distribution get thicker, and it increases as they become thinner although the effect is nonlinear. Third, as the share of worker skill increases in the production function (α) towards more typical values in macroeconomics, the optimal AI investment gets decimated. This is because such technology become much less important in determining future worker productivity.

Table 4: Optimal AI capability M

	Shape parameter for ξ:		
	2	**4** (baseline)	**6**
Results for $\alpha = 0.14$ (baseline):			
$\eta = 0.15$ (ambiguity aversion)	57.6	68.7	71.4
$\eta \to \infty$ (no ambiguity aversion)	91.1	100.0	101.4
Results for $\alpha = 0.67$: (typical macro)			
$\eta = 0.15$ (ambiguity aversion)	10.8	11.5	11.8
$\eta \to \infty$ (no ambiguity aversion)	18.9	19.6	19.7

The finding that optimal policy involves taxing AI investments to reduce its stock by about 30% can appear rather unsettling. After all, the model we present is very simple

and ignores many crucial aspects of AI, such as its potential to raise aggregate productivity (rather than changing the distribution of it) or the fact that it is an intangible technology that easily defuses across borders. For this reason, in Figure 6 we consider a refinement of the previous wedge analysis. Specifically, we now solve the optimal Mirrleesian system under an additional constraint that optimal AI quality (with ambiguity aversion) must be the same as without ambiguity aversion (which itself targets our best guess about its values in the real world). Effectively, we solve the optimal model in which M is a parameter, not a choice. Figure 6(a) shows that such a modification has almost no effect on the optimal labor distortions. On the other hand, Figure 6(b) demonstrates that all of the impact of this change is focused on the optimal savings wedge, which shifts up significantly and turns positive. This means that it is optimal to tax savings in the first period and by doing so incentivize agents to work harder in the future.



(a) Labor wedge

(b) Savings wedge (subjective)

Figure 6: Period 0 wedges with constrained AI quality

5.3 Quantifying belief manipulation

In this section, we analyze the degree of belief manipulation involved in the optimal policy. The possibility of a planner achieving welfare increase by the means of influencing the agents' beliefs is among the core theoretical insights of the present paper. How much it is actually used in practice is then a natural open question.

Figure 7 presents a counterpart to Figure 4(b) that plots likelihood ratios, contrasted with their counterfactual "no ambiguity" beliefs. To compute the counterfactual beliefs, we take the optimal Mirrleesian allocations from the variant of our model without ambiguity aversion, and then we use these allocations to *evalue* the agents' utility with ambiguity aversion (we do not optimize the tax system). We save the likelihood ratios and plot them in the figure.[13] The graph shows that optimal allocations that take ambiguity aversion into account generally reduce the likelihood ratio of a low ("bad") technological shock and increase the likelihood ratio of a high ("good") shocks. As such, the planner acts to reduce the discrepancy between the pessimistic and optimistic subjective beliefs of agents, moving the system towards the one without ambiguity aversion.

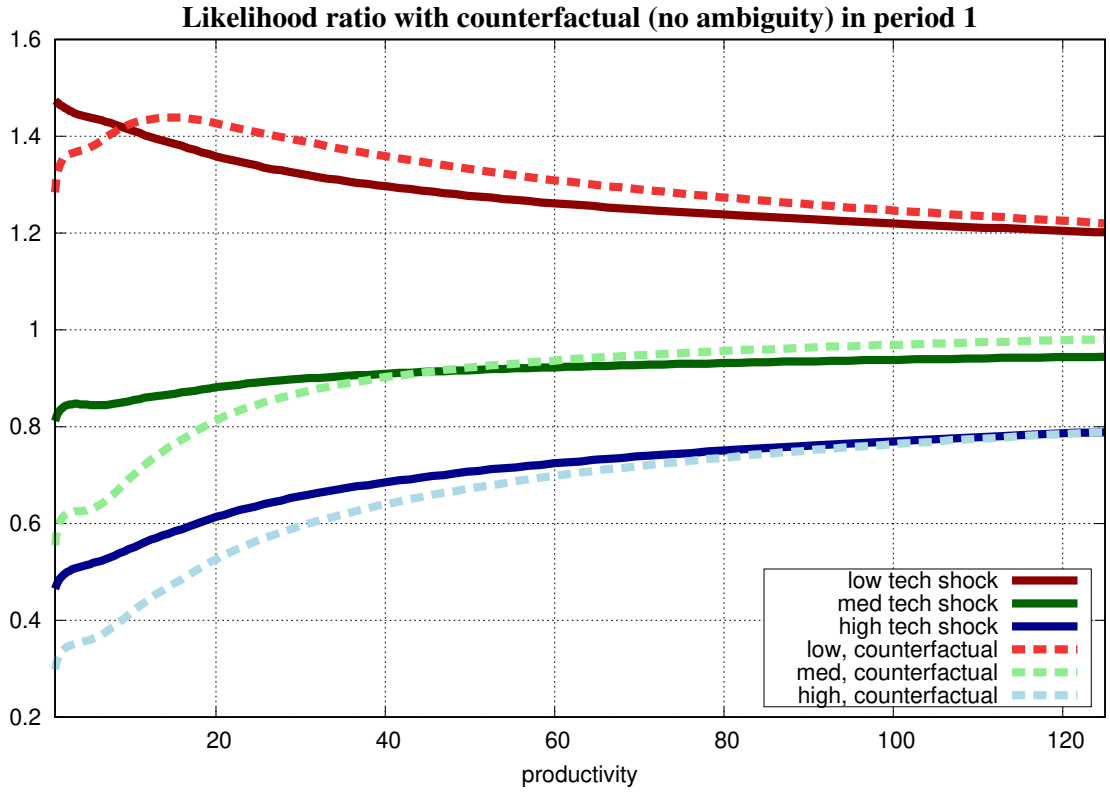


Figure 7: Likelihood ratio in the baseline and counterfactual

Table 5 quantifies the amount of income growth and pessimism under the three scenarios of interest: no ambiguity aversion, and ambiguity aversion with or without a restriction on

[13]Note that in such evaluations, all resource constraints are guaranteed to hold, while incentive compatibility constraints are potentially violated.

AI quality. Under no ambiguity aversion, there is no pessimism in our model (as defined in formula (15)) because the subjective and objective expectations of income growth are equal. In a baseline ambiguity aversion variant of our model, income growth expectations generally fall and a gap opens betwee the subjective and the objective expectations, resulting in non-zero measures of pessimism. Most interestingly, this model generates a positive regression coefficient between pessimism and log income of 0.16, while the same number in the data was -0.06! This means that the planner is able to invert the regressive nature of AI pessimism. Curiously, for the lowest quartile we actually observe the agents exhibiting negative pessimism, which means they are optimistic. This is not possible in the "current policies" model we use for calibration. Finally, the bottom panel of the table shows that if AI quality M is restricted to be high, the belief manipulation actions of the planner are further reinforced, with the regression coefficient increasing further and more agents becoming even more optimistic. There results show that the belief manipulation channel can be a significant force in our model, both qualitatively and quantitatively.

Table 5: Distribution of growth and pessimism

Statistic	Mean	Std Dev	P10	P25	P50	P75	P90
No ambiguity aversion							
y1/y0 (Objective)	12.74	1.87	14.88	12.08	11.68	12.70	13.31
y1/y0 (Subjective)	12.74	1.87	14.88	12.08	11.68	12.70	13.31
Pessimism	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pessimism-income reg coeff	0.00						
Baseline ambiguity aversion							
y1/y0 (Objective)	9.83	1.18	10.98	9.56	9.59	10.05	9.65
y1/y0 (Subjective)	9.47	3.08	14.33	10.36	8.39	7.79	7.33
Pessimism	0.05	0.21	-0.31	-0.08	0.13	0.23	0.24
Pessimism-income reg coeff	0.16						
Ambiguity aversion with high AI							
y1/y0 (Objective)	9.90	1.35	11.31	9.38	9.29	10.01	10.17
y1/y0 (Subjective)	9.77	3.78	15.59	10.58	8.26	7.72	7.58
Pessimism	0.03	0.23	-0.38	-0.13	0.11	0.23	0.25
Pessimism-income reg coeff	0.17						

6 Conclusion

This paper documents that workers are divided in the expectations about the future impacts of AI on labor markets. Pessimism is prevalent and mostly shared among low-income workers. We use this evidence to build a quantitative Mirrleesian model of optimal taxation under ambiguity over the nature of future technology. Beyond managing the traditional efficiency-equity tradeoff, government policy now plays a role in managing workers' expectations and alleviating their concerns about a potentially disadvantageous realization of uncertainty. We calibrate the model to the micro data and show that the ambiguity-alleviation channel is quantitatively significant in the context of AI, calling for a potentially more progressive labor income tax schedule, combined with a savings subsidy.

References

Acemoglu, Daron and Pascual Restrepo, "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 2018, *108* (6), 1488–1542.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz, "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology," *NBER Working Paper*, 2025.

Bhandari, Anmol, "Doubts, Asymmetries, and Insurance," *Working Paper*, 2015.

_ , Jaroslav Borovicka, and Yuki Yao, "Robust Bounds on Optimal Tax Progressivity," *Working Paper*, 2023.

Boyd, Stephen and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

Costinot, Arnaud and Ivan Werning, "Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation," *Review of Economic Studies*, 2023, *90* (5), 2261–2291.

Cover, Thomas M. and Joy A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

Diamond, Peter, "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 1998, *96* (3), 694–719.

Farhi, Emmanuel and Ivan Werning, "Optimal Savings Distortions with Recursive Preferences," *Journal of Monetary Economics*, 2008, *55*, 21–42.

_ and _ , "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 2013, *80*.

Ferriere, Axelle and Anastasios Karantounias, "Fiscal Austerity in Ambiguous Times," *American Economic Journal: Macroeconomics*, 2019, *11* (1), 89–131.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski, "Redistribution and Social Insurance," *American Economic Review*, 2016, *106* (2), 359–386.

Grochulski, Borys and Narayana Kocherlakota, "Nonseparable Preferences and Optimal Social Security Systems," *Journal of Economic Theory*, 2010, *145*, 2055–2077.

Guerreiro, Joao, Sergio Rebelo, and Pedro Teles, "Should Robots be Taxed?," *Review of Economic Studies*, 2022, *89* (1), 279–311.

Guvenen, Fatih, Luigi Pistaferri, and Giovanni L Violante, "The Global Repository of Income Dynamics," 2022. https://www.grid-database.org/. Accessed 07.01.2025.

Hansen, Lars and Thomas Sargent, "Robust Control and Model Uncertainty," *American Economic Review*, 2001, *91* (2), 60–66.

_ and _ , *Robustness*, Princeton University Press, 2008.

Heathcote, Jonathan and Hitoshi Tsujiyama, "Optimal Income Taxation: Mirrlees Meets Ramsey," *Journal of Political Economy*, 2021, *129*, 3141–3184.

_ , Kjetil Storesletten, and Giovanni L Violante, "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics*, 2017, *132*, 1693–1754.

Hellwig, Christian, "Static and Dynamic Mirrleesian Taxation with Non-separable Preferences: A Unified Approach," *Working Paper*, 2021.

Kapicka, Marek, "Optimal Fiscal Policy with Preference Heterogeneity," *Working Paper*, 2024.

Karantounias, Anastasios, "Managing Pessimistic Expectations and Fiscal Policy," *Theoretical Economics*, 2013, *8* (1), 193–231.

_ , "Doubts About the Model and Optimal Policy," *Journal of Economic Theory*, 2023, *210*, 105643.

Kina, Ozlem, Ctirad Slavik, and Hakki Yazici, "Redistributive Capital Taxation Revisited," *American Economic Journal: Macroeconomics*, 2024, *16* (2).

Koehne, Sebastian and Moritz Kuhn, "Optimal Taxation in a Habit Formation Economy," *Journal of Public Economics*, 2015, *122*, 31–39.

Lensman, Todd and Maxim Troshkin, "Implications of Uncertainty for Optimal Policies," *Journal of Economic Theory*, 2022, *199*, 105206.

Loebbing, Jonas, "Redistributive Income Taxation with Directed Technical Change," *Working Paper*, 2019.

Miao, Jianjun and Alejandro Rivera, "Robust Contracts in Continuous Time," *Econometrica*, 2016, *84* (4), 1405–1440.

Mirrlees, James, "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 1971, *38* (2), 175–208.

Paluszynski, Radoslaw and Pei Cheng Yu, "Efficient Consolidation of Incentives for Education and Retirement Savings," *American Economic Journal: Macroeconomics*, 2023, *15* (3), 153–190.

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 2001, *68*, 205–229.

Slavik, Ctirad and Hakki Yazici, "Machines, Buildings, and Optimal Dynamic Taxes," *Journal of Monetary Economics*, 2014, *66*, 47–61.

Strzalecki, Tomasz, "Axiomatic Foundations of Multiplier Preferences," *Econometrica*, 2011, *79* (1), 47–73.

Thuemmel, Uwe, "Optimal Taxation of Robots," *Journal of the European Economic Association*, 2023, *21* (3), 1154–1190.

Vairo, Maren, "Robustly Optimal Income Taxation," *Working Paper*, 2024.

Werning, Ivan, "An Elementary Proof of Positive Optimal Marginal Taxes," Technical Report, MIT 2000.

Taxation and Technological Uncertainty

Anastasios Karantounias and Radoslaw Paluszynski and Pei Cheng Yu

APPENDIX FOR ONLINE PUBLICATION ONLY

A Proofs

Proof of Lemma 1: Given that $m_{M,\theta}(\xi_i) = \frac{\exp\left(-\frac{1}{\eta}\left[u(c_1(\theta,\xi_i))-h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\right]\right)}{\sum_{i=0}^{S} p(\xi_i)\exp\left(-\frac{1}{\eta}\left[u(c_1(\theta,\xi_i))-h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\right]\right)}$, we have

$$\frac{\partial m_{M,\theta}(\xi_i)}{\partial c_1(\theta,\xi_i)} = \frac{1}{\eta}u'(c_1(\theta,\xi_i))\, m_{M,\theta}(\xi_i)\left[p(\xi_i)\, m_{M,\theta}(\xi_i) - 1\right],$$

$$\frac{\partial m_{M,\theta}(\xi_i)}{\partial y_1(\theta,\xi_i)} = \frac{1}{\eta}\frac{1}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right) m_{M,\theta}(\xi_i)\left[1 - p(\xi_i)\, m_{M,\theta}(\xi_i)\right],$$

and for $\xi_k \neq \xi_i$,

$$\frac{\partial m_{M,\theta}(\xi_k)}{\partial c_1(\theta,\xi_i)} = \frac{1}{\eta}m_{M,\theta}(\xi_k)\, u'(c_1(\theta,\xi_i))\, p(\xi_i)\, m_{M,\theta}(\xi_i),$$

$$\frac{\partial m_{M,\theta}(\xi_k)}{\partial y_1(\theta,\xi_i)} = -\frac{1}{\eta}\frac{m_{M,\theta}(\xi_k)}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right) p(\xi_i)\, m_{M,\theta}(\xi_i).$$

Therefore, the results immediately follow from the definitions of the belief elasticities. ■

Let $\chi(\theta)$, $\lambda(\theta)$, and $\mu(\xi_i) \geq 0$ for all $\xi_i \in \Xi$ denote the multipliers on (1), (5), and (6), respectively. Using standard Hamiltonian techniques, the optimum of the relaxed problem is characterized by (1), (5), (6), and the first-order conditions: for any $\theta \in \Theta$ and $\xi_i \in \Xi$,

$$\lambda'(\theta) = f(\theta) - \chi(\theta), \tag{16}$$

$$\chi(\theta)\, u'(c_0(\theta)) = f(\theta)\sum_{i=0}^{S}\mu(\xi_i), \tag{17}$$

$$\chi(\theta)\frac{1}{\theta}h'\left(\frac{y_0(\theta)}{\theta}\right) + \lambda(\theta)\left[\frac{1}{\theta^2}h'\left(\frac{y_0(\theta)}{\theta}\right) + \frac{y_0(\theta)}{\theta^3}h''\left(\frac{y_0(\theta)}{\theta}\right)\right] = f(\theta)\sum_{i=0}^{S}\mu(\xi_i), \tag{18}$$

$$\chi(\theta) u'(c_1(\theta,\xi_i)) p(\xi_i) m_{M,\theta}(\xi_i)$$

$$-\lambda(\theta)\sum_{k=0}^{S} p(\xi_k)\frac{y_1(\theta,\xi_k)}{[\omega(M,\theta,\xi_k)]^2}h'\left(\frac{y_1(\theta,\xi_k)}{\omega(M,\theta,\xi_k)}\right)\frac{\partial\omega(M,\theta,\xi_k)}{\partial\theta}\frac{\partial m_{M,\theta}(\xi_k)}{\partial c_1(\theta,\xi_i)} = f(\theta)\mu(\xi_i), \tag{19}$$

$$\frac{\chi(\theta)}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)p(\xi_i) m_{M,\theta}(\xi_i)$$

$$+\lambda(\theta)\left\{\sum_{k=0}^{S} p(\xi_k)\frac{y_1(\theta,\xi_k)}{[\omega(M,\theta,\xi_k)]^2}h'\left(\frac{y_1(\theta,\xi_k)}{\omega(M,\theta,\xi_k)}\right)\frac{\partial\omega(M,\theta,\xi_k)}{\partial\theta}\frac{\partial m_{M,\theta}(\xi_k)}{\partial y_1(\theta,\xi_i)}\right.$$

$$\left.+\frac{p(\xi_i) m_{M,\theta}(\xi_i)}{[\omega(M,\theta,\xi_i)]^2}\left[h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)+\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}h''\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\right]\frac{\partial\omega(M,\theta,\xi_i)}{\partial\theta}\right\}$$

$$= f(\theta)\mu(\xi_i), \tag{20}$$

$$\int_\Theta \frac{\chi(\theta)}{RM}\sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i)\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\epsilon_M^{\omega}(M,\theta,\xi_i)\,d\theta$$

$$-\int_\Theta \frac{\lambda(\theta)}{RM}\sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i)\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\frac{\epsilon_\theta^{\omega}(M,\theta,\xi_i)}{\theta}\left\{\epsilon_M^{m}(M,\theta,\xi_i)\right.$$

$$\left.+\epsilon_M^{\omega}(M,\theta,\xi_i)\left[\left(\epsilon^H(M,\theta,\xi_i)-1\right)-\left(1+\frac{1}{\gamma_1(M,\theta,\xi_i)}\right)\right]\right\}d\theta = \phi\sum_{i=0}^{S}\mu(\xi_i), \tag{21}$$

and the boundary conditions: $\lambda(\underline{\theta}) = \lambda(\overline{\theta}) = 0$. Below, we show that our theoretical results follow from these conditions.

Proof of Proposition 1: When θ is observable, then $\lambda(\theta) = 0$ for all $\theta \in \Theta$. Immediately, from (16), (17), and (19), we have the belief-adjusted Euler equation: $u'(c_0(\theta)) = \sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i) u'(c_1(\theta,\xi_i))$. From (17) and (18), we can show that output is efficient:

$\theta u'(c_0(\theta)) = h'\left(\frac{y_0(\theta)}{\theta}\right)$, and we can derive a similar condition in $t = 1$ from (19) and (20). Also, from (16) and (17), we can immediately find that full insurance prevails at $t = 0 : c_0(\theta) = c_0(\theta')$ for all $\theta, \theta' \in \Theta$. While at $t = 1$, from (16) and (19), we have $m_{M,\theta}(\xi_i) u'(c_1(\theta, \xi_i)) = m_{M,\theta'}(\xi_i) u'(c_1(\theta', \xi_i))$ for all $\xi_i \in \Xi$ and $\theta, \theta' \in \Theta$. Finally, from the results before and (21), we have

$$\phi M u'(c_0(\theta)) = \frac{1}{R} \int_\Theta \sum_{i=0}^{S} p(\xi_i) m_{M,\theta}(\xi_i) y_1(\theta, \xi_i) u'(c_1(\theta, \xi_i)) \epsilon_M^\omega(M, \theta, \xi_i) dF(\theta),$$

which completes the proof. ■

Proof of Theorem 1: First, by the definition of belief elasticities, we can express (19) as

$$\chi(\theta) = \frac{1}{p(\xi_i) m_{M,\theta}(\xi_i) u'(c_1(\theta, \xi_i))} \Bigg[f(\theta) \mu(\xi_i)$$

$$+ \frac{\lambda(\theta)}{c_1(\theta, \xi_i)} \sum_{k=0}^{S} p(\xi_k) m_{M,\theta}(\xi_k) \frac{y_1(\theta, \xi_k)}{\theta \omega(M, \theta, \xi_k)} h'\left(\frac{y_1(\theta, \xi_k)}{\omega(M, \theta, \xi_k)}\right) \epsilon_\theta^\omega(M, \theta, \xi_k) \epsilon_c^m(M, \theta, \xi_k, \xi_i) \Bigg].$$

By Lemma 1, we can rewrite the expression for $\chi(\theta)$ as

$$\chi(\theta) = \frac{f(\theta) \mu(\xi_i)}{p(\xi_i) m_{M,\theta}(\xi_i) u'(c_1(\theta, \xi_i))} - \lambda(\theta) \sum_{k=0}^{S} \epsilon_y^m(M, \theta, \xi_i, \xi_k) \frac{\epsilon_\theta^\omega(M, \theta, \xi_k)}{\theta}. \tag{22}$$

Substituting (22) into (16) and using the integrating-factor method to solve the ODE yields

$$\lambda(\theta) = \int_\theta^{\overline{\theta}} \exp\left(- \int_\theta^x \sum_{k=0}^{S} \epsilon_y^m(M, z, \xi_i, \xi_k) \frac{\epsilon_\theta^\omega(M, z, \xi_k)}{z} dz \right)$$

$$\times \left[\frac{\mu(\xi_i)}{p(\xi_i) m_{M,x}(\xi_i) u'(c_1(x, \xi_i))} - 1 \right] dF(x), \tag{23}$$

where the integrating factor used was $\exp\left(\int_\theta^x \sum_{k=0}^{S} \epsilon_y^m(M, z, \xi_i, \xi_k) \frac{\epsilon_\theta^\omega(M, z, \xi_k)}{z} dz \right)$.

Next, by (22), we can rewrite (20) as

$$\frac{f(\theta)\mu(\xi_i)h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}{\omega(M,\theta,\xi_i)u'(c_1(\theta,\xi_i))}-\frac{\lambda(\theta)p(\xi_i)m_{M,\theta}(\xi_i)h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}{\omega(M,\theta,\xi_i)}\sum_{k=0}^{S}\epsilon_y^m(M,\theta,\xi_i,\xi_k)\frac{\epsilon_\theta^\omega(M,\theta,\xi_k)}{\theta}$$

$$+\frac{\lambda(\theta)}{y_1(\theta,\xi_i)}\sum_{k=0}^{S}p(\xi_k)m_{M,\theta}(\xi_k)\frac{y_1(\theta,\xi_k)}{\theta\omega(M,\theta,\xi_k)}h'\left(\frac{y_1(\theta,\xi_k)}{\omega(M,\theta,\xi_k)}\right)\epsilon_\theta^\omega(M,\theta,\xi_k)\epsilon_y^m(M,\theta,\xi_k,\xi_i)$$

$$+\lambda(\theta)p(\xi_i)m_{M,\theta}(\xi_i)\frac{\epsilon_\theta^\omega(M,\theta,\xi_i)}{\theta\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)B_1(M,\theta,\xi_i)=f(\theta)\mu(\xi_i),$$

where $B_1(M,\theta,\xi_i)=1+\frac{\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}h''\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}{h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}$. By Lemma 1, we have

$$\frac{1}{y_1(\theta,\xi_i)}\sum_{k=0}^{S}p(\xi_k)m_{M,\theta}(\xi_k)\frac{y_1(\theta,\xi_k)}{\theta\omega(M,\theta,\xi_k)}h'\left(\frac{y_1(\theta,\xi_k)}{\omega(M,\theta,\xi_k)}\right)\epsilon_\theta^\omega(M,\theta,\xi_k)\epsilon_y^m(M,\theta,\xi_k,\xi_i)$$

$$=\frac{p(\xi_i)m_{M,\theta}(\xi_i)h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}{\omega(M,\theta,\xi_i)}\sum_{k=0}^{S}\epsilon_y^m(M,\theta,\xi_i,\xi_k)\frac{\epsilon_\theta^\omega(M,\theta,\xi_k)}{\theta}.$$

Thus, (20) can be further simplified to

$$\frac{f(\theta)\mu(\xi_i)h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)}{\omega(M,\theta,\xi_i)u'(c_1(\theta,\xi_i))}+\lambda(\theta)p(\xi_i)m_{M,\theta}(\xi_i)\frac{\epsilon_\theta^\omega(M,\theta,\xi_i)}{\theta\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)B_1(M,\theta,\xi_i)$$

$$=f(\theta)\mu(\xi_i).$$

By the definition of the labor wedge τ_1^w, we divide both sides of the equation above by $\frac{1}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)$ and rearrange the terms to gives us

$$\frac{f(\theta)\mu(\xi_i)}{u'(c_1(\theta,\xi_i))}\frac{\tau_1^w(M,\theta,\xi_i)}{1-\tau_1^w(M,\theta,\xi_i)}=\lambda(\theta)p(\xi_i)m_{M,\theta}(\xi_i)B_1(M,\theta,\xi_i)\frac{\epsilon_\theta^\omega(M,\theta,\xi_i)}{\theta}. \tag{24}$$

Substituting (23) into (24) yields

$$\frac{\tau_1^w(M,\theta,\xi_i)}{1-\tau_1^w(M,\theta,\xi_i)} = A_1(M,\theta,\xi_i)\, B_1(M,\theta,\xi_i)\, C_1(M,\theta,\xi_i),$$

where $A_1(M,\theta,\xi_i) = A_0(\theta)\,\epsilon_\theta^\omega(M,\theta,\xi_i)$,

$$C_1(M,\theta,\xi_i) = \int_\theta^{\overline{\theta}} \frac{m_{M,\theta}(\xi_i)\,u'(c_1(\theta,\xi_i))}{m_{M,x}(\xi_i)\,u'(c_1(x,\xi_i))}\left[1-\frac{p(\xi_i)\,m_{M,x}(\xi_i)\,u'(c_1(x,\xi_i))}{\mu(\xi_i)}\right] \times T(M,x,\theta,\xi_i)\frac{f(x)}{1-F(\theta)}dx,$$

$$T(M,x,\theta,\xi_i) = \exp\left(-\int_\theta^x \sum_{k=0}^S \epsilon_y^m(M,z,\xi_i,\xi_k)\frac{\epsilon_\theta^\omega(M,z,\xi_k)}{z}dz\right).$$

Next, note that, by $\tilde{U}_{1,c}(M,\theta,\xi_i) = p(\xi_i)\,m_{M,\theta}(\xi_i)\,u'(c_1(\theta,\xi_i))$,

$$\frac{m_{M,\theta}(\xi_i)\,u'(c_1(\theta,\xi_i))}{m_{M,x}(\xi_i)\,u'(c_1(x,\xi_i))} = \frac{\tilde{U}_{1,c}(M,\theta,\xi_i)}{\tilde{U}_{1,c}(M,x,\xi_i)} = \exp\left(-\int_\theta^x \frac{\partial \ln\left[\tilde{U}_{1,c}(M,z,\xi_i)\right]}{\partial z}dz\right).$$

Since $\tilde{U}_{1,cy}(M,\theta,\xi_i) = \frac{\tilde{U}_{1,cl}(M,\theta,\xi_i)}{\omega(M,\theta,\xi_i)}$, we have

$$\begin{aligned}
-\frac{\partial \ln\left[\tilde{U}_{1,c}(M,z,\xi_i)\right]}{\partial z} &= -\sum_{k=0}^S \frac{\tilde{U}_{1,cc}(M,z,\xi_i,\xi_k)}{\tilde{U}_{1,c}(M,z,\xi_i)}\frac{\partial c_1(z,\xi_k)}{\partial z} \\
&\quad -\sum_{k=0}^S \frac{\tilde{U}_{1,cy}(M,z,\xi_i,\xi_k)}{\tilde{U}_{1,c}(M,z,\xi_i)}\left[\frac{\partial y_1(z,\xi_k)}{\partial z} - \frac{y_1(z,\xi_k)}{z}\epsilon_\theta^\omega(M,z,\xi_k)\right] \\
&= \sigma_1(z,\xi_i)\frac{\frac{\partial c_1(z,\xi_i)}{\partial z}}{c_1(z,\xi_i)} - \sum_{k=0}^S \epsilon_c^m(M,z,\xi_i,\xi_k)\frac{\frac{\partial c_1(z,\xi_k)}{\partial z}}{c_1(z,\xi_k)} \\
&\quad -\sum_{k=0}^S \epsilon_y^m(M,z,\xi_i,\xi_k)\left[\frac{\frac{\partial y_1(z,\xi_k)}{\partial z}}{y_1(z,\xi_k)} - \frac{\epsilon_\theta^\omega(z,\xi_k)}{z}\right].
\end{aligned}$$

where the second equality comes from (2), (3), and Corollary 1. Therefore,

$$\frac{m_{M,\theta}(\xi_i)\,u'(c_1(\theta,\xi_i))}{m_{M,x}(\xi_i)\,u'(c_1(x,\xi_i))}T(M,x,\theta,\xi_i)=\exp\left(\int_\theta^x \sigma_1(z,\xi_i)\frac{\frac{\partial c_1(z,\xi_i)}{\partial z}}{c_1(z,\xi_i)}-\Psi(M,z,\xi_i)\,dz\right),$$

where $\Psi(M,z,\xi_i)=\sum_{k=0}^S \epsilon_c^m(M,z,\xi_i,\xi_k)\frac{\frac{\partial c_1(z,\xi_k)}{\partial z}}{c_1(z,\xi_k)}+\sum_{k=0}^S \epsilon_y^m(M,z,\xi_i,\xi_k)\frac{\frac{\partial y_1(z,\xi_k)}{\partial z}}{y_1(z,\xi_k)}$. Thus, since $\tilde{U}_{1,c}(M,\theta,\xi_i)=p(\xi_i)\,m_{M,\theta}(\xi_i)\,u'(c_1(\theta,\xi_i))$, we can rewrite C_1 as

$$C_1(M,\theta,\xi_i)=\int_\theta^{\overline{\theta}}\exp\left(\int_\theta^x\left[\sigma_1(z,\xi_i)\frac{\frac{\partial c_1(z,\xi_i)}{\partial z}}{c_1(z,\xi_i)}-\Psi(M,z,\xi_i)\right]dz\right)\times\left[1-\frac{\tilde{U}_{1,c}(M,x,\xi_i)}{\mu(\xi_i)}\right]\frac{f(x)}{1-F(\theta)}dx.$$

This completes the proof. ■

Proof of Theorem 2: First, note that by (17) and (22), the Euler equation is

$$u'(c_0(\theta))=\sum_{i=0}^S p(\xi_i)\,m_{M,\theta}(\xi_i)\left[1+\Phi(M,\theta,\xi_i)\right]u'(c_1(\theta,\xi_i)),\tag{25}$$

where $\Phi(M,\theta,\xi_i)=\frac{\lambda(\theta)}{\chi(\theta)}\sum_{k=0}^S \epsilon_y^m(M,\theta,\xi_i,\xi_k)\frac{\epsilon_\theta^\omega(M,\theta,\xi_k)}{\theta}$.

Next, we simplify $\Phi(M,\theta,\xi_i)$. By (22), (23), and the definition of $C_1(M,\theta,\xi_i)$,

$$\frac{\lambda(\theta)}{\chi(\theta)}=\left[\frac{f(\theta)}{[1-F(\theta)]\,C_1(M,\theta,\xi_i)}-\sum_{k=0}^S \epsilon_y^m(M,\theta,\xi_i,\xi_k)\frac{\epsilon_\theta^\omega(M,\theta,\xi_k)}{\theta}\right]^{-1}.$$

Therefore, by Theorem 1,

$$1+\Phi(M,\theta,\xi_i)=\frac{1}{1-\frac{\tau_1^w(M,\theta,\xi_i)}{1-\tau_1^w(M,\theta,\xi_i)}\frac{\sum_{k=0}^S \epsilon_y^m(M,\theta,\xi_i,\xi_k)\epsilon_\theta^\omega(M,\theta,\xi_k)}{B_1(M,\theta,\xi_i)\epsilon_\theta^\omega(M,\theta,\xi_i)}}.\tag{26}$$

Define $\Gamma(M,\theta,\xi_i)=1+\Phi(M,\theta,\xi_i)$, which yields the savings wedge.

Alternatively, we can express $\Gamma(M,\theta,\xi_i)$ in terms of τ_0^w and $\frac{\theta\tilde{U}_{1,\theta c}(M,\theta,\xi_i)}{\tilde{U}_{1,c}(M,\theta,\xi_i)}$. From (16) and

(17), we have

$$\frac{\lambda(\theta)}{\chi(\theta)} = \frac{\tau_0^w(\theta)}{1-\tau_0^w(\theta)} \frac{\theta}{B_0(\theta)}.$$

Next, note that $\tilde{U}_{1,c}(M,\theta,\xi_i) = p(\xi_i) m_{M,\theta}(\xi_i) u'(c_1(\theta,\xi_i))$, so

$$\frac{\theta \tilde{U}_{1,c\theta}(M,\theta,\xi_i)}{\tilde{U}_{1,c}(M,\theta,\xi_i)} = \frac{\partial m_{M,\theta}(\xi_i)}{\partial \theta} \frac{\theta}{m_{M,\theta}(\xi_i)} - \sigma_1(\theta,\xi_i) \frac{\partial c_1(\theta,\xi_i)}{\partial \theta} \frac{\theta}{c_1(\theta,\xi_i)}.$$

Given that $m_{M,\theta}(\xi_i) = \frac{\exp\left(-\frac{1}{\eta}\left[u(c_1(\theta,\xi_i)) - h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\right]\right)}{\sum_{i=0}^{S} p(\xi_i) \exp\left(-\frac{1}{\eta}\left[u(c_1(\theta,\xi_i)) - h\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right)\right]\right)}$, we have

$$\begin{aligned}
\frac{\partial m_{M,\theta}(\xi_i)}{\partial \theta} \frac{\theta}{m_{M,\theta}(\xi_i)} = \frac{1}{\eta} \Bigg[& \sum_{k=0}^{S} p(\xi_k) m_{M,\theta}(\xi_k) c_1(\theta,\xi_k) u'(c_1(\theta,\xi_k)) \frac{\partial c_1(\theta,\xi_k)}{\partial \theta} \frac{\theta}{c_1(\theta,\xi_k)} \\
& - c_1(\theta,\xi_i) u'(c_1(\theta,\xi_i)) \frac{\partial c_1(\theta,\xi_i)}{\partial \theta} \frac{\theta}{c_1(\theta,\xi_i)} \Bigg] \\
- \frac{1}{\eta} \Bigg\{ & \sum_{k=0}^{S} p(\xi_k) m_{M,\theta}(\xi_k) \frac{y_1(\theta,\xi_k)}{\omega(M,\theta,\xi_k)} h'\left(\frac{y_1(\theta,\xi_k)}{\omega(M,\theta,\xi_k)}\right) \left[\frac{\partial y_1(\theta,\xi_k)}{\partial \theta} \frac{\theta}{y_1(\theta,\xi_k)} - \epsilon_\theta^\omega(M,\theta,\xi_k)\right] \\
& - \frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)} h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right) \left[\frac{\partial y_1(\theta,\xi_i)}{\partial \theta} \frac{\theta}{y_1(\theta,\xi_i)} - \epsilon_\theta^\omega(M,\theta,\xi_i)\right] \Bigg\},
\end{aligned}$$

which, by Lemma 1, can be rewritten as

$$\begin{aligned}
\frac{\partial m_{M,\theta}(\xi_i)}{\partial \theta} \frac{\theta}{m_{M,\theta}(\xi_i)} = & \sum_{k=0}^{S} \epsilon_c^m(M,\theta,\xi_i,\xi_k) \frac{\partial c_1(\theta,\xi_i)}{\partial \theta} \frac{\theta}{c_1(\theta,\xi_i)} \\
& + \epsilon_y^m(M,\theta,\xi_i,\xi_k) \left[\frac{\partial y_1(\theta,\xi_k)}{\partial \theta} \frac{\theta}{y_1(\theta,\xi_k)} - \epsilon_\theta^\omega(M,\theta,\xi_k)\right].
\end{aligned}$$

By definition, $\Psi(M,\theta,\xi_i) = \sum_{k=0}^{S} \epsilon_c^m(M,\theta,\xi_i,\xi_k) \frac{\frac{\partial c_1(\theta,\xi_k)}{\partial \theta}}{c_1(\theta,\xi_k)} + \sum_{k=0}^{S} \epsilon_y^m(M,\theta,\xi_i,\xi_k) \frac{\frac{\partial y_1(\theta,\xi_k)}{\partial \theta}}{y_1(\theta,\xi_k)}$,

thus

$$-\frac{\theta\tilde{U}_{1,c\theta}\left(M,\theta,\xi_i\right)}{\tilde{U}_{1,c}\left(M,\theta,\xi_i\right)}=\sigma_1\left(\theta,\xi_i\right)\frac{\partial c_1\left(\theta,\xi_i\right)}{\partial\theta}\frac{\theta}{c_1\left(\theta,\xi_i\right)}-\theta\Psi\left(M,\theta,\xi_i\right)$$
$$+\sum_{k=0}^{S}\epsilon_y^m\left(M,\theta,\xi_i,\xi_k\right)\epsilon_\theta^\omega\left(M,\theta,\xi_k\right).$$

As a result, (26) can be rewritten as

$$\Phi\left(M,\theta,\xi_i\right)=-\frac{\tau_0^w\left(\theta\right)}{1-\tau_0^w\left(\theta\right)}\frac{\theta}{B_0\left(\theta\right)}\left[\frac{\tilde{U}_{1,c\theta}\left(\theta,\xi_i\right)}{\tilde{U}_{1,c}\left(\theta,\xi_i\right)}+\sigma_1\left(\theta,\xi_i\right)\frac{\frac{\partial c_1(\theta,\xi_i)}{\partial\theta}}{c_1\left(\theta,\xi_i\right)}-\Psi\left(M,\theta,\xi_i\right)\right],$$

so $\Gamma\left(M,\theta,\xi_i\right)$ can be expressed as (11). ■

Proof of Theorem 3: First, we show that $\Gamma\left(M,\theta,\xi_i\right)\geq 0$ for all $\theta\in\Theta$, and $\xi_i\in\Xi$. From (25), there exists $\theta\in\Theta$ and $\xi_i\in\Xi$ such that $1+\Phi\left(M,\theta,\xi_i\right)>0$. Suppose $1+\Phi\left(M,\theta',\xi_k\right)<0$ for some $\theta'\in\Theta$ and $\xi_k\in\Xi$. Then, from (22),

$$\chi\left(\theta'\right)p\left(\xi_k\right)m_{M,\theta}\left(\xi_k\right)\left[1+\Phi\left(M,\theta',\xi_k\right)\right]u'\left(c_1\left(\theta',\xi_k\right)\right)=f\left(\theta'\right)\mu\left(\xi_k\right).$$

Since $\mu\left(\xi_i\right)\geq 0$ for all $\xi_i\in\Xi$ and (17) implies $\chi\left(\theta\right)\geq 0$ for all $\theta\in\Theta$, the equation above implies that $\chi\left(\theta'\right)=\mu\left(\xi_k\right)=0$. Thus, by (17) and the fact that $f\left(\theta'\right)>0$, $\sum_{i=0}^{S}\mu\left(\xi_i\right)=0$, which implies that $\chi\left(\theta\right)=\mu\left(\xi_i\right)=0$ for all $\theta\in\Theta$ and $\xi_i\in\Xi$. Hence, by (16), we have $\lambda\left(\theta\right)=-\int_{\theta}^{\overline{\theta}}f\left(x\right)dx<0$ for all $\theta\in\left[\underline{\theta},\overline{\theta}\right)$, This is a contradiction since, by the boundary conditions, $\lambda\left(\underline{\theta}\right)=0$ at the optimum. Therefore, $\Gamma\left(M,\theta,\xi_i\right)=1+\Phi\left(M,\theta,\xi_i\right)\geq 0$ for all $\theta\in\Theta$ and $\xi_i\in\Xi$. By (26), it must be that $1\geq\frac{\tau_1^w(M,\theta,\xi_i)}{1-\tau_1^w(M,\theta,\xi_i)}\frac{\sum_{k=0}^{S}\epsilon_y^m(M,\theta,\xi_i,\xi_k)\epsilon_\theta^\omega(M,\theta,\xi_k)}{B_1(M,\theta,\xi_i)\epsilon_\theta^\omega(M,\theta,\xi_i)}$.

By (26), the assumption that ω is increasing in θ, and the definition of $B_1\left(M,\theta,\xi_i\right)$, to prove our result, it suffices to show that $\frac{\tau_1^w(M,\theta,\xi_i)}{1-\tau_1^w(M,\theta,\xi_i)}\geq 0$, which is equivalent to $\tau_1^w\left(M,\theta,\xi_i\right)\geq 0$ for all $\theta\in\Theta$ and $\xi_i\in\Xi$. (The definition τ_t^w implies that $\tau_t^w\leq 1$.) Note that by (24), $\tau_1^w\left(M,\theta,\xi_i\right)$ for any $\xi_i\in\Xi$ shares the same sign as $\lambda\left(\theta\right)$.

Following Werning (2000), suppose $\lambda\left(\theta\right)<0$ for some interval of skills, then there exists a maximal interval $\left[\theta^0,\theta^1\right]\subset\Theta$ such that $\lambda\left(\theta\right)\leq 0$ for any $\theta\in\left[\theta^0,\theta^1\right]$ with $\lambda\left(\theta^0\right)=\lambda\left(\theta^1\right)=0$

and $\lambda'(\theta^0) \leq 0$ and $\lambda'(\theta^1) \geq 0$. By (16), $\chi(\theta^0) \geq f(\theta^0)$ and $\chi(\theta^1) \leq f(\theta^1)$. Hence, by (17) and (18),

$$u'\left(c_0\left(\theta^0\right)\right) = \frac{1}{\theta^0} h'\left(\frac{y_0\left(\theta^0\right)}{\theta^0}\right) \leq u'\left(c_0\left(\theta^1\right)\right) = \frac{1}{\theta^1} h'\left(\frac{y_0\left(\theta^1\right)}{\theta^1}\right).$$

Similarly, by (19) and (20), for all $\xi_i \in \Xi$,

$$m_{M,\theta^0}(\xi_i)\, u'\left(c_1\left(\theta^0, \xi_i\right)\right) \leq m_{M,\theta^1}(\xi_i)\, u'\left(c_1\left(\theta^1, \xi_i\right)\right)$$

with $u'\left(c_1\left(\theta^0, \xi_i\right)\right) = \frac{1}{\omega(M,\theta^0,\xi_i)} h'\left(\frac{y_1\left(\theta^0,\xi_i\right)}{\omega(M,\theta^0,\xi_i)}\right)$ and $u'\left(c_1\left(\theta^1, \xi_i\right)\right) = \frac{1}{\omega(M,\theta^1,\xi_i)} h'\left(\frac{y_1\left(\theta^1,\xi_i\right)}{\omega(M,\theta^1,\xi_i)}\right)$. The flow utility in $t = 1$, $\tilde{U}_1$, is strictly concave and by the fact that

$$\tilde{U}_{1,c}(M, \theta, \xi_i) = p(\xi_i)\, m_{M,\theta}(\xi_i)\, u'\left(c_1(\theta, \xi_i)\right),$$

$$\tilde{U}_{1,y}(M, \theta, \xi_i) = -\frac{p(\xi_i)\, m_{M,\theta}(\xi_i)}{\omega(M, \theta, \xi_i)} h'\left(\frac{y_1(\theta, \xi_i)}{\omega(M, \theta, \xi_i)}\right),$$

$\frac{\partial \tilde{U}_1}{\partial c_1(\theta^0,\xi_i)} \leq \frac{\partial \tilde{U}_1}{\partial c_1(\theta^1,\xi_i)}$ and $\frac{\partial \tilde{U}_1}{\partial y_1(\theta^0,\xi_i)} \geq \frac{\partial \tilde{U}_1}{\partial y_1(\theta^1,\xi_i)}$ for all $\xi_i \in \Xi$. This implies that the marginal utility in consumption and leisure is weakly higher for θ^1-agents than θ^0-agents. However, $\theta^0 < \theta^1$, so the incentive compatibility constraint for θ^1-agent is violated. As a result, $\lambda(\theta) < 0$ is not possible, so $\tau_1^w(M, \theta, \xi_i) \geq 0$.

Hence, for any $\theta \in \Theta$ and $\xi_i \in \Xi$, $\operatorname{sgn}\left[\tau^k(M, \theta, \xi_i)\right] = \operatorname{sgn}\left[\sum_{k=0}^{S} \epsilon_y^m(M, \theta, \xi_i, \xi_k) \frac{\epsilon_\theta^\omega(M,\theta,\xi_k)}{\theta}\right]$. It also immediately follows that $\Gamma(M, \theta, \xi_i) = 1$ when $\tau_1^w(M, \theta, \xi_i) = 0$. From (19) and (20) and the boundary conditions for λ, we know that $\tau_1^w(M, \underline{\theta}, \xi_i) = \tau_1^w\left(M, \overline{\theta}, \xi_i\right) = 0$ for all $\xi_i \in \Xi$. ■

Proof of Theorem 4: From (16) and (17), we can show that

$$\begin{aligned}
\lambda(\theta) &= \frac{\theta f(\theta) \sum_{i=0}^{S} \mu(\xi_i)}{u'(c_0(\theta))} \frac{1 - F(\theta)}{\theta f(\theta)} \int_\theta^{\overline{\theta}} \frac{u'(c_0(\theta))}{u'(c_0(x))} \left[1 - \frac{\sum_{i=0}^{S} \mu(\xi_i)}{u'(c_0(x))}\right] \frac{f(x)}{1 - F(\theta)} dx \\
&= \frac{1}{B_0(\theta)} \frac{\tau_0^w(\theta)}{1 - \tau_0^w(\theta)} \frac{\theta f(\theta) \sum_{i=0}^{S} \mu(\xi_i)}{u'(c_0(\theta))},
\end{aligned}$$

where the second equality comes from Theorem 1. By (17), $\chi(\theta) = \frac{f(\theta)\sum_{i=0}^{S}\mu(\xi_i)}{u'(c_0(\theta))}$ and by the

definition of $\tau_1^w(M,\theta,\xi_i)$, $\frac{1}{\omega(M,\theta,\xi_i)}h'\left(\frac{y_1(\theta,\xi_i)}{\omega(M,\theta,\xi_i)}\right) = [1-\tau_1^w(M,\theta,\xi_i)]\,u'(c_1(M,\theta,\xi_i))$. Therefore, substituting these expressions into (21) yields

$$\int_\Theta \sum_{i=0}^{S} \frac{p(\xi_i)\,m_{M,\theta}(\xi_i)}{RMu'(c_0(\theta))} y_1(\theta,\xi_i)\,[1-\tau_1^w(M,\theta,\xi_i)]\,u'(c_1(M,\theta,\xi_i))\,\epsilon_M^\omega(M,\theta,\xi_i)\,dF(\theta)$$

$$-\int_\Theta \frac{\frac{\tau_0^w(\theta)}{1-\tau_0^w(\theta)}}{B_0(\theta)} \sum_{i=0}^{S} \frac{p(\xi_i)\,m_{M,\theta}(\xi_i)}{RMu'(c_0(\theta))} y_1(\theta,\xi_i)\,[1-\tau_1^w(M,\theta,\xi_i)]\,u'(c_1(M,\theta,\xi_i))\,\epsilon_\theta^\omega(M,\theta,\xi_i)$$

$$\times\left\{\epsilon_M^m(M,\theta,\xi_i)+\epsilon_M^\omega(M,\theta,\xi_i)\left[\left(\epsilon^H(M,\theta,\xi_i)-1\right)-B_1(M,\theta,\xi_i)\right]\right\}dF(\theta)=\phi.$$

The result follows from rearranging the expression above and the definition Ω. ■

B Additional data figures

In this section, we present a more detailed regression analysis of the main empirical findigs from Section 2, and we provide additional summary figures from the UAS survey data used throughout the paper.

In Table 6 we test the relationship between expectations about future pay change due to AI and income by running a multinomial logit regression of the three qualitative responses on household income and a host of demographic controls and wave fixed effects. We find that higher income is a strong predictor of lower likelihood of being a pessimist (relative to the baseline of "no change"). On the other hand, there is no association of income with the likelihood of being an optimist.

Table 6: Multinomial logit results

	Paid more	Paid less
Household income ('000 USD)	-0.000	-0.003***
	(0.001)	(0.000)
female	-0.507***	0.021
	(0.076)	(0.048)
age	-0.006**	-0.007***
	(0.003)	(0.002)
born_us	-0.419***	-0.239***
	(0.113)	(0.074)
black	1.009***	0.370***
	(0.105)	(0.071)
asian	0.540***	0.424***
	(0.142)	(0.096)
hispanic	0.464***	0.407***
	(0.105)	(0.063)
bachelors_or_higher	0.259***	0.086
	(0.086)	(0.052)
Constant	-1.681***	-0.286
	(0.213)	(0.131)
Observations	18738	

Base outcome: No change. Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table 7 we test the relationship between the perceived probability of future job loss due to AI and income by regressing the survey response on log household income while controlling for a host of demographic characteristics. The results show that income is indeed the strongest predictor of the perception of job loss probability and associates with it negatively. It is also notable that being female or non-white is associated with higher pessimism, while having college degree or being an immigrant is associated with higher optimism.

Table 7: Job loss probability on log income regression

	Job loss probability
log income	-0.047***
	(0.003)
female	0.019***
	(0.005)
age	-0.000**
	(0.000)
born_us	-0.070***
	(0.008)
black	0.092***
	(0.009)
asian	0.056***
	(0.010)
hispanic	0.043***
	(0.007)
bachelors_or_higher	-0.037***
	(0.005)
Constant	0.436***
	(0.019)
Observations	18710

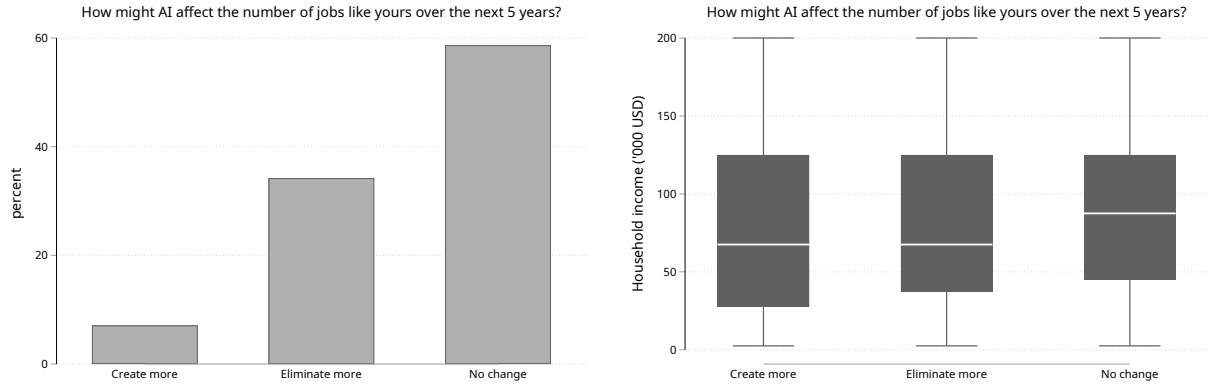
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 8 shows the distribution of responses and their relationship with household income to question $wk002$ which reads

> Thinking of the type of job you do, how do you think AI technologies might affect the number of jobs like yours over the next 5 years?

While most respondents expect no major change in the number of jobs, pessimists dominate optimists among the rest. The three groups of respondents feature a similar distribution of income.



(a) Raw responses

(b) Responses by income

Figure 8: "How might AI affect the number of jobs like yours over the next 5 years?"

Next, Figure 9 summarizes the distribution of responses and their relationship with household income to question $wk003$ which reads

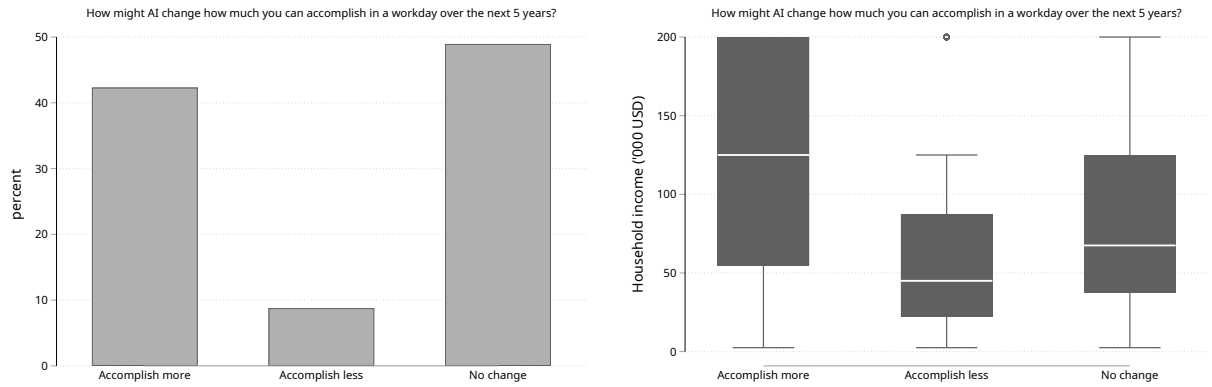
> Thinking of your current job, how do you think AI technologies might change how much you can accomplish in a typical workday over the next 5 years?

In this particular case, optimists dominate the pessimists and are almost as numerous as the "agnostics". Once again, however, optimists, tend to span the widest range of household incomes, while pessimists are strictly concentrated in the lower income range.

Figure 10 summarizes the distribution of responses and their relationship with household income to question $wk004$ which reads

> Thinking of your current job, how do you think AI technologies might change how much you enjoy your job in the next 5 years?

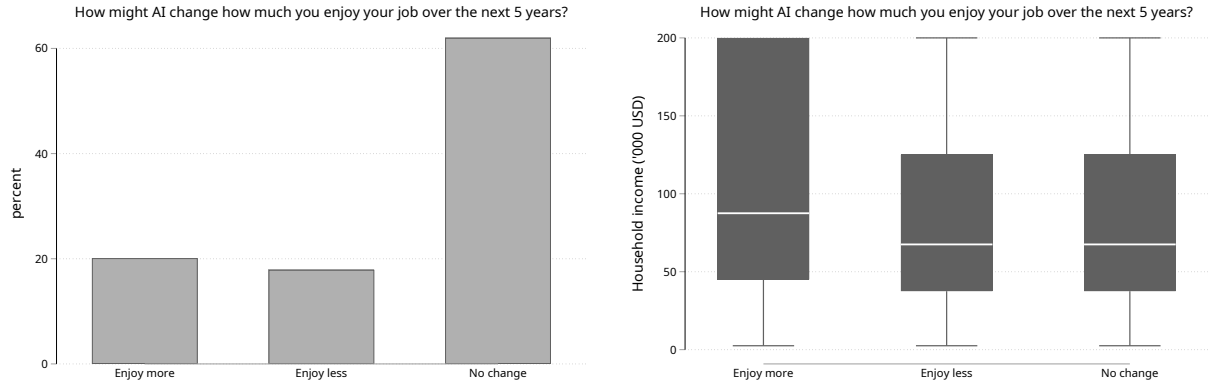
Sixty percent of respondents to this question do not anticipate a major change in their enjoyment of their job due to AI, while the rest are equally divided between optimists and pessimists. We observe a similar relationship of these attitudes with household income where pessimism is concentrated among lower incomes while optimism spans the entire range.



(a) Raw responses

(b) Responses by income

Figure 9: "How might AI change how much you can accomplish in a workday over the next 5 years?"



(a) Raw responses

(b) Responses by income

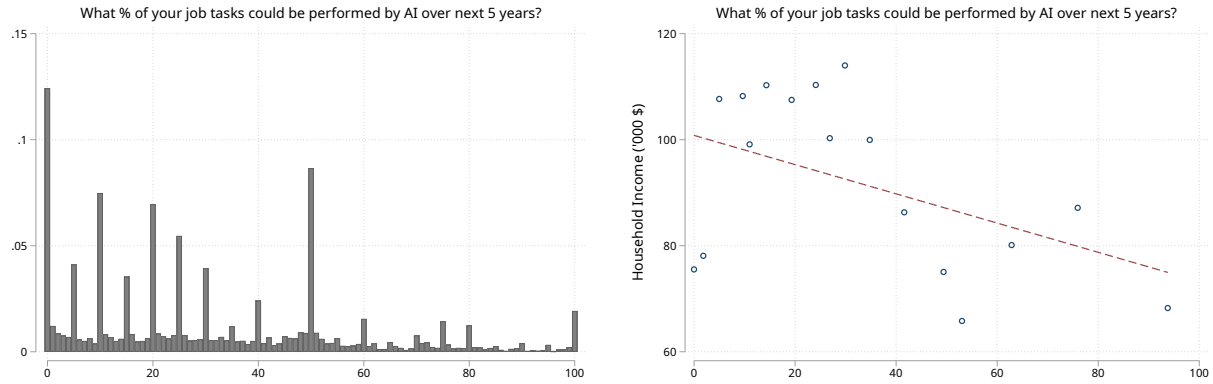
Figure 10: "How might AI change how much you enjoy your job over the next 5 years?"

Figure 11 summarizes the distribution of responses and their relationship with household income to the other numerical question $wk006$ which reads

> Now, think of the tasks that you typically perform as part of your job. On a scale from 0 to 100, where 0 means none and 100 means all of them, what portion of these tasks do you think could be performed either partially or entirely by AI over the next 5 years?

We observe a similar pattern as in the question concerning the probability of job loss, with responses dispersed over the entire $[0, 100\%]$ domain and with bunching around selected "round" numbers. As in the previous case, the responses exhibit a strong negative correlation

with household income.



(a) Raw responses

(b) Responses by income

Figure 11: "What fraction of your job tasks could be performed by AI over the next 5 years?"

Finally, because we have three annual waves of the survey with AI-focused questions, we analyze the time trends in the responses. Figure 12 summarizes the fraction of optimistic and pessimistic responses to the categorical questions $wk002$-$wk005$ over time. The results are broadly stable, with a slight trend towards pessimism when it comes to AI's impact on the number of jobs, and towards optimism when it comes to job enjoyment or ability to accomplish tasks at work.

We similarly look into the time trends in the two numerical questions $wk006$-$wk007$ in Figure 13 and observe a modest pessimistic trend in the expectations of job losses (with the fraction of pessimists rising and the fraction of respondents who believe there is zero probability of losing the job falling). On the other hand, the pattern of responses to the question about the fraction of tasks automated remains stable.

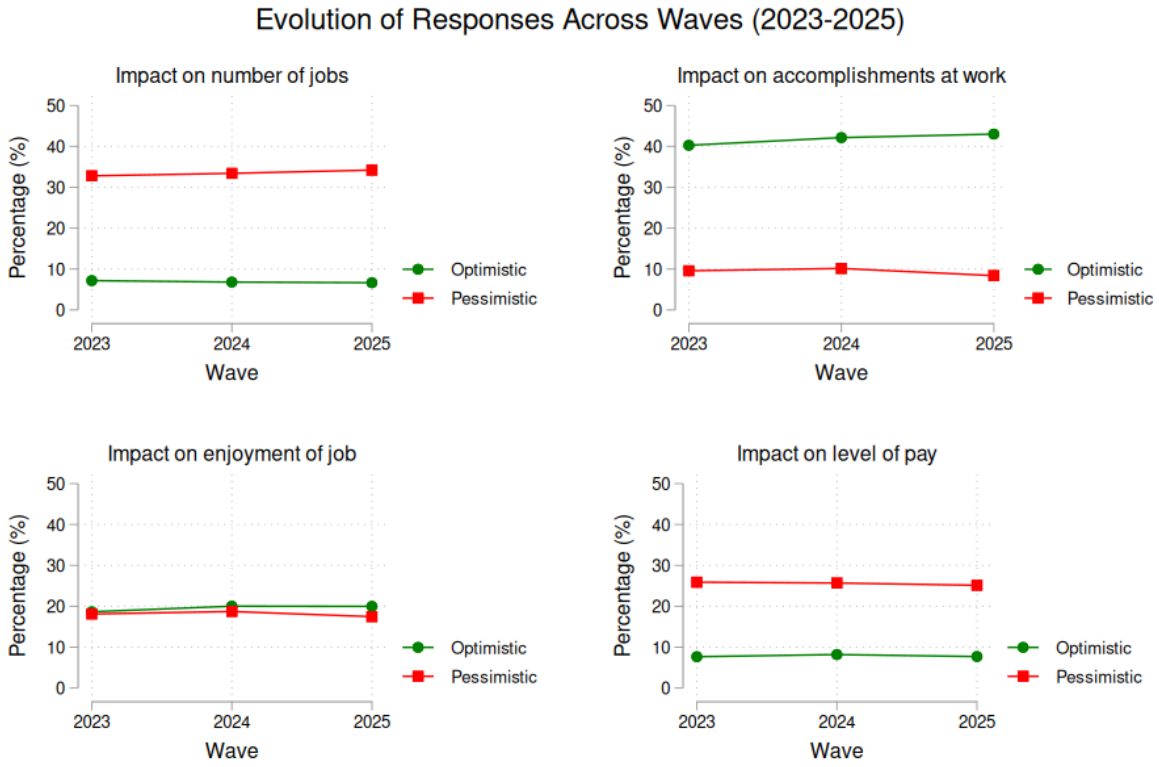
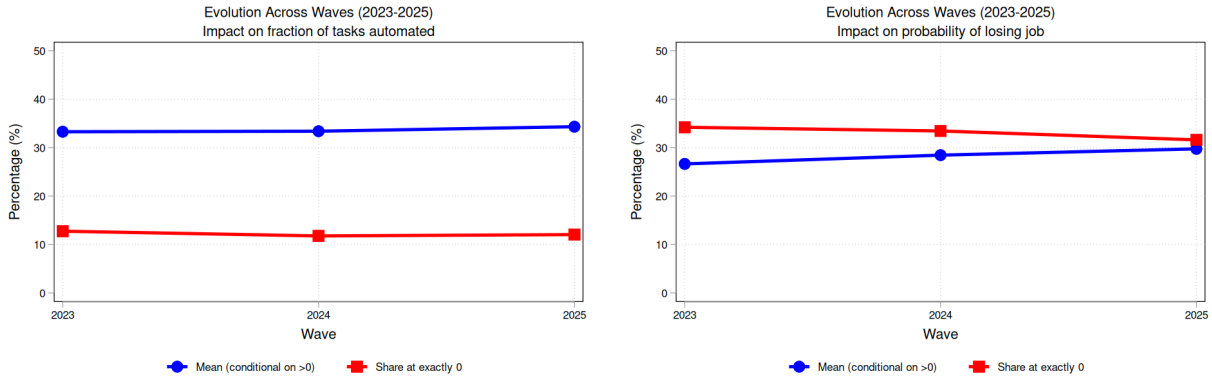


Figure 12: Time trends in responses to categorical questions



(a) Impact on fraction of automated tasks

(b) Impact on probability of job loss

Figure 13: Time trends in responses to quantitative questions