All in the Family: Aggregation and the Interpretation of Tax Bunching*

Anand Bharadwaj† Robert Breunig‡

August 27, 2026

Abstract

Bunching at tax kinks shows that reported income responds to incentives, but not whether total income changes or income is merely reallocated across taxpayers. We show that real responses constrain the distribution of total family income, whereas reallocation can generate individual bunching while leaving family income smooth. Using linked Australian tax records from 2001 to 2022 and a novel global bootstrap Kolmogorov–Smirnov test, we find sharp individual bunching in every year but no detectable bunching in family average income. The results suggest that individual bunching in Australia is driven primarily by intra-family income shifting rather than real responses.

Keywords: bunching; taxable income elasticity; income shifting; family taxation; Kolmogorov–Smirnov test; kernel smoothing
JEL codes: H24, H26, C14, D13.

*We are very grateful to Alan Auerbach, Muhammad Bashir, Tim Cejka, Tim de Silva, Mark Gergen, Michael Jansson, Stephen King, Martin Lettau, Emi Nakamura, Emmanuel Saez, Chris Shannon, Chris Skeels, Joel Slemrod, Jón Steinsson, Dustin Swonder, and seminar participants at the SITE Conference in Public Economics and at UC Berkeley for many valuable comments and suggestions. We also wish to thank the staff at the Australian Taxation Office, particularly Kim Cvetanovska, Son Nguyen, and Justin Holland, for facilitating data access and for ongoing guidance and advice. Bharadwaj gratefully acknowledges support from the Robert D. Burch Center for Tax Policy and Public Finance. This research has received ethics approval from the Human Ethics Committee of the Australian National University under protocol H/2025/0002.

†Department of Economics, University of California, Berkeley. Email: anand_bharadwaj@berkeley.edu.
‡Crawford School of Public Policy, Australian National University. Email: robert.breunig@anu.edu.au.

1 Introduction

Bunching at statutory kinks has become a standard empirical device for measuring how taxable income responds to marginal tax rates (Saez, 2010). In the canonical model, excess mass identifies an elasticity of taxable income and therefore informs the efficiency cost of taxation (Feldstein, 1999; Kopczuk, 2025). Yet the welfare interpretation depends on the behavioral margin underlying that response. A taxpayer may alter labor supply, effort, investment, or production. Alternatively, the same underlying resources may be preserved while only the person or account to which they are attributed, or the date on which they are reported, changes. As Gordon and Slemrod (1998) emphasize, an apparent real response may be income shifting in another guise.

Indeed, observing conspicuous bunching at the individual level does little to resolve this ambiguity. Blomquist et al. (2021) show that unrestricted preference heterogeneity can reconcile essentially any positive elasticity with the same amount of bunching, and even with the same taxable income distribution. Once reported income can also be reassigned across tax bases, the observed distribution reflects preferences, real responsiveness, and the technology that maps resources into reports. Estimating all three from the same individual marginal distribution would be circular because the restrictions used to recover the primitives would also determine the mechanism attributed to the bunching.

Nor is income shifting confined to a single margin. A self-employed taxpayer may move income between personal and business tax bases, while a business owner may retain earnings for a later year. Within a family, investment income, trust distributions, or returns from a closely held business may be assigned among people facing different marginal rates. A member observed at a kink may be a recipient whose lower bracket has just been filled. The same reported income may instead belong to a donor who reduces taxable income to the kink by transferring income that would otherwise fall in the higher bracket to a recipient facing a lower marginal rate. Although the transfers run in opposite directions, the personal return at the threshold is observationally identical.

This donor–recipient ambiguity is especially salient in Australia, where personal income is taxed separately rather than on a joint household basis. Wages are generally attached to the worker who earns them, whereas some business and investment income can be distributed through discretionary trusts, beneficiary allocations, and family controlled entities. A family may therefore face several personal tax prices at once. Filling one member's lower bracket before assigning the next dollar elsewhere can reduce the family's tax liability, yet the return observed at the kink does not reveal who generated that dollar or where the next one went. Linked family records allow us to examine the corresponding aggregate distribution directly.

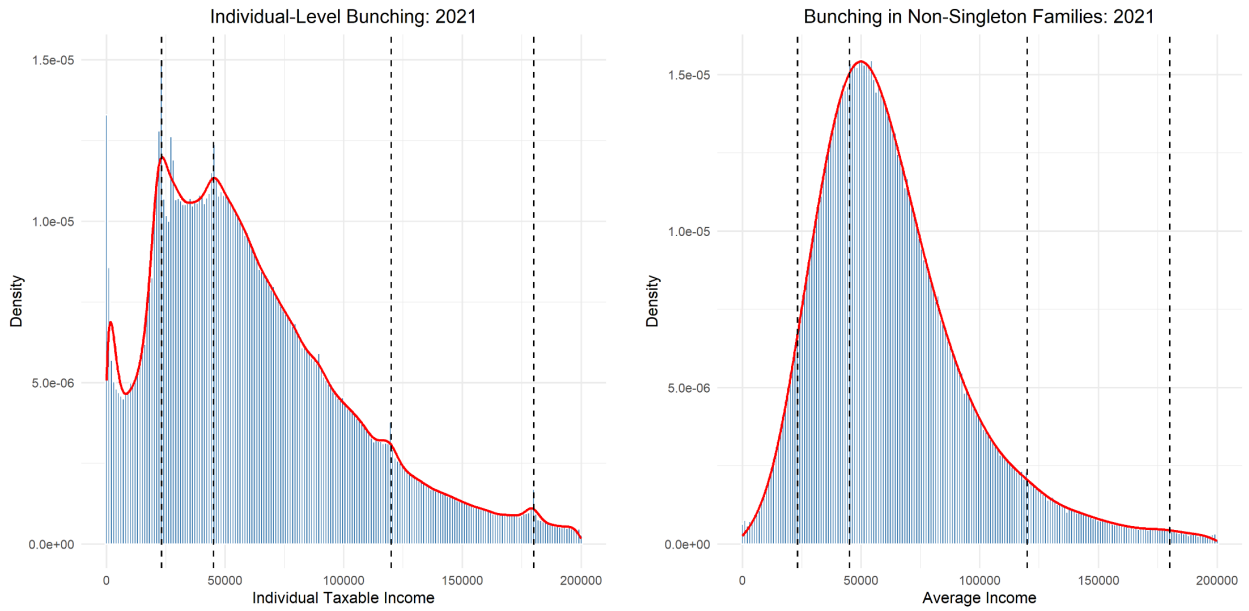
Several strands of existing work sharpen the interpretation of bunching by exploiting dynamics, adjustment frictions, or heterogeneity in reporting opportunities. Le Maire and Schjerning (2013) embed retained earnings in a dynamic model and estimate that intertemporal shifting explains more than half of bunching among the Danish self-employed, while Chetty et al. (2011) and Kleven and Waseem (2013) distinguish structural responsiveness from responses attenuated by adjustment costs. Following Saez (2010) and Chetty et al. (2013), a broad empirical literature also compares groups with different reporting opportunities. Such comparisons are informative, but they cannot exclude the possibility that taxpayers with greater reporting flexibility also have stronger real responses.

Our approach introduces a second observable distribution rather than attempting to recover every primitive from individual bunching. We aggregate income over the economic unit within which reported income can be reassigned. In our application, that unit is the linked family. Transfers among family members alter personal reports while leaving the family total unchanged, whereas changes in production or other underlying resources alter that total. Importantly, the identifying content of aggregation does not depend on a capacity to shift. Even without shifting, real responses impose a restriction on the aggregate distribution: when family members bunch jointly at their respective kinks, total family income has an atom at the sum of those kink incomes. Aggregation can attenuate this atom because the relevant members must bunch jointly, but it does not eliminate the restriction. Mechanisms that generate the same individual income distributions can therefore imply different distributions for the family total.

Coordination can strengthen this restriction. Individual tax kinks can induce kinks in the family's after-tax frontier, so that a positive mass of families choose the same aggregate income. At a common statutory kink, the coordinated models preserve families that already bunch jointly without shifting, while composite kinks can bring additional families to the same aggregate income. Mild economic restrictions on the scope for shifting can therefore sharply strengthen the family signal implied by a given level of individual real responses. This framework supports a falsification approach to explanations based on real responses: individual bunching attributed to real responses must also be consistent with the implied behavior of family income.

Figure 1 illustrates the empirical contrast that motivates the analysis. Australian individual taxable income exhibits pronounced bunching at statutory thresholds, whereas linked family averages are much smoother. Because aggregation can attenuate genuine bunching, the visual contrast alone is not conclusive. The empirical question is whether the family response implied by an explanation based on real responses should be detectable.

For applications beyond the family, we use the term *shiftable syndicate* for a collection



(a) Individual taxable income.

(b) Average taxable income in linked families.

Figure 1: Individual and linked family taxable income distributions in Australian tax data, 2021. Vertical lines mark statutory thresholds. The family panel excludes singleton records and aligns common individual thresholds by dividing total family taxable income by family size. The plots use the wider range from zero to $200,000 for display. Every formal test uses the prespecified interior interval [$1,000, $199,000]. Source: Australian Taxation Office Longitudinal Information Files and linked family records.

of people, accounts, entities, or dates across which income can be reassigned subject to an aggregate accounting identity. A syndicate need not consist of different people. In Le Maire and Schjerning (2013), for example, the relevant units are the same taxpayer in consecutive years, and the tax schedules or kink locations may differ across dates. The choice of syndicate is economic rather than mechanical. It should contain the reports among which reassignment is plausible and exclude unrelated units that merely increase attenuation. In the Australian application, the syndicate consists of family members observed in the same tax year and facing the same statutory personal income tax schedule. The linked family remains the economic unit studied in this paper, but the theory applies much more broadly.

To establish the aggregate restriction and assess its quantitative force, we study three models that differ in the scope for reallocating income within the family. With unrestricted shifting, the family can allocate a given total across personal tax bases, creating a composite kink in its after-tax frontier. A more institutionally plausible setting allows passive income to be allocated across family members while keeping active earnings attached to their earners. We then allow reallocation to carry finite, strictly convex costs. Across these environments, the bunching conditions are not confined to isolated family types: under mild regularity and

support conditions, they hold with positive probability and therefore generate an atom in aggregate family income.

A complementary nesting result shows that the coordinated models also inherit bunching that is already present without reallocation. At any given kink of a common tax schedule, families whose members would all locate at that kink without shifting continue to place every report there under both unrestricted and costly shifting. A parallel result applies when passive income is pooled, provided the required allocations are feasible. The joint mass of these baseline bunchers therefore survives coordination and, where the nesting conditions apply, provides a conservative lower bound on the family atom. Coordination can also bring additional families to the composite kink even when their members would not all bunch there separately. The two results thus capture complementary forces: aggregation can attenuate individual bunching, while the ability to reallocate income preserves the baseline response and potentially generates additional aggregate bunching.

Aggregate bunching need not occur at a single known cutoff corresponding to an individual tax kink. A family may contain several members and face several relevant kinks. More generally, the units being aggregated may face different schedules, while shifting costs can move or disperse bunching across composite thresholds. The empirical restriction is therefore global rather than cutoff specific, motivating us to develop a novel test for smoothness over the relevant part of the aggregate distribution rather than testing separately at a collection of selected thresholds. This differs from local density and manipulation tests at a prespecified cutoff, such as McCrary (2008) and Cattaneo et al. (2018, 2020). Bickel and Rosenblatt (1973) study global deviations of kernel density estimates from a hypothesized density. Our null instead leaves the smooth distribution unspecified.

The statistic compares the empirical CDF with a fitted smooth CDF over a prespecified interval. Its calibration is nonstandard because the benchmark distribution is estimated rather than known and the empirical CDF and its smoothed counterpart are built from the same observations. Their global fluctuations cancel, leaving a smaller local stochastic scale on which ordinary density estimation bandwidths produce first-order bias. We show that two standard solutions, sample splitting and a conventional smooth bootstrap, fail for different reasons. Sample splitting avoids the mechanical dependence created when the same observations are used both to estimate the smooth benchmark and to test deviations from it but replaces that dependence with additional sampling variation from an independently estimated benchmark. A conventional smooth bootstrap creates a different problem. With bandwidths of the usual estimation order, the bootstrap distribution inherits a smoothing drift that is non-negligible at the scale of the test statistic, so standard bandwidth choices generally fail to deliver valid size. Simply resampling from a smoothed estimate is therefore

not enough.

We therefore develop a bootstrap Kolmogorov–Smirnov test that addresses both problems. The procedure uses the full sample to construct the benchmark, quantitatively undersmooths relative to conventional estimation bandwidths so that the remaining smoothing error is asymptotically negligible, and explicitly centers the bootstrap process to remove its own smoothing drift. A Gaussian multiplier implementation avoids repeated re-estimation and makes the test practical in large administrative samples. We prove that under primitive smoothness conditions, the procedure has asymptotically correct size and power against atoms and density discontinuities.

Applied to Australian personal tax returns linked into family units for every year from 2001 through 2022, the procedure yields a strikingly stable contrast. Individual taxable income rejects the null of a smooth distribution in all 22 years, while smoothness is never rejected for family average income under our primary specification. We then use the observed individual bunching to quantify how much aggregation can attenuate a genuine real response. Aggregation alone can make such a response difficult to detect when family members respond independently. Once economically meaningful scope for within-family reallocation is introduced, however, the same real responsiveness implies substantially stronger bunching in family income. The persistent smoothness of the family distribution is therefore difficult to reconcile with individual bunching driven primarily by real responses.

Taken together, the analysis distinguishes among mechanisms that imply different behavior of family totals, even when taxable income elasticities, household preferences, shifting costs, and access to reporting technologies are only partially identified. This is especially useful where conventional elasticity-based interpretations are least informative. The same logic extends beyond family taxation: dates, accounts, related entities, and organizational divisions can provide analogous aggregation tests whenever relabeling disappears on aggregation but genuine changes in income or activity remain visible in the total.

The structure of the paper is as follows. Section 2 describes the Australian setting and construction of linked families, and Section 3 develops the family restriction and its attenuation benchmarks. Sections 4 and 5 present the test and its finite-sample performance. Section 6 applies the procedure to the tax data and calibrates the family implications of a real-response explanation of the individual bunching.

2 Institutional setting and linked family data

2.1 Income shifting within Australian families

Australia is particularly well suited to this exercise because personal income is taxed on an individual rather than a household basis. Spouses file separate returns, so the marginal tax price of an additional dollar depends on the person to whom it is attributed. In 2021–22, for example, the resident marginal rate rose from 19 to 32.5 percent at $45,000 and from 32.5 to 37 percent at $120,000, excluding the Medicare levy (Australian Taxation Office, 2022). At lower incomes, the low-income tax offset (LITO) reduced personal income tax liability. From 2018–19 through 2021–22, taxpayers could also receive the low and middle income tax offset (LMITO). The Medicare levy had a separate low-income threshold, and the beneficiary tax offset reduced tax on qualifying government payments (Australian Taxation Office, 2025). Wages are generally tied to the worker who earns them. Family-controlled business and investment income, however, may be allocated through discretionary trust distributions, beneficiary designations, and closely held entities, subject to tax and trust law. Australia also has no general gift or estate tax. These institutions make the linked family a natural unit for asking whether individual bunching survives the removal of within-family attribution.

Both the tax incentive and the resulting empirical ambiguity are visible in a simple example. Suppose one spouse has $40,000 of active earnings and the family controls distributable trust income. Assigning $5,000 to that spouse fills the 19-percent bracket at $45,000, after which additional distributable income can be assigned to another eligible beneficiary when doing so lowers the family's tax bill. Conversely, income that would otherwise take a member above $45,000 can be allocated away from that member. The person observed at the kink may therefore be either a recipient whose bracket was filled or a donor from whose personal tax base income was diverted. Because individual bunching does not reveal the direction of the transfer, family aggregation is informative precisely because it removes within-family attribution while preserving the family's total income.

The analysis combines the Australian Taxation Office Longitudinal Information Files (ALife), a longitudinal sample of personal income tax returns (Abhayaratna et al., 2022), with the ATO ALife Family File (ALife-Family) (Fischer et al., 2025). ALife-Family begins with a random primary sample from the ATO client register and links those individuals to recorded current and former spouses, parents, children, and siblings, together with secondary links among the relatives themselves. For taxpayer i in year t, let Y_{it} denote taxable income. We use these links to construct two-generation family units consisting of an individual, the individual's spouse, and their children. We deliberately keep these units relatively small to limit attenuation from aggregation. Once the unit contains the family members among whom

the relevant income can plausibly be reallocated, enlarging it further is unnecessary for the identification argument and would generally attenuate a genuine aggregate response. Because the same family can be generated from the records of more than one member, we remove duplicate representations of the same family unit. A person may nevertheless appear in more than one retained family when the recorded links define genuinely different family units. We retain one observation for each resulting family-year pair, sum taxable income across its members, and divide by family size to obtain family average income. Using the family average places families of different sizes on the same income scale, so statutory thresholds appear at the same locations in the individual and family distributions.

Because ALife-Family combines relationship information from several administrative sources, coverage varies across relationship types and cohorts; Fischer et al. (2025) provide a detailed assessment of the sample and its representativeness. Spousal links are updated annually, while parent–child links rely on evidence from the ATO client register, Medicare enrolment, spouse records, and shared addresses. Some tax returns also report spouse income without a corresponding spouse identifier. We therefore use a conservative family definition. A *singleton* is an individual with no recorded spouse or child link in the family file, and our principal family sample excludes singletons and retains only internally consistent family units. The individual distribution gives equal weight to each retained person-year with taxable income strictly between zero and \$200,000. The family distribution instead gives equal weight to each retained family whose average taxable income lies within the same bounds. These weights correspond to the economic unit represented by each distribution: individuals in the first and families in the second. The resulting sample sizes count retained person-years and retained family-year units, respectively. Relationships not recorded in these administrative sources remain outside measured family income.

2.2 Testable implications for the data

Different shifting technologies imply different relationships between the individual and family distributions. If individual bunching reflects a change in total family income, the theory in Section 3 predicts a corresponding departure from smoothness in the family distribution. Without shifting, this arises when family members bunch jointly, potentially at different statutory kinks. At a kink common to all members, those joint bunchers continue to bunch when free or costly shifting is introduced, while coordination can bring additional families to the composite kink. By contrast, reallocating a smooth family income total can generate bunching in personal reports while leaving the family distribution smooth. These theoretical implications lead naturally to annual tests of both individual taxable income and family

average income. Because families combine several reports and may face multiple relevant tax kinks, the family response need not appear at a single known cutoff. We therefore test for departures from smoothness across the relevant part of each distribution.

3 Family aggregation and income shifting

3.1 Aggregation as an identifying restriction

The identification problem is clearest for an individual observed at a tax kink. The individual may be a real responder who chooses generated income at the kink. The same reported value can also arise through income shifting. A person whose generated income lies below the kink may receive enough allocable income from another family member to reach it, while one whose generated income lies above the kink may transfer the excess to another member. The personal return does not reveal whether the individual is a real responder, a recipient, or a donor. Individual bunching therefore cannot by itself distinguish changes in generated income from changes in who reports that income.

Aggregation provides an additional restriction because income shifted within the family nets to zero. Let X_j denote member j's reported taxable income, Y_j the income generated by that member before shifting, and $s_j = X_j - Y_j$ the amount shifted into or out of the report. When shifting only reallocates income within the family,

$$\sum_{j=1}^{J} X_j = \sum_{j=1}^{J} Y_j \qquad \text{because} \qquad \sum_{j=1}^{J} s_j = 0. \tag{1}$$

Income shifting can therefore alter the allocation of reported income across family members without changing the family total. By contrast, changes in generated income alter the Y_j and can therefore alter that total. Two explanations that generate the same individual income distribution can consequently have different implications for family income. This distinction does not require every source of income to be shiftable. It requires only that the reassignment responsible for the individual bunching occur within the unit being aggregated.

The same logic extends beyond families. We call a finite collection of reports across which income can be reassigned while their total is preserved a *shiftable syndicate*. In our application, j indexes family members observed in the same tax year. In an intertemporal application such as Le Maire and Schjerning (2013), the index may instead refer to the same taxpayer in different years, so the relevant tax schedules and kink locations may differ across dates. More generally, the reports may belong to different dates, accounts, or related entities. We write a_j for the after-tax schedule associated with j, k_j for a candidate kink of that

schedule, and

$$k = (k_1, \ldots, k_J), \qquad K = \sum_{j=1}^{J} k_j.$$

The general theory allows both schedules and kink locations to vary across the elements of the syndicate. The main empirical application specializes to family members facing the same statutory schedule.

The identifying value of aggregation does not rest on the premise that individual bunching must reappear unchanged in the aggregate distribution. Rather, an explanation of the individual bunching must also account for what happens to family income. At the individual level, $X_j = k_j$ may reflect either a choice of generated income at k_j or income shifted into or out of member j's report. If individual bunching is generated only by shifting income across family members, aggregation can remove it. If it instead reflects changes in generated income, those changes enter total family income rather than canceling by construction. The models below show when this produces bunching in the family distribution.

We use a family's *type* to summarize the primitives that determine its choices, including member cost parameters, family preferences, and, where relevant, passive income endowments. Under free shifting, the connection between individual tax kinks and family income is particularly transparent. If the selected kinks share a nondegenerate interval of supporting slopes, the family after-tax frontier has a composite kink at K. For families whose effective marginal cost of generating total income lies within the range of marginal benefits created by this composite kink, K is optimal. A convenient sufficient condition for positive mass is that the relevant inequality hold strictly at a reference type $\bar{\omega}$. Continuity then implies that all types in some open neighborhood of $\bar{\omega}$ also choose K. If $\bar{\omega}$ lies in the support of the type distribution, meaning that every open neighborhood of $\bar{\omega}$ has positive probability, this neighborhood contains a positive mass of families and the aggregate distribution has an atom at K. Thus, free reassignment does not eliminate the link between changes in generated income and aggregate bunching. Coordination can itself create a kink in the family problem.

The subsequent models we study retain the same logic under more restricted forms of shifting. With allocable passive income, active income remains with the member who generates it, while a fixed pool of passive income can be assigned across family members. A common shadow value clears the passive income pool among recipients, while members receiving no passive income satisfy the corresponding Kuhn–Tucker inequalities. This also permits different members to be supported at different statutory kinks. With costly shifting, the transfers required to place reports at their selected kinks must instead be consistent with the marginal costs of shifting across the family. In each case, the theorem gives strict conditions at a reference type that, by continuity, continue to hold for an open neighborhood

of nearby types. If that reference type lies in the support of the type distribution, the neighborhood has positive probability and therefore produces an atom in aggregate income.

The case of a common schedule yields a complementary result that is especially useful for the empirical analysis. Suppose every member would choose the same statutory kink k in the absence of shifting. Those families have total income $K = Jk$. Allowing free shifting does not displace them from that aggregate income, and under the maintained conditions the same is true with costly shifting. A parallel nesting argument applies to allocable passive income when the required allocations are feasible. The baseline bunching set without shifting is therefore preserved under these forms of coordination. Coordination can also bring additional families to K even when their members would not all have chosen k separately. Shifting can therefore preserve or strengthen the aggregate implication of individual real responses rather than wash it out.

Aggregation can nevertheless attenuate bunching. Without shifting, suppose member j's income distribution has an atom at its selected kink k_j. If other income configurations sum to

$$K = \sum_j k_j$$

with probability zero, the resulting family atom at K is the joint probability that every relevant family member chooses the corresponding kink k_j. Under independence, this probability is the product of the individual bunching probabilities and can be substantially smaller than any one of them. This calculation therefore provides a conservative benchmark for the aggregate signal. At a common statutory kink, the nesting result carries this joint mass into the free and costly shifting environments as a lower bound, while the composite kink results can add further families to the aggregate bunching set.

These results turn the distinction between changes in generated income and pure reassignment into a restriction on the family income distribution. The benchmark without shifting quantifies how much individual bunching can be attenuated by aggregation alone, while the shifting models show how coordination can preserve or strengthen the family response. The empirical question then is whether the resulting family bunching is large enough to detect. Because family composition, multiple tax kinks, and shifting frictions can generate departures from smoothness at several or imperfectly known locations, the statistical analysis tests for departures from smoothness across the relevant part of the family income distribution rather than at a single cutoff.

3.2 Free shifting and the family after-tax frontier

We now specialize the shiftable syndicate introduced above to a family with J members and allow reported income to be shifted freely among them. Member j generates real taxable income $y_j \geq 0$ at cost $c_j(y_j; \theta_j)$, where θ_j indexes member-specific cost heterogeneity and $c_j(\cdot; \theta_j)$ is differentiable and strictly convex. Reported income $x_j \geq 0$ faces the piecewise-linear tax schedule T_j, and

$$a_j(x) = x - T_j(x)$$

denotes after-tax income. Free shifting requires only that reported income add up to the income generated by the family:

$$\sum_{j=1}^{J} x_j = \sum_{j=1}^{J} y_j.$$

Preferences are summarized by an increasing, concave indirect utility $v(R; \eta)$ over total after-tax income

$$R = \sum_{j=1}^{J} a_j(x_j),$$

where η indexes family preference heterogeneity. Let $\theta = (\theta_1, \ldots, \theta_J)$ and $\omega = (\theta, \eta)$ denote the family type. Throughout, the family chooses both the amount of income generated by its members and the allocation of that income across personal reports.

For later use, if g is concave, write

$$\partial g(z) = \{p \in \mathbb{R} : g(z') \leq g(z) + p(z' - z) \text{ for every } z' \text{ in the domain of } g\}$$

for its superdifferential at z. An element of $\partial g(z)$ is a *supergradient*: the slope of a line that touches g at z and lies weakly above it everywhere. We refer to such a slope as a *supporting slope*. When g is differentiable at z, the superdifferential contains only its derivative. At a kink, it is the interval between the right and left derivatives, which we call the *supporting interval* at that kink. When this notation is applied to $v(\cdot; \eta)$, the superdifferential is taken with respect to its first argument.

Remark 1 (Arbitrary positive Pareto weights). The formulation allows arbitrary fixed positive Pareto weights. Let $w_j > 0$ and let $u_j(z_j, g; \eta_j)$ denote member j's utility from private consumption z_j and public goods g. Conditional on total after-tax income R, let the family choose a feasible allocation of private consumption and public goods to maximize

$$\sum_j w_j u_j(z_j, g; \eta_j).$$

If the u_j are increasing and concave, the resulting value function is itself increasing and concave in R. The reduced form $v(R;\eta)$ can therefore accommodate unequal Pareto weights, nontransferable utility, and family-specific preferences.

If instead the family objective is written directly as

$$\sum_j w_j a_j(x_j) - \sum_j c_j(y_j;\theta_j),$$

the same arguments apply after replacing the supporting interval $\partial a_j(k_j)$ by $w_j \partial a_j(k_j)$ for each member.

Free shifting permits the family problem to be reduced to a choice of total real income. For any total Y, define the after-tax frontier

$$A(Y) = \max_{x_1,\dots,x_J \geq 0} \left\{ \sum_{j=1}^{J} a_j(x_j) : \sum_{j=1}^{J} x_j = Y \right\} \tag{2}$$

and the minimum cost of generating that total,

$$C(Y;\theta) = \min_{y_1,\dots,y_J \geq 0} \left\{ \sum_{j=1}^{J} c_j(y_j;\theta_j) : \sum_{j=1}^{J} y_j = Y \right\}. \tag{3}$$

Thus $A(Y)$ is the greatest after-tax income attainable by allocating a total report Y across family members, while $C(Y;\theta)$ is the minimum cost of generating total real income Y. The family problem becomes

$$\max_{Y \geq 0} \{ v(A(Y);\eta) - C(Y;\theta) \}.$$

This representation separates two choices that need not coincide member by member: how much income the family generates in total and how that total is allocated across personal reports.

Lemma 1 (Family value functions)**.** *Suppose each a_j is increasing, concave, and upper semicontinuous, and each $c_j(\cdot;\theta_j)$ is lower semicontinuous and strictly convex. Then the programs in* (2)–(3) *attain their optima for every finite $Y \geq 0$. Moreover, A is increasing and concave, and $C(\cdot;\theta)$ is strictly convex.*

Proof. See Appendix A.1. □

Let k_j be a kink in member j's tax schedule and define

$$K = \sum_{j=1}^{J} k_j.$$

At k_j, the superdifferential $\partial a_j(k_j)$ is the interval between the marginal net-of-tax rates immediately above and below the kink.

Remark 2 (Common schedules and multiple kinks). A tax schedule may contain several kinks, with different marginal rates at each. The vector

$$k = (k_1, \ldots, k_J)$$

selects one candidate kink for each member. In the general syndicate formulation, these locations may differ when the elements face different schedules or when their kink intervals share a nondegenerate interval of supporting slopes.

In the cross-sectional family application, all members face the same concave statutory schedule a. Jensen's inequality then gives

$$A(Y) = Ja(Y/J).$$

Hence the family after-tax frontier has a kink at $Y = Jk^{(m)}$ whenever the common report Y/J reaches statutory kink $k^{(m)}$. Members can sometimes be placed at different statutory kinks for a given total when the corresponding superdifferentials intersect. Under an ordinary progressive piecewise-linear schedule, however, the intervals associated with different kinks generally intersect in at most a single slope. Such an allocation lies on a linear segment of A and does not generate the nondegenerate composite kink studied below.

To build intuition, let us first consider a simple two-member family. Suppose both members face the same statutory kink k, with marginal net-of-tax rates $q_- > q_+$, family utility is linear in total after-tax income, and member j has the quadratic real income adjustment cost

$$c_j(y; \theta_j) = \frac{(y - \theta_j)^2}{2\tau}, \qquad \tau > 0.$$

The parameter θ_j determines the member's preferred level of income in the absence of tax incentives, and a higher θ_j shifts desired income upward. For this illustration, we interpret incomes and types as deviations from a sufficiently large positive baseline, so the nonnegativity constraints do not bind.

Without shifting, member j chooses income to maximize

$$a(y_j) - c_j(y_j; \theta_j).$$

At the statutory kink,

$$\partial a(k) = [q_+, q_-].$$

The objective is strictly concave, and its superdifferential at k is

$$\partial a(k) - c_{j,y}(k; \theta_j).$$

Hence k is optimal if and only if

$$0 \in \partial a(k) - c_{j,y}(k; \theta_j),$$

or equivalently,

$$c_{j,y}(k; \theta_j) \in \partial a(k).$$

Since

$$c_{j,y}(k; \theta_j) = \frac{k - \theta_j}{\tau},$$

member j bunches if and only if

$$\theta_j \in [k - \tau q_-,\, k - \tau q_+]\,. \tag{4}$$

Thus, without shifting, both members bunch only when each member's type lies in the individual bunching interval.

Free shifting changes the condition because the family can choose who generates a given total income. Conditional on total real income Y, it solves

$$C(Y; \theta) = \min_{y_1 + y_2 = Y} \left\{ \frac{(y_1 - \theta_1)^2}{2\tau} + \frac{(y_2 - \theta_2)^2}{2\tau} \right\}.$$

The first-order conditions equalize marginal earning costs:

$$\frac{y_1 - \theta_1}{\tau} = \frac{y_2 - \theta_2}{\tau}.$$

Together with $y_1 + y_2 = Y$, this gives

$$y_j = \theta_j + \frac{Y - \theta_1 - \theta_2}{2}.$$

Substitution yields

$$C(Y;\theta) = \frac{(Y-\theta_1-\theta_2)^2}{4\tau}, \qquad C_Y(Y;\theta) = \frac{Y-\theta_1-\theta_2}{2\tau}.$$

Because the two members face the same concave after-tax schedule,

$$A(Y) = \max_{x_1+x_2=Y} \{a(x_1) + a(x_2)\} = 2a(Y/2).$$

It follows that

$$\partial A(2k) = \partial a(k) = [q_+, q_-].$$

With linear family utility, the scalar family objective is

$$A(Y) - C(Y;\theta).$$

Its superdifferential at $2k$ is

$$\partial A(2k) - C_Y(2k;\theta).$$

The objective is strictly concave, so $2k$ is optimal if and only if

$$0 \in \partial A(2k) - C_Y(2k;\theta),$$

or equivalently,

$$C_Y(2k;\theta) \in \partial A(2k).$$

Since

$$C_Y(2k;\theta) = \frac{2k-\theta_1-\theta_2}{2\tau} = \frac{k-(\theta_1+\theta_2)/2}{\tau},$$

the free shifting bunching condition is

$$\frac{\theta_1+\theta_2}{2} \in [k-\tau q_-,\, k-\tau q_+]\,. \tag{5}$$

Equations (4) and (5) make the effect of coordination immediate. Without shifting, each θ_j must lie in the individual bunching interval. With free shifting, only their average must lie there. Geometrically, the bunching set without shifting is the square formed by the Cartesian product of the individual intervals, whereas free shifting expands it to the diagonal strip in which the average type lies in that interval. The first condition implies the second, but not conversely. One member can lie below the individual bunching interval and the other above it while their average remains inside.

The corresponding allocation shows how this additional bunching arises. At $Y = 2k$, the

cost-minimizing real incomes are

$$y_j = k + \theta_j - \frac{\theta_1 + \theta_2}{2}.$$

A member with below average θ_j generates less than k and receives income, while a member with above-average θ_j generates more than k and shifts income away. Both nevertheless report k, and total generated and reported income are both equal to $2k$.

The general result preserves this logic without quadratic costs or linear utility. The family need not consist of members who would each bunch separately. The relevant condition is instead aggregate: whether the marginal cost of generating K can be supported by the marginal values created by the composite kink. We return to this quadratic specification in the quantitative analysis.

Suppose now that a nondegenerate interval of slopes supports every selected individual kink:

$$[q_+, q_-] \subseteq \bigcap_{j=1}^{J} \partial a_j(k_j), \qquad q_+ < q_-.$$

When total reports equal K, any slope in this common interval makes the report vector $k = (k_1, \ldots, k_J)$ optimal. To state the corresponding condition for the choice of total income without assuming differentiability of v, write

$$\partial v(A(K); \eta) = [\underline{\nu}(\eta), \overline{\nu}(\eta)]$$

for the superdifferential of $v(\cdot; \eta)$ at $A(K)$. Because $v(\cdot; \eta)$ is increasing,

$$0 \leq \underline{\nu}(\eta) \leq \overline{\nu}(\eta).$$

Define

$$\mathcal{M}(\eta) = \{\nu q : \nu \in \partial v(A(K); \eta), \quad q \in [q_+, q_-]\}.$$

Since the intervals $\partial v(A(K); \eta)$ and $[q_+, q_-]$ contain only nonnegative values,

$$\mathcal{M}(\eta) = [q_+ \underline{\nu}(\eta),\ q_- \overline{\nu}(\eta)].$$

This interval collects the marginal values of total real income that can be supported at the composite kink.

Proposition 1 (Composite family kink)**.** *Suppose*

$$[q_+, q_-] \subseteq \bigcap_{j=1}^{J} \partial a_j(k_j), \qquad q_+ < q_-,$$

and let $K = \sum_{j=1}^{J} k_j$*. Then* $x_j = k_j$ *for every* j *solves* (2) *at* $Y = K$*, and*

$$[q_+, q_-] \subseteq \partial A(K).$$

Let $\bar{\omega} = (\bar{\theta}, \bar{\eta})$ *be a reference family type. Suppose that, throughout a neighborhood of* $\bar{\omega}$*,* $C(\cdot; \theta)$ *is differentiable at* K*, the superdifferential* $\partial v(A(K); \eta)$ *is a nonempty compact interval,* $C_Y(K; \theta)$ *is continuous, and the endpoints* $\underline{\nu}(\eta)$ *and* $\overline{\nu}(\eta)$ *are continuous. If*

$$C_Y(K; \bar{\theta}) \in \operatorname{int} \mathcal{M}(\bar{\eta}), \tag{6}$$

then every family with type in some open neighborhood of $\bar{\omega}$ *uniquely chooses total real income* K*. Under free shifting, total reported income also equals* K*.*

If $\bar{\omega}$ *lies in the support of the distribution of family types, meaning that every open neighborhood of* $\bar{\omega}$ *has positive probability, then total real and reported family income have an atom at* K*.*

Proof. See Appendix A.2. □

The strict interior condition in Proposition 1 is needed for the open neighborhood conclusion, not for optimality of K for a given family type. For any fixed type,

$$C_Y(K; \theta) \in \mathcal{M}(\eta)$$

is already sufficient for K to be optimal by the same supergradient argument. Requiring interior membership at a reference type makes the condition persist under small changes in the family type. If the reference type lies in the support of the type distribution, the resulting neighborhood has positive probability and produces an atom.

The proposition therefore separates two distinct requirements. First, the common interval of supporting slopes makes the report vector k optimal conditional on total reports being K. Second,

$$C_Y(K; \theta) \in \mathcal{M}(\eta)$$

makes K optimal as the family's total real income. The quadratic example makes the economic content of the second condition particularly transparent: coordination allows differences in

members' desired real incomes to offset within the family, so aggregate bunching can occur for families whose members would not all bunch separately.

3.3 Costless allocation of passive income

In many contexts, including our Australian application, active earnings remain attached to the person who earns them even though passive income can be assigned within the family. Member j generates active taxable income $y_j \geq 0$ at cost $c_j(y_j; \theta_j)$ and brings passive income $\iota_j \geq 0$ to the family. The family allocates the passive pool

$$P = \sum_{j=1}^{J} \iota_j$$

through amounts $r_j \geq 0$ satisfying $\sum_j r_j = P$, so that member j reports

$$x_j = y_j + r_j.$$

The family solves

$$\max_{y,r} \left\{ v\left(\sum_{j=1}^{J} a_j(y_j + r_j); \eta \right) - \sum_{j=1}^{J} c_j(y_j; \theta_j) \right\} \quad \text{subject to} \quad r_j, y_j \geq 0, \qquad \sum_j r_j = P. \tag{7}$$

Let $\iota = (\iota_1, \ldots, \iota_J)$ and write

$$\omega = (\theta, \iota, \eta)$$

for the family type. We allow arbitrary dependence across members and heterogeneous passive income endowments.

Every feasible allocation satisfies

$$\sum_j x_j = \sum_j y_j + P.$$

Reallocating a fixed passive pool can therefore create bunching in individual reports without changing total family income. The converse question is whether the family's choice of active income can place total active and passive income at an aggregate kink. Importantly, a costless passive pool need not be allocated to every member. Members who receive positive passive income share the shadow value of the pool, whereas members receiving no passive income satisfy Kuhn–Tucker inequalities at the zero allocation corner.

A simple quadratic example illustrates the role of this corner. Suppose there are two

members, family utility is linear in total after-tax income, and both members face the same statutory kink k, with

$$\partial a(k) = [q_+, q_-], \qquad q_- > q_+.$$

Using the same quadratic specification as above, let

$$c_j(y; \theta_j) = \frac{(y - \theta_j)^2}{2\tau}, \qquad \tau > 0.$$

Suppose $0 < P < k$ and consider an allocation in which member 1 receives the entire passive pool, member 2 receives none, and both report k:

$$r_1 = P, \qquad r_2 = 0, \qquad x_1 = x_2 = k.$$

Their active incomes are then

$$y_1 = k - P, \qquad y_2 = k.$$

Because member 1 receives passive income in the interior, the shadow value of the pool equals that member's marginal earning cost,

$$\mu = c_{1,y}(k - P; \theta_1) = \frac{k - P - \theta_1}{\tau}.$$

Member 2's marginal earning cost is

$$\ell_2 = c_{2,y}(k; \theta_2) = \frac{k - \theta_2}{\tau}.$$

For both reports to be optimal at the kink, these marginal costs must lie in the kink interval:

$$\mu, \ell_2 \in [q_+, q_-].$$

Because member 2 is at the zero allocation corner, the Kuhn–Tucker condition additionally requires

$$\ell_2 \leq \mu.$$

Strict inequality keeps the same recipient set in place under small changes in the family type. In the quadratic specification,

$$\ell_2 < \mu \quad \Longleftrightarrow \quad \theta_2 > \theta_1 + P.$$

Thus passive income is assigned to the member for whom replacing active income saves the greater marginal earning cost, while the other member remains at the zero allocation corner. Both members can nevertheless report the kink. The general result below allows an arbitrary set of recipients, heterogeneous costs, and different candidate kink locations.

Fix a candidate kink vector

$$k = (k_1, \ldots, k_J), \qquad k_j > 0,$$

and let

$$K = \sum_{j=1}^{J} k_j, \qquad R_k = \sum_{j=1}^{J} a_j(k_j).$$

Suppose $v(\cdot;\eta)$ is differentiable at R_k and write

$$\nu(\eta) = v_R(R_k;\eta) > 0.$$

For each member, define the scaled kink interval

$$\Lambda_j(\eta) = \nu(\eta)\,\partial a_j(k_j).$$

This is the interval of marginal values of reported income that can support member j at k_j.

Let

$$\mathcal{J} = \{1, \ldots, J\}$$

and fix a nonempty set $I \subseteq \mathcal{J}$ of members who receive positive passive income. Because every recipient faces the same pool shadow value, define

$$\Lambda_I(\eta) = \bigcap_{j \in I} \Lambda_j(\eta).$$

For the values of (μ, θ_j) considered below, let $y_j(\mu;\theta_j)$ denote the unique positive solution to

$$c_{j,y}(y_j;\theta_j) = \mu, \qquad j \in I.$$

The amount of passive income required to place all recipients at their candidate kinks is therefore

$$Q_I(\mu;\theta) = \sum_{j \in I} \{k_j - y_j(\mu;\theta_j)\}. \tag{8}$$

For a nonrecipient $j \notin I$, define

$$\ell_j(\theta_j) = c_{j,y}(k_j;\theta_j).$$

At a candidate solution, each recipient has marginal earning cost μ and receives

$$r_j = k_j - y_j(\mu; \theta_j) > 0.$$

A nonrecipient generates and reports k_j, receives no passive income, and has marginal earning cost $\ell_j(\theta_j)$ no greater than the shadow value of the passive pool.

The conditions in the theorem have a direct interpretation. The passive pool must clear at a shadow value lying inside the scaled kink interval of every recipient; each recipient must require a strictly positive passive allocation; and each nonrecipient's marginal earning cost must lie inside its own scaled kink interval and below the pool shadow value, so that zero passive income is optimal.

Theorem 1 (Passive income allocation and composite bunching)**.** *Suppose $v(\cdot;\eta)$ is increasing and concave, is differentiable at R_k, and $\nu(\eta) = v_R(R_k;\eta)$ is positive and continuous. Suppose each a_j is increasing and concave with a strict kink at k_j. Suppose each $c_j(\cdot;\theta_j)$ is continuously differentiable and strictly convex, $c_{j,y}(y;\theta_j)$ is jointly continuous, and the inverse marginal cost $y_j(\mu;\theta_j)$ defined above exists, is positive, and is continuous for the values of (μ,θ_j) under consideration.*

Let

$$\bar{\omega} = (\bar{\theta}, \bar{\iota}, \bar{\eta})$$

be a reference family type, let

$$\bar{P} = \sum_j \bar{\iota}_j,$$

and fix a nonempty recipient set I. Suppose there is a scalar $\bar{\mu}$ such that

$$\bar{\mu} \in \operatorname{int} \Lambda_I(\bar{\eta}), \qquad Q_I(\bar{\mu};\bar{\theta}) = \bar{P}, \tag{9}$$

$$0 < y_j(\bar{\mu};\bar{\theta}_j) < k_j \qquad (j \in I), \tag{10}$$

and, for every $j \notin I$,

$$\ell_j(\bar{\theta}_j) \in \operatorname{int} \Lambda_j(\bar{\eta}), \qquad \ell_j(\bar{\theta}_j) < \bar{\mu}. \tag{11}$$

Then every family with type in some open neighborhood of $\bar{\omega}$ has a unique global solution to (7) *satisfying*

$$x_j = k_j \quad (j \in \mathcal{J}), \qquad r_j > 0 \quad (j \in I), \qquad r_j = 0 \quad (j \notin I).$$

Consequently,

$$\sum_j x_j = \sum_j y_j + P = K.$$

If $\bar{\omega}$ lies in the support of the distribution of family types, meaning that every open neighborhood of $\bar{\omega}$ has positive probability, then total reported family income and total active and passive family income have an atom at K.

Proof. See Appendix A.3. □

The reference equation

$$Q_I(\bar{\mu}; \bar{\theta}) = \bar{P}$$

selects the pool shadow value for the reference family. Nearby families choose a nearby value of μ that clears their own passive pool. The strict inequalities in (9)–(11) keep the same recipient set in place. Only members in I must share the pool shadow value. Members at the zero allocation corner can instead remain optimally at their respective statutory kinks according to their own marginal earning costs. This permits different members to bunch at different kinks even though passive income itself is allocated through a common pool. When $I = \mathcal{J}$, the corner conditions are vacuous and the result reduces to the all-recipient case.

3.4 Finite costly shifting

With finite shifting costs, transfers still net to zero within the family, but members need no longer have the same marginal value at the candidate kink. Let

$$s_j = x_j - y_j$$

be income shifted into member j's report, so that

$$\sum_{j=1}^{J} s_j = 0,$$

and let $\phi_j : \mathbb{R} \to \mathbb{R}_+$ be a strictly convex shifting cost. The family solves

$$\begin{aligned} \max_{x,y,s} \quad & v\left(\sum_{j=1}^{J} a_j(x_j); \eta\right) - \sum_{j=1}^{J} c_j(y_j; \theta_j) - \sum_{j=1}^{J} \phi_j(s_j) \\ \text{subject to} \quad & x_j - y_j - s_j = 0 \quad (j = 1, \ldots, J), \\ & \sum_{j=1}^{J} s_j = 0, \qquad x_j, y_j \geq 0. \end{aligned} \tag{12}$$

The feasible set is convex because all constraints are affine. Since v is increasing and concave, the a_j are concave, and the cost functions are convex, the objective is concave. The supporting conditions used below are therefore sufficient for global optimality.

The same quadratic benchmark used above illustrates what finite shifting costs add. Consider two members with linear family utility, a common statutory kink k satisfying

$$\partial a(k) = [q_+, q_-], \qquad q_- > q_+,$$

and real income adjustment costs

$$c_j(y; \theta_j) = \frac{(y - \theta_j)^2}{2\tau}, \qquad \tau > 0.$$

Let shifting costs be symmetric and quadratic:

$$\phi_j(s) = \frac{s^2}{2\rho}, \qquad \rho > 0.$$

Larger ρ makes shifting less costly. As in the preceding quadratic example, interpret incomes and types relative to a sufficiently large positive baseline so that nonnegativity constraints do not bind.

Suppose both members report k. Write

$$s_1 = s, \qquad s_2 = -s.$$

Their active incomes are then

$$y_1 = k - s, \qquad y_2 = k + s.$$

Let λ_j denote the marginal value supporting member j's report at the kink. The active income conditions give

$$\lambda_1 = \frac{k - s - \theta_1}{\tau}, \qquad \lambda_2 = \frac{k + s - \theta_2}{\tau}.$$

If ζ is the multiplier on the adding-up condition $\sum_j s_j = 0$, the shifting conditions are

$$\frac{s}{\rho} = \lambda_1 + \zeta, \qquad -\frac{s}{\rho} = \lambda_2 + \zeta.$$

Subtracting eliminates ζ:

$$\frac{2s}{\rho} = \lambda_1 - \lambda_2.$$

Define

$$\bar{\theta} = \frac{\theta_1 + \theta_2}{2}, \qquad d = \frac{\theta_2 - \theta_1}{2}.$$

Solving the preceding conditions gives

$$s = \frac{\rho}{\tau + \rho}\, d$$

and

$$\lambda_1 = \frac{k - \bar{\theta}}{\tau} + \frac{d}{\tau + \rho}, \qquad \lambda_2 = \frac{k - \bar{\theta}}{\tau} - \frac{d}{\tau + \rho}.$$

Both members can therefore report the kink if and only if

$$\lambda_1, \lambda_2 \in [q_+, q_-],$$

or equivalently,

$$\frac{k - \bar{\theta}}{\tau} \in \left[q_+ + \frac{|d|}{\tau + \rho},\, q_- - \frac{|d|}{\tau + \rho} \right].$$

Finite shifting costs therefore make bunching depend not only on the average family type but also on the difference between members' types. As $\rho \to \infty$, shifting becomes costless and the difference term disappears, leaving

$$\frac{k - \bar{\theta}}{\tau} \in [q_+, q_-],$$

which is exactly the free shifting condition derived above. As $\rho \downarrow 0$, nonzero shifting becomes arbitrarily costly and the condition converges to

$$\frac{k - \theta_j}{\tau} \in [q_+, q_-] \qquad (j = 1, 2),$$

the condition that both members bunch without shifting. Finite shifting costs therefore interpolate smoothly between the no shifting and free shifting benchmarks. The general construction below replaces this explicit two-member calculation with a system of simultaneous clearing conditions.

Fix a candidate kink vector

$$k = (k_1, \ldots, k_J), \qquad k_j > 0,$$

and let

$$K = \sum_{j=1}^{J} k_j, \qquad R_k = \sum_{j=1}^{J} a_j(k_j).$$

Suppose $v(\cdot;\eta)$ is differentiable at R_k and write

$$\nu(\eta) = v_R(R_k;\eta) > 0.$$

For each member, define the scaled kink interval

$$\Lambda_j(\eta) = \nu(\eta)\,\partial a_j(k_j).$$

This is the interval of marginal values of reported income that can support member j at k_j.

For $\lambda_j \in \Lambda_j(\eta)$, let $y_j(\lambda_j;\theta_j)$ denote the unique positive solution to

$$c_{j,y}(y_j;\theta_j) = \lambda_j.$$

Suppose ϕ_j is continuously differentiable, ϕ_j' is strictly increasing and onto $\mathbb{R}$. For a vector

$$\lambda = (\lambda_1, \ldots, \lambda_J),$$

define $\zeta(\lambda)$ as the unique solution to

$$\sum_{j=1}^{J} (\phi_j')^{-1}(\lambda_j + \zeta) = 0.$$

The shift implied by the marginal conditions is

$$S_j(\lambda) = (\phi_j')^{-1}\big(\lambda_j + \zeta(\lambda)\big),$$

whereas the shift required to place member j's report at k_j is

$$F_j(\lambda_j;\theta_j) = k_j - y_j(\lambda_j;\theta_j).$$

Define

$$G_j(\lambda;\theta) = S_j(\lambda) - F_j(\lambda_j;\theta_j).$$

Each object has a direct clearing interpretation. At a candidate vector λ, $S_j(\lambda)$ is the shift implied by the first-order conditions for shifting costs after the adding-up multiplier $\zeta(\lambda)$ has been chosen. By contrast, $F_j(\lambda_j;\theta_j)$ is the shift required to reconcile member j's

active income with report k_j. Thus

$$G_j(\lambda;\theta) = 0$$

states that the implied and required shifts coincide for member j. In the quadratic example these equations can be solved explicitly. In the general model they must be cleared simultaneously across all family members.

Theorem 2 (Costly shifting and simultaneous bunching)**.** *Suppose $v(\cdot;\eta)$ is increasing and concave, is differentiable at R_k, and $\nu(\eta) = v_R(R_k;\eta)$ is positive and continuous. Suppose each a_j is concave with a strict kink at k_j. Suppose each $c_j(\cdot;\theta_j)$ is continuously differentiable and strictly convex, $c_{j,y}(y;\theta_j)$ is jointly continuous, and the inverse marginal cost $y_j(\lambda_j;\theta_j)$ defined above exists, is positive, and is jointly continuous for the values of (λ_j,θ_j) under consideration. Suppose each ϕ_j is continuously differentiable and strictly convex and ϕ_j' is onto $\mathbb{R}$.*

Let

$$\bar{\omega} = (\bar{\theta},\bar{\eta})$$

be a reference family type. Suppose there are numbers

$$\lambda_j^- < \lambda_j^+ \qquad (j = 1,\dots,J)$$

and define

$$\mathcal{B} = \prod_{j=1}^{J}[\lambda_j^-,\lambda_j^+], \qquad \mathcal{B}_{-j} = \prod_{\ell\neq j}[\lambda_\ell^-,\lambda_\ell^+].$$

Suppose

$$[\lambda_j^-,\lambda_j^+] \subset \operatorname{int}\Lambda_j(\bar{\eta}) \qquad (j = 1,\dots,J),$$

and that there is an $\varepsilon > 0$ such that, for every j,

$$G_j(\lambda_j^-,\lambda_{-j};\bar{\theta}) \le -\varepsilon \qquad \text{for every } \lambda_{-j}\in\mathcal{B}_{-j}, \tag{13}$$

$$G_j(\lambda_j^+,\lambda_{-j};\bar{\theta}) \ge \varepsilon \qquad \text{for every } \lambda_{-j}\in\mathcal{B}_{-j}. \tag{14}$$

Then every family with type in some open neighborhood of $\bar{\omega}$ has a unique global solution to (12) *satisfying*

$$x_j = k_j \qquad (j = 1,\dots,J).$$

Consequently,

$$\sum_{j=1}^{J} x_j = \sum_{j=1}^{J} y_j = K.$$

If $\bar{\omega}$ lies in the support of the distribution of family types, meaning that every open neighborhood of $\bar{\omega}$ has positive probability, then total reported family income and total real family income have an atom at K.

Proof. See Appendix A.5. □

The face conditions in (13)–(14) play the same role as the scalar crossing condition in the passive income model. There, one shadow value clears a passive pool. Here, the family must simultaneously reconcile the shifts implied by the shifting costs with the shifts required to place each report at its kink. Poincaré–Miranda supplies a vector at which all of these conditions hold, while the strict face inequalities make the construction persist for an open set of family types. The quadratic example displays the same mechanism in closed form: finite shifting costs allow the marginal values associated with different members' reports to differ, with those differences matched by the marginal costs of the transfers between them. Appendix A shows how a primitive effort-production model can be transformed into the convex real income costs used here without changing the family's feasible choices.

3.5 Nesting across reporting technologies

The preceding results show how coordination can create additional family bunchers. With a common tax schedule, coordination satisfies a stronger restriction: it preserves the families that already bunch jointly without shifting. Consider a family whose no shifting optimum places every member at the same statutory kink k. When free shifting becomes available, the family continues to choose aggregate real and reported income $K = Jk$, and assigning report k to every member remains optimal. With finite convex shifting costs, the same conclusion holds provided the implied real incomes remain positive. Coordination can therefore enlarge the aggregate bunching set without eliminating this baseline set.

Fix a strict kink k of a common increasing, concave after-tax schedule a, and let

$$K = Jk, \qquad R_k = Ja(k), \qquad \nu(\eta) = v_R(R_k; \eta) > 0.$$

Define the scaled kink interval

$$\mathcal{L}_k(\eta) = \nu(\eta)\, \partial a(k) = [\underline{\lambda}_k(\eta), \overline{\lambda}_k(\eta)]. \tag{15}$$

Because every member faces the same schedule, $\mathcal{L}_k(\eta)$ is the common interval of marginal earning costs consistent with optimality at report k.

The quadratic benchmark makes the nesting immediate. With two members and linear family utility, define the marginal earning costs at the no shifting allocation by

$$\ell_1 = \frac{k - \theta_1}{\tau}, \qquad \ell_2 = \frac{k - \theta_2}{\tau}, \qquad \bar{\ell} = \frac{\ell_1 + \ell_2}{2}.$$

Suppose, as in the preceding subsection, that

$$\phi_j(s) = \frac{s^2}{2\rho}, \qquad \rho > 0.$$

The calculation there gives the marginal values associated with the two reports as

$$\lambda_1^\rho = \bar{\ell} + \frac{\tau}{\tau + \rho}(\ell_1 - \bar{\ell}), \qquad \lambda_2^\rho = \bar{\ell} + \frac{\tau}{\tau + \rho}(\ell_2 - \bar{\ell}).$$

Each λ_j^ρ therefore lies between ℓ_j and the family average $\bar{\ell}$. If both ℓ_1 and ℓ_2 already lie in $[q_+, q_-]$, then so do λ_1^ρ and λ_2^ρ. As $\rho \downarrow 0$, shifting becomes prohibitively costly and $\lambda_j^\rho \to \ell_j$. As $\rho \to \infty$, shifting becomes costless and both marginal values converge to $\bar{\ell}$, which also lies between ℓ_1 and ℓ_2. Thus a family in which both members bunch without shifting continues to bunch when finite or free shifting is introduced. The proposition below establishes the same nesting result without quadratic costs.

Proposition 2 (Common schedule nesting)**.** *Suppose $v(\cdot;\eta)$ is increasing and concave, is differentiable at R_k, and $\nu(\eta) = v_R(R_k;\eta)$ is positive and continuous in η. Suppose each $c_j(\cdot;\theta_j)$ is continuously differentiable and strictly convex, $c_{j,y}(y;\theta_j)$ is jointly continuous, and, for every $\lambda \in \mathcal{L}_k(\eta)$ under consideration, the equation*

$$c_{j,y}(y;\theta_j) = \lambda$$

has a unique positive solution $y_j(\lambda;\theta_j)$ that is continuous in (λ, θ_j). Define the baseline bunching set

$$\mathcal{B}_0(k) = \{(\theta, \eta) : c_{j,y}(k;\theta_j) \in \mathcal{L}_k(\eta) \text{ for every } j\}. \tag{16}$$

Every family type in $\mathcal{B}_0(k)$ has the unique no shifting optimum

$$x_j = y_j = k \qquad (j = 1, \ldots, J).$$

Under free shifting, every such family uniquely chooses total real and reported income K, and

$$x_j = k \qquad (j = 1, \ldots, J)$$

is an optimal reporting allocation conditional on that total.

For the finite shifting statement, also impose the shifting-cost assumptions of Theorem 2 and normalize each shifting cost to be minimized at zero, so that

$$\phi_j'(0) = 0.$$

For every $(\theta, \eta) \in \mathcal{B}_0(k)$, suppose the unique solution to

$$\min_s \left\{ \sum_{j=1}^{J} [c_j(k - s_j; \theta_j) + \phi_j(s_j)] : \sum_{j=1}^{J} s_j = 0, \quad k - s_j \geq 0 \text{ for every } j \right\}$$

is interior, in the sense that

$$k - s_j^* > 0 \qquad (j = 1, \ldots, J).$$

Then every family type in $\mathcal{B}_0(k)$ also has the unique finite shifting optimum

$$x_j = k \qquad (j = 1, \ldots, J).$$

Let $m_F^{\text{free}}(k)$ and $m_F^{\text{cost}}(k)$ denote the point masses at K under free and finite shifting, respectively. Then, for any joint distribution of family types,

$$m_F^{\text{free}}(k) \geq \mathbb{P}\{\mathcal{B}_0(k)\}, \qquad m_F^{\text{cost}}(k) \geq \mathbb{P}\{\mathcal{B}_0(k)\}. \tag{17}$$

Independence is not required. If $\theta_1, \ldots, \theta_J$ are independent conditional on η and

$$p_j(k, \eta) = \mathbb{P}(c_{j,y}(k; \theta_j) \in \mathcal{L}_k(\eta) \mid \eta),$$

then

$$\mathbb{P}\{\mathcal{B}_0(k)\} = \mathbb{E}\left[\prod_{j=1}^{J} p_j(k, \eta)\right].$$

If η is fixed, this reduces to

$$\mathbb{P}\{\mathcal{B}_0(k)\} = \prod_{j=1}^{J} p_j(k, \eta).$$

Proof. See Appendix A.6. □

The proof illustrates directly why the baseline bunching set survives. Under free shifting, coordination replaces the member-specific marginal earning costs by a common marginal cost lying between their minimum and maximum. Under finite shifting, marginal earning costs may remain different across members, but none can leave that same interval. Since the entire interval lies in $\mathcal{L}_k(\eta)$, the common kink remains optimal.

There is an analogous nesting argument in the passive income model when every member receives positive passive income. Suppose

$$0 < \iota_j < k \qquad (j = 1, \ldots, J),$$

and suppose each member would report k if their own passive income remained attached. Define

$$\ell_j^P = c_{j,y}(k - \iota_j; \theta_j) \in \mathcal{L}_k(\eta), \qquad P = \sum_j \iota_j,$$

and let

$$\ell_-^P = \min_j \ell_j^P, \qquad \ell_+^P = \max_j \ell_j^P.$$

Since

$$y_j(\ell_j^P; \theta_j) = k - \iota_j,$$

strict convexity implies

$$y_j(\ell_-^P; \theta_j) \leq k - \iota_j \leq y_j(\ell_+^P; \theta_j) \qquad \text{for every } j.$$

Hence the passive income required to place every report at k satisfies

$$Q_{\mathcal{J}}(\ell_-^P; \theta) \geq P \geq Q_{\mathcal{J}}(\ell_+^P; \theta).$$

Since $Q_{\mathcal{J}}(\cdot; \theta)$ is continuous and decreasing, there is some

$$\lambda^* \in [\ell_-^P, \ell_+^P]$$

such that

$$Q_{\mathcal{J}}(\lambda^*; \theta) = P.$$

Because

$$[\ell_-^P, \ell_+^P] \subseteq \mathcal{L}_k(\eta),$$

this common shadow value supports every report at the kink. If, in addition,

$$y_j(\ell_+^P; \theta_j) < k \qquad \text{for every } j,$$

then

$$y_j(\lambda^*; \theta_j) < k \qquad \text{for every } j,$$

so every implied passive allocation

$$r_j^* = k - y_j(\lambda^*; \theta_j)$$

is strictly positive. If each ℓ_j^P lies in

$$\operatorname{int} \mathcal{L}_k(\eta),$$

then λ^* also lies in $\operatorname{int} \mathcal{L}_k(\eta)$. The conditions of Theorem 1 are therefore satisfied at this family type, and the all-recipient solution persists for nearby family types.

Finally, the lower bounds in Proposition 2 are positive whenever $\mathcal{B}_0(k)$ contains a supported interior type. In particular, if

$$\bar{\omega} = (\bar{\theta}, \bar{\eta})$$

lies in the support of the family type distribution and

$$c_{j,y}(k; \bar{\theta}_j) \in \operatorname{int} \mathcal{L}_k(\bar{\eta}) \qquad \text{for every } j,$$

then continuity gives an open neighborhood of $\bar{\omega}$ contained in $\mathcal{B}_0(k)$. That neighborhood has positive probability, so

$$\mathbb{P}\{\mathcal{B}_0(k)\} > 0.$$

The aggregate mass at K is therefore positive under free shifting and, under the finite shifting conditions of Proposition 2, under finite shifting as well. Independence is not required for this conclusion; it is used only for the product representation of the baseline mass.

3.6 Attenuation and coordination under aggregation

The preceding results show how coordination can preserve or create bunching in aggregate family income. A natural concern is that aggregation itself may nevertheless weaken a real response. That concern is valid in the absence of coordination, but it provides a conservative benchmark rather than a general implication of aggregation. If family members respond separately, the family reaches the sum of their kink incomes through the all kink configuration only when all relevant members bunch at their respective kinks. Under independence, the

probability of that event is the product of the individual bunching probabilities. By contrast, the nesting results above show that this baseline set survives when coordination becomes available, while the composite kink results allow additional families to bunch.

Let Y_j^0 denote member j's optimal income in the absence of shifting, let k_j be the relevant kink, and define

$$K = \sum_{j=1}^{J} k_j.$$

Suppose Y_j^0 has an atom of size

$$b_j = \mathbb{P}(Y_j^0 = k_j)$$

at k_j.

Proposition 3 (No shifting attenuation)**.** *Suppose*

$$\mathbb{P}\left(\sum_{j=1}^{J} Y_j^0 = K,\, (Y_1^0, \ldots, Y_J^0) \neq (k_1, \ldots, k_J)\right) = 0.$$

Then

$$\mathbb{P}\left(\sum_{j=1}^{J} Y_j^0 = K\right) = \mathbb{P}\big(Y_1^0 = k_1, \ldots, Y_J^0 = k_J\big)\,.$$

If the member incomes are independent, this mass is

$$\prod_{j=1}^{J} b_j.$$

Proof. See Appendix A.8. □

The assumption rules out a separate atom at K generated by non-kink income configurations that happen to add to the same total. Such configurations can occur, but they do not themselves create an atom when, conditional on the other members' incomes, at least one remaining income is non-atomic at the residual required to reach K. The proposition therefore isolates the attenuation that arises when all relevant members must bunch jointly. If the empirical object is family average income rather than family total income, the same probability mass appears at K/J.

The product can be much smaller than the individual bunching probabilities, making the no shifting calculation a useful conservative benchmark.

The product calculation should not, however, be interpreted as the inevitable effect of aggregation. Proposition 2 shows that, at a common statutory kink, the families whose

members all bunch without shifting remain aggregate bunchers when free shifting becomes available and, under its stated conditions, when shifting carries finite convex costs. Their joint probability therefore remains a lower bound on the aggregate bunching mass. Coordination can then add families outside this baseline set, as the composite kink, passive income, and costly shifting results above demonstrate.

The quadratic specification used earlier makes the quantitative difference particularly transparent. We now return to that benchmark with J members. As before, incomes and types can be interpreted as deviations from a sufficiently large positive baseline so that the nonnegativity constraints do not bind.

Proposition 4 (Quadratic attenuation and coordination benchmark)**.** *Suppose J members face the same statutory kink k, with marginal net-of-tax rates*

$$q_- > q_+,$$

family utility is linear in total after-tax income, and

$$c_j(y;\theta_j) = \frac{(y-\theta_j)^2}{2\tau}, \qquad \tau > 0.$$

Suppose

$$\theta_j \overset{\text{iid}}{\sim} N(\mu,\sigma^2), \qquad \mu = k - \frac{\tau(q_- + q_+)}{2},$$

and define

$$z = \frac{\tau(q_- - q_+)}{2\sigma}.$$

Then the individual atom at k is

$$p = 2\Phi(z) - 1.$$

With no shifting, family average income has an atom of size

$$p^J$$

at k. With free shifting, family average income has an atom of size

$$p_{F,J} = 2\Phi(\sqrt{J}\, z) - 1$$

at k. For every $J > 1$,

$$p_{F,J} > p > p^J.$$

Moreover, as the individual atom becomes small,

$$\frac{p_{F,J}}{p} \longrightarrow \sqrt{J}.$$

Proof. See Appendix A.9. □

The contrast is substantial even for a two-member family. Take

$$J = 2, \qquad q_- = 0.80, \qquad q_+ = 0.60, \qquad \tau = 0.25, \qquad \sigma = 0.40.$$

The individual atom is

$$p = 4.98\%.$$

Without shifting, independence reduces the family atom to

$$p^2 = 0.25\%.$$

With free shifting, however, the family atom is

$$p_{F,2} = 7.04\%.$$

The same underlying primitives can therefore imply severe attenuation when family members act separately and an aggregate atom larger than the individual atom when reports can be coordinated. The difference comes from the bunching set itself. Without shifting, both types must lie in the individual bunching interval. With free shifting, only their average must lie there.

The empirical calibration in Section 6.2 maps this logic into the Australian tax schedule. We first calibrate real responsiveness in a two-member family with no within-family allocation. Holding this responsiveness and the distribution of baseline resources fixed, we then vary the share of family resources that can be allocated across members. At zero allocability, a family reaches the aggregate kink only when the relevant members bunch jointly, so aggregation can substantially attenuate a real response.

Allowing income to be reallocated changes this implication. Under the common schedule nesting result, free shifting or shifting with finite costs preserves the baseline families that bunch at the aggregate kink, while coordination can draw additional families there. The passive income model yields the same possibility even when active earnings remain attached to the members who generate them. Aggregation can therefore attenuate a real response when income cannot be shifted, but even modest scope for within-family allocation can

restore or strengthen bunching in the family distribution. The next section develops a global test designed for these aggregate implications. Because family composition and the kink configuration can generate several possible bunching locations, the test searches a prespecified income interval rather than a single cutoff.

4 A global bootstrap test for distributional smoothness

4.1 Statistic and population geometry

We develop a novel bootstrap Kolmogorov–Smirnov test that evaluates the maximum discrepancy between the empirical and smoothed CDFs over the relevant part of the taxable income distributions. The global construction accommodates multiple composite thresholds as well as aggregate departures from smoothness whose precise locations are not known in advance, while its power analysis places the economic predictions from Section 3 on an explicit detection scale. Let $X_1, \ldots, X_n$ be i.i.d. with common law P and CDF F, and let $I \subset \mathbb{R}$ denote the prespecified interval. The empirical CDF is F_n, and $\mathbb{P}_n$ denotes the empirical measure. For the Gaussian kernel, define

$$\widehat{F}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x - X_i}{h}\right), \qquad D_n(x) = F_n(x) - \widehat{F}_h(x).$$

The interval Kolmogorov–Smirnov discrepancy is $\sup_{x\in I} |D_n(x)|$. Throughout, I lies inside an open neighborhood on which the null density is positive and smooth.

Proposition 5 (Atom separation)**.** *If F has an atom of mass m at $x_0 \in \mathrm{int}(I)$ and G is continuous at x_0, then*

$$\sup_{x\in I} |F(x) - G(x)| \geq \frac{m}{2}.$$

Proof. Choose $x_r \in I$ with $x_r \uparrow x_0$. By the definition of the left limit of a CDF and continuity of G,

$$F(x_r) - G(x_r) \longrightarrow F(x_0-) - G(x_0) =: a.$$

At the atom, $F(x_0) - G(x_0) = a + m$. Hence the supremum over I is at least both $|a|$ and $|a+m|$, in the first case by taking the preceding limit. Since $m = |(a+m)-a| \leq |a+m|+|a|$, at least one is no smaller than $m/2$. □

An atom therefore creates a fixed CDF separation from every continuous benchmark, whereas a density jump leaves the CDF continuous and produces a subtler local signal.

Table 1: Signal and stochastic scale for the interval statistic. Here $\ell_n = \log n \vee \log(1/h_n)$.

Alternative	Population CDF signal	Stochastic scale	Sufficient detection condition
Fixed atom $m > 0$	m	$\sqrt{h_n \ell_n / n}$	automatic
Shrinking atom m_n	m_n	$\sqrt{h_n \ell_n / n}$	$m_n \gg \sqrt{h_n \ell_n / n}$
Density jump Δ_n	$h_n \lvert\Delta_n\rvert$	$\sqrt{h_n \ell_n / n}$	$\lvert\Delta_n\rvert \gg \sqrt{\ell_n / (n h_n)}$
Smooth-null bias	h_n^2	$\sqrt{h_n / n}$	$n h_n^3 \log^2(1/h_n) \to 0$

Proposition 6 (Density jump smoothing drift)**.** *Suppose f has one-sided expansions at $x_0 \in \mathrm{int}(I)$ with jump $\Delta = f(x_0+) - f(x_0-)$. If $F_h(x) = \mathbb{E}\Phi((x - X)/h)$, then*

$$|F(x_0) - F_h(x_0)| = \mu_\phi h |\Delta| + O(h^2), \qquad \mu_\phi = \int_0^\infty u\phi(u)\, du.$$

Proof sketch. Use $F_h(x_0) = \mathbb{E}F(x_0 - hZ)$. For $Z > 0$, expand F from the left of x_0; for $Z < 0$, expand it from the right. Symmetry gives first moments μ_ϕ and $-\mu_\phi$ on the two half-lines, leaving $h\mu_\phi\{f(x_0+) - f(x_0-)\} + O(h^2)$. Appendix C supplies the uniform remainder argument. □

Under a smooth null,

$$F_h(x) - F(x) = \frac{h^2}{2} f'(x) + o(h^2) \tag{18}$$

uniformly on I. Meanwhile the variance of the local contrast $\mathbf{1}\{X \le x\} - \Phi((x - X)/h)$ is of order h. The stochastic scale of D_n is therefore $\sqrt{h/n}$, not $n^{-1/2}$, and the normalized supremum grows at a logarithmic local-maximum rate.

4.2 Issues with standard calibration

With

$$\psi_{x,h}(u) = \mathbf{1}\{u \le x\} - \Phi\left(\frac{x - u}{h}\right),$$

the full-sample discrepancy decomposes as

$$D_n(x) = (\mathbb{P}_n - P)\psi_{x,h} + \{F(x) - F_h(x)\}. \tag{19}$$

The first term is local: the discrepancy between the indicator and its smoothed counterpart is concentrated on scale h around x, so its pointwise variance is of order h/n. The second term is smoothing bias of order h^2. After normalization by $\sqrt{n/h}$, a conventional $h \asymp n^{-1/5}$ bandwidth leaves bias of order $n^{1/5}$, which overwhelms the local stochastic process. The observed statistic must therefore be undersmoothed.

Undersmoothing does not, however, mean choosing the smallest possible bandwidth. For a shrinking atom, the sufficient detection scale is

$$\sqrt{\frac{h_n \ell_n}{n}},$$

so a smaller bandwidth improves sensitivity to an atom. If the sole objective were to detect excess mass at a known point, one could therefore use a more targeted procedure. That is not the question in our application. The relevant aggregate location can vary with family composition and the kink configuration, so the test searches a prespecified interval rather than a single known point. More generally, alternative shifting technologies or adjustment frictions could generate localized departures from smoothness other than literal atoms, including density jumps. The procedure is therefore intended as a broader test of local smoothness rather than a pointwise test optimized for one particular atom alternative.

More importantly, recording precision places a natural lower scale on what should be interpreted economically. Individual taxable income is recorded in whole dollars, so at one-dollar resolution the observed distribution is mechanically discrete. A procedure designed simply to detect positive probability at individual integer values would therefore partly detect the recording convention rather than economically meaningful bunching. Classical results on kernel estimation with rounded or binned data show that the effect of such discretization becomes small when the recording increment is fine relative to the smoothing bandwidth; see Hall (1982) and Scott and Sheather (1985). This provides a separate reason not to push the bandwidth toward the finest observable scale.

The bandwidths used in the application are comfortably larger than the one-dollar recording increment. Every bandwidth in the $\beta = 0.35$ rounding exercise exceeds \$119, so one dollar is less than one percent of the smoothing scale. In 2021, under the preferred $\beta = 0.40$ specification, the bandwidth is \$130.53 for family average income and \$114.36 for individual income. Rounding family averages to whole dollars leaves the results essentially unchanged, adding only a marginal family rejection in 2011 to the 2007 rejection under the sensitivity specification. Appendix Table 12 reports the results.

Two other natural calibrations fail for different reasons. Splitting the sample avoids same-observation dependence but leaves first-order variation from the independently estimated benchmark; even after correct mixed-bridge calibration, it detects atoms and density jumps only at the coarser global scales $n_B^{-1/2}$ and $(h_A\sqrt{n_B})^{-1}$. Smooth resampling from the fitted KDE creates a separate re-smoothing drift unless the bootstrap process is explicitly centered. The feasible procedure addresses the three problems directly: undersmoothing removes observed bias, full-sample construction preserves cancellation of common empirical-CDF

variation, and centering removes bootstrap drift. Appendix B gives the formal holdout results, the pilot bandwidth justification, and the minimax benchmark for density jump detection.

4.3 The centered frozen-bandwidth procedure

Procedure 1 (Centered frozen-bandwidth Kolmogorov–Smirnov test)**.**

1. Prespecify an interior compact interval I.
2. Choose $h_n = cn^{-\beta}$ with $\beta \in (1/3, 1)$. Alternatively, start from a pilot bandwidth of order $n^{-1/5}$ and multiply it by $n^{-\delta}$ with $\delta \in (2/15, 4/5)$. Freeze the resulting bandwidth across resamples.
3. Compute
$$U_n = \sqrt{\frac{n}{h_n}} \sup_{x \in I} \left| F_n(x) - \widehat{F}_{h_n}(x) \right|.$$
4. Draw $X_1^*, \ldots, X_n^*$ from the fitted KDE law $\widehat{F}_{h_n}$ and recompute both F_n^* and $\widehat{F}_{h_n}^*$ using the same frozen bandwidth.
5. Center the bootstrap process by its conditional mean,
$$\mathbb{E}^* D_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \left[\Phi\left(\frac{x - X_i}{h_n} \right) - \Phi\left(\frac{x - X_i}{\sqrt{2} h_n} \right) \right].$$
6. Form
$$S_n^{*,c} = \sqrt{\frac{n}{h_n}} \sup_{x \in I} |D_n^*(x) - \mathbb{E}^* D_n^*(x)|$$
and reject at level α when U_n exceeds the conditional $(1 - \alpha)$ quantile of $S_n^{*,c}$.

The construction combines three features. Using the full sample preserves the cancellation of common empirical CDF variation, undersmoothing makes the smooth null bias in the observed statistic asymptotically negligible, and explicit centering removes the additional smoothing drift generated by drawing from the fitted KDE. The selected bandwidth is held fixed across resamples.

Fast Gaussian multiplier implementation. The centered local empirical process admits a direct Gaussian multiplier approximation. Draw independent $\xi_i \sim N(0, 1)$, let

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i,$$

and compute

$$S_n^{\xi} = \sup_{x \in I} \left| \frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} (\xi_i - \bar{\xi}) \psi_{x,h_n}(X_i) \right|. \tag{20}$$

The conditional quantiles of S_n^{ξ} may be used in place of those of $S_n^{*,c}$. This implementation is centered by construction and avoids drawing and smoothing bootstrap samples. It is therefore particularly convenient for large administrative data sets and is the implementation used in the empirical analysis.

Assumption 1 (Smooth null class)**.** *The interval I is compact and there is an open set $V \supset I$ on which*

$$0 < c_f \leq f \leq C_f < \infty.$$

The density is globally bounded and has two bounded uniformly continuous derivatives. The kernel is Gaussian, and the bandwidth is

$$h_n = cn^{-\beta}$$

for fixed $c > 0$ and $\beta \in (1/3, 1)$.

Assumption 1 provides sufficient conditions for the approximation used below. Twice differentiability controls the CDF smoothing bias, positivity prevents degeneration of the local variance, and $\beta \in (1/3, 1)$ makes the bias negligible while leaving a growing number of observations in each bandwidth neighborhood. Appendix C establishes the Gaussian approximation, covariance consistency, and anti-concentration results.

Theorem 3 (Size and power of the centered test)**.** *Let*

$$\ell_n = \log n \vee \log(1/h_n).$$

The centered frozen-bandwidth test has the following properties.

(i) Under Assumption 1, let $p_n^{*,c}$ denote the conditional p-value obtained from Procedure 1. Then

$$p_n^{*,c} \rightsquigarrow \text{Uniform}(0, 1).$$

Hence the test has asymptotic size α.

(ii) Let

$$P_n = (1 - m_n)P_0 + m_n \delta_{x_0},$$

where P_0 has a bounded density and $x_0 \in \mathrm{int}(I)$. Power tends to one if

$$m_n \sqrt{\frac{n}{h_n \ell_n}} \to \infty.$$

In particular, the test is consistent against every fixed atom in the interior of I.

(iii) Let P_n have uniformly bounded densities with uniformly Lipschitz one-sided expansions at $x_0 \in \mathrm{int}(I)$ and density jump

$$\Delta_n = f_n(x_0+) - f_n(x_0-).$$

Power tends to one if

$$|\Delta_n| \sqrt{\frac{n h_n}{\ell_n}} \to \infty.$$

In particular, the test is consistent against every fixed nonzero interior density jump.

Corollary 1 (Fast multiplier calibration)**.** *Under Assumption 1, let $c^{\xi}_{n,1-\alpha}$ be the conditional $(1-\alpha)$ quantile of (20). Then*

$$\mathbb{P}\Big\{ U_n > c^{\xi}_{n,1-\alpha} \Big\} \to \alpha.$$

The atom and density jump power conclusions in Theorem 3 also hold with multiplier critical values. Thus the centered smooth bootstrap and the Gaussian multiplier procedure provide two asymptotically valid calibrations of the same local empirical process.

The application uses a bandwidth obtained by undersmoothing a pilot rather than a fixed deterministic sequence. The same size and power conclusions continue to hold under the selector conditions stated in Appendix B.

Proof sketch. Undersmoothing makes the normalized h_n^2 bias negligible. A Gaussian approximation and anti-concentration argument then transfer the centered bootstrap or multiplier critical values to the observed local process. An atom creates a fixed CDF separation, while a density jump generates a smoothing discrepancy of order $h_n|\Delta_n|$. Appendix C gives the full argument.

For the family application, part (ii) gives the relevant asymptotic detection scale directly. With n_F families and bandwidth h_{n_F}, an atom of mass m_F is detected with probability tending to one whenever

$$m_F \gg \sqrt{\frac{h_{n_F} \ell_{n_F}}{n_F}}, \qquad \ell_{n_F} = \log n_F \vee \log(1/h_{n_F}).$$

The economic results in Section 3 imply family atoms or lower bounds on their mass under the different reporting technologies. The empirical analysis therefore compares the quantitative benchmarks implied by those results with the finite-sample detection scale of the family test.

5 Finite-sample performance

The simulations evaluate the finite-sample performance of the statistical procedure. They serve two purposes. First, they show why several natural calibration strategies fail and verify size control for the centered undersmoothed procedure. Second, they assess power and illustrate the loss from sample splitting even when holdout is calibrated correctly. Unless otherwise stated, rejection rates refer to nominal five-percent tests. The target interval is fixed before simulation, the supremum over that interval is evaluated on a 2,001-point grid, and the main centered experiments use 500 Monte Carlo replications with 399 bootstrap draws. Detailed local power results, holdout comparisons, and additional null designs appear in Appendix D and Appendix E.

5.1 Calibration and power

Table 2 isolates the calibration problems that motivate the procedure by comparing several implementations under a smooth normal null. Naive holdout overrejects because its critical values treat the estimated training benchmark as fixed and therefore omit its sampling variation. A centered smooth bootstrap used with an unmodified $n^{-1/5}$ bandwidth also overrejects because centering the bootstrap does not remove the smoothing bias from the observed statistic. Combining centering with sufficient undersmoothing restores size control.

Table 2: Smooth-null calibration on a fixed interval. The DGP is $N(0,1)$, $n = 10{,}000$, and the statistic is evaluated on $I = [-1,1]$. Each row uses 300 Monte Carlo replications; bootstrap rows use 199 bootstrap draws. Ordinary holdout ignores the training-sample bridge, while a conventional bandwidth leaves first-order bias in the centered statistic. The mixed-bridge holdout and centered undersmoothed calibrations address the corresponding failures.

Procedure	Bandwidth	Calibration	I	Rejection rate
Holdout Kolmogorov–Smirnov	Silverman	50/50 split, ordinary one-sample critical value	$[-1,1]$	0.220
Calibrated holdout Kolmogorov–Smirnov	Silverman $\to n^{-0.35}$	50/50 split, mixed-bridge simulation, 999 draws	$[-1,1]$	0.033
Centered frozen Kolmogorov–Smirnov	Raw Silverman	Centered bootstrap, 199 draws	$[-1,1]$	0.667
Centered frozen Kolmogorov–Smirnov	Silverman $\to n^{-0.35}$	Centered bootstrap, 199 draws	$[-1,1]$	0.063

The calibrated holdout row simulates 999 Brownian-bridge paths after plugging in the pooled empirical CDF on I.

Holdout can be calibrated correctly by accounting for the sampling variation in the

estimated training benchmark. This removes the size distortion of the naive procedure, but it does not undo the statistical cost of splitting the sample. In the full-sample statistic, the empirical CDF and the smoothed CDF are constructed from the same observations, so their common empirical variation largely cancels. Their difference is a local empirical process concentrated at scale h_n, with pointwise stochastic variation of order $\sqrt{h_n/n}$. With holdout, the empirical CDF and the estimated benchmark are based on independent samples, so this cancellation is lost. The usual sampling variation of the two distribution estimates remains, of order

$$\sqrt{\frac{1}{n_B} + \frac{1}{n_A}},$$

which is of order $n_B^{-1/2}$ when the split ratio is fixed. As Proposition 7 shows, holdout therefore has the coarser atom detection scale $n_B^{-1/2}$, compared with

$$\sqrt{\frac{h_n \ell_n}{n}}$$

for the full-sample statistic.

The simulations illustrate how substantial this difference can be. The local alternatives are indexed by a multiplier c of the theoretical detection scale. At $n = 10{,}000$ and $c = 1$, the centered full-sample test rejects the atom alternative in 76.2 percent of replications, compared with 3.0 percent for correctly calibrated holdout. At $c = 1.5$, the corresponding rejection rates are 99.6 percent and 3.2 percent. The appendix shows a similarly large power loss for the density-jump alternatives. Thus mixed-bridge calibration corrects the size distortion of naive holdout, but it does not recover the power lost by splitting the sample.

Figure 2 illustrates the alternatives considered in the power analysis. An atom creates a jump in the CDF relative to the continuous KDE benchmark. A density jump leaves the CDF continuous but generates a localized smoothing discrepancy. Table 3 shows that the centered procedure remains close to nominal size while retaining substantial power against both alternatives. The more extensive results in the appendix trace the power transition as the signal moves through the theoretical detection scale.

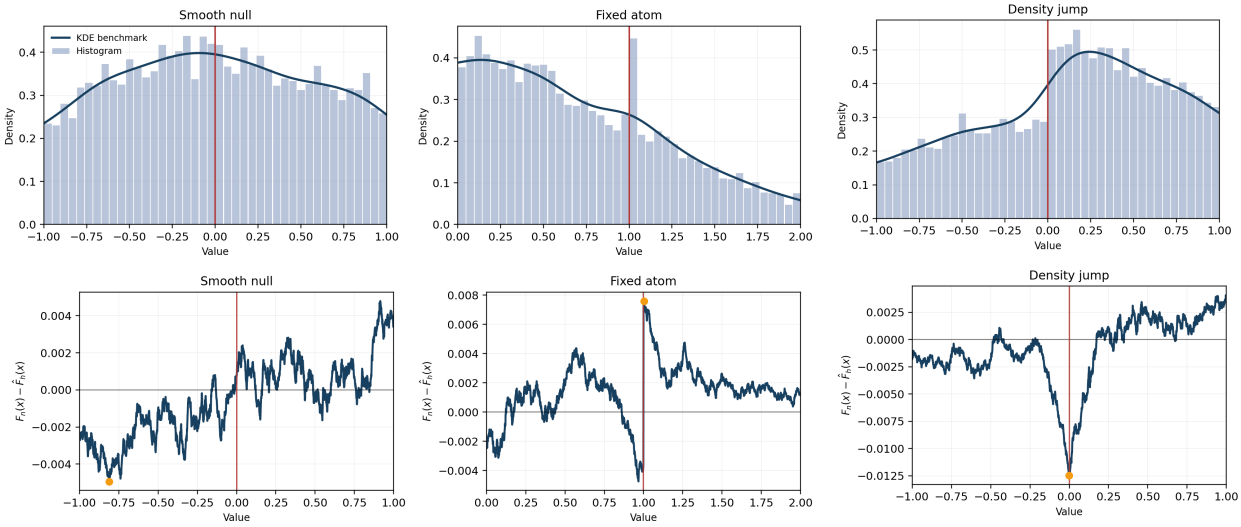


Figure 2: Smooth distributions, atoms, and density jumps. The top row shows histograms and Gaussian KDE benchmarks, and the bottom row shows the corresponding signed empirical-minus-KDE CDF discrepancies. The vertical marker is fixed by the data-generating process. The statistic maximizes the absolute CDF discrepancy over the prespecified target interval.

Table 3: Interval-restricted centered frozen-bandwidth Kolmogorov–Smirnov simulations. Each cell uses 500 Monte Carlo replications and 399 bootstrap draws. The statistic is evaluated only on the prespecified interval shown in the table, matching the form of the formal procedure.

Design	I	n	Reps	Reject	Mean p
Smooth normal null	$[0, 2]$	2000	500	0.014	0.575
Smooth normal null	$[0, 2]$	10000	500	0.038	0.523
One-percent atom	$[0, 2]$	2000	500	0.308	0.234
One-percent atom	$[0, 2]$	10000	500	1.000	0.003
Smooth normal null	$[-1, 1]$	2000	500	0.048	0.477
Smooth normal null	$[-1, 1]$	10000	500	0.052	0.455
Density jump, $c = 1$	$[-1, 1]$	2000	500	0.596	0.091
Density jump, $c = 1$	$[-1, 1]$	10000	500	0.630	0.093
Density jump, $c = 1.5$	$[-1, 1]$	2000	500	0.946	0.014
Density jump, $c = 1.5$	$[-1, 1]$	10000	500	0.976	0.009

6 Australian tax evidence

6.1 Implementation and annual results

For each tax year from 2001 through 2022, the statistic is evaluated on the common interval

$$I = [\$1{,}000, \$199{,}000].$$

This prespecified interval contains every nonzero statutory personal income tax threshold in the sample years. We apply the multiplier implementation in Corollary 1 separately to individual taxable income and family average taxable income. The preferred specification uses the Silverman scale estimate (Silverman, 1986)

$$\widehat{s}_n = \min\left\{\widehat{\sigma}_n, \frac{\widehat{\mathrm{IQR}}_n}{1.34898}\right\}, \qquad h_n = 0.9\widehat{s}_n n^{-0.40},$$

and freezes the resulting bandwidth across 1,999 Gaussian multiplier draws. The exponent 0.40 is chosen based on the difficult smooth-null calibration in Appendix Table 10. A sensitivity specification uses $h_n = 0.9\widehat{s}_n n^{-0.35}$ and otherwise repeats the test unchanged. The main family analysis uses the non-singleton, internally consistent family units defined in Section 2. Annual samples contain 1,311,566 to 1,809,227 individual returns and 381,086 to 529,474 linked family units.

Under the preferred specification, individual taxable income rejects the maintained smooth class in all 22 years, with every p-value equal to the plus-one Monte Carlo floor of 0.0005. Family average income does not reject in any year; family p-values range from 0.0760 to 0.9295. Table 4 reports the annual results. The $\beta = 0.35$ sensitivity specification again rejects for every individual distribution and yields one family rejection, in 2007 with $p = 0.0050$.

Table 4: Centered Kolmogorov–Smirnov tests in Australian tax data, 2001–2022.

			Individual taxable income		Family average income	
Tax year	N_I	N_F	Statistic	p-value	Statistic	p-value
2001	1,311,566	381,086	0.06979	0.0005	0.01064	0.1915
2002	1,338,894	391,021	0.07709	0.0005	0.01109	0.0760
2003	1,373,642	411,716	0.07857	0.0005	0.00880	0.7135
2004	1,416,813	421,523	0.05535	0.0005	0.00825	0.8670
2005	1,463,432	435,809	0.06021	0.0005	0.00901	0.4210
2006	1,503,532	448,538	0.07342	0.0005	0.00898	0.3215
2007	1,549,398	462,276	0.08229	0.0005	0.00924	0.1620
2008	1,616,357	483,181	0.07307	0.0005	0.00810	0.5515
2009	1,616,172	482,306	0.07235	0.0005	0.00831	0.3295
2010	1,631,962	487,589	0.06570	0.0005	0.00858	0.1750
2011	1,660,606	495,899	0.09655	0.0005	0.00809	0.2670
2012	1,673,961	496,305	0.07890	0.0005	0.00777	0.3470
2013	1,666,463	496,988	0.08096	0.0005	0.00829	0.1095
2014	1,699,498	505,277	0.09287	0.0005	0.00707	0.5930
2015	1,733,189	512,942	0.09000	0.0005	0.00669	0.7950
2016	1,759,647	518,861	0.07519	0.0005	0.00739	0.3195
2017	1,787,031	524,705	0.10303	0.0005	0.00721	0.3980
2018	1,800,007	527,204	0.07901	0.0005	0.00615	0.9295
2019	1,809,227	529,474	0.11139	0.0005	0.00628	0.8395
2020	1,801,176	526,588	0.06066	0.0005	0.00661	0.5630
2021	1,785,516	520,246	0.07410	0.0005	0.00666	0.4665
2022	1,753,042	510,322	0.10346	0.0005	0.00604	0.7285

Notes: The table reports the preferred $\beta = 0.40$ specification. N_I counts retained person-years and N_F counts retained non-singleton family-year units after the strict $(0, \$200{,}000)$ income filter. The individual and family distributions are respectively person-weighted and family-weighted. Both tests use the prespecified interval $I = [\$1{,}000, \$199{,}000]$. The reported statistic is $U_n = \sqrt{n/h}\sup_{x\in I}|F_n(x) - \widehat{F}_h(x)|$; its reference law is the centered Gaussian-multiplier process in Corollary 1. Each p-value is the finite-simulation plus-one value based on 1,999 multiplier draws; 0.0005 is therefore the smallest reportable value. Tests are two-sided at the five-percent level.

The annual contrast is therefore stable across the two bandwidths. Individual income rejects in all 44 year-by-bandwidth tests, whereas family average income rejects only once. Every other raw-family p-value remains above five percent. Rounding family averages to whole dollars under the $\beta = 0.35$ sensitivity specification preserves the broad contrast. The rounded family distribution rejects in 2007 ($p = 0.0080$) and 2011 ($p = 0.0470$); every other rounded-family p-value is at least 0.0925. Appendix Table 12 reports the complete comparison, and Appendix Figures 4 and 5 show the individual and linked-family distributions in four additional years.

6.2 Calibrated family responses and statistical power

The 2021 sample provides a concrete benchmark for the family test. Under the preferred specification, the family sample has $n_F = 520{,}246$, an average family size of 2.45, bandwidth

$h_F = 130.5313$, and conditional 95 percent critical value $c_F = 0.0078045$. Proposition 5 implies that an atom of mass m creates a population CDF discrepancy of at least $m/2$ from the continuous smoothed benchmark. On the normalized scale of the test, its signal is therefore at least

$$\frac{m}{2}\sqrt{\frac{n}{h}}.$$

Equating this signal to the conditional critical value gives the single-atom benchmark

$$m_{\text{crit}} = 2c_F\sqrt{\frac{h_F}{n_F}} = 0.0247\%.$$

This benchmark corresponds to roughly 129 families in the 2021 sample. It is not a hard threshold for alternatives whose mass is spread across several composite locations.

We next calibrate a transparent two-member real response benchmark to the 2021 individual distribution. We construct a smooth empirical distribution of baseline resources by interpolating through neighborhoods of the four principal statutory and effective-schedule markers, \$23,226, \$45,000, \$120,000, and \$180,000, and smoothing the resulting density. At zero allocability an individual with baseline type w chooses taxable income to maximize

$$x - T(x) - \frac{(x-w)^2}{2\tau}$$

under the effective 2020–21 tax schedule. We choose the common response parameter τ to match jointly the observed probability mass within combined \$300 and \$500 windows around the four markers. These raw window masses are 1.1270 percent and 1.7713 percent, respectively. They measure concentration around the kinks rather than conventional excess bunching. The preferred calibration gives $\widehat{\tau}_0 = 822.415$ and fitted masses of 1.1634 percent and 1.6904 percent. Appendix F.4 gives the construction in full.

Table 5: Calibration of the real-response parameter to the 2021 individual distribution.

	Observed	Calibrated model
Mass in combined \$300 kink windows	1.1270%	1.1634%
Mass in combined \$500 kink windows	1.7713%	1.6904%
$\widehat{\tau}_0 = 822.415$.		

Notes: The observed moments are union probabilities across neighborhoods of the four principal markers, \$23,226, \$45,000, \$120,000, and \$180,000. The preferred calibration minimizes the sum of squared relative errors in the two moments under the effective 2020–21 tax schedule.

At zero allocability, aggregation substantially attenuates the calibrated real response. We draw independent two-member families from the calibrated smooth benchmark, hold $\widehat{\tau}_0$ fixed,

simulate fresh samples of 520,246 families, and apply the preferred family test. With $\alpha = 0$, only 2 of 100 replications reject. The exact 95 percent binomial interval for the rejection probability is 0.24 to 7.04 percent. A real response calibrated to the individual evidence can therefore become statistically difficult to detect after aggregation when family members respond independently.

We then introduce within-family allocation while holding both the calibrated responsiveness and the benchmark distribution fixed. For member j with baseline resources w_j, let

$$\theta_j = (1 - \alpha)w_j, \qquad P = \alpha(w_1 + w_2),$$

where P is the family pool that can be allocated across members. The parameter α therefore measures the share of baseline family resources made allocable, not the fraction of realized income that is mechanically shifted. The population exercise solves the family problem for one million independently drawn families at each value of α.

Figure 3 traces the resulting selected composite atoms over the economically relevant lower part of the allocability range. The diagnostic counts mass within one cent of composite locations formed from the four principal calibration markers. It therefore does not presume that these are the only kinks generated by the effective tax schedule. The largest selected atom rises from 0.0003 percent at $\alpha = 0$ to 0.0197 percent at $\alpha = 0.10$, reaches the 0.0247 percent single-atom benchmark at $\alpha = 0.125$, and is 0.0375 to 0.0443 percent at $\alpha = 0.175$ to 0.20. The complete selected population results are reported in Appendix Table 13.

For scale, the median passive income share in the linked-family data is approximately 18 percent. This accounting share is not identical to α: the model parameter measures the fraction of baseline resources that can be allocated within the family and is not estimated from observed passive income. The comparison is nevertheless informative about magnitude. The population points bracketing 18 percent imply a largest selected family atom between 0.0375 and 0.0443 percent, approximately 1.5 to 1.8 times the 2021 single-atom benchmark. The calibrated exercise therefore shows that aggregation can sharply attenuate the detectable family response at $\alpha = 0$, but even modest allocability can rapidly produce an economically meaningful atom that is large relative to the test's detection scale.

In a more aggressive sensitivity check, we assign the full 1.1270 percent of probability observed within the \$300 windows as point mass at the four principal markers. This raises the implied response parameter to $\tau = 5294.7$. The family test then rejects in 43 of 100 replications at $\alpha = 0$ and in all 100 replications at $\alpha = 0.01$. This calibration deliberately attributes the smooth component of probability within the windows to bunching and therefore provides an upper bound rather than an alternative estimate of the individual response. The resulting increase in family rejection rates shows how quickly the aggregate implications

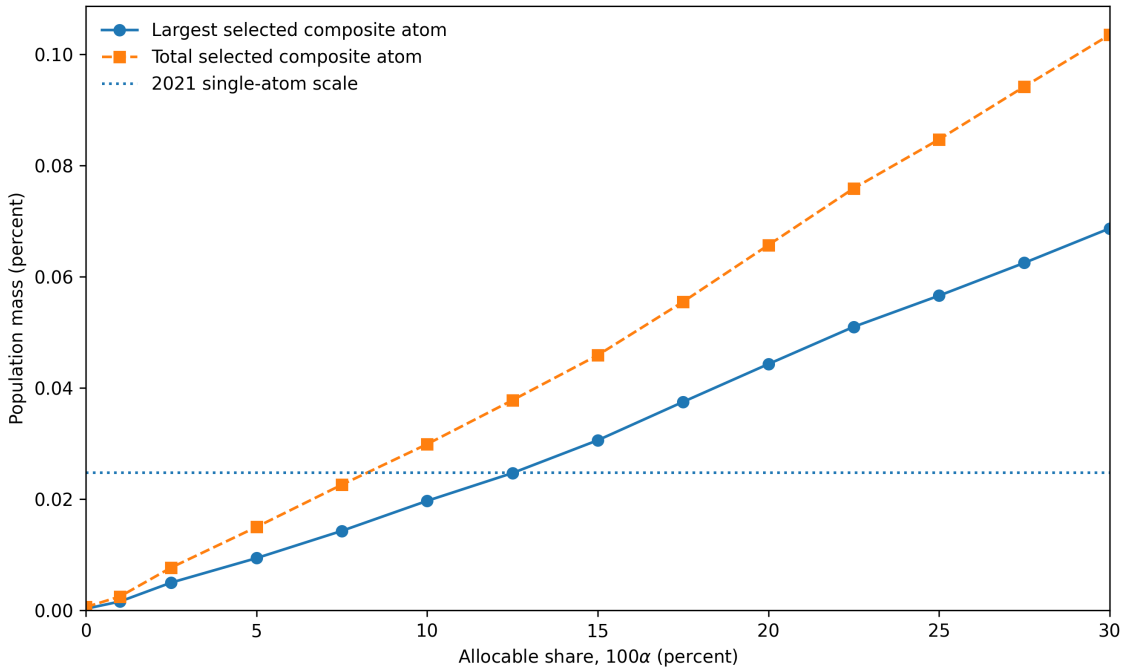


Figure 3: Population atoms in the calibrated two-member family model. The response parameter and smooth benchmark distribution are calibrated once at $\alpha = 0$ and then held fixed. The solid line reports the largest selected composite atom and the dashed line the sum across selected composite locations formed from the four principal calibration markers. The horizontal line is the 2021 single-atom benchmark of 0.0247 percent. The atom diagnostics are descriptive population objects; the statistical test remains global and searches the full prespecified interval.

strengthen under a larger calibrated individual response. Appendix Table 14 reports the results.

The calibrated results sharpen the interpretation of the annual evidence. Aggregation alone can substantially attenuate a genuine individual response, so family smoothness cannot be read mechanically as proof of shifting. But once the same calibrated real responsiveness is combined with meaningful scope for within-family allocation, the model generates increasingly large family atoms, while the actual linked family distribution continues not to reject smoothness. The passive income model is particularly relevant for families in which active earnings remain attached to their earners while trust, business, or investment income can be allocated across members. Johnson et al. (2024) document substantially greater bunching among self-employed taxpayers receiving trust income, especially those with spouses and children. The contrast between strong bunching in individual taxable income and the smoothness of family income suggests that the individual bunching reflects reallocation of income between individuals rather than changes in total family income.

7 Conclusion

Bunching at an individual tax kink shows that reported income responds to incentives, but it does not reveal what is changing. The same spike in individual taxable income may arise because people change how much income they generate or because income is reassigned across taxpayers. The individual distribution alone cannot distinguish these mechanisms.

Aggregation supplies an additional restriction. Income shifting within a family changes the allocation of income across personal returns without changing aggregate family income, whereas a real response changes the family total. This restriction does not depend on a capacity to shift. Even without shifting, individual real responses imply bunching in family income when the relevant family members bunch jointly. Aggregation can attenuate that response when members respond independently. With even modest potential for income shifting within the family, however, coordination preserves the baseline bunching response and can generate additional aggregate bunching. The implied response in family income then becomes economically and statistically large.

The Australian evidence strongly favors income shifting as the explanation for individual bunching. Individual taxable income rejects smoothness in every year from 2001 through 2022, while linked family average income does not reject in any year under the preferred specification. Individual bunching largely disappears once income is aggregated over the family. The evidence therefore indicates that the observed individual bunching is driven primarily by the reassignment of taxable income within families rather than by changes in total family income.

The same logic extends beyond families. Reported income may be reassigned across dates, accounts, related entities, or organizational units. Whenever such reassignment disappears after combining the linked reports while genuine changes in economic activity remain visible in the aggregate, the aggregate distribution contains information that the component distributions do not. The relevant unit is the smallest economically coherent collection of linked reports that contains the reallocation and closes the corresponding accounting identity. Aggregation can therefore distinguish changes in economic activity from changes in where that activity is reported.

A Economic proofs and extensions

A.1 Proof of Lemma 1

Proof. The feasible set in each program is a compact simplex, so semicontinuity gives attainment. If x^r solves the frontier problem at Y_r, then $\lambda x^1 + (1-\lambda)x^0$ is feasible at $\lambda Y_1 + (1-\lambda)Y_0$. Concavity of the a_j therefore gives concavity of A, while monotonicity follows because each a_j is increasing.

An analogous argument applied to cost minimizers gives convexity of C. If $Y_0 \neq Y_1$, the corresponding allocation vectors differ in at least one component. Strict convexity of the component cost functions then makes the convexity inequality strict. Hence $C(\cdot;\theta)$ is strictly convex. □

A.2 Proof of Proposition 1

Proof. Fix $q \in [q_+, q_-]$. Because $q \in \partial a_j(k_j)$ for every member,

$$a_j(x_j) \leq a_j(k_j) + q(x_j - k_j).$$

For any report vector satisfying $\sum_j x_j = K$, summing these inequalities gives

$$\sum_j a_j(x_j) \leq \sum_j a_j(k_j),$$

so $x = k$ solves the reporting problem at $Y = K$. In particular,

$$A(K) = \sum_j a_j(k_j).$$

More generally, if $\sum_j x_j = Y$, then

$$\sum_j a_j(x_j) \leq A(K) + q(Y - K).$$

Maximizing the left-hand side over report vectors summing to Y gives

$$A(Y) \leq A(K) + q(Y - K).$$

Hence $q \in \partial A(K)$. Since the argument applies to every $q \in [q_+, q_-]$,

$$[q_+, q_-] \subseteq \partial A(K).$$

Condition (6) places $C_Y(K;\bar{\theta})$ strictly inside $\mathcal{M}(\bar{\eta})$. Continuity of $C_Y(K;\theta)$, $\underline{\nu}(\eta)$, and $\overline{\nu}(\eta)$ therefore gives an open neighborhood $\mathcal{N}$ of $\bar{\omega}$ such that

$$C_Y(K;\theta) \in \operatorname{int} \mathcal{M}(\eta)$$

for every $\omega = (\theta, \eta) \in \mathcal{N}$.

Fix any $\omega = (\theta, \eta) \in \mathcal{N}$. By the definition of $\mathcal{M}(\eta)$, there exist

$$\nu \in \partial v(A(K);\eta) \qquad \text{and} \qquad q \in [q_+, q_-]$$

such that

$$C_Y(K;\theta) = \nu q.$$

Because ν is a supergradient of $v(\cdot;\eta)$ at $A(K)$,

$$v(A(Y);\eta) \leq v(A(K);\eta) + \nu\{A(Y) - A(K)\}.$$

Since $v(\cdot;\eta)$ is increasing, $\nu \geq 0$. Because $q \in \partial A(K)$,

$$A(Y) - A(K) \leq q(Y - K).$$

Multiplying by ν and combining the two supergradient inequalities yields

$$v(A(Y);\eta) \leq v(A(K);\eta) + \nu q(Y - K) = v(A(K);\eta) + C_Y(K;\theta)(Y - K).$$

Convexity and differentiability of $C(\cdot;\theta)$ at K give

$$C(Y;\theta) \geq C(K;\theta) + C_Y(K;\theta)(Y - K).$$

Subtracting yields

$$v(A(Y);\eta) - C(Y;\theta) \leq v(A(K);\eta) - C(K;\theta).$$

Hence K is a global maximizer of the family's scalar problem.

Because A is concave and $v(\cdot;\eta)$ is increasing and concave, $v(A(\cdot);\eta)$ is concave. Lemma 1 makes $C(\cdot;\theta)$ strictly convex, so

$$v(A(Y);\eta) - C(Y;\theta)$$

is strictly concave. Hence K is the unique maximizer for every family type in $\mathcal{N}$.

If $\bar{\omega}$ lies in the support of the distribution of family types, then $\mathcal{N}$ has positive probability. Every family with type in $\mathcal{N}$ chooses total real income K, producing an atom at K. Under free shifting,

$$\sum_j x_j = \sum_j y_j,$$

so total reported family income has the same atom. □

A.3 Proof of Theorem 1

Proof. Strict convexity of $c_j(\cdot;\theta_j)$ makes $c_{j,y}(\cdot;\theta_j)$ strictly increasing. Hence $y_j(\mu;\theta_j)$ is strictly increasing in μ, and

$$Q_I(\mu;\theta) = \sum_{j\in I}\{k_j - y_j(\mu;\theta_j)\}$$

is continuous and strictly decreasing in μ.

Choose

$$\mu^- < \bar{\mu} < \mu^+$$

sufficiently close to $\bar{\mu}$ that

$$[\mu^-, \mu^+] \subseteq \operatorname{int}\Lambda_I(\bar{\eta}),$$

$$y_j(\mu^+;\bar{\theta}_j) < k_j \qquad (j \in I),$$

and

$$\ell_j(\bar{\theta}_j) < \mu^- \qquad (j \notin I).$$

Because

$$Q_I(\bar{\mu};\bar{\theta}) = \bar{P}$$

and $Q_I(\cdot;\bar{\theta})$ is strictly decreasing,

$$Q_I(\mu^+;\bar{\theta}) < \bar{P} < Q_I(\mu^-;\bar{\theta}).$$

Continuity of ν, the inverse marginal cost functions, $c_{j,y}$, and $P = \sum_j \iota_j$ implies that, after shrinking to some open neighborhood $\mathcal{N}$ of $\bar{\omega}$ if necessary, all of these strict inequalities persist. In particular, for every

$$\omega = (\theta, \iota, \eta) \in \mathcal{N},$$

we have

$$[\mu^-, \mu^+] \subseteq \operatorname{int} \Lambda_I(\eta),$$

$$y_j(\mu^+; \theta_j) < k_j \qquad (j \in I),$$

$$\ell_j(\theta_j) \in \operatorname{int} \Lambda_j(\eta) \qquad (j \notin I),$$

$$\ell_j(\theta_j) < \mu^- \qquad (j \notin I),$$

and

$$Q_I(\mu^+; \theta) < P < Q_I(\mu^-; \theta).$$

Fix any $\omega = (\theta, \iota, \eta) \in \mathcal{N}$. By continuity and strict monotonicity of $Q_I(\cdot; \theta)$, the intermediate value theorem gives a unique

$$\mu^* \in (\mu^-, \mu^+)$$

such that

$$Q_I(\mu^*; \theta) = P.$$

Define

$$(y_j^*, r_j^*) = \begin{cases} \big(y_j(\mu^*; \theta_j),\, k_j - y_j(\mu^*; \theta_j)\big), & j \in I, \\ (k_j, 0), & j \notin I, \end{cases}$$

and set

$$x_j^* = k_j \qquad (j \in \mathcal{J}).$$

Because $\mu^* < \mu^+$ and $y_j(\mu; \theta_j)$ is increasing in μ,

$$y_j(\mu^*; \theta_j) < y_j(\mu^+; \theta_j) < k_j \qquad (j \in I),$$

so $r_j^* > 0$ for every recipient. Moreover,

$$\sum_j r_j^* = Q_I(\mu^*; \theta) = P.$$

Thus the candidate allocation is feasible.

For each member, define

$$\lambda_j^* = \begin{cases} \mu^*, & j \in I, \\ \ell_j(\theta_j), & j \notin I. \end{cases}$$

The maintained strict conditions imply

$$\lambda_j^* \in \operatorname{int} \Lambda_j(\eta) \qquad \text{for every } j,$$

and

$$\lambda_j^* < \mu^* \qquad (j \notin I).$$

By construction,

$$c_{j,y}(y_j^*; \theta_j) = \lambda_j^* \qquad \text{for every } j.$$

We now show that the candidate is globally optimal. Let

$$\nu = \nu(\eta).$$

Since

$$\lambda_j^* \in \nu \, \partial a_j(k_j),$$

we have

$$\frac{\lambda_j^*}{\nu} \in \partial a_j(k_j).$$

Hence, for every x_j,

$$\nu \left\{ a_j(x_j) - a_j(k_j) \right\} \leq \lambda_j^* (x_j - k_j).$$

Concavity of $v(\cdot; \eta)$ and differentiability at R_k give

$$v\left(\sum_j a_j(x_j); \eta \right) - v(R_k; \eta) \leq \nu \sum_j \left\{ a_j(x_j) - a_j(k_j) \right\},$$

while convexity of $c_j(\cdot; \theta_j)$ gives

$$c_j(y_j; \theta_j) - c_j(y_j^*; \theta_j) \geq \lambda_j^* (y_j - y_j^*).$$

Combining these inequalities, for any feasible (y, r) with $x_j = y_j + r_j$,

$$\begin{aligned} v\left(\sum_j a_j(x_j); \eta \right) - \sum_j c_j(y_j; \theta_j) - v(R_k; \eta) + \sum_j c_j(y_j^*; \theta_j) \\ \leq \sum_j \lambda_j^* \left\{ (x_j - k_j) - (y_j - y_j^*) \right\} \\ = \sum_j \lambda_j^* (r_j - r_j^*). \end{aligned}$$

Because

$$\sum_j r_j = \sum_j r_j^* = P,$$

we may subtract μ^* times this equality to obtain

$$\begin{aligned}\sum_j \lambda_j^*(r_j - r_j^*) &= \sum_j (\lambda_j^* - \mu^*)(r_j - r_j^*) \\ &= \sum_{j \notin I} (\lambda_j^* - \mu^*) r_j \\ &\leq 0.\end{aligned}$$

The second equality uses $\lambda_j^* = \mu^*$ for $j \in I$ and $r_j^* = 0$ for $j \notin I$. The final inequality follows from

$$\lambda_j^* < \mu^* \qquad (j \notin I).$$

Thus the candidate is a global optimum.

It remains to establish uniqueness. If any nonrecipient receives positive passive income, then

$$\sum_{j \notin I} (\lambda_j^* - \mu^*) r_j < 0,$$

so such an allocation cannot be optimal. Hence every optimum must have

$$r_j = 0 \qquad (j \notin I).$$

Strict convexity of the earning costs makes the supporting inequality for c_j strict whenever $y_j \neq y_j^*$, so any optimum must also satisfy

$$y_j = y_j^* \qquad \text{for every } j.$$

Finally, because

$$\lambda_j^* \in \operatorname{int} \Lambda_j(\eta),$$

the corresponding supporting slope

$$\frac{\lambda_j^*}{\nu}$$

lies in the interior of the strict kink interval $\partial a_j(k_j)$. The supporting inequality for a_j is therefore strict unless

$$x_j = k_j.$$

Hence every optimum must satisfy

$$x_j = k_j \qquad \text{for every } j,$$

which, together with $x_j = y_j + r_j$, also pins down $r_j = r_j^*$. The candidate solution is therefore unique.

Every family with type in $\mathcal{N}$ consequently satisfies

$$\sum_j x_j = K.$$

Since every feasible allocation obeys

$$\sum_j x_j = \sum_j y_j + P,$$

total active and passive family income also equals K. If $\bar{\omega}$ lies in the support of the distribution of family types, then $\mathcal{N}$ has positive probability. Total reported family income and total active and passive family income therefore each have an atom at K. □

A.4 Real income coordinates for costly shifting

The costly shifting problem in the main text can be derived from a primitive effort model without retaining a nonlinear production equality in the optimization variables. Let member j choose effort $e_j \geq 0$, produce real taxable resources

$$y_j = f_j(e_j; \theta_j),$$

and incur effort cost $d_j(e_j; \theta_j)$. Suppose $f_j(\cdot; \theta_j)$ is continuous, strictly increasing, concave, and onto the relevant real-income range, while $d_j(\cdot; \theta_j)$ is increasing and convex. Define the induced real-income cost

$$c_j(y; \theta_j) = d_j\left(f_j^{-1}(y; \theta_j); \theta_j\right). \tag{21}$$

The inverse of an increasing concave function is increasing and convex. Because d_j is increasing and convex, (21) is convex. It is strictly convex, for example, if d_j is strictly convex, or if f_j^{-1} is strictly convex and d_j is strictly increasing. The map between e_j and y_j is one-to-one, so optimizing in (x, e, s) subject to

$$x_j = f_j(e_j; \theta_j) + s_j$$

is equivalent to optimizing in (x, y, s) subject to the affine identity

$$x_j - y_j - s_j = 0.$$

Thus the real-income formulation in Theorem 2 is a reparameterization of this primitive effort problem.

A.5 Proof of Theorem 2

Proof. For every λ, consider

$$H(z;\lambda) = \sum_{j=1}^{J} (\phi_j')^{-1}(\lambda_j + z).$$

Because each ϕ_j' is continuous, strictly increasing, and onto $\mathbb{R}$, its inverse is continuous and strictly increasing. Hence $H(\cdot;\lambda)$ is continuous and strictly increasing, with

$$H(z;\lambda) \to -\infty \quad \text{as } z \to -\infty, \qquad H(z;\lambda) \to +\infty \quad \text{as } z \to +\infty.$$

It therefore has a unique zero $\zeta(\lambda)$.

The map $\zeta(\lambda)$ is continuous. To see this, let $\lambda^m \to \lambda$ and choose $a < b$ such that

$$H(a;\lambda) < 0 < H(b;\lambda).$$

For all sufficiently large m,

$$H(a;\lambda^m) < 0 < H(b;\lambda^m),$$

so

$$\zeta(\lambda^m) \in [a, b].$$

Every convergent subsequence of $\zeta(\lambda^m)$ must solve

$$H(z;\lambda) = 0$$

in the limit and therefore must converge to the unique solution $\zeta(\lambda)$. Hence

$$\zeta(\lambda^m) \to \zeta(\lambda).$$

It follows that each S_j and each G_j is continuous.

Because

$$[\lambda_j^-, \lambda_j^+] \subset \operatorname{int} \Lambda_j(\bar{\eta})$$

and $\nu(\eta)$ is continuous, these inclusions persist for all η sufficiently close to $\bar{\eta}$. Because the face inequalities (13)–(14) hold with the strict margin ε, continuity of G and compactness of the faces of $\mathcal{B}$ imply that they also persist, with a possibly smaller positive margin, for all θ sufficiently close to $\bar{\theta}$. Hence there is an open neighborhood $\mathcal{N}$ of $\bar{\omega}$ such that, for every

$$\omega = (\theta, \eta) \in \mathcal{N},$$

the box $\mathcal{B}$ lies inside the relevant kink intervals and the corresponding face inequalities hold.

Fix any $\omega = (\theta, \eta) \in \mathcal{N}$. By the Poincaré–Miranda theorem, there exists

$$\lambda^* \in \mathcal{B}$$

such that

$$G_j(\lambda^*; \theta) = 0 \qquad (j = 1, \dots, J).$$

Define

$$x_j^* = k_j, \qquad y_j^* = y_j(\lambda_j^*; \theta_j), \qquad s_j^* = S_j(\lambda^*).$$

The definition of $\zeta(\lambda^*)$ gives

$$\sum_{j=1}^{J} s_j^* = 0.$$

Moreover,

$$G_j(\lambda^*; \theta) = 0$$

implies

$$s_j^* = k_j - y_j^*,$$

so

$$x_j^* - y_j^* - s_j^* = 0.$$

The candidate allocation is therefore feasible.

By construction,

$$\lambda_j^* \in \nu(\eta)\, \partial a_j(k_j),$$

$$c_{j,y}(y_j^*; \theta_j) = \lambda_j^*,$$

and

$$\phi_j'(s_j^*) = \lambda_j^* + \zeta(\lambda^*).$$

We now show directly that these conditions imply global optimality.

Write

$$\nu = \nu(\eta), \qquad \zeta^* = \zeta(\lambda^*).$$

Because

$$\frac{\lambda_j^*}{\nu} \in \partial a_j(k_j),$$

the supporting inequality for a_j gives, for every x_j,

$$\nu \left\{ a_j(x_j) - a_j(k_j) \right\} \leq \lambda_j^*(x_j - k_j).$$

Concavity of $v(\cdot;\eta)$ and differentiability at R_k imply

$$v\left(\sum_j a_j(x_j); \eta \right) - v(R_k;\eta) \leq \nu \sum_j \left\{ a_j(x_j) - a_j(k_j) \right\}.$$

Combining these inequalities yields

$$v\left(\sum_j a_j(x_j); \eta \right) - v(R_k;\eta) \leq \sum_j \lambda_j^*(x_j - k_j).$$

Convexity of $c_j(\cdot;\theta_j)$ gives

$$c_j(y_j;\theta_j) - c_j(y_j^*;\theta_j) \geq \lambda_j^*(y_j - y_j^*),$$

while convexity of ϕ_j gives

$$\phi_j(s_j) - \phi_j(s_j^*) \geq (\lambda_j^* + \zeta^*)(s_j - s_j^*).$$

Therefore, for any feasible (x, y, s),

$$\begin{aligned} v\left(\sum_j a_j(x_j); \eta \right) &- \sum_j c_j(y_j;\theta_j) - \sum_j \phi_j(s_j) \\ &- v(R_k;\eta) + \sum_j c_j(y_j^*;\theta_j) + \sum_j \phi_j(s_j^*) \\ &\leq \sum_j \lambda_j^* \left\{ (x_j - k_j) - (y_j - y_j^*) - (s_j - s_j^*) \right\} - \zeta^* \sum_j (s_j - s_j^*). \end{aligned}$$

For both the candidate allocation and any feasible alternative,

$$x_j - y_j - s_j = 0$$

for every j, and

$$\sum_j s_j = 0.$$

Hence the right-hand side is zero. The candidate is therefore a global optimum.

Strict convexity of every c_j and ϕ_j makes the objective strictly concave in (y, s) on the affine feasible set. Since

$$x = y + s,$$

any two distinct feasible allocations must differ in y or s. The global optimizer is therefore unique.

Every family with type in $\mathcal{N}$ consequently satisfies

$$x_j = k_j \qquad (j = 1, \ldots, J),$$

and hence

$$\sum_j x_j = K.$$

Because

$$\sum_j s_j = 0$$

and $x_j = y_j + s_j$,

$$\sum_j y_j = \sum_j x_j = K.$$

If $\bar{\omega}$ lies in the support of the distribution of family types, then $\mathcal{N}$ has positive probability. Total reported family income and total real family income therefore each have an atom at K. □

A.6 Proof of Proposition 2

Proof. Fix

$$(\theta, \eta) \in \mathcal{B}_0(k)$$

and define

$$\ell_j = c_{j,y}(k; \theta_j), \qquad q_j = \frac{\ell_j}{\nu(\eta)} \in \partial a(k).$$

Consider first the no shifting problem. For any feasible income vector y, concavity of $v(\cdot;\eta)$ gives

$$v\left(\sum_j a(y_j);\eta\right) - v(R_k;\eta) \le \nu(\eta)\sum_j\{a(y_j)-a(k)\}.$$

Because $q_j \in \partial a(k)$,

$$\nu(\eta)\{a(y_j)-a(k)\} \le \ell_j(y_j-k)$$

for each member. Convexity of $c_j(\cdot;\theta_j)$ gives

$$c_j(y_j;\theta_j) - c_j(k;\theta_j) \ge \ell_j(y_j-k).$$

Combining these inequalities,

$$\begin{aligned} v\left(\sum_j a(y_j);\eta\right) &- \sum_j c_j(y_j;\theta_j) - v(R_k;\eta) + \sum_j c_j(k;\theta_j) \\ &\le \sum_j \ell_j(y_j-k) - \sum_j\{c_j(y_j;\theta_j) - c_j(k;\theta_j)\} \\ &\le 0. \end{aligned}$$

Thus $y_j = k$ for every j is optimal. Strict convexity of the earning costs makes the final inequality strict whenever $y \ne (k,\dots,k)$, so the no shifting optimum is unique.

Now consider free shifting. Let

$$\ell_- = \min_j \ell_j, \qquad \ell_+ = \max_j \ell_j.$$

Since each inverse marginal cost $y_j(\lambda;\theta_j)$ is increasing in λ and

$$y_j(\ell_j;\theta_j) = k,$$

we have

$$y_j(\ell_-;\theta_j) \le k \le y_j(\ell_+;\theta_j) \qquad \text{for every } j.$$

Therefore,

$$\sum_j y_j(\ell_-;\theta_j) \le K \le \sum_j y_j(\ell_+;\theta_j).$$

If $\ell_- < \ell_+$, continuity gives some

$$\lambda^* \in [\ell_-,\ell_+]$$

such that

$$\sum_j y_j(\lambda^*;\theta_j) = K.$$

If $\ell_- = \ell_+$, the same equality holds at $\lambda^* = \ell_-$. Because all $y_j(\lambda^*;\theta_j)$ are positive, differentiability and strict convexity of the earning costs imply that λ^* is the common marginal earning cost in the cost-minimizing decomposition of total real income K. Hence

$$C_Y(K;\theta) = \lambda^*.$$

Since

$$\ell_j \in \mathcal{L}_k(\eta) \qquad \text{for every } j$$

and $\mathcal{L}_k(\eta)$ is an interval,

$$\lambda^* \in [\ell_-, \ell_+] \subseteq \mathcal{L}_k(\eta).$$

Thus

$$\lambda^* = \nu(\eta) q^*$$

for some

$$q^* \in \partial a(k).$$

The fixed-type optimality argument in Proposition 1 then implies that K is optimal under free shifting. Strict convexity of $C(\cdot;\theta)$ makes this optimal total unique, and assigning $x_j = k$ to every member is an optimal reporting allocation.

Finally, consider finite shifting. Fix the report vector

$$x_j = k \qquad (j = 1, \ldots, J)$$

and let s^* solve

$$\min_s \left\{ \sum_j \left[c_j(k - s_j;\theta_j) + \phi_j(s_j) \right] : \sum_j s_j = 0, \quad k - s_j \geq 0 \text{ for every } j \right\}.$$

By assumption the solution is interior. Set

$$y_j^* = k - s_j^*, \qquad \lambda_j^* = c_{j,y}(y_j^*;\theta_j).$$

The first-order conditions give a scalar ζ such that

$$\phi_j'(s_j^*) = \lambda_j^* + \zeta \qquad (j = 1, \ldots, J).$$

We show that

$$\ell_- \le \lambda_j^* \le \ell_+ \qquad \text{for every } j.$$

Suppose, to the contrary, that

$$\lambda_j^* < \ell_-.$$

Since $c_{j,y}(\cdot;\theta_j)$ is strictly increasing and

$$c_{j,y}(k;\theta_j) = \ell_j \ge \ell_-,$$

we must have

$$y_j^* < k, \qquad s_j^* > 0.$$

Because the shifts add to zero, there is some i with

$$s_i^* < 0.$$

Since $\phi_j'(0) = 0$ and each ϕ_j' is strictly increasing,

$$\phi_j'(s_j^*) > 0,$$

and hence

$$\zeta > -\lambda_j^* > -\ell_-.$$

For member i,

$$s_i^* < 0$$

implies

$$y_i^* > k$$

and therefore

$$\lambda_i^* > \ell_i \ge \ell_-.$$

Moreover,

$$\phi_i'(s_i^*) < 0,$$

so

$$\zeta < -\lambda_i^* < -\ell_-,$$

a contradiction. Thus

$$\lambda_j^* \ge \ell_-.$$

The upper bound follows symmetrically. If

$$\lambda_j^* > \ell_+,$$

then

$$y_j^* > k, \qquad s_j^* < 0.$$

The adding-up constraint then gives some i with

$$s_i^* > 0.$$

The shifting first-order conditions imply

$$\zeta < -\lambda_j^* < -\ell_+,$$

whereas $s_i^* > 0$ implies

$$y_i^* < k, \qquad \lambda_i^* < \ell_i \leq \ell_+,$$

and hence

$$\zeta > -\lambda_i^* > -\ell_+,$$

again a contradiction. Therefore

$$\lambda_j^* \leq \ell_+$$

for every j.

It follows that

$$\lambda_j^* \in [\ell_-, \ell_+] \subseteq \mathcal{L}_k(\eta) \qquad \text{for every } j.$$

The candidate (x^*, y^*, s^*), with $x_j^* = k$, therefore satisfies

$$\lambda_j^* \in \nu(\eta)\, \partial a(k), \qquad c_{j,y}(y_j^*; \theta_j) = \lambda_j^*, \qquad \phi_j'(s_j^*) = \lambda_j^* + \zeta.$$

The supporting inequality argument in the proof of Theorem 2 then shows that this candidate is globally optimal. Strict convexity of the earning and shifting costs makes the optimum unique.

We have therefore shown that every family type in $\mathcal{B}_0(k)$ continues to choose aggregate income K under free and finite shifting. The bounds in (17) follow immediately.

If $\theta_1, \ldots, \theta_J$ are independent conditional on η, then

$$\mathbb{P}\{\mathcal{B}_0(k) \mid \eta\} = \prod_{j=1}^{J} p_j(k, \eta).$$

Taking expectations over η yields

$$\mathbb{P}\{\mathcal{B}_0(k)\} = \mathbb{E}\left[\prod_{j=1}^{J} p_j(k, \eta)\right].$$

□

A.7 Passive-income nesting and positivity of the baseline mass

There is an analogous nesting argument in the passive income model when every member receives positive passive income. Suppose

$$0 < \iota_j < k \qquad (j = 1, \ldots, J),$$

and suppose each member would report k if their own passive income remained attached. Define

$$\ell_j^P = c_{j,y}(k - \iota_j; \theta_j) \in \mathcal{L}_k(\eta), \qquad P = \sum_j \iota_j,$$

and let

$$\ell_-^P = \min_j \ell_j^P, \qquad \ell_+^P = \max_j \ell_j^P.$$

Since

$$y_j(\ell_j^P; \theta_j) = k - \iota_j,$$

strict convexity implies

$$y_j(\ell_-^P; \theta_j) \le k - \iota_j \le y_j(\ell_+^P; \theta_j) \qquad \text{for every } j.$$

Hence the passive income required to place every report at k satisfies

$$Q_{\mathcal{J}}(\ell_-^P; \theta) \ge P \ge Q_{\mathcal{J}}(\ell_+^P; \theta).$$

Since $Q_{\mathcal{J}}(\cdot; \theta)$ is continuous and decreasing, there is some

$$\lambda^* \in [\ell_-^P, \ell_+^P]$$

such that

$$Q_{\mathcal{J}}(\lambda^*;\theta) = P.$$

Because

$$[\ell_-^P, \ell_+^P] \subseteq \mathcal{L}_k(\eta),$$

this common shadow value supports every report at the kink. If, in addition,

$$y_j(\ell_+^P;\theta_j) < k \qquad \text{for every } j,$$

then

$$y_j(\lambda^*;\theta_j) < k \qquad \text{for every } j,$$

so every implied passive allocation

$$r_j^* = k - y_j(\lambda^*;\theta_j)$$

is strictly positive. If each ℓ_j^P lies in

$$\operatorname{int} \mathcal{L}_k(\eta),$$

then λ^* also lies in $\operatorname{int} \mathcal{L}_k(\eta)$. The conditions of Theorem 1 are therefore satisfied at this family type, and the all-recipient solution persists for nearby family types.

Finally, the lower bounds in Proposition 2 are positive whenever $\mathcal{B}_0(k)$ contains a supported interior type. In particular, if

$$\bar{\omega} = (\bar{\theta}, \bar{\eta})$$

lies in the support of the family type distribution and

$$c_{j,y}(k;\bar{\theta}_j) \in \operatorname{int} \mathcal{L}_k(\bar{\eta}) \qquad \text{for every } j,$$

then continuity gives an open neighborhood of $\bar{\omega}$ contained in $\mathcal{B}_0(k)$. That neighborhood has positive probability, so

$$\mathbb{P}\{\mathcal{B}_0(k)\} > 0.$$

The aggregate mass at K is therefore positive under free shifting and, under the finite shifting conditions of Proposition 2, under finite shifting as well. Independence is not required for this conclusion; it is used only for the product representation of the baseline mass.

A.8 Proof of Proposition 3

Proof. The event

$$\left\{\sum_j Y_j^0 = K\right\}$$

is the disjoint union of the event

$$\left\{Y_1^0 = k_1, \ldots, Y_J^0 = k_J\right\}$$

and the event in the assumption of the proposition. The latter has probability zero, which gives the first equality. Under independence, the probability that every member bunches at the corresponding kink factors as

$$\prod_{j=1}^{J} \mathbb{P}(Y_j^0 = k_j) = \prod_{j=1}^{J} b_j.$$

□

A.9 Proof of Proposition 4

Proof. Without shifting, member j bunches at k if and only if

$$\theta_j \in [k - \tau q_-,\, k - \tau q_+].$$

By the choice of μ, this interval is centered at μ and has half-width

$$\frac{\tau(q_- - q_+)}{2} = \sigma z.$$

Hence

$$p = \mathbb{P}(|\theta_j - \mu| \leq \sigma z) = 2\Phi(z) - 1.$$

Independence and Proposition 3 then give the no shifting family atom

$$p^J.$$

Under free shifting, conditional on total real income Y, the family minimizes

$$\sum_{j=1}^{J} \frac{(y_j - \theta_j)^2}{2\tau}$$

subject to

$$\sum_{j=1}^{J} y_j = Y.$$

The minimizing allocation satisfies

$$y_j = \theta_j + \frac{Y - \sum_{\ell=1}^{J} \theta_\ell}{J},$$

and substitution gives

$$C(Y;\theta) = \frac{\left(Y - \sum_{j=1}^{J} \theta_j\right)^2}{2J\tau}.$$

Therefore

$$C_Y(Jk;\theta) = \frac{Jk - \sum_j \theta_j}{J\tau}.$$

As in the earlier quadratic example, aggregate income Jk is optimal if and only if

$$C_Y(Jk;\theta) \in [q_+, q_-],$$

or equivalently,

$$\frac{1}{J}\sum_{j=1}^{J} \theta_j \in [k - \tau q_-,\ k - \tau q_+].$$

Since

$$\frac{1}{J}\sum_{j=1}^{J} \theta_j \sim N\left(\mu, \frac{\sigma^2}{J}\right),$$

the probability of this event is

$$p_{F,J} = 2\Phi(\sqrt{J}\, z) - 1.$$

For $J > 1$ and $z > 0$,

$$2\Phi(\sqrt{J}\, z) - 1 > 2\Phi(z) - 1,$$

so

$$p_{F,J} > p.$$

Because $0 < p < 1$,

$$p > p^J.$$

Finally,

$$2\Phi(z) - 1 \sim \sqrt{\frac{2}{\pi}}\, z \qquad \text{as } z \downarrow 0,$$

and therefore

$$\frac{p_{F,J}}{p} = \frac{2\Phi(\sqrt{J}\,z) - 1}{2\Phi(z) - 1} \longrightarrow \sqrt{J}.$$

□

B Additional statistical results

B.1 Implementation details

The conditional probabilities and quantiles in Theorem 3 and Corollary 1 refer to the ideal reference law. With B simulated multiplier statistics

$$S_{n,1}^{\xi}, \ldots, S_{n,B}^{\xi},$$

we report

$$\widehat{p}_{n,B}^{\xi} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{S_{n,b}^{\xi} \geq U_n\}}{B+1}.$$

Let

$$p_n^{\xi} = \mathbb{P}^{\xi}\big(S_n^{\xi} \geq U_n \mid X_1, \ldots, X_n\big)$$

denote the ideal conditional multiplier p-value. Conditional on the data,

$$\widehat{p}_{n,B}^{\xi} - p_n^{\xi} = O_p(B^{-1/2}),$$

so the ideal asymptotic result carries over when $B = B_n \to \infty$. For fixed B, the reported value is the corresponding discrete plus-one Monte Carlo approximation, with minimum $1/(B+1)$.

Corollary 2 (Pilot-undersmoothed implementation)**.** *Suppose a pilot bandwidth satisfies*

$$\widehat{h}_n^{\mathrm{pilot}} = c_0 n^{-1/5}\{1 + O_p(r_n)\}, \qquad c_0 \in (0, \infty), \qquad r_n \ell_n^2 \to 0.$$

For any fixed $\beta \in (1/3, 1)$, set

$$\widehat{h}_{n,\beta} = \widehat{h}_n^{\mathrm{pilot}} n^{-(\beta - 1/5)}$$

and freeze the resulting bandwidth across resamples or multiplier draws. Then all three conclusions of Theorem 3, together with the corresponding multiplier conclusions in Corollary 1, continue to hold with h_n replaced by $\widehat{h}_{n,\beta}$.

Appendix C.7 proves the corollary by controlling the sample and multiplier processes

under the bandwidth perturbation. The Silverman scale pilot used in the application satisfies the displayed expansion with $r_n = n^{-1/2}$ when the fourth moment is finite, the limiting scale is positive, and the density is positive and continuous at the population quartiles.

B.2 Holdout calibration and power

Split the sample into training observations A and test observations B. Let $\widehat{F}_A$ denote the Gaussian KDE CDF estimated from the training sample with bandwidth h_A, and let F_B denote the empirical CDF of the test sample. The holdout discrepancy satisfies

$$\begin{aligned}\sqrt{n_B}(F_B - \widehat{F}_A) = \sqrt{n_B}(F_B - F) \\ - \sqrt{\frac{n_B}{n_A}}\sqrt{n_A}(\widehat{F}_A - F_{h_A}) \\ - \sqrt{n_B}(F_{h_A} - F).\end{aligned} \tag{22}$$

Undersmoothing makes the final term negligible, but the independently estimated training distribution continues to contribute at first order. Ordinary one-sample Kolmogorov–Smirnov calibration omits this variation.

Theorem 4 (Mixed-bridge holdout limit)**.** *Suppose F has a globally bounded density f with two bounded uniformly continuous derivatives and*

$$0 < c_f \le f \le C_f < \infty$$

on a neighborhood of I. Suppose also that

$$\frac{n_B}{n_A} \to \lambda \in (0, \infty), \qquad h_A \to 0, \qquad \sqrt{n_B} h_A^2 \to 0.$$

Then

$$\sqrt{n_B}\{F_B(x) - \widehat{F}_A(x)\} \rightsquigarrow B_1(F(x)) - \sqrt{\lambda} B_2(F(x))$$

in $\ell^\infty(I)$, where B_1 and B_2 are independent Brownian bridges. Consequently,

$$\sup_{x \in I} \left| B_1(F(x)) - \sqrt{\lambda} B_2(F(x)) \right| \overset{d}{=} \sqrt{1+\lambda} \sup_{x \in I} |B(F(x))|.$$

The mixed-bridge law therefore permits valid holdout calibration. Correct calibration does not, however, recover the cancellation of common empirical variation that occurs when the empirical and smoothed CDFs are constructed from the same observations.

Proposition 7 (Power cost of calibrated holdout)**.** *Suppose*

$$\frac{n_A}{n} \to \pi_A \in (0,1), \qquad \frac{n_B}{n} \to \pi_B = 1 - \pi_A, \qquad h_A \to 0, \qquad \sqrt{n_B} h_A^2 \to 0.$$

Calibrate the holdout statistic using the plug-in mixed-bridge law in Theorem 4.

(i) Under an atom array

$$P_n = (1 - m_n)P_0 + m_n \delta_{x_0},$$

with $x_0 \in \mathrm{int}(I)$ and smooth P_0, power tends to one if

$$m_n \sqrt{n_B} \to \infty.$$

If

$$m_n \sqrt{n_B} \to 0,$$

rejection tends to α.

(ii) Under the uniformly regular single-jump arrays of Theorem 3, power tends to one if

$$h_A |\Delta_n| \sqrt{n_B} \to \infty.$$

If this quantity tends to zero and the array has no other departure from the smooth limit, rejection tends to α.

Thus calibrated holdout has atom and density-jump boundaries

$$n_B^{-1/2} \qquad \textit{and} \qquad (h_A \sqrt{n_B})^{-1}.$$

With a fixed split and bandwidths of the same order, each boundary exceeds its full-sample centered counterpart in Theorem 3 by the diverging factor

$$\frac{1}{\sqrt{h_n \ell_n}}.$$

The difference comes from the stochastic scale. Holdout compares two independently estimated CDFs and therefore retains global sampling variation of order $n_B^{-1/2}$. The full-sample statistic constructs the empirical and smoothed CDFs from the same observations, so their common variation cancels and the supremum has stochastic scale

$$\sqrt{\frac{h_n \ell_n}{n}}.$$

Appendix C proves Proposition 7.

B.3 Minimax density-jump benchmark

Theorem 3 gives a sufficient detection condition for the proposed procedure. We now compare that condition with the best possible uniform rate for detecting a density jump at an unknown location.

Fix a nonempty compact interval

$$I_0 \subset \operatorname{int}(I)$$

and constants

$$0 < c_f < C_f < \infty.$$

For $0 < s \leq 1$ and $L > 0$, let $\mathcal{H}_s(L)$ contain densities bounded above by C_f on $\mathbb{R}$, bounded below by c_f on a neighborhood of I, and satisfying

$$|f(x) - f(y)| \leq L|x - y|^s, \qquad x, y \in \mathbb{R}.$$

Let $\mathcal{J}_s(\Delta, L)$ contain densities satisfying the same bounds, with an unknown jump point $x_0 \in I_0$, absolute jump magnitude at least Δ, and Hölder constant at most L separately on each side of x_0. Define

$$r_{n,s}(L, C_f) = L^{1/(2s+1)} C_f^{s/(2s+1)} \left(\frac{\log n}{n}\right)^{s/(2s+1)}$$

and

$$h^\star_{n,s} = \left(\frac{C_f \log n}{nL^2}\right)^{1/(2s+1)}.$$

For $Z \sim N(0,1)$, let

$$B_{s,\phi} = \frac{\mathbb{E}|Z|^{s+1}}{s+1}.$$

The fixed constants are taken to leave the null class with nonempty interior. In particular, $\mathcal{H}_s(L)$ contains a smooth density that is constant and strictly between the density bounds on a neighborhood of I_0 and has Hölder-s seminorm at most $L/2$.

Write

$$U_n(h) = \sqrt{\frac{n}{h}} \sup_{x \in I} |F_n(x) - \widehat{F}_h(x)|.$$

Theorem 5 (Minimax detection of an unknown density jump)**.** *For each fixed $s \in (0, 1]$, $L > 0$, and $\alpha \in (0, 1)$, the following statements hold.*

(i) There is a constant $c > 0$ such that every sequence of tests φ_n satisfying

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{H}_s(L)} \mathbb{E}_f\varphi_n \le \alpha$$

also satisfies

$$\limsup_{n\to\infty} \inf_{f\in\mathcal{J}_s(cr_{n,s}(L,C_f),L)} \mathbb{E}_f\varphi_n \le \alpha.$$

Thus no asymptotic level-α test is uniformly consistent below a sufficiently small multiple of $r_{n,s}(L, C_f)$.

(ii) Let $q^{\xi}_{n,1-\alpha}(h)$ be the conditional $(1-\alpha)$ quantile of the multiplier statistic in (20), evaluated at bandwidth h. The bias-aware test

$$\varphi_n^{\mathrm{BA}} = \mathbf{1}\Big\{U_n(h^\star_{n,s}) > q^{\xi}_{n,1-\alpha}(h^\star_{n,s}) + B_{s,\phi}L\sqrt{n}\,(h^\star_{n,s})^{s+1/2}\Big\} \tag{23}$$

satisfies

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{H}_s(L)} \mathbb{E}_f\varphi_n^{\mathrm{BA}} \le \alpha.$$

For a sufficiently large constant $C < \infty$,

$$\inf_{f\in\mathcal{J}_s(Cr_{n,s}(L,C_f),L)} \mathbb{E}_f\varphi_n^{\mathrm{BA}} \to 1.$$

Consequently,

$$r_{n,s}(L, C_f)$$

is the minimax separation rate, up to constants, for detecting an unknown interior density jump.

The bias allowance in (23) permits the rate-balancing bandwidth $h^\star_{n,s}$ while maintaining uniform size over the specified Hölder class. Procedure 1 instead uses undersmoothing and does not require the Hölder radius L to be specified.

Corollary 3 (Distance of the primary procedure from the jump frontier)**.** *For $s = 1$, fixed L and C_f, and*

$$h_n = cn^{-\beta}, \qquad \beta \in (1/3, 1),$$

the direct multiplier implementation without the bias allowance has uniformly asymptotic size α over $\mathcal{H}_1(L)$. Its sufficient density-jump boundary exceeds $r_{n,1}(L, C_f)$ by the factor

$$O\big(n^{(\beta-1/3)/2}(\log n)^{1/6}\big)\,.$$

At the preferred empirical choice $\beta = 0.40$, this factor is

$$O\big(n^{1/30}(\log n)^{1/6}\big)\,.$$

C Statistical proofs

C.1 Smoothing identities and local scale

Let H be the CDF of a symmetric random variable Z with $\mathbb{E}Z^2 < \infty$, and define

$$F_h(x) = \mathbb{E}H\left(\frac{x - X}{h}\right), \qquad \psi_{x,h}(u) = \mathbf{1}\{u \le x\} - H\left(\frac{x-u}{h}\right).$$

Equivalently, $F_h(x) = \mathbb{E}F(x - hZ)$. For the Gaussian kernel used in the paper, $H = \Phi$, $Z \sim N(0,1)$, and $\mathbb{E}Z^2 = 1$.

Lemma 2 (Uniform CDF bias)**.** *If f' is bounded and uniformly continuous and $\kappa_2 = \mathbb{E}Z^2$, then*

$$\sup_{x\in\mathbb{R}}\left|F_h(x) - F(x) - \frac{\kappa_2 h^2}{2}f'(x)\right| = o(h^2).$$

For the Gaussian kernel, $\kappa_2 = 1$.

Proof. Using $F_h(x) = \mathbb{E}F(x - hZ)$, Taylor's formula with integral remainder gives

$$F(x - hz) = F(x) - hzf(x) + \frac{h^2z^2}{2}f'(x) + h^2z^2R_h(x,z),$$

where, for every fixed M, $\sup_x \sup_{|z|\le M}|R_h(x,z)| \to 0$ by uniform continuity of f'. Symmetry makes $\mathbb{E}Z = 0$. Integrating over $|Z| \le M$, then sending $h \to 0$ and $M \to \infty$, controls the remainder by boundedness of f' and $\mathbb{E}Z^2 < \infty$. This yields the display uniformly in x. □

Proof of Proposition 6. Choose $\delta > 0$ so that the one-sided expansions hold on $(x_0 - \delta, x_0 + \delta)$ with a quadratic remainder. Let $A_h = \{|Z| \le \delta/h\}$. On A_h, for $Z > 0$,

$$F(x_0 - hZ) = F(x_0) - hZf(x_0-) + O(h^2Z^2),$$

while for $Z < 0$,

$$F(x_0 - hZ) = F(x_0) - hZf(x_0+) + O(h^2Z^2).$$

Because the kernel is Gaussian,

$$\mathbb{P}(A_h^c) = o(h^2), \qquad \mathbb{E}[|Z|\mathbf{1}\{A_h^c\}] = o(h),$$

so the contribution from A_h^c is $o(h^2)$. Integrating the two local expansions and using symmetry therefore gives

$$F_h(x_0) - F(x_0) = h\mu_\phi\{f(x_0+) - f(x_0-)\} + O(h^2), \qquad \mu_\phi = \mathbb{E}[Z\mathbf{1}\{Z > 0\}].$$

Taking absolute values proves the result. □

Lemma 3 (Local variance)**.** *If f is bounded, then*

$$\sup_{x\in I} \mathbb{E}\psi_{x,h}(X)^2 \le Ch.$$

If in addition f is continuous and bounded away from zero on a neighborhood of I, then for all small h,

$$ch \le \inf_{x\in I} \mathrm{Var}\{\psi_{x,h}(X)\} \le \sup_{x\in I} \mathrm{Var}\{\psi_{x,h}(X)\} \le Ch.$$

Proof. Let $k_0(v) = \mathbf{1}\{v \ge 0\} - H(v)$. Since $|\psi|^2 \le |\psi|$, changing variables $u = x - hv$ gives

$$\mathbb{E}\psi_{x,h}(X)^2 \le h\|f\|_\infty \int |k_0(v)|\, dv = O(h).$$

For the lower bound, change variables directly:

$$h^{-1}\mathbb{E}\psi_{x,h}(X)^2 = \int k_0(v)^2 f(x - hv)\, dv \to f(x) \int k_0(v)^2\, dv$$

uniformly on I. The squared mean is $O(h^2)$ and is negligible. The integral of k_0^2 is positive and finite. □

C.2 Holdout calibration

Proof of Theorem 4. Write P_A for the empirical measure of the training sample. Then

$$\sqrt{n_B}(F_B - \widehat{F}_A) = \sqrt{n_B}(F_B - F) - \sqrt{\frac{n_B}{n_A}}\sqrt{n_A}(\widehat{F}_A - F_{h_A}) - \sqrt{n_B}(F_{h_A} - F).$$

The first term converges in $\ell^\infty(I)$ to $B_1 \circ F$ by the empirical-process Donsker theorem. The third is $o(1)$ uniformly by Lemma 2 and $\sqrt{n_B}h_A^2 \to 0$.

For the training term,

$$\widehat{F}_A(x) - F_{h_A}(x) = (P_A - P)\Phi((x - \cdot)/h_A).$$

The smoothed indicator differs in $L^2(P)$ from $\mathbf{1}\{\cdot \le x\}$ by $O(h_A^{1/2})$ uniformly on I, by Lemma 3. The joint class

$$\{u \mapsto \Phi((x-u)/h) : x \in I, h > 0\} \cup \{u \mapsto \mathbf{1}\{u \le x\} : x \in I\}$$

is a bounded VC-subgraph class: after the monotone transformation $z = \Phi^{-1}(t)$, its subgraphs are half-spaces of the form $u + hz \le x$, with the indicators obtained as $h \downarrow 0$. It is therefore P-Donsker. Asymptotic equicontinuity in its $L^2(P)$ metric yields

$$\sqrt{n_A}(\widehat{F}_A - F_{h_A}) = \sqrt{n_A}(F_A - F) + o_p(1)$$

uniformly on I. It converges to $B_2 \circ F$. The two bridges are independent because the split samples are independent. Since $n_B/n_A \to \lambda$, the process limit follows.

Its covariance is

$$(1+\lambda)\{F(x \wedge y) - F(x)F(y)\},$$

the covariance of $\sqrt{1+\lambda}B(F(x))$. The continuous mapping theorem for the supremum proves the final statement. □

Proof of Proposition 7. Let $F^{(n)}$ denote the CDF in the triangular array and let $F_{h_A}^{(n)}$ be its Gaussian smoothing. The decomposition in the proof of Theorem 4, now centered at $F^{(n)}$, gives

$$\begin{aligned}\sqrt{n_B}(F_B - \widehat{F}_A) = \sqrt{n_B}(F_B - F^{(n)}) - \sqrt{\frac{n_B}{n_A}}\sqrt{n_A}(\widehat{F}_A - F_{h_A}^{(n)}) \\ + \sqrt{n_B}(F^{(n)} - F_{h_A}^{(n)}).\end{aligned} \tag{24}$$

The first two terms are jointly tight in $\ell^\infty(I)$. Whenever $F^{(n)} \to F_0$ uniformly, the same VC-Donsker and asymptotic-equicontinuity argument used for Theorem 4 makes their sum converge to the mixed bridge indexed by F_0. The pooled empirical CDF consistently estimates the endpoints of $F^{(n)}(I)$, so continuity of Gaussian bridge quantiles makes the plug-in critical value converge to the corresponding finite mixed-bridge quantile.

For the atom array, write $J_{x_0}(x) = \mathbf{1}\{x_0 \le x\}$ and let $J_{x_0,h}$ be its Gaussian smoothing. Then

$$F^{(n)} - F_{h_A}^{(n)} = (1-m_n)(F_0 - F_{0,h_A}) + m_n(J_{x_0} - J_{x_0,h_A}).$$

The first term is $O(h_A^2)$ uniformly. Because J_{x_0} has a unit jump and J_{x_0,h_A} is continuous,

$$\frac{1}{2} \le \|J_{x_0} - J_{x_0,h_A}\|_I \le 1$$

for all sufficiently small h_A. Hence

$$\frac{m_n}{2} - Ch_A^2 \le \|F^{(n)} - F_{h_A}^{(n)}\|_I \le m_n + Ch_A^2. \tag{25}$$

If $m_n\sqrt{n_B} \to \infty$, (24)–(25), tightness of the empirical terms, and boundedness in probability of the plug-in critical value imply power one. If $m_n\sqrt{n_B} \to 0$, the deterministic term in (24) is $o(1)$ because $\sqrt{n_B}h_A^2 \to 0$, while $F^{(n)} \to F_0$ uniformly. The statistic and critical value therefore have the same joint limit as under F_0, giving rejection probability α.

For a density jump, Proposition 6 gives

$$\|F^{(n)} - F_{h_A}^{(n)}\|_I \ge ch_A|\Delta_n| - Ch_A^2 \tag{26}$$

for all large n once $h_A = o(|\Delta_n|)$. Thus $h_A|\Delta_n|\sqrt{n_B} \to \infty$ and $h_A^2\sqrt{n_B} \to 0$ imply that the deterministic signal in (24) dominates the tight empirical terms, proving power one. For the low-signal direction, the proposition assumes that the array has no other departure from the smooth limit. Under that condition,

$$\|F^{(n)} - F_{h_A}^{(n)}\|_I = O\{h_A|\Delta_n| + h_A^2\},$$

so $h_A|\Delta_n|\sqrt{n_B} \to 0$ makes the deterministic term $o(1)$ and again yields the mixed-bridge null limit. This proves both claims.

Finally, for fixed split shares and $h_A \asymp h_n$, the ratio of the holdout atom boundary to the centered atom boundary is

$$\frac{n_B^{-1/2}}{\sqrt{h_n\ell_n/n}} \asymp \frac{1}{\sqrt{h_n\ell_n}},$$

and the same ratio follows for the two density-jump boundaries. Since $h_n\ell_n \to 0$, it diverges. □

C.3 Growing-class approximation for the centered test

For the Gaussian kernel, write

$$k_0(v) = \mathbf{1}\{v \ge 0\} - \Phi(v), \qquad f_{x,n}(u) = h_n^{-1/2}k_0\left(\frac{x-u}{h_n}\right),$$

and define

$$\mathcal{F}_n = \{f_{x,n} : x \in I\} \cup \{-f_{x,n} : x \in I\}.$$

Then the centered observed process satisfies

$$S_n := \sqrt{\frac{n}{h_n}} \sup_{x \in I} |D_n(x) - \mathbb{E}D_n(x)| = \sup_{g \in \mathcal{F}_n} \mathbb{G}_n g.$$

Let $\widehat{Q}_n$ denote the fitted KDE law with density $\widehat{f}_{h_n}$. Conditional on the sample, the centered bootstrap statistic has the identical representation with the bootstrap empirical process centered under $\widehat{Q}_n$.

For independent Gaussian multipliers $\xi_1, \ldots, \xi_n$, define

$$\mathbb{G}_n^\xi g = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i - \bar{\xi}) g(X_i), \qquad S_n^\xi = \sup_{g \in \mathcal{F}_n} \mathbb{G}_n^\xi g.$$

Conditional on the observations, $\mathbb{G}_n^\xi$ is exactly a centered Gaussian process whose covariance is the empirical covariance of the local contrasts.

Lemma 4 (Entropy, moments, and Gaussian coupling)**.** *Under Assumption 1, $\mathcal{F}_n$ is VC-type with constants independent of n. It has envelope $b_n \lesssim h_n^{-1/2}$ and, for every fixed $r \geq 2$,*

$$\sup_{g \in \mathcal{F}_n} P|g|^r \lesssim h_n^{1-r/2}.$$

Its variances are bounded above and away from zero. For any finite

$$q > \max\left\{4, \frac{2}{1-\beta}\right\}, \qquad h_n = cn^{-\beta},$$

the Gaussian approximation error in Theorem 2.1 of Chernozhukov et al. (2016) tends to zero.

Proof of Lemma 4. For the Gaussian kernel, k_0 is bounded, of bounded variation, square integrable, and has exponentially decaying tails. The bounded-variation translate result in van der Vaart and Wellner (1996, Section 2.6.3) makes the class VC-type; multiplication by $h_n^{-1/2}$ changes the envelope but not the entropy exponent. Pointwise measurability follows by restricting x to a countable dense subset of I and using the one-sided limits of k_0. Thus Conditions (A) and (B) of Theorem 2.1 in Chernozhukov et al. (2016) hold. The moment bound follows from

$$P|f_{x,n}|^r = h_n^{1-r/2} \int |k_0(v)|^r f(x - h_n v)\, dv.$$

The variance statement follows from Lemma 3. Let B_n denote the constant envelope b_n of

$\mathcal{F}_n$. Taking a fixed σ^2 above the variance bound, the preceding display gives

$$\sup_{g\in\mathcal{F}_n} P|g|^k \le \sigma^2 b_n^{k-2}, \qquad k=2,3,4, \qquad \|\mathsf{B}_n\|_{P,q} = b_n,$$

after increasing the fixed envelope constant if necessary. This is Condition (C) of that theorem.

Because the deterministic functional is zero, its covering number is one for every approximation tolerance. In the notation of Chernozhukov et al. (2016), the remaining logarithmic complexity is therefore $K_n = O(\log n)$ and $K_n^3 \le n$ eventually. With $b_n \asymp h_n^{-1/2}$ and $\sigma \asymp 1$, the two terms in their equation (7), before the tuning factors, are

$$a_n = \frac{K_n}{h_n^{1/2} n^{1/2-1/q}}, \qquad d_n = \frac{K_n^{2/3}}{(nh_n)^{1/6}}.$$

Equivalently,

$$a_n \to 0 \quad \Longleftarrow \quad \frac{n^{1-2/q}h_n}{K_n^2} \to \infty, \qquad d_n \to 0 \quad \Longleftarrow \quad \frac{nh_n}{K_n^4} \to \infty.$$

For $h_n = cn^{-\beta}$, the second condition follows from $\beta < 1$, and the first follows from $1-2/q-\beta > 0$, which is exactly the stated choice $q > 2/(1-\beta)$.

The theorem also multiplies these terms by $\gamma_n^{-1/q}$ and $\gamma_n^{-1/3}$ and bounds the coupling failure probability by a constant times $\gamma_n + n^{-1}$. Set

$$\gamma_n = \max\{a_n^{q/2}, d_n^{3/2}\}.$$

Then $\gamma_n \to 0$, $a_n\gamma_n^{-1/q} \le a_n^{1/2} \to 0$, and $d_n\gamma_n^{-1/3} \le d_n^{1/2} \to 0$. Since the process is centered, the theorem's covering number for the deterministic functional is one; its approximation tolerance may therefore be taken to zero independently. This verifies both the coupling error and its failure probability. □

Lemma 5 (Fitted-density and covariance consistency)**.** *Under Assumption 1,*

$$\|\widehat{f}_{h_n} - f\|_\infty = O_p\left(\sqrt{\frac{\log n}{nh_n}} + h_n^2\right) = o_p((\log n)^{-2}).$$

If

$$\Sigma_n(x,y) = \mathrm{Cov}_P(f_{x,n}(X), f_{y,n}(X)), \quad \widehat{\Sigma}_n(x,y) = \mathrm{Cov}_{\widehat{Q}_n}(f_{x,n}(X), f_{y,n}(X)),$$

then

$$\sup_{x,y\in I} |\widehat{\Sigma}_n(x,y) - \Sigma_n(x,y)| = o_p((\log n)^{-2}).$$

The empirical covariance kernel

$$\widetilde{\Sigma}_n(x,y) = \mathbb{P}_n\{(f_{x,n} - \mathbb{P}_n f_{x,n})(f_{y,n} - \mathbb{P}_n f_{y,n})\}$$

also satisfies

$$\sup_{x,y\in I} |\widetilde{\Sigma}_n(x,y) - \Sigma_n(x,y)| = o_p((\log n)^{-2}).$$

With probability tending to one, the fitted density is bounded above and away from zero on every compact V_0 satisfying $I \subset V_0 \Subset V$.

Proof. The first display follows from the VC-type uniform KDE bound of Gine and Guillou (2002) plus the second-order density bias; the bounded uniformly continuous second derivative makes the bias $O(h_n^2)$. Both terms are polynomially small for $\beta \in (0,1)$ and hence $o((\log n)^{-2})$.

For the covariance products,

$$\left| \int f_{x,n} f_{y,n} (\widehat{f}_{h_n} - f) \right| \leq \|\widehat{f}_{h_n} - f\|_\infty \int |f_{x,n} f_{y,n}|.$$

The final integral is uniformly $O(1)$ by change of variables and Cauchy–Schwarz. Likewise $\int |f_{x,n}| = O(h_n^{1/2})$. Expanding the covariance into second moments and products of first moments proves the uniform fitted-covariance bound.

For the empirical covariance, the product class $\mathcal{F}_n\mathcal{F}_n$ is VC-type, has envelope $O(h_n^{-1})$, and satisfies

$$\sup_{g_1,g_2\in\mathcal{F}_n} P(g_1 g_2)^2 \lesssim h_n^{-1}.$$

The VC maximal inequality and Bernstein's inequality therefore give

$$\sup_{g_1,g_2\in\mathcal{F}_n} |(\mathbb{P}_n - P)(g_1 g_2)| = O_p\left(\sqrt{\frac{\log n}{n h_n}} + \frac{\log n}{n h_n}\right) = o_p((\log n)^{-2}).$$

For the first moments, the same maximal inequality applied to $\mathcal{F}_n$ gives

$$\sup_{g\in\mathcal{F}_n} |(\mathbb{P}_n - P)g| = O_p\left(\sqrt{\frac{\log n}{n}} + \frac{\log n}{n\sqrt{h_n}}\right).$$

Also $\sup_g |Pg| = O(\sqrt{h_n})$. Consequently,

$$\sup_{g_1,g_2} |\mathbb{P}_n g_1 \mathbb{P}_n g_2 - P g_1 P g_2| = o_p((\log n)^{-2}).$$

Expanding the empirical and population covariance kernels proves the claim for $\widetilde{\Sigma}_n$. Finally, uniform consistency and $f \geq c_f$ on V imply $\widehat{f}_{h_n} \geq c_f/2$ on V_0 with probability tending to one; uniform consistency also supplies the upper bound. □

Lemma 6 (Gaussian comparison and anti-concentration)**.** *Let G_n be the centered Gaussian process on $\mathcal{F}_n$ with covariance Σ_n. Let $\overline{G}_n$ have covariance $\overline{\Sigma}_n$, where $\overline{\Sigma}_n$ is either the fitted-law covariance $\widehat{\Sigma}_n$ or the empirical covariance $\widetilde{\Sigma}_n$. Under Assumption 1,*

$$\sup_t \left| \mathbb{P}\left(\sup_{g \in \mathcal{F}_n} G_n g \leq t \right) - \mathbb{P}^\circ \left(\sup_{g \in \mathcal{F}_n} \overline{G}_n g \leq t \right) \right| \to_p 0.$$

Moreover,

$$\mathbb{E} \sup_{g \in \mathcal{F}_n} G_n g = O(\sqrt{\log(1/h_n)}),$$

and the same conditional bound holds for $\overline{G}_n$ in probability. Here $\mathbb{P}^\circ$ denotes its conditional law. For any ε_n with $\varepsilon_n \sqrt{\log(1/h_n)} \to 0$, both Gaussian suprema have anti-concentration modulus $o(1)$ over intervals of length ε_n.

Proof. For the canonical metric ρ_n, bounded variation of k_0 gives

$$\rho_n(x,y)^2 \leq C \min\{1, |x-y|/h_n\}.$$

Hence

$$\log N(\eta, I, \rho_n) \leq C\{\log(1/h_n) + \log(1/\eta)\}.$$

Dudley's entropy integral yields the expected-supremum bound. For either admissible $\overline{\Sigma}_n$, Lemma 5 gives covariance error $\delta_{\Sigma,n} = o_p((\log n)^{-2})$.

We make the passage from a finite net to I explicit. Put $\delta_n = (\log n)^{-3}$ and take a Euclidean grid with spacing at most $c h_n \delta_n^2$. The metric bound above makes it a δ_n-net for ρ_n, and its cardinality satisfies

$$p_n \leq C(h_n \delta_n^2)^{-1}, \qquad \log p_n = O(\log n).$$

For every $\delta > 0$, Dudley's inequality applied to the increment class gives

$$\mathbb{E} \sup_{\rho_n(x,y)\le\delta} |G_n(x) - G_n(y)| \le C \int_0^\delta \sqrt{\log\{C/(h_n u^2)\}}\, du.$$

At $\delta = \delta_n$ this is $o(1)$, and Borell's inequality makes the corresponding modulus $o_p(1)$. On the event in Lemma 5, the conditional canonical metric $\bar{\rho}_n$ obeys

$$\bar{\rho}_n(x,y)^2 \le \rho_n(x,y)^2 + 4\delta_{\Sigma,n}.$$

The same entropy integral, now at radius $(\delta_n^2 + 4\delta_{\Sigma,n})^{1/2} = o_p((\log n)^{-1})$, makes the conditional modulus $o_p(1)$ in outer probability.

The Gaussian comparison inequality of Chernozhukov et al. (2014) bounds the Kolmogorov distance between the maxima on this common net by

$$C\delta_{\Sigma,n}^{1/3}\{1 \vee \log(p_n/\delta_{\Sigma,n})\}^{2/3} = o_p(1),$$

because $\log p_n = O(\log n)$ and $\delta_{\Sigma,n} = o_p((\log n)^{-2})$. The two modulus bounds remove the discretization by the triangle inequality. Finally, the Gaussian anti-concentration inequality in the same paper bounds concentration over an interval of length ε by a constant times $\varepsilon(\mathbb{E}\sup G_n + 1)$. The conditional expected-supremum bound and the high-probability variance bounds from Lemma 5 give the identical conclusion for $\overline{G}_n$. □

Lemma 7 (Centered bootstrap and multiplier approximation)**.** *Under Assumption 1,*

$$\sup_t |\mathbb{P}(S_n \le t) - \mathbb{P}^*(S_n^{*,c} \le t)| \to_p 0.$$

Moreover,

$$\sup_t |\mathbb{P}(S_n \le t) - \mathbb{P}^\xi(S_n^\xi \le t)| \to_p 0,$$

where $\mathbb{P}^\xi$ denotes probability over the Gaussian multipliers conditional on the observations. If $\varepsilon_n = (nh_n^3)^{1/4}$, all three distributions have anti-concentration modulus $o(1)$ over intervals of length ε_n.

Proof. Lemma 4 and Theorem 2.1 of Chernozhukov et al. (2016) couple S_n to $\sup_{g\in\mathcal{F}_n} G_n g$ with error $o(1)$.

For the conditional coupling, fix V_0 with $I \subset V_0 \Subset V$ and let $\mathcal{E}_n$ be the event on which $\widehat{f}_{h_n}$ is bounded above on $\mathbb{R}$, bounded above and away from zero on V_0, and the fitted covariance error in Lemma 5 is at most a deterministic sequence $\bar{d}_n = o((\log n)^{-2})$. Lemma 5 implies

$\mathbb{P}(\mathcal{E}_n) \to 1$. Conditional on a sample in $\mathcal{E}_n$, the bootstrap observations are i.i.d. from $\widehat{Q}_n$. The VC entropy bound is distribution-free, and a change of variables gives, for every fixed $r \geq 2$,

$$\sup_{g \in \mathcal{F}_n} \widehat{Q}_n |g|^r \leq C h_n^{1-r/2} \int |k_0(v)|^r \, dv.$$

The fitted covariance lower bound supplies a fixed positive variance lower bound, while the envelope remains $Ch_n^{-1/2}$. Thus every conditional entropy, moment, and envelope bound used in Lemma 4 holds on $\mathcal{E}_n$ with fixed constants. Applying Theorem 2.1 of Chernozhukov et al. (2016) under the conditional law yields the same deterministic coupling rates a_n, d_n and a conditional failure probability tending to zero. Consequently,

$$\mathbb{P}^* \left(\left| S_n^{*,c} - \sup_{g \in \mathcal{F}_n} G_n^* g \right| > \eta_n \right) \to_p 0$$

for some $\eta_n \downarrow 0$, where G_n^* has covariance $\widehat{\Sigma}_n$. Lemma 6 compares the two Gaussian suprema, and the triangle inequality proves the first Kolmogorov approximation.

For the multiplier statistic, no second empirical-process coupling is needed. Conditional on the observations, S_n^ξ is the supremum of an exactly Gaussian process with covariance $\widetilde{\Sigma}_n$. The empirical-covariance part of Lemma 5, followed by the Gaussian comparison and anti-concentration argument in Lemma 6, compares this conditional law directly with $\sup_{g \in \mathcal{F}_n} G_n g$. Combining that comparison with the CCK coupling for S_n proves the multiplier display.

Because $h_n = cn^{-\beta}$ with $\beta > 1/3$,

$$n h_n^3 \log^2(1/h_n) \to 0.$$

Therefore $\varepsilon_n \sqrt{\log(1/h_n)} \to 0$, and Lemma 6 plus the distributional approximations transfers anti-concentration to S_n, $S_n^{*,c}$, and S_n^ξ. □

Lemma 8 (Distributional transfer to feasible p-values)**.** *Let U_n be an observed statistic and V_n^* a bootstrap statistic. Suppose*

$$\sup_t |\mathbb{P}(U_n \leq t) - \mathbb{P}^*(V_n^* \leq t)| \to_p 0$$

and the largest atom of the law of U_n tends to zero. Then

$$\mathbb{P}^*(V_n^* \geq U_n) \rightsquigarrow \text{Uniform}(0,1).$$

Proof. Let Γ_n and Γ_n^* be the unconditional and conditional CDFs. The first assumption gives

$$1 - \Gamma_n^*(U_n-) = 1 - \Gamma_n(U_n-) + o_p(1).$$

If $j_n = \sup_t \mathbb{P}(U_n = t)$, the generalized probability-integral-transform inequality gives

$$\sup_{a\in[0,1]} |\mathbb{P}\{1 - \Gamma_n(U_n-) \le a\} - a| \le j_n.$$

The right side tends to zero, and Slutsky's theorem completes the proof. □

C.4 Size, calibration, and critical values

Size part of Theorem 3. By Lemma 2,

$$|U_n - S_n| \le \sqrt{\frac{n}{h_n}} \sup_{x\in I} |F(x) - F_{h_n}(x)| = O(\sqrt{nh_n^3}).$$

With $\varepsilon_n = (nh_n^3)^{1/4}$, the right side is $o(\varepsilon_n)$. Lemma 7 gives the conditional distributional approximation and anti-concentration on the ε_n scale. Hence

$$\sup_t |\mathbb{P}(U_n \le t) - \mathbb{P}^*(S_n^{*,c} \le t)| \to_p 0.$$

The largest atom of the distribution of U_n tends to zero by the same anti-concentration bound. Lemma 8 then yields

$$p_n^{*,c} = \mathbb{P}^*(S_n^{*,c} \ge U_n) \rightsquigarrow \text{Uniform}(0,1).$$

This proves asymptotic size. □

Lemma 9 (Reference critical values under the power arrays)**.** *Let $c_{n,1-\alpha}^{*,c}$ and $c_{n,1-\alpha}^{\xi}$ denote fixed upper conditional quantiles of the centered smooth-bootstrap and Gaussian-multiplier statistics. Under the atom array in Theorem 3,*

$$c_{n,1-\alpha}^{*,c} \vee c_{n,1-\alpha}^{\xi} = O_p\left(\sqrt{\left(1 + \frac{m_n}{h_n}\right)\ell_n} + \frac{\ell_n}{\sqrt{nh_n}}\right). \tag{27}$$

Under the uniformly bounded density-jump arrays, both critical values are $O_p(\sqrt{\ell_n})$.

Proof. Let M_n be the empirical count at the atom. Decompose the empirical measure into its smooth and atomic observations. The VC maximal inequality for the bounded-density

observations and direct evaluation of the atomic contribution give

$$\sup_{g\in\mathcal{F}_n} \mathbb{P}_n g^2 = O_p\left(1 + \frac{M_n}{nh_n}\right). \tag{28}$$

For the smooth bootstrap, its fitted density is the sum of a bounded-density KDE component and the Gaussian bump generated by the M_n repeated observations. Uniform KDE bounds for the first component and $\sup_u h_n^{-1}\phi((u-x_0)/h_n) \le C/h_n$ for the second imply

$$\|\widehat{f}_{h_n}\|_\infty = O_p\left(1 + \frac{M_n}{nh_n}\right).$$

Since $\int f_{x,n}^2 \le C$ uniformly in x,

$$\sup_{g\in\mathcal{F}_n} \widehat{Q}_n g^2 = O_p\left(1 + \frac{M_n}{nh_n}\right). \tag{29}$$

The conditional VC maximal inequality, applied under $\mathbb{P}_n$ for the multiplier covariance and under $\widehat{Q}_n$ for the centered bootstrap empirical process, bounds the corresponding expected supremum by

$$C\left\{\sqrt{\sigma_n^2 \ell_n} + \frac{b_n \ell_n}{\sqrt{n}}\right\}, \qquad \sigma_n^2 = O_p\left(1 + \frac{M_n}{nh_n}\right), \qquad b_n \le C h_n^{-1/2}.$$

For the multiplier process this is also an immediate conditional Dudley bound, followed by Borell's inequality. Markov's inequality for a fixed upper bootstrap quantile, or the conditional concentration form of the same maximal inequality, gives the identical order for $c^{*,c}_{n,1-\alpha}$. Finally,

$$\frac{M_n}{n} = O_p(m_n + n^{-1}),$$

which proves (27). Under a uniformly bounded density array, (28)–(29) hold with $\sigma_n^2 = O_p(1)$ and the displayed order reduces to $O_p(\sqrt{\ell_n})$, because $nh_n/\ell_n \to \infty$. □

Proof of Corollary 1. The feasible-bias comparison in the preceding proof gives $|U_n - S_n| = o(\varepsilon_n)$. Use the multiplier display in Lemma 7 in place of its smooth-bootstrap display. The same anti-concentration argument and Lemma 8 yield asymptotic size.

For power, Lemma 9 gives the required multiplier critical-value orders. The signal comparisons in the atom and density-jump proofs are therefore unchanged. □

The Gaussian centering formula used in Procedure 1 follows directly. Conditional on the

data, X^* has CDF $\widehat{F}_h$. Thus

$$\mathbb{E}^* F_n^*(x) = \widehat{F}_h(x),$$

and convolution of two $N(0, h^2)$ kernels gives

$$\mathbb{E}^* \widehat{F}_h^*(x) = \widehat{F}_{\sqrt{2}h}(x).$$

Subtracting yields the displayed expression.

C.5 Power proofs

The power statements in Theorem 3 have the following precise triangular-array interpretation. For atoms, let

$$P_n = (1 - m_n)P_0 + m_n \delta_{x_0},$$

where P_0 has a bounded density and $x_0 \in \mathrm{int}(I)$. For density jumps, let P_n have uniformly bounded densities with one-sided expansions

$$f_n(x_0 + t) = f_{-,n} + O(|t|), \quad t < 0, \qquad f_n(x_0 + t) = f_{+,n} + O(|t|), \quad t > 0,$$

uniformly in n, and set $\Delta_n = f_{+,n} - f_{-,n}$. The proofs below establish power one for these arrays. Fixed atoms and fixed jumps are immediate special cases.

Atom part of Theorem 3. Let $x_0 \in \mathrm{int}(I)$ carry mass m_n and let M_n be the number of observations equal to x_0. Under the stated signal condition, $nm_n/\ell_n \to \infty$, so $M_n/(nm_n) \to_p 1$. Since the fitted KDE-CDF is continuous, Proposition 5 applied to F_n gives

$$U_n \geq \frac{M_n}{2n}\sqrt{\frac{n}{h_n}} \geq \frac{m_n}{3}\sqrt{\frac{n}{h_n}}$$

with probability tending to one.

Lemma 9 gives, for either smooth resampling or multiplier calibration,

$$c^*_{n,1-\alpha} = O_p\left(\sqrt{\left(1 + \frac{m_n}{h_n}\right)\ell_n} + \frac{\ell_n}{\sqrt{nh_n}}\right).$$

Relative to $A_n = m_n\sqrt{n/h_n}$, the three terms are bounded by

$$\frac{\sqrt{\ell_n}}{A_n} \to 0, \qquad \frac{\sqrt{(m_n/h_n)\ell_n}}{A_n} = \sqrt{\frac{\ell_n}{nm_n}} \to 0, \qquad \frac{\ell_n}{nm_n} \to 0.$$

Thus the critical value is $o_p(A_n)$ while the observed statistic exceeds $A_n/3$, proving power one. Fixed atoms are the special case $m_n \equiv m > 0$. □

Density-jump part of Theorem 3. Proposition 6 gives population signal at least $ch_n|\Delta_n|$ once $h_n = o(|\Delta_n|)$. This last relation follows from the signal condition and $nh_n^3/\ell_n \to 0$. After normalization, the signal is at least

$$c|\Delta_n|\sqrt{nh_n}.$$

The observed centered empirical process is $O_p(\sqrt{\ell_n})$ by the VC maximal inequality, and Lemma 9 gives the same order for both reference critical values. Since $|\Delta_n|\sqrt{nh_n/\ell_n} \to \infty$, the deterministic signal dominates both, proving power one. Fixed nonzero jumps satisfy the condition because $nh_n/\ell_n \to \infty$. □

C.6 Minimax jump detection

This subsection proves Theorem 5 and Corollary 3. The minimax claim concerns the unknown location of a density jump. It is distinct from general goodness-of-fit separation in an L_p norm and from estimation of the jump location.

Lemma 10 (Hölder bias and jump drift)**.** *Let $0 < s \le 1$, let $Z \sim N(0,1)$, and set*

$$B_{s,\phi} = \frac{\mathbb{E}|Z|^{s+1}}{s+1}, \qquad \mu_\phi = \mathbb{E}\{Z\mathbf{1}(Z > 0)\}.$$

If f is globally Hölder-s with constant L, then

$$\sup_{x\in\mathbb{R}} |F_h(x) - F(x)| \le B_{s,\phi} L h^{s+1}. \tag{30}$$

If instead f has a jump $\Delta = f(x_0+) - f(x_0-)$ and is Hölder-s with constant L separately on the two sides of x_0, then

$$|F_h(x_0) - F(x_0) - \mu_\phi h\Delta| \le B_{s,\phi} L h^{s+1}. \tag{31}$$

Proof. For a continuity point x, write

$$R_x(t) = F(x-t) - F(x) + tf(x).$$

For $t > 0$,

$$R_x(t) = -\int_0^t \{f(x-u) - f(x)\}\, du,$$

and the analogous identity holds for $t < 0$. Consequently,

$$|R_x(t)| \le \frac{L}{s+1}|t|^{s+1}.$$

Since $F_h(x) = \mathbb{E}F(x - hZ)$ and $\mathbb{E}Z = 0$, taking expectations proves (30).

At x_0, use $f_- = f(x_0-)$ for $Z > 0$ and $f_+ = f(x_0+)$ for $Z < 0$. The one-sided Hölder restrictions give

$$F(x_0 - hZ) - F(x_0) = -hZf_- + R_-(hZ), \quad Z > 0,$$

and

$$F(x_0 - hZ) - F(x_0) = -hZf_+ + R_+(hZ), \quad Z < 0,$$

where each remainder is bounded in absolute value by $Lh^{s+1}|Z|^{s+1}/(s+1)$. Symmetry gives

$$-hf_-\mathbb{E}\{Z\mathbf{1}(Z > 0)\} - hf_+\mathbb{E}\{Z\mathbf{1}(Z < 0)\} = \mu_\phi h(f_+ - f_-).$$

Taking expectations and bounding the remainders proves (31). □

For a bandwidth h, let

$$S_n(h) = \sqrt{\frac{n}{h}} \sup_{x \in I} |(\mathbb{P}_n - P)\psi_{x,h}|,$$

and let $S_n^\xi(h)$ be its Gaussian-multiplier analogue in (20).

Lemma 11 (Uniform multiplier calibration at the rate-balancing bandwidth)**.** *Fix $s \in (0, 1]$ and $L > 0$, and let*

$$h_n = \left(\frac{C_f \log n}{nL^2}\right)^{1/(2s+1)}.$$

For every $\varepsilon > 0$,

$$\sup_{f \in \mathcal{H}_s(L)} \mathbb{P}_f\left\{\sup_{t \in \mathbb{R}} \left|\mathbb{P}_f\{S_n(h_n) \le t\} - \mathbb{P}_f^\xi\{S_n^\xi(h_n) \le t\}\right| > \varepsilon\right\} \to 0. \tag{32}$$

Moreover, there is a finite constant C_0 such that, uniformly over all bounded densities in $\mathcal{H}_s(L) \cup \mathcal{J}_s(\Delta, L)$,

$$\mathbb{P}_f\left\{q_{n,1-\alpha}^\xi(h_n) > C_0\sqrt{\log(1/h_n)}\right\} \to 0. \tag{33}$$

Proof. Write

$$f_{x,n}(u) = h_n^{-1/2}k_0((x-u)/h_n), \qquad k_0(v) = \mathbf{1}\{v \ge 0\} - \Phi(v).$$

As in Lemma 4, the signed translate class is VC-type with constants independent of n, envelope $b_n \lesssim h_n^{-1/2}$, and moments

$$\sup_x P_f |f_{x,n}|^r \le C_r h_n^{1-r/2}, \qquad 2 \le r < \infty,$$

uniformly over the stated density classes. Its variances are bounded above uniformly and, over $\mathcal{H}_s(L)$, bounded away from zero because $f \ge c_f$ on a neighborhood of I.

Set $\beta_s = 1/(2s+1) < 1$. The bandwidth is $n^{-\beta_s}$ times a fixed power of $\log n$. Choose a finite $q > \max\{4, 2/(1-\beta_s)\}$. With $K_n = O(\log n)$, the two coupling terms from Theorem 2.1 of Chernozhukov et al. (2016) are, up to fixed powers of the theorem's slowly vanishing tuning sequence,

$$\frac{h_n^{-1/2} K_n}{n^{1/2-1/q}} \quad \text{and} \quad \frac{(h_n^{-1/2} K_n^2)^{1/3}}{n^{1/6}}.$$

The first is a logarithmic factor times $n^{-(1-\beta_s)/2+1/q} = o(1)$ by the choice of q; the second is a logarithmic factor times $(nh_n)^{-1/6} = o(1)$. Choosing the tuning sequence as in the proof of Lemma 4 makes both the coupling error and its failure probability tend to zero uniformly.

For the multiplier process, the product class has envelope $O(h_n^{-1})$ and second moments $O(h_n^{-1})$. The same VC maximal inequality used in Lemma 5 gives

$$\sup_{x,y \in I} |\widetilde{\Sigma}_n(x,y) - \Sigma_n(x,y)| = O_p\left(\sqrt{\frac{\log n}{nh_n}} + \frac{\log n}{nh_n} \right) = o_p((\log n)^{-2})$$

uniformly over $\mathcal{H}_s(L)$. The canonical metric obeys $\rho_n(x,y)^2 \le C \min\{1, |x-y|/h_n\}$. The same finite-net Gaussian comparison, entropy-modulus, and anti-concentration argument used in Lemma 6 proves (32).

Finally, on an event whose probability tends to one uniformly over every bounded density in the displayed union, conditional Dudley integration gives $\mathbb{E}^\xi S_n^\xi(h_n) \le C\sqrt{\log(1/h_n)}$, and the conditional variances are bounded by a fixed constant. Borell's inequality then supplies a fixed C_0 for which (33) holds. □

Proof of Theorem 5. We first prove the lower bound. Choose a closed interval $J \subset \operatorname{int}(I_0)$ and a smooth density $f_0 \in \mathcal{H}_s(L)$ that equals a constant $a > 0$ on a neighborhood of J and whose Hölder seminorm is at most $L/2$. Such a density exists under the class interiority condition stated in the main text.

Define the compactly supported odd function

$$q_s(t) = \frac{1}{2} \operatorname{sgn}(t)(1 - |t|^s) \mathbf{1}\{0 < |t| \le 1\}.$$

It integrates to zero, has jump one at zero, and has a finite one-sided Hölder-s seminorm C_s. For a proposed jump height Δ, take

$$w = \left(\frac{2C_s \Delta}{L} \right)^{1/s}$$

and choose $M \geq c_J / w$ points $x_1, \ldots, x_M$ in J whose intervals $[x_j - w, x_j + w]$ are disjoint. Set

$$g_j(x) = \Delta q_s((x - x_j)/w), \qquad f_j = f_0 + g_j.$$

For all sufficiently small Δ, every f_j is nonnegative, integrates to one, obeys the fixed upper and lower density bounds, belongs to $\mathcal{J}_s(\Delta, L)$, and has its sole jump inside I_0. The factor two in the definition of w leaves room for the strict Hölder seminorm of f_0.

Let P_j have density f_j, let P_0 have density f_0, and let $L_j = dP_j^n / dP_0^n$. Since $f_0 = a$ on every perturbation support and the supports are disjoint,

$$\int \frac{g_j g_k}{f_0} = 0 \quad (j \neq k), \qquad \int \frac{g_j^2}{f_0} = C_q \Delta^2 w$$

for a fixed $C_q > 0$. Therefore

$$\mathbb{E}_0(L_j L_k) = \left(1 + \int \frac{g_j g_k}{f_0} \right)^n$$

equals one for $j \neq k$ and is at most $\exp(C_q n \Delta^2 w)$ for $j = k$. For the uniform mixture $\overline{P}_n = M^{-1} \sum_{j=1}^M P_j^n$,

$$\chi^2(\overline{P}_n, P_0^n) \leq \frac{\exp(C_q n \Delta^2 w) - 1}{M}. \tag{34}$$

Take $\Delta = c r_{n,s}(L, C_f)$. Then

$$n \Delta^2 w = C L^{-1/s} n \Delta^{2+1/s} = C' c^{2+1/s} \log n,$$

whereas

$$\log M = \frac{1 + o(1)}{2s + 1} \log n.$$

Choosing $c > 0$ sufficiently small makes the right side of (34) tend to zero. Total variation is bounded by one half the square root of chi-square. Hence, for any test with asymptotic size at most α over $\mathcal{H}_s(L)$,

$$\frac{1}{M} \sum_{j=1}^M \mathbb{E}_{f_j} \varphi_n \leq \mathbb{E}_{f_0} \varphi_n + \| \overline{P}_n - P_0^n \|_{\mathrm{TV}} \leq \alpha + o(1).$$

At least one alternative has power at most $\alpha + o(1)$, proving part (i).

For the upper bound, abbreviate $h = h^{\star}_{n,s}$ and let $q^{\xi}_{1-\alpha} = q^{\xi}_{n,1-\alpha}(h)$. Under every $f \in \mathcal{H}_s(L)$, Lemma 10 gives

$$U_n(h) \le S_n(h) + B_{s,\phi} L\sqrt{n}\, h^{s+1/2}.$$

Consequently,

$$\mathbb{P}_f\{\varphi_n^{\mathrm{BA}} = 1\} \le \mathbb{P}_f\{S_n(h) > q^{\xi}_{1-\alpha}\}.$$

Lemma 11 and the definition of the conditional quantile imply that the supremum of the right side over $\mathcal{H}_s(L)$ is at most $\alpha + o(1)$. This proves uniform size.

Now take $f \in \mathcal{J}_s(\Delta, L)$ with jump point x_0. Since the absolute jump is at least Δ, Lemma 10 gives

$$|P_f \psi_{x_0,h}| \ge \mu_\phi h\Delta - B_{s,\phi} L h^{s+1}.$$

The centered fluctuation at this fixed point is $O_p(1)$ after multiplication by $\sqrt{n/h}$, uniformly over the alternative class. Thus

$$U_n(h) \ge \mu_\phi \Delta\sqrt{nh} - B_{s,\phi} L\sqrt{n}\, h^{s+1/2} - O_p(1). \tag{35}$$

At the selected bandwidth,

$$L\sqrt{n}\, h^{s+1/2} = \sqrt{C_f \log n}, \qquad r_{n,s}(L, C_f)\sqrt{nh} = \sqrt{C_f \log n}. \tag{36}$$

By (33), the multiplier critical value is at most $C_0\sqrt{\log n}$ with probability tending to one, uniformly over the alternative class. Combining (35)–(36), the signal is a fixed multiple of $C\sqrt{\log n}$, while the multiplier quantile and both bias terms are bounded by fixed multiples of $\sqrt{\log n}$. Choosing C sufficiently large therefore makes the observed statistic exceed the bias-aware critical value with probability tending to one whenever $\Delta \ge C r_{n,s}(L, C_f)$. This proves part (ii). □

Proof of Corollary 3. The proof of Lemma 11 applies with $h_n = cn^{-\beta}$ for every fixed $\beta < 1$: choose $q > \max\{4, 2/(1-\beta)\}$ in its coupling argument. Under $f \in \mathcal{H}_1(L)$, Lemma 10 gives normalized bias at most $B_{1,\phi} L\sqrt{nh_n^3}$. Because $\beta > 1/3$,

$$\sqrt{nh_n^3 \log(1/h_n)} \to 0.$$

The Gaussian anti-concentration bound in Lemma 6 therefore transfers the centered multiplier quantile to the feasible uncentered statistic uniformly over $\mathcal{H}_1(L)$, proving asymptotic size α.

For fixed L and C_f, the minimax rate is proportional to $n^{-1/3}(\log n)^{1/3}$. With $h_n = cn^{-\beta}$, Theorem 3 requires a jump larger than

$$\sqrt{\frac{\ell_n}{nh_n}} \asymp n^{-(1-\beta)/2}(\log n)^{1/2}.$$

Their ratio is $n^{(\beta-1/3)/2}(\log n)^{1/6}$. Substituting the preferred empirical choice $\beta = 0.40 = 2/5$ gives $(\beta - 1/3)/2 = 1/30$. □

C.7 Pilot undersmoothing

Proof of Corollary 2. Put

$$h_{0n} = c_0 n^{-\beta}, \qquad \ell_n = \log n \vee \log(1/h_{0n}),$$

and

$$\rho_{h,n} = \left| \widehat{h}_{n,\beta}/h_{0n} - 1 \right|.$$

The assumed pilot expansion gives $\rho_{h,n} = O_p(r_n)$ and hence $\rho_{h,n}\ell_n^2 = o_p(1)$. We first control the statistic under this random perturbation.

For $h > 0$, write

$$f_{x,h}(u) = h^{-1/2} k_0((x-u)/h).$$

For a in a fixed neighborhood of one, a change of variables and the mean-value theorem on each half-line give

$$\sup_{x \in I} P\{f_{x,ah_{0n}} - f_{x,h_{0n}}\}^2 \le C(a-1)^2. \tag{37}$$

Indeed, after setting $t = (x-u)/h_{0n}$ the relevant difference is $a^{-1/2}k_0(t/a) - k_0(t)$. Scaling never moves t across zero, and its derivative in a is bounded by an $L^2(\mathbb{R})$ function proportional to $|k_0(t)| + |t|\phi(t)$. The same calculation supplies the fixed-r moment bounds needed below.

The class of these bandwidth differences, indexed by $x \in I$ and $|a-1| \le \rho$, is VC-type. Its $L^2(P)$ radius is $C\rho$ by (37), and its localized envelope has the same $h_{0n}^{-1/2}$ order as the original class, multiplied by $C\rho$. The local VC maximal inequality therefore yields, uniformly for deterministic $\rho \downarrow 0$,

$$\sup_{|a-1| \le \rho} \sup_{x \in I} |\mathbb{G}_n(f_{x,ah_{0n}} - f_{x,h_{0n}})| = O_p\left(\rho\sqrt{\ell_n} + \frac{\rho\ell_n}{\sqrt{nh_{0n}}}\right). \tag{38}$$

The uniform CDF-bias expansion also gives

$$\sup_{|a-1|\le\rho}\sup_{x\in I}\left|\sqrt{\frac{n}{ah_{0n}}}\{F(x)-F_{ah_{0n}}(x)\}-\sqrt{\frac{n}{h_{0n}}}\{F(x)-F_{h_{0n}}(x)\}\right|\le C\rho\sqrt{nh_{0n}^3}+o(\sqrt{nh_{0n}^3}). \tag{39}$$

Since $nh_{0n}/\ell_n^2\to\infty$, $nh_{0n}^3\ell_n^2\to 0$, and $\rho_{h,n}\ell_n^2=o_p(1)$, (38)–(39) imply

$$|U_n(\widehat{h}_{n,\beta})-U_n(h_{0n})|=o_p(\ell_n^{-1/2}). \tag{40}$$

Conditional on the data, the multiplier version of the same local maximal inequality gives

$$\sup_{|a-1|\le\rho_{h,n}}\sup_{x\in I}\left|\mathbb{G}_n^{\xi}(f_{x,ah_{0n}}-f_{x,h_{0n}})\right|=o_p(\ell_n^{-1/2}) \tag{41}$$

in conditional probability. For smooth resampling, apply the conditional coupling argument in Lemma 7 uniformly over $a\in[1/2,2]$. The entropy and moment constants are uniform on this band. The covariance kernel at ah_{0n} differs from that at h_{0n} by $O(|a-1|)$, by the same change of variables used in (37); fitted-density covariance error adds $o_p(\ell_n^{-2})$. Thus the total covariance discrepancy at $a=\widehat{h}_{n,\beta}/h_{0n}$ is $o_p(\ell_n^{-2})$. The finite-net Gaussian comparison and modulus argument in Lemma 6 then compare the selected-bandwidth smooth-bootstrap law with the h_{0n} Gaussian law in Kolmogorov distance.

Equations (40)–(41), the corresponding smooth-bootstrap comparison, and Gaussian anti-concentration transfer the size conclusion at h_{0n} to the selected bandwidth. Because $\widehat{h}_{n,\beta}/h_{0n}\to_p 1$, the atom and density-jump signals and the critical-value orders in Lemma 9 change only by $1+o_p(1)$. Their power conclusions transfer as well. □

For the Silverman scale rule used in the application, the sample standard deviation is root-n under a finite fourth moment, and the sample interquartile range is root-n when the density is positive and continuous at the population quartiles. The minimum of these two scale estimates is Lipschitz, including when their population values coincide. If the limiting scale is positive, the pilot therefore has $r_n=n^{-1/2}$ and satisfies the corollary.

D Additional simulation evidence

D.1 Holdout calibration and power

Table 6 reports the split-ratio experiment for the holdout procedure. Ordinary one-sample calibration becomes increasingly distorted as n_B/n_A rises, whereas mixed-bridge calibration remains close to nominal size across the three splits, consistent with Theorem 4.

Table 6: Holdout split-ratio calibration under a smooth normal null with total sample size $n = 5{,}000$. The statistic is evaluated on $I = [-1, 1]$ with the training bandwidth undersmoothed to the $n_A^{-0.35}$ rate. Ordinary holdout uses the one-sample Kolmogorov–Smirnov critical value, while calibrated holdout simulates the mixed-bridge law with scale $\sqrt{1 + n_B/n_A}$. Each row uses the replication and bridge-simulation counts shown.

Train share	n_B/n_A	MC reps	Bridge draws	Mean h_A	Ordinary reject	Calibrated reject
0.25	3.00	500	999	0.0742	0.532	0.030
0.50	1.00	500	999	0.0582	0.272	0.042
0.75	0.33	500	999	0.0505	0.122	0.050

Notes: Rejection rates are nominal five-percent rates. Monte Carlo standard errors are at most 0.022. The table isolates the split-ratio implication of the mixed-bridge limit; ordinary one-sample critical values become increasingly mis-scaled as the test-to-training ratio changes.

Table 7 compares the full-sample and correctly calibrated holdout procedures under the same shrinking-atom and density-jump alternatives. Correct mixed-bridge calibration restores holdout size but does not recover the power lost through sample splitting, consistent with Proposition 7.

Table 7: Power of the full-sample centered test and calibrated holdout against the same local alternatives. The sample size is $n = 10{,}000$, each entry uses 500 Monte Carlo replications, and all statistics are evaluated on the prespecified intervals. Atom masses are $m_n = c\sqrt{h_n \ell_n / n}$ on $[0, 2]$; density jumps are $\Delta_n = c\sqrt{\ell_n/(nh_n)}$ on $[-1, 1]$, where $h_n = 0.9n^{-0.35}$. The full-sample rows use 399 centered-bootstrap draws. Holdout uses a 50/50 split, $h_A = 0.9n_A^{-0.35}$, and the correctly scaled mixed-bridge critical value approximated with 50,000 paths.

Panel A: Shrinking atoms on $[0, 2]$

Procedure	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$	$c = 2.5$	$c = 3$
Full-sample centered	0.030	0.060	0.762	0.996	1.000	1.000	1.000
Calibrated holdout	0.040	0.032	0.030	0.032	0.034	0.036	0.050

Panel B: Density jumps on $[-1, 1]$

Procedure	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$	$c = 3$
Full-sample centered	0.062	0.132	0.612	0.966	1.000	1.000
Calibrated holdout	0.024	0.032	0.030	0.034	0.018	0.030

Notes: Entries are rejection frequencies for nominal five-percent tests. Monte Carlo standard errors are at most 0.022. Because these alternatives are indexed at the smaller full-sample detection scale, Proposition 7 predicts that calibrated holdout remains near its null rejection probability as n grows.

D.2 Local power transitions

Tables 8 and 9 report rejection frequencies as the atom and density-jump signals move through the detection scales in Theorem 3.

Table 8: Interval-restricted shrinking-atom power for the centered frozen-bandwidth Kolmogorov–Smirnov test. The statistic is evaluated on the prespecified interval $I = [0, 2]$. Each cell uses 500 Monte Carlo replications and 399 bootstrap draws. The atom mass is $m_n = c\sqrt{h_n \ell_n / n}$ with $h_n = 0.9n^{-0.35}$.

Interval	n	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$	$c = 2.5$	$c = 3$
$[0, 2]$	2,000	0.034	0.124	0.882	0.998	1.000	1.000	1.000
$[0, 2]$	10,000	0.030	0.060	0.762	0.996	1.000	1.000	1.000

Table 9: Fixed-interval density-jump power for the centered frozen-bandwidth test. The statistic uses $I = [-1, 1]$. Each cell has 500 Monte Carlo replications and 399 bootstrap draws. The local rows set $\Delta_n = c\sqrt{\ell_n/(nh_n)}$ with $h_n = 0.9n^{-0.35}$.

Interval	n	$c = 0$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 2$	$c = 3$
$[-1, 1]$	2,000	0.040	0.144	0.560	0.934	0.998	1.000
$[-1, 1]$	10,000	0.062	0.132	0.612	0.966	1.000	1.000

D.3 Additional smooth-null results

Table 10 evaluates size under several smooth distributions with skewness or local curvature and also varies the numerical resolution used to approximate the continuum supremum.

The $\beta = 0.35$ specification can overreject under pronounced skewness or curvature, whereas stronger undersmoothing at $\beta = 0.40$ substantially improves size. Refining the numerical grid has little effect on the rejection frequencies. The same pattern remains at the larger sample size used to approximate the scale of the administrative application. These results motivate $\beta = 0.40$ as the preferred specification in the empirical analysis.

E Simulation designs

E.1 Centered-test designs

The baseline centered-test simulations use the Gaussian kernel,

$$h_n = 0.9n^{-0.35},$$

and the fixed target interval specified for each design.

The smooth null is

$$X \sim N(0, 1).$$

Table 10: Size of the centered Gaussian-multiplier test under difficult smooth nulls. Every test uses 399 multiplier draws and the prespecified interval shown. Panel A uses 500 Monte Carlo replications and repeats the calculation at 4,096 and 8,192 bins. Panel B uses 300 replications at an administrative-data sample size and reports the finer grid for each bandwidth exponent.

Panel A: $n = 10{,}000$

		$\beta = 0.35$		$\beta = 0.40$	
Null design	I	4,096 bins	8,192 bins	4,096 bins	8,192 bins
Gaussian	$[-1, 1]$	0.060	0.046	0.032	0.026
Standardized lognormal	$[-0.5, 1.5]$	0.108	0.098	0.060	0.066
Separated Gaussian mixture	$[-1.5, 1.7]$	0.138	0.142	0.072	0.076
High-curvature Gaussian mixture	$[-0.75, 0.75]$	0.128	0.144	0.060	0.068
Standardized Student t_5	$[-2, 2]$	0.070	0.060	0.056	0.054
Gaussian, off-center interval	$[0.5, 2]$	0.076	0.066	0.050	0.052

Panel B: $n = 400{,}000$

Null design	I	$\beta = 0.35$ 8,192 bins	$\beta = 0.40$ 16,384 bins
Gaussian	$[-1, 1]$	0.040	0.040
Standardized lognormal	$[-0.5, 1.5]$	0.153	0.077
Separated Gaussian mixture	$[-1.5, 1.7]$	0.147	0.060
High-curvature Gaussian mixture	$[-0.75, 0.75]$	0.167	0.067
Standardized Student t_5	$[-2, 2]$	0.077	0.050
Gaussian, off-center interval	$[0.5, 2]$	0.067	0.047

Notes: Entries are rejection rates for nominal five-percent tests. The Gaussian, lognormal, mixture, and Student designs are infinitely differentiable on a neighborhood of the displayed interval; the mixture with a narrow component creates high local curvature. All bandwidths use a Silverman scale pilot converted to $n^{-\beta}$. In Panel A, the maximum average bin-width-to-bandwidth ratio is 0.037 for $\beta = 0.35$ and 0.056 for $\beta = 0.40$. In the reported fine-grid columns of Panel B it is 0.061 and 0.058, respectively. Grid refinement therefore matches the resolution used in the administrative application and separates bandwidth bias from numerical discretization. Monte Carlo standard errors are at most 0.016 in Panel A and 0.022 in Panel B.

The fixed-atom design has distribution

$$0.99\,N(0,1) + 0.01\,\delta_1$$

and uses the target interval

$$I = [0, 2].$$

The shrinking-atom designs use

$$m_n = c\sqrt{\frac{h_n \ell_n}{n}}, \qquad \ell_n = \log n \vee \log(1/h_n),$$

with the atom at one and the same target interval.

The density-jump design uses a two-piece normal distribution. Conditional on the sign, the absolute value of the observation is half-normal. If p_+ is the probability of a positive draw, the right and left density limits at zero are

$$2p_+\phi(0) \qquad \text{and} \qquad 2(1-p_+)\phi(0),$$

so the density jump is

$$\Delta = 2\phi(0)(2p_+ - 1).$$

For

$$\Delta_n = c\sqrt{\frac{\ell_n}{nh_n}},$$

the simulation therefore sets

$$p_+ = \frac{1}{2} + \frac{\Delta_n}{4\phi(0)}.$$

The target interval is

$$I = [-1, 1].$$

E.2 Holdout designs

The holdout experiments draw from

$$N(0,1)$$

and evaluate the statistic on

$$I = [-1, 1].$$

The training distribution is estimated using a Gaussian KDE. Ordinary calibration compares

$$\sqrt{n_B}\sup_{x\in I}|F_B(x)-\widehat{F}_A(x)|$$

with the one-sample Brownian-bridge reference. Mixed-bridge calibration instead uses the factor

$$\sqrt{1+\frac{n_B}{n_A}}$$

from Theorem 4 and estimates the probability range $F(I)$ using the pooled empirical CDF.

The split-ratio experiment varies n_B/n_A. The power comparison uses the same shrinking-atom and density-jump arrays as the full-sample experiments at

$$n = 10{,}000.$$

Calibrated holdout uses a 50/50 split and

$$h_A = 0.9n_A^{-0.35}.$$

E.3 Difficult smooth-null designs

The additional size experiments use six smooth designs.

The first is the standard Gaussian,

$$X\sim N(0,1), \qquad I=[-1,1].$$

For the standardized lognormal design,

$$Z\sim N(0,1), \qquad X=\frac{e^Z-e^{1/2}}{\sqrt{e(e-1)}}.$$

Thus

$$\mathbb{E}X=0, \qquad \mathrm{Var}(X)=1.$$

The lower endpoint of the support is approximately -0.763, so

$$I=[-0.5,1.5]$$

lies strictly inside the support.

The separated Gaussian mixture is

$$X \sim \begin{cases} N(-1, 0.45^2), & \text{with probability } 0.65, \\ N(1.15, 0.55^2), & \text{with probability } 0.35. \end{cases}$$

This distribution is not restandardized. The target interval is

$$I = [-1.5, 1.7].$$

The high-curvature Gaussian mixture is

$$X \sim 0.80N(0, 1) + 0.20N(0, 0.18^2),$$

with target interval

$$I = [-0.75, 0.75].$$

The narrow component generates substantial local curvature around zero while the density remains smooth.

For the standardized Student design,

$$T \sim t_5, \qquad X = T\sqrt{\frac{3}{5}}.$$

Since

$$\mathrm{Var}(T) = \frac{5}{3},$$

X has unit variance. The target interval is

$$I = [-2, 2].$$

The final design again uses

$$X \sim N(0, 1),$$

but evaluates the statistic on the off-center interval

$$I = [0.5, 2].$$

For these experiments the bandwidth is

$$h_n = 0.9 \min\left\{\widehat{\sigma}, \frac{\widehat{\mathrm{IQR}}}{1.34898}\right\} n^{-\beta}.$$

F Additional empirical results and robustness

F.1 Bandwidth sensitivity

Table 11 reports the annual results under the $\beta = 0.35$ sensitivity specification. Individual taxable income rejects in every year, while family average income rejects only in 2007.

Table 11: Bandwidth sensitivity of the centered Kolmogorov–Smirnov tests in Australian tax data, 2001–2022.

			Individual taxable income		Family average income	
Tax year	N_I	N_F	Statistic	p-value	Statistic	p-value
2001	1,311,566	381,086	0.06481	0.0005	0.01069	0.1465
2002	1,338,894	391,021	0.06503	0.0005	0.00886	0.7270
2003	1,373,642	411,716	0.06986	0.0005	0.00880	0.6390
2004	1,416,813	421,523	0.04732	0.0005	0.00857	0.6440
2005	1,463,432	435,809	0.04992	0.0005	0.00784	0.8705
2006	1,503,532	448,538	0.05540	0.0005	0.00817	0.6195
2007	1,549,398	462,276	0.08874	0.0005	0.01127	0.0050
2008	1,616,357	483,181	0.06348	0.0005	0.00771	0.6515
2009	1,616,172	482,306	0.05900	0.0005	0.00843	0.2275
2010	1,631,962	487,589	0.07037	0.0005	0.00832	0.2130
2011	1,660,606	495,899	0.09229	0.0005	0.00849	0.1100
2012	1,673,961	496,305	0.06989	0.0005	0.00673	0.8380
2013	1,666,463	496,988	0.07643	0.0005	0.00734	0.3915
2014	1,699,498	505,277	0.09066	0.0005	0.00702	0.5250
2015	1,733,189	512,942	0.08688	0.0005	0.00620	0.9235
2016	1,759,647	518,861	0.07498	0.0005	0.00668	0.6300
2017	1,787,031	524,705	0.08122	0.0005	0.00696	0.4350
2018	1,800,007	527,204	0.07079	0.0005	0.00683	0.4375
2019	1,809,227	529,474	0.09202	0.0005	0.00621	0.7605
2020	1,801,176	526,588	0.05367	0.0005	0.00673	0.3755
2021	1,785,516	520,246	0.07369	0.0005	0.00617	0.6770
2022	1,753,042	510,322	0.08188	0.0005	0.00707	0.0950

Notes: The table reports the $\beta = 0.35$ sensitivity specification. N_I counts retained person-years and N_F counts retained non-singleton family-year units after the strict $(0, \$200{,}000)$ income filter. The individual and family distributions are respectively person-weighted and family-weighted. Both tests use the prespecified interval $I = [\$1{,}000, \$199{,}000]$. The reported statistic is $U_n = \sqrt{n/h}\sup_{x\in I}|F_n(x) - \widehat{F}_h(x)|$; its reference law is the centered Gaussian-multiplier process in Corollary 1. Each p-value is the finite-simulation plus-one value based on 1,999 multiplier draws; 0.0005 is therefore the smallest reportable value. Tests are two-sided at the five-percent level.

F.2 Additional annual distributions

Figures 4 and 5 show the individual and family average taxable income distributions in four additional years spanning the sample.

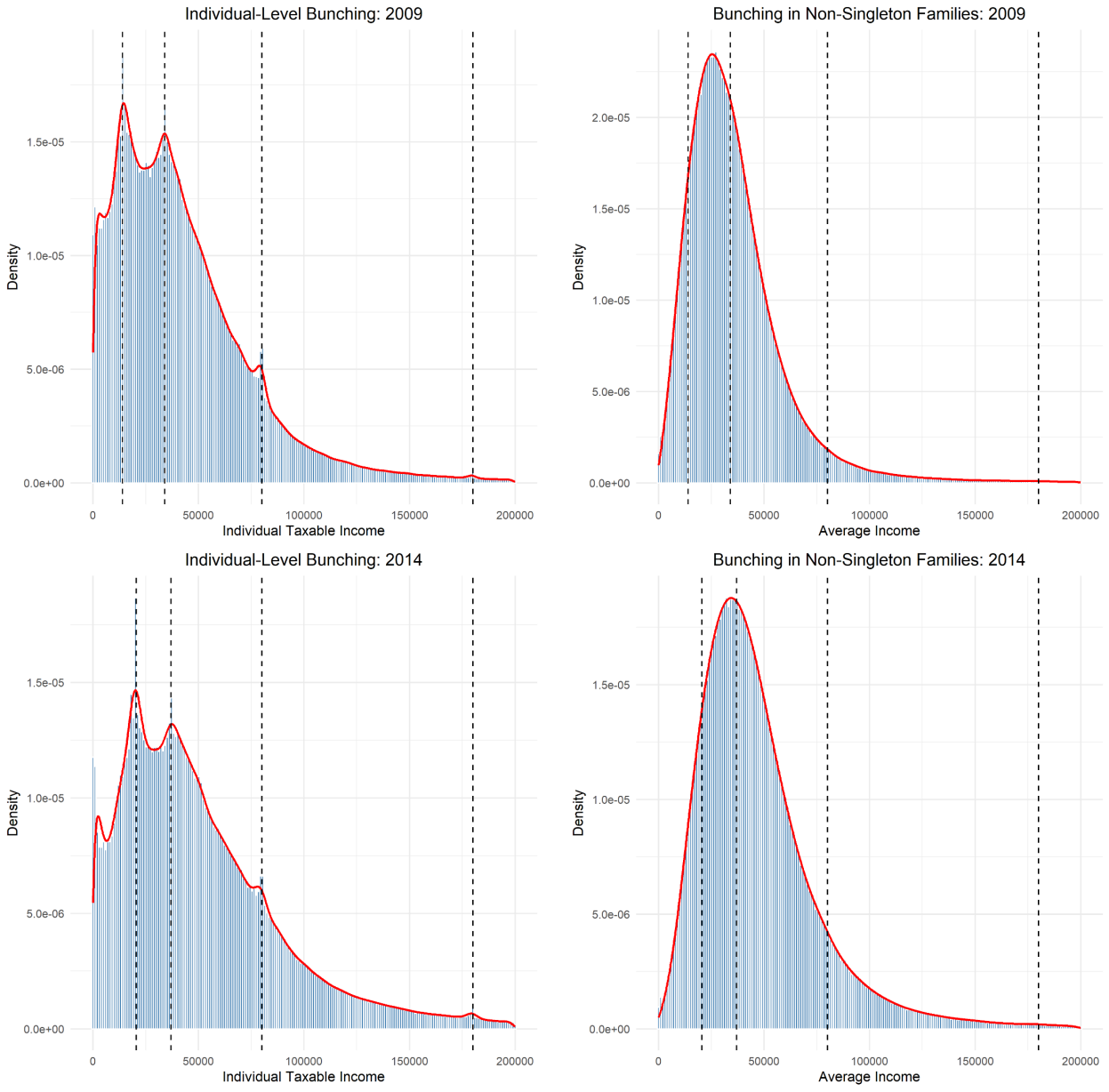


Figure 4: Individual and family average taxable income distributions in 2009 and 2014. The left column shows individual taxable income and the right column shows average taxable income in linked non-singleton families. Vertical dashed lines mark tax thresholds, and the solid curve shows the kernel density. The plots use the range from zero to $200,000 for display.

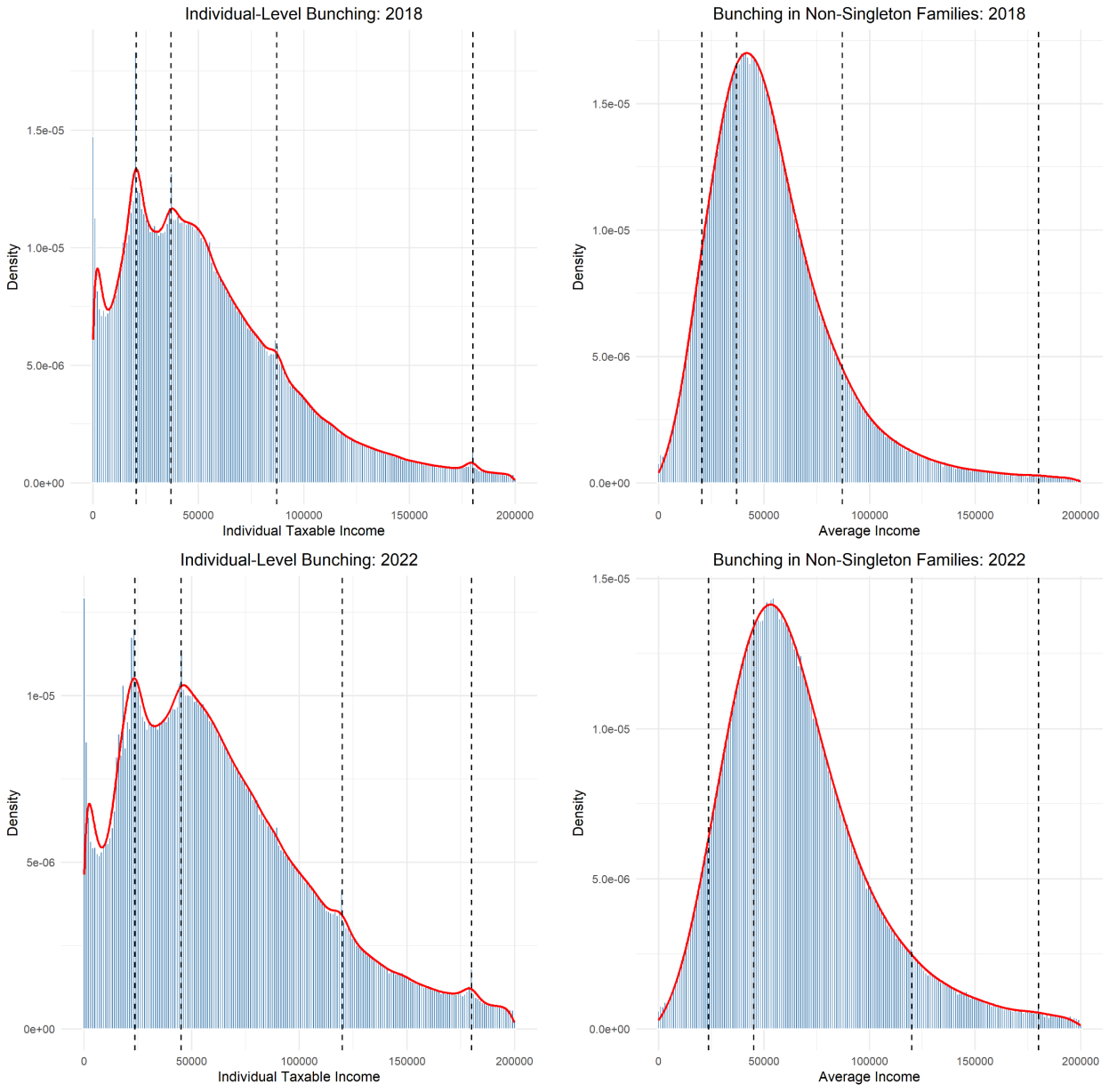


Figure 5: Individual and family average taxable income distributions in 2018 and 2022. The left column shows individual taxable income and the right column shows average taxable income in linked non-singleton families. Vertical dashed lines mark tax thresholds, and the solid curve shows the kernel density. The plots use the range from zero to $200,000 for display.

F.3 Robustness to dollar rounding

Table 12 repeats the $\beta = 0.35$ sensitivity specification after rounding family averages to the one-dollar grid. The raw family distribution rejects only in 2007. The rounded family distribution rejects in 2007 and, marginally, in 2011; every other rounded-family p-value is at least 0.0925.

Table 12: Robustness of the Australian tax-data results to dollar rounding, 2001–2022.

	Individual income		Family average		Dollar-rounded family average	
Tax year	Statistic	p-value	Statistic	p-value	Statistic	p-value
2001	0.06481	0.0005	0.01069	0.1465	0.01052	0.1830
2002	0.06503	0.0005	0.00886	0.7270	0.00814	0.9450
2003	0.06986	0.0005	0.00880	0.6390	0.00872	0.6745
2004	0.04732	0.0005	0.00857	0.6440	0.00847	0.6975
2005	0.04992	0.0005	0.00784	0.8705	0.00814	0.7370
2006	0.05540	0.0005	0.00817	0.6195	0.00864	0.4045
2007	0.08874	0.0005	0.01127	0.0050	0.01097	0.0080
2008	0.06348	0.0005	0.00771	0.6515	0.00825	0.3750
2009	0.05900	0.0005	0.00843	0.2275	0.00859	0.1800
2010	0.07037	0.0005	0.00832	0.2130	0.00809	0.3010
2011	0.09229	0.0005	0.00849	0.1100	0.00896	0.0470
2012	0.06989	0.0005	0.00673	0.8380	0.00684	0.7790
2013	0.07643	0.0005	0.00734	0.3915	0.00743	0.3520
2014	0.09066	0.0005	0.00702	0.5250	0.00722	0.4160
2015	0.08688	0.0005	0.00620	0.9235	0.00636	0.8515
2016	0.07498	0.0005	0.00668	0.6300	0.00709	0.3945
2017	0.08122	0.0005	0.00696	0.4350	0.00729	0.2775
2018	0.07079	0.0005	0.00683	0.4375	0.00683	0.4375
2019	0.09202	0.0005	0.00621	0.7605	0.00605	0.8465
2020	0.05367	0.0005	0.00673	0.3755	0.00735	0.1320
2021	0.07369	0.0005	0.00617	0.6770	0.00625	0.6290
2022	0.08188	0.0005	0.00707	0.0950	0.00708	0.0925
Years rejected	22 of 22		1 of 22		2 of 22	

Notes: The table uses the $\beta = 0.35$ sensitivity specification. The reported statistic is $U_n = \sqrt{n/h}\sup_{x\in I}|F_n(x) - \widehat{F}_h(x)|$, and all p-values use 1,999 draws from the centered Gaussian-multiplier reference law with the plus-one convention. Individual taxable income is already recorded in whole dollars, so its rounded specification is identical and is not repeated. Family averages may take fractional-dollar values; the last two columns recompute the complete test after rounding those averages to the one-dollar grid.

F.4 Calibrated family response simulation

The quantitative exercise uses the 2021 individual distribution to construct a smooth benchmark for heterogeneous baseline income opportunities. Let w denote baseline resources. We form an 8,192-cell histogram on $(0, \$200{,}000)$, interpolate the log density through $\pm\$750$ neighborhoods of the four principal markers \$23,226, \$45,000, \$120,000, and \$180,000, and smooth the resulting density with a Gaussian kernel. The smoothing bandwidth is $0.9sn^{-1/5}$, where $s = \min\{\widehat{\sigma}, \widehat{\mathrm{IQR}}/1.34898\}$. The resulting distribution is used as a smooth empirical benchmark rather than interpreted as a structural distribution of baseline income opportunities.

At $\alpha = 0$, an individual with type w chooses taxable income x to maximize

$$x - T(x) - \frac{(x-w)^2}{2\tau},$$

where T is the effective 2020–21 Australian personal income tax schedule including LITO, LMITO, and the Medicare levy. On a linear segment with net-of-tax slope q, the interior solution is $x = w + \tau q$; intervals of types can therefore bunch at upward marginal-tax kinks. We choose τ to match the observed concentration of individual income around the four principal markers. Let p_{300} denote the probability of lying within ±\$150 of at least one marker and p_{500} the analogous probability within ±\$250. In the data these probabilities are 1.1270 percent and 1.7713 percent. The preferred one-parameter calibration minimizes

$$\left(\frac{p_{300}(\tau) - p_{300}^{\text{obs}}}{p_{300}^{\text{obs}}}\right)^2 + \left(\frac{p_{500}(\tau) - p_{500}^{\text{obs}}}{p_{500}^{\text{obs}}}\right)^2.$$

It gives $\widehat{\tau}_0 = 822.415$, with fitted probabilities 1.1634 percent and 1.6904 percent. The benchmark distribution and $\widehat{\tau}_0$ are then held fixed throughout the family exercise.

For each simulated two-member family, draw w_1 and w_2 independently from the calibrated smooth benchmark. At allocable share α, member j has active-income type

$$\theta_j = (1-\alpha)w_j,$$

and the family has an allocable passive-income pool

$$P = \alpha(w_1 + w_2).$$

If r_j denotes passive income assigned to member j, with $r_1 + r_2 = P$, the family chooses the allocation and members' taxable incomes to maximize the sum of

$$x_j - T(x_j) - \frac{(x_j - r_j - \theta_j)^2}{2\widehat{\tau}_0}.$$

Thus α changes the scope for within-family allocation while leaving fixed both the distribution of baseline resources and the calibrated real responsiveness. The quasi-linear specification abstracts from wealth effects of passive income.

The population calculations use 1,000,000 independently drawn families at $\alpha = 0$, 0.01, and increments of 0.025 from zero to one. Table 13 reports selected cells. The atom diagnostics count mass within one cent of pairwise composite locations formed from the four principal calibration markers. They are useful for relating the economic model to the single-atom

benchmark but are not the global test statistic and do not exhaust all effective-schedule kinks.

Table 13: Population implications of the calibrated family-response model.

α	Total selected atom	Largest selected atom	Mean transfer share
0.0%	0.0006%	0.0003%	0.0%
1.0%	0.0025%	0.0016%	0.7%
2.5%	0.0077%	0.0050%	1.6%
5.0%	0.0150%	0.0094%	3.2%
10.0%	0.0299%	0.0197%	6.1%
12.5%	0.0378%	0.0247%	7.4%
15.0%	0.0459%	0.0306%	8.6%
17.5%	0.0555%	0.0375%	9.7%
20.0%	0.0657%	0.0443%	10.8%
25.0%	0.0847%	0.0566%	12.6%
30.0%	0.1035%	0.0687%	14.2%
50.0%	0.1776%	0.1102%	18.9%
100.0%	0.1776%	0.1102%	20.7%

Notes: The response parameter and smooth benchmark distribution are calibrated once at $\alpha = 0$ and held fixed. Entries use 1,000,000 independently drawn two-member families. The atom columns count mass within one cent of composite locations formed from the four principal calibration markers; they are descriptive diagnostics rather than the global test statistic. The transfer share is the absolute reallocation relative to total baseline family resources.

For the power calculation, every Monte Carlo replication draws a fresh sample of 520,246 families, solves the family problem, and applies the preferred $\beta = 0.40$ global smoothness test on $[\$1{,}000, \$199{,}000]$. The completed regular $\alpha = 0$ cell uses 100 outer replications and 1,999 Gaussian multiplier draws per test and rejects in 2 replications. The exact 95 percent binomial interval is 0.0024 to 0.0704.

F.5 Literal-atom upper-bound sensitivity

The regular calibration uses raw probability in narrow kink windows rather than a conventional estimate of excess bunching. As an upper-bound sensitivity, we instead treat the entire observed p_{300} union mass of 1.1270 percent as point mass at the four principal markers. This construction attributes ordinary smooth probability within the windows to bunching and therefore intentionally overstates the individual response.

The resulting calibration requires $\tau = 5294.7$, about 6.44 times the regular calibrated value. Table 14 reports the corresponding family-test power. At zero allocability, 43 of 100 family samples reject. At one percent allocability, all 100 reject. The comparison shows how rapidly the family implications strengthen under this more aggressive interpretation of the

observed individual concentration.

Table 14: Power of the family test under the regular calibration and upper-bound sensitivity

Calibration	α	τ	Rejections	Power	Exact 95% interval
Regular p_{300}/p_{500}	0.0%	822.415	2/100	0.02	[0.0024, 0.0704]
Upper-bound sensitivity	0.0%	5294.699	43/100	0.43	[0.3314, 0.5329]
Upper-bound sensitivity	1.0%	5294.699	100/100	1.00	[0.9638, 1.0000]

Notes: Each row reports 100 Monte Carlo samples of 520,246 two-member families, with 1,999 Gaussian multiplier draws per test. The regular calibration matches the observed combined p_{300} and p_{500} window masses. The upper-bound sensitivity instead assigns the full observed p_{300} mass of 1.1270 percent as point mass at the four principal markers.

References

Abhayaratna, Thomas, Andrew Carter, and Shane Johnson, "The ATO Longitudinal Information Files (ALife): Individuals: A new dataset for public policy research," *Australian Economic Review*, 2022, *55* (4), 541–557.

Australian Taxation Office, "Individual Income Tax Rates and Thresholds, 2021–22," 2022. Australian Government.

_ , "Individual Income Tax Rates, 1950–51 to 2022–23 Income Years," Taxation Statistics 2022–23, Snapshot Table 1 2025. Australian Government.

Bickel, Peter J. and Murray Rosenblatt, "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1973, *1* (6), 1071–1095.

Blomquist, Soren, Whitney K. Newey, Anil Kumar, and Che-Yuan Liang, "On Bunching and Identification of the Taxable Income Elasticity," *Journal of Political Economy*, 2021, *129* (8), 2320–2343.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Manipulation Testing Based on Density Discontinuity," *The Stata Journal*, 2018, *18* (1), 234–261.

_ , _ , and _ , "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato, "Gaussian Approximation of Suprema of Empirical Processes," *Annals of Statistics*, 2014, *42* (4), 1564–1597.

_ , _ , and _ , "Empirical and Multiplier Bootstraps for Suprema of Empirical Processes of Increasing Complexity, and Related Gaussian Couplings," *Stochastic Processes and their Applications*, 2016, *126* (12), 3632–3651.

Chetty, Raj, John N Friedman, and Emmanuel Saez, "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 2013, *103* (7), 2683–2721.

_ , _ , Tore Olsen, and Luigi Pistaferri, "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records," *The Quarterly Journal of Economics*, 2011, *126* (2), 749–804.

Feldstein, Martin, "Tax avoidance and the deadweight loss of the income tax," *Review of Economics and Statistics*, 1999, *81* (4), 674–680.

Fischer, Daniel, Justin Holland, Marc Chan, Cain Polidano, Ha Vu, and Roger Wilkins, "The ATO ALife Family File," *Australian Economic Review*, 2025, *58* (2), 163–172.

Gine, Evarist and Armelle Guillou, "Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators," *Annales de l'Institut Henri Poincare: Probabilites et Statistiques*, 2002, *38* (6), 907–921.

Gordon, Roger H. and Joel Slemrod, "Are "Real" Responses to Taxes Simply Income Shifting between Corporate and Personal Tax Bases?," NBER Working Paper 6576, National Bureau of Economic Research, Cambridge, MA 1998.

Hall, Peter, "The Influence of Rounding Errors on Some Nonparametric Estimators of a Density and its Derivatives," *SIAM Journal on Applied Mathematics*, 1982, *42* (2), 390–399.

Johnson, Shane, Robert Breunig, Miguel Olivo-Villabrille, and Arezou Zaresani, "Individuals' responsiveness to marginal tax rates: Evidence from bunching in the Australian personal income tax," *Labour Economics*, 2024, *87*, 102461.

Kleven, Henrik J and Mazhar Waseem, "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan," *The Quarterly Journal of Economics*, 2013, *128* (2), 669–723.

Kopczuk, Wojciech, "Shifting and Other Problems with Taxable Income Elasticity: Joel Slemrod's Contributions to What We Know About Taxing Ourselves," *National Tax Journal*, 2025, *78* (3), 751–766.

Maire, Daniel Le and Bertel Schjerning, "Tax bunching, income shifting and self-employment," *Journal of Public Economics*, 2013, *107*, 1–18.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Saez, Emmanuel, "Do taxpayers bunch at kink points?," *American Economic Journal: Economic Policy*, 2010, *2* (3), 180–212.

Scott, David W. and Simon J. Sheather, "Kernel Density Estimation with Binned Data," *Communications in Statistics – Theory and Methods*, 1985, *14* (6), 1353–1359.

Silverman, Bernard W., *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall, 1986.

van der Vaart, Aad W. and Jon A. Wellner, *Weak Convergence and Empirical Processes*, New York: Springer, 1996.