Impatience or illiquidity? The case of early 401(k) withdrawals

Aaron Goodman*

November 2025

Abstract

A third of 401(k) participants cash out their balances when leaving their jobs. While the prevalence of pre-retirement withdrawals is well documented, empirical evidence on underlying mechanisms is lacking. Using a novel paycheck-level income panel covering millions of participants, I show that workers with greater sub-annual income volatility are significantly more likely to cash out. The relationship between earnings volatility and early withdrawals is conditional on annual income and holds using both within-firm and within-worker identification strategies. The empirical results are consistent with a model in which workers use – and behaviorally biased workers overuse – retirement assets as insurance against short-run income shocks. A linked survey confirms that many participants with volatile incomes lack liquid savings, despite contributing regularly to illiquid retirement accounts.

Keywords: retirement saving, early withdrawals, liquidity constraints, income volatility

1 Introduction

A significant share of assets in U.S. defined-contribution retirement accounts are withdrawn before retirement. The implications of this fact – for public policy, for the design of employer-sponsored retirement plans, and for analysis of household saving behavior more broadly – depend crucially on the reasons for early withdrawals. If retirement savings leakage arises solely from present bias, then early-withdrawal penalties raise welfare. Alternatively, if households tap retirement accounts because they lack the liquid resources to weather short-term shocks, then they benefit from policies that facilitate liquid wealth formation and make retirement assets available during emergencies.

*Vanguard, 100 Vanguard Blvd, Malvern, PA, 19355, aaron_goodman2@vanguard.com. I thank John Ameriks, John Beshears, Guillermo Carranza, Marcos Catao, James Choi, Taha Choukhmane, Jeff Clark, Paulo Costa, Tom De Luca, Malena de la Fuente, Peter Ganong, Fiona Greig, Kelly Hahn, David Laibson, Spencer Look, Anna Madamba, Marsella Martino, Pascal Noel, Cormac O'Dea, David Pakula, Erich Pingel, Andy Reed, Lawrence Schmidt, Fu Tan, Jack VanDerhei, Xiao Xu, Nicky Zhang, and colleagues at Vanguard's Investment Strategy Group for valuable feedback. The views expressed here are those of the author and not necessarily those of the Vanguard Group, Inc. **This material is provided for informational purposes only and is not intended to be investment advice or a recommendation to take any particular investment action.**

This paper presents the strongest evidence to date that liquidity constraints drive early withdrawals. The empirical analysis centers on a novel paycheck-level income panel covering 3.4 million 401(k) plan participants during a 16-year period. Because I construct the paycheck panel from administrative 401(k) recordkeeping data, I can relate granular features of participants' income histories to their saving and withdrawal decisions. When I do so, a consistent finding emerges: participants with greater sub-annual income volatility are significantly more likely to take penalized early withdrawals. The relationship between earnings volatility and withdrawal rates is conditional on annual income, most pronounced among low- and middle-income workers, and driven by both mean-zero and asymmetric negative shocks. These results are consistent with a framework in which income shocks (occurring *during* stable employment relationships, and for many hourly workers *re*curring on a monthly basis) create liquidity shortages that prompt early withdrawals.

To formalize the theoretical framework and discipline the empirical analysis, I develop a model of retirement saving and withdrawal decisions that reflects the institutional reality of the U.S. defined-contribution (DC) system. In the model, a worker earns stochastic labor income over multiple periods and saves for retirement in a 401(k) account. The worker may withdraw from the 401(k) account before retirement, but early withdrawals incur a penalty and are only permitted in a designated subset of periods during working life. The designated early-withdrawal periods represent times when the worker changes jobs and thus becomes eligible to take a penalized cash distribution from her 401(k) account. Because withdrawals during employment are subject to relatively strict regulation, the post-separation cashouts on which the model focuses are the most common form of 401(k) leakage: in 2023, 33% of all separators in Vanguard-administered plans cashed out (Clark, 2025). The modeled worker also has access to a lower-yielding liquid savings account, from which withdrawals can be taken at any time without penalty.

The model makes three core predictions that structure the empirical analysis and shape my interpretation of the empirical results. First, negative income shocks cause cashouts. When the worker receives sufficiently poor draws from the stochastic income process, the consumption-smoothing benefit of cashing out exceeds the early-withdrawal penalty. Second, mean-zero income volatility causes cashouts. Even when negative income shocks

are accompanied by equally sized positive shocks during the same employment spell, the worker uses her liquid wealth to smooth consumption, has less liquidity at separation, and may cash out in order to replenish the liquid account before beginning her next job. Third, the relationship between income volatility and cashout propensity holds in a baseline case where the worker is fully rational, but becomes stronger when the worker is present-biased (Laibson, 1997). This is because liquidity is partly endogenous: behavioral workers build smaller precautionary buffers in the liquid account, have less liquidity on hand when income shocks arrive, and must resort to penalized 401(k) withdrawals more often.

Motivated by the model of withdrawal behavior, I build a dataset uniquely suited to test its predictions. High-frequency income observations are crucial: even short-lived negative shocks can deplete liquidity buffers, and mean-zero volatility within a given year would be completely undetectable in annual income data. Using administrative 401(k) recordkeeping data from Vanguard, I rely on the fact that for most of the plans it administers, Vanguard tracks both the dollar amounts of employee contribution transactions and employees' underlying contribution rates. Combining the information on contribution amounts and rates allows me to infer individual paychecks. For example, a participant who made a $50 contribution with a 5% contribution rate on January 1 likely received a $1,000 paycheck on January 1. Focusing on cases where I can verify that the annual sum of inferred paychecks matches official W-2 income figures, I obtain complete paycheck-level income histories for 10.6 million participant-years across 1,259 plans during the 2008-2023 period.

The paycheck data confirm the model's theoretical predictions. I first divide the sample of participants into those who are likely hourly workers and those who are likely salaried, based on the frequency with which their paycheck amounts change by at least $1 relative to the lagged paycheck amount.[1] Hourly workers – who account for over half of my sample – have post-separation cashout rates 10–15 percentage points higher than salaried workers with similar annual incomes. To measure volatility quantitatively, I work at the monthly level and consider monthly earnings fluctuations *within* a given participant-year. In regressions that control for annual income, age, employment tenure, and plan-level

[1] I thank Peter Ganong, Pascal Noel, and their coauthors in Ganong et al. (2024) for providing statistics from external payroll data that help make this hourly-salaried classification as accurate as possible.

fixed effects, I find that a 10% increase in monthly income volatility raises post-separation cashout probability by 8 percentage points for middle-income workers (a 24% increase from the baseline cashout rate of 34%). When I control separately for the second moment and the third moment (i.e., skewness) of monthly income, I find that asymmetric negative shocks have even stronger effects on cashout probability than mean-zero volatility.

Present bias plays a nuanced role in the model of cashout behavior. If hourly and salaried workers have the same present bias, and this common level of present bias prevents hourly workers from holding the additional liquid wealth needed to smooth their more volatile incomes, then the cross-sectional analysis accurately identifies the effect of income volatility. If hourly workers are more present-biased, then their cashout rates would be higher even in the absence of income volatility, and the cross-sectional estimates are biased upward in magnitude. The cross-sectional analysis controls for annual income, which the literature has shown to be a primary predictor of present bias (Lawrance, 1991; Meier and Sprenger, 2015). The identification threat is thus the less obvious possibility that income volatility correlates with present bias *conditional* on annual earnings. To address this possibility, I adopt alternative identification strategies using the subset of participants who make observable changes between hourly and salaried employment.

First, I study participants who make hourly-salaried changes within employment spells, exploiting variation in the timing of those changes. For example, consider two participants a and b who both appear in my sample for n years, begin as hourly workers, and switch to salaried employment during their tenures. Participant a becomes salaried in year $n-1$, while participant b becomes salaried in year n. As long as present bias does not correlate with the precise timing of hourly-salaried changes, a and b cleanly identify the effect of income volatility. I show that the extra year participant b spends as an hourly worker raises her cashout probability by 1.0–1.2 percentage points (still controlling for annual income). Extrapolating to a median employment tenure of five years, this implies a 5–6 percentage-point gap in cashout rates between hourly and salaried workers. The result persists when I restrict to apparently exogenous events where many workers at the same firm make the same change between hourly and salaried employment in the same year.

Second, I leverage a small number of workers who participate in two different Vanguard-

administered plans during the sample period, allowing me to observe multiple cashout outcomes for a given worker. I show that monthly income volatility continues to predict cashout rates in regressions with worker-level fixed effects, which absorb time-invariant characteristics like present bias and are identified from variation in hourly-salaried status between distinct employment spells.

Finally, to verify the theoretical link between liquidity and 401(k) withdrawals, I use a 2024 survey of participants in Vanguard-administered plans, with responses linked to the administrative recordkeeping data. Among a sample of roughly 4,000 survey respondents, 28% report being less than fully confident in their ability to meet a $2,000 emergency expense. Hourly workers with volatile incomes are more likely to lack liquidity, and self-reported liquidity strongly predicts 401(k) withdrawal rates. Over half of low-liquidity participants contribute above their employer's match cap, suggesting that these workers are not strategically exploiting match policies and that their marginal savings dollars might be better allocated to liquid accounts. Unlike the wealthy hand-to-mouth who rationally concentrate their portfolios in higher-yielding illiquid assets (Kaplan and Violante, 2014; Kaplan et al., 2014), many low- and middle-income 401(k) participants could likely benefit from improved liquidity management, instead of paying costly early-withdrawal penalties to insure themselves against recurring short-run income shocks.

I contribute to an empirical literature studying pre-retirement leakage from retirement accounts. Most prior work in this literature computes leakage rates using federal tax-return data (Coyne et al., 2024; Choukhmane et al., 2024; Goodman et al., 2021; Argento et al., 2015; Amromin and Smith, 2003). In addition to studying average leakage rates, Goodman et al. (2021),[2] Argento et al. (2015), and Coyne et al. (2024) find evidence that households are more likely to take early withdrawals in years where their annual income drops by a large amount or they receive unemployment benefits. By using recordkeeping data to study the effect of income volatility, I provide a more thorough analysis that overcomes two limitations of annual tax-return data. First, the correlation between income shocks and withdrawals in tax-return data is partly mechanical: job changes make workers eligible

[2]I share a last name with the first-listed author on this paper but we are not related. I was not invited to collaborate on Goodman et al. (2015).

to take 401(k) distributions and may simultaneously reduce annual income if there is a transition or unemployment period between jobs. I *condition* on separation, then relate cashout decisions to income volatility *during* the preceding employment spell.[3] Second, the sub-annual volatility I study is fundamentally different (and likely more common) than large shocks to annual income. I can detect the mean-zero, within-year volatility that matters for workers' liquidity positions but is not observable in annual income data.

I show that liquidity constraints are an important explanation for early withdrawals, but policymakers generally face a tradeoff between mitigating behavioral biases and providing flexibility to constrained households. Beshears et al. (2025) study this tradeoff in a generalized planning problem and find that the optimal retirement savings system involves a high level of mandatory savings that are completely illiquid before retirement. The optimality of mandating high illiquid savings levels depends on an assumption that present bias is sufficiently heterogeneous relative to the distribution of pre-retirement marginal-utility shocks. The substantial monthly income volatility I document among 401(k) participants establishes that pre-retirement liquidity needs are highly variable.[4] Future work should provide more precise estimates of the empirical distribution of present bias,[5] allowing for data-driven evaluations of the commitment-versus-flexibility tradeoff.

I also contribute to a literature studying the optimal design of employer-sponsored retirement plans. Since the seminal contribution of Madrian and Shea (2001), a substantial body of empirical work has shown that automatic enrollment increases participation and saving rates in 401(k) plans (Choi et al., 2002, 2004; Beshears et al., 2009). More recent studies show that the beneficial effects of automatic enrollment partially dissipate over time (Choukhmane, 2025) as workers reduce non-retirement saving (Choukhmane and Palmer, 2024), take on debt (Beshears et al., 2024a), and cash out at separation (Derby et al., 2023; Choi et al., 2024). The link between income volatility and cashouts suggests liquidity constraints are a partial explanation for attenuated long-run savings effects. Theoretically, the model I develop demonstrates the value of plan-design choices that account for workers'

[3]For completeness, I provide an additional analysis of layoffs in Appendix D.

[4]Checking and credit card account data (Farrell and Greig, 2015) and household survey data (Fulford and Low, 2024) suggest expense shocks are even more volatile than monthly income. The income volatility I document could therefore understate the total variance of marginal-utility shocks.

[5]See the literature reviews in Lockwood (2020) and Cohen et al. (2020).

liquidity needs. To the extent that behavioral biases inhibit retirement saving, they also cause workers to undersave in precautionary liquid accounts (Ganong and Noel, 2019; Ameriks et al., 2007), which in turn hinders consumption smoothing of typical labor income shocks (Ganong et al., 2025). Promising interventions include those that enroll workers in employer-sponsored short-run savings accounts (Berk et al., 2023) and allow limited amounts of emergency 401(k) liquidity during employment (Beshears et al., 2024b).

Finally, I add to a growing literature on heterogeneous labor income risk (Guvenen et al., 2021; Busch et al., 2022; Meeuwis et al., 2025). While this literature relies primarily on annual income data and studies earnings risk that arises largely from job changes and unemployment, I show that sub-annual volatility within a given employment spell can have important household financial implications. In using paycheck-level data and analyzing monthly earnings volatility, I follow Ganong et al. (2024), who are the first to do so with U.S. administrative data. Relative to the payroll data in Ganong et al. (2024), the earnings information I construct from 401(k) data is valuable because it allows me to relate income volatility to workers' retirement saving and withdrawal decisions.

The paper proceeds as follows. Section 2 develops the model of retirement saving and withdrawal decisions. Section 3 describes the 401(k) recordkeeping data and the process by which I infer paychecks. Section 4 presents the main empirical results linking sub-annual income volatility to early 401(k) withdrawals. Section 5 studies cashouts using alternative within-person identification strategies. Section 6 analyzes a linked survey gauging workers' liquid savings. Section 7 discusses policy implications and Section 8 concludes.

2 Model

2.1 Setup

To motivate the empirical analysis, I introduce a simple model of retirement saving and withdrawal decisions. The model describes the partial-equilibrium behavior of a worker who saves in a 401(k) account and maximizes discounted lifetime utility. As in prior literature, I allow the worker to be naively present-biased (Laibson, 1997; Beshears et al., 2022). The distinguishing feature of the model is labor income volatility, which can

shift the marginal utility of consumption during working life and potentially induce early withdrawals from the 401(k) account.[6]

The worker's life consists of T periods. In each period t, the worker's objective is to maximize discounted utility, which is given by

$$u(c_t) + \beta \sum_{x=t+1}^{T} \delta^{x-t} u(c_x) \tag{1}$$

The exponential discount factor is $\delta < 1$ and $\beta \in (0, 1]$ controls the degree of present bias.

For $t < t_R$, the worker is employed and earns stochastic labor income y_t that is independently distributed across periods. For $t \geq t_R$, the worker is retired and receives fixed Social Security income y_R. Consumption may differ from current income because the worker has access to two savings accounts. The first account, with balance at the beginning of period t denoted by a_t^L, is a completely liquid account from which the worker can withdraw at any time without penalty. The only constraint applying to the liquid account is that its balance may not fall below a fixed floor, $a_t^L \geq \underline{a^L}$. As a baseline I consider a simple no-borrowing case, with $\underline{a^L} = 0$. The return on the liquid account is normalized to zero.

The second savings account is a partially illiquid 401(k) account, with balance at the beginning of period t denoted by a_t^I. The 401(k) account must always have a weakly positive balance and earns a return of r between periods. Contributions to the account are matched by the worker's employer at rate m up to a per-period cap of $\bar{m}$. During the retirement period for $t \geq t_R$, the worker may withdraw from the account without penalty. Additionally, for a subset of periods during working life $t \in W$, the worker may withdraw from the 401(k) account but must pay a proportional early-withdrawal penalty of τ. The pre-retirement withdrawal periods W capture instances where the worker changes jobs and thus becomes eligible to take a penalized cash distribution.[7]

Let s_t^I and s_t^L respectively be the worker's 401(k) and liquid saving at time t. Additionally, let $V_t(a_t^L, a_t^I)$ be the period-t value function representing the value of optimal behavior

[6] Beshears et al. (2025) and Amador et al. (2006) consider multiplicative marginal-utility shocks $\theta u(c)$, while I consider shocks to the consumption argument c. Moser and de Souza e Silva (2019) consider persistent heterogeneity in earnings ability but not time-varying, within-person shocks.

[7] A simplifying assumption is that job changes occur in predetermined periods rather than stochastically.

from period t onward, taking the two state variables (liquid and illiquid account balances) as arguments. Because any present bias with $\beta < 1$ is naive, the value function is computed assuming that the worker will make optimal choices according to $\beta = 1$ from period t onward. The fully stated problem for the worker at any time t is thus:

$$\max_{c_t, s_t^L, s_t^I} \quad u(c_t) + \beta\delta V_{t+1}(a_{t+1}^L, a_{t+1}^I), \tag{2}$$

$$a_{t+1}^L = a_t^L + s_t^L, \tag{3}$$

$$a_{t+1}^I = (1+r)(a_t^I + s_t^I + \max(0, m * \min(\bar{m}, s_t^I))), \tag{4}$$

with constraints given by:

$$c_t + s_t^L + s_t^I + p_t = y_t, \tag{5}$$

$$a_{t+1}^L \geq \underline{a^L}, \tag{6}$$

$$s_t^I \begin{cases} \geq -a_t^I & \text{if } t \in W \text{ or } t \geq t_r \\ \geq 0 & \text{otherwise,} \end{cases} \tag{7}$$

$$p_t = \begin{cases} \tau s_t^I & \text{if } s_t^I < 0 \text{ and } t \in W \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

2.2 Simple three-period case

Consider first the simplest meaningful case: two working periods, one retirement period, and one job change. Formally, this corresponds to $T = 3$, $t_R = 3$, and $W = \{2\}$.

Now characterize the worker's behavior when making the cashout decision at $t = 2$. Because 401(k) withdrawals are not penalized at $t = 3$, there is no reason to forgo the 401(k) return r by carrying a liquid account balance into $t = 3$. The worker therefore depletes the liquid account and the budget constraint in (5) becomes $c_2 = y_2 + a_2^L - s_2^I - p_2$. A necessary[8]

[8]By taking a 401(k) withdrawal (and therefore not contributing to the 401(k) account) at $t = 2$, the worker forgoes the employer match. This makes possible a case where (9) holds but the employer match is sufficiently generous to make a cashout suboptimal. As the match rate m or match cap $\bar{m}$ approaches zero, (9) becomes both necessary and sufficient for a cashout to occur.

condition for the worker to choose $s_2^I < 0$ and take a penalized 401(k) withdrawal is:

$$u'\left(y_2 + a_2^L\right) > \beta\delta\left(\frac{1+r}{1-\tau}\right)u'\left(y_R + (1+r)a_2^I\right). \tag{9}$$

From (9), we can see that three factors govern the cashout decision:

- *Current liquidity*: When liquid resources $y_2 + a_2^L$ are low, the marginal utility of consumption at $t = 2$ is high, incentivizing the worker to pull consumption forward in time by withdrawing from the 401(k) account. Low liquidity can arise from an exogenous shock to labor income y_2. It can also arise endogenously through a_2^L if the worker chose a low level of liquid saving at $t = 1$.
- *Future liquidity*: When the stock of illiquid retirement wealth $y_R + (1+r)a_2^I$ is large, the worker anticipates relatively low marginal utility of consumption at $t = 3$ and has a stronger incentive to tap 401(k) wealth before retirement.
- *Discount rate*: Withdrawn 401(k) funds forgo the market return r and incur the early-withdrawal penalty τ. Fixing r, τ, and the exponential discount factor δ, the effective discount rate is determined by the present-bias parameter: as β decreases below 1, the cashout condition becomes more likely to hold.

2.3 Expanded case with many periods

2.3.1 Setup and cashout decision

Now consider an expanded case that aligns more closely with the empirical setting of Section 4. The empirical analysis will consider mid-career 401(k) participants who earn income over many periods, then separate from employment and decide whether to cash out. Income is no longer observable in the Vanguard data following separation, but after making their cashout decision workers presumably begin another job and eventually retire.

In an expanded but computationally tractable case, I scale the time horizon by a factor of five, with $T = 15$, $t_R = 11$, and $W = \{5\}$. The periods preceding the cashout decision $t \leq 5$ represent the observable income history during a worker's employment spell at a Vanguard-administered plan. The worker separates and makes a cashout decision at $t = 5$,

earns income at a different job for $6 \leq t \leq 10$, then retires at $t = 11$. Labor income y_t is independently and lognormally distributed with mean zero and standard deviation σ.[9]

The cashout decision at $t = 5$ is driven by the same factors from the three-period case (current liquidity, future retirement liquidity, present bias) but has two important differences. First, the worker's liquidity position at $t = 5$ represents the cumulative effects of many random income draws during the first job. Second, because working periods remain before retirement, cashouts need not fund consumption exclusively in the current $t = 5$ period. If the worker has a low liquid account balance a_5^L and finds the continuation value of marginal liquid wealth to be high, she may cash out some of her 401(k) balance to deposit in the liquid account as protection against income shocks during the second job.

2.3.2 Numerical simulations

Panel A of Table 1 summarizes workers' cashout decisions in the many-period case. For each present-bias value $\beta \in \{0.6, 0.8, 1.0\}$ and each $\sigma \in \{0, 0.1, 0.2, 0.3\}$, I conduct 10,000 simulations.[10] With individual simulations as observations, I estimate regressions (separately for each β value) where the dependent variable is an indicator equal to 1 if the worker takes a penalized 401(k) withdrawal at $t = 5$. The independent variables are σ (representing the volatility of the underlying income process) and realized average income through $t = 5$ (representing the exogenous realizations of the income process).

The results in the top panel of Table 1 are intuitive. First, the regression intercepts decline monotonically with β, indicating that more patient workers have lower baseline cashout rates regardless of income volatility. Second, the coefficients on the average income realization are negative and decline in magnitude with β. This is a generalization of the simpler three-period logic: workers with lower average income during their first job have less liquidity at $t = 5$ and are more likely to cash out. The relationship attenuates for higher β values because patient workers build larger precautionary buffers in the liquid account

[9] Additionally, utility is CRRA with $\gamma = 2$, $\delta = 0.97$, $r = 0.08$, $\tau = 0.1$, $m = 0.5$, $\bar{m} = 0.06$, and $y_R = 0.5$. See Appendix Appendix C for detail on the simulation process.

[10] The literature generally estimates empirical present-bias values to lie in this range; see Lockwood (2020) and Beshears et al. (2025). Very low β values would produce couterfactual empirical predictions. With $\beta = 0$, the worker would never save in the 401(k) account; for small but nonzero β, the worker would save but cash out in every simulation. My empirical income volatility measures lie in the 0-30% range (see Table 4).

Table 1: Cashout determinants in the many-period case of the model

	Dependent variable: cashout probability		
	$\beta = 0.6$	$\beta = 0.8$	$\beta = 1$
Panel A: without borrowing			
Intercept	0.99	0.81	0.51
Std. dev. of income process (σ)	1.33	0.88	0.49
Realized avg. income through job change period	-0.97	-0.83	-0.53
Panel B: with borrowing			
Intercept	1.80	0.74	0.36
Std. dev. of income process (σ)	2.29	1.07	0.49
Realized avg. income through job change period	-1.86	-0.79	-0.39

Note: This table summarizes cashout decisions made by workers in the expanded case of the model with $T = 15$. See Sections 2.1-2.3 for a description of the model. Income y_t is independently and lognormally distributed with mean 0 and standard deviation σ. For each value of the present-bias parameter β and each $\sigma \in \{0, 0.1, 0.2, 0.3\}$, I conduct 10,000 simulations. The table shows regression results where the observation is a simulation and the dependent variable is an indicator equal to 1 if the worker takes a penalized 401(k) withdrawal when switching jobs at $t = 5$. The independent variables are the standard deviation of the income process σ and the worker's realized average income through the job change period (i.e., $\frac{1}{5}\sum_{t=1}^{5} y_t$). Panel A shows results in the case without borrowing ($\underline{a}^L = 0$) while Panel B shows results in the case with borrowing ($\underline{a}^L = -1$). Standard errors are determined arbitrarily by the number of simulations and are therefore omitted.

that provide better insurance against negative shocks. Appendix Figure A.1 explicitly shows liquid account balances at $t = 5$, which increase sharply with β.

The many-period case is conceptually valuable because it allows us to estimate the remaining regression coefficient on σ. This coefficient is positive, indicating that cashouts are more likely when the underlying data-generating process is more volatile (*conditional* on the average income realization through $t = 5$). This relationship holds for two reasons. The first is backward-looking: fixing realized average income, workers facing higher σ values experience greater mean-zero volatility through $t = 5$ and are therefore more likely to deplete their liquid wealth by intertemporally smoothing consumption during the first job. The second reason is forward-looking: workers facing higher σ values anticipate greater volatility during $6 \leq t \leq 10$ and thus perceive a higher continuation value for liquid wealth. The coefficient magnitudes again decline with β because patient workers build larger liquidity buffers that are more likely to withstand volatility during the first job and remain sufficient for the second job.

2.3.3 Allowing borrowing

I have thus far considered a simplified no-borrowing case with the liquid-account floor $\underline{a}^L = 0$. Empirically, many 401(k) participants have access to short-term credit (e.g., credit

cards) and can use both credit and precautionary savings to smooth consumption. To capture this reality and illustrate the effect of credit access on cashout patterns, I close the theoretical discussion by adding borrowing to the model. In the borrowing case, I set the liquid-account floor $\underline{a}^L = -1$, so the worker can borrow up to one period's worth of expected income. If the worker enters period t with a negative liquid-account balance, she must pay an interest cost $d_t = i * |a_t^L|$. I set $i = 0.2$, representing a typical credit-card interest rate and reflecting the reality that interest rates on unsecured debt are generally higher than the nominal 401(k) early-withdrawal penalty $\tau = 0.1$.[11]

Panel B of Table 1 summarizes cashout determinants when borrowing is allowed. The qualitative patterns from the no-borrowing case persist: cashout rates increase with volatility and decrease with realized income, and the magnitudes of these relationships attenuate at higher β values. The model's main predictions thus continue to hold in an environment with short-term borrowing.

The important insight comes from comparing the coefficient magnitudes in Panel B to those from the no-borrowing case. For $\beta = 1$, credit access *weakens* the relationship between income volatility and cashouts, as the regression coefficients on σ and realized income weakly decrease in magnitude. This result is intuitive: introducing another option to smooth income shocks can only decrease rational workers' reliance on cashouts. Alternatively, when the worker is sufficiently present-biased, credit access can *strengthen* the volatility-cashout relationship: for $\beta = 0.6$, both regression coefficients increase in magnitude. This is because debt can itself motivate cashouts. With smaller precautionary savings buffers, present-biased workers overrely on borrowing to smooth shocks during $t \leq 4$ and enter $t = 5$ with suboptimally high levels of debt. The additional incentive to pay off high-interest debt then raises cashout rates relative to the no-borrowing case.

2.4 Key conclusions and empirical predictions

The model demonstrates that stochastic labor income can induce early 401(k) withdrawals, even in the absence of behavioral biases. When the worker receives negative

[11] If the worker carries debt into retirement, her Social Security income y_R is reduced by the debt amount, prorated across the five retirement periods (i.e., $\frac{1}{5} * |a_{11}^L|$).

income shocks, her current liquid resources can become small relative to her illiquid retirement wealth, making her willing to tap 401(k) accounts (and pay early-withdrawal penalties) to support current consumption. Additionally, even if negative shocks are accompanied by positive shocks in other periods and the first-job income draws are symmetrically distributed around their expected value, the resulting mean-zero volatility can deplete the worker's liquidity buffer and prompt an early withdrawal. By assembling a paycheck-level income panel for 401(k) participants, I can test these empirical predictions and evaluate the effects of both symmetric and asymmetric shocks on cashout probability.

The model also accommodates behavioral biases, but allows them to play a more nuanced role in shaping withdrawal decisions. In standard models of naive β-δ preferences, the worker saves in a retirement account at $t = 1$, overestimates her future patience, and unambiguously lowers her lifetime utility by taking an early withdrawal at $t = 2$. Here, a present-biased worker's decision to cash out at $t = 2$ can be optimal *ex post* if an income shock raises her marginal utility of consumption at $t = 2$. But the same withdrawal can be suboptimal in an *ex ante* sense if a non-present-biased worker would have saved enough in the liquid account at $t = 1$ to provide insurance against the income shock. Similarly, high-interest debt raises ex-post cashout incentives, but the debt itself is suboptimal if a rational worker with a larger liquidity buffer would have smoothed consumption without short-term borrowing.

3 Data and summary statistics

3.1 Vanguard administrative data

I use administrative 401(k) recordkeeping data from Vanguard. At the plan level, I observe characteristics of the plan sponsor (i.e., employer) and plan rules. At the participant level, I observe all financial transactions (both contributions and withdrawals), account balances, hire and separation dates, and age. I focus on a subset of plans in which I can accurately infer participants' paychecks, using the methodology described below.[12]

[12] A small number of plans in the sample are highly similar but legally distinct 403(b) plans. For brevity, I refer throughout the paper to the sample as being composed of "401(k) plans."

Table 2: Income inference examples

Likely hourly worker

Date	Contribution amount	Contribution rate	Inferred income
2022-01-14	56.54	0.06	942.33
2022-01-21	47.60	0.06	793.33
2022-01-28	57.91	0.06	965.17
2022-02-04	60.40	0.06	1,006.67

Likely salaried worker with bonus

Date	Contribution amount	Contribution rate	Inferred income
2022-11-11	153.42	0.06	2,557.00
2022-11-25	885.42	0.06	14,757.00
2022-12-09	178.99	0.07	2,557.00
2022-12-23	178.99	0.07	2,557.00

Note: These fictional examples are for illustrative purposes only; they do not represent any particular participant in the data.

3.2 Income inference

Employees contribute to 401(k) plans by deferring money from their paychecks. Usually, these payroll deferrals are specified in proportional terms: at any given time, a 401(k) participant has an "elected contribution rate" C such that C% of their next paycheck will be contributed to their 401(k) account.[13] If we observe both the dollar amount of an employee contribution transaction and the underlying contribution rate, we can infer the size of the corresponding paycheck. For example, if a participant with a 5% contribution rate made a $50 contribution on January 1, we can infer that they received a $1,000 paycheck on January 1. Because participants specify contribution rates as a percentage of *gross* pay, the $1,000 figure is a gross amount representing pay before tax withholding and other deductions.[14]

Table 2 illustrates the simple logic of the income inference process (using fictional transactions). The left-hand panel shows a participant whose contribution amounts vary substantially from paycheck to paycheck while the contribution rate remains constant; the resulting variance in inferred income values likely indicates that this participant is an hourly worker. The right-hand panel shows a participant whose inferred income amounts are usually identical except for one large outlying transaction; this participant is likely a salaried worker who receives infrequent bonuses.

There are two important limitations to consider when inferring income from contribution activity. First, employees' paychecks are not observable unless they contribute to

[13]The contribution rate can be set by active participant choice or by default in automatic-enrollment plans.

[14]This logic holds for both traditional tax-deferred and Roth contributions. In the above example, a 5% contribution rate implies $1,000 in gross pay regardless of whether the contribution is traditional or Roth. A participant making Roth contributions would have $1,000 in taxable income (rather than $950 for a participant making traditional contributions) and would therefore have higher tax withholding and lower net pay.

their 401(k) accounts; the empirical sample is therefore limited to 401(k) *participants* and does not include non-participating employees. Even if an employee is participating in the plan, their contribution transactions may cease abruptly if they separate from employment, voluntarily set their contribution rate to zero, or hit the annual employee contribution limit.[15] As discussed in more detail below, I generally restrict attention to participant-years in which contributions occur in both January and December, which helps ensure that the resulting sequence of paychecks represents the participant's complete income history for the corresponding calendar year. The main consequence of limiting to full calendar years is that I drop some high earners who hit the annual contribution limit and therefore have unobservable paychecks for some portion of the year. I control for annual income throughout the empirical analysis, so dropping these high earners does not affect the estimates I obtain for low- and middle-income participants.[16]

The other limitation of inferred paychecks is that I cannot directly observe the hourly versus salaried nature of the participant's employment contract. To make the binary hourly-salaried classification, I rely on statistics from the payroll-processing data used by Ganong et al. (2024), in which workers are explicitly identified as hourly or salaried. The optimal binary-classification rule in the Ganong et al. (2024) data is to classify a worker as hourly if at least 45% of their paychecks change by at least $1 relative to the lagged paycheck value; this rule accurately classifies about 84% of workers in their data.[17] I adopt this rule throughout the empirical analysis, classifying a participant-year as hourly if at least 45% of the participant's paychecks in the corresponding year change by at least $1. Similarly, I classify a participant as hourly if at least 45% of all of the participant's observable paychecks change by at least $1. I often use the terms "likely hourly" or "likely salaried" to emphasize that contract status is not directly observable. In cases where I simply say "hourly" and "salaried," I do so only for brevity; the classification is always based on the same 45% cutoff.

While indirect, I show that this binary classification produces two groups with very

[15] For example, the limit on employee contributions in 2023 was $22,500. If a participant reached $22,500 in employee contributions in 2023, their contributions were suspended and did not restart until January 2024. Across all Vanguard plans in 2023, 14% of participants reached the limit (Clark, 2025).

[16] Among Vanguard-administered plans, less than 2% of participants with annual income below $100,000 hit the annual contribution limit. See Clark (2025).

[17] I thank Peter Ganong, Pascal Noel, and coauthors for generously sharing these statistics with me.

different monthly income volatility (see Table 4) and roughly matches the share of hourly versus salaried workers reported in external datasets (see Table 5 and the surrounding discussion in Section 3.4). I also obtain similar results when using continuous volatility measures that do not require cutoff assumptions (see Section 4.3).

The recordkeeping data also allow some independent checks of my hourly-salaried classifications. First, in a small number of cases, plans' official legal names give an explicit indication of their participants' contract types (e.g., "XYZ 401(k) Plan for Hourly Employees"). I show in Table A.1 that among plans with explicit hourly names, the 45% rule reassuringly classifies 99.5% of participant-years as hourly. In plans with explicit salaried names, the 45% rule classifies only 21.4% of participant-years as hourly.[18] The plan-name test is useful but limited in scope because less than 2% of participant-years are in plans with an explicit name. Second, I show in Table A.2 that participant-years with weekly pay schedules are classified as hourly 90.9% of the time. The hourly classification share decreases to 55.3% among biweekly and semimonthly pay schedules, and to 32.5% among monthly pay schedules. Since hourly workers tend to have more frequent pay schedules, and some states require hourly workers to be paid on a weekly basis (ADP, 2019), this heterogeneity by pay schedule is to be expected of an accurate classification rule.

3.3 Sample filtering

I study the 2008-2023 period, which is the longest time period for which all necessary components of the administrative data are available. As part of its recordkeeping responsibilities, Vanguard maintains information about the dollar amount of every contribution transaction for every plan. As discussed above, I must also observe the underlying contribution rate in order to infer paychecks. Vanguard tracks participants' contribution rates for about two-thirds of the plans it administers. During the 2008-2023 period, Vanguard tracked contribution rates for 1,655 distinct plans.

For a distinct but highly overlapping subset of about two-thirds of plans, Vanguard performs annual nondiscrimination testing mandated under the Employee Retirement

[18]The 21.4% figure likely overstates the frequency of incorrect hourly classifications, as some plans with "salaried" in their official legal names include non-unionized hourly workers as well as salaried workers.

Table 3: Sample filtering steps

Filtering step	Plans	Participant-years	
		Number	Share of original
Contribution rate and W-2 income observable	1,298	26,051,745	100.0%
Dropped: hire or separation during the year		4,779,859	18.3%
Dropped: hit annual contribution limit or voluntary cessation		3,874,152	14.9%
Contributions in Jan., Dec., and $\geq$ 9 distinct months	1,266	17,397,734	66.8%
Dropped: non-deferrable income > 3% of total		3,252,052	12.5%
Dropped: remainder, including unobservable non-deferrable income		3,571,722	13.7%
Inferred annual income within 3% of W-2 value	1,259	10,573,960	40.6%

Note: The sample period is 2008-2023. I restrict to plans with observable contribution rates so I can infer paychecks according to the methodology discussed in Section 3.2. I further restrict to plans with W-2 income data so I can verify the sum of inferred paychecks against the annual W-2 figure. Non-deferrable income is income from which participants are not allowed to make 401(k) contributions (and is thus impossible to infer from contribution activity).

Income Security Act. As part of the testing process, these plans provide Vanguard with official annual income figures for all employees (those reported on Internal Revenue Service form W-2). Although the empirical analysis focuses on *sub*-annual income volatility, the annual W-2 figures are useful as a final check on the income inference process: for each participant-year, I can sum the inferred paychecks and verify that the inferred annual income amount is sufficiently close to the official W-2 figure. I therefore restrict to the 1,298 distinct plans for which Vanguard tracked contribution rates and collected W-2 income figures during the 2008-2023 period.[19]

Table 3 shows the steps I take to arrive at the final sample of participant-years. The starting point is the universe of 26 million participant-years for the 1,298 eligible plans across the 2008-2023 period. Requiring employee contribution transactions in January, December, and at least nine distinct months drops partial years with hire or separation activity (18.3% of the initial 26 million) and cases where participants hit the annual contribution limit or voluntarily cease contributions (14.9%).[20] Since roughly 14% of Vanguard participants reach the the limit in a given year (Clark, 2025), contribution limits account for most of the latter set of dropped observations and voluntary cessations are rare.

From the intermediate sample of 17.4 million complete calendar years, I restrict further to participant-years where the sum of inferred paychecks is within 3% of the official W-2

[19]The coverage of the W-2 income data is not perfect, even for these 1,298 plans. A plan might appear in the recordkeeping data (giving information on transactions and contribution rates) continuously for some set of years, but only use Vanguard's nondiscrimination-testing service for a subset of those years (depriving me of W-2 income information in the other years). I drop participant-years with missing W-2 income information.

[20]A small number of participants are not paid during the summer, so I use a 9-month cutoff instead of a 12-month cutoff for the purpose of identifying full calendar years.

value. Observations dropped at this final filtering step fall into two groups. First, in roughly 3.3 million cases, the participant receives non-deferrable income that is at least 3% of their total W-2 income. Non-deferrable income is compensation (often annual bonuses) from which employees are not permitted to make 401(k) contributions. Since it is impossible to infer non-deferrable income from contribution activity, dropping these 3.3 million participant-years is a mechanical result of the 3% W-2 accuracy filter. Some firms do not separately report deferrable and non-deferrable income to Vanguard, instead providing only a single total W-2 figure. The remaining 3.6 million observations dropped at the final step thus likely contain many additional (but not explicitly identifiable) cases where non-deferrable income creates a mechanical gap between the inferred and W-2 values.

The final sample of 10.6 million participant-years are cases where I can be highly confident that the sequence of inferred paychecks is a complete and accurate representation of the participant's income history. To alleviate concerns that the observations dropped at the final filtering step (primarily due to non-deferrable income) are selected, I show in the appendix that the main empirical results hold when removing the W-2 accuracy condition and considering the larger sample of 17.4 million complete years.

3.4 Summary statistics

3.4.1 Monthly income volatility

Table 4 shows summary statistics for the final sample of roughly 10.6 million participant-years. The classification rule is based on a paycheck-change cutoff of 45%, but most participant-years classified as hourly have a substantially higher share of paychecks change by at least $1: the 10th percentile among the hourly group is 73%. The average hourly participant sees 92% of their paychecks in a given year change by at least $1; the corresponding average for the salaried group is 16%. Unsurprisingly, annual income is significantly lower for hourly participants: the median is $66,300 for the hourly group and $107,700 for the salaried group (in 2023 dollars).

To measure sub-annual income volatility, I follow Ganong et al. (2024) in focusing on monthly earnings per paycheck. For a given participant-year, I aggregate earnings within

Table 4: Summary statistics by participant-year

	Likely hourly	Likely salaried
Participant-years	7,069,907	3,504,053
Annual income (2023 $)		
10th	33,612	63,328
50th	66,341	107,667
90th	125,854	184,287
Mean	74,640	117,920
Share of paychecks with $\geq$ \$1 change		
10th	73.1%	4.0%
50th	97.5%	13.0%
90th	100.0%	33.7%
Mean	92.1%	16.1%
Median absolute first difference of monthly earnings/paycheck		
10th	3.1%	0.0%
50th	9.3%	0.0%
90th	21.7%	4.6%
Mean	11.5%	2.1%
Standard deviation of log monthly earnings/paycheck		
10th	0.05	0.01
50th	0.12	0.09
90th	0.27	0.29
Mean	0.15	0.13
Skewness of log monthly earnings/paycheck		
10th	-1.12	-1.58
50th	0.25	1.90
90th	1.71	3.01
Mean	0.26	1.24
Annual contribution rate		
10th	2.9%	3.2%
50th	6.1%	6.6%
90th	12.3%	14.2%
Mean	7.1%	8.0%
Share with plan loan origination	13.0%	8.8%
Share with hardship withdrawal	2.5%	0.8%
Hardship withdrawal amount (cond. nonzero)		
Mean	7,648	11,776
Median	3,687	5,481

Note: Participant-years are classified as "likely hourly" if at least 45% of the participant's paychecks in the relevant year change by at least \$1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least \$1 are classified as "likely salaried." See Section 3.2 for a discussion of this classification rule. Annual income figures are given in real 2023 terms.

each month and compute volatility measures (standard deviation, skewness, median absolute first difference) across the resulting sequence of 12 monthly values. Normalizing each monthly earnings total by the number of paychecks eliminates mechanical variation arising from common pay schedules (e.g., participants paid on biweekly schedules receive three paychecks in two months per calendar year). For the standard deviation and skewness

measures, I apply a log transformation and consider the log of monthly earnings per paycheck. This allows convenient proportional interpretations, especially when volatility measures appear in regression specifications in Section 4.

Table 4 confirms that the typical hourly participant experiences greater monthly income volatility than the typical salaried participant. For the average participant-year in the hourly group, the standard deviation of log monthly earnings per paycheck is 0.15, indicating that monthly volatility is roughly 15%. For the salaried group, the corresponding average figure is 0.13. The hourly and salaried figures are more similar than one might expect because roughly a third of the salaried group receive annual bonuses, creating a right tail of salaried participants with large monthly standard deviations (see the discussion of the bonus-identification algorithm in Appendix B). The difference between hourly and salaried volatility becomes clearer when looking at the third moment, i.e., the skewness of log monthly earnings per paycheck. The average hourly skewness value is 0.26, while infrequent positive bonus shocks push the corresponding salaried value up to 1.24.

The median first difference measure creates a cleaner division between the hourly and salaried groups, capturing the fact that fluctuations for salaried workers are rare (though large conditional on occurrence). Letting $e_{i,t}$ be monthly earnings per paycheck for participant i in month t, I compute $\frac{|e_{i,t}-e_{i,t-1}|}{e_{i,t-1}}$ for each month and take the median value within a participant-year. The resulting within-year medians represent participants' typical monthly earnings changes and are not affected by annual bonuses paid in one month per year. The average value for the hourly group is 11.5%, versus 2.1% for the salaried group (and the median salaried participant has a typical monthly earnings change of 0%). Appendix Figures A.2-A.4 show the full distributions of the monthly volatility measures.

3.4.2 Cashout rates

Table 5 summarizes separation and cashout outcomes for the roughly 3.4 million distinct participants in the final sample (i.e., those with at least one yearly observation in the final sample of 10.6 million participant-years). Using the binary-classification rule discussed in Section 3.2, 65% of participants are classified as likely hourly. Because salaried participants are more likely to hit the annual contribution limit and to receive non-deferrable bonuses,

Table 5: Summary statistics by distinct participant

	Likely hourly	Likely salaried
Distinct participants	2,243,079	1,186,705
Share auto-enrolled	45.4%	36.8%
Share separated	51.4%	53.9%
With complete spell observed	17.3%	12.4%
Age at separation		
10th	27	30
50th	44	47
90th	64	64
Mean	45.3	47.3
Tenure at separation		
10th	1	2
50th	5	6
90th	22	24
Mean	8.6	9.8
Share with cashout (cond. separation)	41.8%	21.0%
Cashout amount (cond. nonzero)		
Mean	12,704	17,109
Median	3,849	5,224
Cashout share of balance (cond. nonzero)		
Mean	91.0%	75.9%
Median	100.0%	100.0%
Share of cashouts for entire balance	88.2%	71.5%

Note: Participants are classified as "likely hourly" if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as "likely salaried." See Section 3.2 for a discussion of this classification rule.

this slightly overstates the share of the broader recordkeeping population that is hourly. Among the initial unfiltered sample (i.e., the distinct participants comprising the 26 million yearly observations in the first row of Table 3), I classify 62% of participants as hourly. This share is comparable to, though still slightly higher than, the Ganong et al. (2024) dataset (where 60% of workers are hourly) and nationally representative government statistics (where 56% are hourly; see U.S. Bureau of Labor Statistics 2024).

About half of participants separate from employment during the sample period. Separating participants can do one of three things with their 401(k) balance: they can leave the balance in the plan, roll the balance over to another tax-privileged retirement account (an individual IRA acccount or a new employer's 401(k) plan), or take a cash distribution. They can also combine these options, for example by taking a partial cash distribution and rolling the remaining balance to an IRA. I consider a participant to have cashed out if they take any nonzero cash distribution after their separation date.[21]

[21]The sample period is 2008-2023 but the transaction data are complete through August 2025, so I have an

Table 6: Cashout rates by annual income and paycheck volatility

	Cashout rate		Separated participants	
Annual income (2023 $)	Likely hourly	Likely salaried	Likely hourly	Likely salaried
$\leq$ 25,000	45.8%	38.3%	95,183	1,836
25,001 – 50,000	49.4%	37.7%	367,053	25,676
50,001 – 75,000	43.1%	28.1%	339,693	106,738
75,001 – 100,000	35.6%	22.8%	177,869	146,570
100,001 – 125,000	31.0%	19.3%	82,493	123,906
125,001 – 150,000	26.6%	16.9%	42,477	89,834
$>$ 150,000	22.4%	14.7%	48,840	145,472

Note: This table considers participants who separate during the sample period. Participants are classified by their average annual income across the sample period. Participants are classified as "likely hourly" if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as "likely salaried." See Section 3.2 for a discussion of this classification rule. Annual income figures are in real 2023 terms.

The baseline cashout rates in Table 5 already give evidence of the model's main empirical prediction: conditional on separating, 42% of hourly participants cash out, compared to 21% of salaried participants. The large majority of cashouts are for the participant's entire vested balance, so cashout amounts are substantial in dollar terms: the median value is $3,849 for the hourly group and $5,224 for the salaried group. Of course, hourly workers have lower average incomes, so the difference in baseline cashout rates is driven both by the level and the volatility of income. Table 6, which again shows cashout rates conditional on separation, disentangles these two effects. Cashout rates decline with annual income in both the hourly and salaried groups, but for each income level, hourly cashout rates are significantly higher. In the $25,000-$150,000 range with the most overlap between the hourly and salaried income distributions, hourly cashout rates are 10-15 percentage points higher. In the next section, I turn to more formal tests of the volatility-cashout relationship.

4 The effect of income volatility on early withdrawals

4.1 Within-firm identification

In this section, I present formal regression-based evidence on the relationship between income volatility and early 401(k) withdrawals. All estimates in this section come from

extra 20 months of data to observe cashout transactions for the latest cohort of separators. When identifying cashouts, I only consider total distributions (i.e., cases where the participant's entire balance is distributed through a cash withdrawal, rollover, or combination) and ignore partial withdrawals (e.g., installment payments taken by retirees). Among participants who cash out, 69% cash out within 6 months of separation, 81% cash out within 12 months, and 87% cash out within 18 months.

within-firm regressions that control for plan-level fixed effects. Because my sample contains many large plans with both hourly and salaried participants, I can compare the withdrawal behavior of workers who have been at the same firm for the same period of time and have the same annual income, but who have different degrees of monthly income volatility. Section 5 considers smaller subsamples of workers who move between hourly and salaried employment, permitting alternative within-person identification strategies.

4.2 Empirical specification

I consider two different withdrawal outcomes: one varies at the participant-year level and the other at the participant level. First, at the participant-year level, I consider the probability that a given participant takes a hardship withdrawal in a given year. Hardship withdrawals, which are generally subject to the 10% early-withdrawal penalty, allow participants experiencing financial distress to withdraw 401(k) savings during employment. Because they are subject to relatively strict regulation – participants must provide documentation of a specific financial hardship and may only withdraw an amount necessary to cover the documented hardship – they are rarely used. As shown in Table 4, hardship withdrawals occur in only 2.5% of hourly participant-years and 0.8% of salaried participant-years. Despite being rare, hardship withdrawals are an informative empirical outcome because they allow me to draw a contemporaneous link between income volatility experienced in a given year and penalized 401(k) withdrawals taken in the *same* year.

The second and more important outcome is cashout probability at separation. Theoretically, post-separation cashout decisions are of primary interest because they are the explicit subject of the model discussed in Section 2. Empirically, cashouts are the most important form of leakage because they are the most common form of leakage. Again referencing Table 4, 42% of separating participants in the hourly group and 21% of separating participants in the salaried group take a cash distribution. Cashouts are more prevalent than hardship withdrawals in part because they are less regulated: any participant who has separated from their employer is eligible to take a cash distribution from their 401(k) balance, without providing a specific reason or documentation of financial hardship (though the withdrawal

is still subject to the 10% penalty).[22] Since cashout decisions are a post-separation outcome, the analysis will condition on separation and relate cashout probability to measures of income volatility that are averaged across the participant's employment spell.

Let i index participants, j index plans, and t index years. Additionally, let h_{ijt} be an indicator equal to one if participant i in plan j takes a hardship withdrawal in year t. The regression specification I use to study hardship withdrawals is:

$$h_{ijt} = \sum_k \delta_k * \mathbb{1}\{e_{ijt} \in b_k\} + \sum_k \gamma_k * \mathbb{1}\{e_{ijt} \in b_k\} * v_{ijt} + \rho X_{ijt} + \alpha_j + \theta_t, \tag{10}$$

where e_{ijt} is annual income, b_k are buckets for annual income, v_{ijt} is a measure of sub-annual income volatility, X_{ijt} are covariates, α_j are plan fixed effects, and θ_t are year fixed effects. The coefficients of interest are γ_k, which give the effect of sub-annual income volatility on hardship-withdrawal probability. The estimated effect of income volatility is conditional on nonparametric controls for annual income (via the b_k buckets), quadratics of age and plan tenure (via X_{ijt}), and the plan fixed effects (which absorb industry, plan rules, and any other time-invariant firm characteristics). Interacting the income volatility measure v_{ijt} with the income buckets b_k allows me to estimate heterogeneous effects across the annual income distribution. For the hardship-withdrawal analysis, I use buckets b_k that are \$5,000 wide (using annual income figures converted to real 2023 dollars).

The regression specification for the cashout analysis has the same basic structure but is specified at the participant level. Let c_{ij} be an indicator equal to one if participant i in plan j takes a nonzero cash distribution after separation. Considering only participants who separate during the sample period, the specification is:

$$c_{ij} = \sum_k \delta_k * \mathbb{1}\{\bar{e}_{ij} \in b_k\} + \sum_k \gamma_k * \mathbb{1}\{\bar{e}_{ij} \in b_k\} * \bar{v}_{ij} + \rho X_{ij} + \alpha_j + \theta_t. \tag{11}$$

Average annual income $\bar{e}_{ij}$ is now taken across all observable years for participant i in plan j, as is the volatility measure $\bar{v}_{ij}$. The age and tenure values in X_{ij} are taken as of the

[22] Cashouts are only subject to the 10% early-withdrawal penalty if the participant is younger than 55 at the end of the calendar year in which the separation occurs. I therefore restrict the baseline cashout analysis to those who are younger than 55 at the end of their separation year.

separation year. Because the sample size for (11) is significantly smaller (participant-years are collapsed to participants and non-separating participants are dropped), I expand the income buckets b_k to be \$10,000 wide.[23]

As v_{ijt}, I use the volatility measures introduced in Section 3.4. In some specifications, I let v_{ijt} be simpler, one-dimensional measures of income volatility. I consider both an hourly dummy (defined according to the binary cutoff rule from Section 3.2) and the median absolute first difference of monthly earnings per paycheck (representing typical monthly earnings fluctuations). These measures capture the extensive-margin difference in mean-zero volatility between hourly and salaried workers.

In other specifications, v_{ijt} is two-dimensional and includes both the standard deviation and the skewness of log monthly earnings per paycheck. The two-dimensional case allows me to distinguish the effects of mean-zero and asymmetric volatility. Conditional on skewness, an increase in standard deviation occurs through larger mean-zero shocks; conditional on standard deviation, a decrease in skewness results from an asymmetric shift toward larger negative shocks. Finally, it is worth emphasizing for the cashout regressions that the standard deviation and skewness values are computed within participant-year and then averaged across years to reach $\bar{v}_{ij}$. The $\bar{v}_{ij}$ values therefore represent the average level of within-year volatility, not annual income growth across years.[24]

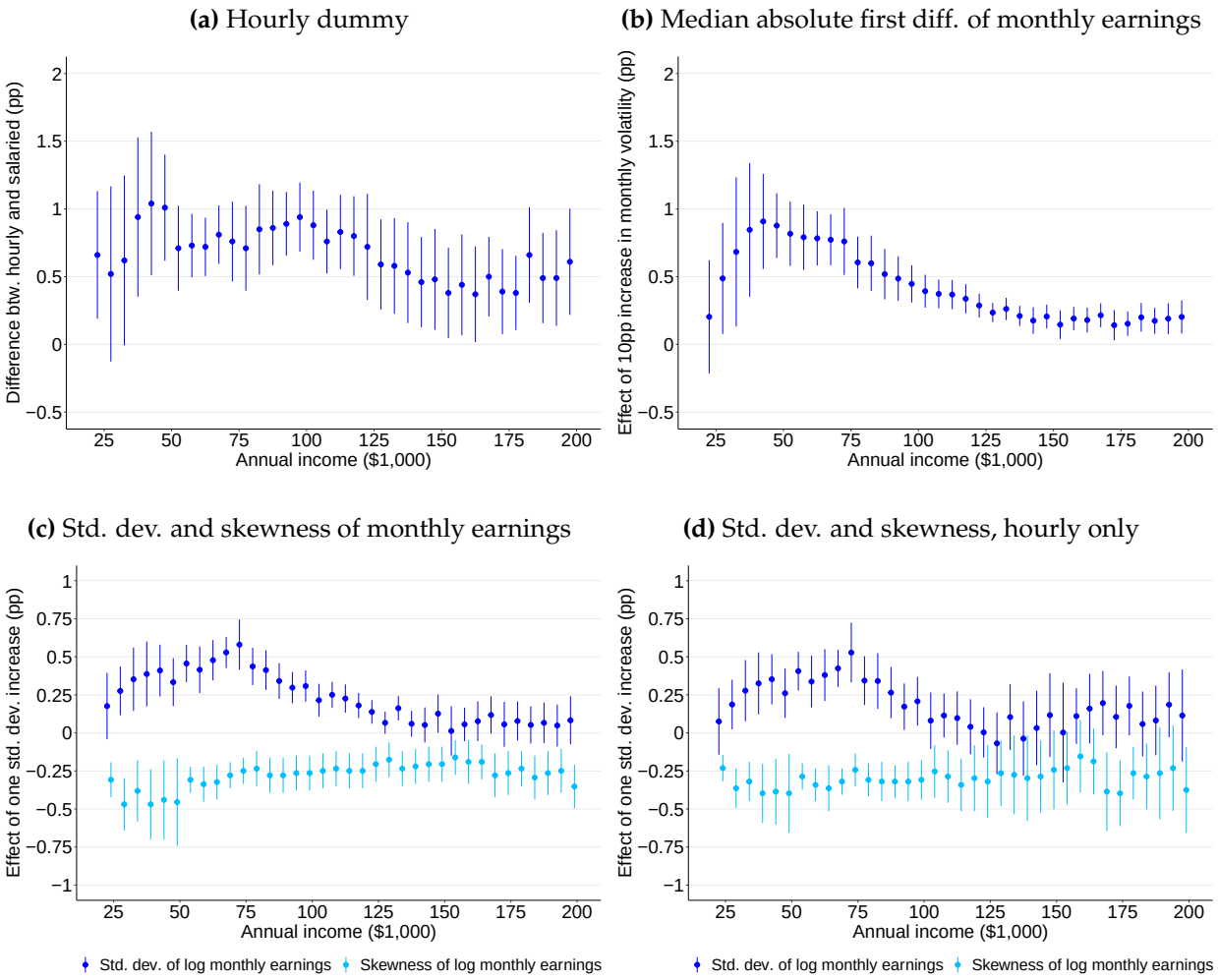
4.3 Results

Figure 1 shows the results of the hardship-withdrawal regressions from (10). The plotted coefficients are the γ_k, which give the heterogeneous effects of income volatility within each annual income bucket. Panel (a) considers the entire sample of participant-years and adopts the simpler specification where the independent variable of interest is a dummy for

[23] A final consideration is that most 401(k) plans execute automatic cashouts for balances smaller than \$1,000 at separation and automatic IRA rollover transactions for balances between \$1,000 and \$5,000. In the X_{ij} vector in (11), I therefore include dummy variables controlling for balances less than \$1,000 and for balances between \$1,000 and \$5,000 at separation. The results are not sensitive to this choice; because I require at least one year of complete contribution history (see Table 3), 97% of participants in my sample have balances above the \$1,000 automatic-cashout threshold at separation.

[24] The median absolute first difference measure compares earnings in consecutive months and is thus minimally affected by annual growth. For the cashout regressions, I simply take the median absolute first difference across all of the participant's observable months (rather than taking the median within each year and averaging across years).

Figure 1: The effect of income volatility on hardship-withdrawal probability



Note: Observations are participant-years and the monthly income volatility measures are as defined in Section 3.4. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (10) includes annual income buckets (interacted with the volatility measures), a quadratic of age, a quadratic of plan tenure, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $5,000 wide and based on values in real 2023 dollars. Standard errors are clustered by plan.

participant-years classified as likely hourly. The coefficients are positive throughout the annual income distribution and indicate that hourly participants (conditional on annual income, age, tenure, and plan fixed effects) have hardship withdrawal rates that are 0.5–1.0 percentage points higher than salaried participants. Given the low baseline hardship-withdrawal rate (1.9% for the overall sample), the coefficients represent substantial proportional effects of 26–53%.

In panel (b), I replace the hourly dummy with the median absolute first difference of

monthly earnings per paycheck. The coefficients in panel (b) are scaled to give the effect of a 10-percentage-point increase, which roughly corresponds to the average difference between the hourly and salaried groups (see again Table 4). The estimates are therefore similar in magnitude to those in panel (a), but using a more precise quantitative measure reduces standard errors. The heterogeneity by annual income is also clearer for the median first difference measure, with the effect of monthly volatility peaking in the \$50,000-\$75,000 range before decreasing steadily toward zero.

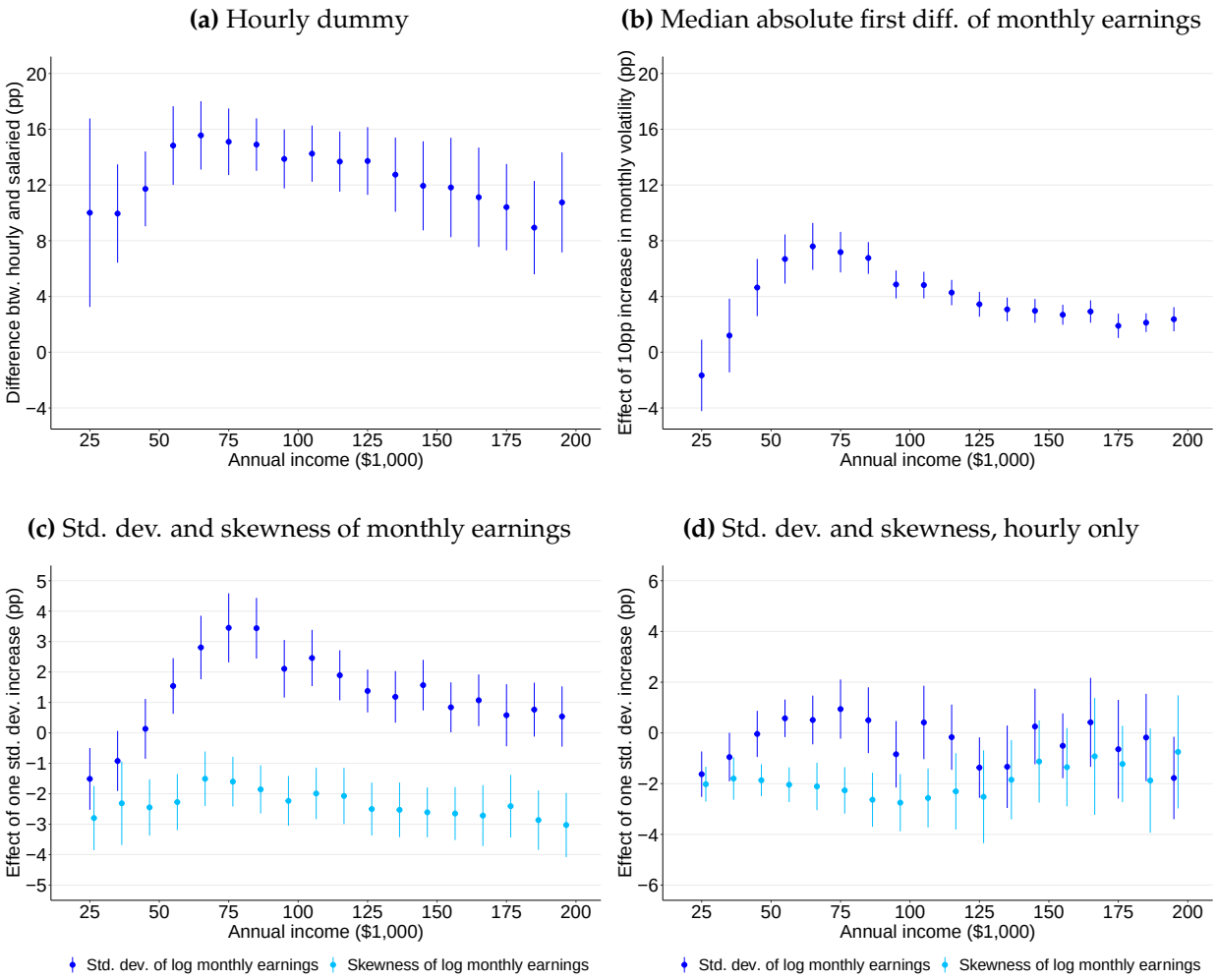
Panel (c) of Figure 1 adopts the fuller specification with both the standard deviation (second moment) and the skewness (third moment) of log monthly earnings per paycheck included on the right-hand side. To put the standard deviation and skewness measures on equal footing, I scale the coefficients so that they each represent the effect of a one-standard-deviation increase in the corresponding measure.[25] The second-moment effects are slightly smaller than those in panel (b) because they now represent the effect of mean-zero volatility conditional on the symmetry of the monthly earnings history. The new third-moment effects are consistently negative, indicating that negative monthly earnings shocks (which decrease the skewness measure) have the strongest effect on hardship-withdrawal rates.

Figure 2 has the same structure but reports the results of the cashout regressions from (11). In the simpler specification with the hourly dummy, participants classified as hourly are consistently 10-15 percentage points more likely to cash out (echoing the cashout rates summarized in Table 6). The median absolute first difference coefficients in panel (b) again peak in the \$50,000-\$75,000 range, where a 10-percentage-point increase in typical monthly volatility (measured across the participant's entire employment spell) raises cashout probability by about 8 percentage points. Given the sample-wide average cashout rate of 34% among separating participants, this is a proportional increase of 24%.

Moving to panels (c) and (d) of Figure 2, the coefficients show that negative income shocks have the largest effect on cashout probability. Considering the full sample in panel (c) and focusing on the peak effect at the \$70,000-\$80,000 bucket, a symmetric perturbation

[25] This requires an unfortunate composition of standard deviations. The dark-blue dots are scaled by the standard deviation of the standard deviation of log monthly earnings per paycheck; the light-blue dots are scaled by the standard deviation of the skewness of log monthly earnings per paycheck. I compute these standard deviations over all participant-years for panel (c) and over hourly participant-years for panel (d). The same logic applies to panels (c) and (d) of Figure 2.

Figure 2: The effect of income volatility on post-separation cashout probability



Note: Observations are participants and the sample is restricted to participants who separate from employment during the sample period. I further restrict to participants for whom cashouts are penalized (i.e., those who are younger than 55 at the end of the calendar year in which they separate). The monthly income volatility measures are as defined in Section 3.4 and are taken across all observable years for a given participant. Participants are classified as hourly if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (11) includes buckets for average annual income during the employment spell (interacted with the volatility measures), a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $10,000 wide and based on values in real 2023 dollars. Standard errors are clustered by plan.

that raised the second moment by one standard deviation but held fixed the third moment would increase cashout probability by about 3.5 percentage points. An asymmetric perturbation that raised the second moment by one standard deviation *and* lowered the third moment by one standard deviation would instead raise cashout probability by about 5 percentage points. When restricting to hourly participants in panel (d) (and therefore conditioning on the higher level of mean-zero volatility among this group), the second-moment

effects shrink toward zero. Among hourly participants, the intensive-margin relationship between income volatility and cashout probability is thus primarily a third-moment effect driven by asymmetric negative shocks.

Taken together, the estimates in Figures 1 and 2 are consistent with the key model predictions from Section 2. For both in-service hardship withdrawals and post-separation cashouts, participants with more volatile incomes are more likely to take penalized early distributions from their 401(k) accounts. The hourly dummy variable and the median absolute first difference measure capture the extensive-margin difference in mean-zero volatility between the hourly and salaried groups. When I allow for both second- and third-moment effects, negative shocks that decrease the skewness of monthly income have the strongest effect on withdrawal probability.[26]

4.4 Heterogeneity by enrollment policy

The effect of income volatility on 401(k) leakage may differ by plan enrollment policy. The set of automatic enrollees (about 40% of participants in the final sample were automatically enrolled; see Table 5) likely includes some workers who would not not have participated in their plan under a voluntary enrollment scheme.[27] If 401(k) contributions crowd out liquid saving among these automatic-enrollment compliers (Choukhmane and Palmer, 2024), we might expect to see stronger withdrawal effects for automatic enrollees.

Focusing on the quantitative median absolute first difference measure, I estimate the regression specifications in (10)-(11) separately for automatic and voluntary enrollees and plot the results in Appendix Figure A.12. With the annual hardship withdrawal rate as the dependent variable, middle-income automatic enrollees do show noticeably stronger effects. The cashout results, however, are statistically indistinguishable between the two groups, with effect sizes for automatic enrollees at most 1-2 percentage points higher (relative to the baseline effect of 8 percentage points in Figure 2b). Since the estimated

[26] Among other additional results in the appendix, Figure A.11 shows regressions where the employee contribution rate is the dependent variable. In contrast to the strong withdrawal effects documented above, income volatility has muted effects on contribution behavior: contribution rates are about 0.25 percentage points lower for middle-income hourly participants than for similar salaried participants.

[27] The share of Vanguard-administered plans using automatic enrollment grew from 20% in 2008 to 59% in 2023. See Clark (2025). I consider participants to be automatically enrolled if their plan had an automatic-enrollment policy on their hire date.

effects remain large and statistically significant among voluntary-enrollment plans, the relationship between monthly income volatility and early withdrawals is not simply a story of automatic enrollment crowding out liquid saving among marginal participants: whatever the underlying mechanism, it is also present for workers who actively choose to participate in their 401(k) plans.

5 Identification from within-person variation

5.1 Identification strategies

The empirical analysis in Section 4 establishes the relationship between monthly income volatility and early withdrawals, controlling for annual income, age, employment tenure, and plan fixed effects. The conclusion is that income volatility correlates with significantly higher cashout rates, even among workers who have the same annual earnings, are at the same point in the lifecycle, have worked at the same firm for the same period of time, and are subject to the same 401(k) plan rules.

The remaining identification threat is the possibility that unobservable worker-level determinants of cashout probability are correlated with monthly earnings volatility. Besides income volatility, the other main factor affecting cashout rates – in the model from Section 2 and the literature generally – is present bias. If hourly workers have greater present bias than salaried workers, then hourly workers would have higher withdrawal rates even in the absence of income volatility, and the coefficient estimates in Figures 1-2 are biased upward in magnitude.

Two points are worth emphasizing. First, the cross-sectional analysis controls for annual income. Prior work suggests present bias is higher among lower-income households (Lawrance, 1991; Paserman, 2008; Meier and Sprenger, 2015), but there is no existing evidence suggesting present bias correlates with income volatility conditional on annual earnings. Second, the identification concern is the *correlation* between present bias and income volatility, not the average *level* of present bias among 401(k) participants. Letting β_h and β_s be the present-bias factors for hourly and salaried workers in a simple two-type case, then $\beta_h < \beta_s$ would cause bias in the estimates from Section 4. If instead $\beta_h = \beta_s < 1$,

then the Section 4 analysis is unbiased, though the effect of income volatility on cashout rates would be smaller in a counterfactual world with $\beta_h = \beta_s = 1$. This is the main insight from the model: if behavioral biases cause the average 401(k) participant to hold too little liquid wealth, then cashouts are overly sensitive to short-run income shocks.

In this section, I consider two alternative identification strategies designed to confront the potential correlation between income volatility and present bias. In each case, I study participants in my sample who make observable changes between hourly and salaried employment, giving within-person variation in income volatility. This allows me to net out the effects of present bias, which evidence suggests is a temporally stable individual characteristic (Meier and Sprenger, 2015). The main tradeoff is sample size: relatively few participants make hourly-salaried changes, so the subsamples here are smaller than the baseline sample in Section 4.

5.2 Within-spell variation

First, I study participants who change between hourly and salaried classifications within a given employment spell. Analyzing single employment spells still leaves me with one cashout outcome per participant, but I exploit variation in the *timing* of the hourly-salaried change and thus in the total amount of time the participant spends as an hourly worker.

As a motivating example, consider two participants a and b who each appear in my sample for three years before separating. Both participants begin as hourly workers and then change to salaried, but participant a becomes salaried in her second year while participant b becomes salaried in her third year. Does the extra year that participant b spends as an hourly worker increase her cashout probability? Because both participants spend at least one year under each classification, it is difficult to argue that their paycheck histories signal different time preferences. As long as present bias does not correlate with the precise timing of hourly-salaried changes, then participants a and b cleanly identify the effect of income volatility on cashout probability.

Generalizing this logic, I consider participants who meet three conditions. First, they appear in the main analysis sample (bottom row of Table 3) in at least three years. I cannot

Table 7: Within-spell changes in hourly/salaried classifications

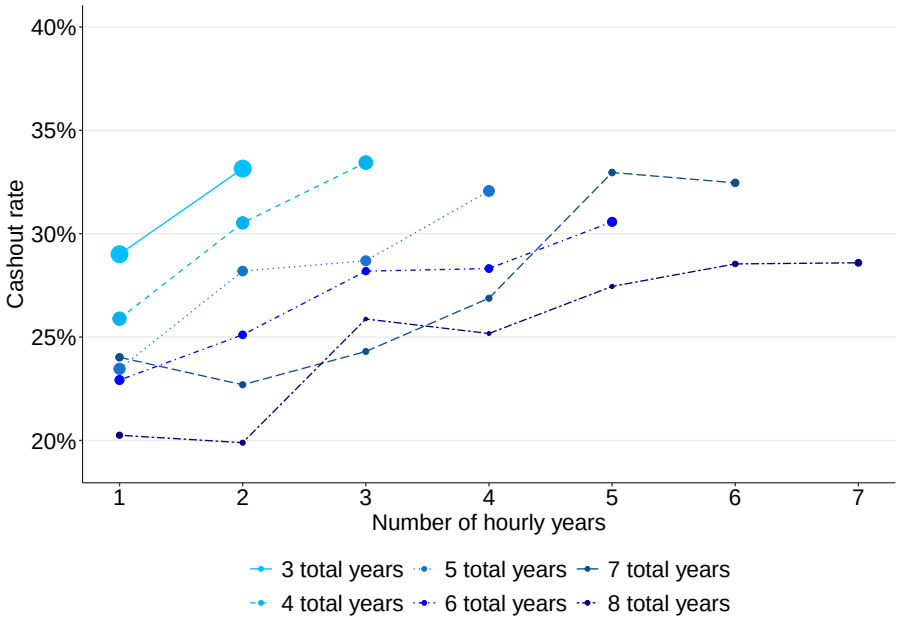
	Hourly to salaried	Salaried to hourly
Number of participants	29,683	21,871
	(57.6%)	(42.4%)
Share of paychecks with $\geq$ \$1 change		
Before change	72.3%	28.9%
After change	26.4%	69.0%
Median abs. first diff. of monthly earnings/paycheck		
Before change	8.9%	4.0%
After change	3.5%	9.0%
Median annual income		
Before change	75,137	74,023
After change	81,421	78,883
Difference (at indiv. level)	4.5%	6.2%
Cashout rate	27.6%	28.5%

Note: Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least \$1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. This table considers the subset of participants who satisfy three conditions. First, they appear in the main analysis sample (bottom row of Table 3) in at least three years. Second, their yearly hourly/salaried classification changes exactly once (i.e., in all years before some year T they are classified as hourly and in all years after and including year T they are classified as salaried, or vice versa). Third, they have separated from employment, making cashout decisions observable. I exploit variation in the number of hourly years conditional on the number of total years (participants with two observable years are excluded because no such conditional variation exists). Annual income figures are in real 2023 dollars.

consider two-year histories because they yield no conditional timing variation (all such histories involve one year of hourly employment and one year of salaried employment). Second, participants' yearly hourly-salaried classifications must change exactly once: in all years before some year T they are classified as hourly and in all years after and including year T they are classified as salaried (or vice versa). To classify participant-years as hourly or salaried, I continue to use the 45% rule (discussed extensively in Section 3.2): if at least 45% of the participant's paychecks in year t change by at least \$1 relative to the lagged paycheck value, then year t is classified as hourly. Third, participants must have separated from employment, so their cashout decisions are observable.

Table 7 summarizes the participants who meet these three conditions. The total sample size is 51,554 participants, with 58% changing from hourly to salaried and 42% changing from salaried to hourly. Most importantly, the 45% rule performs well in capturing discrete changes in paycheck volatility. The share of paychecks changing by at least \$1 increases by about 45 percentage in hourly years, and the median absolute first difference of monthly earnings increases by about 5 percentage points (a proportional increase of more than 100%). One might worry that participants changing from salaried to hourly are being demoted into a lower job rank, but this does not appear to be the case: for both types of

Figure 3: Cashout patterns among participants with hourly/salaried classification changes



Note: Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. See text and note to Table 7 for the sample-selection criteria in identifying hourly-salaried classification changes. I exploit variation in the number of hourly years conditional on the number of total years (participants with two observable years are excluded because no such conditional variation exists). Dots convey the relative sizes of the subsamples. Relatively small subsamples with greater than 8 total years are not shown here but are included in the regression analysis in Table 8.

changes, annual income for the median participant is 4%-6% higher after the change.

Figure 3, which plots cashout rates against the number of hourly years, shows the result graphically. Each line represents a subsample with a different total number of observable years (combining hourly-to-salaried and salaried-to-hourly changes), with dots conveying relative sample sizes. Cashout rates decrease with tenure, so the lines are shifted vertically. But each line slopes upward, confirming that cashout rates increase with time spent in hourly employment. Returning to the three-year example, the additional year participant *b* spends as an hourly worker raises her raw cashout probability from 29% to 33%.

Table 8 gives formal regression estimates. I continue to control for average annual income during the employment spell (though given the smaller sample size I use a simple log transformation rather than the nonparametric dummies in Section 4). I also include fixed effects for the total number of observable years (absorbing the level shifts in Figure 3) and for the calendar year of separation. Using the entire sample, each additional year of hourly employment raises cashout probability by a statistically significant 1 percentage point. The effect size remains similar when considering only hourly-to-salaried changes

Table 8: Regression analysis of participants with hourly/salaried classification changes

	Dependent variable: cashout probability					
	Overall		Hourly to salaried		Salaried to hourly	
	All	Plan events	All	Plan events	All	Plan events
	(1)	(2)	(3)	(4)	(5)	(6)
Number of hourly years	0.010	0.012	0.011	0.011	0.009	0.010
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Controls						
Annual income (avg. during spell)	X	X	X	X	X	X
FE for number of total years	X	X	X	X	X	X
FE for calendar year of separation	X	X	X	X	X	X
Number of participants	51,554	21,402	29,683	11,201	21,871	10,201

Note: Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. See text and note to Table 7 for the sample-selection criteria in identifying hourly-salaried classification changes. I control for the log of average annual income during the employment spell, fixed effects for the total number of observable years, and fixed effects for calendar year of separation. Standard errors are clustered by plan. The "plan events" columns only consider cases where participants change classification in the same year that at least 10% of similarly situated coworkers make the same change. For example, column (4) considers cases where a participant in plan j changes from hourly to salaried in year T *and* at least 10% of all participants at plan j who are classified as hourly in year $T-1$ and remain in the sample in year T change from hourly to salaried in year T.

(1.1 percentage points) or salaried-to-hourly changes (0.9 percentage points).

The estimates also remain stable when I impose a filtering condition to identify plausibly exogenous plan-level movements of workers between the hourly and salaried classifications. Endogenous classification changes are not necessarily a problem for the timing-based identification strategy: even if a worker switches from hourly to salaried because her job performance earned her a new position, the identification assumption is only that she has the same present bias as a similar worker who earned the hourly-to-salaried change *one year earlier*. Even so, plan-level events where many workers make the same change in the same year are particularly interesting cases. In the even-numbered columns of Table 8, I restrict to cases where participants change classification in the same year that at least 10% of similarly situated coworkers make the same change (e.g., 10% of participants at plan j who are classified as hourly in year $T-1$ become salaried in year T). This restriction cuts the sample size by over half, but the estimates are stable and, if anything, increase.

The comparison between Table 8 and the main estimates in Figure 2 is not direct, since the underlying samples are different and the Table 8 analysis is not powered to estimate heterogeneous effects across the annual income distribution. A rough extrapolation would take the overall estimates from Table 8 (1.0–1.2 percentage points per hourly year) and multiply them by a typical tenure length for hourly participants (5 years at the median and 8 years at the mean; see Table 5). This implies an hourly-salaried gap in cashout

rates of 5–6 percentage points for a median-length spell and 8–10 percentage points for an average-length spell, compared to the 10–15 percentage-point effect sizes in Figure 2a (which aggregate over all tenure lengths).

5.3 Between-spell variation

Next, I consider workers who make hourly-salaried changes between distinct employment spells. A subset of workers in the main sample participate in multiple 401(k) plans during the sample period. These are cases in which a worker separates from a job at one Vanguard-administered plan and eventually moves to another job at a different Vanguard-administered plan. The sample of job switchers allows me to observe multiple cashout outcomes for a given participant and therefore estimate cashout regressions with participant-level fixed effects, using differences in monthly income volatility between employment spells to identify the coefficient of interest.

Concretely, I construct the job-switcher sample by identifying the participants who meet three conditions. First, they must appear in the main analysis sample in at least two distinct plans. Second, they must have separated from each plan-level employment spell, making cashout decisions observable. Finally, hire dates in the second and subsequent spells must occur after the separation date of the preceding spell, so that spells are nonoverlapping.

Because it is a relatively rare occurrence to begin and end employment spells at two different Vanguard-administered plans within a fixed time period (and meet all of the additional sample filtering conditions in Table 3 in each spell), the set of observable job switchers is relatively small: it contains 10,835 participants, or 0.32% of the 3.4 million distinct participants in the main sample. The vast majority of participants are observable in exactly two spells. For a subset of 2,162 participants, one spell is classified as hourly and one spell is classified as salaried (again using the 45% rule discussed in Section 3.2). This smaller subset is of primary interest, since it contains within-person variation in the hourly-salaried classification that matters most for cashout probability. Appendix Table A.7 gives additional summary statistics for the job-switcher sample.

Table 9 presents the results of the job-switcher analysis. I estimate regressions where observations are employment spells and the dependent variable is an indicator equal to

Table 9: Regression analysis of subsample with multiple employment spells

	Dependent variable: cashout probability			
	All participants with multiple spells		Participants with hourly and salaried spells	
	(1)	(2)	(3)	(4)
Hourly dummy	0.029		0.032	
	(0.016)		(0.016)	
Median abs. first diff. of monthly earnings		0.033		0.211
		(0.060)		(0.098)
Controls				
Participant FE	X	X	X	X
Year-of-separation FE	X	X	X	X
Annual income	X	X	X	X
Age at separation	X	X	X	X
Observations (spells)	21,911	21,911	4,392	4,392
Participants	10,835	10,835	2,162	2,162

Note: Spells are classified as hourly if at least 45% of the participant's observable paychecks in that spell change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. Observations are employment spells and the dependent variable is an indicator equal to one if the participant took a cash distribution after separating from the employment spell. The regression specification controls for participant fixed effects, year-of-separation fixed effects, the log of average annual income during the employment spell, and age at separation. See Section 3.4 for a discussion of the median absolute first difference measure. Standard errors are clustered by participant.

one if the participant cashed out after separating. I control for participant fixed effects, year-of-separation fixed effects, the log of annual income (averaged across the employment spell), and age at separation (which now also exhibits within-person variation). As the independent variable of interest, I consider a dummy for the spell being classified as hourly, as well as the median absolute first difference of monthly earnings (computed within spell).

Focusing on the subset with both hourly and salaried spells, the volatility-cashout relationship remains positive and statistically significant. The hourly dummy in column (3) implies that cashout rates are 3.2 percentage points higher in hourly spells; the median first difference coefficient in column (4) implies that a 10-percentage-point increase in monthly income volatility raises cashout rates by 2.1 percentage points. These estimates, along with the results of the within-spell analysis from Table 8, help rule out the possibility that the main estimates are driven by a correlation between income volatility and present bias.

6 Survey evidence on liquid savings

6.1 Survey sample

Theoretically, the link between income volatility and 401(k) leakage runs through liquid savings: when workers do not have sufficient liquid wealth to smooth income shocks,

they resort to penalized retirement withdrawals. If the model from Section 2 explains the empirical results, then it should be the case that some 401(k) participants lack liquid savings, and those with low liquidity should be more likely to take penalized withdrawals. Because liquid assets are not observable in 401(k) recorkeeping data, I cannot test these predictions systematically using the main analysis sample.

To evaluate participants' liquidity buffers, I rely instead on survey evidence. I leverage a large survey of Vanguard clients, conducted with randomized email invitations sent in July 2024. The survey posed several questions about household savings and financial well-being, including questions specifically gauging clients' liquid savings. Invitations were sent to both retail brokerage clients and participants in Vanguard-administered 401(k) plans; I analyze only the subset of 401(k) participants. Costa et al. (2025) provide a detailed discussion of the survey design and a fuller analysis of the complete set of questions.

Mirroring the sample-selection criteria from Section 3.3, I consider the set of survey respondents who made employee contributions to their 401(k) accounts in all 12 months of 2024.[28] This yields a sample size of 4,041 respondents, whose survey responses I analyze jointly with 2024 information from the administrative recordkeeping data.[29]

6.2 Results

The survey's primary measure of liquidity is a question that asked:

> *How confident are you that you could find the money to pay for an emergency that costs about $2,000?*

The answer choices were "entirely confident," "somewhat confident," and "not at all confident." Participants could also choose to leave their answer blank (the sample of 4,041 respondents includes those who answered *any* question in the broader survey). The question wording is adapted from leading surveys on household financial well-being (Clark et al., 2021; Sergeyev et al., 2025) and gauges participants' general ability to meet unexpected expenses. The broad wording of the question ("find the money to pay for

[28] Survey invitations were restricted to plans with W-2 income data, so the plan population for the survey is the same population used to construct the main sample. However, W-2 income data are not yet available for 2024, so I cannot impose the W-2 accuracy condition as I do for the main sample (which ends in 2023).

[29] The overall response rate for the survey was 1.66%.

Table 10: Participant demographics and behavior by survey response

		Confidence in meeting $2,000 emergency expense			
	Overall	Entirely confident	Somewhat confident	Not at all confident	Did not answer
Number of survey respondents	4,041	1,233 (30.5%)	569 (14.0%)	554 (13.7%)	1,685 (41.7%)
Annual income in 2024					
10th	41,957	53,345	42,997	37,310	40,630
50th	81,609	103,807	77,741	60,332	76,377
90th	156,302	187,206	138,745	108,992	150,023
Median age	45	46	46	43	45
Median tenure	6	8	6	5	6
Likely hourly	62.6%	42.7%	67.8%	80.5%	69.4%
401(k) behavior in 2024					
Median contribution rate	8.0%	9.9%	7.9%	6.6%	7.7%
Contributed above match cap	67.4%	75.2%	68.5%	57.6%	64.6%
Plan loan origination	21.1%	9.2%	26.5%	35.2%	23.3%
Hardship withdrawal	15.6%	3.6%	15.6%	33.2%	18.5%
Separated in 2025	8.1%	7.9%	8.6%	9.9%	7.4%
Share with cashout (cond. separation)	30.1%	11.3%	40.8%	45.5%	33.6%

Note: This table considers 4,041 participants in Vanguard-administered 401(k) plans who responded to a survey in July 2024 and made employee contributions to their plan in all 12 months in 2024. The column groupings are based on the following survey question (which participants could choose to leave blank): "How confident are you that you could find the money to pay for an emergency that costs about $2,000?" Participants are classified as hourly if at least 45% of their paychecks in 2024 change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. Aside from the column groupings, all information in this table comes directly from the administrative recordkeeping data (i.e., I do not rely on self-reporting in the survey). Separations are observable through August 2025.

an emergency") invites respondents to consider their own liquid savings as well as other sources of liquidity (e.g., short-term credit). Appendix Table A.8 shows similar results for a secondary question that had a lower response rate but asked more narrowly how many months' worth of expenses participants held as emergency savings.

Table 10 summarizes responses to the $2,000 question and linked information from the administrative data. Only 31% of participants reported being "entirely confident" in their ability to meet a $2,000 emergency expense; 14% chose "somewhat confident" and another 14% chose "not all confident." A plurality of 42% left their answer to this question blank. Because the group with blank responses is dissimilar to the "entirely confident" group (in terms of income, hourly share, and withdrawal rates), it appears that many participants with low liquidity preferred not to state this explicitly. The share who are less than fully confident in their ability to access $2,000 is thus at least 28%, but observable characteristics of the missing-response group suggest that 28% is a conservative lower bound.

The low liquidity levels documented in Table 10 are consequential, as many low-liquidity participants have volatile incomes. Of those who chose "somewhat confident" or "not at all confident," 74% are classified as likely hourly based on their 2024 paycheck history. Indeed, Appendix Table A.9 shows that hourly workers (conditional on annual income, age, tenure, and plan fixed effects) were about 16 percentage points less likely to choose "entirely confident."

Another important fact about low-liquidity participants is that their marginal 401(k) contributions are often unmatched. Among the "somewhat confident" and "not at all confident" groups, 63% of participants had a contribution rate in 2024 that was strictly higher than their employer's match cap. The 50%-100% return on matched contributions could rationalize low liquidity for a participant who contributes below the cap (even if they anticipated paying the 10% early-withdrawal penalty with certainty). But most low-liquidity participants do not appear to be making this strategic calculation and could reallocate existing savings flows toward liquid accounts without forgoing match dollars.

Finally and most importantly, liquidity predicts withdrawal behavior. The 2024 hardship-withdrawal rate rises from 3.6% in the "entirely confident" group to 33.2% in the "not at all confident" group.[30] Among those who separated in 2025, 11.3% of the "entirely confident" group cashed out, versus 45.5% of the "not at all confident" group. Appendix Table A.10 shows the differences in withdrawal rates remain large and statistically significant after controlling for age, annual income, tenure, hourly-salaried status, and plan fixed effects.

Given the point-in-time survey responses and lack of exogenous variation in liquid savings, I cannot use the survey data to draw a causal link between liquidity and early withdrawals. Even so, the survey evidence is sufficient to establish a few relevant findings. First, many 401(k) participants – particularly those with volatile incomes – lack liquid savings. Second, many low-liquidity workers nonetheless contribute regularly to 401(k) accounts (recall that all survey respondents analyzed here contributed in all 12 months in 2024) and are often above their employer's match cap. Finally, lower liquidity predicts

[30]It is worth noting that the overall hardship-withdrawal rate in Table 10 is 15.6%, versus 4.8% for all Vanguard participants in 2024 (Clark, 2025). It is possible that participants who recently took a withdrawal (and therefore interacted with Vanguard via website or phone) became more likely to respond to an emailed survey invitation. The survey sample is representative in terms of other observable characteristics (compare Table 10 to Tables 4-5).

higher 401(k) withdrawal rates, consistent with the model's theoretical mechanism.

7 Discussion and policy implications

Monthly income volatility – identified from both within-firm and within-person variation – induces penalized early withdrawals from 401(k) accounts. At the same time, survey responses indicate that many 401(k) participants with volatile incomes lack liquid savings, even though they contribute regularly to illiquid retirement accounts. What are the implications of these findings for policy and plan design?

There are cases in which households with substantial financial wealth could rationally choose to hold little or no liquid savings. Kaplan and Violante (2014) establish one such case: if the welfare benefit from smoothing income shocks is small relative to the return premium on illiquid assets, then households are better off holding zero liquid wealth and consuming their income each period. The hourly workers in my sample do not fit neatly into this framework. First, even during stable employment relationships, they face recurring income shocks that imply large utility gains from consumption smoothing (monthly volatility is on the order of 10%; see again Table 4). Second, over half of those who lack liquidity have already exhausted the strongest 401(k) saving incentive, since they contribute above their employer's match cap (see again Section 6). Tax-advantaged stock market returns provide another incentive to save in retirement accounts, but those returns can also be earned in more liquid vehicles like Roth IRAs.[31] Finally, if households intentionally chose not to smooth income shocks, then we would not observe an empirical relationship between income volatility and early withdrawals.

The model from Section 2 generates the observed relationship between income volatility and cashouts, and shows that this relationship can arise from both rational and present-biased workers. The key difference between the rational and behavioral cases is liquidity. Precautionary savings rise with income volatility among rational workers, but present-biased workers hold too little liquid wealth (and when borrowing is allowed, behavioral workers' liquidity can even *fall* with volatility; see again Section 2.3 and Figure A.1). Given

[31] Contributions to Roth IRAs are not tax deductible but can be withdrawn at any time without penalty (earnings may be subject to taxes and penalties if withdrawn before age 59.5).

that hourly workers in my survey sample have less liquidity, the behavioral case appears to describe the saving and withdrawal choices of at least some 401(k) participants. The model parsimoniously represents behavioral biases as $\beta < 1$, but in reality many things besides pure impatience can depress liquid savings. Workers with low financial literacy may simply lack the budgeting and planning skills needed to maintain a precautionary savings buffer (Lusardi and Mitchell, 2023; Kaiser et al., 2022). Workers may also face difficulty arranging automated payroll deferrals to a dedicated short-term savings account (unlike the automatically recurring payroll deferrals that always fund 401(k) accounts).

A promising intervention is for employers to offer workplace access to – or automatically enroll employees in – short-term savings accounts. Small but automatically reucrring payroll deferrals could be used to build liquidity buffers, and could then be redirected to 401(k) accounts once the liquid account reaches a target balance. This was precisely the motivation behind recent legislation allowing 401(k) plan sponsors to establish pension-linked emergency savings accounts (PLESAs), which are funded with Roth payroll deferrals and capped at a maximum balance of $2,500. Beshears et al. (2020) discuss the practical considerations involved in designing employer-sponsored short-term savings accounts, and recent evidence from U.K. firms (Berk et al., 2023) indicates that automatic enrollment into these accounts significantly increases liquid wealth accumulation.

The policy discussion around retirement savings leakage often focuses on early withdrawal penalties and thus introduces a difficult tradeoff: higher penalties mitigate present bias but impose additional costs on liquidity-constrained households (Beshears et al., 2025; Amador et al., 2006; Beshears et al., 2015; Laibson et al., 1998). If the link between behavioral biases and 401(k) leakage operates at least in part through depressed liquid savings, then policies facilitating liquid wealth formation might help make this tradeoff less severe.

8 Conclusion

Using paycheck-level income data for a large sample of 401(k) participants, I demonstrate a systematic relationship between workers' monthly income volatility and their propensity to take penalized early withdrawals. The empirical results are consistent with a simple model in which workers use (and present-biased workers *overuse*) retirement

assets as insurance against income shocks. Survey evidence confirms that many 401(k) participants with volatile incomes hold low levels of liquid savings, an issue that might be addressed with employer-sponsored interventions facilitating liquid wealth accumulation.

There are promising opportunities for future work. First, linked data on liquid accounts (and credit card balances) would improve upon the point-in-time survey evidence here and allow a complete dynamic analysis of participants' liquidity management. Second, policy experiments can test the potential benefit of short-term savings interventions. Prior research has shown that automatic enrollment into 401(k) accounts raises 401(k) saving, and that automatic enrollment into short-term savings accounts raises short-term saving. Future work should study the interaction effect and determine whether automatic enrollment into short-term savings accounts reduces 401(k) leakage.

References

ADP (2019): "Pay Schedule and Methods: What Are My Options?" https://www.adp.com/spark/articles/2019/03/pay-schedule-and-methods-what-are-my-options.aspx.

AMADOR, M., I. WERNING, AND G.-M. ANGELETOS (2006): "Commitment vs. Flexibility," *Econometrica*, 74, 365–396.

AMERIKS, J., A. CAPLIN, J. LEAHY, AND T. TYLER (2007): "Measuring Self-Control Problems," *American Economic Review*, 97, 966–972.

AMROMIN, G. AND P. SMITH (2003): "What Explains Early Withdrawals from Retirement Accounts? Evidence from a Panel of Taxpayers," *National Tax Journal*, 56, 595–612.

ARGENTO, R., V. L. BRYANT, AND J. SABELHAUS (2015): "Early Withdrawals from Retirement Accounts During the Great Recession," *Contemporary Economic Policy*, 33, 1–16.

BERK, S. H., J. J. CHOI, J. GARG, J. BESHEARS, AND D. LAIBSON (2023): "Automating Short-Term Payroll Savings: Evidence from Two Large U.K. Experiments," Working Paper.

BESHEARS, J., M. BLAKSTAD, J. J. CHOI, C. FIRTH, J. GATHERGOOD, D. LAIBSON, R. NOTLEY, J. D. SHETH, W. SANDBROOK, AND N. STEWART (2024a): "Does Pension Automatic Enrollment Increase Debt? Evidence from a Large-Scale Natural Experiment," Working Paper.

BESHEARS, J., J. J. CHOI, C. CLAYTON, C. HARRIS, D. LAIBSON, AND B. C. MADRIAN (2025): "Optimal Illiquidity," *Journal of Financial Economics*, 165.

BESHEARS, J., J. J. CHOI, J. DICKSON, A. GOODMAN, F. GREIG, AND D. LAIBSON (2024b): "Does 401(k) Loan Repayment Crowd Out Retirement Saving? Implications for Plan Design," Wharton Pension Research Council Working Paper No. 2024-22.

BESHEARS, J., J. J. CHOI, J. HURWITZ, D. LAIBSON, AND B. C. MADRIAN (2015): "Liquidity in Retirement Saving Systems: An International Comparison," *American Economic Review*, 105, 420–425.

BESHEARS, J., J. J. CHOI, M. IWRY, D. C. JOHN, D. LAIBSON, AND B. C. MADRIAN (2020): "Building Emergency Savings through Employer-Sponsored Rainy-Day Savings Accounts," *Tax Policy and the Economy*, 34, 43–90.

BESHEARS, J., J. J. CHOI, D. LAIBSON, AND B. C. MADRIAN (2009): "The Importance of Default Options for Retirement Savings Outcomes: Evidence from the United States," in *Social Security Policy in a Changing Environment*, ed. by J. R. Brown, University of Chicago Press, 167–198.

BESHEARS, J., J. J. CHOI, D. LAIBSON, AND P. MAXTED (2022): "Present Bias Causes and Then Dissipates Auto-enrollment Savings Effects," *American Economic Association Papers and Proceedings*, 112, 136–141.

BUSCH, C., D. DOMEIJ, F. GUVENEN, AND R. MADERA (2022): "Skewed Idiosyncratic Income Risk Over the Business Cycle: Sources and Insurance," *American Economic Journal: Macroeconomics*, 14, 207–242.

CHOI, J. J., D. LAIBSON, J. CAMMAROTA, R. LOMBARDO, AND J. BESHEARS (2024): "Smaller than We Thought? The Effect of Automatic Savings Policies," Working Paper.

CHOI, J. J., D. LAIBSON, AND B. C. MADRIAN (2004): "Plan Design and 401(k) Savings Outcomes," *National Tax Journal*, 57, 275–298.

CHOI, J. J., D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2002): "Defined Contribution Pensions: Plan Rules, Participant Choices, and the Path of Least Resistance," in *Tax Policy and the Economy*, ed. by J. M. Poterba, MIT Press, 67–113.

CHOUKHMANE, T. (2025): "Default Options and Retirement Saving Dynamics," *American Economic Review*, Forthcoming.

CHOUKHMANE, T., J. COLMENARES, C. O'DEA, J. ROTHBAUM, AND L. D. SCHMIDT (2024): "Who Benefits from Retirement Savings Incentives in the U.S.? Evidence on Gaps in Retirement Wealth Accumulation by Race and Parental Income," Working Paper.

CHOUKHMANE, T. AND C. PALMER (2024): "How Do Consumers Finance Increased Retirement Savings?" Working Paper.

CLARK, J. W. (2025): "How America Saves 2025," Vanguard Report.

CLARK, R. L., A. LUSARDI, AND O. S. MITCHELL (2021): "Financial Fragility During the COVID-19 Pandemic," *AEA Papers and Proceedings*, 111, 292–296.

COHEN, J., K. M. ERICSON, D. LAIBSON, AND J. M. WHITE (2020): "Measuring Time Preferences," *Journal of Economic Literature*, 58, 299–347.

COSTA, P., M. MARTINO, AND M. DE LA FUENTE (2025): "The relationship between emergency savings, financial well-being, and financial stress," Vanguard White Paper.

COYNE, D., I. FADLON, AND T. PORZIO (2024): "Measuring Valuation of Liquidity with Penalized Withdrawals," Working Paper.

DERBY, E., K. MACKIE, AND J. MORTENSON (2023): "Worker and Spousal Responses to Automatic Enrollment," *Journal of Public Economics*, 223.

FARRELL, D. AND F. GREIG (2015): "Weathering Volatility: Big Data on the Financial Ups and Downs of U.S. Individuals," JPMorgan Chase Institute Report.

FULFORD, S. L. AND D. LOW (2024): "Expense Shocks Matter," Working Paper.

GANONG, P., D. JONES, P. NOEL, D. FARRELL, F. GREIG, AND C. WHEAT (2025): "Liquid Wealth and Consumption Smoothing of Typical Labor Income Shocks," Working Paper.

GANONG, P. AND P. NOEL (2019): "Consumer Spending During Unemployment: Positive and Normative Implications," *American Economic Review*, 109, 2383–2424.

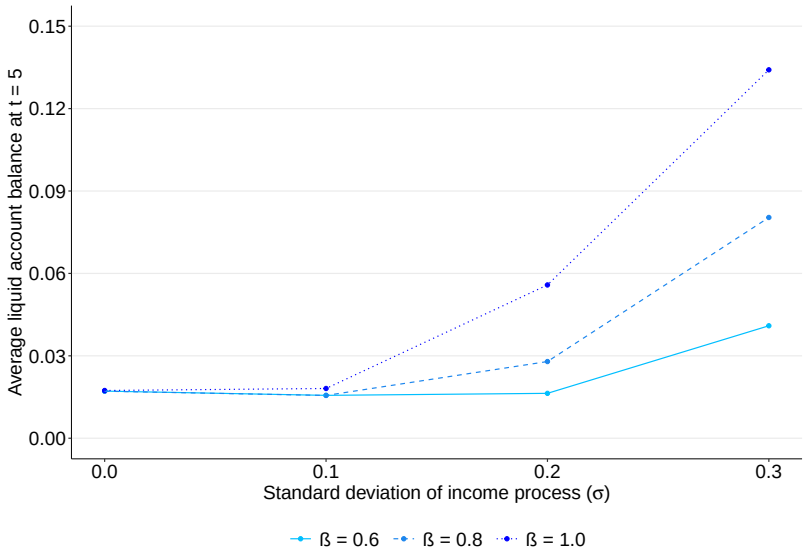
GANONG, P., P. NOEL, C. PATTERSON, J. VAVRA, AND A. WEINBERG (2024): "Earnings Instability," Working Paper.
GANONG, P., P. NOEL, AND J. VAVRA (2020): "US Unemployment Insurance Rates During the Pandemic," *Journal of Public Economics*, 191.
GOODMAN, A. C., J. GOODMAN, L. GOODMAN, AND S. GOODMAN (2015): "A Few Goodmen: Surname-Sharing Economist Coauthors," *Economic Inquiry*, 53, 1392–1395.
GOODMAN, L., J. MORTENSEN, K. MACKIE, AND H. R. SCHRAMM (2021): "Leakage from Retirement Savings Accounts in the United States," *National Tax Journal*, 74, 689–719.
GUVENEN, F., F. KARAHAN, S. OZKAN, AND J. SONG (2021): "What do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica*, 89, 2303–2339.
JACOBSON, L. S., R. J. LALONDE, AND D. G. SULLIVAN (1993): "Earnings Losses of Displaced Workers," *American Economic Review*, 83, 685–709.
KAISER, T., A. LUSARDI, L. MEKNHOFF, AND C. URBAN (2022): "Financial Education Affects Financial Knowledge and Downstream Behaviors," *Journal of Financial Economics*, 145, 255–272.
KAPLAN, G. AND G. L. VIOLANTE (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82, 1199–1239.
KAPLAN, G., G. L. VIOLANTE, AND J. WEIDNER (2014): "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 77–138.
LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–478.
LAIBSON, D., A. REPETTO, J. TOBACMAN, R. E. HALL, W. G. GALE, AND G. A. AKERLOF (1998): "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity*, 91–196.
LAWRANCE, E. (1991): "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, 99, 54–77.
LOCKWOOD, B. B. (2020): "Optimal Income Taxation with Present Bias," *American Economic Journal: Economic Policy*, 12, 298–327.
LUSARDI, A. AND O. S. MITCHELL (2023): "The Importance of Financial Literacy: Opening a New Field," *Journal of Economic Perspectives*, 37, 137–154.
MADRIAN, B. C. AND D. F. SHEA (2001): "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116, 1149–1187.
MEEUWIS, M., D. PAPANIKOLAOU, J. ROTHBAUM, AND L. D. SCHMIDT (2025): "Time-Varying Risk Premia and Heterogeneous Labor Market Dynamics," Working Paper.
MEIER, S. AND C. D. SPRENGER (2015): "Temporal Stability of Time Preferences," *Review of Economics and Statistics*, 97, 273–286.
MOSER, C. AND P. O. DE SOUZA E SILVA (2019): "Optimal Paternalistic Savings Policies," Working Paper.
PASERMAN, M. D. (2008): "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," *The Economic Journal*, 118, 1418–1452.
SERGEYEV, D., C. LIAN, AND Y. GORODNICHENKO (2025): "The Economics of Financial Stress," *Review of Economic Studies*, 92, 3403–3437.
U.S. BUREAU OF LABOR STATISTICS (2024): "Characteristics of Minimum Wage Workers," https://www.bls.gov/opub/reports/minimum-wage/2023/.

Online Appendix

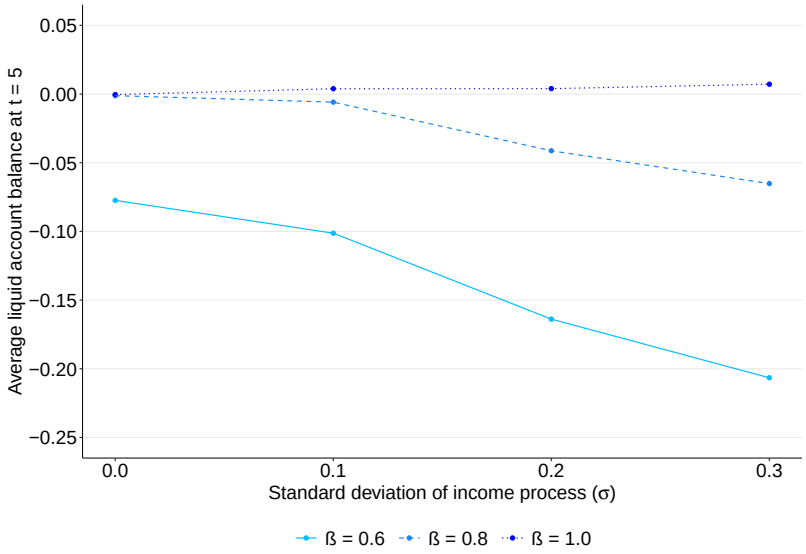
Appendix A Additional tables and figures

Figure A.1: Liquid account balances in the model of cashout behavior

(a) Without borrowing



(b) With borrowing



Note: See Sections 2.1-2.3 for a description of the model. Income in each period is independently and lognormally distributed with mean 0 and standard deviation σ. For each $\sigma \in \{0, 0.1, 0.2, 0.3\}$ and each present-bias parameter $\beta \in \{0.6, 0.8, 1\}$, I conduct 10,000 simulations. The figures show average liquid account balances entering the $t = 5$ period (a_5^L) when workers make cashout decisions. The averages are taken across simulations for each combination of β and σ. Panel (a) shows results in the case without borrowing ($\underline{a}^L = 0$) while Panel (b) shows results in the case with borrowing ($\underline{a}^L = -1$). Small positive values can occur even in the no-borrowing case with $\sigma = 0$ because of discretization of the choice space.

Table A.1: Checking the hourly/salaried classification using explicit plan names

	Participant-years	
Type of plan name	Number	Share classified as hourly
Explicit hourly	121,502	99.5%
Explicit salaried	23,860	21.4%
Rest	10,428,598	66.6%

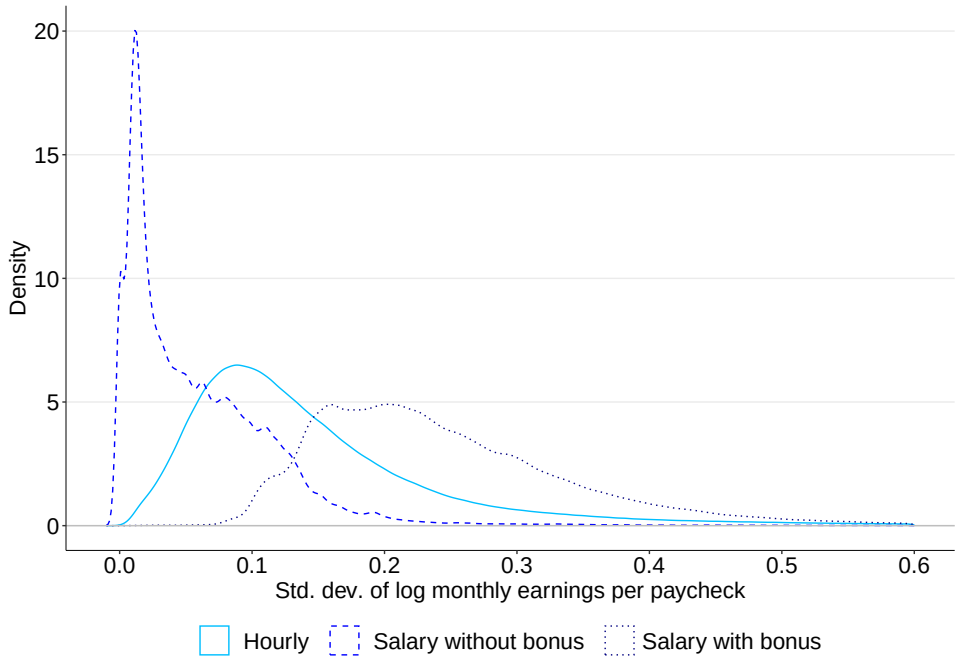
Note: This table considers the final sample of participant-years (bottom row of Table 3). Plans are classified as having an explicit hourly name if their official legal name contains "hourly" or "non-salaried." Plans are classified as having an explicit salaried name if their official legal name contains "salaried" and does not contain "non-salaried." Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule.

Table A.2: Checking the hourly/salaried classification using pay frequency

	Participant-years	
Pay frequency	Number	Share classified as hourly
Weekly ($\geq$ 50 paychecks)	2,856,260	90.9%
Change to/from weekly (29-49 paychecks)	526,509	92.5%
Biweekly/semi-monthly (24-28 paychecks)	6,894,280	55.3%
Monthly (12-13 paychecks)	122,019	32.5%
Rest	174,892	78.5%

Note: This table considers the final sample of participant-years (bottom row of Table 3). The second row is primarily composed of cases in which the participant switched between a weekly pay schedule and some other pay schedule during the year. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule.

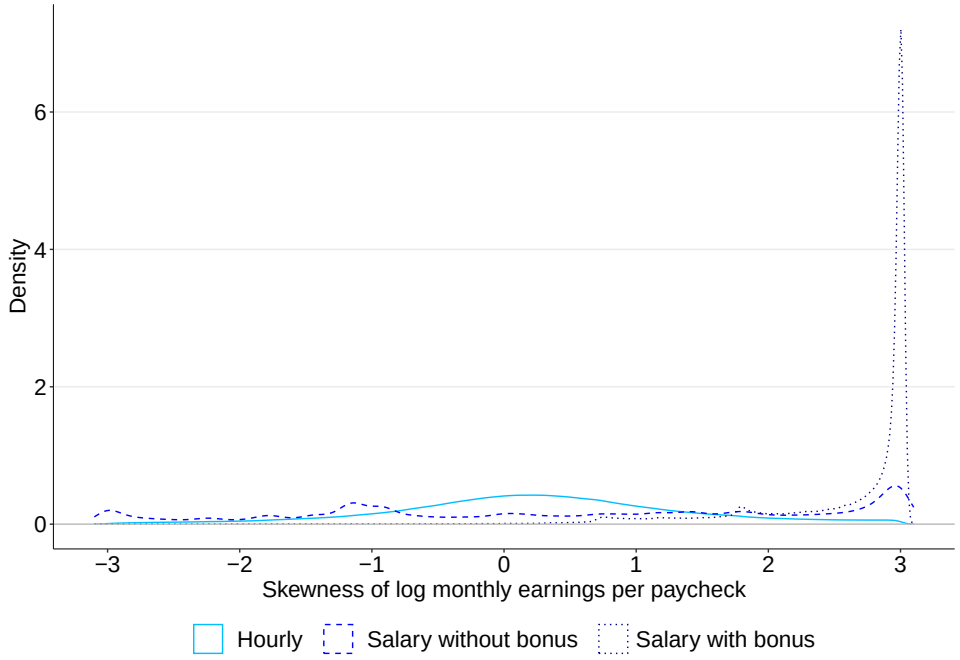
Figure A.2: The standard deviation of monthly income, within participant-year



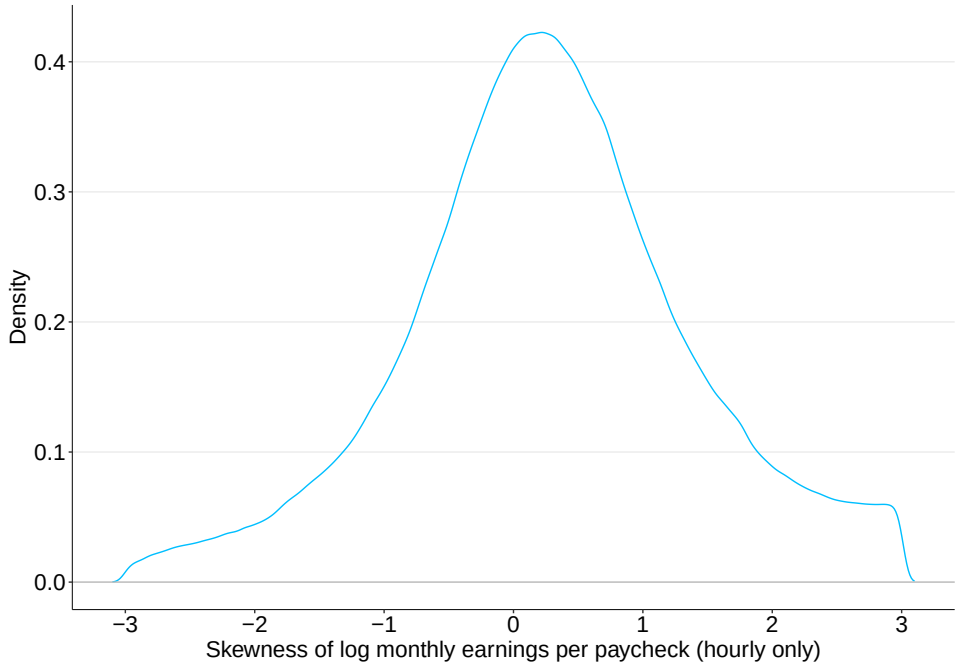
Note: Observations are participant-years; volatility measures computed over monthly values within a given participant-year. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. Bonuses are identified using the algorithm discussed in Appendix B.

Figure A.3: The skewness of monthly income, within participant-year

(a) All participants

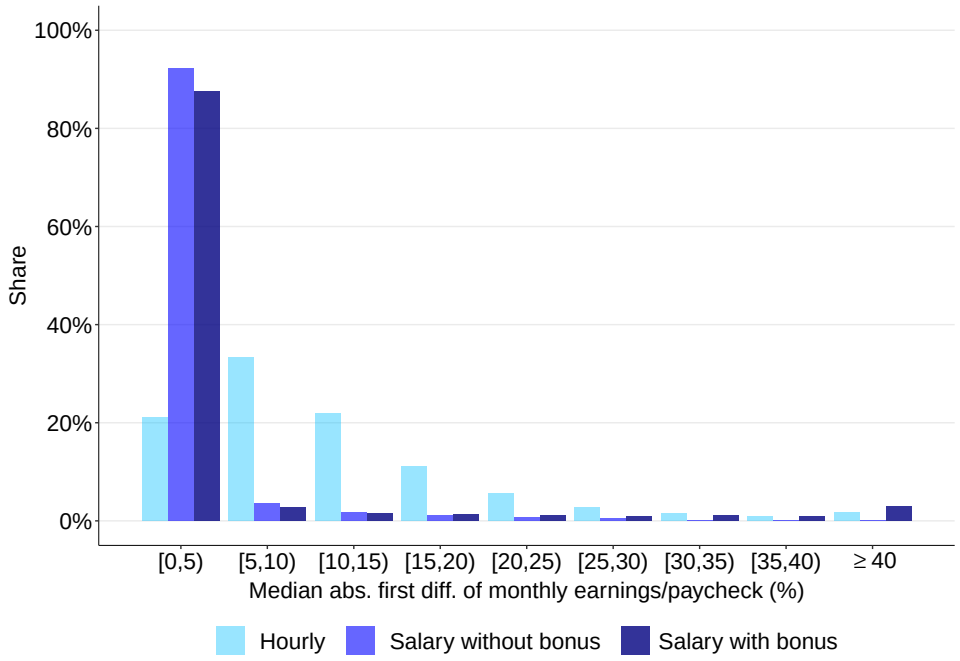


(b) Hourly participants



Note: Observations are participant-years; volatility measures computed over monthly values within a given participant-year. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. Bonuses are identified using the algorithm discussed in Appendix B. If X is a vector of 12 values, 11 of which are identical and strictly less than the 12th, then the skewness of X is approximately 3. Salaried participants with bonuses (who often have 11 very similar monthly income values and an outlying 12th) therefore bunch around 3.

Figure A.4: Median absolute monthly earnings changes, within participant-year



Note: Observations are participant-years. Letting $e_{i,t}$ be monthly earnings per paycheck for participant i in month t, I compute $\frac{|e_{i,t}-e_{i,t-1}|}{e_{i,t-1}}$ for each month and take the median value within each participant-year. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least \$1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least \$1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. Bonuses are identified using the algorithm discussed in Appendix B.

Table A.3: Summary statistics by participant-year, removing the W-2 accuracy condition

	Likely hourly	Likely salaried
Participant-years	10,357,869	7,039,865
Annual income (2023 $)		
10th	33,515	65,354
50th	68,591	114,613
90th	134,558	205,506
Mean	78,842	130,426
Share of paychecks with $\geq$ \$1 change		
10th	70.4%	3.9%
50th	97.3%	12.4%
90th	100.0%	32.2%
Mean	91.5%	15.1%
Median absolute first difference of monthly earnings/paycheck		
10th	3.1%	0.0%
50th	9.5%	0.0%
90th	23.1%	3.5%
Mean	12.0%	1.8%
Standard deviation of log monthly earnings/paycheck		
10th	0.06	0.01
50th	0.13	0.08
90th	0.31	0.30
Mean	0.17	0.12
Skewness of log monthly earnings/paycheck		
10th	-1.18	-2.01
50th	0.23	1.14
90th	1.75	3.00
Mean	0.25	0.78
Share negative	40.3%	37.3%
Annual contribution rate		
10th	2.8%	3.3%
50th	6.0%	7.0%
90th	13.1%	14.8%
Mean	7.3%	8.4%
Share with plan loan origination	13.3%	8.5%
Share with hardship withdrawal	2.5%	0.8%
Hardship withdrawal amount (cond. nonzero)		
Mean	7,743	12,581
Median	3,703	5,702

Note: This table shows the same summary statistics as Table 4, but considers all participant-years in which employee contributions occur in January, December, and at least 9 distinct months. The baseline sample summarized in Table 4 is subject to an additional filter requiring the inferred annual income amount to be within 3% of the official W-2 value. See Table 3 for an accounting of these filtering steps. Participant-years are classified as "likely hourly" if at least 45% of the participant's paychecks in the relevant year change by at least \$1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least \$1 are classified as "likely salaried." See Section 3.2 for a discussion of this classification rule. Monthly income figures are normalized by the number of paychecks to eliminate mechanical variation (e.g., two- versus three-paycheck months). Annual income figures are given in real 2023 terms.

Table A.4: Summary statistics by participant-year, for those dropped by the W-2 accuracy condition

	Likely hourly	Likely salaried
Participant-years	3,287,962	3,535,812
Annual income (2023 $)		
10th	34,244	67,462
50th	72,376	122,157
90th	154,653	225,604
Mean	87,877	142,820
Share of paychecks with $\geq$ \$1 change		
10th	65.1%	3.8%
50th	96.5%	11.7%
90th	100.0%	30.6%
Mean	90.1%	14.2%
Median absolute first difference of monthly earnings/paycheck		
10th	2.9%	0.0%
50th	10.2%	0.0%
90th	26.4%	2.4%
Mean	13.1%	1.5%
Standard deviation of log monthly earnings/paycheck		
10th	0.06	0.01
50th	0.15	0.06
90th	0.38	0.32
Mean	0.20	0.12
Skewness of log monthly earnings/paycheck		
10th	-1.37	-2.35
50th	0.22	0.14
90th	1.83	2.98
Mean	0.22	0.31
Share negative	41.8%	47.9%
Annual contribution rate		
10th	2.9%	3.4%
50th	6.1%	7.3%
90th	14.1%	15.3%
Mean	7.6%	8.8%
Share with plan loan origination	13.8%	8.2%
Share with hardship withdrawal	2.6%	0.9%
Hardship withdrawal amount (cond. nonzero)		
Mean	7,966	13,363
Median	3,702	5,980

Note: This table shows the same summary statistics as Table 4, but considers only participant-years in which: i) employee contributions occur in January, December, and at least 9 distinct months, but ii) the inferred annual income amount is not within 3% of the official W-2 value. These participant-years are dropped from the baseline sample by the final filtering step in Table 3. This table demonstrates that the participant-years dropped at the final filtering step have higher incomes than the baseline sample and are more likely to be classified as salaried, and the salaried observations have lower standard deviation and skewness measures than the baseline sample. These patterns are all consistent with non-deferrable income creating mechanical gaps between the inferred and W-2 values: non-deferrable payments are more likely to be made as bonuses to higher-income salaried participants, and missing these bonus payments causes me to understate the volatility and positive skewness of the dropped participants' paycheck histories.

Table A.5: Summary statistics by distinct participant, removing the W-2 accuracy condition

	Likely hourly	Likely salaried
Distinct participants	2,800,909	1,949,385
Share auto-enrolled	46.0%	36.4%
Share separated	53.0%	53.7%
With complete spell observed	23.9%	19.1%
Age at separation		
10th	27	29
50th	44	46
90th	64	64
Tenure at separation		
10th	1	1
50th	5	6
90th	21	23
Share with cashout (cond. separation)	41.5%	19.9%
Cashout amount (cond. nonzero)		
Mean	12,179	16,353
Median	3,597	4,935
Cashout share of balance (cond. nonzero)		
Mean	91.0%	74.5%
Median	100.0%	100.0%
Share of cashouts for entire balance	88.3%	70.5%

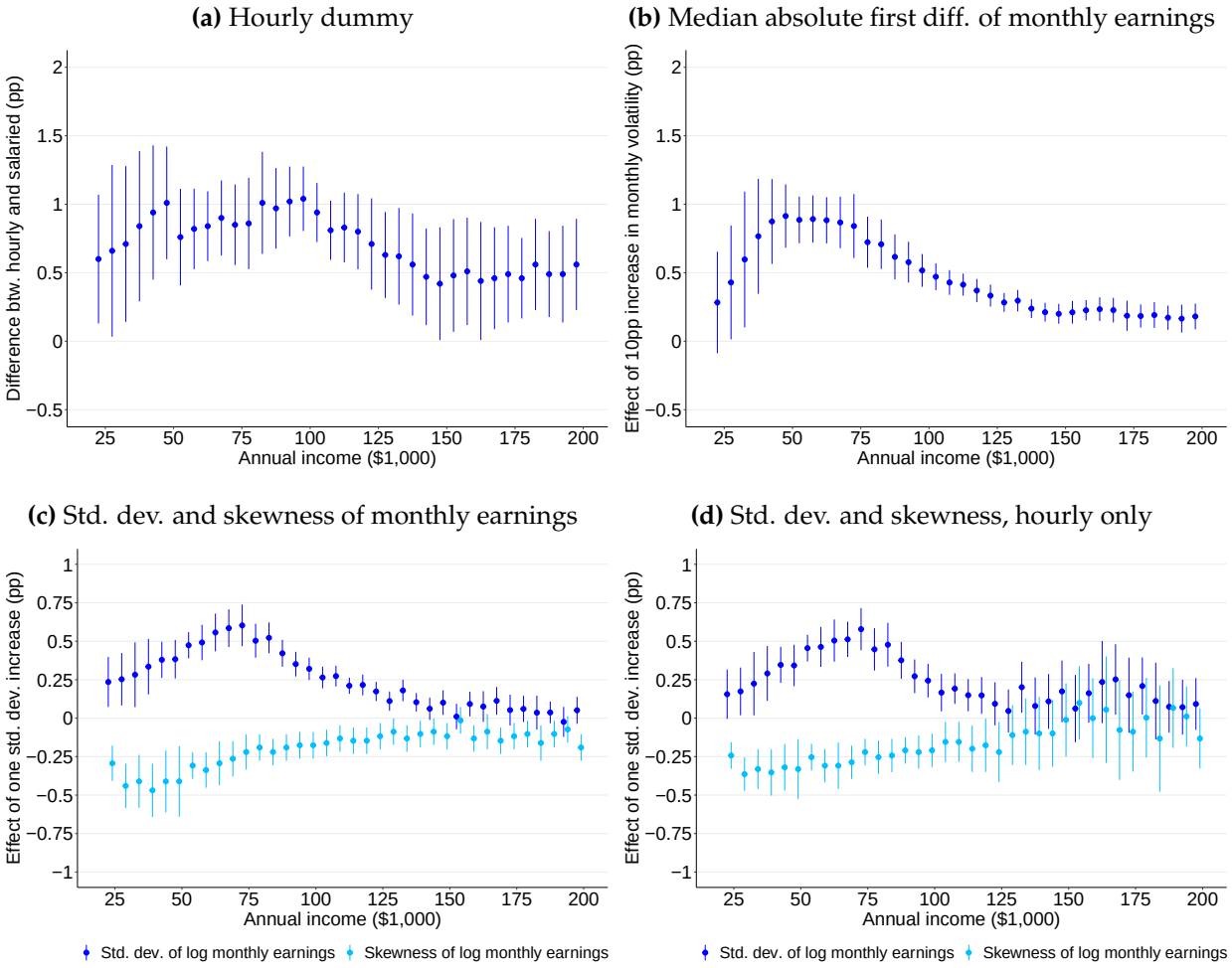
Note: This table shows the same summary statistics as Table 5, but considers all participants who have at least one year in which employee contributions occur in January, December, and at least 9 distinct months. The baseline sample summarized in Table 5 is subject to an additional filter requiring at least one year in which the inferred annual income amount is within 3% of the official W-2 value. See Table 3 for an accounting of these filtering steps. Participants are classified as "likely hourly" if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants where fewer than 45% of paychecks change by at least $1 are classified as "likely salaried." See Section 3.2 for a discussion of this classification rule.

Table A.6: Cashout rates by annual income and paycheck volatility, removing the W-2 accuracy condition

	Cashout rate		Separated participants	
Annual income (2023 $)	Likely hourly	Likely salaried	Likely hourly	Likely salaried
$\leq$ 25,000	46.1%	38.3%	125,789	3,192
25,000 – 50,000	49.3%	38.4%	466,463	39,246
50,000 – 75,000	42.9%	28.4%	416,074	150,814
75,000 – 100,000	36.2%	22.6%	225,229	215,393
100,000 – 125,000	31.4%	19.2%	111,628	188,983
125,000 – 150,000	26.5%	16.3%	59,498	145,758
$>$ 150,000	21.6%	13.4%	81,175	303,242

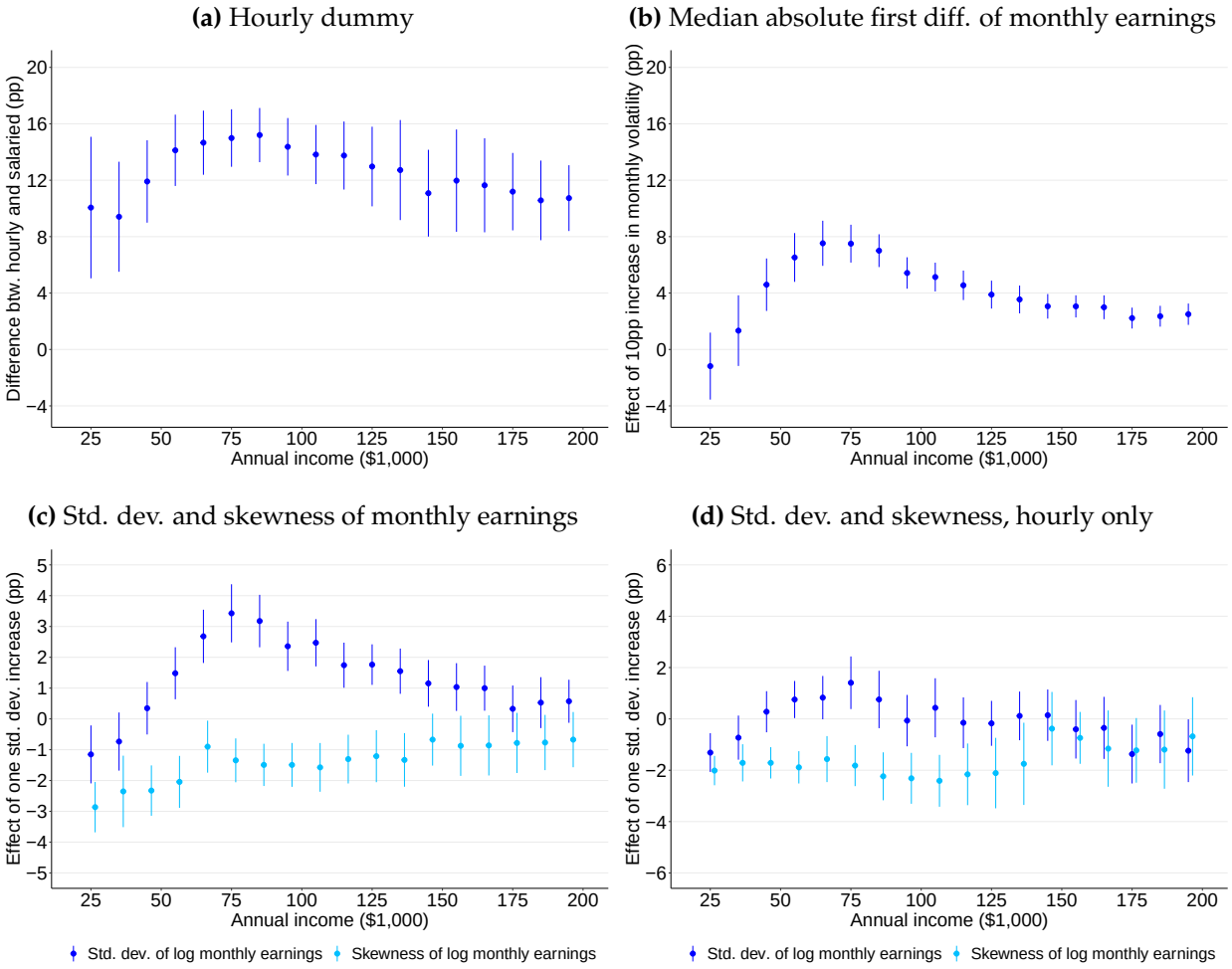
Note: This table shows the same summary statistics as Table 6, but considers all separated participants who have at least one year in which employee contributions occur in January, December, and at least 9 distinct months. The baseline sample summarized in Table 6 is subject to an additional filter requiring at least one year in which the inferred annual income amount is within 3% of the official W-2 value. See Table 3 for an accounting of these filtering steps. Participants are classified by their average annual income across the sample period. Participants are classified as "likely hourly" if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as "likely salaried." See Section 3.2 for a discussion of this classification rule.

Figure A.5: The effect of income volatility on hardship-withdrawal probability, removing the W-2 accuracy condition



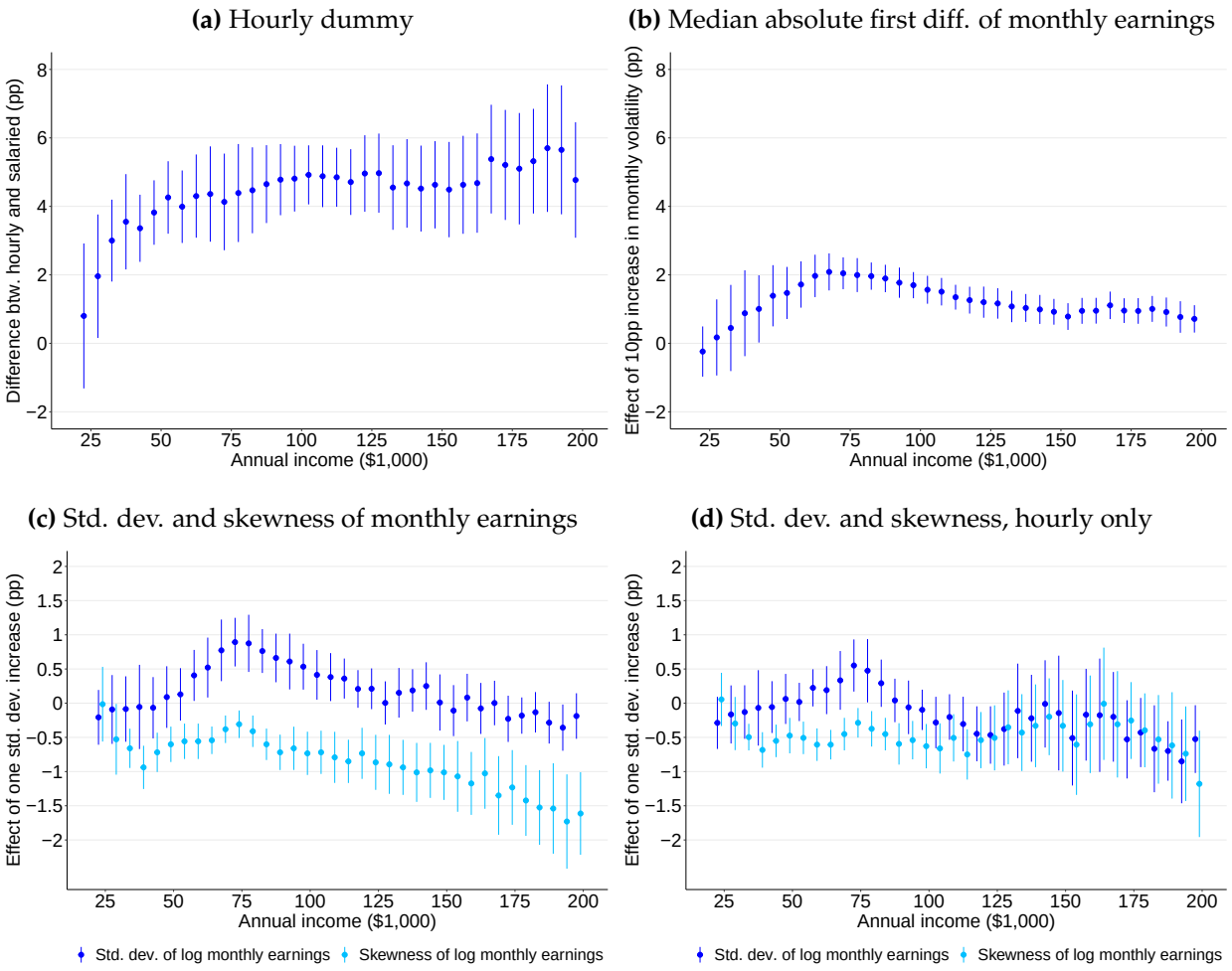
Note: Observations are participant-years and the monthly income volatility measures are as defined in Section 3.4. I consider all participant-years in which employee contributions occur in January, December, and at least 9 distinct months. The baseline analysis in Figure 1 imposes an additional filter requiring the inferred annual income amount to be within 3% of the official W-2 value. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (10) includes annual income buckets (interacted with the volatility measures), a quadratic of age, a quadratic of plan tenure, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $5,000 wide and based on values in real 2023 dollars. Standard errors are clustered by plan.

Figure A.6: The effect of income volatility on post-separation cashout probability, removing the W-2 accuracy condition



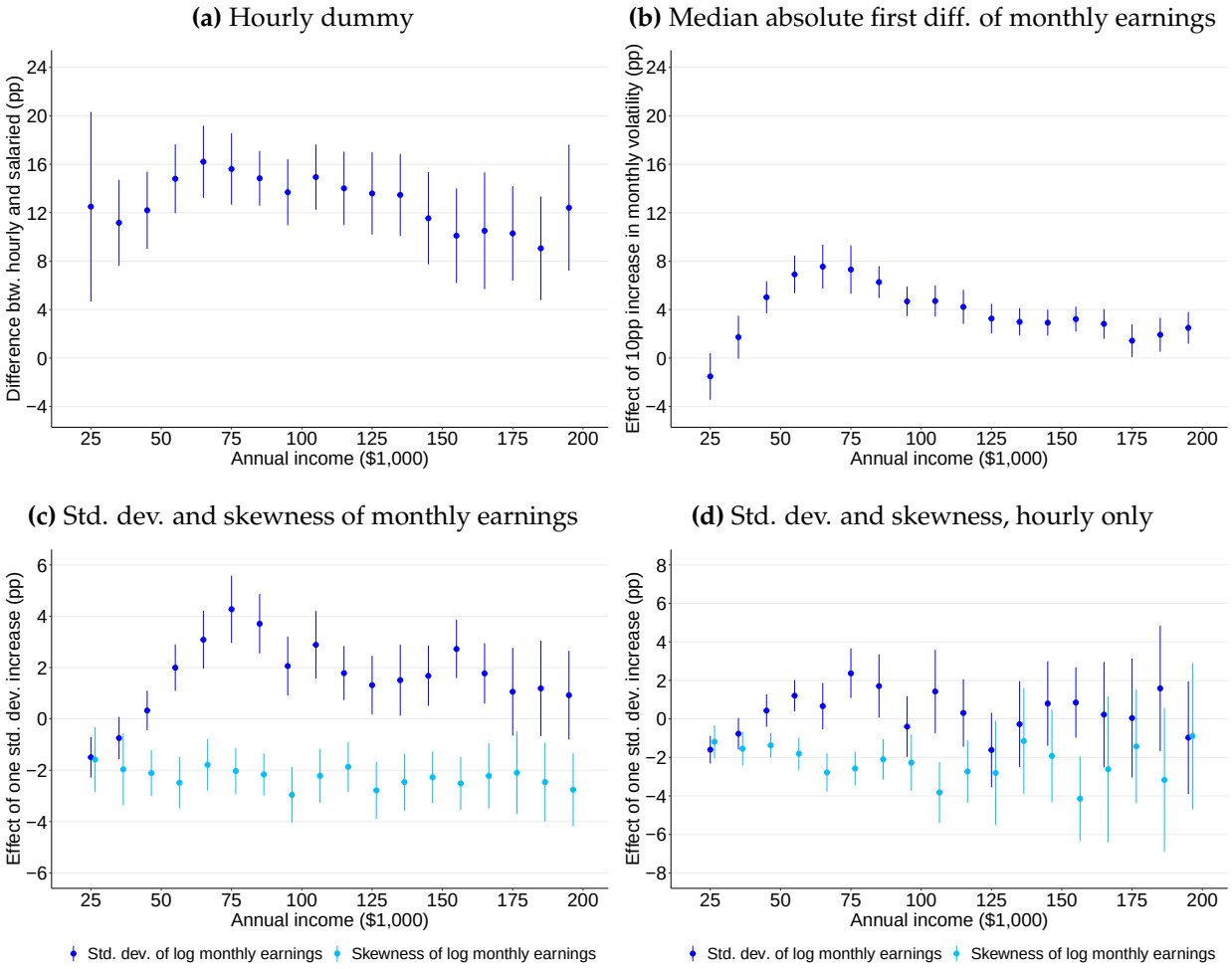
Note: Observations are participants and the sample is restricted to participants who separate from employment during the sample period. I consider all separated participants who are younger than 55 at separation and have at least one year in which employee contributions occur in January, December, and at least 9 distinct months. The baseline analysis in Figure 2 imposes an additional filter requiring at least one year in which the inferred annual income amount is within 3% of the official W-2 value. The monthly income volatility measures are as defined in Section 3.4 and are taken across all observable years for a given participant. Participants are classified as hourly if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (11) includes buckets for average annual income during the employment spell (interacted with the volatility measures), a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $10,000 wide and are based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

Figure A.7: The effect of income volatility on loan-origination probability



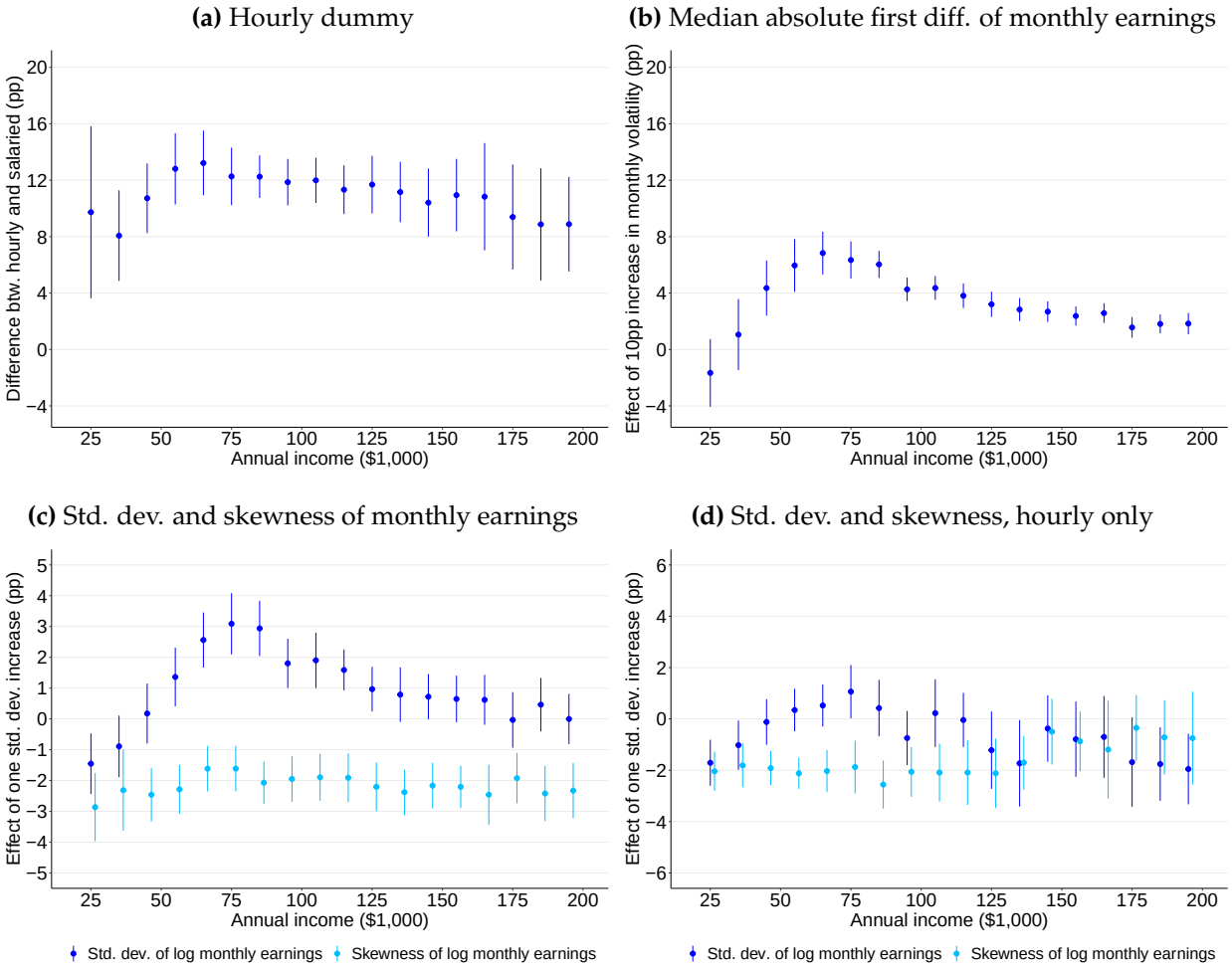
Note: Observations are participant-years and the monthly income volatility measures are as defined in Section 3.4. The dependent variable is an indicator equal to one if the participant originates a 401(k) plan loan in the relevant year. Plan loans allow participants to borrow against the vested balance in their 401(k) account, paying back the account by making additional payroll deferrals over a prespecified repayment period. If a participant fails to make the repayments on schedule, the remaining loan balance is considered a cash distribution on which the participant must pay income taxes and a 10% penalty (the penalty only applies to participants younger than 59.5). Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (10) includes annual income buckets (interacted with the volatility measures), a quadratic of age, a quadratic of plan tenure, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $5,000 wide and based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

Figure A.8: The effect of income volatility on post-separation cashout probability, restricting to balanced panel of participant-years



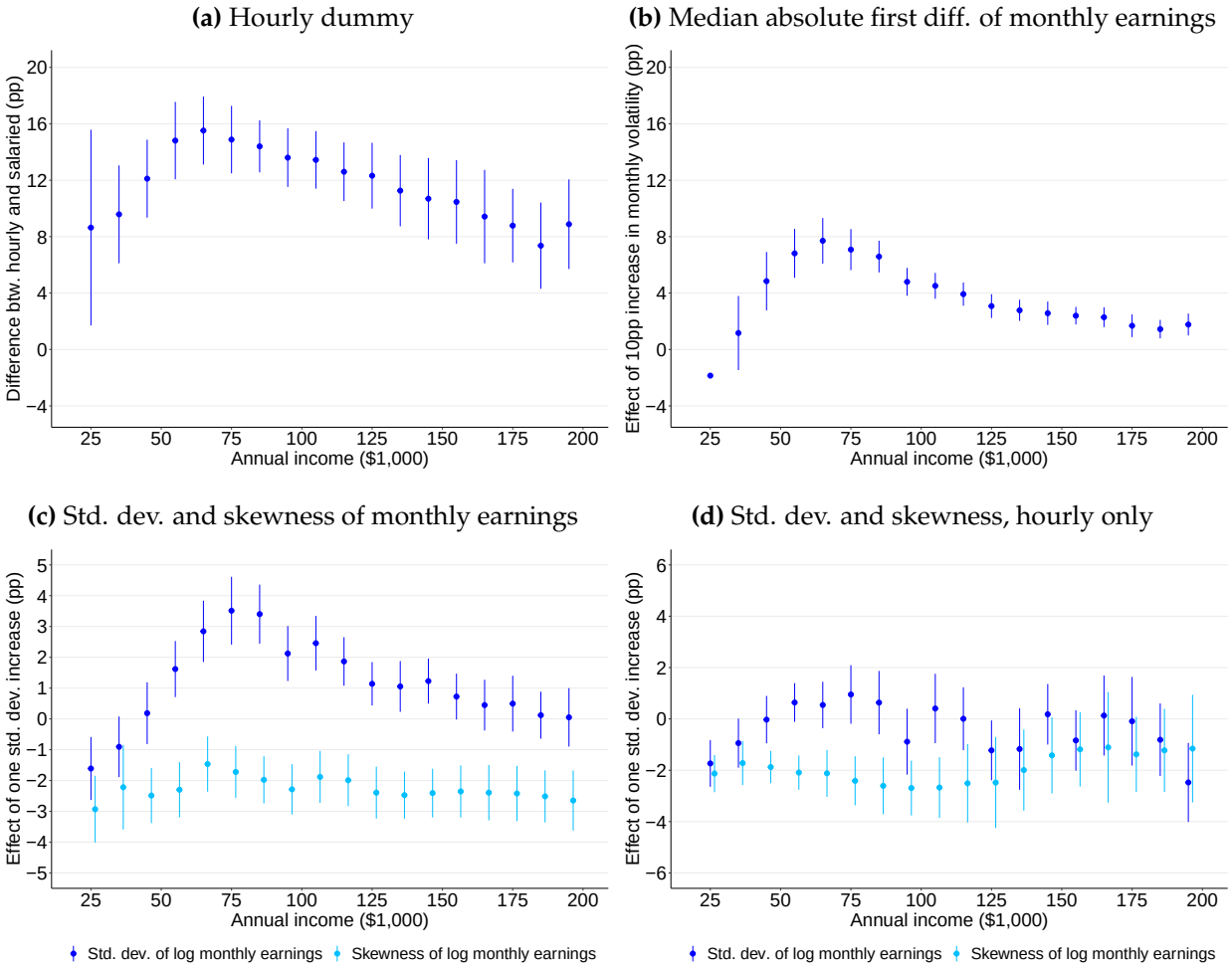
Note: Observations are participants and the sample is restricted to participants who separate from employment during the sample period. I further restrict to participants for whom cashouts are penalized (i.e., those who are younger than 55 at the end of the calendar year in which they separate). Unlike the baseline analysis in Figure 2, I also restrict to participants who appear in the final sample of participant-years for their entire employment spell (i.e., they begin and separate from employment within the sample period and their inferred annual income is within 3% of the corresponding W-2 value for each year of their employment spell). The monthly income volatility measures are as defined in Section 3.4 and are taken across all observable years for a given participant. Participants are classified as hourly if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (11) includes buckets for average annual income during the employment spell (interacted with the volatility measures), a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $10,000 wide and are based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

Figure A.9: The effect of income volatility on post-separation cashout probability, including separators older than 55



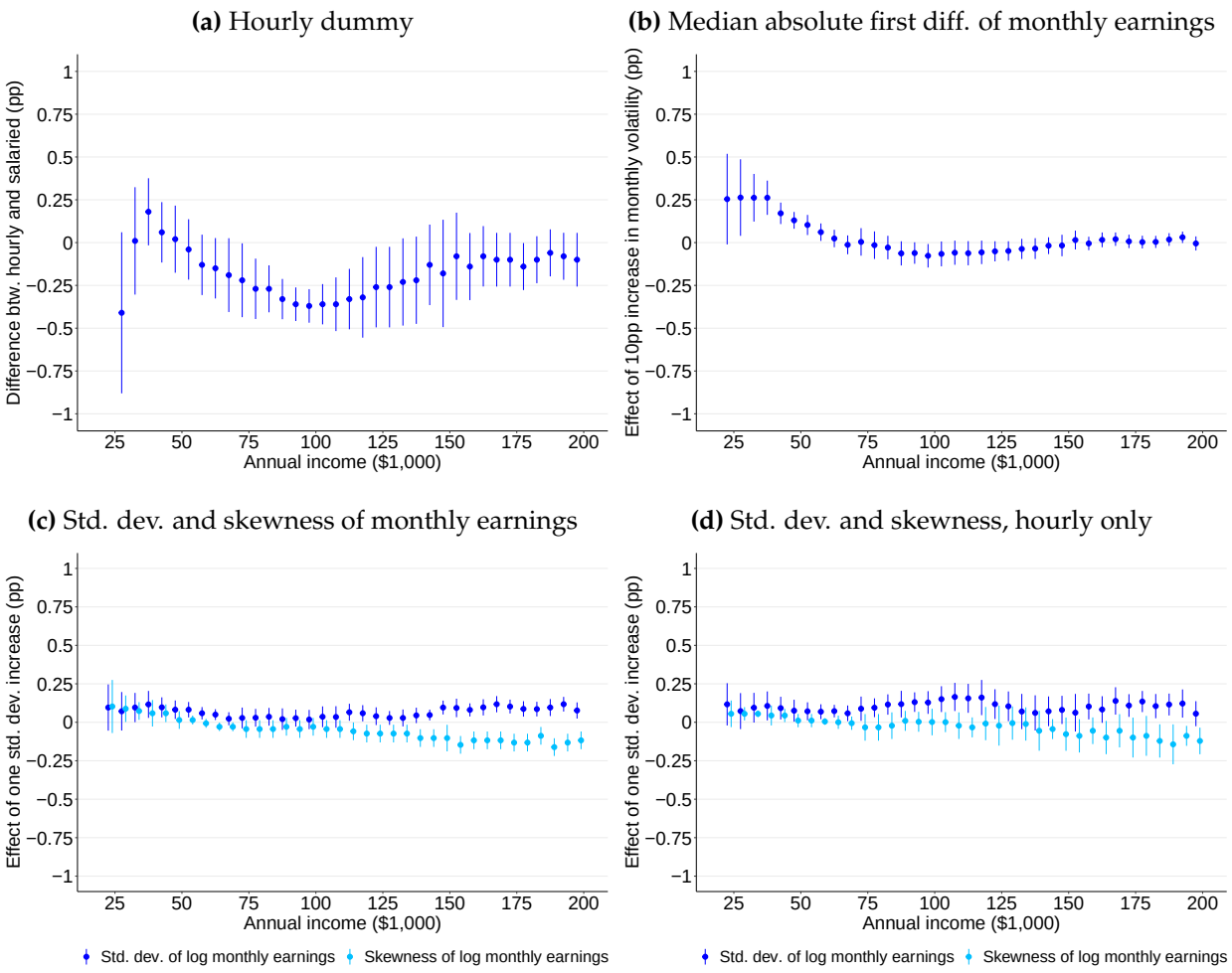
Note: Observations are participants and the sample is restricted to participants who separate from employment during the sample period. Unlike the baseline analysis in Figure 2, I consider all separators, even those who are 55 or older at the end of the calendar year in which they separate and for whom cashouts are not penalized. The monthly income volatility measures are as defined in Section 3.4 and are taken across all observable years for a given participant. Participants are classified as hourly if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (11) includes buckets for average annual income during the employment spell (interacted with the volatility measures), a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $10,000 wide and are based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

Figure A.10: The effect of income volatility on post-separation cashouts, with share of balance cashed out as the dependent variable



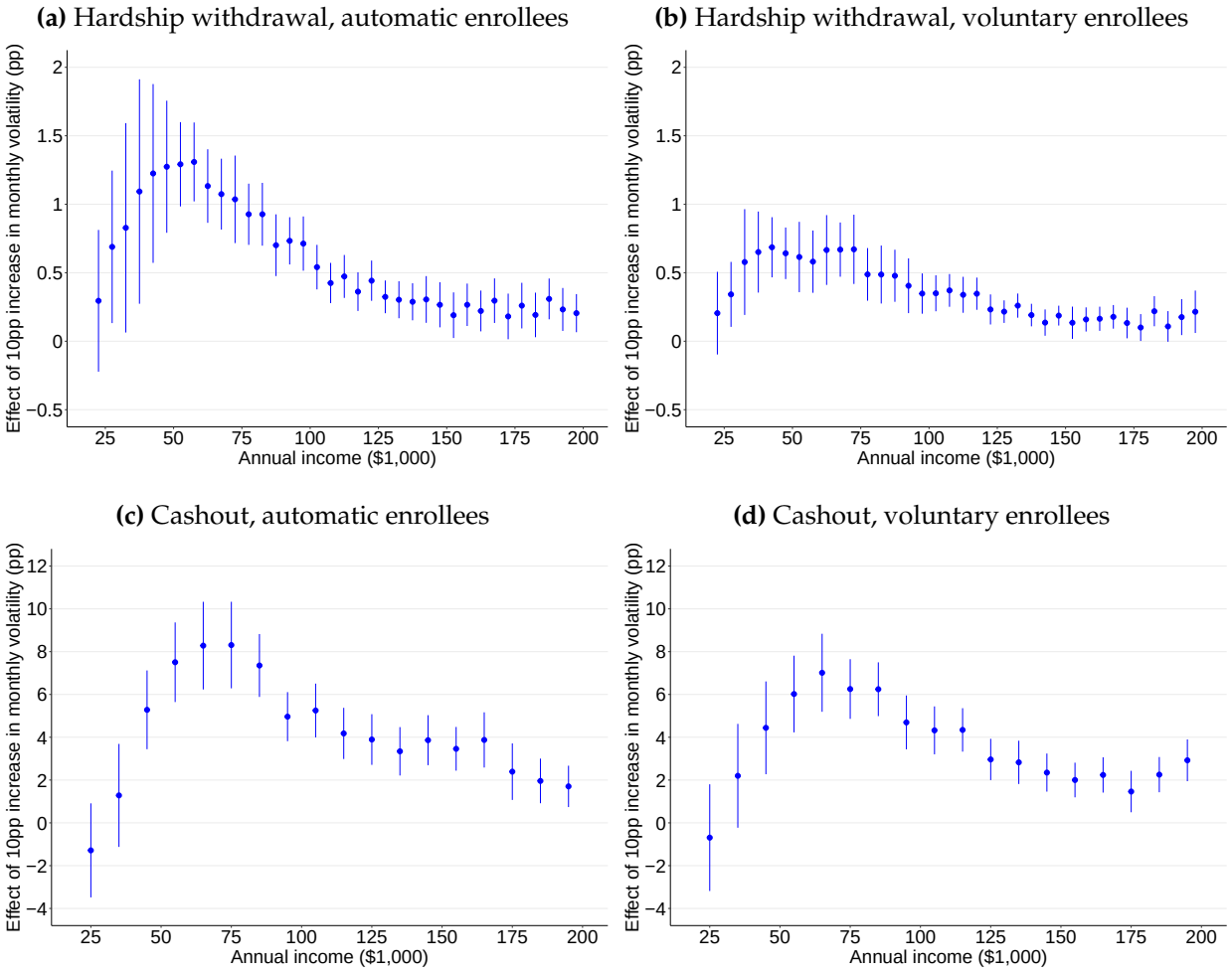
Note: Observations are participants and the sample is restricted to participants who separate from employment during the sample period. The dependent variable is the post-separation cashout amount, expressed as a share of the participant's final balance. If the participant does not take a post-separation cash distribution, the dependent variable is 0. The baseline analysis in Figure 2 uses a binary indicator (equal to 1 if the participant takes a post-separation cash distribution and 0 otherwise) as the dependent variable. The monthly income volatility measures are as defined in Section 3.4 and are taken across all observable years for a given participant. Participants are classified as hourly if at least 45% of their observable paychecks change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (11) includes buckets for average annual income during the employment spell (interacted with the volatility measures), a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $10,000 wide and are based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

Figure A.11: The effect of income volatility on employee contribution rates



Note: Observations are participant-years and the monthly income volatility measures are as defined in Section 3.4. The dependent variable is the participant's annual contribution rate (i.e., total employee contributions divided by annual income). Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The underlying regression specification in (10) includes annual income buckets (interacted with the volatility measures), a quadratic of age, a quadratic of plan tenure, plan fixed effects, and year fixed effects. The coefficient estimates in panels (c) and (d) are scaled by the standard deviation of the corresponding volatility measure. Annual income buckets on the horizontal axis are $5,000 wide and based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

Figure A.12: The effect of monthly income volatility on withdrawal probability, automatic versus voluntary enrollees



Note: The monthly income volatility measure (interacted with the annual income buckets) is the median absolute first difference of monthly earnings per paycheck, as defined in Section 3.4. In panels (a)-(b), observations are participant-years and the annual income buckets are $5,000 wide (in real 2023 dollars). The underlying regression specification in (10) includes annual income buckets, a quadratic of age, a quadratic of plan tenure, plan fixed effects, and year fixed effects. In panels (c)-(d), observations are participants and the annual income buckets (based on average annual income during the employment spell) are $10,000 wide. The sample is restricted to participants who separate from employment during the sample period and are younger than 55 at the end of the calendar year in which they separate. The underlying regression specification in (11) includes buckets for average annual income during the employment spell, a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. Panels (a) and (c) consider only participants who were automatically enrolled and panels (b) and (d) consider only those who were voluntarily enrolled. Standard errors are clustered by plan.

Table A.7: Summary statistics for subsample with multiple employment spells

	All participants with multiple spells	Participants with hourly and salaried spells
Number of participants	10,835	2,162
Within-participant distribution of spells		
Two spells	98%	97%
More than two spells	2%	3%
Hourly spells precede salaried spells		54%
Median annual income		
First spells	80,905	74,880
Second spells	90,138	83,837
Hourly spells	60,715	72,185
Salaried spells	108,375	86,139
Within-participant cashout patterns		
Cashout in all spells	15%	13%
Cashout in no spells	59%	60%
Cashout in at least one but not all spells	26%	27%
In hourly spells only		15%
In salaried spells only		12%

Note: The subsample of participants considered in this table meet all of three conditions. First, they appear in the main analysis sample (bottom row of Table 3) in at least two distinct plans. Second, they have separated from each plan-level employment spell, so cashout decisions are observable. Finally, hire dates in the second and subsequent spells always occur after the separation date of the preceding spell, so the spells are nonoverlapping. Spells are classified as hourly if at least 45% of the participant's observable paychecks in that spell change by at least $1 relative to the lagged paycheck value. The remaining spells in which fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. The rightmost column summarizes the smaller subset of participants for whom at least one spell is classified as hourly and at least one spell is classified as salaried. Annual income figures are in real 2023 terms.

Table A.8: Participant demographics and behavior by survey response (alternate emergency savings question)

	Months of expenses in emergency savings				
	Less than 1 month	1-3 months	3-6 months	More than 6 months	Did not answer
Number of survey respondents	495 (12.2%)	578 (14.3%)	353 (8.7%)	475 (11.8%)	2,140 (53.0%)
Annual income					
10th	43,180	46,127	51,616	55,099	38,950
50th	74,393	89,968	101,834	108,520	72,397
90th	135,448	163,069	182,015	204,276	143,983
Median age	45	46	47	47	44
Median tenure	6	8	7	7	6
Likely hourly	65.7%	52.6%	41.6%	43.6%	72.2%
401(k) behavior in 2024					
Median contribution rate	7.4%	8.6%	10.0%	10.2%	7.3%
Contributed above match cap	62.8%	73.5%	78.2%	77.7%	62.8%
Plan loan origination in 2024	28.9%	16.3%	9.9%	5.1%	25.9%
Hardship withdrawal in 2024	19.6%	7.1%	2.6%	2.1%	22.1%
Separated in 2025	8.1%	8.1%	7.6%	7.4%	8.3%
Share with cashout (cond. separation)	40.0%	19.1%	7.4%	8.6%	38.4%

Note: This table considers 4,041 participants in Vanguard-administered 401(k) plans who responded to a survey in July 2024 and made employee contributions to their plan in all 12 months in 2024. The column groupings are based on the following survey question (which participants could choose to leave blank): "How much money do you have that could be used as emergency savings?" Participants are classified as hourly if at least 45% of their paychecks in 2024 change by at least $1 relative to the lagged paycheck value. See Section 3.2 for a discussion of this classification rule. Aside from the column groupings, all information in this table comes directly from the administrative recordkeeping data (i.e., I do not rely on self-reporting in the survey). Separations are observable through August 2025.

Table A.9: Regression analysis of liquid savings survey responses

	Dependent variable			
	"Entirely confident" in meeting $2,000 expense		At least one month worth of expenses saved	
	All participants	Drop non-responses	All participants	Drop non-responses
	(1)	(2)	(3)	(4)
Likely hourly	-0.157	-0.175	-0.133	-0.079
	(0.022)	(0.031)	(0.023)	(0.028)
Log annual income	0.134	0.213	0.122	0.115
	(0.021)	(0.033)	(0.020)	(0.030)
Controls				
Age	X	X	X	X
Tenure	X	X	X	X
Plan FE	X	X	X	X
Number of participants	4,041	2,356	4,041	1,901

Note: This table considers 4,041 participants in Vanguard-administered 401(k) plans who responded to a survey in July 2024 and made employee contributions to their plan in all 12 months in 2024. The first survey question asked, "How confident are you that you could find the money to pay for an emergency that costs about $2,000?" The answer choices were "entirely confident," "somewhat confident," and "not at all confident" (participants could also leave their answers blank). The dependent variable in columns (1)-(2) is the probability that the participant chose "entirely confident." The second survey question asked, "How much money do you have that could be used as emergency savings?" The answer choices were "less than one month," "between one and three months," "between three and six months," and "more than six months" (participants could also leave their answers blank). The dependent variable in columns (3)-(4) is the probability that the participant chose "between one and three months," "between three and six months," or "more than six months." Columns (1) and (3) code the dependent variable as zero for participants who left their answers blank. Columns (2) and (4) drop participants who left their answers blank. Participants are classified as hourly if at least 45% of their paychecks in 2024 change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. All columns control for quadratics of age and tenure and plan fixed effects. Standard errors are clustered by plan.

Table A.10: Regression analysis of withdrawal behavior among survey respondents

	Dependent variable		
	Hardship withdrawal in 2024	Loan origination in 2024	Cashout (cond. separation)
	(1)	(2)	(3)
Confidence in meeting $2,000 expense			
"Not at all confident"	0.220	0.197	0.304
	(0.027)	(0.045)	(0.128)
"Somewhat confident"	0.082	0.140	0.437
	(0.016)	(0.019)	(0.103)
Did not answer	0.105	0.107	0.249
	(0.013)	(0.015)	(0.063)
Likely hourly	0.069	0.084	0.162
	(0.014)	(0.021)	(0.129)
Log annual income	-0.051	-0.023	0.123
	(0.013)	(0.012)	(0.074)
Controls			
Age	X	X	X
Tenure	X	X	X
Plan FE	X	X	X
Number of participants	4,041	4,041	326

Note: This table considers 4,041 participants in Vanguard-administered 401(k) plans who responded to a survey in July 2024 and made employee contributions to their plan in all 12 months in 2024. The survey question asked, "How confident are you that you could find the money to pay for an emergency that costs about $2,000?" The answer choices were "entirely confident," "somewhat confident," and "not at all confident" (participants could also leave their answers blank). The omitted group is the participants who chose "entirely confident." Column (3) considers the smaller subset of 326 participants who separated from employment in 2025 (separations are observable through August 2025). Participants are classified as hourly if at least 45% of their paychecks in 2024 change by at least $1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least $1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. All columns control for quadratics of age and tenure and plan fixed effects. Standard errors are clustered by plan.

Table A.11: Cashout rates by age, annual income, balance, and hourly/salaried

		Ages 20-29		Ages 30-39	
Annual income	Balance at separation	Hourly cashout rate	Salaried cashout rate	Hourly cashout rate	Salaried cashout rate
$<$ 25k	$<$ 1,000	68%	–	71%	–
	1,000-4,999	34%	26%	44%	39%
	5,000-9,999	33%	–	41%	–
	10,000-49,999	23%	–	26%	–
	50,000-99,999	–	–	6%	–
	$\geq$ 100,000	–	–	–	–
25-50k	$<$ 1,000	74%	–	79%	–
	1,000-4,999	40%	27%	53%	43%
	5,000-9,999	45%	25%	55%	41%
	10,000-49,999	41%	22%	47%	33%
	50,000-99,999	29%	–	29%	25%
	$\geq$ 100,000	–	–	14%	–
50-75k	$<$ 1,000	70%	–	68%	70%
	1,000-4,999	32%	21%	45%	41%
	5,000-9,999	41%	20%	54%	36%
	10,000-49,999	39%	14%	47%	26%
	50,000-99,999	29%	9%	34%	16%
	$\geq$ 100,000	21%	–	24%	11%
75-100k	$<$ 1,000	59%	–	63%	–
	1,000-4,999	30%	20%	52%	35%
	5,000-9,999	35%	17%	47%	32%
	10,000-49,999	32%	12%	41%	22%
	50,000-99,999	21%	7%	31%	14%
	$\geq$ 100,000	16%	4%	19%	9%
100-125k	$<$ 1,000	–	–	60%	–
	1,000-4,999	–	–	47%	33%
	5,000-9,999	41%	14%	46%	27%
	10,000-49,999	31%	8%	40%	20%
	50,000-99,999	23%	7%	28%	12%
	$\geq$ 100,000	22%	6%	18%	7%
125-150k	$<$ 1,000	–	–	–	–
	1,000-4,999	–	–	–	–
	5,000-9,999	–	–	33%	33%
	10,000-49,999	24%	10%	32%	18%
	50,000-99,999	20%	6%	26%	11%
	$\geq$ 100,000	14%	7%	16%	7%
$\geq$ 150k	$<$ 1,000	–	–	–	–
	1,000-4,999	–	–	–	–
	5,000-9,999	–	–	–	20%
	10,000-49,999	18%	7%	26%	14%
	50,000-99,999	16%	6%	20%	9%
	$\geq$ 100,000	15%	4%	13%	7%

Note: This table shows detailed cashout rates for separated participants in the final sample (i.e., participants with at least one year of complete contribution history in which the inferred annual income amount is within 3% of the W-2 value). Annual income is averaged across participants' employment spells and is in real 2023 dollars. Age and balance are as of separation. Participants are classified as hourly if at least 45% of their observable paychecks change by at least \$1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least \$1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. Cells are left blank if they contain less than 100 participants.

Table A.12: Cashout rates by age, annual income, balance, and hourly/salaried (continued)

		Ages 40-49		Ages 50-59	
Annual income	Balance at separation	Hourly cashout rate	Salaried cashout rate	Hourly cashout rate	Salaried cashout rate
$<$ 25k	$<$ 1,000	72%	–	71%	–
	1,000-4,999	47%	–	42%	–
	5,000-9,999	45%	–	40%	–
	10,000-49,999	32%	–	30%	–
	50,000-99,999	13%	–	10%	–
	$\geq$ 100,000	5%	–	6%	–
25-50k	$<$ 1,000	81%	–	73%	–
	1,000-4,999	55%	50%	51%	43%
	5,000-9,999	60%	49%	54%	44%
	10,000-49,999	52%	44%	47%	39%
	50,000-99,999	32%	26%	30%	25%
	$\geq$ 100,000	16%	18%	14%	19%
50-75k	$<$ 1,000	69%	–	55%	–
	1,000-4,999	48%	49%	44%	49%
	5,000-9,999	60%	47%	54%	43%
	10,000-49,999	54%	38%	49%	37%
	50,000-99,999	40%	27%	36%	26%
	$\geq$ 100,000	25%	16%	22%	17%
75-100k	$<$ 1,000	58%	–	51%	–
	1,000-4,999	48%	45%	45%	42%
	5,000-9,999	53%	43%	50%	42%
	10,000-49,999	50%	36%	46%	33%
	50,000-99,999	40%	24%	36%	24%
	$\geq$ 100,000	26%	14%	23%	16%
100-125k	$<$ 1,000	59%	–	43%	–
	1,000-4,999	48%	45%	43%	37%
	5,000-9,999	45%	38%	44%	35%
	10,000-49,999	44%	30%	43%	28%
	50,000-99,999	38%	21%	34%	21%
	$\geq$ 100,000	22%	12%	21%	14%
125-150k	$<$ 1,000	–	–	–	–
	1,000-4,999	–	–	–	–
	5,000-9,999	50%	35%	–	38%
	10,000-49,999	40%	25%	35%	25%
	50,000-99,999	32%	19%	32%	21%
	$\geq$ 100,000	18%	10%	19%	12%
$\geq$ 150k	$<$ 1,000	–	–	–	–
	1,000-4,999	–	–	–	–
	5,000-9,999	–	32%	–	25%
	10,000-49,999	34%	22%	32%	22%
	50,000-99,999	27%	15%	25%	16%
	$\geq$ 100,000	16%	9%	17%	11%

Note: This table shows detailed cashout rates for separated participants in the final sample (i.e., participants with at least one year of complete contribution history in which the inferred annual income amount is within 3% of the W-2 value). Annual income is averaged across participants' employment spells and is in real 2023 dollars. Age and balance are as of separation. Participants are classified as hourly if at least 45% of their observable paychecks change by at least \$1 relative to the lagged paycheck value. The remaining participants for whom fewer than 45% of paychecks change by at least \$1 are classified as salaried. See Section 3.2 for a discussion of this classification rule. Cells are left blank if they contain less than 100 participants.

Appendix B Bonus-identification algorithm

To identify annual bonuses among salaried participants, I use the following algorithm. I only apply the algorithm to participant-years that are classified as salaried (i.e., participant-years in which less than 45% of paychecks change by at least \$1 relative to the lagged paycheck value; see the discussion in Section 3.2).

- Let g_{it} be annual base income for salaried participant i in year t. The typical paycheck is therefore $\frac{1}{n_{it}} g_{it}$, where n_{it} is the number of paychecks for participant i in year t.
- When an annual bonus is paid, the paycheck becomes $g_{it}\left(\frac{1}{n_{it}} + \theta\right)$, where θ is the bonus amount as a share of base income.
- Proxy for g_{it} with $n_{it} y_{it}^m$, where y_{it}^m is the median paycheck amount for participant i in year t. I use the median paycheck amount (rather than the mean) so that my estimate of base income is not affected by the annual bonus amount itself.
- Choose a minimum bonus-detection threshold $\underline{\theta}$. I set $\underline{\theta} = 0.05$, so I aim to detect bonuses that are at least 5% of base income.
- Classify a given paycheck y_{itp} as containing a bonus if:

$$\begin{aligned} y_{itp} &\geq n_{it} y_{it}^m \left(\frac{1}{n_{it}} + \underline{\theta}\right) \\ &= y_{it}^m (1 + n_{it}\underline{\theta}) \end{aligned}$$

- If the above condition is satisfied for a given paycheck y_{itp}, the inferred bonus amount is $y_{itp} - y_{it}^m$ and the inferred base amount is y_{it}^m.

Table B.1 shows summary statistics for the identified bonuses. About 35% of salaried participant-years have bonus income. In the vast majority of these cases, the bonus is identified in only one paycheck, consistent with the algorithm's goal of identifying annual bonuses. As a share of base income, bonus amounts range from 5.8% at the 10th percentile to 25.6% at the 90th percentile.

Table B.1: Inferred bonuses among likely salaried workers

	Likely hourly	Likely salaried
Participant-years	7,069,907	3,504,053
Share with inferred bonus	–	34.9%
# of paychecks with bonus income (cond. nonzero)		
Mean	–	1.2
10th	–	1
50th	–	1
90th	–	1
Bonus as share of base income (cond. nonzero)		
Mean	–	14.5%
10th	–	5.8%
50th	–	10.9%
90th	–	25.6%

Note: Observations are participant-years. Participant-years are classified as hourly if at least 45% of the participant's paychecks in the relevant year change by at least $1 relative to the lagged paycheck value. The remaining participant-years where fewer than 45% of paychecks change by at least $1 are classified as salaried.

Appendix C Numerical model simulations

To conduct the model simulations in Section 2.3, I use the following computational process. Throughout the process, I fix the secondary parameters at the values listed in the text: utility is CRRA with $\gamma = 2$, the exponential discount factor is $\delta = 0.97$, the 401(k) return is $r = 0.08$, the early-withdrawal penalty is $\tau = 0.1$, Social Security income is $y_R = 0.5$ (i.e., replacing half of average pre-retirement income), and $m = 0.5$, $\bar{m} = 0.06$ (representing the 50% match on the first the first 6% of income that is the most common match formula among Vanguard-administered plans; see Clark 2025). Then take the following steps:

- Fix a value for the standard deviation of the income process (σ). Take 100 draws from a lognormal distribution with mean zero and stanard deviation σ and store them in a vector $\hat{y}$.

- Each period, the worker makes two choices: consumption c_t and 401(k) saving s_t^I. Liquid saving s_t^L is then implied from the budget constraint in (5). Discretize the choice space as the cross-product of two grids, one for the c_t choice and one for the s_t^I choice. Each grid has a spacing of 0.03 between grid points.

- Discretize the value function $V_t(a_t^L, a_t^I)$, which is defined over the two-dimensional state (a_t^L, a_t^I). I use 28 points for the a_t^I grid and 16 points for the a_t^L grid (because simulated workers accumulate larger amounts of 401(k) wealth and the visited region of the a_t^I space spans a wider range). When borrowing is allowed, I expand the a_t^L grid to be 25 points (because additional points are needed below zero). Both the a_t^L and a_t^I grids are densest in the most-visited regions of the state space, with points slightly further apart at more extreme values.

- Compute the value function by backward induction. I start by defining $V_{11}(a_{11}^L, a_{11}^I)$ because optimal behavior during the retirement period $11 \leq t \leq 15$ is a deterministic function of starting retirement wealth.

- Populate $V_{10}(a_{10}^L, a_{10}^I)$ by solving for optimal behavior in (2), given V_{11}. Because any present bias is naive, β is fixed at 1 when populating the value function.

- Then iterate the backward-induction process to $t = 1$. Integrate over the next-period income distribution by averaging continuation values across the 100 random income draws $\hat{y}$. For the cashout period $t = 5$, I remove the constraint requiring $s_t^I \geq 0$ but impose the early-withdrawal penalty τ if $s_5^I < 0$.

- Finally, with the value function populated, simulate the model forward. Repeat the following simulation process 10,000 times for each $\beta \in \{0.6, 0.8, 1\}$:

  - Initialize $a_1^L = a_1^I = 0$.
  - Take a random income draw from $\hat{y}$ and choose the optimal (c_1, s_1^I, s_1^L). When simulating forward, the continuation value in (2) is now discounted by β.
  - Iterate forward through the rest of the worker's life.

This process is computationally intensive. On an AWS instance with eight virtual CPUs and 32GB of RAM, it takes approximately 5 hours to complete the process for a single value of σ.

Appendix D Additional evidence from layoffs

The main empirical analysis in Section 4 – enabled by the paycheck-level income data I construct – focuses on monthly earnings volatility *within* employment spells. Relating high-frequency, sub-annual earnings fluctuations to 401(k) withdrawals is the principal contribution of this paper. However, for completeness and comparison with prior literature, I also consider *between*-job income shocks caused by job loss.

Goodman et al. (2021) and Coyne et al. (2024) find that households are significantly more likely to take early withdrawals from retirement accounts in years in which they receive unemployment insurance (UI) benefits. These findings are consistent with households using retirement accounts as a source of liquidity after layoffs. Because UI receipt is not observable in 401(k) recordkeeping data, I must use other methods to identify layoffs. The recordkeeping data contain a field explicitly identifying layoffs, but plan sponsors do not always flag layoffs to Vanguard. I therefore use a combination of direct information and indirect inference. I consider the separation of participant i in plan j to be a layoff if i) the separation is explicitly identified as a layoff in the administrative data, or ii) at least L% of currently employed participants in plan j separate in the same month as participant i, for some threshold L. The logic behind the second condition is that tightly bunched separations typically indicate firm-wide layoffs (Jacobson et al., 1993).[32]

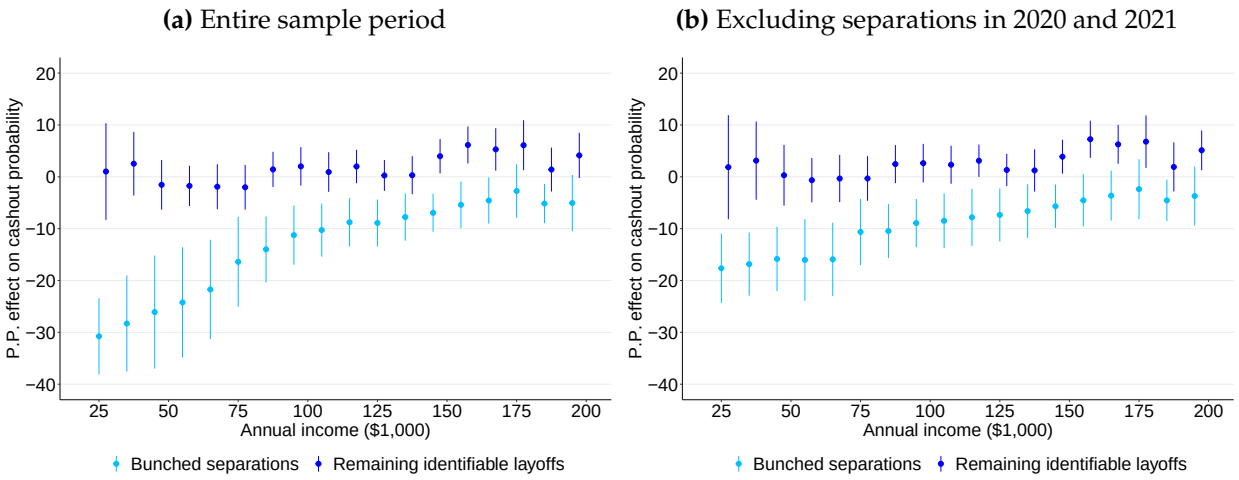
Figure D.1 shows the results of the layoff analysis. I begin with the cashout regression specification in (11), using the simple dummy for hourly participants as the volatility measure $\bar{v}_{ij}$. I then add two indicators, again interacted with the annual income buckets: the first indicator is equal to one for separations occurring during a bunched month (I set $L = 5\%$ in the baseline analysis and consider robustness in Figure D.2), and the second indicator is equal to one for separations that do not occur in bunched months but are explicitly coded as layoffs.[33]

The indicators for explicitly coded layoffs are generally positive, especially when I exclude separations occurring in 2020 or 2021, when unusually generous pandemic-era UI

[32] I only apply the second condition to plans with at least 1,000 participants.

[33] With $L = 5\%$, 10.7% of separations occur in bunched months and an additional 2.3% of separations are explicitly coded as layoffs.

Figure D.1: The effect of layoffs on post-separation cashout probability



Note: Observations are participants and the sample is restricted to participants who separate from employment during the sample period. I further restrict to participants for whom cashouts are penalized (i.e., those who are younger than 55 at the end of the calendar year in which they separate). I begin with the regression specification in (11) and add two indicators (each interacted with the annual income buckets). The first indicator is equal to one if participant i in plan j separates in a month in which at least L% of currently employed participants in plan j also separate (I set $L = 5$% in the baseline analysis and consider robustness in Figure D.2). The second indicator is equal to one if participant i does not separate in a bunched month but their separation is explicitly coded as a layoff in the administrative recordkeeping data. In panel (b), I exclude separations occurring in 2020 or 2021. The underlying regression specification in (11) includes buckets for average annual income during the employment spell, a dummy for participants classified as hourly workers (interacted with the annual income buckets), a quadratic of age at separation, a quadratic of plan tenure at separation, plan fixed effects, and year fixed effects. Annual income buckets on the horizontal axis are $10,000 wide and are based on income values converted to real 2023 dollars. Standard errors are clustered by plan.

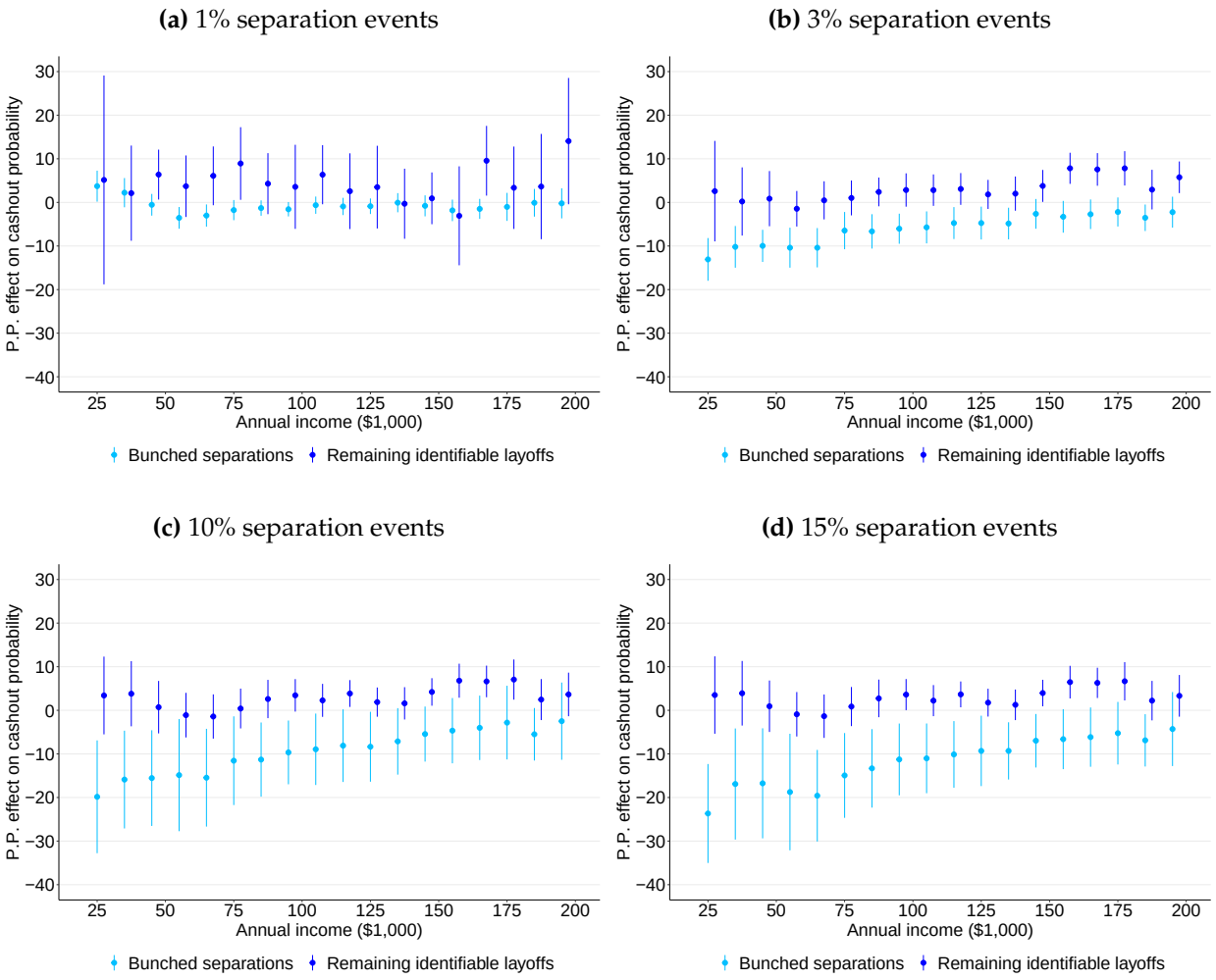
benefits likely made laid-off workers less dependent on retirement accounts for liquidity.[34] The estimates for middle- and higher-income workers (in the 5-10 percentage point range) are roughly consistent with Goodman et al. (2021), who find UI recipients are 11 percentage points more likely to take penalized withdrawals than other job separators. Interestingly, the indicators for bunched separations are uniformly negative and quite large at the bottom of the income distribution, though they again shift upward when I exclude the 2020-2021 period. The recordkeeping data do not allow me to confirm this, but it is possible that employers are more likely to give advance notice (allowing workers to accumulate liquid savings) or award severance payments (which provide post-separation liquidity) during mass layoff events, reducing the affected workers' dependence on 401(k) withdrawals.[35]

When excluding mass-layoff events (which appear to have unique implications for

[34]See Ganong et al. 2020 regarding the pandemic-era UI expansion. The empirical samples for both Goodman et al. (2021) and Coyne et al. (2024) end before 2020.

[35]The federal WARN Act requires employers to give 60 days' notice before certain types of mass layoffs, or to award back pay to laid-off workers if notice is not given.

Figure D.2: The effect of layoffs on post-separation cashout probability, using alternative separation bunching thresholds



Note: This figure reproduces the layoff estimates from Figure D.1, with alternative separation-bunching thresholds L.

cashout behavior), I thus find modest positive cashout effects that are reconcilable with prior literature. The explicit layoff effects in Figure D.1 are nonetheless smaller than the effects of recurring monthly volatility in Figure 2, underscoring the importance of *within*-job income shocks in explaining early 401(k) withdrawals.