A Parsimonious Model of Idiosyncratic Income*

Edmund Crawley Martin B. Holm Håkon Tretvoll

September 2024

Abstract

The standard permanent and transitory income model is known to be misspecified. Estimates of income volatility within this model differ depending on the specific data moments used – whether they are in levels or differences – and how these moments are weighted during estimation. We suggest a simple modification to the standard model: Allowing for two transitory shocks that persist for different lengths of time. Our proposed model, which uses the same number of state variables and introduces only one additional parameter, consistently and accurately identifies the parameters of the income process, regardless of the estimation method used.

JEL: E21, E24, J30
Keywords: Income Uncertainty, Inequality, Household Finance

*We thank Dmytro Hryshko for sharing the code from Daly, Hryshko, and Manovskii (2022) with us and Sisi Zhang for sharing the data and code from Moffitt and Zhang (2018). We thank Johannes H. Alsvik, Michael Graber, Robert Moffitt, two anonymous referees, and the editor Iourii Manovskii for valuable comments and discussions. We also thank seminar and conference participants at the 3D-In-Macro Workshop 2023, CEF 2023, EWMES 2022, the Federal Reserve Board, Johns Hopkins University, NBER Summer Institute 2022, Econometric Society Summer Meetings 2022, NHH Norwegian School of Economics, Norges Bank, Normac 2022, Oslo Macro Group, Sciences Po, SEHO 2022, Statistics Norway, and Vienna Macro Cafe 2022. This project has received funding from the European Research Council (ERC) under the European Union's Horizon2020 research and innovation programme (grant agreement No. 851891). We use the Panel Study of Income Dynamics public use dataset, produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (2021). Viewpoints and conclusions stated in this paper are the responsibility of the authors alone and do not necessarily reflect the viewpoints of Statistics Norway or the Federal Reserve Board.
Crawley: Federal Reserve Board, edmund.s.crawley@frb.gov.
Holm: University of Oslo, m.b.holm@econ.uio.no.
Tretvoll: Statistics Norway, hakon.tretvoll@ssb.no.

1 Introduction

> [T]he key challenge for future work is to develop a specification for the wage process that is both parsimonious enough to be used as an input to incomplete-markets models, and rich enough to account empirically for the covariance structure of wages in both levels and differences.
>
> Heathcote, Perri, and Violante (2010, p. 40)

What is the nature of idiosyncratic income risk faced by households, and how has it changed over time? The related economic literature centers around a model of permanent and transitory income. In this standard model, households face a series of income shocks. Some of these income shocks permanently (or persistently) affect income, such as a job change or promotion. Others only have transitory effects on income, such as a bonus or a period of sick leave. However, although widely used, this model is known to be misspecified. If the model was accurate, the permanent and transitory components' estimated variances would be consistently estimated regardless of how the estimation is performed. In practice, the choice of income moments (levels of differences) and the weighting matrix applied to those moments in a minimum distance estimation (optimally, diagonally, or equally weighted) lead to different conclusions about the importance of permanent versus transitory income risk.

In this paper, we make one change to the standard model of idiosyncratic income: We divide transitory shocks into two flavors. "Bonus" shocks (which can include classical measurement error) display no persistence, while "passing" shocks persist for a stochastic period. Estimation using Norwegian administrative data indicate that transitory income risk is best described by a mixture of bonus and passing shocks.

We simulate our proposed model and show that when we estimate the standard model on the simulated data, we find evidence of the same type of misspecification as in actual data. Indeed, with the most commonly used weighting—diagonally weighted minimum distance (*diagonal*)—the permanent income variance is higher and the transitory income variance is lower when estimating the standard model with difference moments than with level moments. This structure of misspecification is the same as we observe when estimating the standard model using actual data. Hence, the adjustment we make to the income process is sufficient to explain the differences in parameter estimates obtained when estimating the standard model across moments and weighting matrices.

We next estimate our proposed model using Norwegian administrative data and Panel Study of Income Dynamics (PSID) data. We find similar parameter estimates regardless of the moment or weighting matrix applied. This stability of parameter estimates suggests that our proposed model does not have the misspecification problems of the standard model. Moreover, because our proposed model only requires estimating one extra parameter relative to the standard model, it can be estimated even in small samples such as the PSID. Hence, we argue that the benefit of reducing model misspecification outweighs the cost of increased model complexity. We also provide codes online to make the estimation of our proposed model available to other researchers.[1]

Although the standard income model is misspecified and we recommend using our proposed model, our results can be used to interpret the existing literature. Some combinations of moments and weighting matrices yield more reliable estimates than others. Both in simulated and actual data, the combination of level moments and the equally or diagonally[2] weighted minimum distance method provides parameter estimates in the standard model close to the data-generating process or the parameter estimates of the proposed model. The intuitive reason is that with level moments, the permanent variance will be identified from moments for which the transitory shocks play no role ($cov(y_t, y_s)$ with $s >> t$). Hence, the standard model will accurately identify the permanent income variance if these long covariances are weighted sufficiently high in the estimation and the transitory shocks are sufficiently small or short-lived.

We also use our proposed model to investigate how income risk varies by age and how the nature of income risk has changed over time in Norwegian administrative data and the Panel Study of Income Dynamics (PSID). First, regarding lifecycle income risk, we find that neither permanent nor transitory income variance varies much from age 35 to 50. This result is consistent with prior findings in the literature where income risk does not vary much by age (see, e.g., Storesletten, Telmer, and Yaron, 2004, Heathcote, Storesletten, and Violante, 2005, and Guvenen, Karahan, Ozkan, and Song, 2021). Second, we find evidence that 'start-of-working-life' inequality and permanent income risk have increased over time in both the Norwegian data and the PSID.[3] We further show that the estimated time trends of income risk are similar irrespective of whether one uses the standard or our proposed model. Hence, while the standard model tends to yield very different estimates of the *level* of risk depending on the moment or weighting matrix

[1] https://github.com/edmundcrawley/IncomeEstimationToolbox.

[2] In both simulated and actual data, the diagonally weighted minimum distance matrix performs well, but one cannot guarantee its performance in all cases because it might give sufficiently low weights to the covariances with large time differences.

[3] This question has been discussed extensively at least since Gottschalk, Moffitt, Katz, and Dickens (1994). Recent contributions include Moffitt and Gottschalk (2002), Gottschalk and Moffitt (2009), Heathcote, Storesletten, and Violante (2010), Heathcote, Perri, and Violante (2010), Sabelhaus and Song (2010), Moffitt and Gottschalk (2012), Bloom, Guvenen, Pistaferri, Sabelhaus, Salgado, and Song (2017), Moffitt, Bollinger, Hokayem, Wiemers, Abowd, Carr, McKinney, Zhang, and Ziliak (2021), Carr, Moffitt, and Wiemers (2020), Moffitt and Zhang (2020), and McKinney and Abowd (2020).

applied, the *time trends* in income risk are similar.

Related literature. Our paper most closely relates to Daly, Hryshko, and Manovskii (2022), who also ask why estimates of permanent and transitory shocks in the standard model depend on the estimation method used. However, their solution is different. They argue that differences in the sample selection that naturally arise between level and difference estimation in unbalanced panels can explain the differences in estimates. They show that using a balanced panel overcomes these sample selection issues and, as a result, conclude that the standard model can fit the data well. In contrast, we show, both in theory and practice, that the results of Daly, Hryshko, and Manovskii (2022) are sensitive to the weighting matrix used in the estimation. This sensitivity to the weighting matrix is distinct but similarly concerning evidence that the standard model is misspecified. We show that our proposed model is robust to the choice of moments used and the choice of weighting matrix. Nevertheless, we remain convinced by the arguments about sample selection and restrict ourselves to using balanced panels in our analysis.

Several papers are building rich models to match higher-order moments of the income distribution. For example, Guvenen, McKay, and Ryan (2022) build an income process that is rich enough to fit several moments of the income distribution but is still sufficiently tractable to be included in structural models. Similarly, several papers build more complicated models matching higher-order moments of the income distribution.[4] We see our paper as complementary to this literature—while this literature focuses on matching higher-order moments, our paper is targeted at correctly determining the persistence of shocks and their variance and makes no claims about higher-order moments. Instead of building complicated models to match the dynamics of the income innovations, we construct a *parsimonious* income process that is robust to known misspecification issues. Hence, it fits the data well, yet is simple enough to be included in structural models and allows estimation in relatively small data sets.

[4]Other recent examples include Druedahl, Graber, and Jørgensen (2021), whose model matches data on monthly income innovations in Denmark; Guvenen, Karahan, Ozkan, and Song (2021), which build an income process to match moments from administrative data on annual income innovations in the U.S.; and Arellano, Blundell, and Bonhomme (2017), who estimate a process that allows for variation in parameters by individual income levels to match income innovations. Several other papers also estimate variants of non-linear income dynamics, e.g., Browning, Ejrnaes, and Alvarez (2010), Altonji, Smith Jr., and Vidangos (2013), De Nardi, Fella, and Paz-Pardo (2020), Braxton, Herkenhoff, Rothbaum, and Schmidt (2021), and De Nardi, Fella, Knoef, Paz-Pardo, and Van Ooijen (2021). For Norway, Halvorsen, Holter, Ozkan, and Storesletten (2024) present evidence of non-gaussian features of income dynamics.

2 Data

The analysis uses Norwegian administrative data on annual income. We combine this income data with demographic information such as sex, age, country of birth, and years of education. To ensure comparability with the rest of the literature, we restrict attention to males born in Norway with income between 1971 and 2014.

The Norwegian income data have several advantages relative to other available data sets. First, the data are administrative, covering the population of Norwegian tax residents. Second, only a small share of the earnings are right-censored.[5] Therefore, our data include precisely measured earnings for almost everyone. Third, the long panel from 1971 to 2014 allows us to follow individuals for long periods, including some cohorts' complete earnings history from their first job to retirement.

Variable definitions. Our pre-tax earnings measure includes labor income (from wages and self-employment) and work-related cash transfers such as unemployment and short-term sickness benefits. In the analysis, pre-tax earnings are deflated with the consumer price index, indexed to 2011 Norwegian kroner. The observations we use for individual log income are residuals obtained from a regression of log earnings on a full set of dummies for year, age, and years of education.

Sample selection. We select our sample from this data according to the suggestions in Daly, Hryshko, and Manovskii (2022). They argue for the importance of using a balanced sample since gaps in the data can bias the estimation of income processes. We focus on individuals in the middle of their careers, restricting attention to ages 35 to 50.[6] When doing so, we include individuals whose earnings observations are available for all the years from 34 to 51 to ensure a balanced sample.[7] In addition, we remove observations of extreme income changes and observations where the income level is very low. We define *extreme income changes* as observations where income increases by more than 500 percent or

[5]Bhuller, Mogstad, and Salvanes (2017) document that less than 3 percent of the sample is right-censored in any given year. The income series is available from 1967, but Halvorsen, Ozkan, and Salgado (2022) show that top-coding was most prevalent from 1967 to 1970 and that from 1971 less than 1% of observations are right-censored each year.

[6]We start at age 35 because the variance of income by age decreases with age before age 35 in the Norwegian data, driven by the variance of high-skilled individuals (Blundell, Graber, and Mogstad, 2015). Neither the standard nor our proposed model can account for such a pattern. Appendix A.1 provides more detail on our sample selection.

[7]Following Daly, Hryshko, and Manovskii (2022), the first (34) and last (51) observations ensure that we include individuals who worked the entire first (35) and final (50) observations in our sample.

decreases by more than 80 percent, same as Daly, Hryshko, and Manovskii (2022).[8] A *very low* income level is defined as below the Norwegian social security system's definition of a base level (around USD 10,000 in 2011). Observations where an individual has a lower income level than this base level are considered years where the individual is only loosely attached to the labor force and is treated as a missing observation. If an individual ever experiences an extreme income change or a very low income level, all individual observations are dropped from the sample, ensuring that the panel is balanced. Our final sample includes 536,399 Norwegian males from 27 cohorts born between 1937 and 1963.

3 Problems with the Standard Model

This section illustrates the misspecification issues with the standard model using the Norwegian data. We first describe the model and how to identify the model parameters. Next, we present estimation results for all combinations of moments and weighting matrices typically applied.

3.1 The Standard Model

We will use the "standard model" as a benchmark to highlight some of the existing problems in the literature and how our proposed model resolves these problems.[9]

The standard model is typically written at an annual frequency with log income y_t composed of a permanent component p_t with innovations ψ_t and transitory shocks ε_t with a moving-average coefficient θ. The model can be written as

$$y_t = p_t + \varepsilon_t + \theta \varepsilon_{t-1} \tag{1}$$

$$p_t = p_{t-1} + \psi_t \tag{2}$$

where ψ_t and ε_t are independent of each other and across time.

The permanent income component is a random walk at an annual frequency, and the transitory income component follows an MA(1). The process begins at time zero with some existing distribution of permanent income. The parameters of interest in this model are the variance of permanent and transitory income innovations, $\sigma^2_{\psi_t}$ and $\sigma^2_{\varepsilon_t}$ respectively,

[8] In a robustness exercise, in Appendix E.3, we do not exclude these individuals with extreme income changes. There is no substantial effect on our results.

[9] While there are several variants of the standard model—most commonly including some decay in the permanent component—they all suffer from the same problems that our proposed model can resolve. Appendix E.2 estimates both the standard and proposed model allowing for general persistence and finding the same structure of misspecification issues.

as well as the 'start-of-working-life' variance of permanent income $\sigma^2_{p_0}$, and the MA(1) parameter θ.

Researchers can use either level or difference moments to identify the parameters in the model. If they use difference moments, the covariance structure of permanent income is

$$var(\Delta p_t) = \sigma^2_{\psi_t} \tag{3}$$
$$cov(\Delta p_t, \Delta p_s) = 0 \qquad \text{if } s \neq t. \tag{4}$$

and the covariance structure of the transitory component is (defining the transitory component as $q_t = \varepsilon_t + \theta\varepsilon_{t-1}$)

$$var(\Delta q_t) = \sigma^2_{\varepsilon_t} + (1-\theta)^2\sigma^2_{\varepsilon_{t-1}} + \theta^2\sigma^2_{\varepsilon_{t-2}} \tag{5}$$
$$cov(\Delta q_t, \Delta q_{t-1}) = -(1-\theta)\sigma^2_{\varepsilon_{t-1}} + \theta(1-\theta)\sigma^2_{\varepsilon_{t-2}} \tag{6}$$
$$cov(\Delta q_t, \Delta q_{t-2}) = -\theta\sigma^2_{\varepsilon_{t-2}} \tag{7}$$
$$cov(\Delta q_t, \Delta q_s) = 0 \qquad \text{if } s < t-2. \tag{8}$$

Using the independence of the permanent and transitory components of income, the covariance structure of the difference in log income is the sum of the covariance structure of each component

$$cov(\Delta y_t, \Delta y_s) = cov(\Delta p_t, \Delta p_s) + cov(\Delta q_t, \Delta q_s). \tag{9}$$

It is common to use equations (3)-(9) to estimate parameters by minimizing the distance between the empirically observed moments and those implied by the model. The approach is similar when using level moments, deferred to Appendix B.1.

3.2 Results Using the Standard Model

We now estimate the standard model using Norwegian data. We first describe our estimation procedure before we present the estimated parameters.

Estimation procedure. We start with M balanced panels, each starting in a different year. For each panel, we calculate the empirical covariance matrix for either the levels or differences

$$EmpiricalLevels_{t,s} = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}y_{i,s} \tag{10}$$

$$EmpiricalFD_{t,s} = \frac{1}{N}\sum_{i=1}^{N} \Delta y_{i,t}\Delta y_{i,s} \tag{11}$$

where $y_{i.t}$ is residualized log income of individual i at time t as described in Section 2. Our minimum distance procedure for differences uses the loss function:

$$\mathcal{L} = \sum_{j=1}^{M} vech(EmpiricalFD_j - ModelFD_j)^T \Omega_j^{-1} vech(EmpiricalFD_j - ModelFD_j) \tag{12}$$

and equivalently for levels. Here Ω_j is either the full optimal minimum distance weighting matrix for panel j (*optimal*), the optimal minimum distance weighting matrix along the diagonal with all off-diagonal elements set to zero (*diagonal*), or the identity matrix (*identity*).

Results. Table 1 presents estimated parameters using all combinations of moments and weighting matrix typically applied in the literature. There are several notable observations.

First, the estimated variance of permanent shocks differs depending on whether one estimates the model using level or difference moments. For example, looking at the *diagonal* column, the estimated variance of permanent income is 0.004 when estimated using level moments but almost three times as large when using difference moments. Since permanent income risk is of primary importance for household welfare, these differences substantially affect household behavior in structural models. Hence, it is important to understand the origins of this result and how to estimate permanent and transitory income variances consistently.

Second, while the estimated permanent variance is higher when using difference rather than level moments, the estimated transitory variance is lower. For example, when using the diagonal weighting matrix (*diagonal*), the estimation using level moments yields a lower estimate of permanent risk but a higher estimate of the variance of transitory risk than the estimation using difference moments. Contrary movements in parameter estimates suggest that part of the variation in the data is not correctly allocated to the transitory or permanent variance when using the standard model.

The third observation is that when using the full optimal minimum distance weighting

	Identity		Diagonal		Optimal	
	Level	Difference	Level	Difference	Level	Difference
σ^2_{perm}	0.004	0.011	0.004	0.011	0.005	0.007
σ^2_{tran}	0.032	0.020	0.033	0.020	0.021	0.021
θ	0.570	0.070	0.574	0.071	0.163	0.145
σ^2_{init}	0.062	✗	0.062	✗	0.058	✗

Notes: The table presents estimated parameters using the standard model on Norwegian data. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonal of the optimal, and full optimal weighting matrix, respectively. The parameter θ is the moving-average coefficient. The table shows the mean of parameter estimates over time and age.

Table 1: Estimated parameters using the standard model, Norwegian data.

matrix (*optimal*), one gets more similar parameter estimates regardless of which moments are used. Daly, Hryshko, and Manovskii (2022) note this and argue that this estimation method, combined with a balancing of the panels, yields unbiased estimates of the permanent and transitory variance using both level and difference moments. However, although parameter estimates are similar irrespective of the moments used, it does not imply that the resulting parameter estimates are correct, a point we return to in Section 5.

We have now established the main misspecification issues of the standard model. The rest of the paper is devoted to providing an alternative yet simple model that does not feature these misspecification issues.

4 The Proposed Model

We now present the proposed model, focusing on how it differs from the standard model. Our innovation is that we allow for three types of income shocks. This section shows that this change is sufficient to resolve the misspecification issues in the standard model discussed in Section 3. Moreover, since we only add one additional parameter, our proposed model is parsimonious and can be estimated in small data sets such as the PSID.

4.1 The Shocks

The innovation is to include three types of income shocks in our model. Permanent shocks are the exact analog of those in the standard model. Bonus shocks are the analog of transitory shocks in the standard model without an MA(1) component and can also be considered to encompass classical measurement error. Finally, in place of the MA(1) process for transitory shocks, we propose combining bonus shocks with "passing" income shocks persisting over a stochastic period. The model for log income, y_t is written as follows:

$$y_t = p_t + v_t + \zeta_t \tag{13}$$

$$p_t = p_{t-1} + \psi_t = p_0 + \sum_{i=1}^{t} \psi_i \tag{14}$$

$$v_t = \begin{cases} v_{t-1} & \text{with probability } \rho_{\text{passing}} \\ \varepsilon_t^v & \text{with probability } 1 - \rho_{\text{passing}} \end{cases} \tag{15}$$

where p_t is the permanent component, v_t is the passing component, and ζ_t is the bonus component. As is common for the standard model, our proposed model is written at an annual frequency.[10] The shocks to each component, ψ_t, ζ_t, and ε_t^v are each i.i.d. across time and from each other and have potentially time-varying variance.[11] The rest of this subsection describes each income component type in detail, how parameters are identified in the data using both level and difference moments, and provides intuition about how the standard model may misinterpret these shocks.

Income component 1: Permanent. A permanent shock to income can be thought of as a promotion, a wage rise, or a job change that sets an individual on a permanently higher income trajectory.[12] In levels, the covariance matrix for the permanent shock is, for $t \leq s$:

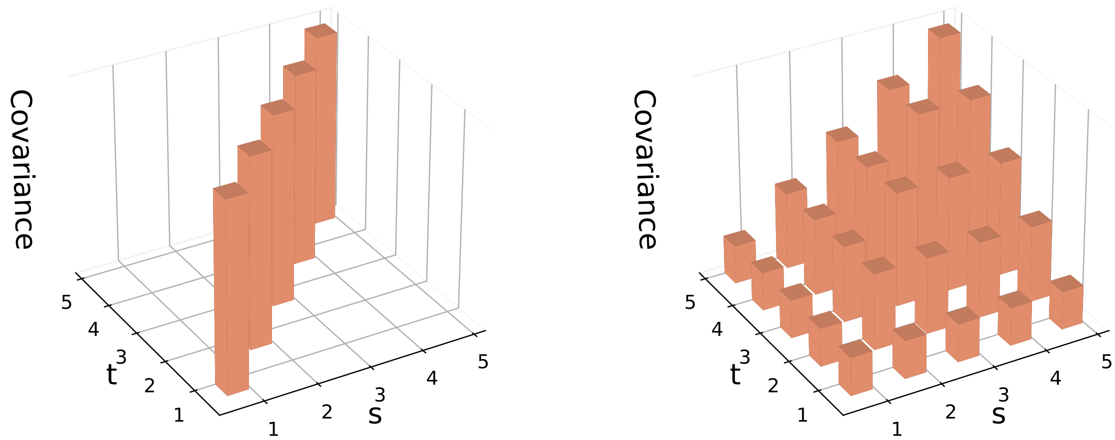
$$cov\left(p_t, p_s\right) = cov\left(p_0 + \sum_{i=1}^{t} \psi_i, p_0 + \sum_{i=1}^{s} \psi_i\right) = var\left(p_0 + \sum_{i=1}^{t} \psi_i\right) = \sigma_{p_0}^2 + \sum_{i=1}^{t} \sigma_{\psi_i}^2 \tag{16}$$

where $\sigma_{\psi_i}^2$ is the permanent shock variance at time i. For any component of income, x, we can always calculate the difference covariance matrix from the levels covariance by the

[10]Since the model is written in discrete time, it suffers from some of the problems of time aggregation highlighted in Crawley (2020). In a previous version of this paper Crawley, Holm, and Tretvoll (2022), we investigated the importance of time aggregation by modeling the income process in continuous time. Our results indicate that this matters for accurately estimating the relative importance of the different transitory shocks in the model, but it is not crucial for estimating permanent versus transitory risk. For economic applications, this latter distinction is the key issue, and hence, we focus on the discrete-time formulation of our proposed model in this version of the paper.

[11]Note that our estimation procedure only uses the covariance structure of the income process; therefore, our model allows for arbitrary higher-order moments. In particular, there is no restriction that the shocks in the model are Gaussian.

[12]In the body of the paper, we focus on versions of the model where the permanent component of income is a unit root. In Appendix E.2, we estimate models in which the permanent component is allowed to mean-revert and can be highly persistent but possibly not permanent.



(a) Difference covariances.

(b) Level covariances.

Figure 1: Covariance structure for the permanent component of income.

relationship:

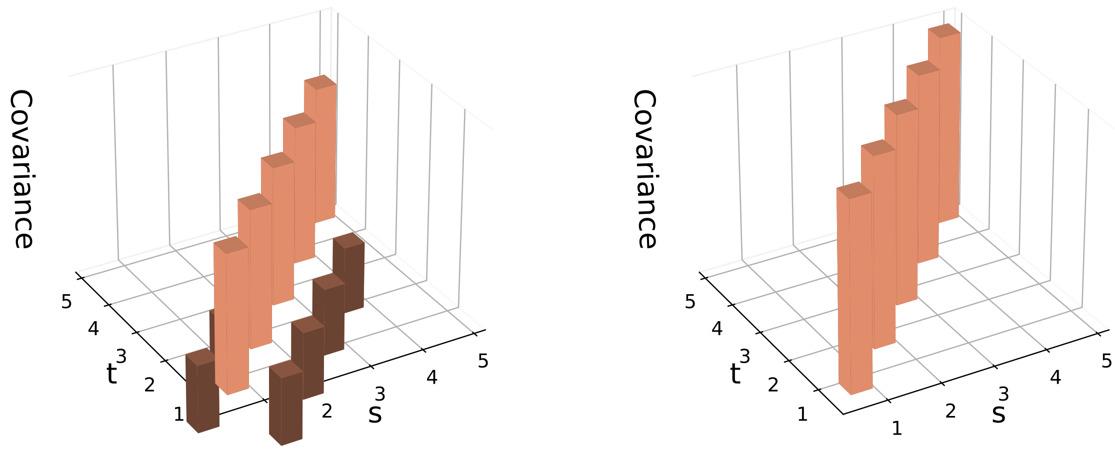
$$cov\left(\Delta x_t, \Delta x_s\right) = cov\left(x_t, x_s\right) - cov\left(x_{t-1}, x_s\right) - cov\left(x_t, x_{s-1}\right) + cov\left(x_{t-1}, x_{s-1}\right). \tag{17}$$

In the case of the permanent income component, this reduces to:

$$cov\left(\Delta p_t, \Delta p_s\right) = \begin{cases} \sigma^2_{\psi_t} & \text{if } t = s \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Figure 1 graphically shows the covariance structure of the permanent component of income in differences (panel a) and levels (panel b), assuming the variance of permanent shocks does not vary over time. The key features to notice are that the permanent component difference moments are constant on the diagonal and zero off the diagonal. By contrast, the permanent component level moments grow at a constant rate along the diagonal—a feature that is unique to the permanent shock component, as we shall see.

Income component 2: Bonus. The bonus shock consists of a one-time shock to income, like a bonus that increases the annual income that year but has no correlation with income in other years. It can be either positive or negative. In levels, the covariance matrix for



(a) Difference covariances.

(b) Level covariances.

Figure 2: Covariance structure for the bonus component of income, differences and levels.

the bonus shock is:

$$cov\left(\zeta_t, \zeta_s\right) = \begin{cases} \sigma^2_{\zeta_t} & \text{if } t = s \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

As with the permanent shock, we can calculate the first difference covariance matrix from the level covariance matrix, it is:

$$cov\left(\Delta\zeta_t, \Delta\zeta_s\right) = \begin{cases} \sigma^2_{\zeta_t} + \sigma^2_{\zeta_{t-1}} & \text{if } t = s \\ -\sigma^2_{\zeta_t} & \text{if } t = s-1 \\ -\sigma^2_{\zeta_{t-1}} & \text{if } t = s+1 \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Figure 2 shows the covariance structure for the bonus component of income assuming non-time-varying shock variance. In differences (panel a), the diagonal is positive and the off-diagonal elements are negative and half the size of the diagonal elements. Note that the sum of each row or column of the matrix is zero—a feature of all transitory shocks. In levels (panel b), the bonus component covariance matrix is constant along the diagonal and zero elsewhere.

Income component 3: Passing. The final income shock we introduce is a passing shock. In this stochastic process, income jumps at a fixed hazard rate. When this component of income jumps, it moves to a new level drawn from an independent distribution with a mean of zero. As a result, the level of this shock remains fixed for an exponentially distributed period of time before jumping to a new level.

One can think of the passing shock as representing a longer unemployment spell or a temporary switch to part-time employment. Alternatively, the passing covariance structure is consistent with pay rises that are predictable over the long term, but whose exact timing is subject to randomness.[13]

Another income process that results in an identical covariance structure to the passing shock process described here is an AR(1) process.[14] However, we describe the model as having passing shocks for two reasons. First, Druedahl, Graber, and Jørgensen (2021) estimate a monthly model where they specify a shock type that is general enough to potentially contain both the passing variant and the AR(1) variant. They reject the AR(1) version and end up with an income process similar to the passing shock in the current paper. Second, Arellano, Blundell, and Bonhomme (2017) show that individuals with very low income realizations tend to experience sudden and large increases in income. These sudden large increases are more consistent with the end of passing-type shocks when income jumps back toward the mean, rather than slowly drifting toward the mean.

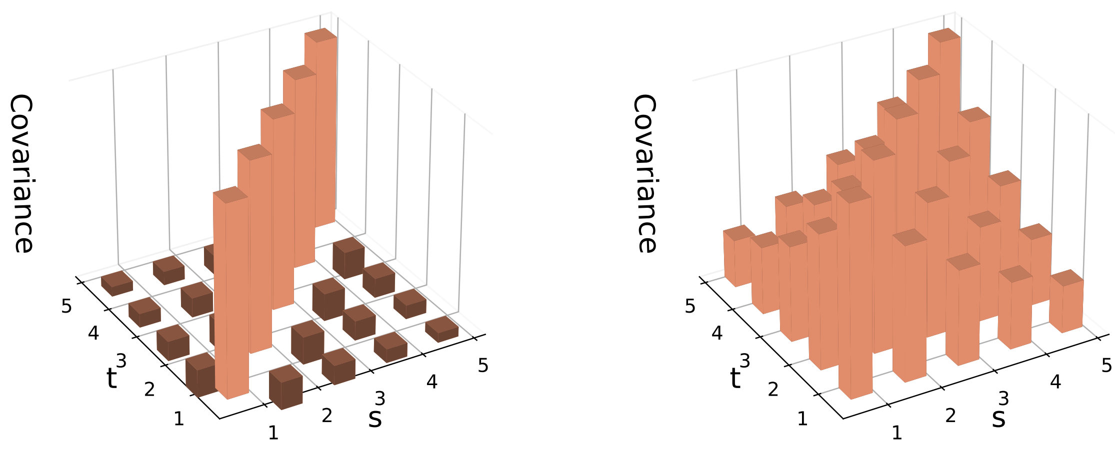
The covariance structure of the passing component of income in levels for $s \geq t$ is:

$$cov\left(v_t, v_s\right) = \rho_{\text{passing}}^{s-t}\sigma_{v_t}^2 = \rho_{\text{passing}}^{s-t}\left(\sum_{i=1}^{t}\left(1-\rho_{\text{passing}}\right)\rho_{\text{passing}}^{t-i}\sigma_{\varepsilon_i^v}^2 + \rho_{\text{passing}}^{t}\sigma_{\varepsilon_0^v}^2\right) \tag{21}$$

where $\sigma_{v_t}^2 = \sum_{i=1}^{t}\left(1-\rho_{\text{passing}}\right)\rho_{\text{passing}}^{t-i}\sigma_{\varepsilon_i^v}^2 + \rho_{\text{passing}}^{t}\sigma_{\varepsilon_0^v}^2$ is the variance of the passing shock component of income at time t. This variance is equal to the weighted mean of the variance of the passing shock at each time in the past, where the weights are given by the probability that the passing component of income has not changed since that time. Note that this specification of the passing shock implies that it has a half-life of $\tau = -\frac{\log(2)}{\log(\rho_{\text{passing}})}$. In our estimation results, we will show the half-life estimate as it is easier to interpret. In

[13] Appendix C.2 contains a model of such an income process with the same covariance matrix as the passing shock process described here.

[14] Appendix C.3 proves this claim.



(a) Difference covariances.

(b) Level covariances.

Figure 3: Covariance structure for the passing component of income, differences and levels.

differences, the passing component covariance matrix is given by:

$$cov\left(\Delta v_t, \Delta v_s\right) = \begin{cases} \sigma^2_{v_t} + \left(1 - 2\rho_{\text{passing}}\right)\sigma^2_{v_{t-1}} & \text{if } t = s \\ -\left(1 - \rho_{\text{passing}}\right)\rho_{\text{passing}}^{s-t-1}\left(\sigma^2_{v_t} - \rho_{\text{passing}}\sigma^2_{v_{t-1}}\right) & \text{if } s > t. \end{cases} \tag{22}$$

Figure 3 shows the covariance structure of a passing component on income in which the variance of the passing shock is constant over time. The difference covariance matrix (panel a) shows a tall column along the diagonal. In contrast to the bonus shock, which had off-diagonal elements that were half as tall as those on the diagonal, the off-diagonal elements for the passing shock are only a little negative. It is still the case that the (infinite) sum of each row is equal to zero, a necessary feature of the difference moments of a transitory income process. The levels covariance matrix (panel b) is constant along the diagonal and falls away exponentially away from the diagonal.

Figure 3 also reveals how to identify ρ_{passing} from the covariances. For both level and difference covariances, the ratio between two off-diagonal covariances along a row of the variance-covariance matrix can identify the passing shock persistence. For the level covariances, it is straightforward to compute that $cov(v_t, v_{s+1})/cov(v_t, v_s) = \rho_{\text{passing}}$. For the difference moments, it is more cumbersome and this is relegated to Appendix C.1.

Adding up. Using the independence of the permanent and transitory components of income, the covariance structure of log income is the sum of the covariance structure of each component:

$$cov\left(y_t, y_s\right) = cov\left(p_t, p_s\right) + cov\left(v_t, v_s\right) + cov\left(\zeta_t, \zeta_s\right). \tag{23}$$

For estimation, we will make use of all the available covariance elements and as a result our parameters are over-identified. In Appendix C.1 we write down an example of how to derive each parameter from an exactly-identified equation which also serves the purpose of showing that all the parameters are identified in our model.

Intuition for the problems with the standard model. To gain intuition for why the standard model can fail, assume that the proposed model is the data generating process with constant shock variances. Furthermore, assume that we have an infinite sample such that our empirical covariance matrix exactly matches that of the proposed model. We now consider what happens when we estimate the standard model parameters on this covariance matrix.

First, consider the case of estimating the standard model using the difference covariance matrix. The standard model assumes that all elements of the difference covariance matrix beyond order 2 is zero. The three parameters of the standard model (σ^2_ϕ, σ^2_ϵ, and θ) are set such that the first three covariances ($var(\Delta y_t)$, $cov(\Delta y_t, \Delta y_{t-1})$, and $cov(\Delta y_t, \Delta y_{t-2})$) are exactly matched. The model cannot do better than this because none of the parameters affect covariances beyond order 2.

For any transitory shock, the contribution to the diagonal elements is equal to minus the sum of the other elements in each row. Considering only the two first covariances implies that the model will underestimate the presence of transitory shocks. Instead, the standard model will interpret a large share of the passing shocks as permanent shocks. Hence, if our proposed model is the data generating process, the standard model estimated using difference moments will underestimate the variance of transitory shocks and overestimate the variance of permanent shocks.

The size of the underestimation of the transitory shock is equal to the sum of all covariances beyond order 2. In practice, as can be seen in Figure 3, each of these covariances are small and likely to be statistically insignificant outside of administrative datasets (see, e.g., Abowd and Card, 1989). However, the sum is the relevant statistic to understand the extent of the bias, and this sum may be large. This observation motivates our test in the PSID data that the sum of higher-order covariances is statistically significant.

Second, consider the case of estimating the standard model using the level covariance matrix. A distinguishing feature of the permanent income component's covariance matrix is that it grows along the diagonal, while the two transitory income components are constant along the diagonal. This unique feature of the permanent component's covariance matrix results in the permanent shock variance being correctly estimated even when estimating the standard model. The total size of the transitory shock variance (the sum of the bonus and the passing component) is also correctly estimated with a long panel because both transitory components share the feature that their levels covariance matrix falls to zero as you move further from the diagonal.

4.2 Mapping Between the Standard and the Proposed Model

Our proposed model has five parameters. Three of these parameters have almost exact counterparts in the standard model: initial permanent income variance, permanent income variance, and transitory income variance.[15] The two remaining parameters, the half-life of the passing shock τ and the fraction of the transitory variance that is of the "bonus" variety ($b = \sigma^2_{bonus}/\sigma^2_{tran}$ where $\sigma^2_{tran} = \sigma^2_{bonus} + \sigma^2_{passing}$), roughly serve the same purpose as the MA(1) parameter in the standard model in determining how persistent transitory shocks are, although they are not numerically comparable. Table 2 shows how the parameters in the two models compare.

Parameter Description	Proposed	Standard
Permanent income variance	σ^2_{perm}	$\text{Var}(\psi)$
Transitory income variance	σ^2_{tran}	$(1+\theta^2)\text{Var}(\varepsilon)$
Half-life of 'passing' shock	τ	✗
'Bonus' fraction of σ^2_{tran}	b	✗
MA(1) coefficient	✗	θ
Initial permanent income variance	σ^2_{init}	$\text{Var}(p_0)$

Table 2: Parameters in the proposed and standard models.

[15] The transitory income variance is defined as the total variance induced by transitory income processes. That is, in the standard model, transitory income variance at time T is $(1+\theta^2)\text{Var}(\varepsilon_T)$, not $\text{Var}(\varepsilon_T)$ which does not have a clear counterpart in the proposed model.

5 Simulation Results

This section presents a simulation exercise to illustrate how our proposed model can reconcile the misspecification problems in the standard model discussed in Section 3. The main result is that, in data simulated from our proposed income process, the standard model has the same type of misspecification issues as in actual data. Hence, the simulation exercise indicates that the standard model's problems can be explained by an insufficient description of transitory income shocks.

Simulation details. We simulate data according to the proposed model in equations (13), (14), and (15). Our simulation contains a panel of 200,000 individual log income histories over 16 years. We choose parameters to match those that we estimate using Norwegian data in Section 6.1. We then calculate the covariance moments of both income levels and differences, that is $cov(y_t, y_s)$ and $cov(\Delta y_t, \Delta y_s)$ for all t and s between 0 and 15. For each set of moments, we estimate the standard model using three different weighting matrices: equally weighted minimum distance (*identity*), diagonally weighted minimum distance (*diagonal*), and optimally weighted minimum distance (*optimal*).

Simulation results. Table 3 presents parameter estimates using either the standard model (Panel A) or the proposed model (Panel B). The column "True Value" shows the parameters of the data-generating process.

The columns using *identity* and *diagonal* in panel A of Table 3 show the main misspecification problem: relative to estimation using level moments, estimation using difference moments tends to overestimate permanent income variance and underestimate transitory income variance. This pattern of misspecification is the same as we found when estimating the standard model in Norwegian data in Table 1. Indeed, panel A of Table 1 and Table 3 are challenging to tell apart. Hence, the simulation exercise indicates that the standard model's problems can be explained by an insufficient description of transitory income shocks.

Moreover, reflecting the results in Daly, Hryshko, and Manovskii (2022), there is no difference between the parameter estimates obtained using level and difference moments when one uses the optimal weighting matrix (*optimal*). However, estimation using *optimal*—albeit similar across moments used—does not obtain the true parameters. Hence, not being sensitive to the moments used is insufficient to claim that the model is no longer misspecified.[16]

[16]In Appendix D.1 we further investigate the results presented in Daly, Hryshko, and Manovskii (2022).

Parameter	True Value	Identity Level	Identity Difference	Diagonal Level	Diagonal Difference	Optimal Level	Optimal Difference
Panel A: Standard Model							
σ^2_{perm}	0.005	0.005	0.012	0.005	0.012	0.008	0.008
σ^2_{tran}	0.038	0.033	0.020	0.033	0.020	0.024	0.023
τ	2.0 years	✗	✗	✗	✗	✗	✗
b	0.40	✗	✗	✗	✗	✗	✗
θ	✗	0.39	0.07	0.40	0.07	0.14	0.13
σ^2_{init}	0.065	0.068	✗	0.068	✗	0.068	✗
Panel B: Proposed Model							
σ^2_{perm}	0.005	0.005	0.005	0.005	0.005	0.005	0.005
σ^2_{tran}	0.038	0.038	0.039	0.038	0.039	0.038	0.038
τ	2.0 years	2.0	2.1	2.0	2.1	2.0	2.1
b	0.40	0.40	0.40	0.40	0.40	0.40	0.40
θ	✗	✗	✗	✗	✗	✗	✗
σ^2_{init}	0.065	0.065	✗	0.065	✗	0.065	✗

Notes: Panel A presents estimated parameters using the standard model on data generated for 200,000 individuals from the proposed model. 'True Value' refers to the parameters used in the data-generating process. *Level* and *Difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. The parameter θ is the moving-average coefficient. Panel B presents the estimated parameters using the proposed model on the same simulated data.

Table 3: Estimated parameters using simulated data.

Proposition 1 sheds some light on why the optimal weighting matrix may yield similar results independent of the moments used.

Proposition 1. *For any data-generating process with an ergodic difference distribution, as $T \to \infty$ and $N \to \infty$ the optimal weighted minimum distance estimator will yield the same estimate irrespective of the type of moments – levels or differences – used.*

The intuition for Proposition 1 is as follows. The *optimal* estimation procedure is invariant to any invertible linear mapping of the moments used for estimation. Hence, if

We follow the sample selection criteria they used for their Danish data as closely as possible to obtain a similar selection from the Norwegian registry data. Then we replicate the estimations they present using both level and difference moments with *optimal* and show the same result: with a balanced panel, there is no difference between estimates from the two sets of moments. However, we also present estimates using *identity* and *diagonal*, and for those weighting matrices, the parameter estimates again depend on the choice of moments. Our simulation exercise indicates that we cannot have confidence in the estimates obtained with the standard model using *optimal*, even though they do not depend on the choice of moments.

level moments were an invertible linear mapping of difference moments, the parameter estimates from the *optimal* estimation procedure would be identical irrespective of the moments used. Importantly, level moments are not an invertible linear mapping of difference moments because the difference moments have a lower dimension (there is one less covariance term) than the level moments. However, one can construct an invertible linear mapping between the level and difference moments by adding an additional row of level moments (either time 0 or time T) to the difference moments. The level and difference moments are thus 'almost' invertible. As T converges to infinity, the relative importance of these additional terms in the covariance matrix converges to zero, and the estimated parameters become similar. Appendix D.2 provides a sketch proof of this proposition.[17]

Proposition 1 requires the data-generating process to be *ergodic* to make sure that the importance of the terms added to the difference covariance matrix to make it invertible converges to zero. For the standard and proposed models in this paper, this holds. However, in some other common applications, this is not true. For example, if the income process contains 'types' where fixed effects are correlated with the variances of permanent shocks or if the income process contains heterogeneous trends (Guvenen, 2009).

Proposition 1 requires $T \to \infty$. Importantly, with small T, as is common in practical applications, the *optimal* estimation procedure will yield different parameter estimates depending on the type of moments used. In practice, we therefore need to know how large T would have to be for the parameter estimates to converge. Figure A.2 in Appendix D.3 illustrates how the parameter estimates of permanent and transitory variance converge as we increase T when the standard model is estimated on simulated data from the proposed model (as in Table 3). As suggested, the parameter estimates are relatively different for small T but converge as we increase T. For T equal to 16, as in the simulation exercise above, the parameter estimates are already reasonably close (less than 10% different).

Another notable observation from Table 3 is that some combinations of weighting matrix and moments yield parameter estimates close to the data-generating process even though the standard model is misspecified. For example, using the combination of level moments and either *identity* or *diagonal* weighting matrix provide estimates that are close to the true permanent and transitory variances.[18] Proposition 2 shows that as long as the data-generating process has a random walk component and the transitory income shocks

[17]Proposition 1 assumes the existence of a data-generating process that yields a set of balanced panels, that is, each individual n has an income panel of length T. The importance of balanced panels in estimation is shown by Daly, Hryshko, and Manovskii (2022).

[18]However, we note that the persistence of transitory shocks is too low. The parameters for the persistence of the transitory shock—θ in the standard model, b and τ in our proposed model—are not directly comparable. However, the θ estimated in the table suggests the transitory shocks are less persistent than in the data-generating process.

are sufficiently transitory, the standard model will consistently estimate the permanent income variance, and therefore also the total transitory income variance when estimating using level moments and the identity weighting matrix.

Proposition 2. *Take any data-generating process in which income is made up of a random walk (permanent) component and a potentially persistent but transitory component – our proposed model is an example. The equal-weighted minimum distance estimator of the standard model with level moments consistently estimates the permanent income variance of this data-generating process as the panel length of the data tends to infinity.*

The intuition for Proposition 2 is as follows. When $s >> t$, $cov(y_t, y_s)$ only contains information about the variance of permanent income – the covariance of the transitory component tends to zero by definition. Therefore, as the panel length gets large, there will be enough of these 'long' covariances to estimate the permanent income variance. When the permanent income variance is estimated consistently, the model identifies transitory income variance residually. This intuition is fleshed out in Appendix D.4.

Proposition 2 requires a panel length that tends to infinity. In practice, the panel length of the data has to be sufficiently long relative to the persistence and size of the transitory shock. The parameter estimates will be biased if the panel dimension is too short relative to the persistence and size of the passing shock. Indeed, when estimating the standard model, the transitory shock is assumed to be an MA(1) such that level covariances beyond order 1 are used to identify the variance of the permanent shock only. However, the passing shock in our proposed income process affects covariances beyond order 1. These covariances stemming from the passing shocks will bias the estimates of permanent income risk, and this bias increases with the persistence and variance of innovations in the passing shock process. Hence, the more persistent or risky the passing shocks are, the longer the panel dimension of the data must be to provide a consistent estimate of permanent variance, and it might need to be longer than the work-life of employees.

In Appendix D.5, we illustrate the role of passing shock size and persistence for the bias of estimated permanent variance. The biases are small for the half-life values and variance of passing shock innovations we estimate in the paper (used in the simulation). Still, if the passing shocks have a higher variance or the passing shocks have a longer half-life, the bias can become large, especially when the length of the panel is shorter.

Proposition 2 also sheds light on another case when estimating the standard model using level moments may be problematic. If the permanent shocks decay slowly over time – a common version of the standard model – the level estimation using the identity weighting matrix will no longer provide unbiased estimates. Indeed, if the permanent component of the income process is not a unit root, the permanent income variance can no longer be identified from the 'long' covariances. In this case, it is particularly important to estimate our proposed model rather than relying on estimates from the standard model using level moments.

In Appendix D.6, we show the results of estimating the standard model on a shorter panel of simulated data and on a simulated panel where the 'permanent' shocks decay slowly. These changes have little effect on the difference estimates (which are already biased) but can also introduce bias in the level estimates. In particular, the table suggests that when the permanent income shocks slowly decay, the permanent income variance is underestimated when using level moments.

Overall, our recommendation is to estimate our proposed model to avoid any of the issues discussed above related to estimation of the standard model. However, Proposition 2 does form the basis of our secondary recommendation that *if* you are estimating the standard model, then it should be estimated using level moments and an equally-weighted minimum distance estimator.[19]

Panel B of Table 3 shows the parameter estimates for the proposed model using the same set of simulated data. This set of results has no conflict between the level and difference estimates or the weighting method applied.

6 Data Results

We now proceed to estimate our proposed model using data. We first show that the parameter estimates for our proposed model are relatively insensitive to the moments and weighting matrix applied.[20] Hence, enriching the description of transitory income shocks is sufficient to significantly reduce the extent of misspecification. To illustrate that our proposed model is still parsimonious, we first estimate how income risk varies by age and time using Norwegian data. We also estimate our proposed model using the PSID data from the U.S.

6.1 Results using Norwegian Data

Estimation details. The Norwegian data contains 27 balanced panels, one for each cohort of our data. We allow for the permanent and transitory shock variance to vary year-to-year

[19] We can also note that, in practice, we have found that the diagonally-weighted minimum distance estimator tends to work well for the standard model when combined with level moments.

[20] The estimation of all models in the body of the paper assumes that the permanent component of income is a unit root. In Appendix E.2, we show that all our main results also prevail if we relax the assumption of a unit root.

and also linearly with age. Specifically, we assume that variances in year t for individual i born in cohort c are defined as $\sigma^2_{i,t} = (\alpha_0 + \alpha_1(t-c)) \cdot \sigma^2_t$. That is, for a 40 year old in 1982, we assume they experience shocks equal to an average for individuals in 1982 multiplied by a factor specific to 40 years that does not change over time.

The loss function is the sum of 27 sub-loss-functions, one for each cohort c. To reduce oscillatory behavior in the permanent volatility estimate, we add a regularization penalty to the loss function that penalizes changes in the permanent shock variance from year to year.[21] In addition, we adjust the model to allow for an institutional feature in Norway where a share (approx. 10 percent but time-varying) of labor income is paid in the following year as vacation pay. We adjust the model-implied moments to be associated with $(1-\eta_t)y_t + \eta_t y_{t-1}$ where η_t is the vacation pay share and compare these with the data. In practice, this adjustment has only minor effects on our results.

Income risk. Panel A of Table 4 shows the parameter estimates using the proposed model with each of the six combinations of weighting and covariance matrices. To ease comparison, Panel C of Table 4 also includes the estimates using the standard model (same as Table 1). Our proposed model's estimated parameters are relatively similar across all six combinations of data moments and weighting matrices. The permanent income variance is between 0.003 and 0.005, while the transitory income variance varies between 0.035 and 0.042. In contrast, the parameter estimates vary widely depending on the moment and method applied when using the standard model (Panel C).

Similar to Daly, Hryshko, and Manovskii (2022), we also find that the parameter estimates of transitory and permanent variance become more similar if one estimates the standard model using the optimally weighted minimum distance method. However, as shown using simulated data, this does not imply that the estimated variances are correct. Indeed, the estimated permanent variance using the standard model in the simulated data was upward biased. In Table 4, we find the same pattern. The permanent income variance is higher when estimated using the standard model and the optimal weighting matrix compared with estimation using the proposed model.

In two of the six combinations of moments and weighting matrix in Table 4, the standard model provides similar permanent income variance estimates as the proposed model. The proposed model's permanent income variance is between 0.003 and 0.005. The standard model estimated using the equally or diagonally weighted minimum distance method with level moments provides parameter estimates of permanent income variance

[21] The regularization term, $\mathcal{R} = R\sum_{t=2}^{T-1}\left(\sigma^2_{\phi_t} - (\sigma^2_{\phi_{t-1}} + \sigma^2_{\phi_{t+1}})/2\right)^2$, penalizes deviations of $\sigma^2_{\phi_t}$ from the mean of the points before and after it. R is a constant.

	Identity		Diagonal		Optimal	
	Level	Difference	Level	Difference	Level	Difference
Panel A: Proposed Model						
σ^2_{perm}	0.003	0.005	0.003	0.005	0.003	0.004
σ^2_{tran}	0.042	0.035	0.041	0.038	0.039	0.035
τ	2.200	1.865	2.362	2.046	2.221	1.865
b	0.288	0.434	0.290	0.395	0.369	0.394
σ^2_{init}	0.061	✗	0.060	✗	0.058	✗
Panel B: Proposed Model (no bonus shock)						
σ^2_{perm}	0.003	0.007	0.003	0.008	0.005	0.006
σ^2_{tran}	0.039	0.025	0.038	0.025	0.026	0.024
τ	1.333	0.431	1.425	0.431	0.541	0.468
b	.	.	.	.	.	.
σ^2_{init}	0.063	✗	0.062	✗	0.059	✗
Panel C: Standard Model						
σ^2_{perm}	0.004	0.011	0.004	0.011	0.005	0.007
σ^2_{tran}	0.032	0.020	0.033	0.020	0.021	0.021
θ	0.570	0.070	0.574	0.071	0.163	0.145
σ^2_{init}	0.062	✗	0.062	✗	0.058	✗

Notes: The table presents estimated parameters using the proposed and standard models on Norwegian data. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. The parameter θ is the moving-average coefficient.The table shows the mean of parameter estimates over time and age.

Table 4: Estimated parameters using the Norwegian data.

within the bounds of the proposed model. These are the same combination of moments and weighting matrices that provided estimates close to the data-generating process also in the simulation exercises. Hence, while we caution about using the standard model, our results indicate that if one does use it, one should estimate the standard model using level moments and the equally or diagonally weighted minimum distance method. However, we warn that the transitory income variance is always lower when estimated using the standard model compared with the estimates using the proposed model.

We estimate the half-life of the passing shock to be about two years. What kind of labor market events might this passing shock represent? We first note that while we specify the model as a passing shock, we show in Appendix C.2 and C.3 that the covariance structure of the passing shock is isomorphic to other types of shock processes, for example, an asymmetric passing shock or an AR(1). Hence, the underlying economic mechanism does not have to satisfy the passing shock specification exactly, only generate similar covariances.

A half-life of two years is typically too long to represent an unemployment spell. Instead, the passing shock we estimate most likely represents other events. Indeed, Halvorsen, Holter, Ozkan, and Storesletten (2024) investigate the same data and find that while changes in wages and increases in hours are persistent, "moderate and large reductions in hours worked tend to be transitory and have mostly disappeared five years after the initial fall." Hence, our passing shock most likely represents "transitions in and out of long-term sickness, transitions between full-time and part-time work, and job changes."

Another type of labor market event that our passing shock potentially represent is suggested by Low, Meghir, and Pistaferri (2010). They estimate a stochastic process for income that consists of a permanent shock, a fully transitory shock, and a firm-worker match fixed effect. The match between a firm and a worker persists for a while, but may be broken by firm-level shocks leading to job destruction or by workers moving after receiving a better offer of employment. In our model, the passing shock could thus be interpreted as connected with such a firm-worker match-specific effect.

Income risk by age and time. To investigate whether our proposed model yields different trends in income risk from the standard model, we provide age-varying and time-varying estimates of income risk using both models. Figure 4 presents age profiles of income risk estimated using the Norwegian data. We restrict attention in this exercise to results using the diagonally weighted minimum distance method. Figure 4 thus displays the results for four different combinations of models (standard and proposed) and moments (levels and differences).

The main takeaway from Figure 4 is that the misspecification of the standard model might mislead researchers to spurious conclusions about age patterns. This observation is illustrated in the top right panel showing the estimates of the transitory income variance. First, the misspecification of the standard model is visible as the large discrepancy between the estimates using the standard model. Moreover, the age profiles estimated using the standard model differ depending on the moments used. With level moments, transitory income risk decreases in age, while it is approximately flat if one estimates using difference moments. In contrast, our proposed model is less sensitive to the type of moments used in the estimation. Indeed, irrespective of the moments used in the estimation, the proposed model suggests a similar age pattern: transitory income variance declines slightly by age.

Figure 5 shows how estimates of income risk have changed over time in Norway. Panel (a) shows $\sigma^2_{\phi_t}$, permanent income variance ; Panel (b) shows $\sigma^2_{b_t} + \sigma^2_{\upsilon_t}$, transitory income variance, or the sum of passing shock variance and bonus shock variance; Panel (c) shows $\sigma^2_{p_0}$, the start-of-working-life variance. Again, the figure highlights how estimation using the standard model may yield different estimated variances depending on the type of moments used. However, the time trends of parameter estimates are relatively similar across specifications. The permanent income variance has increased over time, the transitory income variance has remained relatively stable, and the start-of-working-life income variance has increased from around 0.05 to almost 0.08 from 1972 to 1998. Hence, our results suggest that although the standard model is misspecified in that it may provide biased estimates of the *level* of income risk, the time trends of income risk are relatively similar across models.[22]

[22]Note that, in levels, a few years give rise to estimates of zero permanent income variance – a lower bound in our estimation. Heathcote, Perri, and Violante (2010) point to the negative estimates using levels as evidence of misspecification. Our proposed model is still likely to be misspecified. Nevertheless, this misspecification seems worthwhile to accept considering the parsimony of the model.

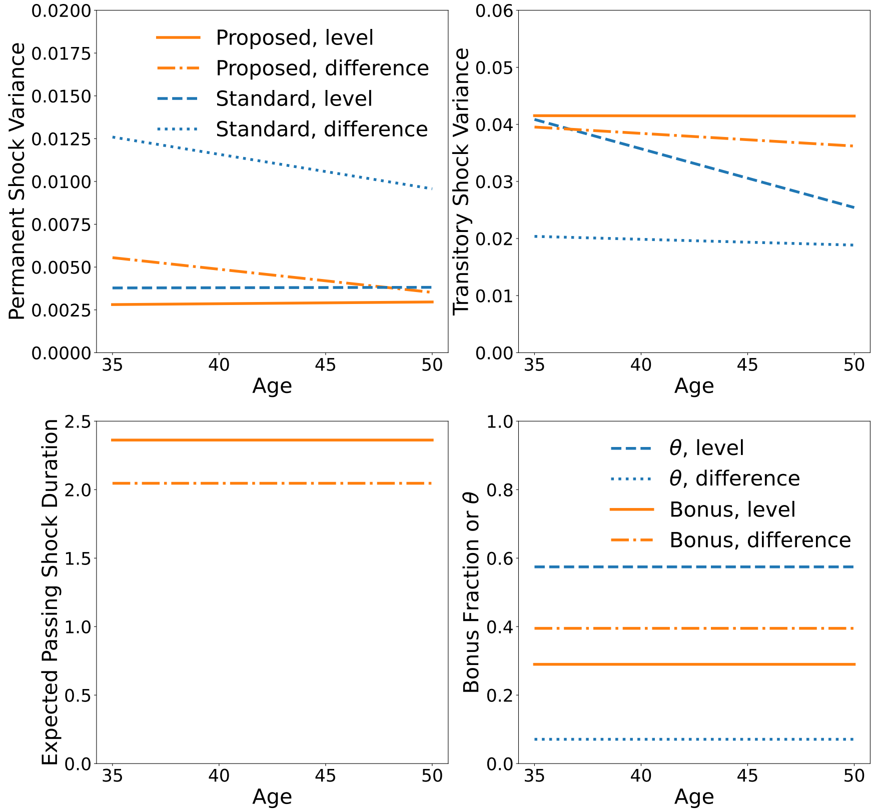
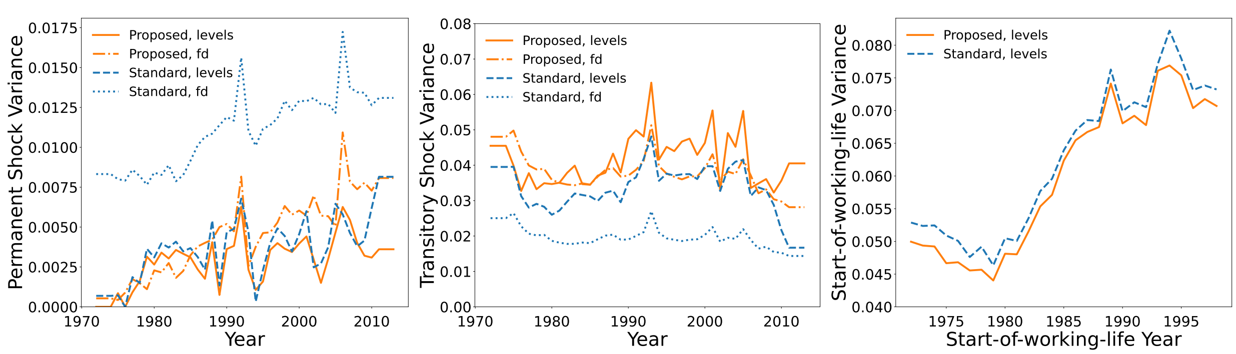


Figure 4: Age-varying estimates, Norwegian data.



(a) Permanent income variance **(b)** Transitory income variance **(c)** Start-of-working-life variance

Figure 5: Time-varying variance estimates, Norwegian data.

6.2 Results using the Panel Study of Income Dynamics

Above, we illustrated that our proposed model provides stable parameter estimates of income risk in the Norwegian administrative data. In this section, we show that our proposed model performs well also when applied to the much smaller sample sizes in the Panel Study of Income Dynamics (PSID). We first describe the data source and sample selection before we estimate our proposed model using data from the PSID.

PSID. The PSID has been the main source of data to estimate the idiosyncratic income process of households in the United States. In our analysis, we follow the data selection criteria of Moffitt and Zhang (2018), "the dataset consists of male heads from 1970 to 2014, 30-59 years old who were not full-time students, had positive weeks worked and wage and salary earnings, and which excludes non-sample men and all in PSID over-samples." Moreover, we only consider even year observations so that our estimation consistently uses data every two years.[23]

The importance of estimating using a balanced panel is laid out in Daly, Hryshko, and Manovskii (2022). Taking this lesson on board, we create 16 balanced panels from our underlying data, covering different time periods. Each panel spans 14 years (8 observations, each 2 years apart). The first spans from 1970 to 1984, the next from 1972 to 1986, and the last panel from 2000 to 2014. The idea is that 14 years is long enough to estimate the model, while the requirement that panels be balanced means there would be too few observations in longer panels.[24] Our panels in the PSID are far smaller than those from the Norwegian registry data. As a result, we do not allow the parameters to vary by age.[25] Furthermore, since the data only contain even-year observations, we do not identify θ in the standard model.

Income risk. Table 5 presents the estimated parameters using the proposed and standard models in PSID data.[26] There are four main takeaways.

First, as illustrated in the Norwegian data, the estimated parameters differ depending on the moments used when estimating using the standard model. For example, focusing

[23]The PSID was run annually until 1997 after which it has been run only every other year.

[24]Note that, unlike in the Norwegian data where cohorts were non-overlapping, in the set up here an individual can be used in two or more panels. For example, if they have complete income data from 1970 to 1986, their data will be used in both the 1970-1984 panel and the 1972-1986 panel. To calculate bootstrapped standard errors, we sample before building the panels.

[25]Somewhat reassuringly, the estimates from Norwegian data imply little age variation.

[26]Appendix E.1 provides bootstrapped confidence intervals for the PSID estimates. The large sample size of the Norwegian data makes the parameter estimate confidence bands small, and negligible next to the large model uncertainty.

	Identity		Diagonal		Optimal	
	Level	Difference	Level	Difference	Level	Difference
Panel A: Proposed Model						
σ^2_{perm}	0.012	0.013	0.009	0.013	0.008	0.008
σ^2_{tran}	0.049	0.066	0.059	0.062	0.043	0.044
τ	1.069	1.137	1.170	1.169	1.195	1.084
b	0.324	0.331	0.316	0.337	0.330	0.350
σ^2_{init}	0.067	✗	0.084	✗	0.089	✗
Panel B: Standard Model						
σ^2_{perm}	0.012	0.021	0.010	0.019	0.009	0.010
σ^2_{tran}	0.047	0.044	0.054	0.042	0.034	0.036
σ^2_{init}	0.069	✗	0.086	✗	0.089	✗

Notes: The table presents estimated parameters using the proposed and standard models on the PSID. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. Since the PSID data only contain even-year observations, we do not identify θ (the moving-average coefficient) in the standard model. The table shows the mean of parameter estimates over time.

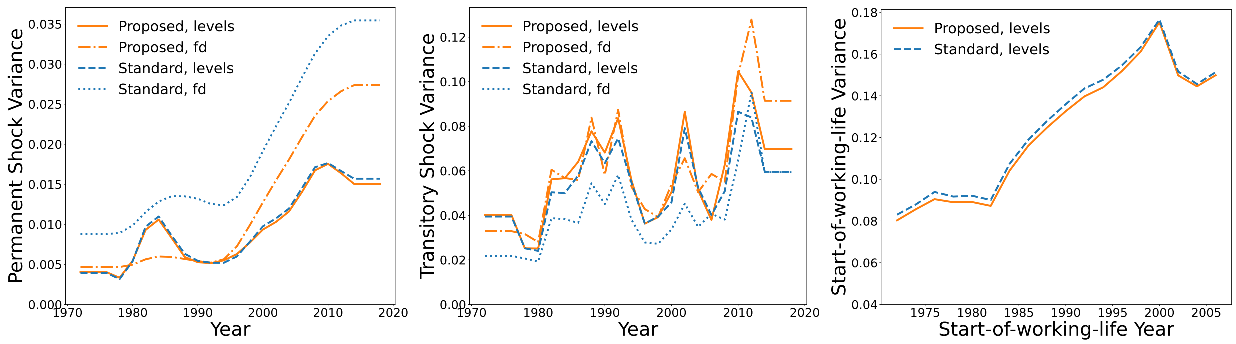
Table 5: Estimated parameters using PSID data.

on *diagonal*, permanent variance is almost twice as large when estimated using difference moments compared with level moments. In contrast, parameter estimates using our proposed model are more similar. For example, the estimated permanent income variance varies between 0.008 and 0.013, compared with 0.009 and 0.021 in the standard model.

Second, despite the two-year gap between surveys, we estimate the half-life of passing shocks to be a little over one year using the proposed model. We also find that these passing shocks make up more than half of the observed transitory income variance.

Third, as shown by Daly, Hryshko, and Manovskii (2022) and discussed extensively above, the standard model parameters are more similar when using the optimal weighting matrix.

Fourth, the parameter estimates from the standard model with level moments are closest to the estimates from the proposed model, suggesting that using level moments is the best available option if one has to rely on standard model estimates.



(a) Permanent income variance **(b)** Transitory income variance **(c)** Start-of-working-life variance

Figure 6: Time-varying variance estimates, PSID data.

How do we identify the persistence of passing shocks in PSID data? The existing literature that uses first difference moments in the PSID data has found that the covariance of income growth between differences more than two years apart is statistically insignificant. Therefore an MA(1) process is often chosen for the transitory component of income. Our estimations find a non-zero passing shock component, despite using PSID data with two-year intervals. One question is why our estimation finds these passing shocks to be present if the covariances on which it is estimated are statistically insignificant. The answer lies in the fact that our estimation makes use of *all* the covariances of income growth.

Section 4.1 demonstrated that the size of the variance for any transitory shock – an element in the diagonal of the covariance matrix – is equal to minus the sum of all the other covariances in that row or column. It is therefore the sum of these higher order covariances ($cov(\Delta y_t, \Delta y_{t+n})$ for all $n \geq 2$) that is relevant to measure by how much the standard model will underestimate the size of the transitory variance.

Although each of the higher-order covariances is individually insignificant, their sum is statistically significant. To demonstrate this, for each of our bootstrap samples we calculate $cov(\Delta y_t, \Delta y_{t+2})$ for all cohorts and years and sum these up. This value is negative in 65 out of 100 bootstrap samples, indicating that this individual covariance is not statistically significant. However, when we calculate $cov(\Delta y_t, \Delta y_{t+n})$ for all $2 \leq n \leq 6$ and sum up over all n, cohorts, and years, we find a negative value for all but one of our 100 bootstrap samples. That is, there is clear negative covariance between income growth in this period and future periods more than two years out, which is ruled out in an MA(1) specification.[27]

[27] We also test if all higher-order covariances above the second order are jointly equal to zero using an adapted version of the test in Hryshko and Manovskii (2022) and Abowd and Card (1989). We reject the

Income risk over time. Figure 6 displays the estimated permanent, transitory, and start-of-working-life variance over time estimated using PSID data. The main takeaway here is the same as for the Norwegian data. While the standard model estimated with difference moments tends to produce estimates that differ from the proposed model in levels of income risk, the time trends are relatively similar.

7 Conclusion

The standard permanent-transitory income process has a well-known problem: it is misspecified because the parameter estimates depend on the type of moments used and the weighting matrix applied. This paper proposes a parsimonious model of income dynamics that is more robust to the choice of moments or weighting matrix. Importantly, the model only requires one additional parameter compared with the standard model. Hence, one can estimate our proposed model using small data sets and include it in heterogeneous-agent models without additional state variables.

We reiterate our conclusion for practitioners here: use our proposed model. This model is robust when using specific moments (level or difference) and the weighing matrix (optimally, diagonally, or equally weighted) applied. When interpreting the existing literature, parameter estimates of the standard model using level moments and the identity or diagonal weighting matrix likely do a good job of decomposing permanent and transitory income risk. In contrast, parameter estimates of the standard model using first difference moments are likely to be biased to an economically significant extent.

null hypothesis at the 5 percent level. This test and its results are described in Appendix E.1.

References

Abowd, J. M., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

Altonji, J. G., A. A. Smith Jr., and I. Vidangos (2013): "Modeling Earnings Dynamics," *Econometrica*, 81(4), 1395–1454.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica*, 85(3), 693–734.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2017): "Life-Cycle Earnings, Education Premiums, and Internal Rates of Return," *Journal of Labor Economics*, 35(4), 993–1030.

Bloom, N., F. Guvenen, L. Pistaferri, J. Sabelhaus, S. Salgado, and J. Song (2017): "The great micro moderation," *Working paper*.

Blundell, R., M. Graber, and M. Mogstad (2015): "Labor income dynamics and the insurance from taxes, transfers, and the family," *Journal of Public Economics*, 127, 58–73.

Braxton, J. C., K. F. Herkenhoff, J. L. Rothbaum, and L. Schmidt (2021): "Changing Income Risk across the US Skill Distribution: Evidence from a Generalized Kalman Filter," Working Paper 29567, National Bureau of Economic Research.

Browning, M., M. Ejrnaes, and J. Alvarez (2010): "Modelling income processes with lots of heterogeneity," *Review of Economic Studies*, 77(4), 1353–1381.

Carr, M. D., R. A. Moffitt, and E. E. Wiemers (2020): "Reconciling Trends in Volatility: Evidence from the SIPP Survey and Administrative Data," *Working Paper*.

Crawley, E. (2020): "In search of lost time aggregation," *Economics Letters*, 189, 108998.

Crawley, E., M. B. Holm, and H. Tretvoll (2022): "A Parsimonious Model of Idiosyncratic Income," Finance and Economics Discussion Series 2022-026, Board of Governors of the Federal Reserve System.

Daly, M., D. Hryshko, and I. Manovskii (2022): "Improving the Measurement of Earnings Dynamics," *International Economic Review*, 63(1), 95–124.

De Nardi, M., G. Fella, M. Knoef, G. Paz-Pardo, and R. Van Ooijen (2021): "Family and government insurance: Wage, earnings, and income risks in the Netherlands and the US," *Journal of Public Economics*, 193, 104327.

De Nardi, M., G. Fella, and G. Paz-Pardo (2020): "Nonlinear household earnings dynamics, self-insurance, and welfare," *Journal of the European Economic Association*, 18(2), 890–926.

Druedahl, J., M. Graber, and T. Jørgensen (2021): "High Frequency Income Dynamics," *Working Paper*.

Gottschalk, P., and R. Moffitt (2009): "The rising instability of US earnings," *Journal of Economic Perspectives*, 23(4), 3–24.

Gottschalk, P., R. Moffitt, L. F. Katz, and W. T. Dickens (1994): "The growth of earnings instability in the US labor market," *Brookings Papers on Economic Activity*, 1994(2), 217–272.

Guvenen, F. (2009): "An Empirical Investigation of labor income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of US Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

Guvenen, F., A. McKay, and C. Ryan (2022): "A Tractable Income Process for Business Cycle Analysis," *Working Paper*.

Halvorsen, E., H. A. Holter, S. Ozkan, and K. Storesletten (2024): "Dissecting idiosyncratic earnings risk," *Journal of the European Economic Association*, 22(2), 617–668.

Halvorsen, E., S. Ozkan, and S. Salgado (2022): "Earnings Dynamics and Its Intergenerational Transmission: Evidence from Norway," *Quantitative Economics*, 13(4), 1707–1746.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51.

Heathcote, J., K. Storesletten, and G. L. Violante (2005): "Two views of inequality over the life cycle," *Journal of the European Economic Association*, 3(2-3), 765–775.

——— (2010): "The macroeconomic implications of rising wage inequality in the United States," *Journal of Political Economy*, 118(4), 681–722.

Hryshko, D., and I. Manovskii (2022): "How much consumption insurance in the U.S.?," *Journal of Monetary Economics*, 130, 17–33.

Low, H., C. Meghir, and L. Pistaferri (2010): "Wage Risk and Employment Risk over the Life Cycle," *American Economic Review*, 100(4), 1432–67.

McKinney, K. L., and J. M. Abowd (2020): "Male Earnings Volatility in LEHD before, during, and after the Great Recession," *Working Paper*.

Moffitt, R., C. Bollinger, C. Hokayem, E. Wiemers, J. Abowd, M. Carr, K. McKinney, S. Zhang, and J. Ziliak (2021): "Reconciling Trends in US Male Earnings Volatility," *Working Paper*.

Moffitt, R., and S. Zhang (2018): "Income Volatility and the PSID: Past Research and New Results," *AEA Papers and Proceedings*, 108, 277–80.

Moffitt, R. A., and P. Gottschalk (2002): "Trends in the transitory variance of earnings in the United States," *The Economic Journal*, 112(478), C68–C73.

Moffitt, R. A., and P. Gottschalk (2012): "Trends in the Transitory Variance of Male Earnings Methods and Evidence," *Journal of Human Resources*, 47(1), 204–36.

Moffitt, R. A., and S. Zhang (2020): "Estimating Trends in Male Earnings Volatility with the Panel Study of Income Dynamics," *Working Paper*.

Sabelhaus, J., and J. Song (2010): "The great moderation in micro labor earnings," *Journal of Monetary Economics*, 57(4), 391–403.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51(3), 609–633.

Online Appendix

A Appendix to Section 2

A.1 Further Details on the Norwegian Sample Selection

We restrict our sample to Norwegian-born males in the most stable part of their working life – ages 35 to 50. The age restriction is somewhat more restrictive than other studies in the literature, but we do this for several reasons. First, our large sample size allows us to restrict our analysis to the prime of working life without sacrificing accuracy. Second, the main focus of our paper is to reconcile known problems with the standard model, which is best done without introducing too many age-varying complications. Third, we see strong evidence in the Norwegian data that the type of model we analyze here – both the standard and proposed model – does not fit our data for the young and the old. Figure A.1 shows how the variance of income changes with age. In our models, this variance is expected to increase with age, which it does in the age range we restrict to. However, we see income variance declining in the 30-35 age range. Furthermore, we have found this cannot be explained by a reduction in transitory variance. Rather, we conjecture that it might come from mean-reversing permanent shocks as the young converge toward stable jobs from informal or part-time work. In addition, the age group above 50 shows an increase in permanent shocks that appears to be associated with early retirement. Further work in these directions may be fruitful but is beyond the scope of this paper.

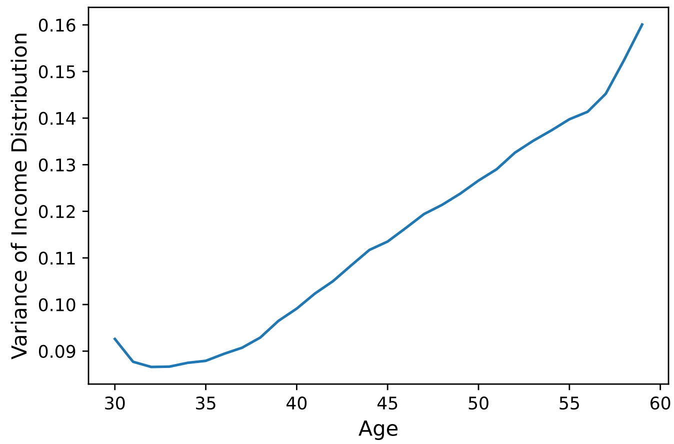


Figure A.1: The variance of the income distribution by age, averaged across cohorts.

B Appendix to Section 3

B.1 Identifying the Standard Model using Level Moments

If we use level moments, the covariance structure of the permanent component is, for $t < s$,

$$cov(p_t, p_s) = var(p_t) = \sigma^2_{p_0} + \sum_{i=1}^{t} \sigma^2_{\psi_i},$$

and the covariance structure of the transitory component is (define the transitory component as $q_t = \varepsilon_t + \theta\varepsilon_{t-1}$)

$$\begin{aligned} var(q_t) &= \sigma^2_{\varepsilon_t} + \theta^2\sigma^2_{\varepsilon_{t-1}} \\ cov(q_t, q_{t+1}) &= \theta\sigma^2_{\varepsilon_t} \\ cov(q_t, q_s) &= 0 \qquad \text{if } s > t+1. \end{aligned}$$

Using the independence of the permanent and transitory components of income, the covariance structure of log income is the sum of the covariance structure of each component

$$cov(y_t, y_s) = cov(p_t, p_s) + cov(q_t, q_s).$$

One can then use these theoretical moments and find the parameters that minimize the distance between the empirically observed moments and those implied by the model.

C Appendix to Section 4

C.1 An Example of Exact Identification of the Proposed Model

It is possible from equations (18), (20), and (22) to identify all the parameters, except the initial variances, from the difference moments – in fact, our model is overidentified using all the available moments. Here, we provide a way to identify the parameters exactly if the difference moments are known.

$$
\begin{aligned}
cov\left(\Delta y_t, \Delta y_{t+2}\right) &= -\left(1-\rho_{\text{passing}}\right)\rho_{\text{passing}}\left(\sigma^2_{v_t} - \rho_{\text{passing}}\sigma^2_{v_{t-1}}\right) \\
cov\left(\Delta y_t, \Delta y_{t+3}\right) &= -\left(1-\rho_{\text{passing}}\right)\rho^2_{\text{passing}}\left(\sigma^2_{v_t} - \rho_{\text{passing}}\sigma^2_{v_{t-1}}\right) \\
\implies \rho_{\text{passing}} &= \frac{cov\left(\Delta y_t, \Delta y_{t+3}\right)}{cov\left(\Delta y_t, \Delta y_{t+2}\right)}.
\end{aligned}
$$

We assume $\sigma^2_{v_1} = \sigma^2_{v_0}$ and solve the above two equations with $t = 1$ for $\sigma^2_{v_0}$ and then iterate forward to identify $\sigma^2_{v_t}$.

To identify $\sigma^2_{\zeta_t}$ (the bonus shock variance), we can use the moment

$$
cov\left(\Delta y_t, \Delta y_{t+1}\right) = -\sigma^2_{\zeta_t} - \left(1-\rho_{\text{passing}}\right)\left(\sigma^2_{v_t} - \rho_{\text{passing}}\sigma^2_{v_{t-1}}\right)
$$

where only $\sigma^2_{\zeta_t}$ is now unknown. Finally, the permanent variance is now the only remaining parameter and can be identified from:

$$
Var(\Delta y_t) = \sigma^2_{\psi_t} + \sigma^2_{\zeta_t} + \sigma^2_{\zeta_{t-1}} + \sigma^2_{v_t} + (1-2\rho_{\text{passing}})\sigma^2_{v_{t-1}}.
$$

C.2 An Alternative Passing Shock Process

In the main text, we define the passing shock process as remaining at a fixed level until it jumps up or down to a new fixed level. We estimate the time between jumps to be about two years, which seems too long to represent an unemployment spell. Another process that gives rise to the same moments is one in which shocks are always positive, but when there are no positive shocks, income returns to trend. These shocks can be thought of as promotions – immediately after the promotion, income jumps above others. Still, the other workers also get promoted, which reduces the size of the income residual until the time of the next promotion. This type of process may be a better fit for the length of time between shocks that we estimate. Such a process can be modeled as follows:

$$v_t = \begin{cases} \theta_{\text{passing}} v_{t-1} & \text{with probability } \rho_{\text{passing}} \\ \varepsilon_t^v & \text{with probability } 1 - \rho_{\text{passing}}. \end{cases} \tag{A.1}$$

The covariance matrix of this process is given by

$$cov(v_t, v_s) = (\theta_{\text{passing}} \rho_{\text{passing}})^{s-t} \sigma_{v_t}^2.$$

That is the moments are identical to the process described in the main text except the parameter ρ_{passing} is multiplied by θ_{passing}. Only the product of the two parameters is identified with our method.

C.3 AR(1) and Passing Shock Covariance Matrices

Consider an AR(1) process

$$v_t = \rho_{\text{passing}} v_{t-1} + \varepsilon_t.$$

The covariance matrix of this process is

$$cov(v_t, v_s) = \rho_{\text{passing}}^{s-t} \sigma_{v_t}^2,$$

which is the same as the covariance restriction with a passing shock in equation (21) in the paper. Hence, the covariance matrix of an AR(1) process is isomorphic to the passing shock process we describe in the main text, and one can reinterpret our estimate of the passing shock as the persistence of an AR(1) process.

D Appendix to Section 5

D.1 Replicating Daly, Hryshko, and Manovskii (2022)

As mentioned in Section 1, the paper most closely related to ours is Daly, Hryshko, and Manovskii (2022) (DHM). Therefore, this section replicates their results on Norwegian data by comparing estimates using moments in levels and differences. They use registry data from Denmark and Germany, and here we aim to mimic as closely as possible the three different sample selection criteria they apply to the Danish data.

We again use the Norwegian registry data described in section 2 and restrict our attention to males born in Norway. As DHM do in the Danish data, we further restrict attention to those born from 1951 to 1955 and only use income data from 1981 to 2006. We also drop individuals whose educational status has changed during their longest spell (discussed further below). Outliers are defined as year-to-year earnings increases of more than 500 percent or a decrease of more than 80 percent. Individuals with earnings outliers within their longest spell are dropped.

There are two selection criteria that DHM apply to the Danish data that we cannot replicate exactly in the Norwegian data:

1. DHM drop records where individuals worked less than 10 percent of the year as a full-time employee.
2. DHM remove individuals who were ever self-employed.

We handle both of these by referring to the Norwegian social security system's definition of a base level of income ("grunnbeløp" which is abbreviated to 'G'), used as a basis for calculating various social security and pension benefits. The first criterion mentioned above is handled by dropping observations where income is below 1G (approximately USD 10,000). This should capture individuals who are only loosely attached to the labor force during the year. The second criterion drops observations where business income is above 1G which should capture those who are self-employed.[28]

DHM's focus is on the importance of using a balanced sample, and to highlight this they contrast estimates obtained from three different samples in their paper. The first sample ("Balanced") only keeps individuals where observations are available for all 26 years they consider. The second sample ("9 consec.") constructs spells of consecutive observations for each individual and only keeps those individuals whose longest spell contains at least 9 consecutive observations. Only the observations within that longest

[28]However, the measure of business income is only available from 1993. Hence, this sample selection criterion is only applied for the years between 1993 and 2006.

Sample	Norwegian data	Danish data (from DHM)
Sample 1 - Balanced	71,825	67,008
Sample 2 - 9 consec.	98,078	102,825
Sample 3 - 20 not nec. consec.	90,305	90,668

Table A.1: Number of individuals in the three samples in Norwegian and Danish data

spell are kept so there are no gaps in the resulting data set. The third sample ("20 not nec. consec.") keeps individuals with at least 20 income observations available but does not require that these are consecutive. So in this sample, the "longest spell" is irrelevant, and the data can contain gaps. Table A.1 shows the number of individuals we obtain in the three different samples and compares them to the numbers obtained by DHM in the Danish data.

Tables A.2, A.3, and A.4 present results from estimating the standard model from Section 3 adjusted to allow for persistent but not permanent shocks

$$p_t = \rho p_{t-1} + \psi_t \tag{A.2}$$

where ρ can be less than 1. The estimation uses both moments in levels and differences for each sample.

Table A.2 shows that we get similar results to DHM when we follow their approach and use the optimal weighting matrix – that is, the inverse of the variance-covariance matrix of the data moments – in the estimation. Columns (1) – (4) show results for the samples with 9 or more consecutive observations and 20 or more observations that are not necessarily consecutive. For both of these samples, we see the same patterns that DHM report for the Danish data: The estimated persistence and variance of the permanent shock as well as the estimated persistence of the transitory shock are higher when using difference moments, while the estimated variance of the transitory shock is higher when using moments in levels. Columns (5) and (6) show that these differences disappear when using the balanced sample where individuals are only included if data is available for all 26 years. Thus the estimation of Norwegian registry data gives results very similar to the ones obtained by DHM for Danish (and German) data.

Tables A.3 and A.4 show that using a different weighting matrix in the estimation—respectively the inverse of a diagonal weighting matrix using only the variance of the data moments and the identity matrix—does not yield the same results as in DHM. In both estimations, we get that using difference moments leads to a higher estimated variance of the permanent shock and a lower estimated variance of the transitory shock even in the

	9 consec.		20 not nec. consec.		Balanced	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
ρ	0.952	0.990	0.967	0.981	0.970	0.975
	(0.001)	(0.0003)	(0.0006)	(0.0005)	(0.0006)	(0.0008)
σ^2_{perm}	0.010	0.015	0.008	0.013	0.006	0.006
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
θ	0.221	0.250	0.195	0.263	0.273	0.272
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
σ^2_{tran}	0.017	0.009	0.019	0.010	0.009	0.009
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
σ^2_{init}	0.025	—	0.027	—	0.025	—
	(0.0004)	—	(0.0004)	—	(0.0004)	—

Notes: The parameter ρ is the persistence of the permanent component, and θ is the moving-average coefficient. Asymptotic standard errors are in parentheses.

Table A.2: Estimated parameters using Norwegian data and DHM's sample selection criteria. Weighting matrix = DHM (= optimal weighting matrix).

balanced sample. The estimated values of persistence also depend on the moments used, but the ranking differs depending on the weighting matrix.

That the choice of weighting matrix affects the estimation results in this way is yet another indication that the standard model is misspecified and that simply using a balanced panel does not fix the issue. We remain convinced that using a balanced panel is important for the reasons pointed out by DHM, but the observation that the estimated values in columns (5) and (6) of table A.2 are the same does not necessarily imply that these estimates are correct. As shown in the estimation on simulated data presented in table 3, obtaining the same estimated values using moments in levels and differences does not guarantee that the estimations recover the true parameter values.

	9 consec.		20 not nec. consec.		Balanced	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
ρ	0.968	0.990	0.977	0.889	0.979	0.828
	(0.0008)	(0.0003)	(0.0007)	(0.0022)	(0.0008)	(0.0036)
σ^2_{perm}	0.008	0.017	0.006	0.017	0.005	0.011
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
θ	0.477	0.233	0.440	0.206	0.737	0.211
	(0.006)	(0.003)	(0.006)	(0.005)	(0.019)	(0.005)
σ^2_{tran}	0.023	0.009	0.026	0.008	0.012	0.007
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)
σ^2_{init}	0.025	—	0.029	—	0.026	—
	(0.0004)	—	(0.0004)	—	(0.0004)	—

Notes: The parameter ρ is the persistence of the permanent component, and θ is the moving-average coefficient. Asymptotic standard errors are in parentheses.

Table A.3: Estimated parameters using Norwegian data and DHM's sample selection criteria. Weighting matrix = diagonal.

	9 consec.		20 not nec. consec.		Balanced	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
ρ	0.969	0.992	0.981	0.987	0.983	0.988
	(0.0008)	(0.0003)	(0.0008)	(0.0004)	(0.0008)	(0.0005)
σ^2_{perm}	0.007	0.018	0.006	0.018	0.005	0.011
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
θ	0.466	0.239	0.460	0.239	0.726	0.245
	(0.006)	(0.003)	(0.006)	(0.003)	(0.017)	(0.004)
σ^2_{tran}	0.024	0.010	0.028	0.010	0.014	0.008
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
σ^2_{init}	0.026	—	0.031	—	0.027	—
	(0.0004)	—	(0.0004)	—	(0.0004)	—

Notes: The parameter ρ is the persistence of the permanent component, and θ is the moving-average coefficient. Asymptotic standard errors are in parentheses.

Table A.4: Estimated parameters using Norwegian data and DHM's sample selection criteria. Weighting matrix = identity.

D.2 Sketch Proof of Proposition 1

This proof aims to show that using the optimal minimum distance estimator will give the same parameter estimates regardless of using level or difference moments when each individual has sufficient panel observations. There are two steps in the proof. We first show that the optimal minimum distance estimator gives the same result when the moments are transformed by an invertible linear mapping. We next show that one can construct an invertible linear mapping that transforms level moments into difference moments but includes some additional terms. However, as we add more panel observations, the contribution of these additional terms converges to zero such that the optimal minimum distance estimator will yield the same result regardless of which moments are used.

Assume that $Y = (y_1, ..., y_T)'$ is a random variable, that $f(Y)$ is a function that generates data moments from Y, and that $g(\theta)$ is a function generating model moments for a set of parameters θ. We can then formulate the original optimal minimum distance problem (OMD) as

$$\underset{\theta}{\text{argmin}}\, \mathbb{E}(f(Y) - g(\theta))' \Omega^{-1} \mathbb{E}(f(Y) - g(\theta))$$

where $\Omega = \mathbb{E}\Big((f(Y) - \mathbb{E}(f(Y)))(f(Y) - \mathbb{E}(f(Y)))'\Big)$.

Now, for any invertible linear map A, we have

$$\begin{aligned}
\text{Objective} &= \mathbb{E}(f(Y) - g(\theta))' \Omega^{-1} \mathbb{E}(f(Y) - g(\theta)) && \text{(A.3)}\\
&= \mathbb{E}(f(Y) - g(\theta))' A' (A')^{-1} \Omega^{-1} A^{-1} A \mathbb{E}(f(Y) - g(\theta)) && \text{(A.4)}\\
&= \mathbb{E}(Af(Y) - Ag(\theta))' (A\Omega A')^{-1} \mathbb{E}(Af(Y) - Ag(\theta)) && \text{(A.5)}\\
&= \mathbb{E}(Af(Y) - Ag(\theta))' \tilde{\Omega}^{-1} \mathbb{E}(Af(Y) - Ag(\theta)) && \text{(A.6)}
\end{aligned}$$

where $\tilde{\Omega} = \mathbb{E}\Big((Af(Y) - \mathbb{E}(Af(Y)))(Af(Y) - \mathbb{E}(Af(Y)))'\Big)$. Hence, solving the optimal minimum distance problem of $f(Y)$ and $g(\theta)$ is equivalent to solving it for $Af(Y)$ and $Ag(\theta)$.

To show that the level and difference moments give the same estimates under the optimal minimum distance estimator, we have to show that there exists such an A transforming level moments into difference moments. The level moments are defined as

$$f(Y) = vech\left(Y'Y\right)$$

and the difference moments are defined as

$$f(\Delta Y) = vech\left((\Delta Y)'(\Delta Y)\right).$$

Since Y is of length T and ΔY is of length $T-1$, $f(Y)$ and $f(\Delta Y)$ have different dimensions and there does not exist any invertible A such that $Af(Y) = f(\Delta Y)$. However, we can construct an invertible linear mapping A with some extra terms:

$$Af(Y) = [y_1^2, y_1\Delta y_1, y_1\Delta y_2, ..., y_1\Delta y_T, f(\Delta Y)].$$

Given that ΔY is an ergodic process, as $T \to \infty$ the variance of parameter estimates that can be identified using the optimal weighting matrix on the difference moments eventually declines linearly with T. Ergodicity also implies that $y_1\Delta y_T$ goes to zero in expectation as $T \to \infty$. Thus, as $T \to \infty$, these extra terms available to the optimal minimum distance estimator eventually do not help to reduce the parameter variance. Therefore, as $T \to \infty$, the optimal minimum distance parameter estimates using levels must converge to those using first differences.

This proof does not apply to other weighting methods, such as equal or diagonal. For a generic weighting matrix Ω, it is not the case that $A\Omega A'$ is the counterpart weighting matrix under an invertible linear map A.

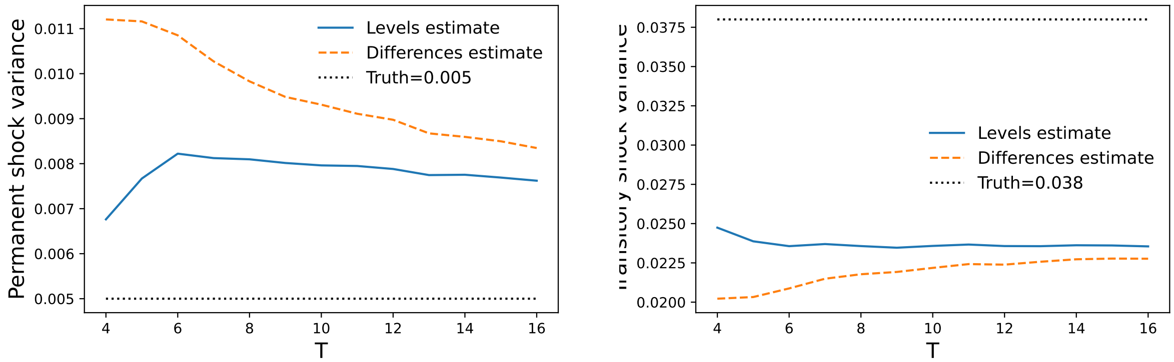
D.3 Importance of T for Proposition 1

Proposition 1 explains how the parameter estimates of the standard model become similar when using the optimal weighting matrix. The result hinges on the assumption that $T \to \infty$, while T is typically small in practical applications. We, therefore, include Figure A.2 which illustrates how the parameter estimates of permanent and transitory variance depend on the size T when estimating the standard model using the optimal weighting matrix in simulated data from the proposed model. We use the same parameters in the simulation as in Section 5, but vary the T.

For small T, the difference between the parameter estimates is quite large. This difference decreases gradually as we increase T. For $T = 16$, as in Section 5, the parameter estimates are reasonably close. Figure A.2 also illustrates that while the parameter estimates converge, they do not converge to the correct value in the simulation, illustrating the general point that convergence of parameter estimates does not imply correct parameter estimates if the model is misspecified.

D.4 Sketch Proof of Proposition 2

Suppose the data-generating process for income consists of initial permanent income, a permanent income component that follows a random walk, and transitory shocks. The



(a) Permanent income variance. **(b)** Transitory income variance.

Notes: The figures show estimated permanent and transitory shock variance using the standard model on simulated data with varying panel length (T). Level and difference denote the type of moments used, corresponding to (10) and (11), respectively.

Figure A.2: Importance of T for Proposition 1.

transitory component, v_t, is defined such that the covariances of order T tends to zero as $T \to \infty$. Formally, given any $\delta > 0$, there exists a T_δ such that $cov(v_t, v_{t+k}) < \delta$ for all $k > T_\delta$. Assume that $Y = (y_0, y_1, ..., y_T)$ is random variable of length T generated from this income process. The covariance matrix for this process can be summarized as:

$$var(y_t) = \sigma^2_{init} + t\sigma^2_{perm} + \sigma^2_{trans} \tag{A.7}$$

$$cov(y_t, y_{t+s}) \text{ is unconstrained} \qquad \text{if } s \leq T_\delta \tag{A.8}$$

$$|cov(y_t, y_{t+s}) - \sigma^2_{init} - t\sigma^2_{perm}| < \delta \qquad \text{if } s > T_\delta \tag{A.9}$$

Then, $cov(y_t, y_{t+s})$ for $s > T_\delta$ will provide information only about the initial permanent variance and the variance of permanent innovations. Any model that assumes a unit root will therefore identify the parameters of this process from these covariances. Using the identity matrix as the weighting matrix ensures that these covariances always have a non-zero weight.

For example, consider the standard model with an MA(1) transitory process but where all parameters are constant across time. Then

$$var(y_t) = \sigma^2_{p_0} + t\sigma^2_\psi + (1+\theta^2)\sigma^2_\varepsilon \quad \text{(A.10)}$$

$$cov(y_t, y_{t+1}) = \sigma^2_{p_0} + t\sigma^2_\psi + \theta\sigma^2_\varepsilon \quad \text{(A.11)}$$

$$cov(y_t, y_{t+s}) = \sigma^2_{p_0} + t\sigma^2_\psi \qquad \text{if } 1 < s \leq T_\delta \quad \text{(A.12)}$$

$$cov(y_t, y_{t+s}) = \sigma^2_{p_0} + t\sigma^2_\psi \qquad \text{if } s > T_\delta \quad \text{(A.13)}$$

In this case, the transitory part of the income process is misspecified because (A.12) assumes that there is no influence of transitory shocks on this covariance, while there is such an influence in the data-generating process. Using this model, the permanent variance will be identified from equations (A.12) and (A.13). As $T \to \infty$, it will be entirely identified by (A.13) because the relative weight assigned to (A.12) will go to zero ($T \to \infty$ but T_δ is fixed, and the weight on $cov(y_t, y_{t+s})$ will tend to a fixed positive number under the *identity* weighting matrix). As (A.13) is identical to (A.9) up to an arbitrarily small error, δ, the estimation of σ^2_{perm} and σ^2_{init} will be consistent.

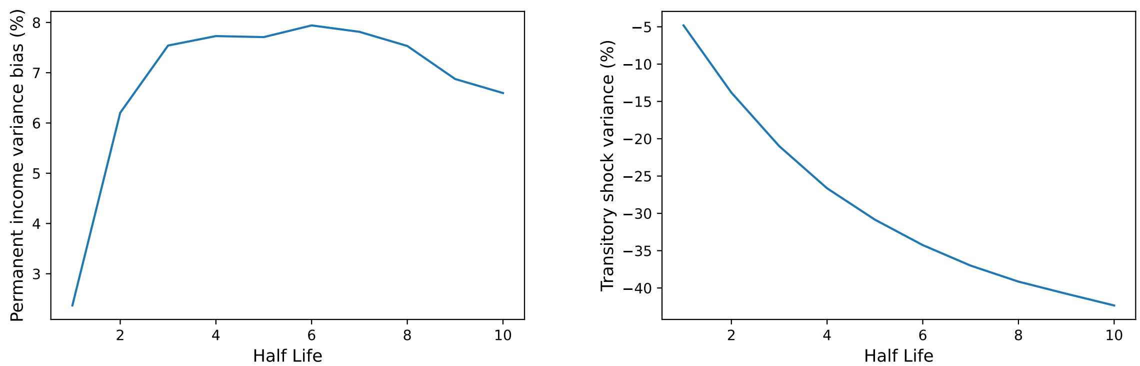
D.5 The Relevance of Proposition 2 in Practice

Proposition 2 states that the standard model, under the equally-weighted estimator using level covariances, provides unbiased estimates for the permanent variance as the panel length tends to infinity. However, it does not guide how long the panel needs to be in practice for reasonable parameter values. Intuitively, if the panel dimension is too short or the passing shock is too large and persistent, the parameter estimates will be biased because there are not enough 'long' covariances to identify the permanent variance.

In the context of our proposed model, the extent of bias depends on a combination of model parameters and the length of the panel. Here, we test the extent of the bias for different values of the parameters along three dimensions: persistence of the passing shock, size of the passing shock, and length of the panel.

Passing shock persistence. Here, we estimate the standard model using the identity matrix on simulated data from the proposed model. We use the same parameters as in Section 5 but vary the half-life of the passing shock τ.

Figure A.3 illustrates the bias on permanent and transitory shock variance as we increase the half-life of the passing shock. For the values of half-life we estimate in the paper and used in the simulation, the bias is relatively small, around 5% for the permanent shock and 10% for the transitory shock. This bias increases as the half-life of the passing shock increases, especially for the transitory shock variance.



(a) Permanent income variance bias.

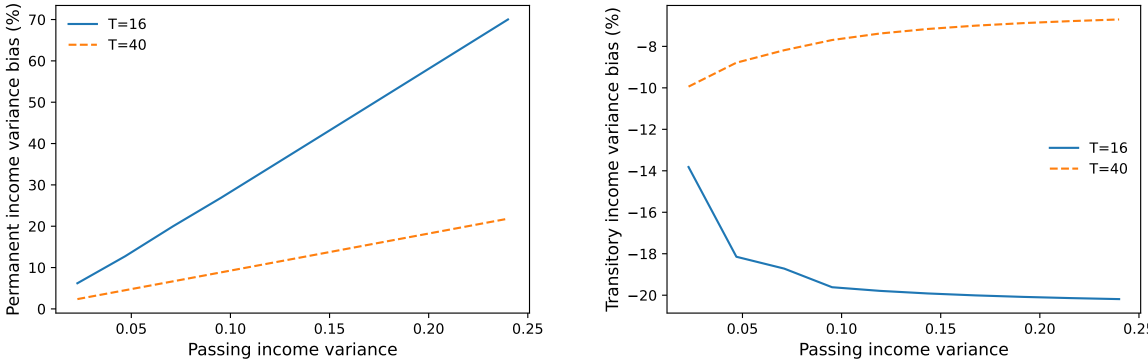
(b) Transitory income variance bias.

Notes: The figures show estimated permanent and transitory shock variance bias using the standard model on simulated data with varying passing shock half-life (τ). The estimation uses level moments (10) and the equally-weighted minimum distance method.

Figure A.3: Importance of τ (Half-life of Passing Shock) for Proposition 2.

Size of passing shock and length of panel. Figure A.4 shows the size of the bias—for $T = 16$ and $T = 40$—in the case that the passing shock variance is much larger than we estimate in the Norwegian data, up to 0.24, while the permanent and bonus shock variances remain unchanged relative to our simulations in section 5. The size of the permanent shock variance bias rises almost linearly with the size of the passing shock variance. As a result, this bias is small for the size of passing shock that we estimate in the Norwegian data, but can grow large for extreme values for the size of the passing shock variance. When the passing shock variance is 0.24, this bias is as large as 70 percent for $T = 16$ and 20 percent for $T = 40$.

D.6 Further Simulation Results



(a) Permanent income variance bias. **(b)** Transitory income variance bias.

Notes: The figures show estimated permanent and transitory shock variance bias using the standard model on simulated data with varying passing shock variance. The estimation uses level moments (10) and the equally-weighted minimum distance method.

Figure A.4: Importance of the size of the passing shock variance for Proposition 2.

This Appendix present results of estimating the standard model in simulated data when the simulated model has a shorter panel (T=5) or includes a persistent, but not permanent income process.

Specifically, in the simulation of a persistent income process, we replace (14) with the following

$$p_t = \rho p_{t-1} + \psi_t \tag{A.14}$$

with $\rho \in [0, 1)$.

Table A.5 shows estimation results for the standard model for simulations with varying panel length and when we introduce some decay to the permanent shock in the simulation. The estimation results reported are for the standard model where we assume in the estimation that $\rho = 1$.

Two results are noteworthy. First, reducing the panel size to $T = 5$ results in the level estimation mildly underestimating the transitory income variance, while the difference estimates are unchanged and remain far from the true parameter values. Second, if the 'permanent' income shock decays even slowly over time, estimating the standard model with $\rho = 1$, especially for long panels, results in an overestimation of the transitory variance and an underestimation of the permanent variance when using level moments. The difference estimates are little changed.

Parameter	True Value	Identity Level	Identity Difference	Diagonal Level	Diagonal Difference	Optimal Level	Optimal Difference
Panel A: $T = 5$, $\rho = 1$							
σ^2_{perm}	0.005	0.005	0.012	0.005	0.012	0.008	0.011
σ^2_{tran}	0.038	0.028	0.020	0.028	0.020	0.024	0.021
τ	2.0 years	✗	✗	✗	✗	✗	✗
b	0.40	✗	✗	✗	✗	✗	✗
θ	✗	0.25	0.08	0.25	0.08	0.15	0.09
σ^2_{init}	0.065	0.074	✗	0.073	✗	0.071	✗
Panel B: $T = 16$, $\rho = 0.97$							
σ^2_{perm}	0.005	0.002	0.011	0.002	0.011	0.005	0.007
σ^2_{tran}	0.038	0.040	0.020	0.041	0.020	0.026	0.024
τ	2.0 years	✗	✗	✗	✗	✗	✗
b	0.40	✗	✗	✗	✗	✗	✗
θ	✗	0.55	0.07	0.57	0.07	0.18	0.15
σ^2_{init}	0.065	0.049	✗	0.048	✗	0.041	✗
Panel C: $T = 5$, $\rho = 0.97$							
σ^2_{perm}	0.005	0.003	0.011	0.003	0.011	0.006	0.011
σ^2_{tran}	0.038	0.031	0.020	0.032	0.020	0.026	0.021
τ	2.0 years	✗	✗	✗	✗	✗	✗
b	0.40	✗	✗	✗	✗	✗	✗
θ	✗	0.28	0.08	0.28	0.08	0.18	0.10
σ^2_{init}	0.065	0.059	✗	0.059	✗	0.056	✗

Notes: The table presents estimated parameters using the standard models on data generated for 200,000 individuals from the proposed model. 'True Value' refers to the parameters used in the data-generating process. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. The parameter θ is the moving-average coefficient.

Table A.5: Estimated standard-model parameters using simulated data.

E Appendix to Section 6

E.1 Further detail on the PSID data and results

Bootstrapped standard errors. Table A.6 shows the PSID estimation results for both the proposed and standard model, level and difference moments, along with their bootstrapped confidence intervals in parentheses below.

Statistical significance of higher-order covariances. We run two tests to show that higher-order covariances are statistically significantly different from zero. The first test, described in the main text, is to sum up all the higher order covariances in all our bootstrap

	Identity		Diagonal		Optimal	
	Level	Difference	Level	Difference	Level	Difference
Panel A: Proposed Model						
σ^2_{perm}	0.012	0.013	0.009	0.013	0.008	0.008
	(0.010, 0.014)	(0.008, 0.016)	(0.008, 0.011)	(0.009, 0.015)	(0.006, 0.008)	(0.005, 0.009)
σ^2_{tran}	0.049	0.066	0.059	0.062	0.043	0.044
	(0.042, 0.056)	(0.052, 0.075)	(0.050, 0.064)	(0.049, 0.067)	(0.030, 0.041)	(0.033, 0.047)
τ	1.069	1.137	1.170	1.169	1.195	1.084
	(0.967, 1.131)	(1.038, 1.336)	(1.006, 1.244)	(1.057, 1.489)	(1.023, 1.419)	(0.922, 1.657)
b	0.324	0.331	0.316	0.337	0.330	0.350
	(0.319, 0.338)	(0.294, 0.418)	(0.312, 0.337)	(0.307, 0.390)	(0.302, 0.360)	(0.287, 0.447)
σ^2_{init}	0.067	✗	0.084	✗	0.089	✗
	(0.056, 0.085)	✗	(0.072, 0.098)	✗	(0.076, 0.098)	✗
Panel B: Standard Model						
σ^2_{perm}	0.012	0.021	0.010	0.019	0.009	0.010
	(0.010, 0.014)	(0.016, 0.025)	(0.008, 0.011)	(0.015, 0.022)	(0.007, 0.009)	(0.008, 0.011)
σ^2_{tran}	0.047	0.044	0.054	0.042	0.034	0.036
	(0.040, 0.053)	(0.036, 0.051)	(0.046, 0.058)	(0.032, 0.047)	(0.022, 0.033)	(0.025, 0.034)
σ^2_{init}	0.069	✗	0.086	✗	0.089	✗
	(0.058, 0.088)	✗	(0.074, 0.101)	✗	(0.068, 0.090)	✗

Notes: The table presents estimated parameters using the proposed and standard models on the PSID. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. Since the PSID data only contain even-year observations, we do not identify θ (the moving-average coefficient) in the standard model. The table shows the mean of parameter estimates over time.

Table A.6: Estimation using PSID with bootstrapped standard errors.

samples and count how many are negative. The second, described in detail below, is an adapted version of the test for higher-order covaraînces being jointly equal to zero, as found in Hryshko and Manovskii (2022) and Abowd and Card (1989).

We cannot directly run the test in Abowd and Card (1989) because of the way that we create our balanced panels in the PSID data. As described in the main text, for each panel we select all individuals who have a complete set of data for the panel's time period and, as such, an individual can appear in more than one panel. For the adapted test we run here, we create one large panel in a slightly modified way. First, for each individual we select at random one of the periods in which he has a complete set of data for six consecutive waves of the PSID. We then stack all these individual's income outcomes in one large panel and calculate a single 6 by 6 covariance matrix. We then run the test of joint significant for all higher-order covariances ($n \geq 2$). Note that this test does not use all the data, as we only allow each individual's data for 6 waves so that we can achieve a balanced panel. We run this test 40 times, each time with a different random choice for the period we choose for each individual. The average of these tests rejects the null hypothesis at the 5 percent level (the chi-squared statistic 18.8, greater than the 18.3 level for statistical significance at the 5 percent level with 10 degrees of freedom).

Sample in relation to existing literature. We start with the same selection criteria as that in Moffitt and Zhang (2018) and then select for individuals who have at least six consecutive waves of data available in the PSID. Figure A.5 shows the variance of two-year changes in the log income residual in four ways. First, using the sample selection criteria of Moffitt and Zhang (2018). This line closely matches that in Figure 1 of Moffitt and Zhang (2018) except we only show the even-year variances. Second, we show the two-year variances under our extra criteria that we maintain balanced panels over 6-wave periods. This criterion reduces the level of volatility in our sample relative to Moffitt and Zhang (2018). The final two lines show the model-implied two-year variances for both our proposed model and the standard model when both are estimated on the difference moments. Both models closely match the two-year variances.

The time-series of permanent and transitory shock variance estimates shows a similar pattern that that in Moffitt and Zhang (2018), despite the extra selection criterion. The pattern is different to that in Heathcote, Perri, and Violante (2010) because they use different underlying data. The primary difference is that, unlike our paper and Moffitt and Zhang (2018), Heathcote, Perri, and Violante (2010) estimate the standard model on hourly wages instead of total labor income.

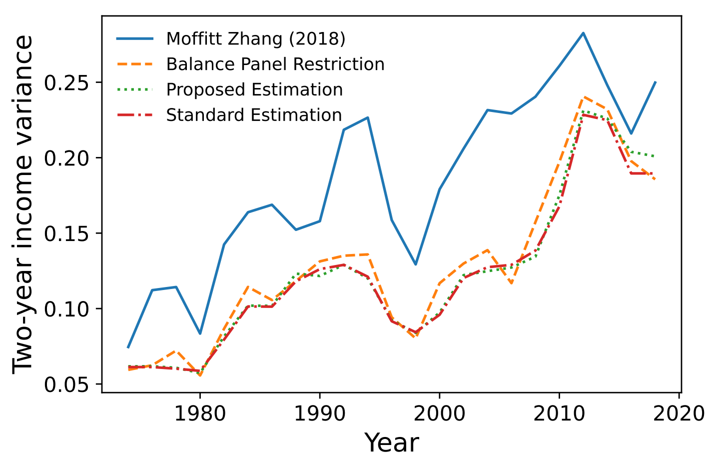


Figure A.5: Two-year income variance.

E.2 Estimation using persistent (but not permanent) shocks

The paper restricts attention to estimating income processes where the permanent component is a unit root. In this appendix, we present the paper's main results when we allow the persistent part of the income process to have an arbitrary persistence.

	Identity		**Diagonal**		**Optimal**	
	Level	Difference	Level	Difference	Level	Difference
Panel A: Proposed Model						
σ^2_{perm}	0.003	0.008	0.003	0.005	0.003	0.005
σ^2_{tran}	0.041	0.028	0.041	0.035	0.038	0.032
τ	2.238	1.469	2.328	1.910	2.206	1.626
b	0.296	0.541	0.297	0.426	0.372	0.434
ρ	1.000	0.965	1.000	0.988	0.998	0.988
σ^2_{init}	0.061	✗	0.060	✗	0.058	✗
Panel B: Standard Model						
σ^2_{perm}	0.010	0.014	0.010	0.015	0.011	0.010
σ^2_{tran}	0.022	0.017	0.022	0.016	0.019	0.019
θ	0.216	0.026	0.233	0.000	0.088	0.092
ρ	0.923	0.894	0.921	0.809	0.926	0.952
σ^2_{init}	0.063	✗	0.062	✗	0.058	✗

Notes: The table presents estimated parameters using the proposed and standard models on Norwegian data. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. The parameter ρ is the persistence of the permanent component, and θ is the moving-average coefficient. The table shows the mean of parameter estimates over time and age.

Table A.7: Estimation with persistent, but not permanent shock.

Specifically, we adjust the permanent component of both income processes to be

$$p_t = \rho p_{t-1} + \psi_t \tag{A.15}$$

where the estimation allows for a ρ that can be different from 1.

Table A.7 presents the estimated income processes on Norwegian data when we allow for a $\rho \leq 1$. The proposed model provides consistent estimates of the model parameters even in this more general case. In particular, across all combinations of moments and weighting matrices applied, the estimates of ρ are close to 1, and the estimated permanent and transitory variances are stable.

In contrast, estimates of the standard model vary more depending on moments, and weighting matrices applied. For example, the estimates of ρ vary from 0.81 to 0.95. The estimates of the permanent variance also vary more. However, when adjusted for the persistence of the permanent component (annual variance = $\sigma^2_{\text{perm}}/(1-\rho^2)$), they are only slightly higher than the permanent variance estimates from the proposed model. Moreover, when we estimate using the optimal weighted minimum distance method, we get similar results regardless of the moments used, as discussed in Proposition 1.

		Identity		**Diagonal**		**Optimal**	
Parameter	True Value	Level	Difference	Level	Difference	Level	Difference
σ^2_{perm}	0.005	0.008	0.014	0.008	0.014	0.012	0.012
σ^2_{tran}	0.038	0.029	0.018	0.029	0.018	0.020	0.020
τ	2.0 years	✗	✗	✗	✗	✗	✗
b	0.40	✗	✗	✗	✗	✗	✗
θ	✗	0.29	0.05	0.31	0.05	0.09	0.08
ρ	1.00	0.97	0.94	0.97	0.94	0.95	0.96
σ^2_{init}	0.065	0.065	✗	0.066	✗	0.067	✗

Notes: The table presents estimated parameters using the standard models on data generated for 200,000 individuals from the proposed model. 'True Value' refers to the parameters used in the data-generating process. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. The parameter θ is the moving-average coefficient, and ρ is the persistence of the permanent component.

Table A.8: Estimation of the standard model with persistent shocks using simulated data.

Table A.8 also presents the results when we simulate the proposed model and estimate the standard model. Compared with the data-generating process, the standard model, in this case, tends to underestimate the persistence of the permanent component. Furthermore, the standard model tends to overestimate permanent variance somewhat and underestimate the transitory variance, suggesting that some of the transitory shocks in the model are misrepresented as permanent shocks when estimated using the standard model. This pattern of misrepresented shocks was also present in the simulation exercise in the body of the paper.

	Identity		Diagonal		Optimal	
	Level	Difference	Level	Difference	Level	Difference
Panel A: Proposed Model						
σ^2_{perm}	0.003	0.004	0.003	0.005	0.003	0.004
σ^2_{tran}	0.044	0.046	0.043	0.045	0.044	0.040
τ	1.572	2.377	1.779	2.230	2.279	1.900
b	0.377	0.478	0.366	0.489	0.467	0.503
σ^2_{init}	0.063	✗	0.063	✗	0.060	✗
Panel B: Standard Model						
σ^2_{perm}	0.004	0.012	0.004	0.012	0.005	0.007
σ^2_{tran}	0.038	0.023	0.037	0.023	0.025	0.024
θ	0.480	0.055	0.571	0.056	0.157	0.138
σ^2_{init}	0.063	✗	0.063	✗	0.060	✗

Notes: The table presents estimated parameters using the proposed and standard models on Norwegian data. *Level* and *difference* denote the type of moments used, corresponding to (10) and (11), respectively. *Identity*, *diagonal*, and *optimal* denote the weighting matrix applied, corresponding to identity, diagonally optimal, and full optimal weighting matrix, respectively. The parameters τ and b are the half-life of the passing shock and the fraction of the transitory variance that consists of a "bonus" shock. The parameter θ is the moving-average coefficient. The table shows the mean of parameter estimates over time and age.

Table A.9: Estimated parameters using the Norwegian data. Sample including extreme-income-change outliers.

E.3 Estimation results without removing outliers

In the text, the sample from the Norwegian administrative data contains 536,399 Norwegian males from 27 cohorts born between 1937 and 1963. We arrive at this sample after dropping individuals that ever experience an extreme income change as described in Section 2. Table A.9 contains estimates of both the proposed and the standard model on a similar sample where we do not exclude individuals who experience an extreme income change. This sample contains 544,341 individuals, so excluding outliers in the main sample removes about 1.5 percent.

Comparing the results in Table A.9 with those presented in panels A and D in Table 4, we see that the results do not change much. The main difference is that when the extreme income changes are included in the sample, the estimates of the transitory variance are slightly higher. None of the conclusions regarding the choice of moments and weighting matrix in either the standard or the proposed model are affected.