# Estimating Life-Cycle Income Processes including Means-Tested Transfers

David Lander*

February 22, 2019

**Abstract**

This paper evaluates whether the 'persistent-transitory' model, which is commonly used to model household income, effectively measures government transfer policies. When this model is estimated using data comprised of both labour earnings and public transfers, it implicitly includes potentially nonlinear means-tested transfer policies typically designed to benefit low-income households. To assess how accurately this model approximates such policies, I undertake a Monte Carlo experiment with a simulated panel of income, where means-tested transfers are generated by an income floor. I show that means-tested transfers are not well represented by a stationary stochastic process for income with the canonical persistent-transitory structure, since means-testing generates age dependence in the autocovariance structure of income. The consequence, in terms of economic inference, is demonstrated by using the method of simulated moments to estimate key behavioural parameters (i.e., the discount factor and coefficient of risk aversion), with target moments taken from the model with an income floor. While the bias of these parameter estimates, relative to their 'true' values, is increasing in the income floor, the magnitude and direction of the bias depends on the choice of target moments.

**JEL codes:** C22, D15, D31, E21, H31.
**Keywords:** disposable income, income floor, consumption floor, persistent-transitory.

*Peking University HSBC Business School, Shenzhen, China 518055; email: lander@phbs.pku.edu.cn.

# 1 Introduction

An exogenous process is often used in dynamic economic models to represent the evolution of household income over the life cycle. Estimating such a process using household-level data involves making an important choice about which sources of income to include. Specifically, whether it is solely comprised of labour earnings, or also includes non-labour sources of income (e.g., public transfers). While the appropriateness of this choice depends on the economic question, the latter approach has significant advantages. Primarily, it avoids the alternative of explicitly modelling transfer (and taxation) policies and unemployment, thereby simplifying and hence reducing the dimensionality of the model. Consequently, it has been employed in many studies relating to household finance and income risk (e.g., Storesletten, Telmer, and Yaron (2001, 2004a,b, 2007), Cocco, Gomes, and Maenhout (2005), Blundell, Pistaferri, and Preston (2008), Guiso and Sodini (2013), and Cooper and Zhu (2016)).[1] Yet, given that many public transfer policies are means tested, since they are designed to benefit low-income households, it is unclear whether this approach adequately captures the effect of the policy—particularly potential non-linearities in the distribution of income introduced by means testing. While the aforementioned benefits of including transfers in the income process are compelling, the drawbacks are not so clear, beyond it being an approximation. Therefore, in this paper, I undertake a Monte Carlo experiment to better understand the trade-off associated with including means-tested transfers in the income process.

Regardless of which income definition is used, most models decompose log income into a deterministic and stochastic component, following the work of Lillard and Willis (1978), Lillard and Weiss (1979), MaCurdy (1982), and Abowd and Card (1989), and use a panel of income to estimate the parameters of these components. While many studies estimate a general ARMA model for the stochastic component, it is increasingly common to consider a 'persistent-transitory' model (e.g., Ejrnæs and Browning (2014)), i.e., an AR(1) with an iid normally-distributed shock. The former part captures 'persistent' variations in income, which remain over time, while the latter captures 'transitory' variations, which are relatively short-lived. This structure is attractive due to its computational simplicity—it is characterized by only three parameters—and its efficacy at modelling the distribution of labour earnings over the life cycle. This model can be used to directly represent the sum of labour earnings and transfers minus taxes without introducing any additional model complexity. Furthermore, it implicitly captures the dynamics of unemployment, since the estimated process is no longer conditional on having positive labour earnings. While this approach is relatively simple and computationally efficient, its overall efficacy will

[1] Note that some other studies include non-means-tested transfers in the exogenous income process (e.g., Hubbard, Skinner, and Zeldes (1995) and Heathcote, Storesletten, and Violante (2014)), but exclude means-tested transfers since they explicitly model state-contingent transfers.

now depend on how well a stationary autoregressive process captures the effect of means-tested transfers on the distribution of income over the life cycle.

There has been a significant amount of recent work studying the use of autoregressive processes estimated with labour-earnings data (e.g., Guvenen (2009), Hryshko (2012), and Karahan and Ozkan (2013)). However, relatively little attention has been directed toward analysing how these processes perform when estimated using data including public transfers, despite the fact that this is indeed a common occurrence. Including or excluding public transfers leads to fundamental differences in the distribution of income, particularly in the left tail, given the means-tested nature of many government transfer policies. Therefore, it is unclear to what extent this difference can be captured by such a parsimonious model. I seek to address this by demonstrating how means-tested transfers affect the ability of the estimation procedure to generate a distribution of income, over the life cycle, consistent with the underlying data. Therefore, I isolate the effect of means-tested transfers on the parameter estimates of the persistent-transitory model, and quantify how these estimates affect life-cycle behaviour of economic agents—relative to a model where the transfer policy is explicitly modelled.

To accomplish this, I undertake a Monte Carlo experiment with a 'baseline' model, which comprises a data-generating process (DGP) for labour earnings and a government transfer policy in the form of an income floor. This is used to simulate a panel of life-cycle earnings and transfers, which is used to estimate the parameters of a DGP for income (i.e., the 'alternative' model). The goal of this exercise is to evaluate how well the alternative model approximates the baseline model, using both statistical and economic criteria. I focus on several measures of bias, i.e., the difference between moments or parameter estimates of the alternative model relative to the baseline model. The first statistical measure involves comparing the first two moments of the cross-sectional distribution of income over the life cycle. This approach directly compares the extent to which the alternative model approximates the underlying DGP for income.

While this provides us with insight into the statistical accuracy of the alternative model, it does not demonstrate how substantially it may affect economic outcomes. Therefore, I employ a dynamic economic model of household consumption and savings, together with an income floor as the transfer policy, in order to obtain economic measures of model bias. The first economic measure is also statistical, since it compares the first two moments of the cross-sectional distribution of consumption and savings over the life cycle. The second measure, however, evaluates how the model affects inference by a researcher. It entails using the method of simulated moments to estimate the behavioural parameters of the alternative model (i.e., the discount factor and coefficient of risk aversion), with target moments taken from the simulated panels of consumption and savings in the baseline model. Therefore, this criteria provides us with a measure of bias in terms of the parameter estimates governing the behaviour of households.

The results of the Monte Carlo experiment reveal that incorporating transfers into the earnings process may result in severe bias for the parameters governing future discounting ($\beta$) and risk aversion ($\gamma$), and the direction and magnitude of this bias depends on the choice of target moments in the estimation. The bias in these estimates directly follows from the bias in the moments of income over the life cycle: the bias in mean income is relatively small over the life cycle, while the bias in life-cycle income variance is substantial. Specifically, the alternative model underestimates the variance of income for most ages over the life cycle, and hence consequently underestimates household savings. Therefore, when mean savings over the life cycle are used as target moments in the method of simulated moments estimation, the model overestimates $\gamma$ to ensure there is sufficient savings to match the target moments; i.e., the households need to be more risk averse to generate the additional savings. If mean-savings moments are used to only estimate $\beta$, by the same logic, this will lead to $\beta$ being overestimated; i.e., the households need to be more patient to generate additional savings. On the other hand, when only mean consumption over the life cycle is used as a target moment, the bias is relatively small. This occurs because mean consumption closely follows mean income over the life cycle, and the alternative model closely approximates mean income.

The paper proceeds by defining the baseline model for earnings and transfers in Section 2. The simulated income data from the baseline model is used in Section 3 to estimate the parameters of the alternative model of income. This section also studies the impact of the transfer policy on the parameter estimates, relative to the parameters of the DGP for labour earnings. Section 4 introduces the dynamic model of household consumption and savings to quantify, in economic terms, the bias induced by using the alternative model of income. Section 5 offers concluding remarks.

## 2 Baseline model

This section outlines the baseline model of the Monte Carlo experiment, which includes a data-generating process (DGP) for labour earnings and a government transfer policy. Together, these form the DGP for income, which is used to generate a simulated panel of income, i.e., the sum of labour earnings and transfers, net of taxes. The role of unemployment and the tax policy used to finance the transfers is also discussed in this section.

### 2.1 Earnings

The DGP for labour earnings contains both deterministic and stochastic components, where the specification of the shocks follows the canonical 'persistent-transitory' structure. For a given household $i$, pre-retirement log earnings, $\ln y_{i,j}$, is comprised of a common deterministic component, $m_j$, and household-specific persistent and transitory components, $z_{i,j}$ and $u_{i,j}$, respectively, all of which are indexed by age $j$:

$$\ln y_{i,j} = m_j + z_{i,j} + u_{i,j}, \tag{1}$$

where $z_{i,j}$ follows an AR(1) process, $z_{i,j} = \rho z_{i,j-1} + \eta_{i,j}$, with $\rho \leq 1$, $\eta_{i,j} \sim$ iid $\mathcal{N}(0, \sigma_\eta^2)$, $u_{i,j} \sim$ iid $\mathcal{N}(0, \sigma_u^2)$, $\eta_{i,j} \perp\!\!\!\perp u_{i,j}$. The deterministic component, $m_j$, and the parameters of the stochastic process, $\theta = \{\rho, \sigma_\eta, \sigma_u\}$, are estimated using U.S. household earnings data from the Panel Study of Income Dynamics (PSID), from 1970 to 2012.[2] The estimate of the deterministic component is plotted in the left panel of Figure 1 (the relevant series without any transfers is labelled '0%'), while the point estimates of $\theta$ are in Table 1.[3]

Table 1: Parameter estimates for the stochastic component of labour earnings.

| **Parameter** | **Value** | **Description** |
|---|---|---|
| $\rho$ | 0.9735<br>(0.0033) | Persistence of labour earnings |
| $\sigma_\eta$ | 0.1272<br>(0.0082) | Standard deviation of persistent labour-earnings shock |
| $\sigma_u$ | 0.2102<br>(0.0075) | Standard deviation of transitory labour-earnings shock |

**Note:** Parameter estimates obtained using PSID residual household earnings data, following Guvenen (2009). Standard errors are presented in parentheses.

The estimates in Table 1 are broadly in the range of those from previous studies: estimates of $\sigma_\eta$ include 0.14–0.16 (Hubbard, Skinner, and Zeldes (1995)), 0.15–0.17 (Gourinchas and Parker (2002)), and 0.12–0.21 (Storesletten, Telmer, and Yaron (2004b)); while estimates of $\sigma_u$ include 0.17 (Hubbard, Skinner, and Zeldes (1995)), 0.21 (Gourinchas and Parker (2002)), and 0.235–0.257 (Storesletten, Telmer, and Yaron (2004b)).

The structure of the DGP for labour earnings, with normal transitory and persistent innovations, imposes zero skewness and (excess) kurtosis in the distribution of log earnings. While recent work, e.g., Guvenen, Ozkan, and Song (2014), departs from the persistent-transitory stochastic structure with the goal of modelling higher-order moments, I maintain this structure for several reasons. First, this approach is still commonly employed in studies where higher-order moments are not the primary focus. Additionally, given we are evaluating a technique

[2]The deterministic component is comprised of age dummies and year fixed-effects. Household earnings is defined as the sum of head and wife labour earnings, while age refers to that of the household head. Additionally, the SEO sample is excluded to ensure the sample is representative of the population.

[3]The parameter estimates for the stochastic component of labour earnings in Table 1 will be referred to as $\theta$, rather than $\hat{\theta}$, since the latter notation will be used to refer to the corresponding parameter estimates for a DGP for income.

which assumes zero skewness and kurtosis, introducing these features into the DGP for labour earnings complicates the experiment since the technique cannot possibly match them. Finally, eliminating skewness and kurtosis in the DGP for log earnings allows us to clearly demonstrate one of the effects of the policy: it introduces skewness and kurtosis in the distribution of income.

An important difference between estimating a process for earnings vs. income is the treatment of zero-earnings observations. It is standard when estimating a process for earnings to condition on having positive earnings, i.e., discarding observations with zero earnings. When following an analogous procedure for income, however, we discard zero-income observations, i.e., zero earnings and zero transfers. Therefore, the latter procedure potentially discards less observations, since it includes zero-earnings observations that received transfers. Therefore, in principle, we should account for this sampling difference by specifying a data-generating process for zero (annual) labour earnings. Henceforth, I refer to this event as a household being 'unemployed', while acknowledging that when we observe zero (annual) earnings in the data, it may not necessarily indicate that an individual in the household is seeking work.

I proceed by demonstrating the effect of the income floor by continuing to exclude these unemployed households. The primary reason is that we can infer the direction of the effect (of introducing unemployment), and therefore can think of the case without unemployment as a lower-bound. Furthermore, since (in the following section) I specify the floor at some arbitrary quantile of the distribution of income, we can generate results without zeros equivalent to a mass of zeros (uniformly distributed across age) by simply increasing the level of the floor. To understand this, suppose in addition to our sample of households with positive labour earnings, we specify a mass of unemployed households with zero earnings. We know that every household with zero earnings will obtain a transfer due to the income floor policy. Hence, as we increase the mass of zeros, relative to non-zeros, we are increasing the fraction of households receiving a transfer. This is demonstrated in Appendix A, which contains results obtained with a DGP for unemployment. Additionally, if we allow the likelihood of unemployment to vary by age (which generally increases in age, especially as retirement nears), the numerical results demonstrated later in this section become amplified.

## 2.2 Transfers

The means-tested government transfer policy takes the form of an income floor, $\underline{y} \geq 0$. These transfers are financed by a flat tax, $\tau \geq 0$, which the government sets to balances its budget, in expectation, across the life-cycle, i.e., it does not balance at each time period, therefore $\tau$ is a constant.[4] With this policy, income for household $i$, in levels, is given by,

$$\tilde{y}_i = \max(y_i, \underline{y}) - \tau. \tag{2}$$

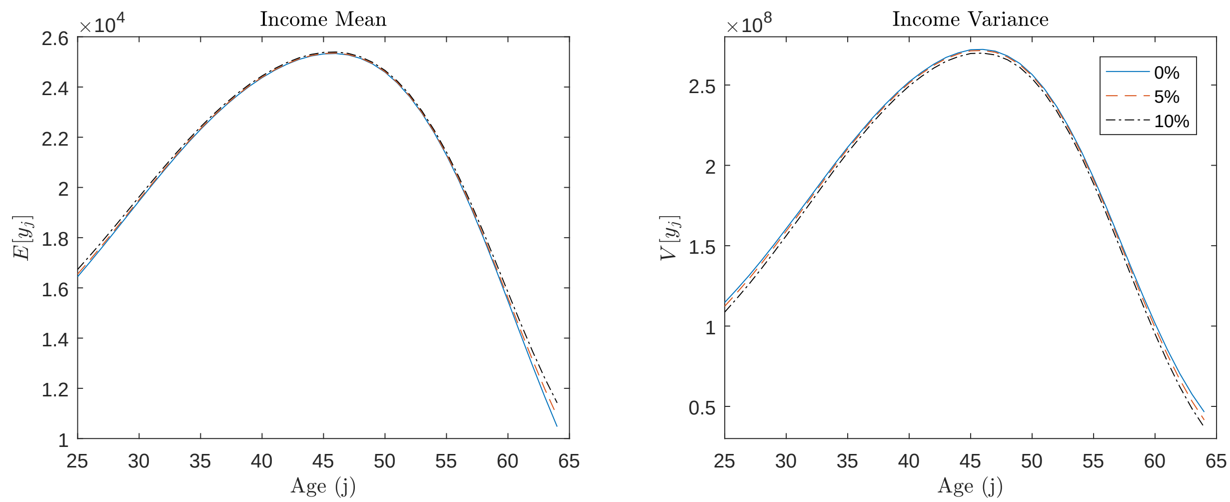
An income floor is a simple representation of the types of policies enacted by governments with the general goal of mitigating poverty by insuring individuals against spells of low or zero earnings. Although it is a stark representation of government transfers, it is commonly used in economic models to model a 'social safety net'. Furthermore, this structure captures the key essence of means-tested transfer policies, and is sufficient to understand the general consequences of estimating a process for income using data on labour earnings and means-tested government transfers. The tax specification in (2) is chosen so that it does not affect the estimation of $\theta = \{\rho, \sigma_\eta, \sigma_u\}$, which allows us to isolate the effect of the income floor on these parameters. This occurs because the tax only affects the mean of the cross-sectional distribution of income, at every age in the life cycle, and hence only affects the estimation of the deterministic component of income. The income floor is set, in levels, at a value corresponding to some positive quantile, (e.g., at a 5 percent level), which is calculated from the simulated labour-earnings data, pooled across ages.

Denote the simulated panel of labour earnings $Y_\theta = \{y_{i,j}\}_{i=1,\ldots,N}^{j=25,\ldots,64}$, which comprises $N = 500{,}000$ households working for $T = 40$ years, i.e., for age $j = 25, \ldots, 64$. The simulated panel of income, resulting from the policy defined in (2), is denoted $\tilde{Y}_\theta = \{\tilde{y}_{i,j}\}_{i=1,\ldots,N}^{j=25,\ldots,64}$. The moments from these two panels will be compared to understand the effect of the transfer policy on the distribution of income over the life cycle. The first two moments of the income distribution, in levels, are contained in Figure 1. Note that these figures exclude taxes (i.e., $\tau = 0$), in order to isolate the life-cycle effects of the transfer. As noted above, this tax policy is chosen deliberately to not affect the estimation of the stochastic parameters: it will equally shift mean earnings down at every age and hence be captured in the parameter estimates of the deterministic component.

For a given income floor, it is clear that any increase in the level of the floor causes a monotonic shift up in the mean profile (excluding taxes), and a downward shift in the variance profile. Note, however, that the policy is disproportionately benefiting households early and late in life. We observe a relatively large increase in mean income (left panel of Figure 1), and a relatively large decrease in income variance (right panel of Figure 1). Therefore, the policy is introducing life-cycle effects in the distribution of income. Specifically, it is differentially affecting the cross-sectional distribution of income, by age. This reflects an important feature of the policy: although the policy definition is independent of age, the transfers are disproportionately being received by the relatively young and old due to the hump-shape in the mean profile of earnings.

[4]Specifically, $\tau = \int_0^{\underline{y}} (\underline{y} - y) dG(y)$, where $G(\cdot)$ is the distribution of labour earnings.

Figure 1: Effect of the income floor on life-cycle income moments.




This figure demonstrates the effect of an income floor on the mean and variance of life-cycle income, with the income floor set at the 5 and 10% quantiles of labour earnings. The 0% floor corresponds to mean and variance of labour earnings.

The direction of the effect of the income floor on the first two moments of income, i.e., increasing the mean of income and decreasing the variance of income, both relative to labour earnings, is demonstrated in Appendix B for a general distribution of labour earnings. The effect on the mean profile of earnings is not particularly important for the purposes of this exercise. However, the life-cycle effects on the variance will impact the parameter estimates of $\theta$. As mentioned above, the transfer policy also introduces skewness and kurtosis into the distribution of income over the life cycle, even though these are absent in the labour-earnings process, by construction. Since I will use the same procedure to estimate the parameters of the stochastic process for income, it will not match these higher-order moments. Hence, I will only use the first two moments to evaluate the efficacy of this procedure.

## 3 Alternative model

It is useful to briefly review the standard approach to estimating a life-cycle process for labour earnings before describing the alternative approach. First, labour earnings are regressed on demographic characteristics to obtain parameter estimates that can be used to specify a deterministic life-cycle labour-earnings profile. Second, the residuals from this regression (i.e., residual earnings) are used to estimate the parameters of a stochastic process for earnings. This approach invariably estimates an earnings equation in logs, which necessitates dropping zero-earnings observations. Therefore, it provides us with a life-cycle process for log labour earnings, conditional on having positive earnings.

The alternative approach, however, deviates from this procedure by estimating a process for income, i.e., the sum of labour earnings and transfers. This avoids the need for explicitly specifying a transfer policy in an economic model, and dropping zero-earnings observations. The justification for including observations with zero earnings follows from the goal of studying transfers to the household. Since this procedure is in terms of income, and not labour earnings, zero-income observations are dropped. However, with the income floor specified in (2), there will be no households with zero income, hence no observations are dropped in this procedure.

To estimate a DGP for income, rather than labour earnings, I will follow the same procedure described in Section 2.1, except (1) is redefined to be in terms of log income, $\ln \tilde{y}$, rather than log earnings, $\ln y$. As before, we estimate the parameters of the deterministic component by regressing log income on observable characteristics, which is now solely comprised of age dummies. That is, we pool the panel $\tilde{Y}_\theta$ into a cross-section of $N \times T$ observations and estimate $\ln \tilde{y}_i = \alpha D_i + \epsilon_i$, where $D_i$ is a vector of dummies for each age $j$. The estimate $\hat{\alpha}$ is used to generate the deterministic component of income, which is also given by $\tilde{m}_j = \frac{1}{N}\sum_{i=1}^{N} \ln \tilde{y}_{i,j}$. The residuals from this regression are used to estimate the parameters of the stochastic component, $\hat{\theta} = (\hat{\rho}, \hat{\sigma_\eta}, \hat{\sigma_u})$, which is identified from the covariance matrix of the residuals. I employ an equally weighted minimum distance estimator, i.e., $\hat{\theta}$ minimizes the distance between the target moments (generated from $\tilde{Y}_\theta$) and the model moments (generated from $\tilde{Y}_{\hat{\theta}}$) from the covariance matrix, with an identity weighting matrix. The estimate of $\hat{\theta}$ obtained with an income floor set at the 5 and 10% quantiles are contained in Table 2.

Table 2: Parameter estimates for the stochastic component of income.

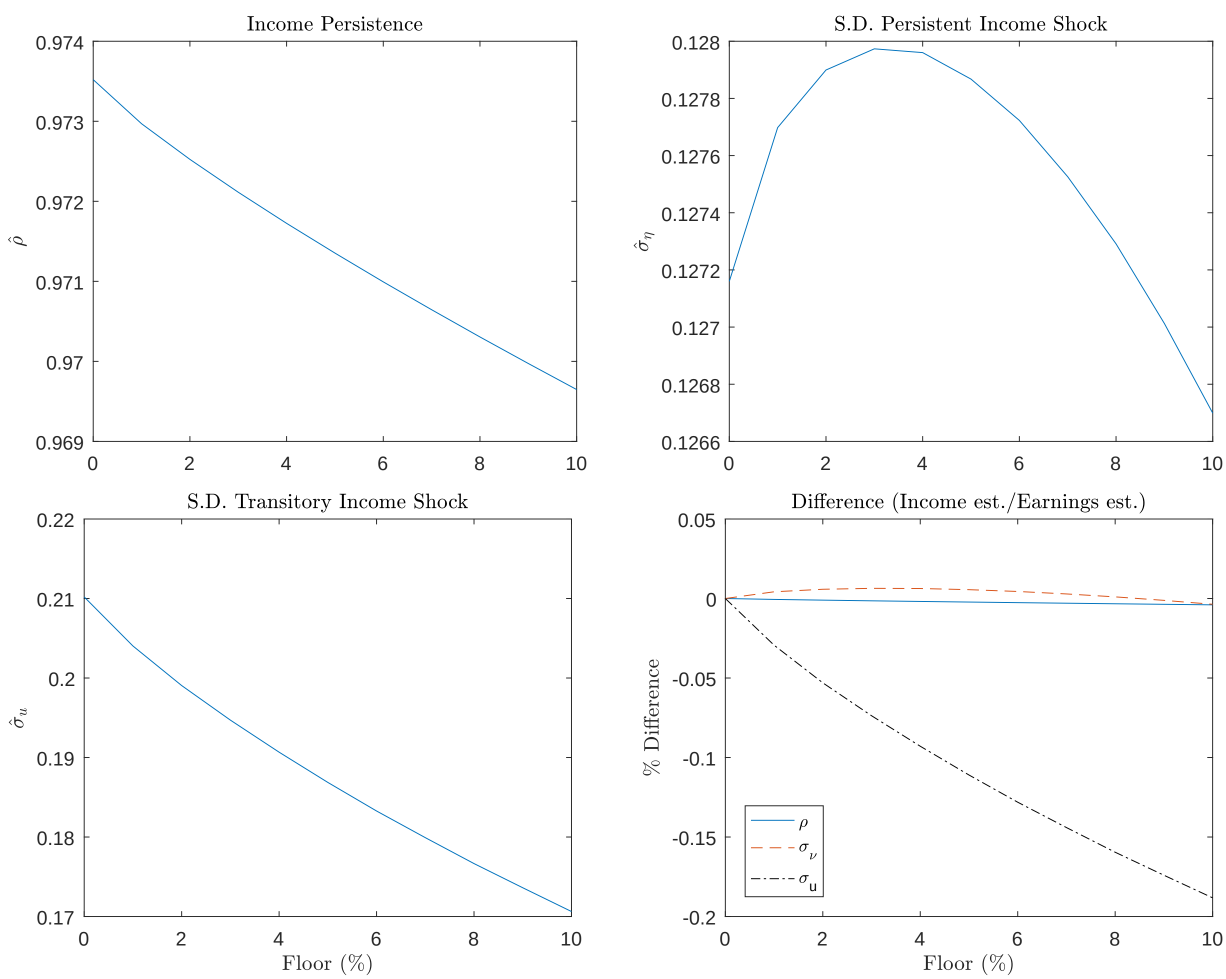
| **Parameter** | **0% floor** | **5% floor** | **10% floor** | **Description** |
|---|---|---|---|---|
| $\hat{\rho}$ | 0.9735 | 0.9714 | 0.9697 | Persistence of income |
| | (0.00005) | (0.00005) | (0.00006) | |
| $\hat{\sigma}_\eta$ | 0.1272 | 0.1278 | 0.1266 | Std. deviation of persistent income shock |
| | (0.00012) | (0.00011) | (0.00011) | |
| $\hat{\sigma}_u$ | 0.2102 | 0.1869 | 0.1707 | Std. deviation of transitory income shock |
| | (0.00019) | (0.00018) | (0.00020) | |

**Note:** These parameter estimates of $\hat{\theta}$ were obtained with an income floor set at the 5 and 10% quantiles of the simulated panel of labour earnings; while the 0% floor values are the corresponding estimates of $\theta$, from Table 1.

The results in Table 2 reveal that $\hat{\rho}$ and $\hat{\sigma}_u$ are monotonically decreasing in the value of the income floor, while $\hat{\sigma}_\eta$ is nonmonotonic. Specifically, the standard deviation of the persistent labour earnings shock, $\sigma_\eta$, is less than the estimate of $\hat{\sigma}_\eta$ with a 5% income floor, but greater than the estimate of $\hat{\sigma}_\eta$ with a 10% income floor. These relationships are more clearly depicted in Figure 2, which plots the parameter estimates for $\hat{\theta}$ against different levels of the income floor. As the level of the floor increases, $\hat{\rho}$ and the standard deviation of the transitory shock, $\hat{\sigma}_u$, unambiguously decrease. However, the effect on $\hat{\sigma}_\eta$ is ambiguous: from Figure 2, it is increasing at low levels of the floor, but decreasing at higher levels.

At first glance, a higher value of $\hat{\sigma}_u$ appears to be inconsistent with the empirical estimates in Blundell, Pistaferri, and Preston (2008), where the estimated variance of both the *permanent* and transitory shocks are much lower when taxes and transfers are included—a result consistent with our intuition that transfers are providing insurance and hence reducing the variance of income shocks. However, this discrepancy appears to be driven by the fact their income process follows a 'permanent-transitory' model, i.e., restricting $\hat{\rho} = 1$. I show in Section 3.1 that at least one of $\hat{\rho}$ and $\hat{\sigma}_u$ must be decreasing, hence if I impose this same restriction here, an unambiguous reduction in $\hat{\sigma}_u$ will result. Interestingly, this also implies that, for a given variance of a permanent income shock, transfers may provide insurance to the household by reducing the permanence of the shock.

Figure 2: Effect of the income floor on parameter estimates.




It is important to note that, as the value of the income floor increases, the parameter estimate of the transitory shock ($\hat{\sigma}_u$) appears to absorb the majority of the changes introduced by the transfer policy. This is evident from the estimates in Table 2, and it is clearly depicted in the bottom-right panel of Figure 2, which plots the percentage difference between the parameter estimates and their labour-earnings counterparts. For example, the percentage difference between the persistence of income, $\hat{\rho}$, and the persistence of labour earnings, $\rho$, is given by $\frac{\hat{\rho}}{\rho} - 1$ in this figure. This figure indicates that, for the relatively large 10% income floor, $\hat{\rho}$ and $\hat{\sigma}_\eta$ only vary approximately 1 percent from their labour-earnings counterparts, whereas $\hat{\sigma}_u$ decreases by almost 20 percent.

## 3.1 Identification

To can gain some further understanding into how the income floor affects $\hat{\theta}$, I will briefly discuss how the parameter estimates are identified. Figure 1 indicates that the variance of income is conditional on age, and hence nonstationary.[5] Therefore, we know that this procedure, which assumes stationarity of the underlying DGP will fail to capture life-cycle variations in the variance of income. Additionally, since identification of $\hat{\theta}$ comes from the autocovariance matrix of income, the point estimate of $\hat{\theta}$ will now vary depending on the range of ages in the sample. Maintaining the assumption that the underlying DGP is stationary, we can use the identification coming from the autocovariance structure of income to evaluate how the parameter estimates for $\hat{\theta}$ varies over the life cycle. With the serial structure of income, from Section 3, and $\theta = \{\rho, \sigma_\eta, \sigma_u\}$, let the cross-sectional moment between agents of age $j$ and $n$ be denoted $m_{j,n}(\theta) = \mathbb{E}[y_{i,j} \cdot y_{i,j+n}] = \mathbb{E}[(\mu_j + z_{i,j} + u_{i,j}) \cdot (\mu_{j+n} + z_{i,j+n} + u_{i,j+n})]$. Therefore, we obtain:

$$m_{0,n}(\theta) = \rho^n \tfrac{\sigma_\eta^2}{1-\rho^2} \tag{3}$$

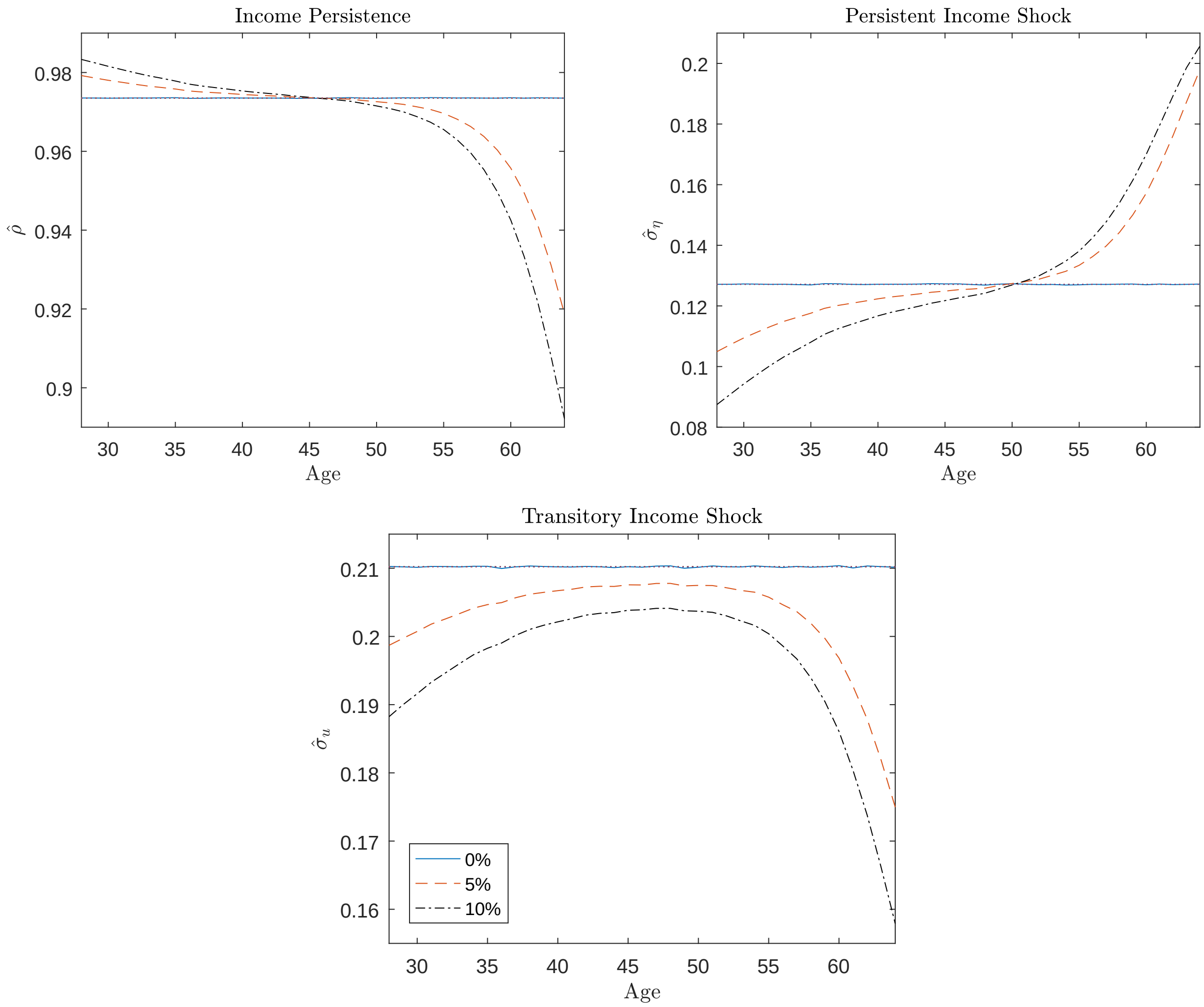
$$m_{0,0}(\theta) = \tfrac{\sigma_\eta^2}{1-\rho^2} + \sigma_u^2 \tag{4}$$

where $\sigma_u^2$ only appears in $m_{0,0}(\theta)$, since it is iid. The parameter $\rho$ is identified using the slope of the cross-sectional moment function, (3), $\frac{m_{0,3}(\theta)-m_{0,2}(\theta)}{m_{0,2}(\theta)-m_{0,1}(\theta)} = \rho$, while the identification of $\sigma_\eta$ and $\sigma_u$ follows from (3) and (4), respectively. We can use plot the empirical cross-sectional moments that identify each parameter in $\theta$ over the life cycle, in order to demonstrate any potential nonstationarity. This is depicted in Figure 3.

Figure 3 provides us with some insight into the functions depicted in Figure 2. Specifically, regarding the direction the parameter estimates change as the floor increases. First, it should be noted that when there is no income floor, i.e., 0% floor, each of the estimated parameters is a constant, since the covariance of labour earnings is stationary over the life cycle. At all ages

[5] An analytical demonstration of this can be found in Appendix B.

Figure 3: Life-cycle effects on parameter estimates.




This figure demonstrates how the empirical cross-sectional moments vary over the life cycle for each parameter in $\theta$. Note that identification depends on up to 3 lags, hence this is for $j = 28, \ldots, 64$.

we see that $\hat{\sigma}_u$ decreases, hence the value of $\hat{\sigma}_u$ that minimizes the distance between the target moments and simulated moments will be lower than $\sigma_u$. However, $\hat{\rho}$ and $\hat{\sigma}_\eta$ can potentially increase or decrease relative to $\rho$ and $\sigma_\eta$, respectively. Although, unlike $\hat{\sigma}_u$, they may increase, but they cannot both increase: Figure 3 demonstrates that $\hat{\rho}$ and $\hat{\sigma}_\eta$ are moving in opposite directions, relative to $\rho$ and $\sigma_\eta$; e.g., if $\hat{\rho} > \rho$ then $\hat{\sigma}_\eta < \sigma_\eta$. This follows from the fact that the autocovariance with 1 lag, for example, is unambiguously decreasing for $j = 26, \ldots, 64$. Since this corresponds to (3) with $j = 1$, i.e., $\rho \frac{\sigma_\eta^2}{1-\rho^2}$, it must be that one of $\rho$ or $\sigma_\eta$ is decreasing. This is consistent with Figure 3, which shows that, for all ages, at least one of $\hat{\rho}$ and $\hat{\sigma}_\eta$ is decreasing relative to $\rho$ and $\sigma_\eta$. Additionally, the magnitude of each parameter estimate is increasing (or decreasing) monotonically as the level of the floor increases.

The figure also demonstrates the consequences of assuming a stationary structure for these shocks: $\hat{\sigma}_u$ is overestimated early and late in life, and underestimated in mid life; $\hat{\sigma}_\eta$ is overestimated early in life, and underestimated late in life; $\hat{\rho}$ is underestimated early in life, and overestimated late in life. This reveals how a stationary process fails to capture impact of the transfer policy on income variance over the life cycle. As an aside, the distortions introduced by means-tested transfers provide a motivation for possibly employing age-dependent shocks when using data on labour earnings together with transfers. This is an interesting implication which is separate from, but consistent with, recent studies suggesting that we dispense with the stationarity assumption when estimating processes for labour earnings; see, for example, Karahan and Ozkan (2013).

Nevertheless, the key result of this section is that a constant income floor together with an age-dependent profile for mean earnings over the life cycle will lead to age dependence in the variance of income over the life cycle. Furthermore, the shape of the life-cycle income variance will follow that of life-cycle mean earnings.[6] Therefore, if mean earnings is hump-shaped over age, then life-cycle income variance will also be hump-shaped.

This effect is not specific to the stark example of an income floor—it will also arise with other means-tested transfer policies when the proportion of eligible transfer recipients varies by ages (where eligibility could depend on income, wealth or consumption, for example), even if the level of the transfer is constant across all ages. For example, in the case of a consumption floor, an age-dependent profile for consumption over the life cycle will lead to transfers across age groups, truncating the cross-sectional distribution of income differentially across ages.

Finally, it is important to recognise that it is not particularly meaningful to directly compare $\hat{\theta}$ (DGP for income) with $\theta$ (DGP for earnings), since the distance between these estimates is not a measure of bias in this context. Instead, we should be comparing moments coming from $\hat{\theta}$ with the baseline panel of income. The moments of income alone, however, are not particularly helpful for evaluating the economic consequences of failing to capture the nonstationarity in life-cycle income induced by the transfer policy. Therefore, to evaluate the economic consequences of assuming a stationary structure of these shocks, I study the effect of using these DGPs as inputs into a life-cycle heterogeneous-agents model with an income floor. This model can be used to define an economic measure of bias, which clearly quantifies the difference between the baseline and alternative models of income.

[6] See Appendix B for further details.

# 4 Economic analysis

To provide us with an economic interpretation for the consequences of failing to capture the underlying nonstationarity of the income process, I evaluate the alternative income process using a standard dynamic economic model of household consumption and savings. This life-cycle model, with discounting and CRRA preferences, is parameterised by $\beta$ and $\gamma$, respectively. These parameters are used to evaluate how the income estimation techniques affect the decisions of economic agents. The logic of the experiment is as follows: the baseline model, parameterised by $(\beta_0, \gamma_0)$ and characterized by exogenous labour earnings and the income floor (from Section 2), is used to generate a series of consumption and savings over the life cycle. These series are used as target moments in the method of simulated moments estimation procedure. This procedure estimates the pair $(\hat{\beta}, \hat{\gamma})$, which minimizes the distance between the moments of the alternative model, characterized by exogenous income (from Section 3), and the target moments from the baseline model.

The baseline model captures the effect of the policy on the consumption decisions of agents, given the exogenous labour earnings process and parameters, while the alternative model is an approximation. If the estimation process in Section 3 accurately measures the baseline DGP of income, these two models should be equivalent. However, we know that they will differ, since the procedure cannot account for the underlying nonstationarity of income. Therefore, the goal of this exercise is to quantify this difference in economic terms. To this end, I make use of the bias in the estimated parameters, defined as $\left(\frac{\hat{\beta}}{\beta_0}, \frac{\hat{\gamma}}{\gamma_0}\right)$, which reveals how using this approach will lead to incorrect inference regarding household behaviour.

## 4.1 Model

Agents live for $T$ periods, and work for $T_r < T$ periods. Following Section 2, we set $T_r = 40$ to correspond with agents working from age 25 to 64, and set $T = 60$, implying a retirement duration of $T - T_r = 20$ periods. The agent maximizes the discounted present value of expected lifetime utility, given by

$$E_0 \sum_{t=1}^{T} \beta^{t-1} u(c_t), \tag{5}$$

subject to the period-$t$ budget constraint,

$$\begin{aligned} c_t &= \tilde{y}_t + s_t - R s_{t+1}, \qquad (6) \\ s_t &\geq 0 \text{ for } t = 1, \ldots, T. \end{aligned}$$

Consumption is given by $c_t$, savings is denoted $s_t$, which accumulates at the gross interest rate $R > 1$, and income is defined in (2), i.e., $\tilde{y}_t = \max(y_t, \underline{y}) - \tau$. The initial and terminal conditions are $s_0 = 0$ and $s_T \geq 0$, respectively. The Bellman equation corresponding to (5) and (6) is

$$V(y, s, j) = \max_{s' \geq 0} u(c) + \beta E_{y'|y} V(y', s', j+1), \tag{7}$$

where the age-$j$ budget constraint is given by $c = \tilde{y}(j) + s - Rs'$.

This setup encompasses the two alternative approaches: the baseline model involves jointly specifying a Markov process for $y_t$ with a transfer policy $\underline{y}$, while the alternative model directly specifies a Markov process for $\tilde{y}_t$. It is important to note that this difference only applies pre-retirement, since this is the relevant range of ages considered in the estimation in Section 2. The post-retirement period is characterised by no income uncertainty. However, in order to make sure the post-retirement period across models is consistent, we have to take a stance on whether the income floor takes effect during retirement for both approaches. I dispense with the income floor, and set post-retirement income is equal to 57.9 percent of the deterministic and persistent component of earnings in the final year of work (i.e., it excludes the transitory component).[7]

The baseline model is defined in terms of labour earnings, $y$, and a transfer and tax policy pair, $(\underline{y}, \tau)$. The income floor, $\underline{y}$, is set at the 5 percent quantile of labour earnings, and the tax levied to finance this floor, $\tau$, is calculated so that the government balances its budget exactly, i.e., the tax is conditional on the pooled sample of $N \times T$ households.[8] I set the preference parameters $\beta_0 = 0.95$ and $\gamma_0 = 2.0$, and the exogenous return on savings $R = 1.0344$, following Gourinchas and Parker (2002). Earnings and income, $y$ and $\tilde{y}$, are both governed by AR(1) processes, each with a transitory component. Hence these represent vectors of state variables $\mathbf{y} = (y_z, y_u)$ and $\tilde{\mathbf{y}} = (\tilde{y}_z, \tilde{y}_u)$, where the persistent AR(1) and the iid transitory component are denoted by the indices $z$ and $u$, respectively, consistent with (1). The distribution of each of the shocks is discretised, following Rouwenhorst (1995), into 10 states. I use the baseline model to obtain policy rules and simulate a panel of consumption and savings for $N = 500,000$ households. This dataset is used to generate a series of moments that are used as 'target' moments in a method of simulated moments estimation, which is undertaken in the following section. The alternative model has the same parameterisation as the baseline model, except the income process, $\tilde{\mathbf{y}}$, is directly specified using the parameter estimates from Table 2 (for the 5% income floor), i.e., $(\hat{\rho}, \hat{\sigma}_\eta, \hat{\sigma}_u) = (0.9714, 0.1278, 0.1869)$.

[7] This fraction is calibrated to ensure the mean post-retirement income matches the average sum of income and public transfers in the PSID for households with a household head at least 65 years old.

[8] That is, $\tau = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=25}^{64} \max(\underline{y} - y_{i,j}, 0)$. This results in $\underline{y}$ and $\tau$ of \$5961 and \$657, respectively.

### 4.2 Estimation and analysis

In this section, the preference parameters, $\Theta = (\beta, \gamma)$, are estimated using the method of simulated moments, where the target moments come from the simulated panel of consumption and assets from the baseline model. This will inform us of whether including transfers implicitly in the income process is a good approximation to explicitly defining an income floor in the model. Furthermore, the difference between the parameter sets, $\Theta_0 = (\beta_0, \gamma_0)$ and $\hat{\Theta} = (\hat{\beta}, \hat{\gamma})$, can be used to measure bias with an economic interpretation. This is a useful exercise for several reasons. First, as discussed in Section 3.1, we cannot measure bias in terms of the difference between the parameters of the DGP for labour earnings and the parameter estimates of the DGP for income. Additionally, our main concern is evaluating the economic implications of using this approach, and this procedure allows us to clearly quantify how this procedure may lead to incorrect inference.

I use a minimum distance estimator that minimises the distance between the target and model moments, $\hat{\Theta} = \arg\min_{\Theta}[M - \hat{M}(\Theta)]'W[M - \hat{M}(\Theta)]$, where $M$ represents the target moments, $\hat{M}(\Theta)$ represents the model moments, and $W$ is the weighting matrix, which is set as the identity matrix. The parameter estimates, obtained using several combinations of parameter estimates and target moments, are given in Table 3.

Table 3 contains three sets of parameter estimates, (1) $\beta$ only (with $\gamma = \gamma_0$), (2) $\gamma$ only (with $\beta = \beta_0$), and (3) $(\beta, \gamma)$, which are estimated using target moments based on either consumption ($\mathbb{E}[C]$), savings ($\mathbb{E}[S]$), or both consumption and savings. The target moment $\mathbb{E}[C]$ is a vector containing $T = 40$ (pre-retirement) moments from the panel of consumption, defined as $\{\mathbb{E}[c_j]\}_{j=25,\ldots,64}$, where $\mathbb{E}[c_j] = \frac{1}{N}\sum_{i=1}^{N} c_{i,j}$. The same construction applies to the target moment for savings, $\mathbb{E}[S]$. When the target is comprised of both $\mathbb{E}[C]$ and $\mathbb{E}[S]$, these respective vectors are concatenated to yield $2T$ moments.[9]

The estimates in Table 3 demonstrate that the alternative model does a reasonable job approximating $\beta$ and $\gamma$ when the target moments are mean consumption. In every case targeting only $\mathbb{E}[C]$, the parameter estimates, $\hat{\beta}$ and/or $\hat{\gamma}$, have less than 1 percent bias. However, when the target moments include savings, significantly more bias is introduced, except in the case where only $\beta$ is estimated. This is because when $\gamma$ is fixed, the model cannot adequately adjust $\beta$ to provide a good fit with the target moment for savings, hence the model fit is very poor. Allowing $\gamma$ to adjust significantly improves the model fit, and it does so by increasing $\hat{\gamma} > \gamma_0$. In the case where only $\gamma$ is estimated, and $\beta = \beta_0$, we obtain a bias of around 3 percent. When both $(\hat{\beta}, \hat{\gamma})$ are estimated, the bias remains low for $\beta$ (approximately 1 percent), but the bias for $\gamma$

[9] Results obtained using a weighting matrix modified so that the optimization minimises the sum of percentage differences between the simulated and target moments produced parameter estimates with significantly more bias, since the weighting matrix places relatively more weight on the savings moments.

Table 3: Parameters estimates and bias.

| | **Estimates** | | **Bias** | |
|---|---|---|---|---|
| Target Moments | $\hat{\beta}$ | $\hat{\gamma}$ | $\frac{\hat{\beta}}{\beta_0}$ | $\frac{\hat{\gamma}}{\gamma_0}$ |
| $\mathbb{E}[C]$ | 0.950 | – | 1.000 | – |
| | (0.0001) | | | |
| $\mathbb{E}[S]$ | 0.953 | – | 1.003 | – |
| | (0.0001) | | | |
| $\mathbb{E}[C]$ & $\mathbb{E}[S]$ | 0.953 | – | 1.003 | – |
| | (0.0001) | | | |
| $\mathbb{E}[C]$ | – | 1.99 | – | 0.997 |
| | | (0.0023) | | |
| $\mathbb{E}[S]$ | – | 2.07 | – | 1.034 |
| | | (0.0017) | | |
| $\mathbb{E}[C]$ & $\mathbb{E}[S]$ | – | 2.07 | – | 1.033 |
| | | (0.0019) | | |
| $\mathbb{E}[C]$ | 0.950 | 1.99 | 1.000 | 0.997 |
| | (0.0003) | (0.0002) | | |
| $\mathbb{E}[S]$ | 0.942 | 2.26 | 0.991 | 1.132 |
| | (0.0026) | (0.0025) | | |
| $\mathbb{E}[C]$ & $\mathbb{E}[S]$ | 0.940 | 2.30 | 0.990 | 1.152 |
| | (0.0016) | (0.0016) | | |

**Note:** Income floor set at the 5 percent quantile of pooled life-cycle earnings. Estimates obtained with $N = 500,000$ simulated households. The target moments $\mathbb{E}[C]$, $\mathbb{E}[S]$, and $\mathbb{E}[C]$ & $\mathbb{E}[S]$, refer to the mean (pre-retirement) moments for consumption, savings, and both consumption and savings, respectively. Standard errors are presented in parentheses.

ranges from 13 to 15 percent, depending on whether only savings or both consumption and savings are targeted. These results have two important implications. First, if a researcher uses the alternative approach and only targets mean consumption, the bias will be small, but the model fit is relatively poor. Second, if the researcher targets savings (with or without consumption), the model fit improves, but at the cost of mistakenly overestimating the risk aversion of households.

In order to understand precisely how the alternative model of income generates the bias observed in parameter estimates of $\gamma$ in Table 3, it is useful to simulate the alternative model with the parameters fixed at their true values, $(\beta_0, \gamma_0)$. Directly comparing the series for consumption, savings and income generated by the baseline and alternative model, with the income floor set at 5%, enables us to see clearly how the moments of income translate into consumption and savings in each model. These life-cycle profiles are depicted in Figure 4. The ratio between these series can be used as a measure of bias, which is contained in Figure 5.[10]

Figure 5 demonstrates how the bias in income variance is greatest among the categories of consumption, savings and income, while the bias in mean savings is significantly greater than for consumption and income, and it persists for most of the life cycle. The income variance is

[10]Using consumption as an example, at every age $j$, this ratio is defined as $\frac{1}{N}\sum_{i=1}^{N}\tilde{c}_{i,j}$ divided by $\frac{1}{N}\sum_{i=1}^{N}c_{i,j}$.

Figure 4: Model comparison, with 5% income floor.

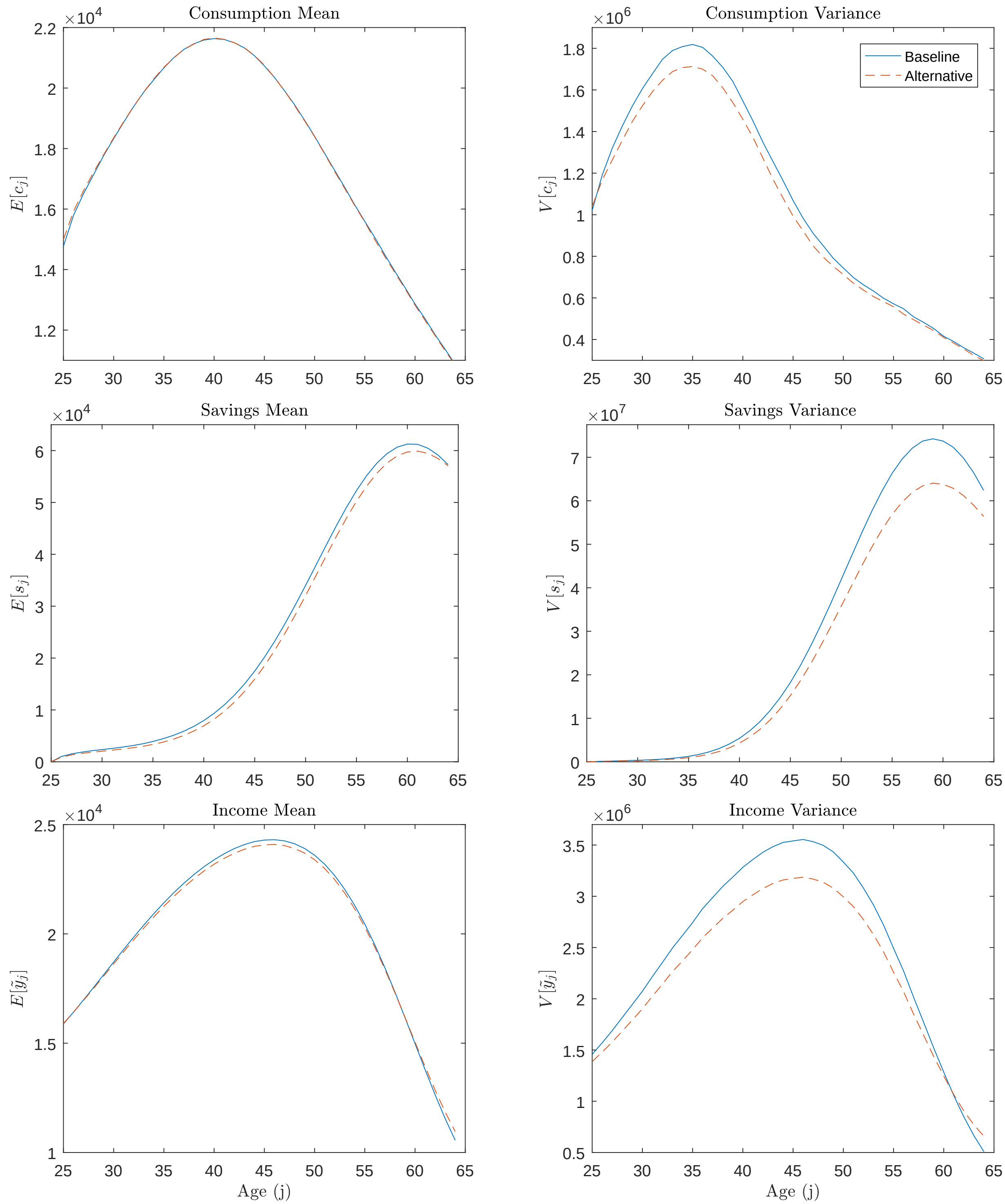
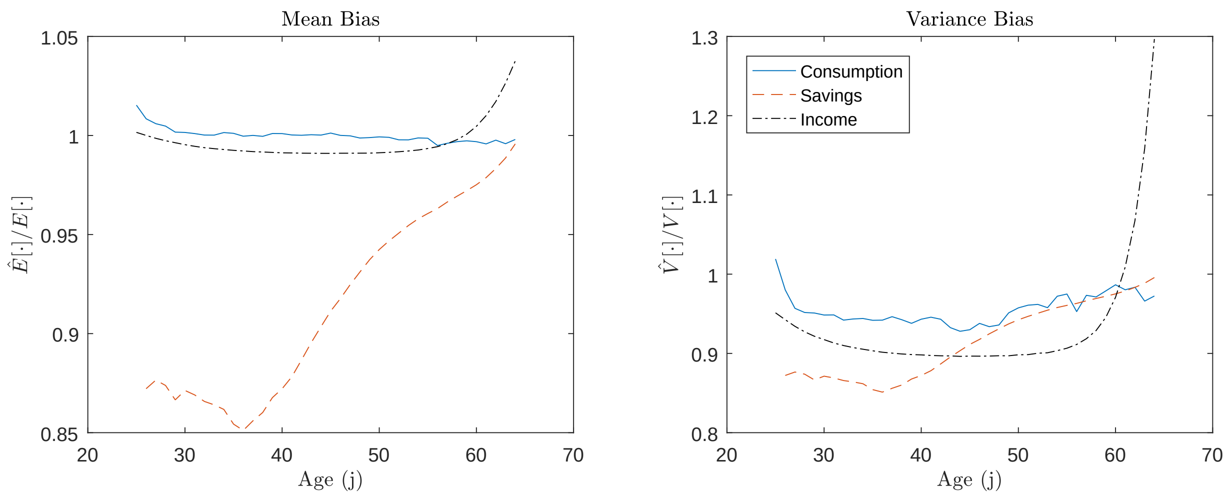


Figure 5: Bias in alternative model, with 5% floor.




consistently underestimated over the course of the life cycle, but as retirement approaches it is severely overestimated. This is consistent with the analysis in Section 3.1, which demonstrated that $\hat{\sigma}_u$ is likely to be overestimated early and late in life. It is the consistent underestimation of the income variance, over the life cycle, which causes household savings to be consistently lower in the alternative model than in the baseline model for all ages. This explains why there is severe bias in the estimated parameter of the coefficient or risk aversion, $\hat{\gamma}$: in order for the alternative model to correct the downward bias in mean savings, it increases $\gamma$ (risk aversion) to generate larger savings over the life cycle. Additionally, the lack of bias in mean consumption also reveals why targeting consumption did not cause significant bias in parameter estimates.

# 5 Conclusion

This paper studies the effect of an income floor on the distribution of income, i.e., the sum of labour earnings and transfers, and addresses whether these features can be adequately approximated by a stochastic process for income with the canonical 'persistent-transitory' AR(1) structure. The income floor, which represents a means-tested government transfer policy, is shown to introduce life-cycle effects into the distribution of income. While this policy does not have any explicit age dimension, its means-testing component interacts with the hump-shape in mean earnings over the life cycle to introduce age dependence into the variance of income over the life cycle. By design, this cannot be captured using a stationary stochastic process. Failing to account for this nonstationarity will lead to the life-cycle income variance being consistently underestimated, resulting in misleading inference regarding key preference parameters, such as the coefficient of risk aversion. Therefore, a researcher should take care to either explicitly model the transfer policy, if known, or attempt to account for the nonstationary component of income, particularly the transitory income shock. This is important for life-cycle studies involving low-income households and means-tested transfer policies.

## References

ABOWD, J. M., AND D. CARD (1989): "On the covariance structure of earnings and hours changes," *Econometrica*, 57(2), 411–445.

BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2008): "Consumption inequality and partial insurance," *American Economic Review*, 98(5), 1887–1921.

CARROLL, C. D. (1992): "The buffer-stock theory of saving: Some macroeconomic evidence," *Brookings Papers on Economic Activity*, 1992(2), 61–156.

COCCO, J. F., F. J. GOMES, AND P. J. MAENHOUT (2005): "Consumption and portfolio choice over the life cycle," *The Review of Financial Studies*, 18(2), 491–533.

COOPER, R., AND G. ZHU (2016): "Household finance over the life-cycle: What does education contribute?," *Review of Economic Dynamics*, 20, 63–89.

EJRNÆS, M., AND M. BROWNING (2014): "The persistent–transitory representation for earnings processes," *Quantitative Economics*, 5(3), 555–581.

GOURINCHAS, P.-O., AND J. A. PARKER (2002): "Consumption over the life cycle," *Econometrica*, 70(1), 47–89.

GUISO, L., AND P. SODINI (2013): "Household finance: An emerging field," in *Handbook of the Economics of Finance*, vol. 2, pp. 1397–1532. Elsevier.

GUVENEN, F. (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12(1), 58–79.

GUVENEN, F., S. OZKAN, AND J. SONG (2014): "The nature of countercyclical income risk," *Journal of Political Economy*, 122(3), 621–660.

HANSEN, G. D., AND A. IMROHOROĞLU (1992): "The role of unemployment insurance in an economy with liquidity constraints and moral hazard," *Journal of Political Economy*, 100(1), 118–142.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Consumption and labor supply with partial insurance: an analytical framework," *American Economic Review*, 104(7), 2075–2126.

HRYSHKO, D. (2012): "Labor income profiles are not heterogeneous: Evidence from income growth rates," *Quantitative Economics*, 3(2), 177–209.

HUBBARD, R. G., J. SKINNER, AND S. P. ZELDES (1995): "Precautionary saving and social insurance," *Journal of Political Economy*, 103(2), 360–99.

KARAHAN, F., AND S. OZKAN (2013): "On the persistence of income shocks over the life cycle: Evidence, theory, and implications," *Review of Economic Dynamics*, 16(3), 452–476.

LILLARD, L. A., AND Y. WEISS (1979): "Components of variation in panel earnings data: American scientists 1960–70," *Econometrica*, pp. 437–454.

LILLARD, L. A., AND R. J. WILLIS (1978): "Dynamic aspects of earning mobility," *Econometrica*, pp. 985–1012.

MACURDY, T. E. (1982): "The use of time series processes to model the error structure of earnings in a longitudinal data analysis," *Journal of Econometrics*, 18(1), 83–114.

ROUWENHORST, G. (1995): "Asset pricing implications of equilibrium business cycle models," in *Frontiers of Business Cycle Research*, ed. by T. Cooley, pp. 294–330. Princeton University Press, Princeton, NJ.

STORESLETTEN, K., C. I. TELMER, AND A. YARON (2001): "How important are idiosyncratic shocks? Evidence from labor supply," *American Economic Review*, 91(2), 413–417.

——— (2004a): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51(3), 609–633.

——— (2004b): "Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy*, 112(3), 695–717.

——— (2007): "Asset pricing with idiosyncratic risk and overlapping generations," *Review of Economic Dynamics*, 10(4), 519–548.

## Appendix A Unemployment

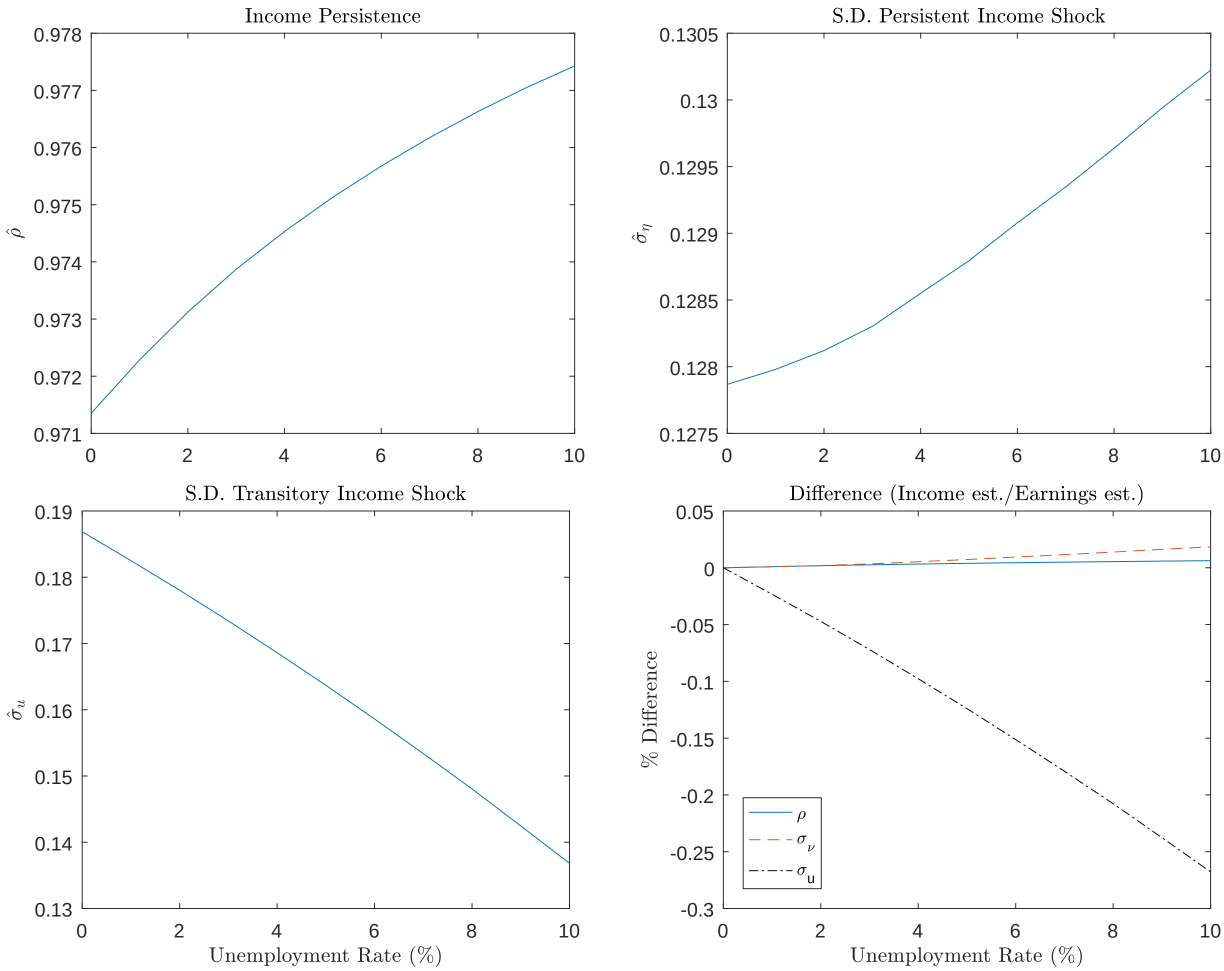
In this section, I demonstrate the effect of unemployment on the parameter estimates of the stochastic income process, from Section 3. The labour earnings process from Section 2, which is conditional on having positive earnings, is complemented by a fraction of households obtaining zero annual earnings. These households correspond to the observations that are typically dropped when this approach is used to estimate the stochastic process of labour earnings. Several previous studies, e.g., Carroll (1992) and Hansen and Imrohoroğlu (1992), have explicitly modelled the probability of unemployment to complement these earnings processes. Instead of specifying a model for unemployment, I demonstrate how the probability of unemployment impacts the parameter estimates of the stochastic income process.[11] Additionally, to prevent the unemployment rate from introducing additional nonstationarities into the distribution of income over the life cycle, I focus exclusively on a constant unemployment rate over the life cycle, i.e., the probability of a household being unemployed is independent of age.

Once unemployment is introduced, an important detail needs to be specified regarding the definition of the income floor: I need to take a stance of whether the income floor is set at a quantile conditional on being employed, or unconditionally. In Section 2.2 this distinction does not appear, since all households are employed. In general, the choice of definition does not matter—it simply affects the interpretation of the value of the income floor. However, it will affect the exercise in this section, since the latter (unconditional) definition implies that the level of the income floor is a function of the unemployment rate. Therefore, if we want to evaluate how the parameter estimates vary as we increase the unemployment rate, this effect will be confounded by a decrease in the income floor. Therefore, I proceed by defining the quantiles to be conditional on the labour earnings distribution with positive earnings, and hence the level of the income floor is independent of the unemployment rate. Finally, it is also worth noting that this approach ensures the income floor is set above the minimum non-zero earnings in the simulated sample, which implies that transfer recipients will include both unemployed and low-earnings employed individuals, consistent with what we observe in the PSID.

Figure 6 depicts the effect of the unemployment rate (i.e., the proportion of individuals at each age $j$ that have zero annual earnings) on the parameter estimates. The income floor is 5%, and hence the parameter estimates for $\hat{\theta}$ at a 0 percent unemployment rate are identical to the estimates found in Table 2, and Figure 2 (at the 5% floor). As the unemployment rate increases, persistence, $\hat{\rho}$, and the standard deviation of the persistent income shock, $\hat{\sigma}_\eta$, both monotonically increase; and, the standard deviation of the transitory income shock, $\hat{\sigma}_u$ monotonically decreases. As the unemployment rate increases, the fraction of households at the income floor increases.

[11] Since the distribution over ages is uniform, the probability of unemployment at every age coincides with the overall unemployment rate in the simulated panel of earnings.

Figure 6: Effect of unemployment on parameter estimates, with 5% floor.




This appears to have two effects: the variance of income declines, indicated by the relatively large reduction in $\hat{\sigma}_u$, and the persistence of income marginally increases, reflected by increases in $\hat{\rho}$ and $\hat{\sigma}_\eta$.

This figure also enables us to infer how unemployment affects the life-cycle consumption and savings decisions of the household, and parameter-estimate bias of $(\hat{\beta}, \hat{\gamma})$, from Section 4. The bottom-right panel of Figure 6 demonstrates that, at all levels of the unemployment rate, the magnitude of the increase in $\hat{\sigma}_\eta$ and $\hat{\rho}$ is insignificant compared to the reduction in $\hat{\sigma}_\eta$. Since $\hat{\sigma}_\eta$ and $\hat{\rho}$ increase as the unemployment rate increases, but generally decreases when the income floor rises, the net effect on $\hat{\sigma}_\eta$ and $\hat{\rho}$, relative to the labour-earnings parameters, is ambiguous. However, since $\hat{\sigma}_u$ is decreasing for all levels of the unemployment rate and income floor, it is unambiguously decreasing. Furthermore, as in Section 3.1, the magnitude of its change is substantially larger than the change in $\hat{\rho}$ and $\hat{\sigma}_\eta$. Therefore, the ambiguity of the net effect on $\hat{\rho}$ and $\hat{\sigma}_\eta$ is unlikely to have a significant effect on the variance, relative to the reduction in $\hat{\sigma}_u$. The total reduction in $\hat{\sigma}_u$ coming from unemployment, in addition to the income floor, will reduce life-cycle variance beyond the estimates obtained in Section 4. Consequently, this amplifies the underestimated life-cycle savings, leading to even greater bias in the (over)estimate of $\hat{\gamma}$.

## Appendix B Distributional effects of an income floor

This section characterises the effect of an income floor on the cross-sectional distribution of income at some age $j > 0$ over the life cycle. The main result is that the first two moments of the resulting income distribution move in opposing directions, relative to the earnings distribution, following the introduction of an income floor. Additionally, these results imply that, with a constant income floor, the magnitude of the variance change will follow the level of mean earnings; i.e., the shape of the life-cycle profile for income variance will match the shape of the mean-profile for labour earnings.

Let age-$j$ labour earnings, $y_j$, be distributed according to cdf $F_j(\cdot)$, with pdf $f_j(\cdot)$. When an income floor is imposed at $\underline{y} \geq 0$, age-$j$ income is given by $\tilde{y}_j = \max(\underline{y}, y_j)$. This policy effectively truncates the distribution of labour earnings from below, and shifts the truncated mass to the truncation point. I will therefore make use of a truncated version of labour earnings, denoted $y_j^*$, with a cdf and pdf $\underline{f}_j(\cdot)$ and $\underline{F}_j(\cdot)$, respectively, defined on $[\underline{y}, +\infty]$. The resulting pdf for income is a mixture of the (truncated) earnings pdf and a probability mass at the income floor. The first two moments for income, $\mathbb{E}_j[\tilde{y}_j]$ and $\mathbb{V}_j[\tilde{y}_j]$, are:

$$
\begin{aligned}
\mathbb{E}_j[\tilde{y}_j] &= F_j(\underline{y})\underline{y} + (1 - F_j(\underline{y}))\mathbb{E}_j[y_j^*] \\
&= F_j(\underline{y})\underline{y} + (1 - F_j(\underline{y})) \int_{\underline{y}}^{\infty} y \underline{f}_j(y) dy \\
&= F_j(\underline{y})\underline{y} + \int_{\underline{y}}^{\infty} y f_j(y) dy \\
\mathbb{V}_j[\tilde{y}_j] &= F_j(\underline{y})(\underline{y} - \mathbb{E}_j[\tilde{y}_j])^2 + (1 - F_j(\underline{y}))[(\mathbb{E}_j[y_j^*] - \mathbb{E}_j[\tilde{y}_j])^2 + \mathbb{V}_j[y_j^*]] \\
&= F_j(\underline{y})(\underline{y} - \mathbb{E}_j[\tilde{y}_j])^2 + \int_{\underline{y}}^{\infty} (y - \mathbb{E}_j[\tilde{y}_j])^2 f_j(y) dy
\end{aligned}
$$

The effect of the policy on these moments (e.g., the difference between $\mathbb{E}_j[\tilde{y}_j]$ and $\mathbb{E}_j[y_j]$) can be characterised as follows:

(1) $\mathbb{E}_j[\tilde{y}_j] \geq \mathbb{E}_j[y_j]$, with equality at $\underline{y} = 0$.

$$\begin{aligned}
\frac{\partial \mathbb{E}_j[\tilde{y}_j]}{\partial \underline{y}} &= F_j(\underline{y}) + f_j(\underline{y})\underline{y} + \frac{\partial}{\partial \underline{y}} \left( \int_{\underline{y}}^{\infty} y f_j(y) dy \right) \\
&= F_j(\underline{y}) + f_j(\underline{y})\underline{y} - f_j(\underline{y})\underline{y} \\
&= F_j(\underline{y}) > 0
\end{aligned}$$

Although (1) indicates that the mean of income is increasing in the floor (and age-dependent, if $\mathbb{E}_j[y_j]$ varies across $j$), this will be captured by the deterministic component of income, if we include age-dummies in the regression on income.

(2) $\mathbb{V}_j[\tilde{y}_j] \leq \mathbb{V}_j[y_j]$, with equality at $\underline{y} = 0$.

$$\begin{aligned}
\frac{\partial \mathbb{V}_j[\tilde{y}_j]}{\partial \underline{y}} &= 2F_j(\underline{y}) \left\{ (\underline{y} - \mathbb{E}_j[\tilde{y}_j])(1 - F_j(\underline{y})) - \int_{\underline{y}}^{\infty} (y - \mathbb{E}_j[\tilde{y}_j]) f_j(y) dy \right\} \\
&= 2F_j(\underline{y}) \int_{\underline{y}}^{\infty} (\underline{y} - y) f_j(y) dy \\
&= 2F_j(\underline{y})(1 - F_j(\underline{y}))(\underline{y} - \mathbb{E}_j[y_j^*]) < 0
\end{aligned}$$

While the variance of income is decreasing in the level of the floor, this can be captured by the estimation process in Section 3 if the resulting change is identical for all $j$. This will not be true, however, if $\mathbb{E}_j[y_j]$ varies across $j$; i.e., mean earnings is age-dependent. Therefore, the variance of income is age-dependent and cannot be captured using a stationary stochastic process. Specifically, as mean earnings increases over the life cycle, the floor binds for fewer households, and hence the variance decrease is small, relative to a lower mean (or higher floor). Consequently, the shape of the life-cycle profile for income variance will match the shape of the mean-profile for labour earnings. This also gives us some insight into why a constant income floor is sufficient to generate this effect.