Reconciling Estimates of Earnings Processes in Growth Rates and Levels

Moira Daly[*]
Copenhagen Business School

Dmytro Hryshko[†]
University of Alberta

Iourii Manovskii[‡]
University of Pennsylvania

June 10, 2015

Abstract

The stochastic process for earnings is the key element of incomplete markets models in modern quantitative macroeconomics. It determines both the equilibrium distributions of endogenous outcomes and the design of optimal policies. Yet, there is no consensus in the literature on the relative magnitudes of the permanent and transitory innovations in earnings. When estimation is based on the earnings moments in levels, the variance of transitory shocks is found to be relatively high. When the moments in differences are used, the variance of the permanent component is relatively high instead. We show theoretically that the difference can be induced by the fact that earnings at the start or at the end of earnings spells are lower and more volatile than the observations in the interior of earnings histories. Using large administrative datasets from Denmark and Germany, we show that this property of earnings spells quantitatively accounts for the full amount of discrepancy in the estimates. Using data from the Panel Study of Income Dynamics, we show that this property of earnings induces a substantial upward bias in the estimate of consumption insurance against permanent shocks.

JEL Classifications: D52, D91, E21, J31.

[*]Department of Economics, Copenhagen Business School, Porcelaenshaven 16A, 2000 Frederiksberg, Denmark. E-mail: moda.eco@cbs.dk.

[†]Department of Economics, University of Alberta, 8–14 HM Tory Building, Edmonton, AB, T6G2H4, Canada. E-mail: dhryshko@ualberta.ca.

[‡]Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297 USA. E-mail: manovski@econ.upenn.edu.

1 Introduction

The central element of many models in modern quantitative macroeconomics with heterogeneous agents is either an exogenously specified or an endogenously determined stochastic process for individual earnings. For example, in the models with incomplete insurance markets the properties of the earnings process serve as key determinants of the evolution of consumption, assets, and other economic choices over the life cycle and across individuals.[1] Following the seminal contribution by Friedman (1957), modern consumption theory recognizes that consumption should respond more to the longer lasting or permanent than to transitory innovations in earnings. This explains the keen interest in the literature in measuring the variances of these components using the variants of the permanent/transitory earnings decomposition pioneered by Friedman and Kuznets (1954) and later found to have sound empirical support in, e.g., MaCurdy (1982), Abowd and Card (1989), and Meghir and Pistaferri (2004).[2] In its basic form, such earnings process can be written as:

$$\begin{aligned} y_{it} &= \alpha_i + p_{it} + \tau_{it} \\ p_{it} &= \phi_p p_{it-1} + \xi_{it} \\ \tau_{it} &= \theta(L)\epsilon_{it}, \end{aligned} \tag{1}$$

where log-earnings y_{it} of individual i at time t consist of the permanent component p_{it}, and the transitory component, τ_{it}. If ϕ_p is close to one, the shocks ξ_{it} are highly persistent (truly permanent if ϕ_p is one), and if $\theta(L) = 1$ (where $\theta(L)$ is a moving average polynomial in the lag operator L), the shocks ϵ_{it} are completely transitory.

In addition to determining equilibrium consumption and wealth distributions, the variance and persistence of the shocks ξ_{it} and ϵ_{it} have important implications for policy design. For example, they are key for determining the optimal design of the bankruptcy code in Livshits, MacGee, and Tertilt (2007), they govern the impact of the welfare system on household savings in Hubbard, Skinner, and Zeldes (1995), stimulus effects of fiscal policy in Heathcote (2005), as well as the optimal design of the tax system in Banks and Diamond (2010) and Farhi and Werning (2012). Moreover, there is great interest in understanding whether the

[1]See, e.g., Deaton (1991), Carroll (1997), Castañeda, Díaz-Giménez, and Ríos-Rull (2003).

[2]A prominent alternative in the literature allows for less persistent shocks but individual-specific trends in earnings. Guvenen (2009) is a leading recent example.

dramatic increase in earnings dispersion over the last few decades in the U.S. is due to the increase in the variances of persistent or transitory shocks, e.g., Gottschalk and Moffitt (1994). This is relevant for understanding why consumption inequality did not increase nearly as much, e.g., Krueger and Perri (2006), Blundell, Pistaferri, and Preston (2008), Heathcote, Storesletten, and Violante (2010), Attanasio, Hurst, and Pistaferri (2012). Knowing the stochastic nature of earnings is also essential for the design of active labor market policies. For example, Meghir and Pistaferri (2011) suggest that income maintenance policies might be an appropriate response to changes in inequality driven by transitory shocks, while training programs are potentially more relevant to counteract the effects of permanent shocks.

Unfortunately, despite their manifest importance, there is no consensus in the vast existing empirical literature on the sizes of the shocks ϵ_{it} and ξ_{it}. The key problem confronting this literature is that, using the same data, the estimates of the earnings process in equation (1) when targeting the moments of log-earnings in levels are dramatically different from the estimates obtained when fitting the moments of log-earnings in differences. This led Heathcote, Perri, and Violante (2010) to conclude that the widely used model of earnings dynamics in equation (1) is misspecified. However, the nature of this potential misspecification is unknown. Consequently, the conclusions of the models that use this earnings process as a primitive cannot be fully relied upon. Even if this process is used as a primitive due to the lack of a better alternative, there is no consensus on whether the parameter values estimated in levels or differences should be used. Relatedly, in the literature that endogenizes the earnings process (e.g., Huggett, Ventura, and Yaron (2011) and Postel-Vinay and Turon (2010)), it is unclear whether the implied process generated by the model should be compared to the one estimated in the data using the specification in levels or in differences, given that estimating the reduced-form process (1) on the model-generated data does not give rise to the observed discrepancy.

In this paper we uncover an important source of this misspecification. While the mechanism we describe applies to survey-based and administrative data alike, our primary focus is on understanding the source of the discrepancy in large administrative datasets that are becoming central in the literature.[3] These datasets are typically orders of magnitude larger than survey-based ones, they are free

[3]Recent contributions include Blundell, Graber, and Mogstad (2015), DeBacker, Heim, Panousi, Ramnath, and Vidangos (2013), Domeij and Flodén (2010), Guvenen, Ozcan, and Song (2014), among others.

of sampling issues, they do not suffer from the typical issues of attrition, except what is due to international migration and death, they are based on administrative sources, such as tax records, and are considered highly reliable and free of issues of systematic non-response or measurement error typically plaguing survey-based data. However, despite numerous attractive properties, we show that these datasets have features that generically bias the estimates of earnings processes and generate the large discrepancy in the estimates based on moments in growth rates and in levels. Fortunately, we show that it is relatively easy to account for these features in estimation in order to eliminate the discrepancy and to obtain consistent estimates.

Estimation of the parameters of the earnings process in the literature is based on fitting the (entire) set of autocovariance moments for levels or differences of earnings. However, even when estimation is based on the same set of observations in the data, computation of the autocovariance moments in levels and differences is effectively based on different information. To clarify with an extreme example, consider an individual with a single earnings observation in the sample. This observation will contribute to the estimated variance of earnings in levels, but it will not contribute to any moment in differences. More generally, some individual contributions towards the autocovariance moments are not defined because there are no earnings observations before the start of an individual's earnings history, nor subsequent to its end, or due to missing data in the interior of the earnings history. We show theoretically that the discrepancy in the estimates arises since individual contributions to different autocovariance moments are not defined due to missing data when earnings are taken in levels and in differences, and since earnings observations surrounding missing observations are not random. Indeed, we document that in the data the earnings at the time an individual permanently enters or exits the sample, or the earnings surrounding the missing observations, are systematically different. In particular, they are considerably lower on average and substantially more volatile. This can be expected. For example, the data on earnings are typically recorded at an annual frequency. An individual, say, entering the sample for the first time is (statistically) expected to enter in the middle of the year, but may enter at any point throughout the year. Thus, earnings in that year are expected to be lower and have a larger variance than interior earnings observations from contiguous earnings histories. We will show formally below that the low mean and high variance of earnings surrounding missing observations raises the variance of transitory shocks when estimation relies on the moments in levels and the variance of permanent shocks recovered by estimation based on the moments in differences.[4]

We quantitatively assess the magnitude of these biases using large administrative datasets from Denmark and Germany and find that they fully account for the discrepancy between the estimates using data in levels and in differences. The Danish data contain complete earnings histories of each resident of Denmark from 1981 through 2006. The German data are a 2% random sample of social security numbers. For these individuals, the complete earnings history from 1975 through 2008 is available. These samples are sufficiently large to allow analysis at the level of particular age cohorts making it possible to focus on a parsimonious earnings model in (1), sidestepping the issue of modelling cohort effects. Moreover, the large size of the data enables reliable estimation when replicating the design of samples typically used in the literature. Specifically, we consider a balanced sample spanning 25 (26) years in German (Danish) data, a sample with 9 or more consecutive observations as in e.g., Browning, Ejrnæs, and Alvarez (2010) and Meghir and Pistaferri (2004), and a sample with 20 or more not necessarily consecutive observations as in e.g., Guvenen (2009). Our smallest Danish sample is comprised of about 67,000 individuals and 1.7 million observations, while our smallest German sample contains about 10,000 individuals with more than 200,000 observations.

Using the unbalanced samples in both datasets, we find, consistently with the literature, a substantially higher estimated variance of permanent (transitory) shocks targeting the moments of earnings in growth rates (levels). Perhaps more surprisingly, we find that the discrepancy is nearly absent in balanced samples drawn from the two datasets. For the vast majority of individuals in the balanced sample their first year in the sample does not coincide with the first year of their earnings history. Similarly, their last year in the sample mechanically truncates earnings histories, implying that it is not the last year of the earnings spell of individuals in the sample. Thus, the mean and the variance of earnings in the first and the last sample years are similar to those of the other years. By definition, the balanced sample also does not contain missing observations. This suggests that it is the non-randomness of earnings surrounding missing observations in the unbalanced samples that drives the discrepancy between the estimates in

[4]Consistently with this interpretation, we find much smaller differences in the estimated variances of permanent and transitory shocks when using the moments in levels and differences in *wages*, as much of the variability in earnings in our administrative datasets at the start and end of contiguous spells is due to the variability in hours.

levels and differences on the data from the unbalanced samples. However, it is still possible that the earnings processes of individuals in the unbalanced samples are fundamentally different and misspecified in some other way. To exclude this possibility, we proceed in three steps.

First, we quantify the contribution of the low mean and high variance of earnings surrounding missing observations in the unbalanced samples drawn from German and Danish data to the subset of theoretical autocovariance moments on which the identification argument in levels and differences is based, and confirm that they induce the observed discrepancy in the estimates. Second, using unbalanced samples, we drop a few observations at the start and at the end of the earnings history, as well as observations surrounding missing records. We find that estimating the earnings process in levels and in differences on the remaining data yields virtually identical estimates of the variances of permanent and transitory shocks. Third, we simulate artificial data based on these estimates of the earnings process while replicating the structure of the unbalanced samples (by design of this experiment, first and last observations as well as those surrounding missing observations are not systematically different from observations in the rest of the earnings histories). We find no discrepancy of the estimates in levels and differences in these artificial data. We then draw an additional transitory shock ("rare transitory shock") at the start and end of the earnings history and surrounding missing observations to replicate the mean and the variance of earnings in those periods in the data. We find that in this case the estimates of the variance of permanent and transitory shocks are very different when moments in levels and differences are used, but are very close to those in the data from the corresponding unbalanced samples.

Having established that the rare transitory shocks at the start and end of earnings histories and surrounding interior missing observations are the source of discrepancy in the estimated variances of permanent and transitory shocks in our large administrative datasets using the moments for earnings growth rates or levels, we illustrate the importance of accounting for these shocks for understanding consumption responses to earnings shocks. To this end, we follow Blundell, Pistaferri, and Preston (2008) and estimate consumption insurance coefficients for permanent and transitory idiosyncratic earnings shocks, i.e., the fraction of those shocks that does not translate into movements in consumption. Using their male earnings data drawn from the Panel Study of Income Dynamics (PSID), we find that the presence of rare transitory shocks at the start and end of earnings histories leads to a substantial upward bias in the estimated insurance against permanent shocks. We show theoretically that the bias is driven by the same forces that cause overestimation of the variance of permanent shocks using the earnings moments in growth rates. The rare transitory (and highly insurable) shocks are effectively "misinterpreted" by those moments as being permanent.

While the mechanism described in this paper is exceptionally powerful in reconciling the estimates of the earnings process in growth rates and levels, it is not the only mechanism that can generate such discrepancy. For example, Hryshko and Manovskii (2015) show that this mechanism does not eliminate the full amount of discrepancy in the estimates of the stochastic process for household disposable income in PSID data. Instead, they show that the remaining discrepancy is primarily driven by the typical restriction on the persistence of the permanent component, which limits its heterogeneity in the sample. This heterogeneity, however, is shown to be largely induced by the PSID sampling procedures. Importantly, this type of misspecification cannot generate the difference between the theoretical moments that we use to establish identification in levels and differences in this paper because they make use of exactly the same earnings data and are identically affected by any such misspecification. These theoretical identifying moments can only differ if the underlying autocovariance moments on which they are based disagree and we show that this is indeed the consequence of the low mean and high variance of observations at the start and end of earnings spells. We find that this accounts for virtually all discrepancy of the estimates in growth rates and levels in the earnings data we consider.

The rest of the paper is organized as follows. In Section 2 we discuss identification of the permanent-transitory decomposition of earnings, and derive theoretically the biases in the estimated variances of permanent and transitory shocks when using the moments in levels and differences constructed from an unbalanced panel. In Section 3 we describe the data and the estimation procedure. In the same section we present basic estimation results and document that earnings are typically lower and more volatile in the periods surrounding missing observations. In Section 4 we show that this property of earnings quantitatively accounts for the difference in estimates of earnings processes in levels and differences. In Section 5 we study theoretically and quantitatively the bias induced by this property of earnings on the insurance coefficients against permanent and transitory shocks. Section 6 concludes.

2 Sources of the Differences

Estimation of the parameters of the earnings process in the literature typically relies on the minimum-distance method. In particular, estimation based on the moments in levels targets the entire set of autocovariance moments in levels $E[y_{it}y_{it+j}]$, where $i \in [1, N]$ denotes individuals in the sample, t denotes time, and j denotes all the leads and lags of earnings observed in the data. In differences, estimation targets the full set of autocovariance moments in differences $E[\Delta y_{it}\Delta y_{it+j}]$, where Δ is the difference operator between two consecutive observations, so that $\Delta y_{it} \equiv y_{it} - y_{it-1}$.

While all available autocovariance moments are used in estimation, the identification is usually established using only a subset of autocovariance moments, e.g., Meghir and Pistaferri (2004), Blundell, Pistaferri, and Preston (2008), Hryshko (2012), and Heathcote, Storesletten, and Violante (2014). For example, consider the earnings process that consists of a random walk and an iid transitory shock—this corresponds to setting $\theta(L)$ and ϕ_p to 1 in equation (1). This process was considered in Heathcote, Perri, and Violante (2010) who proposed the following moments to identify the variances of permanent and transitory shocks at time t:

`Differences:`

$$\sigma^2_{\xi,t} = E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}], \tag{D1}$$

$$\sigma^2_{\epsilon,t} = -E[\Delta y_{it}\Delta y_{it+1}]. \tag{D2}$$

Note that (D1) and (D2) represent linear combinations of autocovariance moments for earnings growth rates. For clarity, we will refer to individual autocovariance moments as simply "moments," and to a linear combination of autocovariance moments used for identification such as (D1) and (D2) as "identifying moments."

Expanding (D1) and (D2), we obtain the identifying moments for the variances of permanent and transitory shocks, based on autocovariance moments in levels, at time t:

`Levels:`

$$\sigma^2_{\xi,t} = E[y_{it}y_{it+1}] - E[y_{it+1}y_{it-1}] - E[y_{it}y_{it-2}] + E[y_{it-1}y_{it-2}], \tag{L1}$$

$$\sigma^2_{\epsilon,t} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}] - E[y_{it-1}y_{it}] + E[y_{it-1}y_{it+1}]. \tag{L2}$$

As identifying moments (D1)-(D2) and (L1)-(L2) are based on exactly the same earnings information, they are expected to deliver identical estimates of the variance of permanent and transitory shocks at time t in a sample of individuals whose earnings are nonmissing for the periods $t-2$ through $t+1$.[5]

Importantly, each autocovariance moment is measured as the average across all available observations that contribute to it. This implies that, while the identifying moments (D1)-(D2) and (L1)-(L2) are based on the same earnings data, the autocovariance moments used in estimation of (D1)-(D2) and (L1)-(L2) are computed using different sets of observations. To take an extreme example, consider an individual who appears in the sample only once, in period t. This individual will contribute to the autocovariance moment $E[y_{it}y_{it}]$ and thus his only earnings observation will affect the identifying moment (L2) but it will not contribute to any autocovariance moment used to construct the corresponding identifying moment in differences (D2). If earnings of individuals who appear in the sample only once are systematically different, this will induce the difference between identifying moments (L2) and (D2) and lead to different estimates of the variance of transitory shocks using the moments in levels and differences.

While the preceding example seems pedagogically insightful, empirically we find that the difference between the autocovariance moments used in computing (D1)-(D2) and (L1)-(L2) is driven by the fact that in our data, earnings at the time an individual permanently enters or exits the sample, or earnings surrounding the missing observations, are systematically different (they are typically lower and substantially more volatile). As earnings observations at the time individuals (re-)enter and exit the sample contribute differently to the autocovariance moments on which the identifying moments (D1)-(D2) and (L1)-(L2) are based, this leads to systematic differences in estimated variances of permanent and tran-

[5]Note that Heathcote, Perri, and Violante (2010) show that identifying moments in levels can be constructed using fewer autocovariance moments such as

$$\sigma^2_{\xi,t} = E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}], \tag{L1-Short}$$

$$\sigma^2_{\epsilon,t} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}]. \tag{L2-Short}$$

These identifying moments in levels do not, however, use the same information as the identifying moments (D1)-(D2) in differences. For example, the information on earnings in $t-2$ is used in (D1) but not in (L1-Short). The assessment of the sources of biases is more transparent using moments (D1)-(D2) and (L1)-(L2) which rely on the same information. Moreover, as the moments (L1)-(L2) simply represent an expansion of the moments (D1)-(D2), they are identically affected by any other potential misspecification of the earnings process. This allows us to isolate and measure the importance of the high variance and low mean of the observations at the start and end of contiguous earnings observations, which, as we show below, contribute differently to the autocovariance moments on which (D1)-(D2) and (L1)-(L2) are based.

sitory shocks using the moments in growth rates and levels. In the rest of this section we formally describe the associated biases. In subsequent sections we will show that they account for the entire difference in the estimates using the identifying moments in levels and differences.

We will consider three types of samples. Consider a dataset with panel data on individual earnings that starts in period t_0 and ends in period T. We refer to the sample as *balanced* if all individuals in the sample have $T - t_0 + 1$ valid earnings observations. While not part of the formal definition, it is convenient to think that earnings spells of individuals in the balanced samples start prior to t_0 and end subsequent to T. In other words, the boundaries of the balanced sample mechanically truncate continuous earnings spells in progress. We refer to samples that include only uninterrupted earnings spells (i.e., no gaps) but with duration of less than $T - t_0 + 1$ for at least some individuals as *consecutive unbalanced* samples. Finally, we refer to unbalanced samples that also include individual earnings spells interrupted by missing observations in any period $t \in (t_0, T)$ as *non-consecutive unbalanced* samples.

Consecutive unbalanced samples. The nature of these samples is such that at least some individuals are observed starting or ending their earnings spells inside the sample. As mentioned above, earnings have a lower mean and are highly volatile in the first and last period of an incomplete earnings history. Consider modeling this through an additional transitory shock in the first and last year of an individual's earnings history, that is

$$y_{it} = \alpha_i + p_{it} + \epsilon_{it} + \nu_{it},$$

where ν_{it} has mean μ_ν (taking a negative value) and variance σ_ν^2 and is uncorrelated with permanent and transitory shocks. Hereafter, we refer to the shock ν_{it} as a rare transitory shock, and call an earnings observation y_{it}, affected by this shock, an outlying earnings observation. We now show that ignoring ν_{it} and estimating the process (1) instead leads to an upward bias in the estimated variance of permanent shocks using the moments in differences and in the estimated variance of transitory shocks using the moments in levels.

For simplicity, assume there is a set of individuals entering first into the sample at time t, in the interior of the sample period $[t_0, T]$, while the remaining individuals are continuously observed throughout the sample. Individuals first appearing at time t will contribute to estimation of the autocovariance moments

$E[y_{it}y_{it}]$ and $E[y_{it}y_{it+1}]$ in the identifying moment (L2), used to back out the variance of the transitory shock at time t by targeting the moments in levels. Since the rare shock at t is assumed to be uncorrelated over time, the estimated moment $E[y_{it}y_{it+1}]$ will be no different for such individuals relative to the rest of the sample, and will equal $\text{var}(p_{it})$. The other moments in (L2), $E[y_{it-1}y_{it}]$ and $E[y_{it-1}y_{it+1}]$ will both equal $\sigma_\alpha^2 + \text{var}(p_{it-1})$. The autocovariance moment $E[y_{it}y_{it}]$ estimated on the full sample, however, will equal $\sigma_\alpha^2 + \text{var}(p_{it}) + \sigma_{\epsilon,t}^2 + s_t(\mu_\nu^2 + \sigma_\nu^2)$, where s_t is the share of individuals, at time t, whose (incomplete) spells start at time t in the total number of individuals at time t with nonmissing earnings. The identifying moment (L2), therefore, will recover an estimate of the variance of transitory shocks equal to $\sigma_{\epsilon,t}^2 + s_t(\mu_\nu^2 + \sigma_\nu^2)$, with an upward bias of $s_t(\mu_\nu^2 + \sigma_\nu^2)$. Since the rare shock is assumed to be uncorrelated, there are no consequences, at any point in time, for the estimated magnitude of the identifying moments (L1), and (D2).

The variance of permanent shocks at time $t + 1$ will, however, be also biased upward. Individuals first appearing at t will contribute to estimation of the autocovariance moments $E[\Delta y_{it+1}\Delta y_{it+1}]$ and $E[\Delta y_{it+1}\Delta y_{it+2}]$ in the identifying moment (D1), used for backing up the variance of permanent shocks at time $t+1$ by targeting the moments in growth rates. For such individuals the autocovariance moment $E[\Delta y_{it+1}\Delta y_{it+2}]$ will be no different from the rest of the sample and will equal $-\sigma_{\epsilon_{t+1}}^2$, while the autocovariance moment $E[\Delta y_{it+1}\Delta y_{it+1}]$ will equal $\sigma_{\xi_{t+1}}^2 + s_{t,t+1}(\mu_\nu^2 + \sigma_\nu^2) + \sigma_{\epsilon_t}^2 + \sigma_{\epsilon_{t+1}}^2$, where $s_{t,t+1}$ is the share of individuals who start (incomplete) earnings spells at time t, with nonmissing earnings at times t and $t + 1$, in the number of individuals with nonmissing earnings both at t and $t + 1$. Since the autocovariance moment $E[\Delta y_{it+1}\Delta y_{it}]$ will be estimated using information for those individuals whose earnings are nonmissing in periods $t - 1$ through $t + 1$ and will equal $-\sigma_{\epsilon_t}^2$, the identifying moment (D1) for time $t + 1$ will recover an estimate of the permanent shock equal to $\sigma_{\xi_{t+1}}^2 + s_{t,t+1}(\mu_\nu^2 + \sigma_\nu^2)$, with an upward bias of $s_{t,t+1}(\mu_\nu^2 + \sigma_\nu^2)$.

It is also worth noting that if the rare shock appears, say, first at time $t+1$, i.e. in the interior of an earnings spell for individuals first entering into the sample at time t, it will simply elevate, by the same magnitude, the estimated variance of transitory shocks in levels and differences at time $t + 1$, with no differential effect on the identifying moments (L2) and (D1).

Summing up, incomplete earnings spells first appearing in the sample at t will bias upward the estimated variance of transitory shocks at time t when targeting the moments in levels, and the variance of permanent shocks at time $t+1$ when targeting the moments in differences.

The same logic extends to the incomplete earnings spells ending at time t, which is different from the last potential sample year T—the presence of such spells will produce upward-biased estimates of permanent variances in differences at t (since these individuals will contribute to estimation of the moment $E[\Delta y_{it} \Delta y_{it}]$ which is part of the identifying moment (D1)) and of transitory variances in levels at t.

Non-consecutive unbalanced samples. We now consider the consequences of missing earnings in the interior points of the earnings history. We assume that individual earnings are realizations of the earnings process (1), with some observations missing in any period $t \in (t_0, T)$. We will show below that such periods are often associated in the data with high variance of earnings in periods $t-1$ and $t+1$. We model this by introducing additional rare transitory shocks with a negative mean μ_ν at the time before and after earnings is missing (ν_{it-1} and ν_{it+1}, respectively) that are assumed to be uncorrelated with permanent and transitory shocks, and each other:[6]

$$
\begin{aligned}
&y_{it-1} = \alpha_i + p_{it-1} + \epsilon_{it-1} + \nu_{it-1}, \\
&y_{it} \ \text{missing}, \\
&y_{it+1} = \alpha_i + p_{it+1} + \epsilon_{it+1} + \nu_{it+1}.
\end{aligned}
$$

Assume there is a set of individuals whose earnings are missing at time t, which is interior to the sample period $[t_0, T]$, while the rest of individuals have continuously observed earnings throughout the whole sample period.

In this case, the variance of transitory shocks at times $t-1$ and $t+1$ using the moments in levels will be biased upward as the autocovariance moments $E[y_{it-1} y_{it-1}]$ and $E[y_{it+1} y_{it+1}]$ in the identifying moment (L2) will be amplified by the variation of the rare shocks. Similarly, the variance of permanent shocks at times $t-1$ and $t+2$ using the moments in differences will be biased upward as the autocovariance moments $E[\Delta y_{it-1} \Delta y_{it-1}]$ and $E[\Delta y_{it+2} \Delta y_{it+2}]$ in the identifying moment (D1) will be amplified by the variation of the rare shocks. Since the rare shocks are assumed to be uncorrelated, the identifying moments (L1) and

[6] For the ease of exposition, we assume that the mean and variance of the rare shock one year before and after earnings are missing are the same, although they are slightly different in the data.

(D2) will not be affected.

Summing up, incomplete earnings spells with missing earnings at t, in the interior of the sample period, will bias upward the estimated variance of transitory shocks at times $t-1$ and $t+1$ when targeting the moments in levels, and the variance of permanent shocks at times $t-1$ and $t+2$ when targeting the moments in differences.

Summary. The analysis above yields two major implications. First, one may expect to recover without any biases the variance of transitory shocks using the moments in growth rates, and the variance of permanent shocks using the moments in levels. Second, the identifying moments in levels tend to produce upward-biased estimates of the variance of transitory shocks, while the identifying moments in differences produce upward-biased estimates of the variance of permanent shocks. The magnitude of the biases depends positively on the variance of the rare shocks and on the difference between their mean from the mean of the shocks in the rest of earnings histories.

3 Data, Estimation Details, and Basic Results

3.1 Data

In this section we describe the data and construction of the samples that we study. Following the literature, we focus on individuals with a strong attachment to the labor market characterized by sufficiently high earnings and time spent working.[7]

[7] The selection rules we adopt are typical of not only the literature that utilizes survey data but also of the recent literature utilizing administrative data. For example, Guvenen, Ozcan, and Song (2014) use US administrative data on individual wage and salary income and make the following sample selection: "For a statistic computed using data for not necessarily consecutive years $t_1, t_2, \ldots, t_n$, an individual observation is included if the following three conditions are satisfied for all these years: the individual (i) is between the ages of 25 and 60, (ii) has annual wage/salary earnings that exceed a time-varying minimum threshold, and (iii) is not self-employed (i.e., has self-employment earnings less than the same minimum threshold). This minimum, denoted $Y_{min,t}$, is equal to one-half of the legal minimum wage times 520 hours... This condition allows us to focus on workers with a reasonably strong labor market attachment and avoids issues with taking the logarithm of small numbers. It also makes our results more comparable to the income dynamics literature, where this condition is standard." Similarly, DeBacker, Heim, Panousi, Ramnath, and Vidangos (2013) "...exclude earnings (or income) observations below a minimum threshold..." and "...take the relevant threshold to be one-fourth of a full-year, full-time minimum wage." In line with our selection of consecutive unbalanced samples (with the difference that we use at least 9 consecutive earnings observations), Blundell, Graber, and Mogstad (2015) "...restrict the sample to individuals with at

3.1.1 Danish data

Several administrative registers provided by Statistics Denmark were used to construct our samples. The tax register from 1980–2006 provides panel data on total earnings for more than 99.9 percent of Danish residents between the ages of 15 and 70. The register was merged with the Danish Integrated Database for Labor Market Research (IDA) so that additional demographic variables, such as educational status could be appended. The population consists of Danish males born in 1951 through 1955. We observe annual earnings over the period of 1980 through 2006. We first remove all individuals who were ever self-employed and drop records in which an individual was making non-positive labor market earnings. Next, we drop records for those individuals who have worked less than 10 percent of the year as a full time employee; this restriction limits our data to the period 1981–2006 as we cannot identify the full-time employment status for the year 1980.[8] Annual earnings in a particular year include all earned labor income, taken from tax records, for that calendar year. This variable is considered "high quality" by Statistics Denmark in that it very accurately captures the earnings of individuals. Earnings are expressed in 1981 monetary units (Danish kroner). We calculate the maximum number of consecutive periods in which an individual has nonmissing earnings and use this information to construct two consecutive samples: a sample in which an individual's maximum spell is at least 9 consecutive periods (102,825 individuals), and a sample in which the individual's maximum spell covers the entire 26 periods, hereafter called "balanced" sample (67,008 individuals). For the sample with 9 or more consecutive observations, periods outside of the longest spell are dropped. Within the longest spell, earnings outliers are defined as an individual with an increase in earnings of more than 500 percent or a fall of more than –80 percent in adjacent years. Individuals with earnings outliers within their longest spell are dropped. The third sample we consider consists of individuals who have at least 20 not necessarily consecutive periods in which they have nonmissing earnings (90,668 individuals). We also drop individuals in this sample if they have earnings growth outliers. Finally, we drop individuals if their educational status has changed during the

least four subsequent observations with positive market income."

[8] We use the variable "erhverv" from the IDAP table provided by Statistics Denmark. This variable calculates work experience as a full time employee since 1980 based on individuals' yearly pension contributions and is available for all members of the population (with the exception of those individuals who have spent time abroad for whom the variable is reset to 0). By taking the first difference of this measure, we can calculate the percent of the year an individual has worked full time, which restricts our observation period to 1981–2006.

spells considered. Table 1 contains basic statistics for selected samples.

3.1.2 German data

We use administrative data from the IABS, a 2% random sample of German social security records for the years 1974–2008. A detailed description of the dataset can be found in Dustmann, Ludsteck, and Schönberg (2009). We use full-time job spells for German males born in 1951–1955, dropping the spells in East Germany. We also drop annual records when an individual was in apprenticeship during any part of the year. Individual real earnings is the sum of earnings from all jobs within a year expressed in 2005 euros. We set individual education to the maximum schooling attained during the sample years, and the number of days worked to the sum of calendar days on all jobs within a year. As individual earnings are right-censored at the highest level subject to social security contributions, we impute earnings exceeding the limit assuming that daily wages in the upper tail follow a Pareto distribution, the parameters of which differ by year and an age group.[9] After 1983, earnings include one-time payments such as bonuses. To make variable definitions consistent throughout, we use only the data since 1984. We further drop individual records on annual earnings if the combined duration of job spells within a year is below 35 calendar days, and records with very low daily earnings.[10] As in the Danish data, we construct three samples—balanced, with 9 or more consecutive, and 20 or more not necessarily consecutive earnings observations—and, similarly to the Danish samples, drop individuals who have earnings growth outliers. The respective samples contain 9,452, 18,130, and 13,635 individuals with 236,300, 379,080, and 330,748 observations, respectively. Table 2 provides some descriptive details for the samples.

[9] We consider the following eight age groups: younger than 25, 6 five-year age groups (ages 25–29, 30–34, up to the group of 50–54 years old), and those who are older than 54. We use a "fixed effects" imputation, keeping a uniform draw for each individual affected by the right-censoring limit fixed, when creating a Pareto variate in different years. We also experimented with imputation based on the assumption that truncated log-wage distribution is normal, and a simpler imputation when daily wage is multiplied by the factor 1.2 if it hits the upper censoring limit—these three imputation methods have been used in Dustmann, Ludsteck, and Schönberg (2009). Our conclusions below are robust to the choice of the imputation method as well as to limiting the sample to individuals whose earnings histories are not affected by the censoring.

[10] The highest marginal part-time income threshold during the sample period was 13.15 euros a day (set for the first time in 2003), and we drop the records with daily earnings below 14 euros in 2003 prices in any year.

3.2 Estimation Details

As is standard in the literature, we estimate the earnings process in equation (1) using the method of minimum distance, fitting the data autocovariance function of log-earnings in levels and first differences to the autocovariance function implied by the model.[11] We allow for an AR(1) transitory component and an unrestricted estimation of the persistence of the permanent component, ϕ_p. We estimate five parameters in total—the persistence and the variance of permanent shocks, ϕ_p and σ_ξ^2; the persistence and the variance of transitory shocks, ϕ_τ and σ_ϵ^2; and the variance of individual fixed effects, σ_α^2.[12] The model, in reduced form, corresponds to an ARMA(2,1) process in levels.[13] The autoregressive part of the reduced form would allow identification of autoregressive parameters of the persistent and transitory processes, while the MA part, containing two unique parameters, would allow identification of the variance of persistent and transitory shocks; see, e.g., Harvey (1989), for identification of models with unobserved components. We assume that individuals start accumulating permanent and transitory shocks at the age of 25 so that part of the estimated variance of fixed effects captures the accumulated permanent and transitory components prior to that age. We remove predictable variation in earnings by estimating cross-sectional regressions of log earnings on educational dummies, a third polynomial in age, and the interactions of the age polynomial with the educational dummies. Our measure of idiosyncratic earnings, consistent with the literature, is the residual from those regressions. Since our samples are large, we estimate the model using the optimal weighting matrix which is an inverse of the variance-covariance matrix of the data moments.

3.3 Basic Results

3.3.1 Samples with 9 or more consecutive observations

Columns (1)–(4) in Table 3 contain estimation results for the samples with 9 or more consecutive observations. The first two columns use German data. The permanent component is estimated to be close to a random walk using the moments

[11] One of the recent exceptions is Browning, Ejrnæs, and Alvarez (2010) who, apart from selected moments in levels and differences, fit a variety of other data moments studied in the literature on earnings dynamics.

[12] In differences, the variance of fixed effects is not identified.

[13] If the permanent component is a random walk, the reduced form model for log earnings in first differences is an ARMA(1,1) process which is supported, for example, in U.S. data—see MaCurdy (2007).

in differences, while the persistence of the permanent component is estimated to be somewhat lower using the moments in levels. The same pattern can be seen in estimations utilizing Danish data—see columns (3) and (4). Importantly, in both datasets the variance of the permanent shock is more than two times larger in the estimation that uses the moments in growth rates, while the variance of the transitory shock is about two times larger using the moments in levels. Thus, our data exhibit the same large discrepancy endemic to this literature.

3.3.2 Samples with 20 or more not necessarily consecutive observations

Columns (5)–(8) in Table 3 contain the results for the samples with 20 or more not necessarily consecutive observations. Relative to the results in columns (1)–(4), the variances of persistent shocks are somewhat smaller, while the variances of transitory shocks are similar in magnitude. Importantly, we still observe that estimations using the moments in differences deliver relatively higher estimates of the variance of permanent shocks, while estimations in levels deliver relatively higher estimates of the variance of transitory shocks, once again confirming the widely documented discrepancy.

3.3.3 Balanced samples

Estimation results based on the balanced samples are reported in Table 4. The use of balanced samples results in at least a 50% reduction of the variance of permanent shocks when using the moments in differences—compare, e.g., columns (6) and (8) in Table 3, and columns (2) and (4) in Table 4. There is a similarly striking reduction of at least 50% in the variance of transitory shocks when using the moments in levels—see columns (1) and (3), and (5) and (7) in Table 3, and columns (1) and (3) in Table 4. It appears that the use of balanced samples eliminates the discrepancy between the estimates of the earnings process in levels and differences.

3.4 A Closer Look at Unbalanced Samples

The results of estimation on balanced and unbalanced samples indicate that the discrepancy between the estimates based on the moments in levels and differences is specific to unbalanced samples. One possible explanation for this finding is that individuals with shorter earnings spells are intrinsically different, and that while permanent/transitory decomposition in equation (1) is appropriate for workers in the balanced sample, it provides a fundamentally misspecified model of the earnings processes for individuals in the unbalanced samples. Alternatively, it is possible that the decomposition is essentially valid but individuals in unbalanced panels either have higher shock variances or represent a selection of workers who experienced a sufficiently unfavorable earnings "shocks" that pushed them out of employment. One consequence of such selection is that the earnings surrounding the missing observations are likely to belong to workers in transit into or out of employment, with a potentially large impact on earnings in those periods. As discussed in Section 2 this can induce the difference in the estimates of the earnings process in growth rates or levels.

To explore the latter possibility, we estimate panel regressions of residual earnings on dummies for the first and last year an individual is observed in the sample. The results are shown in Table 5. Columns (1)–(4) explore the effects in the consecutive unbalanced samples, while columns (5)–(8) do the same for the non-consecutive samples. In the odd-numbered columns of the table, we focus on individuals whose earnings spells begin or end in the interior of the sample period. Consequently, the dummies "Year observed: first"−"Year observed: third" equal one if an individual's first earnings record in the sample occurs later than 1984 in German data and 1981 in Danish data, and zero otherwise, while the dummies "Year observed: second-to-last"−"Year observed: last" equal one if an individual's last earnings record is prior to 2008 in German data and 2006 in Danish data, and zero otherwise. As a useful comparison, we also focus on individuals whose earnings spells begin in the first or end in the last sample year. For the vast majority of these individuals such cutoffs do not represent an actual start or end of the earnings spell; instead, the sample window mechanically truncates their earnings spells in progress. Accordingly, the even-numbered columns of the table equate the first (last) few dummies to one if an individual's fist (last) earnings record is in the first (last) sample year, and to zero otherwise.

In all samples and both datasets, earnings are more than 0.60 log points lower than an individual's average in the first record of the spell while the last earnings record is below an individual's average by about 0.40 to 0.50 log points—columns (1), (3), (5), and (7). Earnings are still lower in the first and second years following the year of the first earnings record, reverting slightly faster to the average in German data; earnings are also lower in the two years preceding the last earnings record, with more pronounced effects in Danish data. Not surprisingly, individuals' earnings residuals in the first (last) few years do not substantially deviate from the unconditional average of zero if their first (last) earnings records occur in the first sample year—columns (2), (4), (6), and (8). Earnings are, on average, lower in the years preceding and following a missing earnings record in the non-consecutive samples—see columns (5)–(8). Clearly, the "shock" in the first year of an individual's spell is transitory, but somewhat persistent.[14] Interestingly, the dummies for the few first and last earnings records within a spell explain 5 to 8 percent of the variation in residual earnings. This number is quite high taking into account that a variety of observable factors normally explain about 30 percent of variation in earnings. As expected, the dummies do not explain any variation in residual earnings if individuals' first earnings records are in the first sample year and last earnings records in the last sample year—columns (2) and (4).

In Table 6, we explore the volatility of idiosyncratic earnings. The size of squared earnings is mechanically higher in the few first and last earnings records since residual earnings are more negative, on average, in those periods, as we illustrated in Table 5. To remove the influence of more negative residual earnings in those periods, we first group individual residual earnings observations by the values of the dummies (e.g., "Year observed: first," etc.) and year. We then calculate the means within these groups, remove the means from residual earnings, and square the result.[15] In German data, the mean of squared residual earnings, calculated this way, is about 0.15, while the standard deviation is about 0.36 in the consecutive and 0.39 in the non-consecutive samples; in Danish data, the mean is 0.11 in both samples and the standard deviation is 0.30 and 0.32 in the consecutive and non-consecutive samples, respectively. The results imply that earnings are not only lower on average in the (few) first and (few) last years of individual spells but are also more volatile. As an example, in German data, the mean of squared residual earnings is about 0.28 higher than the average in the first year which is about 190% larger than the typical size measured by the mean of squared residual earnings in the sample (we use the estimated coefficient in column (1) on the dummy "Year observed: first" and the mean squared residual of 0.15 for the German data to calculate this number). In the German consecu-

[14]If the shock were permanent, it would elevate earnings in all periods, with no distinguishable differences of the first earnings record relative to the individual's average. We cannot make the same conclusion for the last period as one would need a history of earnings after an individual's earnings spell is interrupted to argue if the shock is temporary or persistent.

[15]Similar results are obtained if we remove the negative mean effects of earnings around missing records by taking the residuals from the regressions of columns (1), (3), (5), and (7).

tive sample, about 23% of individuals have their first earnings record after 1984, the first calendar year of the sample, and about 31% of individuals have their last record before 2008, the last year of the sample—see Table 2. The same numbers for Danish data are 18% and 22%, respectively—see Table 1. This is a non-trivial number of individuals with pronounced differences in the level and volatility of residual earnings in the few first and last periods of earnings spells. In the non-consecutive samples, earnings in the periods preceding and following interior missing earnings records are also highly volatile—see columns (5)–(8). In German data, for instance, the volatility of earnings observations one year before a missing record is about 150% larger than the volatility of typical earnings observations.[16] Within shorter spells of the non-consecutive samples, the fraction of missing earnings records can be as high as 5% in German data and 14% in Danish data—see Tables 1 and 2. Columns (2), (4), (6), and (8) provide a contrast to the odd-numbered columns. In both datasets, the volatility of residual earnings is lower in the few first sample years, relative to the mean volatility, and higher in the last few sample years. This captures an increasing profile of earnings inequality over the life cycle—as the first earnings records are more likely to happen early in the life cycle, while the last earnings records are more likely to happen late in the life cycle—which is more pronounced in the German data. Note also that while the volatility of earnings, say, in the third spell year, for those spells with the first earnings record different from the fist sample year, is seemingly small—extra 0.04 points relative to the average volatility for German data—it is about 0.11 points larger relative to the volatility in the third year for those spells which started in the first sample year (see columns (1) and (2)). Similarly, the volatility of earnings in the last year of incomplete spells ending earlier than the last sample year is 0.25 points higher relative to the volatility of spells ending in the last sample year in both datasets—see columns (1)–(4).

Missing observations in applications using small unbalanced panels such as the PSID are typically treated as random. We can explore whether this is the case in our German and Danish samples with 20 or more not necessarily consecutive observations. In Table 7, the dependent variable is a dummy that equals 100 if individual earnings is missing, 0 otherwise. The predictive power of observables—earnings growth rates before and after missing earnings records, together with education dummies and age—on the incidence of missing earnings is small, in

[16] The number is calculated as $100 \times (0.16/0.15)$—the extra estimated effect on the dummy "1 year before earnings missing" in column (5), relative to the average squared residual of 0.15 for the sample of column (5), divided by the average squared residual for the sample.

line with Fitzgerald, Gottschalk, and Moffitt (1998) who made a similar observation using PSID data. Importantly, missing observations are associated with positive earnings growth in the periods following a missing record and negative earnings growth in the periods preceding a missing earnings record implying that individual realizations of residual earnings in the few first and last spell periods, as well as in the few periods preceding and following missing earnings records do not appear to be random draws from the earnings distribution. As pointed out by Moffitt and Gottschalk (2012), little is known about the effect of attrition on the autocovariance function of earnings and, therefore, on the estimates of the earnings process. Our results indicate that the effect can be large. In particular, lower and more volatile earnings observations surrounding missing earnings distort the autocovariance function for growth rates and levels such that the estimated variance of permanent shocks is substantially upward-biased when using the moments for growth rates, while the estimated variance of transitory shocks is substantially upward-biased when relying on the moments in levels.

4 Quantitative Evaluation of the Mechanism

4.1 Direct Evaluation of the Biases using the Permanent-Transitory Decomposition Moments

To evaluate the contribution of outlying observations to the estimated variances of permanent and transitory shocks in levels and differences, we calculate the variances in accordance with (L1)–(L2) and (D1)–(D2). As an example, we calculate an estimate of the permanent variance at time t using the identifying moment in levels (L1) as

$$\sigma^2_{\xi,l,t} = \frac{\sum_i y_{i,t}y_{i,t+1}}{\sum_i I^i_{t,t+1}} + \frac{\sum_i y_{i,t-2}y_{i,t-1}}{\sum_i I^i_{t-2,t-1}} - \frac{\sum_i y_{i,t+1}y_{i,t-1}}{\sum_i I^i_{t-1,t+1}} - \frac{\sum_i y_{i,t}y_{i,t-2}}{\sum_i I^i_{t-2,t}}, \qquad (2)$$

where the subscript l indicates that we are estimating the variance using information on log-earnings in levels, and $I^i_{t,t'}$ is an indicator function taking the value of one if individual earnings observations are nonmissing in both years t and t', and zero otherwise. Note that individual i will not contribute to the estimated variance of the permanent shock at time t only if all of the earnings cross-products for that individual—$y_{it}y_{it+1}$, $y_{it-2}y_{it-1}$, $y_{it+1}y_{it-1}$, and $y_{it}y_{it-2}$—are missing.

Let I^m_{it} be an indicator function that equals one at the times $t = t_m$ when individual i's earnings residual is missing and the years t_{m+j} surrounding it ($j = \pm1, \pm2, \pm3$), and zero in all other periods $t \neq t_m$ and $t \neq t_{m+j}$. We calculate the variance of permanent shocks due to outlying observations surrounding the missing earnings records, $\sigma^2_{\xi,l,o,t}$, as

$$\begin{aligned}\sigma^2_{\xi,l,o,t} = &\frac{\sum_i y_{i,t}y_{i,t+1}I^m_{it}}{\sum_i I^i_{t,t+1}I^m_{it}} + \frac{\sum_i y_{i,t-2}y_{i,t-1}I^m_{it}}{\sum_i I^i_{t-2,t-1}I^m_{it}} \\ &- \frac{\sum_i y_{i,t+1}y_{i,t-1}I^m_{it}}{\sum_i I^i_{t-1,t+1}I^m_{it}} - \frac{\sum_i y_{i,t}y_{i,t-2}I^m_{it}}{\sum_i I^i_{t-2,t}I^m_{it}}.\end{aligned} \tag{3}$$

An estimate of the permanent variance in levels, net of the effects of outliers, $\sigma^2_{\xi,l,n,t}$, can then be calculated as

$$\begin{aligned}\sigma^2_{\xi,l,n,t} = &\frac{\sum_i y_{i,t}y_{i,t+1}(1 - I^m_{it})}{\sum_i I^i_{t,t+1}(1 - I^m_{it})} + \frac{\sum_i y_{i,t-2}y_{i,t-1}(1 - I^m_{it})}{\sum_i I^i_{t-2,t-1}(1 - I^m_{it})} \\ &- \frac{\sum_i y_{i,t+1}y_{i,t-1}(1 - I^m_{it})}{\sum_i I^i_{t-1,t+1}(1 - I^m_{it})} - \frac{\sum_i y_{i,t}y_{i,t-2}(1 - I^m_{it})}{\sum_i I^i_{t-2,t}(1 - I^m_{it})}.\end{aligned} \tag{4}$$

We can similarly define the variances of permanent and transitory shocks in levels and differences for the consecutive unbalanced panels—e.g., the permanent variance utilizing all sample information ($\sigma^2_{\xi,l,t}$ for levels and $\sigma^2_{\xi,d,t}$ for differences), the permanent variance due to outlying observations in the first and last few periods of an individual's earnings spell ($\sigma^2_{\xi,l,o,t}$ for levels and $\sigma^2_{\xi,d,o,t}$ for differences), and the permanent variance net of outlying effects ($\sigma^2_{\xi,l,n,t}$ for levels and $\sigma^2_{\xi,d,n,t}$ for differences).

We present the estimates of those variances, averaged across all sample years, for both datasets in Table 8. For German data, in the consecutive sample, the estimates of the variance of permanent shocks in levels and differences using all sample information are 0.013 and 0.024, respectively.[17] When we drop outliers, the estimated net variances are $\hat{\sigma}^2_{\xi,l,n} = 0.010$ in levels and $\hat{\sigma}^2_{\xi,d,n} = 0.010$ in differences. The unadjusted variances of transitory shocks in levels and differences are estimated at 0.020 and 0.008, respectively, while the variances net of outliers in levels and differences are both estimated at 0.007. The results for Danish data are qualitatively similar. Clearly, the discrepancy between the estimates of

[17]The estimates deviate from the values in Table 3 because we do not impose the exact permanent-transitory decomposition on the data in the minimum-distance estimation of Table 3; the difference between the estimated variance of permanent shocks in levels and differences is not as drastic as in Table 3 because the estimated persistence of the permanent shocks in levels is estimated to be lower than in differences in the minimum-distance estimation.

permanent and transitory shocks in levels and differences is virtually eliminated when netting out the effects of outlying observations on the estimated variances.

For German data, in the non-consecutive sample, the variances of permanent shocks are $\sigma^2_{\xi,l} = 0.0096$, $\sigma^2_{\xi,l,n} = 0.0097$, $\sigma^2_{\xi,d} = 0.018$, $\sigma^2_{\xi,d,n} = 0.0097$, while the variances of transitory shocks are $\sigma^2_{\epsilon,l} = 0.018$, $\sigma^2_{\epsilon,l,n} = 0.007$, $\sigma^2_{\epsilon,d} = 0.007$, $\sigma^2_{\epsilon,d,n} = 0.007$. Netting out the influence of missing observations and the influence of the first and last records in the earnings spells eliminates most of the discrepancy between the variances of permanent and transitory shocks in differences and levels.

4.2 Restricting Unbalanced Samples

In the previous section, we found that the earnings of a few first and last observations, as well as a few observations before and after missing earnings records, are likely to be substantially lower than an individual's average, and are more volatile. In Table 9, we repeat our analysis of Table 3, dropping the first 3 observations for individuals whose earnings spells start later than 1984 in German data and later than 1981 in Danish data, and and last 3 observations for individuals whose earnings spells end earlier than 2008 in German data and 2006 in Danish data, as well as dropping 3 observations before and after a missing earnings record in the non-consecutive samples.

For the sample with 9 or more consecutive observations this barely affects the persistence of permanent shocks, while the variance of transitory (permanent) shocks is substantially reduced in estimations utilizing the moments in levels (growth rates). In German data, the variance of permanent shocks estimated using the moments in differences is reduced by about 70%, while the variance of transitory shocks estimated using the moments in levels is reduced by about 50%. Similar reductions can be seen in the Danish sample.

Dropping the first 3 and last 3 observations, as well as 3 earnings records before and after a year of missing earnings within an individual's earnings spell has a similar effect on the estimated earnings process in the non-consecutive samples—see the results in columns (5)–(8) of Table 9: the variance of the permanent (transitory) shock is reduced substantially in estimations using growth rates (levels). As a result, the estimated earnings process is virtually identical in estimations utilizing the moments for growth rates and levels in Table 9, columns (5)–(6) and (7)–(8) for German and Danish data, respectively.

A comparison between Tables 3 and 9 also indicates, consistently with the analysis in Section 2, that in both datasets the variance of the permanent component is more robustly estimated using the moments in levels while the variance of the transitory component, exclusive of the transitory variation in earnings due to rare shocks, is more robustly estimated using the moments in differences.

4.3 Simulation

Finally, we present a suggestive simulation, consistent with German data, aimed at replicating the results for the consecutive and non-consecutive samples presented above. We replicate our German unbalanced samples with 9 or more consecutive observations and 20 or not necessarily consecutive observations in terms of the number of person-year observations, and assume, consistently with Tables 5 and 6, that incomes in the spells starting (ending) in the years other than the first (last) year of the sample are, in addition, affected by a transitory shock, which has a negative mean and high variance.

For the consecutive sample, we assume that persistence of the permanent component equals 0.988, the variance of permanent shocks is 0.0056, persistence of the transitory component is 0.250, the variance of transitory shocks is 0.010, and the variance of fixed effects is 0.02. These are the estimates of the transitory component using the moments in growth rates and permanent component using the moments in levels in Table 9. We assume that the shocks and fixed effects are drawn from Student t-distributions with four degrees of freedom as our samples have high excess kurtosis.[18] We take the means and variances of the rare shocks in the first and last 3 periods from columns (1) and (2) of Table 5 and 6, and assume that they are independent and normally distributed.[19] The results, averaged across 100 simulations, are in Table 10. Utilizing the full sample results in overestimation of the variance of the permanent (transitory) shock in differences (levels), and it appears that the permanent component is more robustly estimated utilizing the moments in levels while the transitory component is closer to the truth utilizing the moments in differences. Interestingly, our full-sample estimation results are similar to the data results in Table 3. Dropping the first and last 3 observations in an individual's spell aligns the results in levels and

[18]Battistin, Blundell, and Lewbel (2009) document the departure of log-income from normality using survey data from the PSID. Assuming normal shocks instead has no impact on our findings.

[19]We assume a normal distribution rather than a t-distribution for rare shocks since the choice of a t-distribution with high variance and low degrees of freedom would sometimes result in large residual draws uncharacteristic of our empirical distribution of residuals.

differences—see columns (3) and (4) and correctly recovers the parameters of the underlying earnings process.

For the non-consecutive sample, we assume that persistence of the permanent component is 0.996, the variance of permanent shocks is 0.0040, persistence of the transitory component is 0.267, the variance of transitory shocks is 0.009, and the variance of fixed effects is 0.02. This is in line with the estimated permanent component in column (5) and transitory component in column (6) of Table 9. We take the means and variances of rare shocks in the first and last 3 periods, and three years before and after missing earnings records, from columns (3) and (6) of Table 5 and 6, and assume that the shocks are independent. The reported results are averages across 100 simulations. The full-sample estimation results in estimates close to the data estimates in Table 3, and recovers fairly well the permanent component using the moments in levels, and the transitory component using the moments in differences—columns (5) and (6) of Table 10. Dropping the first and last 3 observations in an individual's spell, as well as observations surrounding interior missing records, once again aligns the results in levels and differences as is confirmed in columns (7) and (8).

5 Implications for a Life-Cycle Model of Consumption with Incomplete Insurance Markets

In this section we study whether the presence of additional transitory earnings deviations in the first and last observations of earnings spells, and next to the interior missing observations, affects the estimates of the insurance coefficients against permanent and transitory shocks proposed in Blundell, Pistaferri, and Preston (2008); BPP hereafter. These estimates represent state-of-the-art measures of insurance available to households in the data, providing the key benchmark for assessing the performance of incomplete markets models. Clearly, getting the amount of insurance available to the agents correctly is key for many substantive implications of these models and for assessing, e.g., welfare implications of various economic shocks and policies.

5.1 Insurance Coefficients

We first introduce the insurance coefficients. In the standard consumption-savings model, if the Euler equation holds at equality, consumption growth,

Δc_{it}, can be expressed as

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it},$$

where $1 - \phi_t$ is the amount of insurance of permanent shocks and $1 - \psi_t$ is the amount of insurance of transitory shocks available at time t. Blundell, Pistaferri, and Preston (2008) show that the insurance coefficients for permanent and transitory shocks, in the case of a serially uncorrelated transitory component, can be recovered using the following data moments:

$$\begin{aligned} &\text{Permanent insurance:} \\ &1 - \phi_t = 1 - \frac{E[\Delta c_{it} \Delta y_{it-1}] + E[\Delta c_{it} \Delta y_{it}] + E[\Delta c_{it} \Delta y_{it+1}]}{E[\Delta y_{it} \Delta y_{it-1}] + E[\Delta y_{it} \Delta y_{it}] + E[\Delta y_{it} \Delta y_{it+1}]}, \qquad (5) \\ &\text{Transitory insurance:} \\ &1 - \psi_t = 1 - \frac{E\left[\Delta c_{it} \Delta y_{it+1}\right]}{E\left[\Delta y_{it} \Delta y_{it+1}\right]}. \qquad (6) \end{aligned}$$

where the expectation (averaging) is taken over all individuals used for estimation of each particular moment in the equations. Since the sample sizes utilized in the literature are typically small, and the above identifying moments may be imprecise in small samples, it is common to rely on a minimum-distance procedure for estimation of the model parameters, which utilizes all of the available autocovariance moments in the data. This is, e.g., the route taken in Blundell, Pistaferri, and Preston (2008).

5.2 The Biases in Estimating Insurance Coefficients Due to Presence of Rare Shocks

Since the denominator in equations (5)–(6) utilizes information on earnings data only, we can use our results on biases in the estimated variances of permanent and transitory shocks to earnings from Section 3 to find the biases in the estimated insurance coefficients.

Consider first an unbalanced sample with consecutive earnings observations such that part of the sample is comprised of individuals who start their incomplete earnings spells at t, different from the first sample year, while the rest of individuals have nonmissing earnings and consumption data throughout the whole sample period. Notice that the denominator of equation (5) is equal to the identifying moment (D1), and will therefore result in an estimate of $\sigma^2_{\xi,t+1} + s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)$, as discussed above. If consumption reacts to the current shocks only (which will be the case when an intertemporal shift of resources is allowed to the extent desired by a household), none of the moments in the numerator of equation (5) will be affected such that the bias in the estimated permanent insurance at $t+1$ will equal $\left(1 - \frac{\phi_{t+1}\sigma^2_{\xi,t+1}}{\sigma^2_{\xi,t+1}+s_{t,t+1}(\mu^2_\nu+\sigma^2_\nu)}\right) - (1-\phi_{t+1}) = \lambda_{t+1}\phi_{t+1}$, where $\lambda_{t+1} = \frac{s_{t,t+1}(\mu^2_\nu+\sigma^2_\nu)}{s_{t,t+1}(\mu^2_\nu+\sigma^2_\nu)+\sigma^2_{\xi,t+1}}$, and μ_ν and σ^2_ν are the mean and variance of the rare shock, respectively, and $s_{t,t+1}$ is the share of individuals who have nonmissing earnings and consumption records at t and $t+1$ and started their earnings spells at time t in the total number of individuals who have nonmissing earnings records at both times t and $t+1$.

Consider next an unbalanced sample with consecutive earnings observations such that part of the sample is comprised by individuals who end their incomplete earnings spells at t, different from the last sample year, while the rest of individuals have nonmissing earnings and consumption data throughout the whole sample period. In the following, we will assume that the insurance of the rare shock ν_{it} is the same as the insurance of the transitory shock ϵ_{it}, and equals ψ_t at time t. In this case, the denominator of (5) will equal $\sigma^2_{\xi,t} + s_{t-1,t}(\mu^2_\nu + \sigma^2_\nu)$, as discussed above. Since the rare shock is assumed to occur at t and consumption reacts to the current shocks only, the moment $E[\Delta c_{it}\Delta y_{it-1}]$ equals zero, while the moment $E[\Delta c_{it}\Delta y_{it+1}]$ will be identified by averaging over the sample of individuals who have complete earnings spells, and will equal $-\psi_t\sigma^2_{\epsilon_t}$. The moment $E[\Delta c_{it}\Delta y_{it}]$ will, however, be affected by incomplete earnings spells—averaging over all individuals observed at times $t-1$ and t, the moment will be estimated as $\phi_t\sigma^2_{\xi,t} + \psi_t\sigma^2_{\epsilon,t} + s_{t-1,t}\psi_t(\mu^2_\nu + \sigma^2_\mu)$, where $s_{t-1,t}$ is the share of individuals with incomplete earnings spells in the total sample of individuals observed at times $t-1$ and t. Summing up, the bias in the estimated permanent insurance in this case will equal $\left(1 - \frac{\phi_t\sigma^2_{\xi,t}+s_{t-1,t}\psi_t(\mu^2_\nu+\sigma^2_\mu)}{\sigma^2_{\xi,t}+s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)}\right) - (1-\phi_t) = (\phi_t - \psi_t)\lambda_t$, where $\lambda_t = \frac{s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)}{s_{t-1,t}(\mu^2_\nu+\sigma^2_\nu)+\sigma^2_{\xi,t}}$. The bias is unambiguously positive and potentially large if $\phi \gg \psi$ (which is true for the self-insurance life-cycle model of consumption), and λ is large (in case the mean and/or the variance of the rare shock are large relative to the variance of permanent shocks).

Consider now a sample which comprises individuals with missing earnings records at time t, in the interior of the sample period, and individuals with nonmissing earnings and consumption records throughout the sample period. Individuals with missing earnings records at time t will bias the estimated permanent insurance at times $t+2$ and $t-1$—the biases, respectively, are $\phi_{t+2}\lambda_{t+2}$ and $(\phi_{t-1} - \psi_{t-1})\lambda_{t-1}$ with properly defined λ's.

It can be further shown that the transitory insurance estimated using equations (5)–(6) is not systematically biased in our setup both for consecutive and not necessarily consecutive samples.

5.3 Application to PSID Male Earnings Data

In this Section, we examine if the mechanism described above for administrative data is operational in PSID male earnings data. We will first explore whether earnings residuals differ in the few periods around missing observations, then assess if dropping those observations helps aligning the variances of permanent and transitory shocks in levels and differences, and, lastly, will document the biases in the insurance coefficients for the shocks to male earnings due to outlying earnings observations. We will rely on the male earnings and household consumption data from Blundell, Pistaferri, and Preston (2008) and from which we drop (just a few) top-coded male earnings observations.[20]

5.3.1 Growth rates vs. levels and outlying observations

Table 11 shows the results of a regression of male earnings residuals on the dummies for the first and last observations, and observations surrounding missing records. The data span the reporting period 1979–1993, and earnings recorded in, say, year 1979 reflect male earnings received in 1978. In columns (1) and (3), the dummy "Year observed: first" equals 1 if an individual's first earnings record is after 1979, the first sample year, while the dummy "Year observed: last" equals 1 if an individual's last earnings record is prior to 1993, the last sample year. For comparison, in columns (2) and (4), the dummy "Year observed: first" equals 1 if an individual's first earnings record is in 1979, the first sample year, while the dummy "Year observed: last" equals 1 if an individual's last earnings record is in 1993, the last sample year. Earnings residuals are about 0.10 log points lower in the few first and last periods (if they differ from the first and last sample years) but are substantially lower in a few periods right after male earnings are missing—column (1). In contrast, earnings residuals are not different from the unconditional mean of zero in the few first (if the first record is in the first sample

[20] Although the main focus of Blundell, Pistaferri, and Preston (2008) is on the insurance against permanent and transitory shocks to family disposable income, they also document the estimates of insurance against the shocks to male earnings, which is of interest in itself. As our focus in this paper is on earnings, we restrict attention to insurance against shocks to earnings in this paper. In Hryshko and Manovskii (2015) we extend this analysis to study the insurance of the shocks to family disposable income.

year) and last periods (if the last record is in the last sample year)—column (2). In columns (3) and (4), we net out the mean effects of outlying observations on the residuals, and then regress squared (net) residuals on the same dummies as in columns (1) and (2), respectively. Squared residuals are lower in the few first and higher in the last sample years due to the well-known increase in male earnings inequality over the life cycle—column (4). The volatility of earnings, however, is much higher in the first and last sample years if individuals' first earnings records are not in the first sample year and last earnings records are not in the last sample year, as can be seen by comparing the first six regressors in columns (3) and (4) of Table 11.[21]

Earnings residuals around missing earnings records are not only lower on average but also more volatile than typical residuals—columns (3) and (4). To sum up, the patterns for the mean and variance effects of earnings residuals around missing records in the PSID data are qualitatively similar to those in Danish and German administrative data.

We further estimated the male earnings process using the moments in levels and differences. Since the sample is small, with about 1750 individuals, we restricted the persistence of the permanent component to unity; we also modeled the transitory component as an MA(1) component, both in agreement with Blundell, Pistaferri, and Preston (2008). As consumption data are believed to help identifying earnings shocks, we used both consumption and earnings growth moments, and the covariance between earnings and consumption growth to recover the variance of permanent and transitory shocks to male earnings when using the moments in differences, as in Blundell, Pistaferri, and Preston (2008).[22] Following Blundell, Pistaferri, and Preston (2008), we used a diagonal weighting matrix for weighting the moments in estimation. The variance of permanent shocks estimated using the moments in levels is about 0.017, and about 0.071 using the moments in growth rates. The variance of transitory shocks using the

[21]The coefficients on the dummies measure the variances in the respective periods relative to the average variance in the sample overall measured by a constant. For instance, the estimated constant in column (4) is 0.39 so that the variance of residual earnings in the first year for the earnings spells which start in the first sample year equals 0.21 (=0.39 − 0.18), while the estimated constant in column (3) equals 0.34 such that the volatility of earnings in the first year for earnings spells which start later than the first sample year equals 0.43. Similarly, the difference in the volatility of earnings residuals in the last year for the spells which end earlier than the last sample year and spells ending in the last sample year equals 0.34=(0.57 − 0.18) − (0.39 − 0.34).

[22]Consistently with the results in Blundell, Pistaferri, and Preston (2008), the insurance coefficients for permanent and transitory shocks are assumed to be constant over time and life cycle in estimations.

moments in levels is estimated at about 0.20, and 0.095 using the moments in differences. Next, we drop the first and last 3 earnings observations, if an individual's first record is after 1979 and the last record is prior to 1993, as well as the three earnings observations before and after missing earnings records. In this sample, the average variance of permanent shocks, using the moments in levels, is estimated at about 0.015, barely changing relative to the estimate for the whole sample. The estimated variance of permanent shocks using the moments in differences is, however, substantially reduced from 0.071 to about 0.022. This would drastically reduce prediction about the increase in the variance of log male earnings over time and over the life cycle when using the estimated permanent variance recovered with the growth-rates moments (we will return to this issue shortly). After dropping outlying observations, the numbers for the variance of transitory shocks are 0.108 and 0.084, when using the moments in levels and growth rates, respectively. While the variance of transitory shocks using the moments in growth rates changed little after dropping outlying observations, the variance of transitory shocks using the moments in levels was cut in half. We next turn to the estimated insurance coefficients for permanent shocks to male earnings in PSID data.

5.3.2 Insurance of the shocks to male earnings

Using full sample, the estimated transmission coefficient for permanent shocks to male earnings is about 0.26 (with a standard error of 0.05), implying insurance of about 74%. The number is similar to the one reported in Blundell, Pistaferri, and Preston (2008). After dropping the outlying observations, the transmission coefficient for permanent shocks to male earnings rises to 0.64 (standard error 0.18), implying a much smaller insurance of permanent shocks, in line with the theoretical bias outlined above.

While the preceding estimate was based on dropping a fraction of earnings observations, it is also possible to estimate the variance of permanent and transitory shocks, as well as the variance of rare shocks, and the insurance against permanent, transitory shocks, and rare shocks, retaining all earnings observations. The consumption equation in this case is modified to $\Delta c_{it} = \phi\xi_{it} + \psi\epsilon_{it} + \psi_{\text{rare}}\nu_{it}$, where ν_{it} is an iid rare transitory shock. Besides the standard BPP consumption and income moments, we use in addition the regression coefficients in Table 11 to estimate the mean and variance effects of rare shocks, assuming the rare shocks are iid.[23] The estimated transmission coefficient for permanent shocks to male earnings is about 0.60 (with a standard error of 0.13), not far from the estimated coefficient using the sample with the outlying observations dropped.[24] The insurance against rare transitory shocks which are larger in magnitude than typical transitory shocks is estimated at about 92%, while the estimated insurance against typical transitory shocks binds at 100%—column (2). When we restrict the insurance against rare and typical standard shocks to be equal, it is estimated at about 93%, close to the estimate in the standard BPP model—column (3).

Lastly, Figure 1(a) shows the fit of the BPP model to the variance of log earnings residuals in levels and differences observed in the data, while Figure 1(b) shows the fit of the BPP model, which, in addition, allows for explicit estimation of the means and variances of outlying shocks. The implied variance of log earnings residuals in the last sample year (the short-dash line) is about twice as large as the variance in the data (the solid line) if we rely on BPP moments—the direct consequence of rather large estimates of the variances of permanent shocks recovered from the standard BPP estimation—but is fairly close to the variance in the data if we allow for rare shocks (the line with circles). Similarly, if we rely on income growth moments only, the implied variance of earnings in levels in 1992 is more than 2 times larger than in the data (the long-dash line). Both the standard BPP model and the model which allows for rare shocks provide a good fit to the variance of income growth rates, while estimation in levels substantially overpredicts the observed variances—Figure 1(b).

[23] Specifically, in addition to all of the moments in the original BPP estimation, we target the regression coefficients in two regressions, with residuals and (net) squared residuals on the left-hand side, and the 19 regressors on the right-hand side: six dummies around interior missing earnings observations, three dummies for the first earnings records if the incomplete earnings spells start later than the first sample year, three dummies for the first earnings records if spells start in the first sample year, three dummies for the last earnings records if the incomplete earnings spells end earlier than the last sample year, three dummies for the last earnings records if earnings spells end in the last sample year, and a constant. We estimated the model by the method of simulated minimum distance, assuming that permanent, transitory, and rare transitory shocks are drawn from normal distributions. We verified that the simulated method of moments with the assumption of normal permanent and transitory shocks delivers virtually the same parameter estimates as the standard BPP estimation (the results of which are reported in column (1)) which allows for any distributions of permanent and transitory shocks.

[24] Full estimation results for the standard BPP estimation and the estimation which takes account of rare shocks are given in Table 12.

6 Conclusion

Properties of the earnings process play an important role in various areas of macro and labor economics. Different specifications of this process have been explored in the literature, but the most widely used one is based on decomposing earnings into the sum of persistent and transitory components, where the persistent component is often assumed to follow a random walk. The parameters of such a process can be identified using the moments based on earnings growth rates (first-difference in log earnings) or the moments based on log earnings levels. Historically, the former approach is more common in labor economics, while the latter is more common in the macroeconomics literature. Unfortunately, these two approaches lead to dramatically different estimates of the variances of permanent and transitory components. In particular, using the same set of observations in the data, the variance of the persistent component is typically estimated to be much higher when the moments in growth rates are targeted, while the variance of the transitory component is found to be much higher when the estimation is based on fitting the moments in levels. This has important implications for substantive economic analysis. For example, the earnings process drives the heterogeneity in Bewley-type models with incomplete markets and the variances of earnings components determine not only economic choices, such as e.g., consumption, and savings but also the optimal design of policies, such as taxes and transfers. Moreover, the standard approach to estimating the amount of insurance that individuals have against permanent and transitory shocks in the data relies on the estimated variances of permanent and transitory components. The uncertainty over the size of these variances translates into uncertainty over the right amount of insurance generated by the widely used Bewley-type models, and the associated uncertainty about the results of welfare analysis using those models.

In this paper we uncover the feature of the data that can quantitatively account for the large difference in the estimates based on earnings growth rates and levels in the administrative data from Denmark and Germany. In particular, we found that earnings are lower on average and more volatile at the start and end of continuous earnings spells. We have shown theoretically that these "outlying" earnings observations, that are either preceded or are followed by a missing observation, induce an upward bias in the estimates of the variance of permanent shocks based on the moments in differences and of the variance of transitory shocks when estimation is based on the moments in levels. Thus, even when working with very large administrative datasets with highly reliable information, one must remain vigilant because such natural features of the datasets as low mean and high variance of earnings at the start and end of earnings spells can induce very large biases in the estimated earnings processes.

While the primary focus of this paper is on estimating earnings processes on large administrative datasets that that are becoming central in the literature, the mechanism we describe also applies to survey-based data on earnings. We illustrate the importance of accounting for the high variance and low mean of earnings at the start and end of the earnings spells by replicating the analysis in Blundell, Pistaferri, and Preston (2008) using their PSID male earnings data. We show that not taking these features of the data into account leads to significant biases in the estimated amount of insurance against permanent earnings shocks. In particular, we find a substantially lower extent of insurance available to individuals in the data.

These findings have several practical implications for estimation of the earnings process. To estimate the parameters of the earnings process in equation (1), one can follow several approaches. First, we have shown theoretically and verified empirically, that the variance of the transitory shock is estimated with no bias when estimation is based on the moments for earnings growth rates, and the variance of the permanent shock is unbiased when estimation fits the moments in levels. One could therefore use the estimated permanent component from the moments in levels and the estimated transitory component from the moments in growth rates. An alternative way to proceed would be to estimate the earnings process in equation (1) on the data that do not include the observations surrounding the missing ones. As we have shown, this recovers the true parameters of this process quite well. Finally, one can follow our approach in Section 5.3.2 and explicitly incorporate additional transitory shocks at the beginning and the end of contiguous earnings histories into the analysis—the mean and the variance of these shocks are readily identified from the mean and the variance of earnings in those periods.

References

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

Attanasio, O., E. Hurst, and L. Pistaferri (2012): "The Evolution of Income, Consumption, and Leisure Inequality in the US, 1980–2010," NBER Working Paper # 17982.

Banks, J., and P. Diamond (2010): "The Base for Direct Taxation," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by The Institute for Fiscal Studies (IFS), chap. 6, pp. 548–648. Oxford University Press for The Institute for Fiscal Studies, Oxford.

Battistin, E., R. Blundell, and A. Lewbel (2009): "Why Is Consumption More Log Normal than Income? Gilbrat's Law Revisited," *Journal of Political Economy*, 117, 1140–1154.

Blundell, R., M. Graber, and M. Mogstad (2015): "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family," *Journal of Public Economics*, In Press.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Browning, M., M. Ejrnæs, and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77, 1353–1381.

Carroll, C. D. (1997): "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, 112, 1–55.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111(4), 818–857.

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59, 1121–1248.

DeBacker, J., B. Heim, V. Panousi, S. Ramnath, and I. Vidangos (2013): "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns," *Brookings Papers on Economic Activity*, pp. 67–142.

Domeij, D., and M. Flodén (2010): "Inequality Trends in Sweden 1978–2004," *Review of Economic Dynamics*, 13(1), 179–208.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): "Revisiting the German Wage Structure," *Quarterly Journal of Economics*, 124, 843–881.

FARHI, E., AND I. WERNING (2012): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies, forthcoming.*

FITZGERALD, J., P. GOTTSCHALK, AND R. MOFFITT (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33, 251–299.

FRIEDMAN, M. (1957): *A Theory of the Consumption Function.* Princeton University Press, Princeton, NJ.

FRIEDMAN, M., AND S. KUZNETS (1954): *Income from Independent Professional Practice.* National Bureau of Economic Research, New York.

GOTTSCHALK, P., AND R. A. MOFFITT (1994): "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 2, 217–54.

GUVENEN, F. (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

GUVENEN, F., S. OZCAN, AND J. SONG (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

HARVEY, A. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

HEATHCOTE, J. (2005): "Fiscal Policy with Heterogeneous Agents and Incomplete Markets," *Review of Economic Studies*, 72(1), 161–188.

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 (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118(4), 681–722.

——— (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, 177–209.

HRYSHKO, D., AND I. MANOVSKII (2015): "Reassessing Consumption Responses to Income Shocks in U.S. Data," mimeo, University of Alberta and University of Pennsylvania.

HUBBARD, R. G., J. SKINNER, AND S. P. ZELDES (1995): "Precautionary Saving and Social Insurance," *Journal of Political Economy*, 103(2), 360–399.

Huggett, M., G. Ventura, and A. Yaron (2011): “Sources of Lifetime Inequality,” *American Economic Review*, 101, 2923–2954.

Krueger, D., and F. Perri (2006): “Does Income Inequality Lead to Consumption Inequality? Evidence and Theory,” *Review of Economic Studies*, 73(1), 163–193.

Livshits, I., J. MacGee, and M. Tertilt (2007): “Consumer Bankruptcy: A Fresh Start,” *American Economic Review*, 97, 402–418.

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.

——— (2007): “A Practitioner’s Approach to Estimating Intertemporal Relationships Using Longitudinal Data: Lessons from Applications in Wage Dynamics,” in *Handbook of Econometrics, Vol. 6A*, ed. by J. J. Heckman, and E. E. Leamer, chap. 62, pp. 4057–4167. Elsevier, Amsterdam.

Meghir, C., and L. Pistaferri (2004): “Income Variance Dynamics and Heterogeneity,” *Econometrica*, 72(1), 1–32.

——— (2011): “Earnings, Consumption, and Life Cycle Choices,” in *Handbook of Labor Economics, Vol. 4B*, ed. by D. Card, and O. Ashenfelter, chap. 9, pp. 773–854. Elsevier, Amsterdam.

Moffitt, R. A., and P. Gottschalk (2012): “Trends in the Transitory Variance of Male Earnings: Methods and Evidence,” *Journal of Human Resources*, 47(1), 204–236.

Postel-Vinay, F., and H. Turon (2010): “On-the-Job Search, Productivity Shocks, and the Individual Earnings Process,” *International Economic Review*, 51, 599–629.

Table 1: Danish data, 1981–2006. Summary statistics for selected years.

	9 consec.	20 not nec. consec.	Balanced
Number of individuals	102,825	90,668	67,008
Number of observations	2,367,552	2,298,429	1,742,208
Education			
Less than high school	0.227	0.222	0.206
High school degree	0.032	0.031	0.029
Vocational training	0.505	0.521	0.542
Two-year university degree	0.046	0.046	0.047
Bachelors degree	0.125	0.122	0.124
Master or Ph.D.	0.065	0.059	0.051
Earnings			
1985	40,157	40,227	41,383
	(12,831)	(12,889)	(12,278)
1995	48,197	48,444	50,004
	(20,562)	(20,462)	(19,954)
2005	52,656	51,511	53,298
	(26,635)	(26,279)	(25,917)
Spell counts			
Start 1981, end 2006	67,008	80,787	67008
Start after 1981, end 2006	13,439	4,376	0
Start in 1981, end before 2006	17,723	5,210	0
Start after 1981, end before 2006	4,655	295	0
Total	102,825	90,668	67008
Number of spells with 20 or more not nec. consec. observations, by length			
[Proportion of missing observations within spell in square brackets]			
20		1,634 [0.144]	
21		2,009 [0.119]	
22		2,665 [0.096]	
23		3,296 [0.079]	
24		4,486 [0.054]	
25		9,570 [0.030]	
26		67,008 [0.00]	

Notes: Earnings are expressed in 2005 Euros; the standard deviation of earnings is given in parentheses.

Table 2: German data, 1984–2008. Summary statistics for selected years.

	9 consec.	20 not nec. consec.	Balanced
Number of individuals	18,130	13,635	9,452
Number of observations	379,080	330,748	236,300
Education			
Middle school or no degree	0.05	0.04	0.04
Vocational training	0.72	0.74	0.76
High school degree	0.06	0.05	0.05
College	0.17	0.17	0.15
Earnings			
1985	33,626	33,930	34,559
	(15,876)	(13,323)	(12,881)
1995	45,309	47,180	47,965
	(24,702)	(24,295)	(24,463)
2005	49,121	51,289	52,457
	(36,473)	(37,106)	(37,666)
Spell counts			
Start 1984, end 2008	9,452	11,179	9,452
Start after 1984, end 2008	3,136	1,007	0
Start in 1984, end before 2008	4,463	1,393	0
Start after 1984, end before 2008	1,079	56	0
Total	18,130	13,635	9,452
Number of spells with 20 or more not nec. consec. observations, by length			
[Proportion of missing observations within spell in square brackets]			
20		575 [0.054]	
21		509 [0.054]	
22		623 [0.05]	
23		871 [0.037]	
24		1,605 [0.027]	
25		9,452 [0.00]	

Notes: Earnings are expressed in 2005 Euros; the standard deviation of earnings is given in parentheses.

Table 3: Estimates of the earnings process in unbalanced samples.

	9 consec.				20 not nec. consec.			
	German data		Danish data		German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.980	0.992	0.964	0.990	0.995	0.997	0.967	0.989
	(0.001)	(0.0008)	(0.0008)	(0.0004)	(0.001)	(0.001)	(0.0007)	(0.0006)
$\hat{\sigma}^2_\xi$	0.007	0.019	0.007	0.012	0.0046	0.008	0.0066	0.0103
	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
$\hat{\phi}_\tau$	0.173	0.173	0.289	0.285	0.158	0.316	0.184	0.355
	(0.006)	(0.014)	(0.003)	(0.004)	(0.009)	(0.012)	(0.004)	(0.004)
$\hat{\sigma}^2_\epsilon$	0.025	0.009	0.022	0.014	0.016	0.011	0.023	0.016
	(0.0004)	(0.0003)	(0.0002)	(0.0001)	(0.0003)	(0.0003)	(0.0002)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.026	—	0.020	—	0.029	—	0.023	—
	(0.002)	—	(0.0004)	—	(0.002)	—	(0.0004)	—
χ^2	929.67	725.21	6166.66	4196.83	1518.17	1284.62	6637.87	4799.07
(d.f.)	320	296	346	321	320	296	346	321

Notes: The estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \phi_\tau \tau_{it} + \epsilon_{it+1}$. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table 4: Estimates of the earnings process. Balanced samples.

	German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)
$\hat{\phi}_p$	1	0.998	0.975	0.979
	(0.001)	(0.002)	(0.0007)	(0.0009)
$\hat{\sigma}^2_\xi$	0.0031	0.0033	0.0046	0.0045
	(0.0001)	(0.0001)	(0.0000)	(0.0001)
$\hat{\phi}_\tau$	0.278	0.258	0.311	0.317
	(0.011)	(0.012)	(0.004)	(0.004)
$\hat{\sigma}^2_\epsilon$	0.008	0.0078	0.0104	0.0106
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.024	—	0.018	—
	(0.001)	—	(0.0003)	—
χ^2	1205.84	935.52	6244.18	5094.84
(d.f.)	320	296	346	321

Notes: The estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \phi_\tau \tau_{it} + \epsilon_{it+1}$. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table 5: Dependent variable: residual earnings. Panel regressions.

	9 or more consec.				20 not nec. consec.			
	German data		Danish data		German data		Danish data	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year observed: first	–0.75***	0.00	–0.67***	0.00**	–0.75***	0.00	–0.62***	–0.00***
	(–74.06)	(1.07)	(–144.86)	(2.19)	(–37.73)	(0.14)	(–74.14)	(–3.84)
Year observed: second	–0.30***	0.02***	–0.35***	0.02***	–0.24***	0.01***	–0.35***	0.01***
	(–43.44)	(8.34)	(–93.73)	(18.66)	(–17.86)	(5.42)	(–44.27)	(12.92)
Year observed: third	–0.26***	0.03***	–0.28***	0.03***	–0.19***	0.02***	–0.25***	0.01***
	(–39.23)	(11.10)	(–81.25)	(27.44)	(–14.95)	(10.22)	(–36.41)	(15.66)
Year observed: second-to-last	–0.10***	0.02***	–0.14***	0.01***	–0.08***	0.02***	–0.12***	0.01***
	(–15.82)	(7.50)	(–47.87)	(12.14)	(–6.77)	(6.60)	(–18.58)	(12.24)
Year observed: next-to-last	–0.13***	0.01***	–0.17***	0.01***	–0.13***	0.01***	–0.16***	0.01***
	(–19.37)	(5.01)	(–55.32)	(7.97)	(–9.39)	(4.54)	(–23.54)	(9.49)
Year observed: last	–0.50***	0.00	–0.38***	0.00**	–0.51***	–0.00	–0.37***	–0.00**
	(–56.23)	(1.04)	(–95.64)	(2.52)	(–27.71)	(–0.48)	(–43.71)	(–2.09)
3 years before earn. miss., dummy					–0.12***	–0.12***	–0.10***	–0.10***
					(–12.45)	(–12.06)	(–30.09)	(–29.34)
2 years before earn. miss., dummy					–0.12***	–0.12***	–0.12***	–0.12***
					(–13.25)	(–13.08)	(–38.71)	(–39.67)
1 year before earn. miss., dummy					–0.35***	–0.35***	–0.38***	–0.39***
					(–28.76)	(–28.86)	(–104.00)	(–106.92)
1 year after earn. miss., dummy					–0.47***	–0.48***	–0.55***	–0.55***
					(–34.10)	(–34.25)	(–136.97)	(–137.21)
2 years after earn. miss., dummy					–0.21***	–0.21***	–0.22***	–0.22***
					(–21.39)	(–21.00)	(–65.50)	(–64.95)
3 years after earn. miss., dummy					–0.21***	–0.21***	–0.19***	–0.19***
					(–20.08)	(–19.62)	(–51.05)	(–50.52)
Adj. R sq.	0.068	0.000	0.055	0.000	0.053	0.032	0.081	0.068
No. obs.	379,080	379,080	2,367,552	2,367,552	330,748	330,748	2,298,429	2,298,429
No. indiv.	18,130	18,130	102,825	102,825	13,635	13,635	90,668	90,668

Notes: German data span the period 1984–2008, while Danish data span the period 1981–2006. In columns (1), (3), (5), and (7), the dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is later than in 1984 in German data and 1981 in Danish data, zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is earlier than in 2008 in German data and 2006 in Danish data, zero otherwise. In columns (2), (4), (6), and (8), the dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is in 1984 in German data and 1981 in Danish data, zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is in 2008 in German data and 2006 in Danish data, zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 6: Dependent variable: squared residual earnings. Panel regressions.

	9 or more consec.				20 not nec. consec.			
	German data		Danish data		German data		Danish data	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year observed: first	0.28***	–0.05***	0.28***	–0.01***	0.26***	–0.02***	0.18***	0.01***
	(29.30)	(–14.42)	(65.60)	(–10.70)	(15.89)	(–6.73)	(27.39)	(8.41)
Year observed: second	0.06***	–0.07***	0.16***	–0.02***	0.04***	–0.04***	0.14***	–0.00***
	(7.71)	(–23.25)	(41.37)	(–22.08)	(2.89)	(–15.77)	(17.76)	(–4.51)
Year observed: third	0.04***	–0.07***	0.12***	–0.03***	0.01	–0.05***	0.08***	–0.01***
	(6.02)	(–26.84)	(31.97)	(–27.02)	(0.62)	(–20.22)	(10.99)	(–10.03)
Year observed: second-to-last	0.08***	0.02***	0.09***	0.00	0.07***	0.05***	0.11***	0.02***
	(10.52)	(7.22)	(27.15)	(0.67)	(5.54)	(15.96)	(12.90)	(17.16)
Year observed: next-to-last	0.10***	0.02***	0.12***	0.00	0.11***	0.05***	0.13***	0.02***
	(13.25)	(7.38)	(32.13)	(0.67)	(7.87)	(16.34)	(15.29)	(17.38)
Year observed: last	0.29***	0.04***	0.25***	0.00	0.32***	0.07***	0.26***	0.03***
	(30.14)	(11.92)	(58.60)	(0.30)	(17.04)	(20.09)	(27.09)	(20.98)
3 years before earn. miss., dummy					0.04***	0.04***	0.07***	0.07***
					(4.07)	(4.20)	(20.30)	(20.46)
2 years before earn. miss., dummy					0.03***	0.04***	0.07***	0.07***
					(3.49)	(4.08)	(21.82)	(22.59)
1 year before earn. miss., dummy					0.16***	0.17***	0.14***	0.15***
					(12.95)	(13.39)	(41.43)	(43.36)
1 year after earn. miss., dummy					0.23***	0.24***	0.19***	0.20***
					(15.58)	(15.99)	(50.20)	(51.53)
2 years after earn. miss., dummy					0.06***	0.06***	0.11***	0.11***
					(6.13)	(5.98)	(30.50)	(29.98)
3 years after earn. miss., dummy					0.07***	0.07***	0.07***	0.11***
					(7.15)	(6.63)	(20.30)	(27.49)
Adj. R sq.	0.015	0.017	0.012	0.020	0.003	0.000	0.048	0.074
No. obs.	379,080	379,080	2,367,552	2,367,552	330,748	330,748	2,298,429	2,298,429
No. indiv.	18,130	18,130	102,825	102,825	13,635	13,635	90,668	90,668

Notes: See notes to Table 5.

Table 7: Dependent variable: the incidence of missing earnings observation; panel regressions. 20 not nec. consec. observations.

	German data					Danish data				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Earn. growth ($t-4$ to $t-3$)					–0.43***					–0.58***
					(–3.34)					(–10.65)
Earn. growth ($t-3$ to $t-2$)				–1.02***	–1.16***				–1.06***	–1.23***
				(–6.49)	(–7.04)				(–16.34)	(–18.11)
Earn. growth ($t-2$ to $t-1$)	–3.21***		–3.10***	–3.32***	–3.41***	–3.15***		–3.01***	–3.25***	–3.33***
	(–12.24)		(–12.14)	(–12.18)	(–12.20)	(–32.02)		(–31.24)	(–32.23)	(–32.57)
Earn. growth ($t+1$ to $t+2$)		3.58***	3.49***	3.88***	4.03***		4.38***	4.28***	4.77***	4.98***
		(12.67)	(12.60)	(12.80)	(12.96)		(38.88)	(38.54)	(40.48)	(41.34)
Earn. growth ($t+2$ to $t+3$)				1.32***	1.63***				1.89***	2.29***
				(7.81)	(8.80)				(24.08)	(27.41)
Earn. growth ($t+3$ to $t+4$)					0.85***					1.27***
					(6.45)					(19.03)
Adj. R sq.	0.005	0.007	0.012	0.013	0.014	0.007	0.012	0.017	0.020	0.021
No. obs.	210641	210641	210641	210641	210641	1486308	1486308	1486308	1486308	1486308
No. indiv.	13635	13635	13635	13635	13635	90584	90584	90584	90584	90584

Notes: Education dummy variables and age are also included in the regressions. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 8: Variances of permanent and transitory shocks in the permanent-transitory decomposition of earnings.

	9 consec.				20 not nec. consec.			
	German data		Danish data		German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
Perm. var., full sample, $\hat{\sigma}^2_{\xi}$	0.013	0.024	0.016	0.019	0.0096	0.018	0.013	0.019
Perm. var., outliers, $\hat{\sigma}^2_{\xi,o}$	0.034	0.158	0.053	0.124	–0.009	0.137	–0.004	0.133
Perm. var., net of outliers, $\hat{\sigma}^2_{\xi,n}$	0.010	0.010	0.013	0.013	0.0097	0.0097	0.013	0.013
Trans. var., full sample, $\hat{\sigma}^2_{\epsilon}$	0.020	0.008	0.014	0.009	0.018	0.007	0.019	0.009
Trans. var., outliers, $\hat{\sigma}^2_{\epsilon,o}$	0.143	0.011	0.104	0.022	0.162	0.011	0.173	0.030
Trans. var., net of outliers, $\hat{\sigma}^2_{\epsilon,n}$	0.007	0.007	0.008	0.008	0.007	0.007	0.008	0.008

Notes: The variances are calculated as in equations (2)–(4).

Table 9: Estimates of the earnings process. Drop first and last, and before/after missing earnings records.

	9 or more consec.				20 not nec. consec.			
	German data		Danish data		German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.988	0.998	0.967	0.986	0.996	0.999	0.971	0.985
	(0.001)	(0.001)	(0.0008)	(0.0006)	(0.001)	(0.001)	(0.0008)	(0.0007)
$\hat{\sigma}^2_\xi$	0.0056	0.005	0.0062	0.0061	0.0041	0.0042	0.0053	0.0057
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\hat{\phi}_\tau$	0.305	0.250	0.338	0.313	0.317	0.267	0.339	0.318
	(0.010)	(0.009)	(0.004)	(0.004)	(0.01)	(0.011)	(0.004)	(0.004)
$\hat{\sigma}^2_\epsilon$	0.012	0.010	0.015	0.013	0.011	0.009	0.015	0.013
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.022	—	0.02	—	0.024	—	0.021	—
	(0.002)	—	(0.0004)	—	(0.002)	—	(0.0004)	—
χ^2	1093.77	923.01	5330.86	4844.48	1178.30	905.41	5919.64	5025.96
(d.f.)	320	296	346	321	320	296	346	321

Notes: German data span the period 1984–2008, while Danish data span the period 1981–2006. In German data, if an individual's first (last) earnings observation is not in 1984 (2008), his first (last) three observations are dropped prior to estimation. In Danish data, if an individual's first (last) earnings observation is not in 1981 (2006), his first (last) three observations are dropped prior to estimation. In columns (5)–(8), in addition, we drop three observations preceding and following a missing earnings record. The estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \phi_\tau \tau_{it} + \epsilon_{it+1}$. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses.

Table 10: Estimates of the earnings process in unbalanced samples. Simulated "German" data.

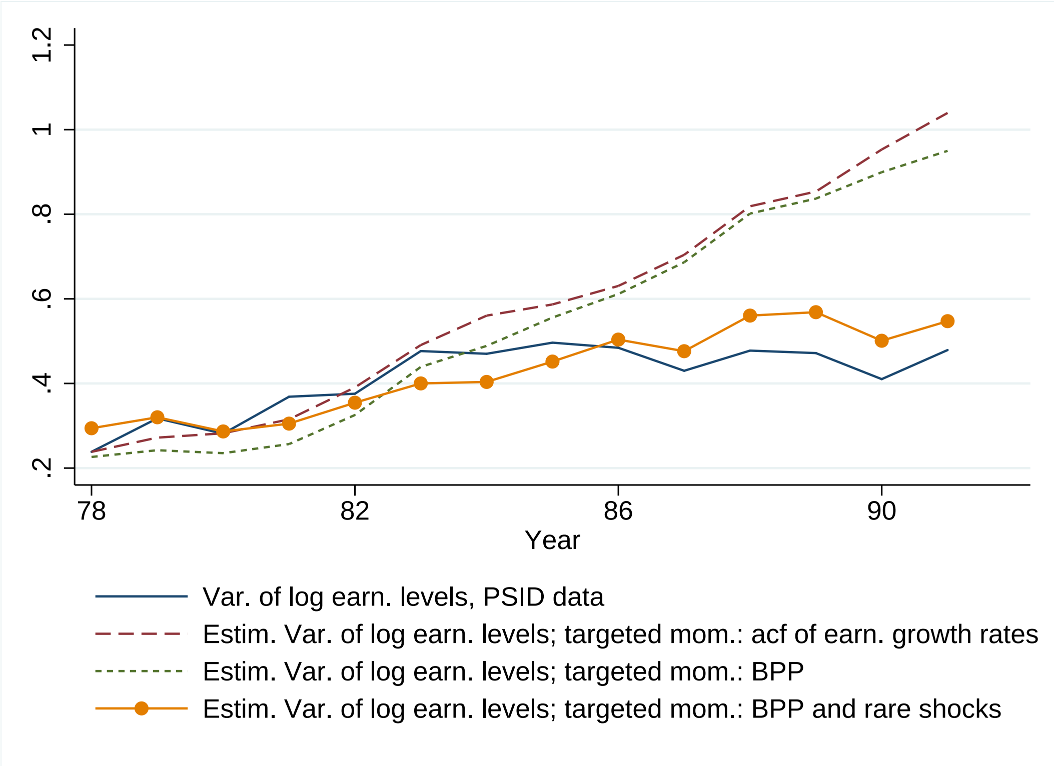
	9 consec.				20 not nec. consec.			
	Full sample		Drop		Full sample		Drop	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.988	0.992	0.988	0.988	0.993	0.990	0.996	0.996
	(0.001)	(0.001)	(0.0008)	(0.001)	(0.0008)	(0.001)	(0.0007)	(0.0009)
$\hat{\sigma}^2_\xi$	0.0054	0.014	0.0055	0.0055	0.0043	0.0083	0.004	0.004
	(0.0001)	(0.0009)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0001)
$\hat{\phi}_\tau$	0.236	0.155	0.250	0.250	0.193	0.203	0.267	0.266
	(0.007)	(0.02)	(0.005)	(0.005)	(0.009)	(0.009)	(0.005)	(0.005)
$\hat{\sigma}^2_\epsilon$	0.019	0.009	0.01	0.01	0.014	0.009	0.0087	0.0088
	(0.0008)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.02	—	0.02	—	0.018	—	0.02	—
	(0.002)	—	(0.001)	—	(0.001)	—	(0.001)	—
χ^2	3058.75	3139.04	324.92	302.79	1416.02	1343.71	331.77	308.51
(d.f.)	320	296	346	321	320	296	346	321

Notes: The true earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \phi_\tau \tau_{it} + \epsilon_{it+1}$. In columns (1)–(4), $\sigma^2_\alpha = 0.02$, $\phi_p = 0.988$, $\sigma^2_\xi = 0.0056$, $\phi_\tau = 0.250$, $\sigma^2_\epsilon = 0.01$, while in columns (5)–(8), $\sigma^2_\alpha = 0.02$, $\phi_p = 0.996$, $\sigma^2_\xi = 0.0040$, $\phi_\tau = 0.267$, $\sigma^2_\epsilon = 0.009$. In columns (3)–(4) the first 3 (last 3) observations are dropped if an individual's earnings spell starts (ends) later (earlier) than in 1984 (2008); in columns (7) and (8), in addition, three observations before and after missing earnings records are dropped. The results are the averages across 100 simulations. The model is estimated using the optimal weighting minimum distance method. Standard errors, calculated as the standard deviations of the estimates across simulations, are in parentheses.

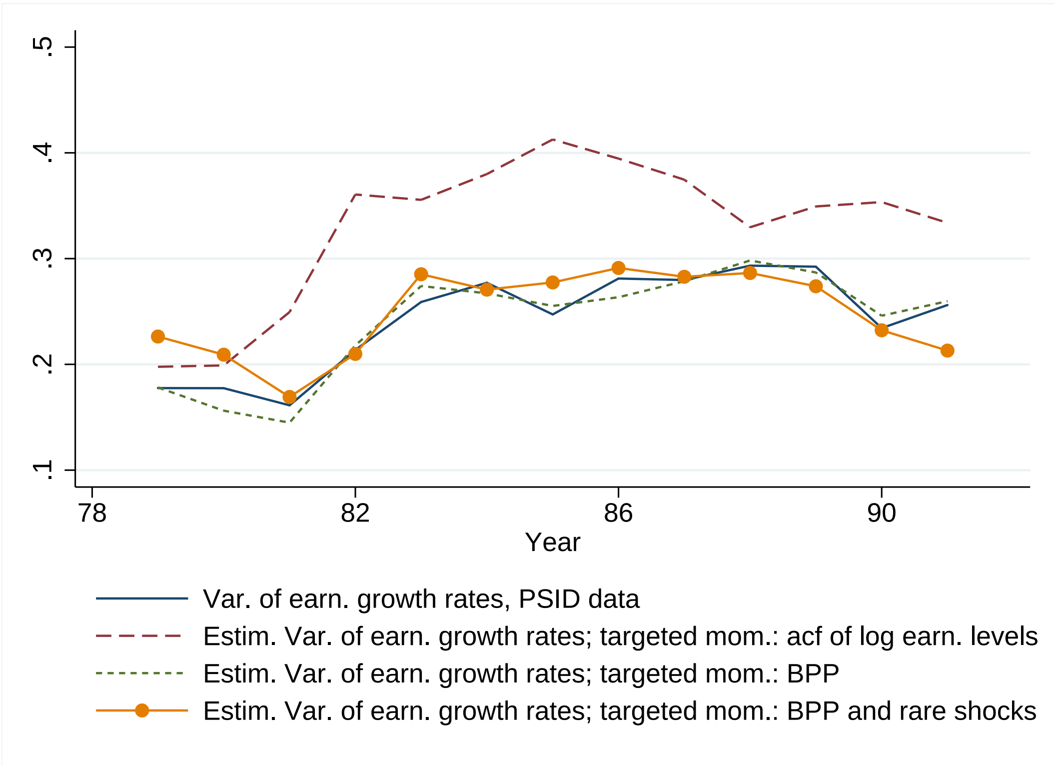
Table 11: Male Earnings Residuals. PSID data

Dependent variable	Residuals		Squared residuals	
	(1)	(2)	(3)	(4)
Year observed: first	–0.11***	–0.01	0.09**	–0.18***
	(–4.56)	(–0.43)	(2.15)	(–7.93)
Year observed: second	–0.07***	0.02	–0.01	–0.12***
	(–3.14)	(1.05)	(–0.23)	(–3.90)
Year observed: third	–0.06**	0.01	0.01	–0.17***
	(–2.31)	(0.68)	(0.32)	(–7.86)
Year observed: two years before last	–0.09***	0.01	0.19***	–0.04*
	(–2.76)	(0.42)	(2.93)	(–1.65)
Year observed: next-to-last	–0.02	0.01	0.23***	0.01
	(–0.74)	(0.72)	(4.09)	(0.20)
Year observed: last	–0.07*	0.00	0.57***	0.18***
	(–1.70)	(0.21)	(6.50)	(3.38)
3 years before earn. miss., dummy	–0.18*	–0.18*	0.45***	0.46***
	(–1.73)	(–1.72)	(2.70)	(2.74)
2 years before earn. miss., dummy	–0.12	–0.11	0.72***	0.69***
	(–1.18)	(–1.09)	(3.21)	(3.11)
1 year before earn. miss., dummy	–0.32**	–0.33**	1.56***	1.57***
	(–2.29)	(–2.34)	(4.62)	(4.62)
1 year after earn. miss., dummy	–1.11***	–1.11***	1.35***	1.40***
	(–9.02)	(–9.00)	(5.17)	(5.40)
2 years after earn. miss., dummy	–0.52***	–0.52***	1.19***	1.23***
	(–3.95)	(–3.92)	(3.24)	(3.31)
3 years after earn. miss., dummy	–0.25**	–0.25**	0.25	0.26
	(–2.10)	(–2.07)	(1.11)	(1.16)
Adj. R sq.	0.039	0.036	0.064	0.057
No. obs.	16496	16496	16496	16496
No. indiv.	1741	1741	1741	1741

Notes: PSID male earnings data span the period 1979–1993. Earnings recorded in year t reflect remuneration received in year $t-1$. In columns (1) and (3), the dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is later than in 1979, and are zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is earlier than in 1993, and are zero otherwise. In columns (2) and (4), the dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is in 1979, and are zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is in 1993, and are zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.



(a) Variances of log earnings levels



(b) Variances of earnings growth rates

Notes: "Acf" stands for autocovariance function; "BPP" moments include the autocovariance functions of earnings and consumption growth rates, and the cross-covariances between earnings and consumption growth rates; "BPP and rare shocks" moments include, in addition, the regression coefficients reported in Table 11. Male earnings and household nondurable consumption data for 1979–1993 from Blundell, Pistaferri, and Preston (2008) are used in estimations.

Figure 1: Fit to the Moments of Male Log Earnings in Levels and Differences. PSID Data.

Table 12: Minimum-Distance Partial Insurance and Variance Estimates

		(1)	(2)	(3)
σ^2_ξ	1979–1981	0.0215	0.0084	0.0097
Variance of perm. shock		(0.0105)	(0.0044)	(0.0040)
	1982	0.0634	0.0301	0.0299
		(0.0158)	(0.0103)	(0.0092)
	1983	0.0916	0.0393	0.0380
		(0.0252)	(0.0144)	(0.0380)
	1984	0.0710	0.0317	0.0312
		(0.0223)	(0.0120)	(0.0105)
	1985	0.0819	0.0446	0.0433
		(0.0195)	(0.0138)	(0.0125)
	1986	0.0899	0.0352	0.0344
		(0.0194)	(0.0122)	(0.0109)
	1987	0.0675	0.0483	0.0468
		(0.0258)	(0.0246)	(0.0245)
	1988	0.0978	0.0458	0.0447
		(0.0316)	(0.0272)	(0.0271)
	1989	0.0740	0.0005	0.0007
		(0.0318)	(0.0153)	(0.0170)
	1990–1992	0.0527	0.0134	0.0163
		(0.0150)	(0.0122)	(0.0111)
σ^2_ϵ	1979	0.0809	0.0726	0.0707
Variance of trans. shock		(0.0189)	(0.0116)	(0.0115)
	1980	0.0640	0.0253	0.0233
		(0.0118)	(0.0130)	(0.0129)
	1981	0.0817	0.0213	0.0219
		(0.0131)	(0.0113)	(0.0115)
	1982	0.0776	0.0246	0.0245
		(0.0151)	(0.0138)	(0.0138)
	1983	0.1075	0.0400	0.0416
		(0.0193)	(0.0182)	(0.0184)
	1984	0.0921	0.0337	0.0331
		(0.0180)	(0.0170)	(0.0169)
	1985	0.1052	0.0467	0.0477
		(0.0167)	(0.0166)	(0.0168)
	1986	0.1088	0.0493	0.0493
		(0.0173)	(0.0166)	(0.0167)
	1987	0.0897	0.0269	0.0271
		(0.0151)	(0.0157)	(0.0156)
	1988	0.1319	0.0785	0.0782
		(0.0225)	(0.0216)	(0.0216)
	1989	0.0965	0.0484	0.0473
		(0.0164)	(0.0162)	(0.0162)
	1990–1992	0.1025	0.0266	0.0263
		(0.0121)	(0.0133)	(0.0133)
Serial corr. trans. shock		0.0406	–0.0182	–0.0221
		(0.0311)	(0.0584)	(0.0596)
Var. unobs. slope heterog.		0.0137	0.00^a	0.00^a
		(0.0037)	—	(0.00)
ϕ		0.2629	0.5988	0.6932
(Partial ins. perm. shock)		(0.0549)	(0.1299)	(0.1270)
ψ		0.0364	0.00^a	0.0270^b
(Partial ins. trans. shock)		(0.0295)	—	(0.0274)
ψ, rare shock		—	0.0818	0.0270^b
(Partial ins. rare trans. shock)		—	(0.0457)	—

Notes: Column (1) contains the results of the original BPP estimation. In column (2), in addition to all of the moments in the original BPP estimation, we target the regression coefficients in two regressions, with residuals and (net) squared residuals on the left-hand side, and 19 regressors on the right-hand side—six dummies around interior missing earnings observations, three dummies for the first earnings records if incomplete earnings spells start later than the first sample year, three dummies for the first earnings records if spells start in the first sample year, three dummies for the last earnings records if incomplete earnings spells end earlier than the last sample year, three dummies for the last earnings records if earnings spells end in the last sample year, and a constant—and the (average) variance of permanent shocks prior to 1979. We estimated the models of columns (2) and (3) by the method of simulated minimum distance, assuming that permanent, transitory, and rare transitory shocks are drawn from normal distributions. In all columns, in addition, we estimated the time-varying variances of measurement error in consumption. [a] binds at zero; [b] parameters are restricted to equal each other in estimation. Standard errors in parentheses.