

Reconciling Estimates of Income Processes in Growth Rates and Levels

Moira Daly
Copenhagen Business School

Dmytro Hryshko
University of Alberta

Iouri Manovskii
University of Pennsylvania
and NBER

Extended Abstract

In this paper, we attempt to improve our understanding of the individual earnings dynamics and economic forces that affect it. There is a vast and very active literature estimating stochastic processes for individual earnings. The keen interest in determining the nature of these processes stems from the fact that the earnings process is a crucial ingredient of most models with incomplete insurance markets; it is typically estimated on the data and then fed into the model. Consider, for example, the following income process:

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

where individual i 's log-income at time t , y_{it} , is composed of the permanent component p_{it} , and the transitory component, τ_{it} (see, e.g., Meghir and Pistaferri 2004). If α_p is close to one, the shocks ξ_{it} are highly persistent (permanent if α_p is one), and if $\theta(L) \equiv 1$ (where $\theta(L)$ is a moving average polynomial in the lag operator L), the shocks ϵ_{it} are completely transitory. In the context of a simple savings model under incomplete insurance markets, household savings decisions, apart from other things such as unemployment insurance and pension provisions, will depend on the size and persistence of the shocks ξ_{it} and ϵ_{it} . It is reasonable to expect that the fit of an incomplete markets model (what wealth (the model) households should have at the retirement) to the data (what wealth households in fact accumulate for the retirement) will depend on the estimates of the variances and persistence of ξ_{it} and ϵ_{it} . However, in the literature on estimation of individual and household income processes there is no consensus on the extent of persistence of the shocks ξ_{it} , as well as on the size of the shocks ϵ_{it} and ξ_{it} . Some papers in the 2010 issue of the Review of Economic Dynamics on the cross-sectional facts for macroeconomists highlighted, for data sets from different countries, that the estimates of the income process in equation (1) when using the moments of log-income in levels are substantially different from

the estimates when using the moments of log-income in differences (see, e.g., Heathcote, Perri, and Violante (2010) for the U.S., and Domeji and Flodén (2010) for Sweden). In particular, the estimated variance of ϵ_{it} (ξ_{it}) is relatively higher when using the moments in levels (differences). As a result, the estimated model when using the moments in levels (differences) is inconsistent with the observed moments in differences (levels). Heathcote, Perri, and Violante (2010) attribute this result to misspecification of the model in equation (1). However, the nature of this potential misspecification is unknown. In this paper, we fill this gap in the literature.

We use a uniquely large and unusually rich source of data, covering 100% of Danish population, in an attempt to improve our understanding of individual earnings dynamics. Various government registers provide very reliable data on earnings and labor market status of each resident of Denmark from 1981 through 2006. The data are a true census such that typical issues of attrition are avoided. Earnings contain wages and salaries from all jobs, including employer contributions to retirement accounts. These data are also linked with extremely rich demographic information.

Our analysis proceeds in three steps.

1. First, we compare the estimates of the standard earnings processes in the U.S. and Danish data. We find that earnings processes estimated in Danish data are very similar to the corresponding estimates using U.S. data, although the variances are generally smaller in Denmark. Various measures of worker mobility across employers, occupations and industries are also very similar between the U.S. and Denmark. This suggests that the lessons we learn from the Danish data are likely more widely applicable.
2. Second, we measure the effects of sample size on the estimates available in the literature. Our primary sample comprises 1,283,973 males, including 37,667 males whose earnings are observed for the entire 26 years covered by the data. Our second sample comprises 20,000 individuals and is comparable in size to the largest longitudinal U.S. data set used to study individual earnings dynamics (a confidential data set on individual earnings extracted from tax returns from the Internal Revenue Service). Our third sample comprises 2,000 individuals and is comparable in cross-sectional and longitudinal dimensions to a typical sample from the Panel Study of Income Dynamics (PSID) on which most research on earnings dynamics is based. We estimate standard models of earnings processes in the literature on each of these data sets and assess the importance of sample size for the interpretation of these estimates.

We find that the broad conclusions about the nature of the earnings process are relatively little affected by the sample size. Despite the hope expressed in the literature that the inconsistency between estimates of the earnings process in growth rates or levels might be due to the small size of the data sets on which earnings processes are traditionally estimated, we find that this is not the case and that the inconsistency remains even on the largest sample we use. To address this issue we propose richer specifications of earnings processes that reconcile the apparent contradiction between the estimates in growth rates and levels.

3. Third, our data contain very rich demographic information and information on various labor market transitions. This allows us to explore the relationship between innovations in the earnings process and other events in workers' lives to gain further insights into the dynamics of earnings. In particular, we have very detailed data on worker mobility across employers, occupations, industries, and geographic regions. We also observe measures of health for individuals (and all their family members). Specifically, we observe data on prescription medication, occurrence and the nature of surgeries and/or hospital/doctor visits in which a procedure is performed allowing us to quantify a measure for health shocks. It is argued that health shocks might be important determinants of innovations in earnings but PSID data, used to address this question, contains only very limited data on the self-reported health status.

References

- DOMEJI, D., AND M. FLODÉN (2010): “Inequality Trends in Sweden 1978–2004,” *Review of Economic Dynamics*, 13(1), 179–208.
- 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.
- MEGHIR, C., AND L. PISTAFERRI (2004): “Income Variance Dynamics and Heterogeneity,” *Econometrica*, 72(1), 1–32.