ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Rahman, Md. Arifur

Working Paper

Heterogeneous labor earning shock process in finite horizon

LIS Working Paper Series, No. 721

Provided in Cooperation with:
Luxembourg Income Study (LIS)

Suggested Citation: Rahman, Md. Arifur (2017) : Heterogeneous labor earning shock process in finite horizon, LIS Working Paper Series, No. 721, Luxembourg Income Study (LIS), Luxembourg

This Version is available at:
https://hdl.handle.net/10419/203008

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

LIS
Working Paper Series

No. 721

Heterogeneous labor earning shock process in finite horizon

Md. Arifur Rahman

November 2017

Luxembourg Income Study (LIS), asbl

Heterogeneous labor earning shock process in finite horizon

Md.Arifur Rahman[♣]

Abstract

This paper investigates the importance of heterogeneity in the labor earning shock processes. We analyze the earning shock process for both male and female workers in several countries. We argue that unlike time series analysis, in a life cycle model the forecasting horizon is finite and in finite horizon cross-sectional variance profiles of labor earning shocks display any of five specific shapes. Since our study is based on cross-sectional data, we do not estimates the time series parameters, rather we put emphasis on the shapes of the variance profiles. A linear variance profile implicates highly persistent labor earning shock. Even though the overall variance profile is non-linear, we can show that earning shock is highly persistent and the curvature of variance profile can be explained by the specification of Storesletten, et. al. (2004) and Guiso, et.al. (2002). Moreover, the study of persistence of female labor earning is a novel contribution of this paper. Earlier studies occasionally assume highly persistent female earning, nevertheless, remain silent on providing empirical evidence of female earning shock process.

Keywords: inequality, income distribution, wages, earnings

1. Introduction

In Life-cycle models with stochastic earnings (or in DSGE models) the labor earning shocks are often assumed to be homogeneous. In other words, individual-earnings receive shocks from same i.i.d. distribution (often assumed to be log-normal). This paper reviews the concern expressed by Browning, Hansen, & Heckman (1999) that while using micro-data evidences in general equilibrium models it is important to distinguish between micro and macro uncertainties of labor earning.

> *"Aggregating earnings across members of households and across income types may create a disconnect between uncertainty as measured by innovations in time series processes of earnings and income equations and actual innovation in information."* (Browning, Hansen, & Heckman, 1999)

[♣] National Graduate Institute for Policy Studies (GRIPS), Tokyo and Office of the Comptroller & Auditor General of Bangladesh.

Author extends sincere thanks and gratitude toward Ponpoje Porapakkarm for advising this project. He gratefully acknowledges the critiques and contributions of Fumio Hayashi in the econometric analysis. He specially thanks Alistair Munro for his idea of gender neutral approach in life-cycle models, and Junichi Fujimoto for the idea of diverse sources of uncertainty. Author is also grateful to Tatsuo Oyama, Minchung Hsu, Yoichi Arai, Shin S Ikeda, Khawado Opoku, Chie Aoyagi and Sharif M Hossain. Takashi Tsuchiya has been contributed to solve *matlab* program and thanks to Niku Maattanen for providing the *matlab* code, which is the extended part of the paper. Special regards to Kritsanee Pisitsupakul for her significant contribution in editing, structure and guidance. Author is also grateful to GRIPS for providing the GRIPS fellowship to conduct the research and to JASSO for its contribution to the completion of the project. He extends thanks to OC&AG, and Ministry of Finance, Bangladesh for providing necessary co-operation throughout the project period.

Taking the concern into account, it can be said that the study of labor earning shock in the field of economics especially in the life-cycle model estimation plays a very important role. Although most of the literature regarding labor earning shocks are focused on the relationship between earning and consumption (Carroll & Summers, 1991), in this paper we solely concentrate our attention on the dynamics of earning shock over the working life of an individual (rather than household) to better understand the heterogeneous sources of earning uncertainty.

It is to be mentioned that, reviewing this concern is not the only interest of the paper. By assuming, different individual's earning-shocks come from different distributions, we prove that highly persistent earning shocks can produce non-linear earning variance profile over the life cycle. Our proof not only sheds lights on the current debate in life-cycle literature that whether labor earning shocks are highly persistent or not, more importantly, it also provides a rare empirical evidence that labor earning of female workers are also highly persistent. Though in many literature female labor earning assumed to be highly persistent (Blundell, Dias, Meghir, & Shaw, 2016), probably because of the non-linear profile of female earning variance, study on the persistence of female labor earning is limited compared to that of male earning.

We study a series of cross-sectional data from Luxembourg Income Study (LIS) database, and construct pseudo-panel (see Section 4.2) data for several countries including United States and Taiwan. Since we cannot observe the same individual across the time points, we cannot estimate the persistence parameter or the time series variance of the earning shock process directly from the data. Therefore, our study is limited to descriptive and analytical methods.

Based on several panel data analysis, there are different views and specifications regarding labor earning shock process in literature. However, it is widely accepted that labor earning shock is an AR (1) process and has a persistent component in it (Meghir & Pistaferri, 2004). Some study finds highly persistent (Storesletten, Telmer, & Yaron, 2004) earning shocks over the life cycle while some other differ (Guvenen, 2007; Allerano, Blundell, & Bonhomme, 2017; Karahan & Ozkan, 2013) the argument.

A highly persistent shock process implicates a linear earning variance profile over the working life. However, the cross-sectional variance profile of earning not only depends on persistence parameter (ρ), but also on the time series variance of earning shocks. Browning, Hansen, & Heckman (1999) expressed concern that aggregating earnings from all sources may not accurately represent the true uncertainty. It also says, though it is widely accepted that residuals of male labor earnings from Mincer earnings equation (Mincer, 1974) follow AR (1) process, there is little information about female labor earnings in existing literature of DSGE models. Each source of earnings has its own component in uncertainty[1] which makes the household earning shocks a complex process (Browning, et.al., 1999). It is also evident from US Panel Study on Income dynamics (PSID) data that, labor earning uncertainty varies depending on occupation across ages or sexes (Carroll & Samwick, 1997).

[1] Uncertainty is defined as the variance of permanent component and transitory component of the earning shock process in Browning, Hansen, & Heckman (1999).

To ease the complexity of household labor earning in this paper we study individual labor earning shocks by segregating them in four groups; by gender and by occupation. For the classification of occupation we follow the LIS database guidelines and classify workers as *Dependent Employees (DE)* and *Independent Employees (IE)* (see Appendix A4).

We analyze the effect of persistence parameter by analyzing the shapes of the variance profiles keeping the shock process homogeneous (or shocks come from same distribution) for all individuals. Later we also assume that the shock process is heterogeneous (i.e. distribution of shocks of *DE*s are different from that of *IE*s) to see the shape of variance profile keeping the persistence parameter constant. It can be said that (from the evidence in data: Figure – 4) an *IE* is exposed to a persistent shock and a transitory shock from broader distributions; whereas by creating a Dependent Employment some part of the persistent component (ε_t) and some part of the transitory component (η_t) of the worker's earning shock can be hedged. In other words a *DE* receives a shock from a narrower distribution than that of an *IE*.

In the following section we study the specifications developed for labor earning shock process in recent literature. Section 3 derives the theoretical shapes of variance profiles depending of different values of persistence parameter (ρ). In Section 4 we describe the data and provide the cross-sectional variance profiles of various countries for both male and female workers. The fifth section of this paper analyses the shapes obtained from the data. Finally we conclude this paper in Section 6.

2. Literature Review

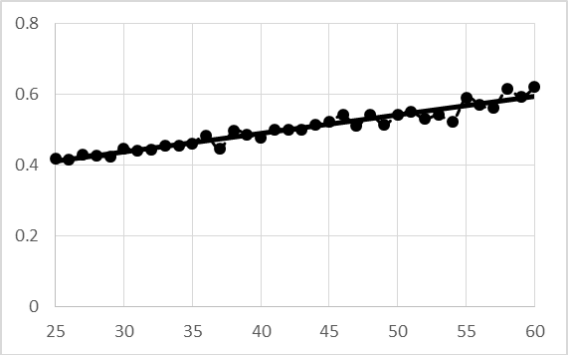
The debate over persistence of labor earning shock constitutes a significant interest in the study of life-cycle models. In many of the literature it has already been proved that there might be a persistent component in the shock of labor earning (Meghir & Pistaferri, 2004). However, in different literature the degree of persistence is often debated.

Labor earning is composed of two components, a random walk and a white noise (Deaton & Paxson, 1994; Blundell & Preston, 1998) component which can be expressed in the following (assuming $\epsilon_t = \log y_t$):

$$\epsilon_t = \theta\epsilon_{t-1} + \xi_t \quad \dots \quad (1)$$

With this simple expression if $\theta = 1$ then according to the time series literature the cross-sectional variance of labor earning increases linearly over time. And in the data from US it has been seen that cross-sectional variance of log earning increases linearly between ages 25 and 60 (see Figure 1). Equation (1) is a very simple approximation of the increase in log-earning variance over life cycle.

Figure 1 Log earning variance profile of the male worker in the United States



However, equation (1) with time invariant ($|\theta| \leq 1$) cannot produce a hump shaped average earning profile for male workers which is found in many studies (Friedman, 1957; Ben-Porath, 1967; Mincer, 1974). Moreover, in case of female workers the earning profile is more complicated, for instance it is M shape in case Japanese female workers (Lise, Sudo, Suzuki, Yamada, & Yamada, 2014). A modified model was then proposed to accommodate both the average earning profile and cross-sectional variance profile (Storesletten, Telmer, & Yaron, 2004; Meghir & Pistaferri, 2004), where the average earning is a function of age (t) and the shock to earning has persistent (ε_t) and transitory (η_t) components.

$$\log y_t = \kappa_t + \alpha + \eta_t + z_t \quad \ldots \quad (2)$$

$$z_t = \rho z_{t-1} + \varepsilon_t \quad \ldots \quad (3)$$

Where, κ_t captures the average income profile, α is the individual fixed effect, η_t is the transitory earning shock and z_t is the AR (1) shock process with the persistence parameter (ρ) and ε_t is the white noise component. The cross-sectional variance profile can be given as

$$Var(\log y_t) = \sigma_\alpha^2 + \sigma_\eta^2 + Var(z_t) \quad \ldots \quad (4)$$

where $\alpha \sim iid(0, \sigma_\alpha^2)$; and $\eta_t \sim iid(0, \sigma_\eta^2)$. Storesletten, et. al. (2004) claims that in the United States ρ is very close to one implying the labor earning shock is highly persistent. Guvenen (2007) opines that the individual specific component in the shock process can also vary over time. Therefore, Storesletten, et. al. (2004) can be called a Restricted Income Profile (RIP). A Heterogeneous Income Profile (HIP) is one where the individual specific component is a function of age (t) and can yield to a different conclusion. It shows that if we estimate the persistence parameter using RIP process ρ is very close to one whereas using HIP process it is around 0.8 (Guvenen, 2007). The HIP process can be written as:

$$\log y_t = \kappa_t + \alpha + \varpi\, t + \eta_t + z_t \quad \ldots \quad (5)$$

$$z_t = \rho z_{t-1} + \varepsilon_t \quad \ldots \quad (6)$$

Where the individual specific component is a linear function of t with a slope of ϖ as well as α and ϖ are correlated, that is the individual specific component is positively correlated with her earning growth. The cross-sectional variance profile is given by

$$Var(\log y_t) = \sigma_\alpha^2 + \sigma_\eta^2 + t^2\sigma_\varpi^2 + 2\sigma_{\alpha\varpi} + Var(z_t) \quad \dots \quad (7)$$

The HIP process has some advantage over the RIP process. Though the RIP process can explain the earning inequality profile, the HIP process can explain both the earning and consumption inequality profile for heterogeneous groups (according to educational attainment) as well as the economy as a whole.

Karahan & Ozkan (2013) finds that the specifications in the literature are too stringent to capture the dynamics of labor earning shock over the life cycle. In the United States the persistence of earning shock is age dependent. For the younger workers the persistence is around 0.75 while it increases over the working life and gets closer to one at the end of working life (Karahan & Ozkan, 2013). The shock process in equation (3) has been modified as

$$z_t = \rho_{t-1} z_{t-1} + \varsigma\varepsilon_t \quad \dots \quad (8)$$

where ς is the co-efficient of the White Noise process and ρ is age dependent. A similar argument is seen in another literature which shows that the persistence is not only age dependent but also it depends on earning shock levels (Allerano, Blundell, & Bonhomme, 2017). It shows that the state dependent persistent parameter ρ changes across the percentiles of earning shocks.

Apart from this debate there is another paper which explains the earning uncertainty based on the employment status (Guiso, Jappelli, & Pistaferri, 2002). In this paper the cross-sectional variance has been divided into intragroup variances and intergroup variance. Given that there are two groups: employed and unemployed, the total cross-sectional variance at a particular age t can be divided into three components. The cross-sectional variance for employed, the cross-sectional variance for unemployed and the intergroup variation between the two, which can be shown in the following equation.

$$Var(\log y_t) = p_t Var(\log y_t^E) + (1 - p_t) Var(\log y_t^U) + p_t(1 - p_t)[E(\log y_t^E) - E(\log y_t^U)]^2 \quad \dots \quad (9)$$

where $p_t$is the proportion of Employed among the total workers. This equation determines the cross-sectional variance in the total by considering the cross-sectional variances of two sub-groups and their proportions in the economy.

Carroll, & Samwick (1997) has also found that in the United States the distribution of labor earning shock differs across occupation as well as age. It opines that largest distribution of shock (or uncertainty) is faced by Farmers, Service workers and Self-employed managers. This group is followed by Laborers, Clerical workers, and Managers. The least distribution of shock is experienced by Professionals and Craftsmen. This paper also claims the distribution also differ across ages; which implies, older workers face higher distribution (or uncertainty) than younger workers.

Browning, et.al. (1999) claims that ignoring the sources of earning shocks, may affect the predictability of the DSGE or Life-cycle model. Regarding the earning uncertainty it spells out three key points: first, household is composed of male and female earners, who are likely to face different uncertainties. Second, though the male earning is widely accepted to follow AR (1) process, there is little evidence on female earning in relevant literature of Life-cycle models. And

finally, the sources of uncertainty can be dominated by occupation, family structure, health etc. which is often ignored while setting up the models.

Considering these observations, we segregate individual earners according to their gender and occupation. We assume that, male earners and female earners are subject to different earning processes. In addition, *IE*s are susceptible to shocks from a broader distribution than that of *DE*s. In Section 4 we plot the variance profiles of log-earning for both male and female workers of United States and Taiwan segregating them into *IE*s and *DE*s (see Figure – 4). We assume that sources of uncertainty are the different based on types of employment and workers' sex.

3. Variance profiles: (Assuming σ_ε^2 is common to all individuals)

The variance profile of an AR (1) shock process is an utmost interest of this paper. In the time series econometric literature the shock structure is given as (Hayashi, 2000, p. 561)

$$z_{t+s} = \rho^s z_t + \rho^{s-1}\varepsilon_{t+1} + \cdots + \rho\varepsilon_{t+s-1} + \varepsilon_{t+s}$$

where z_t is the AR (1) shock process of labor earning and ε_t is the persistent component (a white noise process) that is accumulated at each period over the forecasting horizon of s. We can obtain the cross-sectional variance of the process $\{z_{t+s}\}$ as follows:

$$Var(z_{t+s}) = \rho^{2s} Var(z_t) + \sigma_\varepsilon^2 \left[\frac{1-\rho^{2s}}{1-\rho^2}\right] \quad \ldots \quad (10)$$

where $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. As $Var(z_t)$ is known at time t we can say $Var(z_{t+s})$ depends on s as well as ρ and σ_ε^2.

In the time series analysis the usual assumption is ρ and σ_ε^2 is constant over time and $s \to \infty$. Given that $Var(z_t)$ is known at time t, as s goes to infinity the variance profile of $\{z_{t+s}\}$ produces two shapes. Because for $s \to \infty$ we can show $\rho^{2s} \to 0$ when $\rho^2 < 1$ yielding a constant variance $\sigma_\varepsilon^2 \left[\frac{1}{1-\rho^2}\right]$ over time. As a result $\{z_t\}$ becomes a stationary AR (1) process. For $\rho^2 > 1$ as $\rho^{2s} \to \infty$ the variance $Var(z_{t+s})$ is undefined. And for $\rho^2 = 1$ the variance profile is a linear function in s as it can be written as $Var(z_t) + s\sigma_\varepsilon^2$.

Though it is evident that cross-sectional variance can be obtained from time-series variances, the vice versa is not straight forward. The time series variance of ε_t can only be obtained by the cross-sectional variance with a strong assumption that all individuals' time-series shocks follow same distribution σ_ε^2. Suppose there are n individuals in an economy who are observed for t years. Now from the assumption of the white noise process for the 1st individual we can write $\{\varepsilon_{11}, \varepsilon_{12}, \ldots \varepsilon_{1t}\} \sim iid(0, \sigma_\varepsilon^2)$; for the 2nd individual $\{\varepsilon_{21}, \varepsilon_{22}, \ldots \varepsilon_{2t}\} \sim iid(0, \sigma_\varepsilon^2)$; and similarly for the n^{th} individual $\{\varepsilon_{n1}, \varepsilon_{n2}, \ldots \varepsilon_{nt}\} \sim iid(0, \sigma_\varepsilon^2)$. Now if we assume that all the individuals get a shock from the same distribution σ_ε^2, then we can show at time t the cross-sectional sequence $\{\varepsilon_{1t}, \varepsilon_{2t}, \ldots \varepsilon_{nt}\}$ also follows $iid(0, \sigma_\varepsilon^2)$. Which implies $\sum_{n=1}^{N} \varepsilon_{nt}^2 \to \sum_{t=1}^{T} \varepsilon_{nt}^2$ when $N \to T \to$

∞. Probably for this reason when we calibrate a Life cycle (or DSGE) model we usually generate the earning shocks for every individual from the same distribution[2].

Having said that we can argue, estimation of σ_ε^2 (in the cross-section) is more reliable only when the sample size becomes very large. In case of panel data both of the dimensions 'time horizon' and 'number of individuals' are not very large. Since panel data exhibit certain characteristics of time series data, they have advantage over cross-section data in estimating time series parameters. However, to estimate σ_ε^2 (in the cross-section) if the sample size of the panel data is not very large, then the assumption of individuals facing same distribution of shock might be very strong.

On the other hand from a cross-sectional data we can only have a speculative approximation of time series variance. Therefore, it is necessary to generate a simulated sample replicating the features and characteristics of the cross-sectional data, and then estimate the underlying parameters by method of simulated moments (which is currently not covered by this paper).

Let us argue that, for a life cycle model it is reasonable to assume that $s < \infty$. For a finite s we know that $\rho^{2s} < \infty$ for any real value of ρ. And we state that it is necessary to analyze shape of the variance profile for different values of ρ and σ_ε^2 in the finite horizon of s. From equation (10) we can derive the slope of the variance profile as (for $\rho \neq 1$ and constant over time)

$$Var(z_{t+s}) = \rho^{2s} Var(z_t) - \sigma_\varepsilon^2 \left[\frac{\rho^{2s}}{1-\rho^2}\right] + \sigma_\varepsilon^2 \left[\frac{1}{1-\rho^2}\right]$$

$$Var(z_{t+s}) = \rho^{2s} \left[Var(z_t) - \frac{\sigma_\varepsilon^2}{1-\rho^2}\right] + \sigma_\varepsilon^2 \left[\frac{1}{1-\rho^2}\right]$$

$$\frac{\partial\, Var(z_{t+s})}{\partial\, s} = \rho^{2s} (\log \rho^2) \left[Var(z_t) - \frac{\sigma_\varepsilon^2}{1-\rho^2}\right] \quad \dots \quad (11)$$

The slope of the variance profile depends on ρ and σ_ε^2 simultaneously. If they vary across time (or other dimension) it would be very difficult to assess what is the effect of the two on the shape of the variance profile.

Therefore let us assume ρ and σ_ε^2 are constant over time and the slope of the variance profile is a function of s only. Then equation (11) can be written as

$$\frac{\partial\, Var(z_{t+s})}{\partial\, s} = \rho^{2s} (\log \rho^2) \left[Var(z_t) - \frac{\sigma_\varepsilon^2}{1-\rho^2}\right] = \rho^2 \rho^{2(s-1)} (\log \rho^2) \left[Var(z_t) - \frac{\sigma_\varepsilon^2}{1-\rho^2}\right]$$

$$\frac{\partial\, Var(z_{t+s})}{\partial\, s} = \rho^2 \frac{\partial\, Var(z_{t+s-1})}{\partial\, s-1} \quad \dots \quad (12)$$

Which implies ρ^2 is the ratio of the slope of the variance profile at time s and $s-1$. So the change of slope has a particular pattern for a constant ρ and constant σ_ε^2. Now if $\rho^2 < 1$ then the slope of the profile decreases at every time. However, the sign of the slope depends on $(\log \rho^2) \left[Var(z_t) -\right.$

[2] In literature of life cycle model, the distribution σ_ε^2 is estimated in the time dimension and averaged across individuals using panel data, and during calibration all individuals are given the shock from same distribution.

$\frac{\sigma_\varepsilon^2}{1-\rho^2}\Big]$. Since $(\log \rho^2) < 0$ and $\frac{\sigma_\varepsilon^2}{1-\rho^2} > 0$, the slope could be negative, positive or zero. In case of negative slope the variance profile decreases over time at a decreasing rate, which gives it a downward sloping convex shape. If it is positive then the profile increases at a decreasing rate implicating an upward sloping concave shape. And in the third case when it is zero the variance profile will be a straight line parallel to the horizontal axes.

The time series literature (Hamilton, 1994; Hayashi, 2000) implicate that when $\rho^2 = 1$ the cross-sectional variance profile not only linear but also has a positive slope. Though time series literature have rule out the case of $\rho^2 > 1$, we say that if this case exists the variance profile will be an upward sloping convex curve, unlike the previous case the slope will be always positive as $(\log \rho^2) > 0$ and $\frac{\sigma_\varepsilon^2}{1-\rho^2} < 0$.

We conclude in this section that with a constant ρ and a constant σ_ε^2 over the life-cycle, the variance profile of the labor earning shock process displays any one of the five specific shapes (in a finite horizon) and ρ^2 is the ratio of slopes of the variance profile at two consecutive time periods.

4. Log-earning variance profiles in the data

4.1. The data

In this study we analyze Luxembourg Income Study (LIS) database (LIS, 2017), which is one of the largest database for household and individual income surveys across the world. The database is harmonized across countries and across years, that is why it is easy to compare across times or places. The variable definitions are also standardized and easy to understand. The data are available for more than 50 countries across the world for several decades. However, we select a few countries (see Section 4.3) to represent different regions of the world given that the sample sizes are relatively large. It also important for our selection that individual income data are available for each age group.

For this paper we define a worker based on her labor earning. If labor earning is positive then the person is a worker otherwise a non-worker. Then we pooled the cross-section data that are available after 2001 to 2013 and separated them according to sex. We regress the labor earning with year dummy and remove the year effect from the individual earnings. Afterwards, we top code and bottom code the highest and lowest one percent observations in the data. We take log of the earnings and then construct the variance profiles for every ages.

4.2.The pseudo-panel

We follow Deaton (1985) and Deaton & Paxson (1994) to construct our pseudo-panel from the repeated cross-sectional data. We construct a cohort of individuals consisting members of the same age, or born in the same year. For example in the survey data of 2004 the cohort of age 20 becomes cohort of age 23 in that of 2007. Assuming that both the survey datasets are representative enough

we argue the cohort of age 23 in the data is representative of the cohort population. However, in the cohort of age 23, there are several groups of people now who are born in different years.

The heterogeneity of these groups comes from two sources: income year effects, and birth year effects. We eliminate the income year effect by controlling for year dummy. However, the birth year effect was not eliminated. We assume that the birth year effect is an individual fixed effect and the variations in labor earnings that come from the birth year-effect, do not change across ages. Therefore, it does not affect the persistence of shocks (given that sample is large and representative enough), only increases the variance of transitory components (see Appendix A3).

It can be argued that population belonging to a particular cohort is fixed through time (Deaton A. , 1985). Which implies the cohort mean and cohort variance can be estimated from the micro survey data. However, the sample size remains an issue here. In many literature five-year age groups have been constructed to increase the sample size (Deaton & Paxson, 1994). In this paper we create age-cohort for every age. For the United States and Brazil the cohort sizes are greater than 1000, and for other economies it is over 300. We assume that the cohort sizes are large enough to avoid further adjustments.

4.3.The log-earning variance profiles

The variance profiles are given in Figure – 2 and Figure – 3. Male workers variance profiles are given in Figure – 2 and that of female workers are given in Figure – 3. We have eight profiles for male workers including United States, United Kingdom, Brazil, Mexico, Germany, India, Taiwan and Japan; and four profiles for female workers United States, United Kingdom, Brazil and Taiwan. Except for India's male earning profile we take e-based logarithm to generate log-earning. Only for India's male workers we take 10-based logarithm. The earning profile of Japan male workers has been constructed by using Japan Household Panel Survey (JHPS) data.

In case of India the log-earning variance profile with e-based log is highly fluctuating. In addition, for many younger cohorts in India data the standard deviation is more than one, which generates a mathematical inconvenience. While the variance of a cohort having standard deviation less than one is smaller than its standard deviation, the variance of a cohort having standard deviation more than one is higher than its corresponding standard deviation. Which leads to even higher degree of fluctuation. To avoid this mathematical phenomenon, we take 10-based log, which gives standard deviations less than one for every age cohorts and thus simplify our analysis.

5. Variance profiles: (Assuming σ_ε^2 is heterogeneous based on employment)

Given the theoretical variance profiles of five different shapes (see Section 3), we can say that male earning profiles are somewhat closer to theoretical shapes[3]. However, the female workers

[3] From age 25 to 60 the variance profiles for United States, United Kingdom, Brazil, Mexico, and India are linear or approximately linear (see Figure – 2). For Taiwan, Germany and Japan the profiles are convex. Deaton & Paxson (1994) also documents, variance profiles for United States, United Kingdom and Taiwan, which are similar to the profiles in Figure – 2. Linear or convex shapes are among the five shapes that we discussed in Section 3. We include several countries from several regions across the world, to make a comparison of earning variances across the

earning profiles (especially for United States and United Kingdom) are away from them[4]. Which can raise a question that – *are the specification in literature of a constant ρ and a constant* σ_ε^2 *too stringent* (Karahan & Ozkan, 2013; Allerano, Blundell, & Bonhomme, 2017)?

We argue that the non-linear shapes of the variance profiles in the data can be explained within the restricted specification ($\rho = 1$) in literature. Unlike Storesletten et. al. (2004) where all workers draw a shock from the same distribution, we assume that, there are at least (it can be more) two groups of workers in the economy. The first group constitutes of 'Independent Employees (*IE*)' and the second of 'Dependent Employees (*DE*)'. The *IE*s are subject to shocks from broader distribution, both persistent ($\sigma_{\varepsilon IE}^2 > \sigma_{\varepsilon DE}^2$) and transitory ($\sigma_{\eta IE}^2 > \sigma_{\eta DE}^2$) compared to the *DE*s. The reason behind the assumption is that by creating the dependent employment the economy hedges some parts of the earning shocks of the workers. This argument follows the findings of (Carroll & Samwick, 1997), which shows that Farmers, Laborers, and Self-managers face the highest distribution of earning shocks.

We summarize the assumptions:

i. Labor earning shocks are highly persistent ($\rho = 1$) and follow Storesletten et.al. (2004).
ii. *IE*s are subject to higher persistent shock ($\sigma_{\varepsilon IE}^2 > \sigma_{\varepsilon DE}^2$).
iii. *IE*s are also exposed to higher transitory shocks ($\sigma_{\eta IE}^2 > \sigma_{\eta DE}^2$).

Given these assumptions, following Storesletten, et.al. (2004) the cross-sectional variance profiles for *IE*'s will be a straight line (Hayashi, 2000, p. 561) with a slope of ($\sigma_{\varepsilon IE}^2$). Similarly, *DE*'s profiles will also be a straight line with a slope of ($\sigma_{\varepsilon DE}^2$). Since ($\sigma_{\varepsilon IE}^2 > \sigma_{\varepsilon DE}^2$), the variance profiles of *IE*'s will be steeper compared to that of *DE*'s. In addition as ($\sigma_{\eta IE}^2 > \sigma_{\eta DE}^2$), the point where *IE*'s profile intersects vertical axis will stay above the point where *DE*'s profile intersects. Now if we compare these theoretical shapes with the shapes in Figure – 4 we find a proximity implying that the assumptions may work.

We also express that when a worker moves to *DE* at time (t), her earning process accumulates the shock in the following process:

$$z_{t+s} = z_t + \varepsilon_{t+1}^{DE} + \cdots + \varepsilon_{t+s-1}^{DE} + \varepsilon_{t+s}^{DE} \dots\dots (13)$$

where, $\varepsilon_t^{DE} \sim iid\ (0, \sigma_{\varepsilon DE}^2)$. And similarly when the person moves to *IE* the shock process can be expressed as

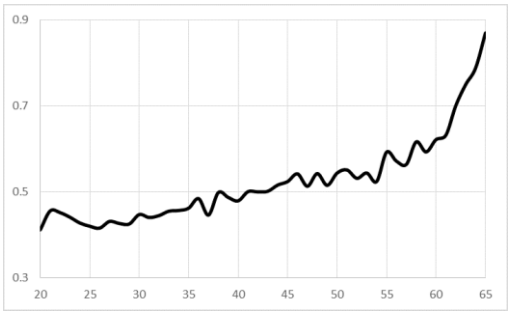
$$z_{t+s} = z_t + \varepsilon_{t+1}^{IE} + \cdots + \varepsilon_{t+s-1}^{IE} + \varepsilon_{t+s}^{IE} \dots\dots (14)$$

where $\varepsilon_t^{IE} \sim iid\ (0, \sigma_{\varepsilon IE}^2)$.

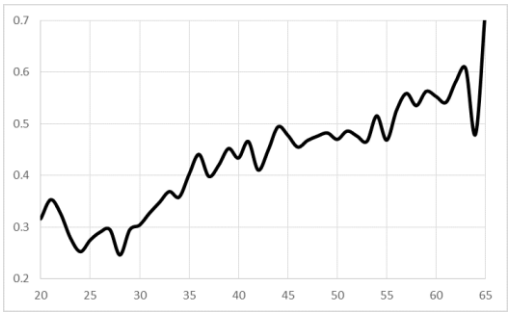
countries. However, the comparison not only shows the differences of the country specific profiles, but also focuses on the similarity of earning shock process, i.e. earning shocks follow AR (1) process irrespective of countries.

[4] Female variance profile from age 25 to 60 is approximately linear in case of Brazil. It is convex in case of Taiwan. However, for the United States and United Kingdom the variance profiles increase till 30 to 35 and then fall until 40 to 45 years of age respectively. Finally, the profiles of U.S. and U.K. again increases till the end of the working life. It is to be noted here that Brazil only reports labor earning for *dependent employees* (for both male and female).

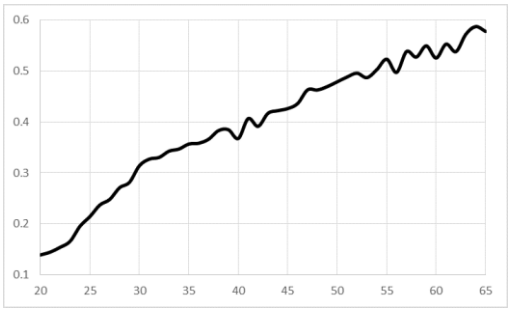
Figure 2 Log-earning variance profiles of male workers



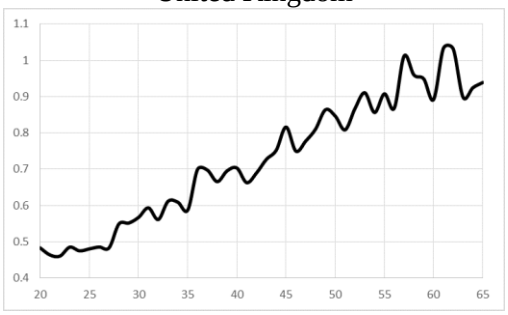
United States



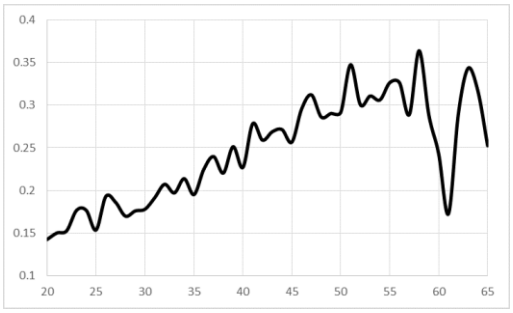
United Kingdom



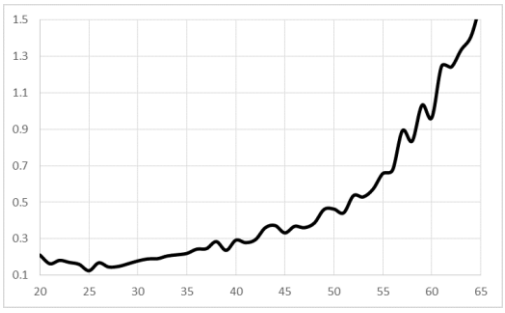
Brazil



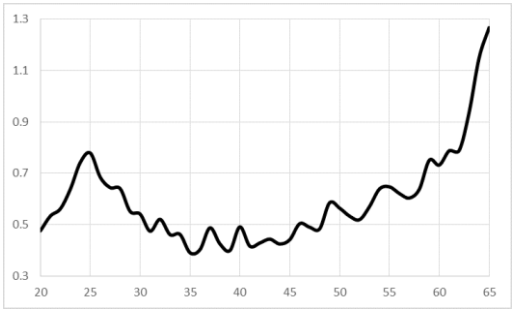
Mexico



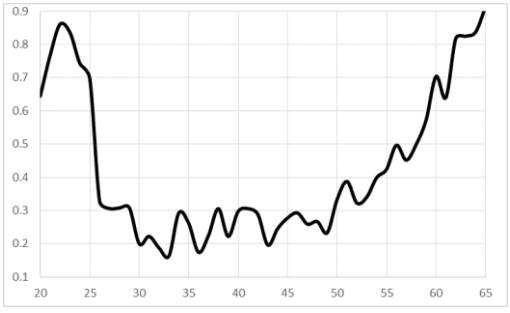
India



Taiwan



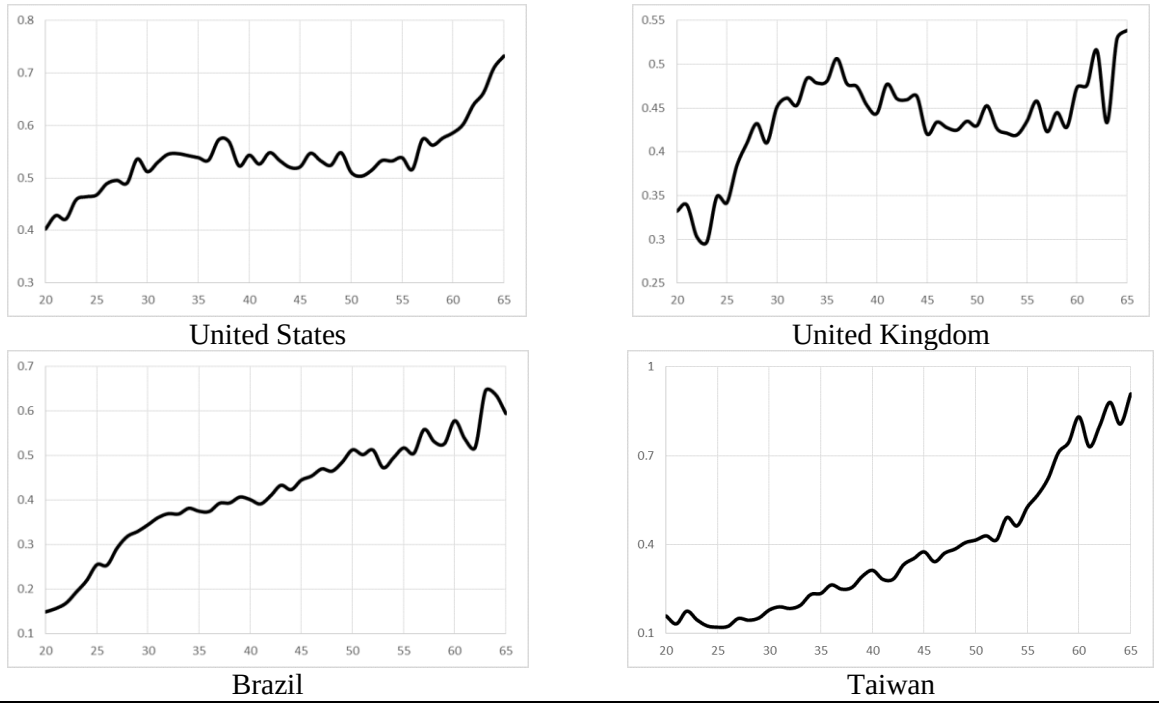
Germany



Japan

Note. Variance profiles have been constructed from Luxembourg Income Study (LIS) database except for Japan. We constructed the profile for Japan using JHPS data, because in LIS (2017) data set only 2008 income data are available for Japan. Moreover, the data set of Japan is not large enough compared to other countries. Except for India we use e-based logarithm, while for India we use 10-based logarithm. In case of India the standard deviations of log-earning (e-based) across the age groups fluctuate to a large extent, for some age-groups they are higher than one, while for some others they are less than one. If the standard deviation is less than one then the variance is smaller than it. On the contrary, if standard deviation is over one then variance becomes larger than standard deviation. Which leads to an increased degree of fluctuation. To avoid this problem we take 10-based log in case of India's income data.

Figure 3 Log-earning variance profiles of female workers

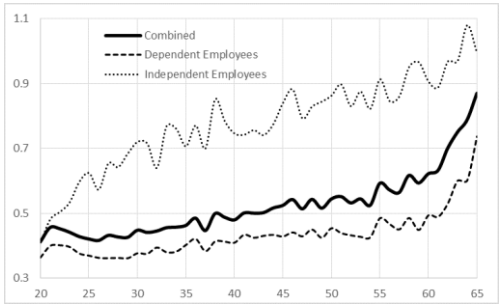


Note. For female workers we only select four countries based on the sample size of the female workers in the data. In case of these countries the sample sizes are large enough to construct the variance profiles. It should be noted here that, Brazil only provides individual income data for *dependent employees*.

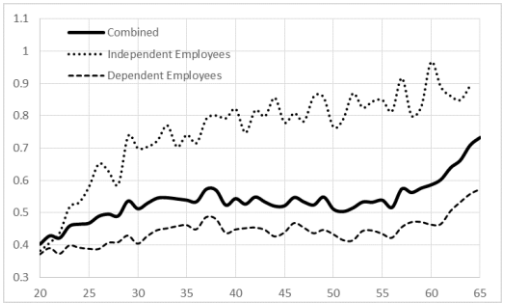
Since at time t we know z_t and we also know the employment status of the individual we can forecast z_{t+s} conditional on the employment status and thus can approximate the time series variances by the cross-sectional variances conditional on employment. Therefore, we argue that the cross-sectional variance profiles given the employment will be two different straight lines. The slope of *IE*'s profile is higher ($\sigma^2_{\varepsilon IE}$) than that of *DE*'s profile ($\sigma^2_{\varepsilon DE}$). We also argue that the transitory shock is also partially hedged for the *DE*'s, which produces a smaller intercept ($\sigma^2_{\eta DE}$) for the *DE*'s profile than that of *IE*'s profile ($\sigma^2_{\eta IE}$).

Figure 4 closely pictures our above explanation. In addition it also exhibits that the shapes of the overall variance profiles are replicated by the shape of the proportion of employees among total workers. Which implies the relationship between share of employees and the overall cross-sectional variance at each age. And this relationship is described by equation (9) following Guiso, et.al. (2002), which shows that the overall cross-sectional variance depends on within group variance, between group variance and the group's proportion among total workers.

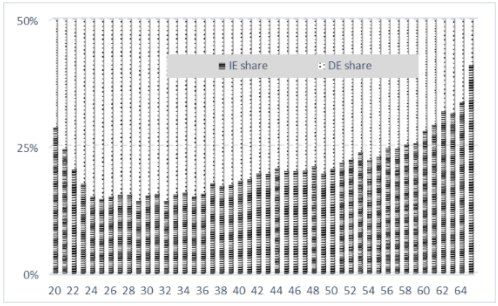
Figure 4 Log earning variance profiles by types of employees in the Unites States and Taiwan



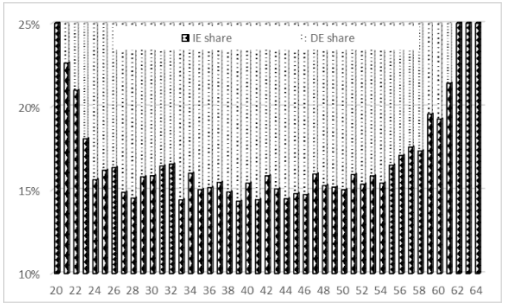
United States male workers



United States female workers



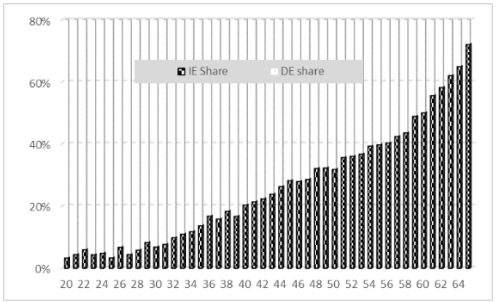
Proportion of male workers (US)



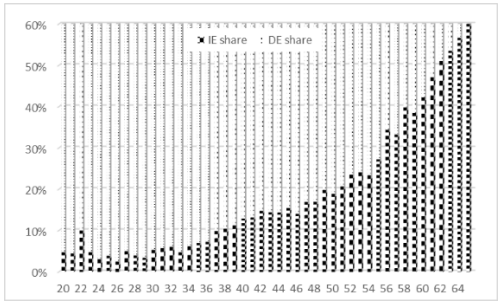
Proportion of female workers (US)

Taiwan male workers

Taiwan female workers



Proportion of male workers (Taiwan)



Proportion of female workers (Taiwan)

Note. The log-earning variance profile for Taiwan male *Independent Employees* is still convex. However, we can split it into two more straight lines. Which implies that this variance profile is also composed of two highly persistent shock process. The figure is given in the appendix (see Appendix A1). We also note that, U.S. and Taiwan are compared because they the two countries among three largest data sets. The third one is Brazil, which is excluded because it does not provides data for *IE*s. Moreover, in many related literature a comparison between U.S. and Taiwan is often observed (Deaton & Paxson, 1994; Deaton & Paxson, 1995; Deaton & Paxson, 1997). While the variance profile (male workers) of U.S. is linear, the profile of Taiwan is convex. And both profiles are composition of two or more linear profiles. Therefore, comparing between U.S. and Taiwan gives us additional insight on the persistence of labor earning shocks.

The within group cross-sectional variance[5] for both *IE*'s and *DE*'s we can obtain by their time series properties given in equation (13) and (14). Then from these within group cross-sectional variances we obtain overall cross-sectional variance by Guiso, et.al. (2002). In this section we conclude that, even though we consider that labor earning shock is highly persistent, the non-linearity of the log-earning variance profile can be explained by assuming heterogeneous shock processes.

6. Conclusion

We conclude that Browning, Hansen, & Heckman's (1999) concern should be taken into account while using estimates of labor earning uncertainty from micro-data in life-cycle models or general equilibrium models. Besides the conventional framework where we assume the uncertainty of labor earning comes from single source, we can introduce heterogeneity in the sources of uncertainty (as also evident in (Carroll & Samwick, 1997)).

The dynamics of the log-earning variance profiles have been explained by various specifications in recent literature. However, the dynamics of earning shock can also be explained by a restricted profile of Storesletten et.al. (2004) with the help of Guiso, et. al. (2002), by introducing heterogeneity in the earning shock process.

The shape of the log-earning variance profile not only depends on the persistence of the shock, but also depends on the heterogeneity of sources of the shock. These sources can be sex, job, place, education, or many other relevant things. Even though we put a strong assumption in place that labor earning shock is highly persistent ($\rho \cong 1$) the curvature of the overall cross-sectional variance profile can be explained by introducing heterogeneity of distribution of earning shocks.

There are two major drawbacks of the study. Since we do not have a panel data we cannot estimate any time series parameter other than explaining it theoretically. Second, this study cannot implicate anything regarding the consumption inequality since it is a household phenomenon, which is the key interest of many economics researches.

Despite the drawbacks we have shown that our explanation is more general in terms of labor earning shock than the earlier studies, since we can explain the earning shock dynamics across gender and across occupations (in other words across various sub-groups in the economy). Moreover, our study provides the rare empirical evidence that, female labor earning is also an AR (1) process and highly persistent. Whereas, earlier studies remain silent on analyzing the persistence of female labor earning, they occasionally assume a highly persistent female labor earning shock.

[5] The within group cross-sectional variance does not depend on the proportion of the group in the sample, rather it can be better approximated by a very large sample size for all the groups in the sample assuming that individuals within the group face the same distribution for the shocks.

References

Allerano, M., Blundell, R., & Bonhomme, S. (2017). Earning and consumption dynamics: A non-linear panel data framework. *Econometrica, 85(3),* 693-734.

Ben-Porath, Y. (1967). The production of human capital and life cycle of earnings. *The Journal of Political Economy, 75(4),* 352-465.

Blundell, R., & Preston, I. (1998). Consumption inequality and income uncertainity. *Quarterly Journal of Economics, 113(2),* 603-640.

Blundell, R., Dias, M. C., Meghir, C., & Shaw, J. (2016). Female labor supply, human capital and welfare reform . *Econometrica, 84(5),* 1705-1753.

Browning, M., Hansen, L. P., & Heckman, J. J. (1999). Micro data and general equilibrium models. In J. Taylor, & M. Woodford, *Handbook of Macroeconomics* (pp. 544-625). Amsterdam: North-Holland.

Carroll, C. D., & Samwick, A. A. (1997). The nature of precautionary wealth. *Journal of Monetary Economics,40(1),* 41-71.

Carroll, C. D., & Summers, L. H. (1991). Consumption growth parallels income growth: some new evidence. *National Saving and Economic Performance, ed. B. D. Bernheim and J. B. Shoven, University Press Chicago*, pp. 305-343.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics, 30,* 109-126.

Deaton, A. S., & Paxson, C. H. (1994). Intertemporal choice and inequality. *Journal of Political Economy, 102(3),* 438-467.

Deaton, A. S., & Paxson, C. H. (1995). Saving, Inequality and Aging: an East Asian perspective . *Asia-Pacific Economic Review, 1(1),* 7-19.

Deaton, A. S., & Paxson, C. H. (1997). The effects of economic and population growth on national saving and inquality. *Demography, 34(1),* 97-114.

Friedman, M. (1957). The permanent income hypothesis. *Theory of consumption function, ed. Friedman M., Princeton University Press for National Bureau of Economic Research* , pp. 20-37.

Guiso, L., Jappelli, T., & Pistaferri, L. (2002). An empirical analysis of earning and employment risk. *Journal of Business and Economic Statistics, 20(2),* 241-253.

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *The American Economic Review, 97,* 687-712.

Hamilton, J. D. (1994). *Time Series Analysis.* 41 William Street, Princeton, New Jersey 08540: Princeton University Press.

Hayashi, F. (2000). *Econometrics* . 41 William Street, Princeton, New Jersey 08540: Princeton University Press.

Karahan, F., & Ozkan, S. (2013). On the persistence of income shock over the life cycle: evidence, theory and implication. *Review of Economic Dynamics, 16*, 452-476.

LIS. (2017). *Luxembourg Income Study Database: http://www.lisdatacenter.org (multiple countries; data retrieved from September to October, 2017).* Luxembourg: LIS.

Lise, J., Sudo, N., Suzuki, M., Yamada, K., & Yamada, T. (2014). Wage, income and consumption inequality in Japan, 1981-2008: from boom to lost decades. *Review of Economic Dynamics, 17*, 582-612.

Meghir, C., & Pistaferri, L. (2004). Income variance dynamics and heterogeneity. *Econometrica*, 1-72.

Mincer, J. A. (1974). The human capital earnings function. In J. A. Mincer, *Schooling, Experience and Earnings* (pp. 83-96). Chicago: National Bureau of Economic Research.

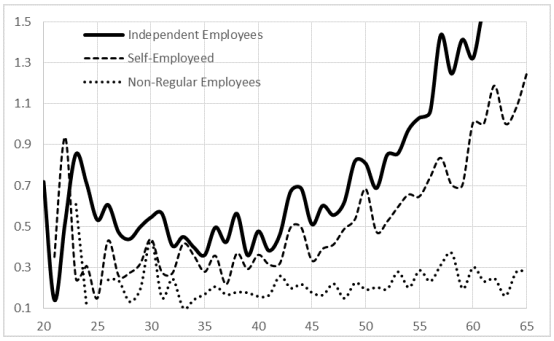
Storesletten, K., Telmer, C. I., & Yaron, A. (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics, 51*, 609-633.

Storesletten, K., Telmer, C. I., & Yaron, A. (2004a). Cyclical dynamic in idiosyncratic labor market risk. *Journal of Political Economy, 112(3)*, 695-717.

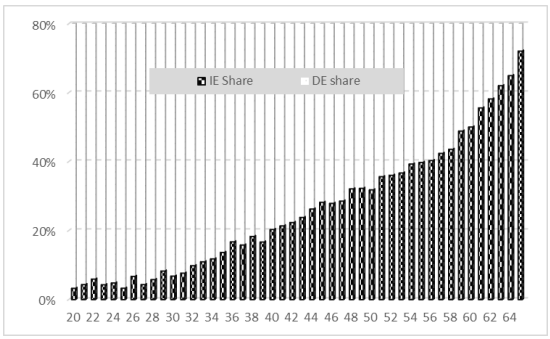
Appendices

A1 Taiwan's male Independent Employees in two groups

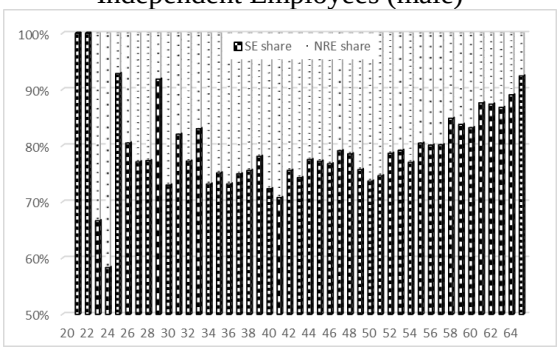
Total male workers



Independent Employees (male)



Share of IE's and DE's



Share of SE's and NRE's

Note. SE stands for self-employees and NRE stands for non-regular employees. The total number of IE's among the male workers are divided into two groups. Self-Employees and Non-Regular Employees. We show that the convex profile of IE's can be split into two straight lines, in a similar fashion that we split the total males into IE's and DE's. Moreover, in the top-right figure the overall variance is above both the profiles of sub-groups. From equation (9) we know that overall variance consists a component called intergroup variance, which is responsible for this. If the intergroup variance (or the difference between average earnings of the two groups) is too high then the overall variance profile stay above the spilt profiles, otherwise in between the two profiles. (More specifically when the minimum of one group is higher than the maximum of another group, the overall variance profile will stay above both the group profiles.)

A2 Sample Code to use data from LIS database for United States male workers

```
append using $us04p $us07p $us10p $us13p
keep if sex==1 & age>19
sum  hid pid year wave relation pil  oddjob_c  status1 status2 sector1 fyft ptime
tab sex  * Check if only male or not
gen worker=0
replace worker=1 if pil>0 & pil<.
tab year if worker==1 , sum(pil)
tab year  status1 if worker==1
gen labinc= .
replace labinc=pil if worker==1
gen y04=0
replace y04=1 if year==2004
gen y07=0
replace y07=1 if year==2007
gen y10=0
replace y10=1 if year==2010
gen y13=0
replace y13=1 if year==2013
reg labinc y04 y07 y10
gen norminc=.
replace norminc=labinc + 8658.355 if year==2004
replace norminc=labinc + 4067.524 if year==2007
replace norminc=labinc + 3902.875 if year==2010
replace norminc=labinc  if year==2013
sum norminc, det
replace norminc= 4514.5 if norminc<4514.5
replace norminc=293902.9 if norminc>293902.9 & norminc<.
gen loginc=.
replace loginc= log(norminc)
gen regwork=.
replace regwork=0 if worker==1
replace regwork=1 if worker==1 & status1<200
tab year regwork
tab age if regwork<. , sum(loginc)
tab age if regwork==0 , sum(loginc)
tab age if regwork==1 , sum(loginc)
```

A3 Transitory component (η_t) with individual fixed effect (α_t)

In fixed effect models with richer specification, one way to eliminate individual – specific fixed is to take first – difference. We take first – difference assuming that individual – specific fixed effect does not change from one period to another. Now when you use a parsimonious model (with one or two regressors) the method of first – difference may give us bias estimates for the model in this paper.

Why? Assume that we study a life-cycle model. Where the earning equation has only one regressor *age*. That implies the effects of all other possible regressors are explained by *age* only. Now consider the model of this paper. We segregate the earnings by types of employments, - that is, we partial out the effect of employment status. Then if we take first – difference the consequences leads to a bias estimate of fixed effect. Because – though individual fixed effect does not vary across time, it its highly likely to vary across employment status. Let us give an example: a university lecturer takes a leave for child care and during the leave period she works independently as a consultant. However, after the leave when she returns she gets her usual salary. Now if we take the first – difference between her working period and leave period, it gives a biased estimate of fixed effect.

Therefore, in this model, we assume that individual – fixed effects produce differential earning profile depending on employment status. In this case Deaton (1985) provides the arguments of unbiased and consistent estimates. We assume that the cohort population is constant conditional on employment status. Though the fixed effects, and transitory components cannot be estimated separately, they can be estimated together given that the sample size is sufficiently large.

A4 Dependent and Independent Employees

Dependent Employees – or simply employees: According to ILO classification for status in employment 'employees' constitute the first major group of ICSE (International Classification of Status in Employment). It is defined as (ICSE – 21) *Employees are defined as 'all those workers who hold the type of job defined as paid employment jobs.' Whilst more detailed categories of employees are not provided as a formal part of the classification, the definition of this group provides guidance on the definition of employees on stable contracts and of 'regular employees'.*

Independent Employees – workers other than Dependent Employees called independent employees in this paper. Which includes the rest four major groups: Employers, Own – account workers, Members of producers cooperatives, and Contributing family workers (see ICSE 23 – 25).