Is the Use of Autocovariances in Level the Best in Estimating the Income Processes? A Simulation Study

Chau, Tak Wai

Shanghai University of Finance and Economics

30 January 2013

Is the Use of Autocovariances in Level the Best in Estimating the Income Processes? A Simulation Study

Tak Wai Chau*
School of Economics
Shanghai University of Finance and Economics
777, Guoding Road, Yangpu District
Shanghai, 200433
China

Abstract

In this simulation study, I compare the efficiency and finite sample bias of parameter estimators for popular income dynamic models using various forms of autocovariances. The dynamic models have a random walk or a heterogeneous growth permanent component, a persistent autoregressive component and a white noise transitory component. I compare the estimators using autocovariances in level, first differences (FD), and autocovariances between level and future first differences (LD), where the last one is new in the literature of income dynamics. To maintain the same information used as in using level covariances, I also augment the FD and LD covariances with level variances in the estimation. The results show that using level covariances can give rise to larger finite sample biases and larger standard errors than using covariances in FD and LD augmented by level variance. Without augmenting the level variances, LD provides more efficient estimators than FD in estimating the non-permanent components. I also show that LD provides a convenient test between random walk and heterogeneous growth models with good power.

JEL codes: C33, C51, J31

Keywords: covariance structure, income dynamics, random walk, heterogeneous growth profile, finite sample bias, efficiency.

*Corresponding author: Email: zhou.dewei@mail.shufe.edu.cn. Tel: 86-21-65902119. Fax: 86-21-65904198

1 Introduction

Researchers estimate income dynamics models to understand the nature income shocks, how the variances of these shocks change over time (such as Moffitt and Gottschalk, 2012 and their previous work) and how they affect individual behavior such as consumption (such as Blundell, Pistaferri and Preston, 2008, Guvenen, 2007).[1] These models decompose the residual earnings into permanent and transitory components, and identify the variance of each component with panel data. The model parameters are usually estimated by Method of Moments through matching the theoretical and sample unconditional autocovariances of various lags. Researchers such as Moffitt and Gottschalk (2012) and Guvenen (2009) use the autocovariances in level, while other researchers, such as MaCurdy (1982), Baker (1997), and Hryshko (2012) use autocovariances in first difference (hereafter FD). However, there is a lack of research in comparing the performance of these estimators. This paper investigates the efficiency and finite sample bias with different forms of autocovariances through Monte-Carlo simulation.

I also investigate the use of covariances between current and future first differences (hereafter LD) in estimation. These autocovariances cancel out the random walk component, while keeping the information at level allows more efficient estimation for other components. Finally, I also augment these moments with the variances in level to add back the missing information for identifying the full model. My results show that LD can deliver estimators with smaller finite sample bias and lower variance, whereas using the common level autocovariances results in a substantial finite sample bias.

To shed light on the debate of whether the permanent component is driven by a random walk process or a heterogeneous growth process[2], I also consider a test by estimating a nested model with only LD covariances with the random walk model as the null hypothesis. This can control for the effect of the persistent AR component, and the results show that its power is reasonably high with good size under the null.

The remaining part of the paper is as follows. Section 2 describes the model and method-

[1] A detailed review of the use of income dynamics models in understanding income and consumption can be found in Meghir and Pistaferri (2011).

[2] See MaCurdy (1982), Baker (1997), Guvenen (2009) and Hryshko (2012).

ology. Section 3 presents the results. I briefly conclude in Section 4.

2 Model and Estimation Methods

2.1 Earnings Dynamics Model

Let y_{iat} be the earnings, income or wage of an individual in a year. The first step is to take away the observable component due to x_{iat} by OLS.

$$y_{iat} = x'_{iat}\beta + u_{iat} \tag{1}$$

x_{iat} usually include polynomial in age, year dummies and education. The second step is to model the dynamics of the residual, which is the deviation from the common profile. In this simulation study, I skip this first stage and focus on the estimators for the dynamics of the residual. A model with three independent additive components of different levels of persistency is used.[3]

$$u_{iat} = p_{iat} + v_{iat} + w_{iat} \tag{2}$$

where p_{iat} is a permanent component, v_{iat} is a persistent component, and w_{iat} is a transitory shock or measurement error.

Two types of permanent components are commonly used in the literature. One is a random walk process

$$p^R_{iat} = p^R_{i,a-1,t-1} + \varepsilon^p_{iat} \tag{3}$$

where ε^p_{iat} is the permanent shock for each period with variance $\sigma^2_{p\varepsilon}$. The other is a heterogeneous growth process (Guvenen, 2009, calls it Heterogeneous Income Profiles, HIP.)

$$p^H_{iat} = \theta_{1i} + \theta_{2i}a \tag{4}$$

where a is age normalized to zero at 25 and θ_{1i} and θ_{2i} are individual heterogeneous factors for initial value and growth rate of earnings with variances σ^2_1 and σ^2_2 respectively with correlation ρ_{12}.

[3] This follows from Guvenen (2009). Many other researchers use a two-component model instead, such as Moffitt and Gottschalk (2012).

For the persistent component, I use an autoregressive process of order 1,

$$v_{iat} = \rho v_{i,a-1,t-1} + \varepsilon^v_{iat} \tag{5}$$

where ε^v_{iat} is the persistent shock with variance $\sigma^2_{v\varepsilon}$, and ρ describes the degree of persistency in this component that satisfies $|\rho| < 1$. Finally, the transitory component w_{iat} is specified as independent and identically distributed over time with variance σ^2_w.

Following the literature, I consider the working life to start at the age 25 and end at 60. For the random walk permanent component, I allow an initial value, at the age 25, p_{i1} with variance σ^2_{p1}. I also allow a general initial value for the persistent autoregressive component v_{i1} with variance σ^2_{v1}.[4]

To simplify the problem, changes in parameters over calender years or birth cohorts are not considered in this simulation exercise. The parameter set to be estimated for a random walk permanent component model is $\Theta_1 = \{\sigma^2_{p1}, \sigma^2_{p\varepsilon}, \sigma^2_{v1}, \sigma^2_{v\varepsilon}, \rho, \sigma^2_w\}$ and the parameter set for a heterogeneous profile model is $\Theta_2 = \{\sigma^2_1, \sigma^2_2, \rho_{12}, \sigma^2_{v1}, \sigma^2_{v\varepsilon}, \rho, \sigma^2_w\}$. While ρ and ρ_{12} are restricted to lie between -1 and 1, all variance parameters are restricted to be non-negative.

2.2 Estimating Model Parameters

2.2.1 Estimation Method: Weighted Non-linear Least Squares

To estimate the dynamic model for u_{it}, the Method of Moments in the form of Classical Minimum Distance (CMD) (Chamberlain, 1984) is commonly used. This estimates the parameter by matching unconditional autocovariances from the data to those implied by the model using an appropriate weight matrix. Since Altonji and Segal (1996) showed that the finite sample properties of using an estimated optimal weight matrix is not desirable in this context, researchers generally use the identity weight matrix. The method can then be expressed in terms of (weighted) non-linear least squares.

First, I demonstrate the estimation method in level covariances. Collecting all possible

[4]In the previous literature, the initial value of this persistent component is often set to zero. However, because the start of the process in this model may not be the true starting point, and it is also plausible that the initial earnings contains a component that is persistent rather than permanent, I allow a general starting variance for the persistent component, though the results show that it is not always precisely estimated.

within-individual pairs of observations from the data, we can obtain an estimator by

$$\min_{\Theta} \sum_{a,s} w_{a,s} \left(\left(\frac{1}{n_{a,s}} \sum_{i} u_{ia} u_{i,a-s} \right) - m(\Theta, a, s) \right)^2 \tag{6}$$

where $u_{ia}u_{ia-s}$ include all available within-individual pairs for all individuals in the data and $m(\Theta, a, s) = Var(u_{ia}, u_{i,a-s}; \Theta)$ is the model moment function given the value of parameters, given by the sum of the covariances of the three independent additive components[5]. $n_{a,s}$ is the number of individuals with the corresponding age-and-lag pair in the sample. $w_{a,s}$ is a weight that may be used to improve efficiency depending on age (a) and lag between the two points (s). The following forms are considered. First, the commonly used identity weight matrix corresponds to $w_{a,s} = 1$ for all age and lags. However, since different age-lag combination contains different number of observations, so it would be more efficient to use $w_{a,s} = n_{a,s}$ to allow for a higher weight for moments calculated from more observations. Finally, we may also give a higher weight to the moments of lower variance. In the same spirit as the Diagonally Weighted Minimum Distance (DWMD) of Blundell, Pistaferri and Preston (2008), the third type of weight is to use the inverse of the within age-lag cell variance:

$$w_{a,s} = \frac{n_{a,s}}{\left[\widehat{Var}(u_{i,a}u_{i,a-s} - m(\Theta; a, s))\right]} = \frac{n_{a,s}}{\widehat{Var}(u_{i,a}u_{i,a-s})}. \tag{7}$$

This can help bringing autocovariances from different moments to be more comparable in size when we mix different types of moments in some specifications.

2.2.2 Covariances Used: Level, First Difference and Level-First-Difference

Besides using level u_{it}, another form is to use first differences $\Delta u_{ia} = u_{ia} - u_{i,a-1}$ (FD). Then the objective function becomes

$$\min_{\Theta} \sum_{a,s} w_{a,s} \left(\left(\frac{1}{n_{a,s}} \sum_{i} \Delta u_{ia} \Delta u_{i,a-s} \right) - m^{FD}(\Theta, a, s) \right)^2 \tag{8}$$

where m^{FD} is the corresponding true moment function, $s \geq 0$. The use of autocovariances in FD has the advantage that it can extract the random walk shocks without regards to the past level, and so, for lagged autocovariances, permanent shocks are uncorrelated.

[5]Formulas are available in the Appendix A.

I also use the covariance between level and future first difference (LD)

$$cov(u_{ia-s}, \Delta u_{ia})$$

for $s \geq 1$. The estimator is obtained by

$$\min_{\Theta} \sum_{a,s} w_{a,s} \left(\left(\frac{1}{n_{a,s}} \sum_{i} u_{ia-s} \Delta u_{ia} \right) - m^{LD}(\Theta, a, s) \right)^2 \tag{9}$$

where m^{LD} is the corresponding true moment function. The main advantage is that the future shocks does not depend on the past value in the random walk component, and so we can drop totally the variance of random walk shock, but yet, it maintains the information in level that enables the estimation of the parameters in the persistent and transitory components more precisely.

Finally, to add back the required information to identify the whole model and improve efficiency in estimating existing parameters, the variances in level are augmented to the FD and LD moments.[6] That means the following term is added to the objective function:

$$\sum_{a} \tilde{w}_a \left(\left(\frac{1}{\tilde{n}_a} \sum_{i} u_{ia}^2 \right) - \tilde{m}(\Theta, a) \right)^2 \tag{10}$$

2.3 Estimation for the Test

I also consider a simple test of random walk model against heterogeneous growth model for the permanent component. Here I do not use the full model, because, as shown in Appendix B, the age profile for level variance are the same between these two models if we allow the shocks or growth rate changes over the lifecycle. So I focus on the most important distinguishing feature about the covariances of longer lags.

Using LD moments only, under the random walk model, $cov(p_{iat}^R, \Delta p_{i,a+s,t+s}^R) = 0$ for all $s > 0$, because the future shocks are all uncorrelated to previous levels. However, for the heterogeneous growth model, $cov(p_{iat-s}^H, \Delta p_{i,a,t}^H) = (a - s)\sigma_2^2$, which is non-zero for all lags s. Therefore, we can estimate the heterogeneous growth model (also a nested model)

[6] In principle, using level covariances and the FD and LD augmented with level variances contain the same information in that using level variances plus the FD or LD covariances can solve for covariances at level for all corresponding lags. However, due to over-identification and a non-linear (sum of squares) criterion function, the results may differ substantially.

and test if $\sigma_2^2 = 0$ against $\sigma_2^2 > 0$. To avoid non-standard distribution under the null, I drop the non-negative restriction of the estimator of σ_2^2. Here, I apply a one-tail test and the null of random walk is rejected when the t-value is above 1.645. I have tried a similar test using FD moments, but some parameter estimators become unstable because the FD moments do not have enough information to estimate the persistent component precisely, so the performance is not good.[7]

3 Simulation Results

3.1 Results for Parameter Estimation

In the simulation exercise, I simulate u_{it} according to the true model with N individuals from the age of 25 to 60. I then extract the segment of T years observed, and apply the above estimation methods to obtain the parameter estimates. T years of data for each individual are used, where I randomly assign the starting observed age from uniform distribution between 25 and $60 - T + 1$.[8] I repeat this process 5000 times and calculate the means and standard deviations of the parameter estimates among these simulations. I assume that all shocks and initial conditions are normally distributed with mean zero with the corresponding variances. The parameters used to simulate the data are chosen with reference to Guvenen (2009) and Moffitt and Gottschalk (2012).

Table 1 reports the results for models with a random walk permanent component under different weighting schemes for all five sets of covariances. The baseline I use is $N = 3000$ and $T = 10$. The three panels show the results from unweighted, weighted by number of observations used, and weighted by number of observations divided by cell variance. The third weighting scheme gives the lowest finite sample bias and variances, especially for the specifications that mix level variance with FD or LD covariances. Table 2 shows the results of different sample length, number of individuals and different parameter values, with the third weighting scheme applied. The use of usual level autocovariances actually results in higher finite sample biases and less efficient estimators, especially for the AR persistent

[7]Results are available upon request.

[8]The age structure in the data may differ, but this setting should be useful for understanding relative performance.

parameter ρ. Using FD or LD augmented with level variance gives lower finite sample biases and standard deviations. Moreover, using only FD results in more noisy estimators of parameters, while using only LD gives us estimator for persistent and transitory components as precise as augmenting these moments by level variances. These findings are robust across variations shown in Table 2.

Table 3 and 4 show the analogous results for models with a heterogeneous growth permanent component. Table 3 shows the results for different weighting schemes. Table 4 shows the results for varying sample structures and parameter values under the third weighting scheme. The results are very similar to those of the random walk model. There is a general tendency that LD gives us more efficient estimators with close to zero bias. However, the advantage is smaller for LD-only versus FD-only under the heterogeneous growth model, except for the initial variance parameter of the transitory component. On the other hand, using only FD covariances can pin down the variance of the heterogeneous growth rate rather precisely when the panel length is long.

In summary, using level autocovariances results in larger finite sample biases and standard errors, while using LD covariances essentially removes this bias and generally provides more efficient estimators.

3.2 Results for Test between Models

Table 5 shows the rejection probability of the test of random walk model against heterogeneous growth model of various data structure, parameter values and models. I use the standard deviations across simulations as the standard error.[9] Under all specifications, the rejection probabilities of random walk models are close to the nominal size of 5%. For heterogeneous growth model, in many cases the rejection probabilities exceed 0.5 and sometimes even close to 1, unless the sample size and length are both small, or the heterogeneous growth rate variance is small. A higher persistent ρ in the AR component reduces the power of the test.

[9] In actual data, we have to use an estimator for the standard error, such as bootstrap, that involves more sampling variations, but this should not change the results substantially.

4 Conclusions

In the simulation study, I find that using level autocovariances is not the best in terms of reducing finite sample bias and standard errors. The use of LD covariances perform the best in these two aspects, so I recommend researchers to use LD covariances in empirically estimating income dynamic models. This paper also introduces a reasonably powerful test to distinguish between random walk and heterogeneous growth permanent component using the LD covariances.

References

[1] Altonji, J.G. and L.M. Segal (1996) "Small-Sample Bias in GMM Estimation of Covariance Structures." *Journal of Business and Economic Statistics*, 14(3), 353-386.

[2] Baker, M. (1997) "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings." *Journal of Labor Economics*, 15(2), 338-375.

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

[4] Chamberlain, G. (1984) "Panel Data." In: Griliches, Z., Intriligator, M. (Eds.), *Handbook of Econometrics*, Volume 1. Amsterdam: North-Holland, 1274-1318.

[5] Guvenen, F. (2007) "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review*, 97(3), 687-712.

[6] Guvenen, F. (2009) "An Empirical Investigation of Labor Income Processes." *Review of Economic Dynamics*, 12, 58-79.

[7] Hryshko, D. (2012) "Labor Income Profiles Are Not Heterogeneous: Evidence from Income Growth Rates." *Quantitative Economics*, 3, 177-209.

[8] MaCurdy, T., (1982) "The Use of Time-Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics*, 18, 83-114.

[9] Meghir, C. and L. Pistaferri (2011) "Earnings, Consumption and Life Cycle Choices." In: Ashenfelter, O. and D. Card (Eds.), *Handbook of Labor Economics*, 4B, 773-854.

[10] Moffitt, R and P. Gottschalk (2012) "Trends in the Transitory Variance of Male Earnings: Methods and Evidence." *Journal of Human Resources*, 47(1), 204-236.

Tables

Table 1: Means and Standard Deviations Across Simulations for Random Walk Model with Various Weights

	σ^2_{p1}	$\sigma^2_{p\epsilon}$	ρ	σ^2_{v1}	$\sigma^2_{v\epsilon}$	σ^2_{w}
True Values	0.1	0.012	0.8	0.1	0.05	0.05
Unweighted						
Level	0.0598	0.0112	0.8538	0.1364	0.0425	0.0588
	(0.0470)	(0.0030)	(0.0732)	(0.0612)	(0.0109)	(0.0120)
FD Only		0.0154	0.7140	0.1006	0.0504	0.0462
		(0.0126)	(0.1990)	(0.0950)	(0.0098)	(0.0115)
FD + Level Variance	0.0668	0.0115	0.8184	0.1312	0.0479	0.0513
	(0.0575)	(0.0028)	(0.0966)	(0.0664)	(0.0069)	(0.0049)
LD Only			0.8001	0.1057	0.0504	0.0500
			(0.0516)	(0.0564)	(0.0056)	(0.0034)
LD + Level Variance	0.0838	0.0116	0.8157	0.1146	0.0480	0.0513
	(0.0415)	(0.0024)	(0.0656)	(0.0491)	(0.0068)	(0.0050)
Weighted by number of observations only						
Level	0.0596	0.0114	0.8530	0.1409	0.0432	0.0572
	(0.0470)	(0.0027)	(0.0672)	(0.0704)	(0.0088)	(0.0089)
FD Only		0.0126	0.7713	0.1121	0.0502	0.0491
		(0.0096)	(0.1014)	(0.1006)	(0.0077)	(0.0045)
FD + Level Variance	0.0583	0.0116	0.8371	0.1421	0.0479	0.0517
	(0.0532)	(0.0022)	(0.0717)	(0.0674)	(0.0053)	(0.0036)
LD Only			0.8005	0.1012	0.0500	0.0500
			(0.0346)	(0.0327)	(0.0029)	(0.0022)
LD + Level Variance	0.0786	0.0117	0.8259	0.1211	0.0474	0.0518
	(0.0414)	(0.0020)	(0.0570)	(0.0538)	(0.0056)	(0.0040)
Weighted by number of observations times inverse of variance of each age-lag cell						
Level	0.0592	0.0114	0.8547	0.1372	0.0445	0.0557
	(0.0427)	(0.0021)	(0.0596)	(0.0605)	(0.0083)	(0.0082)
FD Only		0.0126	0.7784	0.1049	0.0500	0.0490
		(0.0092)	(0.0898)	(0.0831)	(0.0076)	(0.0038)
FD + Level Variance	0.0923	0.0118	0.8075	0.1039	0.0501	0.0497
	(0.0246)	(0.0014)	(0.0380)	(0.0379)	(0.0028)	(0.0021)
LD Only			0.7986	0.0955	0.0499	0.0495
			(0.0349)	(0.0325)	(0.0031)	(0.0022)
LD + Level Variance	0.0984	0.0119	0.8009	0.0972	0.0496	0.0497
	(0.0240)	(0.0012)	(0.0360)	(0.0350)	(0.0033)	(0.0024)

Note: N=3000 and T=10 are used in the above specifications. I repeat the simulation 5000 times. Means across simulations are reported with standard deviations across simulations in the parenthesis. FD stands for using autocovariances of first difference. LD stands for using covariances between level and future first differences.

Table 2: Means and Standard Deviations Across Simulations for Random Walk Model with Different Sample Sizes and Parameter Values

	σ^2_{p0}	$\sigma^2_{p\epsilon}$	ρ	σ^2_{v1}	$\sigma^2_{v\epsilon}$	σ^2_{w}
True Value (Default)	0.1	0.012	0.8	0.1	0.05	0.05
Smaller Number of Observations: $N = 1500, T = 10$						
Level	0.0436	0.0109	0.8706	0.1513	0.0440	0.0565
	(0.0442)	(0.0031)	(0.0609)	(0.0753)	(0.0094)	(0.0093)
FD Only		0.0144	0.7436	0.1057	0.0500	0.0467
		(0.0117)	(0.1586)	(0.1065)	(0.0093)	(0.0093)
FD + Level Variance	0.0845	0.0116	0.8121	0.1090	0.0502	0.0494
	(0.0379)	(0.0021)	(0.0567)	(0.0557)	(0.0040)	(0.0031)
LD Only			0.7973	0.0941	0.0499	0.0491
			(0.0505)	(0.0552)	(0.0047)	(0.0033)
LD + Level Variance	0.0953	0.0116	0.8032	0.09676	0.0492	0.0495
	(0.0354)	(0.0019)	(0.0530)	(0.0500)	(0.0048)	(0.0036)
Longer Panels: $N = 3000, T = 20$						
Level	0.0970	0.0119	0.8110	0.0996	0.0480	0.0523
	(0.0113)	(0.0013)	(0.0301)	(0.0394)	(0.0047)	(0.0054)
FD Only		0.0123	0.7969	0.1000	0.0498	0.0498
		(0.0028)	(0.0276)	(0.0535)	(0.0022)	(0.0014)
FD + Level Variance	0.0981	0.0119	0.8027	0.0997	0.0500	0.0499
	(0.0130)	(0.0010)	(0.0199)	(0.0277)	(0.0018)	(0.0013)
LD Only			0.8000	0.0960	0.0499	0.0498
			(0.0157)	(0.0233)	(0.0017)	(0.0013)
LD + Level Variance	0.1002	0.0119	0.8006	0.0970	0.0498	0.0499
	(0.0136)	(0.0009)	(0.0161)	(0.0253)	(0.0018)	(0.0013)
Lower Persistence in AR component: $\rho = 0.5, \sigma^2_{v\epsilon} = 0.1$. $N = 3000, T = 10$						
Level	0.0874	0.0117	0.5856	0.0991	0.0842	0.0644
	(0.0270)	(0.0016)	(0.1289)	(0.0519)	(0.0236)	(0.0222)
FD Only		0.0132	0.4890	0.1020	0.1008	0.0474
		(0.0047)	(0.0703)	(0.0587)	(0.0081)	(0.0100)
FD + Level Variance	0.0988	0.0120	0.5023	0.0994	0.1003	0.04907
	(0.0134)	(0.0010)	(0.0482)	(0.0351)	(0.0083)	(0.0077)
LD Only			0.4991	0.0962	0.0999	0.0491
			(0.0424)	(0.0330)	(0.0078)	(0.0071)
LD + Level Variance	0.0999	0.0120	0.5007	0.0973	0.0996	0.0493
	(0.0143)	(0.0010)	(0.0432)	(0.0323)	(0.0079)	(0.0072)
Lower Variance in Persistent Shock: $\sigma^2_{p\epsilon} = 0.006$. $N = 3000, T = 10$						
Level	0.0668	0.0056	0.8470	0.1309	0.0451	0.0549
	(0.0390)	(0.0014)	(0.0544)	(0.0563)	(0.0073)	(0.0072)
FD Only		0.0079	0.7776	0.0941	0.0487	0.0491
		(0.0072)	(0.0799)	(0.0687)	(0.0057)	(0.0034)
FD + Level Variance	0.0928	0.0059	0.8075	0.1039	0.0500	0.0497
	(0.0224)	(0.0010)	(0.0347)	(0.0353)	(0.0025)	(0.0020)
LD Only			0.7985	0.0950	0.0499	0.0495
			(0.0302)	(0.0301)	(0.0027)	(0.0021)
LD + Level Variance	0.0991	0.0059	0.8004	0.0967	0.0497	0.0497
	(0.0201)	(0.0009)	(0.0316)	(0.0319)	(0.0029)	(0.0022)

Note: The benchmark is the parameter values on the first row and $N = 3000$, $T = 10$, and the weights are the number of observations times inverse of variance of each age-lag cell. I repeat the simulation 5000 times. Means across simulations are reported with standard deviations across simulations in the parentheses. FD stands for using covariances in first difference. LD stands for using covariances between level and future first differences.

Table 3: Means and Standard Deviations Across Simulations for Heterogeneous Growth Model Under Various Weights

	σ_1^2	$\sigma_2^2 \times 100$	ρ_{12}	ρ	σ_{v1}^2	$\sigma_{v\epsilon}^2$	σ_w^2
True Value	0.1	0.03	0.0	0.8	0.1	0.05	0.05
Unweighted							
Level	0.1073	0.0341	0.0883	0.8274	0.0924	0.0456	0.0541
	(0.0506)	(0.0178)	(0.5936)	(0.0583)	(0.0656)	(0.0094)	(0.0087)
FD Only		0.0398		0.7791	0.1000	0.0504	0.0493
		(0.0361)		(0.0574)	(0.0842)	(0.0029)	(0.0028)
FD + Level Variance	0.1056	0.0355	-0.0108	0.8053	0.0960	0.0493	0.0502
	(0.0484)	(0.0207)	(0.6611)	(0.0873)	(0.0572)	(0.0042)	(0.0035)
LD Only		0.0371		0.8031	0.1034	0.0538	0.0501
		(0.0379)		(0.0541)	(0.0355)	(0.0220)	(0.0030)
LD + Level Variance	0.1023	0.0316	0.1197	0.8029	0.0976	0.0494	0.0503
	(0.0412)	(0.0142)	(0.4991)	(0.0437)	(0.0494)	(0.0048)	(0.0034)
Weighted by number of observations only							
Level	0.1010	0.0329	0.1338	0.8248	0.1014	0.0466	0.0531
	(0.0491)	(0.0161)	(0.5938)	(0.0513)	(0.0721)	(0.0072)	(0.0063)
FD Only		0.0346		0.7926	0.1039	0.0502	0.0497
		(0.0292)		(0.0437)	(0.0816)	(0.0024)	(0.0022)
FD + Level Variance	0.0929	0.0337	0.0500	0.8187	0.1093	0.0492	0.0506
	(0.0457)	(0.0173)	(0.6529)	(0.0656)	(0.0586)	(0.0032)	(0.0026)
LD Only		0.0312		0.7994	0.1014	0.0506	0.0499
		(0.0157)		(0.0356)	(0.0296)	(0.0061)	(0.0019)
LD + Level Variance	0.0989	0.0310	0.1345	0.8083	0.1016	0.0491	0.0506
	(0.0393)	(0.0129)	(0.4965)	(0.0359)	(0.0524)	(0.0038)	(0.0026)
Weighted by number of observations times inverse of variance of each age-lag cell							
Level	0.0960	0.0335	0.0788	0.8363	0.1008	0.0460	0.0535
	(0.0467)	(0.0149)	(0.5684)	(0.0447)	(0.0666)	(0.0063)	(0.0057)
FD Only		0.0364		0.7948	0.0974	0.0501	0.0494
		(0.0303)		(0.0437)	(0.0607)	(0.0023)	(0.0022)
FD + Level Variance	0.0981	0.0302	0.1221	0.8026	0.0987	0.0500	0.0496
	(0.0340)	(0.0110)	(0.4381)	(0.0282)	(0.0456)	(0.0024)	(0.0019)
LD Only		0.0312		0.7991	0.0957	0.0504	0.0496
		(0.0143)		(0.0366)	(0.0294)	(0.0053)	(0.0020)
LD + Level Variance	0.0997	0.0299	0.1025	0.7999	0.0963	0.0497	0.0496
	(0.0331)	(0.0095)	(0.3855)	(0.0279)	(0.0415)	(0.0027)	(0.0020)

Note: N=3000 and T=10 are used in the above specifications. I repeat the simulation 5000 times. Means across simulations are reported with standard deviations across simulations in the parenthesis. FD stands for using covariances in first difference. LD stands for using covariances between level and future first differences.

Table 4: Means and Standard Deviations Across Simulations for Heterogeneous Growth Model with Different Parameter Values

	σ_1^2	$\sigma_2^2 \times 100$	ρ_{12}	ρ	σ_{v1}^2	$\sigma_{v\epsilon}^2$	σ_w^2
True Value (Default)	0.1	0.03	0.0	0.8	0.1	0.05	0.05
Smaller Number of Observations: $N = 1500, T = 10$							
Level	0.0950	0.0359	0.0650	0.8532	0.1001	0.0449	0.0543
	(0.0606)	(0.0188)	(0.6912)	(0.0546)	(0.0850)	(0.0077)	(0.0072)
FD Only		0.0426		0.7896	0.1002	0.0502	0.0488
		(0.0392)		(0.0610)	(0.0832)	(0.0034)	(0.0032)
FD + Level Variance	0.0985	0.0308	0.1760	0.8042	0.0952	0.0499	0.0492
	(0.0432)	(0.0149)	(0.5544)	(0.0417)	(0.0598)	(0.0034)	(0.0027)
LD Only		0.0344		0.7990	0.0924	0.0523	0.0492
		(0.0320)		(0.0541)	(0.0482)	(0.0297)	(0.0029)
LD + Level Variance	0.1012	0.0302	0.1603	0.7990	0.0909	0.0494	0.0493
	(0.0432)	(0.0128)	(0.5033)	(0.0401)	(0.0554)	(0.0039)	(0.0029)
Longer Panels: $N = 3000, T = 20$							
Level	0.0969	0.0297	0.1043	0.8045	0.0998	0.0492	0.0507
	(0.0307)	(0.0084)	(0.3528)	(0.0229)	(0.0445)	(0.0033)	(0.0030)
FD Only		0.0305		0.7999	0.0999	0.0500	0.0498
		(0.0056)		(0.0164)	(0.0429)	(0.0015)	(0.0012)
FD + Level Variance	0.0984	0.0298	0.0658	0.8009	0.0999	0.0499	0.0498
	(0.0229)	(0.0072)	(0.2828)	(0.0152)	(0.0335)	(0.0015)	(0.0012)
LD Only		0.0306		0.7997	0.0958	0.0502	0.0498
		(0.0090)		(0.0202)	(0.0211)	(0.0029)	(0.0011)
LD + Level Variance	0.0990	0.0298	0.0493	0.8003	0.0983	0.0498	0.0499
	(0.0225)	(0.0059)	(0.2279)	(0.0145)	(0.0302)	(0.0016)	(0.0012)
Smaller Persistence at AR Component: $\rho = 0.5, \sigma_{v\epsilon}^2 = 0.1$							
Level	0.0958	0.0298	0.0938	0.5402	0.0962	0.0917	0.0571
	(0.0236)	(0.0094)	(0.3548)	(0.0568)	(0.0505)	(0.0127)	(0.0110)
FD Only		0.0355		0.4994	0.1013	0.1004	0.0489
		(0.0196)		(0.0463)	(0.0556)	(0.0080)	(0.0075)
FD + Level Variance	0.0976	0.0294	0.0896	0.5029	0.1003	0.1000	0.0493
	(0.0211)	(0.0086)	(0.3185)	(0.0437)	(0.0364)	(0.0077)	(0.0072)
LD Only		0.0302		0.4992	0.0964	0.0999	0.0492
		(0.0107)		(0.0471)	(0.0323)	(0.0070)	(0.0071)
LD + Level Variance	0.0986	0.0294	0.0757	0.5020	0.0982	0.0993	0.0497
	(0.0207)	(0.0078)	(0.2879)	(0.0366)	(0.0333)	(0.0071)	(0.0064)
Lower Spread in Growth Rates: $\sigma_2^2 \times 100 = 0.015$							
Level	0.0950	0.0177	0.1079	0.8322	0.0996	0.0464	0.0532
	(0.0389)	(0.0111)	(0.5949)	(0.0404)	(0.06012)	(0.0057)	(0.0052)
FD Only		0.0258		0.7920	0.0951	0.0501	0.0493
		(0.0260)		(0.0411)	(0.0589)	(0.0023)	(0.0021)
FD + Level Variance	0.0990	0.0154	0.1543	0.8026	0.0968	0.0500	0.0496
	(0.0292)	(0.0084)	(0.4958)	(0.0280)	(0.0416)	(0.0023)	(0.0019)
LD Only		0.0170		0.8005	0.0954	0.0506	0.0496
		(0.0123)		(0.0327)	(0.0283)	(0.0048)	(0.0019)
LD + Level Variance	0.1009	0.0152	0.1409	0.7990	0.0945	0.0497	0.0496
	(0.0297)	(0.0074)	(0.4534)	(0.0261)	(0.0384)	(0.0025)	(0.0019)

Note: The benchmark is the parameter values on the first row and $N = 3000$, $T = 10$, and the weights are the number of observations times inverse of variance of each age-lag cell. I repeat the simulation 5000 times. Means across simulations are reported with standard deviations across simulations in the parentheses. FD stands for using covariances in first difference. LD stands for using covariances between level and future first differences.

Table 5: Rejection Probability of the Null of Random Walk Under Various True Models

	Sample Sizes				
N	3000	1500	3000	1500	10000
T	20	20	10	10	10
Data Generated from					
	$\rho = 0.8$, $\sigma^2_{v\varepsilon} = 0.05$				
Random Walk	0.069	0.067	0.052	0.034	0.064
Heterogeneous Growth	0.976	0.761	0.676	0.279	0.994
	$\rho = 0.5$, $\sigma^2_{v\varepsilon} = 0.1$				
Random Walk	0.051	0.061	0.060	0.063	0.054
Heterogeneous Growth	1.000	0.998	0.878	0.592	0.999
	Half Permanent Change				
Random Walk	0.067	0.067	0.072	0.042	0.061
Heterogeneous Growth	0.610	0.374	0.332	0.139	0.692

Note: The nominal size of tests is 5%. For the parameters of the data generating process, the benchmark is $\sigma^2_{p1} = 0.1$, $\sigma^2_{p\varepsilon} = 0.012$, $\sigma^2_1 = 0.1$, $\sigma^2_2 = 0.0003$, $\rho_{12} = 0$, $\sigma^2_{v1} = 0.1$, $\sigma^2_w = 0.05$. For specifications with $\rho = 0.8$, $\sigma^2_{v\varepsilon} = 0.05$. For specifications with $\rho = 0.5$, $\sigma^2_{v\varepsilon} = 0.1$. We half the size of σ^2_2 and $\sigma^2_{p\varepsilon}$ for the third panel. The test is performed by estimating the heterogeneous growth model using LD covariances, but not to restrict the growth variance to be positive, and then perform a one-tail test on this parameter under the null of zero for the random walk model against a positive value for the heterogeneous growth model. The above report the probability of rejection using T test, using standard deviation across all simulations as the standard error.

Appendix

A Theoretical Covariances of Different Sets of Moment Conditions

Consider Error Component Model

$$u_{it} = p_{it} + v_{it} + w_{it}$$

where p_{it} is the permanent component, $v_{it}$is a persistent component that dies down relatively slowly, and $w_{it}$is a short memory component. Here I take it as an independent component over time.

A.1 Permanent component

I have two types of permanent component: random walk and heterogeneous profile.

A.1.1 Random Walk Model

$$p_{iat} = p_{i,a-1,t-1} + \epsilon_{i,a,t}$$

where a is age-24.

For variance of level, we have

$$Var(p_{iat}) = Var(p_{i,a-1,t-1}) + Var(\epsilon_{i,a,t})$$

or

$$\sigma^2_{p,a,t} = \sigma^2_{p,a-1,t-1} + \sigma^2_{p\epsilon,a,t}$$

with initial condition $Var(p_{i,1,t}) = \sigma^2_{p0}$. Thus the variances can be calculated recursively. The covariance is then given by

$$cov(p_{iat}, p_{ia-s,t-s}) = cov(p_{i,a-s,t-s} + \sum_{s'=0}^{s-1} \epsilon_{i,a-s',t-s'}, p_{i,a-s,t-s}) = Var(p_{i,a-s,t-s})$$

since new shocks are uncorrelated to old values.

Then, consider first difference

$$\Delta p_{iat} = p_{iat} - p_{i,a-1,t-1} = \epsilon_{i,a,t}$$

Then,

$$Var(\Delta p_{iat}) = Var(\epsilon_{i,a,t}) = \sigma_{p\epsilon}^2$$

and

$$cov(\Delta p_{iat}, \Delta p_{i,a-s,t-s}) = cov(\epsilon_{i,a,t}, \epsilon_{i,a-s,t-s}) = 0$$

for $s \geqslant 1$

Then, consider the covariance between level and future first difference

$$cov(p_{iat}, \Delta p_{i,a+s,t+s}) = cov(p_{iat}, \epsilon_{i,a+s,t+s}) = 0$$

where $s \geqslant 1$. This is true because future shocks are uncorrelated to previous shocks and thus realized values by definition.

A.1.2 Heterogeneous Growth Profile

Using the usual parameterization,

$$p_{iat} = \theta_1 + \theta_2 a$$

where across individuals, $E(\theta_1) = E(\theta_2) = 0$ and $Var(\theta_1) = \sigma_1^2$ and $Var(\theta_2) = \sigma_2^2$ and $cov(\theta_1, \theta_2) = \sigma_{12} = \rho_{12}\sigma_1\sigma_2$.

If using level,

$$Var(p_{iat}) = Var(\theta_1) + a^2 Var(\theta_2) + 2acov(\theta_1, \theta_2) = \sigma_1^2 + \sigma_2^2 a^2 + 2a\sigma_{12}$$

and

$$cov(p_{iat}, p_{ia-s,t-s}) = cov(\theta_1 + \theta_2 a, \theta_1 + \theta_2(a-s)) = \sigma_1^2 + \sigma_2^2 a(a-s) + (a + (a-s))\sigma_{12}$$

If we use first difference, then

$$\Delta p_{iat} = p_{iat} - p_{i,a-1,t-1} = \theta_2$$

Therefore,

$$Var(\Delta p_{iat}) = Var(\theta_2) = \sigma_2^2$$

and

$$cov(\Delta p_{iat}, \Delta p_{i,a-s,t-s}) = cov(\theta_2, \theta_2) = \sigma_2^2$$

Finally, if we use covariance between level and future first difference, we will use

$$cov(p_{iat}, \Delta p_{i,a+s,t+s}) = cov(\theta_1 + \theta_2 a, \theta_2) = \sigma_{12} + a\sigma_2^2$$

For estimating a complete model, putting σ_{12} as $\sigma_1\sigma_2\rho_{12}$ may be useful to impose appropriate restrictions on the correlation coefficient. But if σ_1 cannot be identified independently, but σ_{12} is involved, it may bring about some unreasonable extreme estimates. In some parts of my paper, I have dropped this part and assume θ_1 and θ_2 are independent, because the identification of the covariance between the two is indeed weak.

A.2 Persistent Component

The persistent component is represented by an AR(1) process,

$$v_{iat} = \rho v_{i,a-1,t-1} + \epsilon_{iat}^v$$

where $-1 < \rho < 1$. I do not assume it to be stationary, so there is an initial variance

$$Var(v_{i,1,t}) = \sigma_{v,1,t}^2$$

and its unconditional variance evolves according to

$$Var(v_{i,a,t}) = \rho^2 Var(v_{i,a-1,t-1}) + Var(\epsilon_{iat}^v) = \rho^2\sigma_{v,a-1,t-1}^2 + \sigma_{\epsilon v,a,t}^2$$

Then, the covariance is given by

$$cov(v_{i,a,t}, v_{i,a-s,t-s}) = cov(\rho^s v_{i,a-s,t-s} + \sum_{s'=0}^{s-1} \rho^{s'} \epsilon_{i,a-s',t-s'}, v_{i,a-s,t-s}) = \rho^s \sigma_{v,a-s,t-s}^2$$

For first differences, the following formulation is the most useful,

$$\Delta v_{i,a,t} = (\rho - 1) v_{i,a-1,t-1} + \epsilon_{iat}^v$$

in which the two terms on the right-hand side are uncorrelated by definition. So,

$$Var(\Delta v_{iat}) = (\rho - 1)^2 Var(v_{i,a-1,t-2}) + Var(\epsilon_{iat}^v) = (\rho - 1)^2 \sigma_{v,a-1,t-1}^2 + \sigma_{\epsilon v,a,t}^2$$

and

$$\begin{aligned} cov(\Delta v_{iat}, \Delta v_{i,a-s,t-s}) &= cov((\rho-1)v_{i,a-1,t-1} + \epsilon^v_{iat}, v_{i,a-s,t-s} - v_{i,a-s-1,t-s-1}) \\ &= cov((\rho-1)\rho^{s-1} v_{i,a-s,t-s} + ..., v_{i,a-s,t-s} - v_{i,a-s-1,t-s-1}) \\ &= \rho^{s-1}(\rho-1)\left[\sigma^2_{v,a-s,t-s} - \rho\sigma^2_{v,a-s-1,t-s-1}\right] \end{aligned}$$

where the second line we omit the future shocks that are uncorrelated to past v, and in the last line, we apply the above covariance results for $s = 1$. Moreover,

$$\begin{aligned} cov(v_{iat}, \Delta v_{i,a+s,t+s}) &= cov(v_{iat}, \rho^{s-1}(\rho-1)v_{iat} + ...) \\ &= \rho^{s-1}(\rho-1)\sigma^2_{v,a,t} \end{aligned}$$

A.3 Transitory Component

Here I assume the transitory component follows identically and independently distributed shocks. This may be measurement errors or very transitory shocks. So,

$$Var(w_{iat}) = \sigma^2_{w,a,t}$$

and

$$cov(w_{iat}, w_{i,a-s,t-s}) = 0$$

and

$$Var(\Delta w_{iat}) = Var(w_{iat} - w_{i,a-1,t-1}) = \sigma^2_{w,a,t} + \sigma^2_{w,a-1,t-1}$$

and

$$cov(\Delta w_{iat}, \Delta w_{i,a-1,t-1}) = cov(w_{iat} - w_{i,a-1,t-1}, w_{i,a-1,t-1} - w_{i,a-2,t-2}) = -\sigma^2_{w,a-1,t-1}$$

and

$$cov(\Delta w_{iat}, \Delta w_{i,a-s,t-s}) = 0$$

for $s > 1$, and

$$cov(w_{iat}, \Delta w_{i,a+1,t+1}) = -\sigma^2_{w,a,t}$$

and

$$cov(w_{iat}, \Delta w_{i,a+s,t+s}) = 0$$

for $s > 1$.

B Variance Profile for the Two Models of Permanent Components

Here I would like to show that the two types of permanent components in the basic form imply different shapes of variance profile, but if we extend the basic form of model to have age-varying shocks or growth rates, they indeed can have the same implication in the shape of variance profile. Therefore, in the tests in this paper, I do not use level covariances that involves the variance profile to test between the two main types of models, and restrict our attention to the long term covariances involving first difference.

Recall that the simple form of random walk model takes the form

$$p^R_{iat} = p^R_{i,a-1,t-1} + \epsilon_{i,a,t}$$

and the variance takes the recursive formula

$$\begin{aligned} Var(p^R_{iat}) &= Var(p^R_{i,a-1,t-1}) + Var(\epsilon_{i,a,t}) \\ &= Var(p^R_{i,1,t-a+1}) + \sum_{s=0}^{a-2} Var(\epsilon_{i,a-s,t-s}) \\ &= \sigma^2_{p1} + (a-1)\sigma^2_{p\epsilon} \end{aligned}$$

assuming that the random walk shocks are of the same variance over the life-cycle. Thus, the variance profile is linear in age.

On the other hand, for a standard heterogeneous growth model

$$p^H_{iat} = \theta_1 + \theta_2 a$$

and the variance formula is given by

$$Var(p^H_{iat}) = \sigma^2_1 + \sigma^2_2 a^2 + 2a\sigma_{12}$$

and so it is a convex quadratic function in age.

Then, let us extend the model so that the shock size or growth rate are age-varying and unrestricted. For the random walk case, the change in variance of the permanent component is

$$Var(p^R_{iat}) - Var(p^R_{i,a-1,t-1}) = \sigma^2_{p\epsilon,a}$$

which is a general positive function of age for all a. For the heterogeneous growth case, an age-varying loading λ_a can be introduced

$$p^H_{iat} = \theta_1 + \theta_2 \sum_{a'=2}^{a} \lambda_{a'}$$

where λ_a is unrestricted besides a normalization such as $\lambda_2 = 1$. The variance is then given by

$$Var(p^H_{iat}) = Var(\theta_1) + Var(\theta_2) \left(\sum_{a'=2}^{a} \lambda_{a'} \right)^2 + 2cov(\theta_1, \theta_2) \left(\sum_{a'=2}^{a} \lambda_{a'} \right)$$

and to make it more clearly, we take the difference

$$Var(p^H_{iat}) - Var(p^H_{i,a-1,t-1}) = Var(\theta_2) \left[\left(\sum_{a'=2}^{a} \lambda_{a'} \right)^2 - \left(\sum_{a'=2}^{a-1} \lambda_{a'} \right)^2 \right] + 2cov(\theta_1, \theta_2)\lambda_a$$

which is also a free function of age. Given the value of $\sigma^2_{p\epsilon,a}$, we can solve for λ_a that gives the same change in variance. This change in variance can even be more general than the random walk model because it can take a negative value.