Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages*

Laura Hospido †
Bank of Spain, Research Division

This version: May 2010

Abstract

This paper presents a model for the heterogeneity and dynamics of the conditional mean and the conditional variance of individual wages. A bias-corrected likelihood approach, which reduces the estimation bias to a term of order $1/T^2$, is used for estimation and inference. The small sample performance of the proposed estimator is investigated in a Monte Carlo study. The simulation results show that the bias of the maximum likelihood estimator is substantially corrected for designs calibrated to the data used in the empirical analysis, drawn from the PSID. The empirical results show that it is important to account for individual unobserved heterogeneity and dynamics in the variance, and that the latter is driven by job mobility. The model also explains the non-normality observed in logwage data.

JEL Codes: C23, J31.
Keywords: Panel data, dynamic nonlinear models, conditional heteroskedasticity, fixed effects, bias reduction, individual wages.

*I would like to thank Manuel Arellano and Stéphane Bonhomme for encouragement and advice. Special thanks also to Richard Blundell for his suggestions and time. Comments from the co-editor, Ed Vytlacil, and three anonymous referees have helped improve this paper. I am grateful for useful comments to Pedro Albarrán, Jesús Carro, Jinyong Hahn, Alfredo Martín, Costas Meghir, Javier Mencía, Pedro Mira, Jean-Marc Robin, Enrique Sentana, Ernesto Villanueva and Gema Zamarro, as well as seminar participants at CEMFI, University College London, Universidad Pablo de Olavide, Universitat Autònoma de Barcelona, Universidad Carlos III de Madrid - Statistics, Universidad Jaime I de Castellón, Université Catholique de Louvain, Maastricht University, Universidad de Alicante, Banco de España, Stockholm School of Economics, the 9th IZA Summer School in Labor Economics, the 13th International Conference on Panel Data, the COST meeting in Essen, the European Winter Meeting of the Econometric Society in Turin, the Simposio de Análisis Económico in Oviedo, the RES Job Market Conference in London, the EEA-ESEM meetings in Budapest and the SOLE meetings in New York. Of course, all errors are my own. The views in this paper are those of the author and do not represent the views of the Bank of Spain or the Eurosystem.
†Bank of Spain, Alcala 48, 28014-Madrid. Email: laura.hospido@bde.es, Telephone: +34 913385625, Fax: +34 913385678.

1 Introduction

Estimates of individual earnings processes are useful for a variety of purposes, which include testing between different models of the determinants of earnings, building predictive earnings distributions, or calibrating consumption and saving models. Having a good description of the individuals' earnings dynamics is crucial since the conclusions of many economically relevant models clearly depend on the properties of the earnings process used as an input.

This paper presents a model for the heterogeneity and dynamics of both the levels and the volatilities of individual wages, given past observations and unobserved characteristics. The motivation behind this specification is related to two strands of the literature on earnings dynamics.

The first one has focused on modelling the heterogeneity and time series properties of the conditional mean of earnings (Lillard and Willis, 1978; MaCurdy, 1982; Abowd and Card, 1989; among others), whereas the modelling of the conditional variance, or higher order moments of the process, has been mostly neglected.[1] However, in many applications it is important to understand also the behaviour of the variance. This is the case if we consider an individual trying to forecast her future earnings, in order to guide savings or other decisions. As the individual faces various sorts of risk, she will be interested in forecasting not only the level of earnings but also its variability. Moreover, this person would act very differently if she knows that the risk she suffers is permanently higher, than if it is only due to a period of higher volatility. The properties of the individual variances are thus fundamental both for describing wage profiles over time and for better understanding what drives fluctuations on them. In fact, some recent studies stress the relevance of considering a variance that varies over time and across individuals (Meghir and Windmeijer, 1999; Chamberlain and Hirano, 1999; Meghir and Pistaferri, 2004; Albarrán, 2004; Alvarez and Arellano, 2004; Jensen and Shore, 2008).

A second literature studies the increase in the cross-sectional variance of earnings in the United States since the late 1970s (Autor *et al.*, 2008). This growth in the variance among individuals is associated with an increase in the aggregate inequality. However, we do not know much how the conditional variance of wages behaves during a period of increasing aggregate inequality.

In this paper I propose a model for the conditional variance of wages with the two main ingredients that are also present in the conditional mean: individual unobserved heterogeneity and dynamics. In addition, the model is estimated on data drawn from the Panel Study of Income Dynamics (PSID).

In particular, I build a dynamic panel data model with linear individual fixed effects in the conditional mean and multiplicative individual effects in an autoregressive conditional heteroskedasticity (ARCH) variance function.[2] It is well known that failure to control for this individual heterogeneity can lead to misleading conclusions. This problem is particularly severe when the unobserved heterogeneity is correlated with explanatory variables. Such a situation arises naturally in a dynamic context. Here, I adopt a fixed effects

[1] One important exception to this is Meghir and Pistaferri (2004) as explained below.

[2] Therefore, with this model, we can say to what extent the time evolution of the variance is determined by state dependence effects or by permanent unobserved individual heterogeneity.

perspective leaving the distribution for the unobserved heterogeneity completely unrestricted and treating each effect as one different parameter to be estimated.

There is an extensive literature on how to estimate linear panel data models with fixed effects (see Chamberlain, 1984, and Arellano and Honoré, 2001, for references), but there are no general solutions for non-linear cases. If the number of individuals n goes to infinity while the number of time periods T is held fixed, estimation of non-linear models with fixed effects by maximum likelihood suffers from the so-called incidental parameters problem (Neyman and Scott, 1948). This problem arises because the unobserved individual characteristics are replaced by inconsistent sample estimates that bias the estimates of the model parameters. In particular, the bias of the maximum likelihood estimator (MLE) is of order $1/T$. The most recent reaction to the fact that micro panels are short is to ask for approximately unbiased estimators as opposed to estimators with no bias at all. This approach has the potential of overcoming some of the fixed-T identification difficulties and the advantage of generality. Methods of estimation of nonlinear fixed effects panel data models with reduced bias properties have been recently developed (see Arellano and Hahn, 2007, for a review). There are automatic methods based on simulation (Hahn and Newey, 2004; Dhaene and Jochmans, 2009), bias corrections based on orthogonalization (Cox and Reid, 1987; Lancaster, 2002) and their extensions (Woutersen, 2002; Arellano, 2003a), and corrections based on bias reducing priors (Bester and Hansen, 2007; Arellano and Bonhomme, 2009a), analytical bias corrections of estimators (Hahn and Newey, 2004; Hahn and Kuersteiner, 2004), of the moment equation (Carro, 2007; Fernández-Val, 2009) and of the concentrated likelihood (DiCiccio and Stern, 1993; Severini, 1998; Pace and Salvan, 2006; Bester and Hansen, 2009).[3]

Following this perspective, I consider a modified likelihood function for estimation and inference. Using a bias-corrected concentrated likelihood makes it possible to reduce the estimation bias to a term of order $1/T^2$, without increasing its asymptotic variance (Arellano and Hahn, 2006). This is very encouraging since the goal is not necessarily to find a consistent estimator for fixed-T, but one with a good finite sample performance and a reasonable asymptotic approximation for the samples used in empirical studies.

The small sample performance of the bias corrected estimator is investigated first in a Monte Carlo exercise. The simulation results show that the bias of the MLE is substantially corrected for sample designs that are broadly calibrated to the one used in the empirical application. Then the empirical analysis is conducted on data on the annual wages of prime-age males, as is typical in this literature. The empirical results show that it is important to account for individual unobserved heterogeneity and dynamics in the variance, and that the latter is driven by job mobility. The model also explains the non-normality observed in logwage data.

In a similar sample for male earnings, Meghir and Pistafferi (2004) also find strong evidence of state

[3] So far there are not general theoretical properties in the literature that would help us to narrowing the choice between these alternative bias reducing estimation methods.

dependence effects as well as evidence of unobserved heterogeneity in the variances.[4] However, there exist two important differences between their paper and this one, both in terms of the model and the estimation method. First, they consider a two-shock model, which is assumed to consist of a permanent and a transitory component, and they propose a quadratic specification for the conditional variance of each shock. On the contrary, I consider a single-shock model and propose an exponential specification for the conditional variance of the observed variable, that is, the individual wages.[5] Second, with respect to the estimation method, Meghir and Pistafferi (2004) recover orthogonality conditions and implement a within-group GMM estimator which is consistent when $T \to \infty$ and has a bias of order $1/T$ in a fixed-T context.[6] On the other hand, the bias-corrected likelihood approach adopted in this paper is consistent when $T \to \infty$, but it also reduces the estimation bias to a term of order $1/T^2$.[7] The method in Meghir and Pistafferi (2004) depends essentially on the linear specification they assume for the conditional variances, whereas the properties of the bias-corrected estimator do not depend on specific assumptions related to functional forms.[8] In this paper I use a particular exponential specification, but the same approach could also be used without major changes in other models.[9]

Summing up, the contributions of the paper are twofold. First, from a methodological perspective, I adapt a version of the modified likelihood based on Arellano and Hahn (2006) to a dynamic conditional variance model. Second, from a practical point of view, I show how to apply this new methodology in a relevant empirical context. Two limitations of the current analysis are the following: (i) so far there is not adjustment for measurement error; and (ii) there is not explicit treatment of job changes. It is known that measurement error may be important for PSID earnings data and that part of the variance in wages may be due to job mobility, so these issues need to be addressed in further work.

The rest of the paper is organized as follows. Section 2 presents the nonlinear dynamic model and the likelihood function. Section 3 reviews alternative approaches for correcting the concentrated likelihood adapted to this particular setting. Section 4 studies the finite sample performance of the bias correction in simulated data. Section 5 shows the estimates from the empirical application on individual earnings. Finally, Section 6 concludes with some remarks on a future research agenda.

[4]Also Lin (2005), using a subsample of the dataset considered by Meghir and Pistaferri (2004), finds statistically significant evidence of ARCH effects in earnings dynamics. He considers an ARCH-fixed effects estimator in a "quasi-linear" setting. Here I consider a different econometric framework, which allows me to handle models with multiple effects and estimators without being constrained to the availability of differencing schemes.

[5]Meghir and Windmeijer (1999) and Albarrán (2004) use single-shock models as well but without an application to data.

[6]Notice that although the MLE also has a bias of order $1/T$, the within-group GMM estimator is asymptotically inefficient because it uses arbitrary non-optimal moment conditions. The reason why they cannot do fixed-T consistent GMM estimation is due to a problem of weak instruments.

[7]The difference between having an estimator with a fixed-T bias of order $1/T$ as opposed to our estimator which has a bias of order $1/T^2$ is not negligible, as shown in Section 4 in the simulations for the MLE.

[8]In fact, one of the main general advantages of the bias-corrected methods over other methods for estimating non-linear panel data models is its generality.

[9]These differences are discussed in more detail below.

2 The Model and the Likelihood Function

In this section, I present a model for the heterogeneity and dynamics of the conditional mean and the conditional variance of individual wages, given past observations and unobserved characteristics. This specification would be useful for estimating the conditional distribution of earnings (Chamberlain and Hirano, 1999) and for describing how shocks propagate along that distribution.

2.1 A Model for the Propagation of Shocks

For the conditional mean of standardized logwages I consider an autoregressive specification where i and t index individuals and time, respectively:[10]

$$y_{it} = \eta_{1i} + \alpha y_{it-1} + e_{it};\ (i = 1, ..., n; t = 1, ..., T),$$

where $\{y_{i0}, ..., y_{iT}\}_{i=1}^{n}$ are the observed data,[11] η_{1i} describes permanent differences across individuals, e_{it} reflects shocks that individuals receive every period, and the parameter α measures the persistence on the level of wages to those shocks (net of individual unobserved heterogeneity).[12]

Most of the literature has focused on the estimation of the conditional mean parameter, α, either by assuming homoskedastic shocks or by using estimators of α that are robust to heteroskedasticity. The conditional variance of the process typically has not been modelled. But I am interested in a model for the propagation of shocks along the distribution of individual wages, not only the conditional mean. So I also consider a model for the conditional variance that changes over time and across individuals according to the following specification:

$$\begin{aligned} y_{it} &= \eta_{1i} + \alpha y_{it-1} + e_{it} = \eta_{1i} + \alpha y_{it-1} + h_{it}^{1/2}\epsilon_{it}, \\ h_{it} &= \exp\left\{\eta_{2i} + \beta\left[|\epsilon_{it-1}| - E\left(|\epsilon_{it-1}|\right)\right]\right\}, \end{aligned}$$

where e_{it} is thus an exponential ARCH process, in which the η_{2i}'s are individual fixed effects, ϵ_{it} are *i.i.d.* shocks with zero mean and unit variance, and β measures the persistence on the volatilities of wages to those shocks (net of individual heterogeneity).[13] This formulation implies that h_{it} is always nonnegative, regardless of the parameter values, and it has a known steady-state distribution (Nelson, 1992).[14]

Similarly to the mean, this model captures two patterns of wage volatility. The first one is individual heterogeneity, η_{2i}, meaning that wages of different individuals can vary differently. For instance, there can be permanent differences on the volatilities of wages between civil servants and workers of a sales department and also between workers of sales departments in big and small firms. The second one is dynamics, β,

[10] In case of unbalanced panels, T_i should be indexed on individuals. I omit the subindex to simplify the notation.

[11] I assume that y_{i0} is observed for notational convenience, so that the actual number of waves in the data is $T + 1$.

[12] I focus on a first-order process to simplify the presentation and because this specification turns out to be a good description of the data used in the empirical analysis for the idiosyncratic part of the variation, net of aggregate shocks (see section 5).

[13] Notice that the estimation method that I consider is not dependent on this particular specification.

[14] In the empirical analysis, we approximate the absolute value function by means of a differentiable function.

reflecting the response on the volatility of wages to idiosyncratic shocks (large shocks may translate into larger subsequent volatilities).

The following two equations summarize the model so far:

$$\begin{aligned} E\left(y_{it}|y_i^{t-1},h_{i1},\eta_i\right) &= \eta_{1i}+\alpha y_{it-1}, \\ Var\left(y_{it}|y_i^{t-1},h_{i1},\eta_i\right) &= h\left(\epsilon_{it-1},\eta_{2i}\right)=\exp\left\{\eta_{2i}+\beta\left[\left|\epsilon_{it-1}\right|-E\left(\left|\epsilon_{it-1}\right|\right)\right]\right\}, \end{aligned}$$

where $\eta_i=(\eta_{1i},\eta_{2i})'$ is the vector of individual fixed effects.[15]

2.2 The Individual Likelihood Function

I complete the specification with a normality assumption[16] and an assumption about initial conditions. Under the assumption that $\epsilon_{it}\sim N(0,1)$ the model, given past observations and individual characteristics, is normal heteroscedastic. Formally,

$$\epsilon_{it}|y_i^{t-1},h_{i1},\eta_i\sim N(0,1)\Rightarrow y_{it}|y_i^{t-1},h_{i1},\eta_i\sim N(\eta_{1i}+\alpha y_{it-1},h_{it}).$$

Then, the individual likelihood, conditioned on initial observations and fixed effects, is:

$$f\left(y_{i1},...,y_{iT}|y_{i0},h_{i1},\eta_{i0}\right)=\prod_{t=1}^{T}f\left(y_{it}|y_i^{t-1},h_{i1},\eta_{i0},\theta_0\right),$$

where $\theta=(\alpha,\beta)'$ denotes the vector of common parameters.

The log-likelihood for one observation, ℓ_{it}, differs from the linear model with normal errors through the time-dependence of the conditional variance. For any individual i and $t>1$, I have:

$$\ln f\left(y_{it}|y_i^{t-1},h_{i1},\eta_i,\theta\right)=\ell_{it}\left(\theta,\eta_i\right)\propto-\frac{1}{2}\ln h\left(\epsilon_{it-1},\eta_{2i}\right)-\frac{1}{2}\frac{\left(y_{it}-\alpha y_{it-1}-\eta_{1i}\right)^2}{h\left(\epsilon_{it-1},\eta_{2i}\right)},$$

but evaluation of the likelihood at $t=1$ requires pre-sample values for ϵ_{it}^2 and h_{it}.

Initial conditions. For $t=1$,

$$y_{i1}=\alpha y_{i0}+\eta_{1i}+h_{i1}^{1/2}\epsilon_{i1},$$

where $h_{i1}=h\left(y_{i0},y_{i,-1},y_{i,-2},...,\eta_i\right)$. This is a model for $f\left(y_{i1}|y_{i0},y_{i(-1)},y_{i(-2)},...,\eta_{i0}\right)$ or for $f\left(y_{i1}|y_{i0},\epsilon_{i0},\eta_{i0}\right)$ where ϵ_{i0} summarizes all the past values of y_{it}. Here, I make the additional assumption that h_{i1} is given by the steady-state unconditional variance of e_{it} given fixed effects:

$$\varphi\left(\eta_i,\theta\right)=plim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(y_{it}-\alpha y_{it-1}-\eta_{1i}\right)^2.$$

Following Bollerslev (1986), I approximate this function by the mean of the squared errors:

$$h_{i1}\approx\frac{1}{T}\sum_{t=1}^{T}e_{it}^2.$$

[15] In the sequel, for any random variable (or vector of variables) Z, z_{it} denotes observation for individual i at period t, and $z_i^t=\{z_{i0},...,z_{it}\}$ the set of observations for individual i from the first period to period t.

[16] See section 5 for a check of the validity of this assumption on real data.

Therefore, the individual likelihood function becomes:

$$\mathcal{L}_i(\theta, \eta_i) = \prod_{t=2}^{T} \frac{1}{h_{it}^{1/2}} \phi\left(\frac{y_{it} - \alpha y_{it-1} - \eta_{1i}}{h_{it}^{1/2}}\right) \cdot \frac{1}{h_{i1}^{1/2}} \phi\left(\frac{y_{i1} - \alpha y_{i0} - \eta_{1i}}{h_{i1}^{1/2}}\right),$$

where

$$h_{it} = \begin{cases} \frac{1}{T}\sum_{t=1}^{T} e_{it}^2 & if \quad t = 1, \\ \exp\{\eta_{2i} + \beta[|\epsilon_{it-1}| - E(|\epsilon_{it-1}|)]\} & if \quad t > 1, \end{cases}$$

and $\phi(\cdot)$ denotes the probability density function of a standard normal variable.

3 Correcting the Likelihood Function

In this section, I adopt an analytically bias corrected approach that deals with dynamics and multiple fixed effects in the estimation of a nonlinear panel data model.

3.1 The Bias-Corrected Concentrated Likelihood

The MLE of θ, concentrating out the η_i, is the solution to:

$$\hat{\theta} \equiv \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}(\theta, \hat{\eta}_i(\theta))\right]; \ \hat{\eta}_i(\theta) \equiv \arg\max_{\eta_i} \frac{1}{T}\sum_{t=1}^{T}\ell_{it}(\theta, \eta_i).^{17} \tag{1}$$

In the context of nonlinear models, fixed effects MLE suffers from the incidental parameters problem noted by Neyman and Scott (1948). In this case, the incidental parameters would be the individual effects η_i. The problem arises because these unobserved individual effects are replaced by noisy sample estimates. As only a finite number T of observations are available to estimate each η_i, the estimation error of $\hat{\eta}_i(\theta)$ does not vanish as the sample size n grows, and this error contaminates the estimate of the common parameter due to the nonlinearity.

Formally, let

$$L(\theta) \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left[\sum_{t=1}^{T}\ell_{it}(\theta, \hat{\eta}_i(\theta))\right].$$

Then, from the usual maximum likelihood properties, for $n \to \infty$ with fixed-T, $\widehat{\theta}_T = \theta_T + o_p(1)$, where $\theta_T \equiv \arg\max_\theta L(\theta)$. In general, $\theta_T \neq \theta_0$, but $\theta_T \to \theta_0$ as $T \to \infty$.

An alternative approach to describe the same problem is the following. Due to the noise in estimating the individual effects, the expectation of the concentrated likelihood is not maximized at the true value of the common parameter, θ_0. In fact, the bias in the expected concentrated likelihood at an arbitrary θ can be expanded in orders of magnitude of T:

$$E\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}(\theta, \hat{\eta}_i(\theta)) - \frac{1}{T}\sum_{t=1}^{T}\ell_{it}(\theta, \bar{\eta}_i(\theta))\right] = \frac{\beta_i(\theta)}{T} + o\left(\frac{1}{T}\right),$$

[17] The ML estimates for the individual fixed effects would be obtained as $\hat{\eta}_i \equiv \hat{\eta}_i\left(\hat{\theta}\right) = \arg\max_{\eta_i} \frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\hat{\theta}, \eta_i\right)$.

where $\bar{\eta}_i(\theta)$ maximizes an unfeasible (and unbiased) target likelihood $\lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^{T}\ell_{it}(\theta,\eta_i)\right]$. The idea behind the analytically bias-adjusted approach is to avoid the problem of having an expected concentrated likelihood that is not maximized at the true value of the θ, by correcting the likelihood itself. Therefore, I will consider an estimator that maximizes a bias-corrected concentrated likelihood function like:

$$\hat{\theta}^{BC} \equiv \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\ell_{it}(\theta,\hat{\eta}_i(\theta)) - \beta_i(\theta,\hat{\eta}_i(\theta))\right].^{18} \qquad (2)$$

Letting β_i be an adjustment term, the bias-corrected MLE (BCE), $\hat{\theta}^{BC}$, will be less biased than the MLE, $\hat{\theta}$. For further discussion on the estimation method and a formal analysis of the asymptotic properties of the bias-corrected estimators when n and T grow at the same rate see Arellano and Hahn (2006).

3.2 Estimation of the Bias

The form of the approximate bias is:

$$\beta_i(\theta) \approx \frac{1}{2}\text{trace}\left(H_i^{-1}(\theta,\eta_i)\Upsilon_i(\theta,\eta_i)\right),$$

where

$$H_i(\theta,\eta_i) \equiv -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\ell_{it}(\theta,\eta_i)}{\partial\eta_i\partial\eta_i'},$$

$$\Upsilon_i(\theta,\eta_i) \equiv \sum_{l=-m}^{m}\omega_{T,l}\Gamma_l(\theta,\eta_i), \text{and}$$

$$\Gamma_l(\theta,\eta_i) \equiv \frac{1}{T}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\left[\frac{\partial\ell_{it}(\theta,\eta_i)}{\partial\eta_i}\cdot\frac{\partial\ell_{it-l}(\theta,\eta_i)}{\partial\eta_i'}\right].^{19}$$

In practice, for estimating the bias I use the corresponding sample counterparts. The quantity m is a bandwidth parameter and $\omega_{T,l}$ denotes a weight that guarantees positive definiteness of $\Upsilon_i(\theta,\eta_i)$.[20]

Interpretation of the Bias Expression. The two objects involved in the expression for the bias (i.e. the inverse hessian and the outer product term) are very familiar in a likelihood setting. In terms of the Information Identity, the bias would have the interpretation of a penalty to the expected concentrated likelihood for being apart from the true value, θ_0 (Bester and Hansen, 2009).

Standard Error Estimates. I calculate standard errors of the estimates using *Individual Block-Bootstrap,* that is fixed-T large-n non parametric bootstrap. The assumption of independence across individuals allows me to draw complete time series for each individual to capture the time series dependence. Therefore, I draw $y_i = (y_{i1},...,y_{iT_i})'$ S times to obtain the simulated data $\left\{y_i^{(s)}, y_{i(-1)}^{(s)}\right\}_{s=1}^{S}$. Then, for each

[18] Now, the corresponding estimates for the individual fixed effects would be obtained as $\hat{\eta}_i^{BC} \equiv \hat{\eta}_i\left(\hat{\theta}^{BC}\right) = \arg\max_{\eta_i}\frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\hat{\theta}^{BC},\eta_i\right)$.

[19] Detailed derivations are given in the Appendix 7.1.

[20] In principle, m could be chosen as a suitable function of T to ensure bias reduction but, given that in practice T will be small and that the procedure is known to fail for values of m at both ends of the admissible range ($m = 0$ and $m = T-1$), m will be chosen equal to $1, 2$ or 3. Regarding $\omega_{T,l}$, I use Bartlett weights.

sample, I compute the corresponding estimates of the common parameters $\left\{\left(\widehat{\theta}, \widehat{\theta}^{BC}\right)^{(s)}\right\}_{s=1}^{S}$, and calculate the empirical distribution as an approximation of the distribution of $\widehat{\theta}$ and $\widehat{\theta}^{BC}$.[21]

3.3 Alternative bias reducing estimation methods

In this subsection, I discuss the relationship between the bias-corrected likelihood approach in this paper and other analytical bias adjusted estimation methods for nonlinear panel data models with fixed effects that have been recently developed.

Carro (2007) considered an estimator of a dynamic probit model with a scalar fixed effect that relied on an analytical bias correction to the moment equation. His correction, inspired in a generalization of Cox and Reid (1987), involved both sample and analytical expected likelihood derivatives. In the present context, a generalization of Carro's estimator to multiple effects would be infeasible because the corresponding expected quantities lack closed form expressions. They would need to be replaced by numerical approximations, leading to a different type of estimator with unknown properties.

Hahn and Kuersteiner (2004) considered a bias corrected estimator for a general dynamic model with a scalar fixed effect using sample likelihood derivative quantities evaluated at maximum likelihood estimates. The method I use is equivalent to using a bias correction of the score as opposed to a bias correction of the estimator, hence implicitly updating in estimation the values of the parameters at which corrections are evaluated. Moreover, formulating the correction at the level of the likelihood, as I do, provides an objective function based method and greatly simplifies the form of the correction term, both relative to corrections of estimators or of moment equations, specially with multiple effects. Thus, the expression for the bias in the likelihood, in the case of multiple fixed effects, is much simpler than in the moment equation or in the estimator itself.

The analytic correction used in this paper is closely related to the penalty function independently obtained by Bester and Hansen (2009).[22] In fact, the estimator in (2) is equivalent to

$$\begin{pmatrix} \hat{\theta}^{BC} \\ \hat{\eta}^{BC} \end{pmatrix} = \arg\max_{\theta,\eta} \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\ell_{it}\left(\theta,\eta\right) - \beta_i\left(\theta, \hat{\eta}_i\left(\theta\right)\right)\right],$$

whereas the one proposed in Bester and Hansen would be

$$\begin{pmatrix} \tilde{\theta}^{BC} \\ \tilde{\eta}^{BC} \end{pmatrix} = \arg\max_{\theta,\eta} \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\ell_{it}\left(\theta,\eta\right) - \beta_i\left(\theta, \eta\right)\right],$$

or, equivalently,

$$\tilde{\theta}^{BC} \equiv \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\ell_{it}\left(\theta,\tilde{\eta}_i\left(\theta\right)\right) - \beta_i\left(\theta, \tilde{\eta}_i\left(\theta\right)\right)\right],$$

[21] Notice that, contrary to the block bootstrap procedure used in the time-series literature (Horowitz, 2003), here I do not need to choose any bandwidth.

[22] More specifically to the HS penalty that they consider.

where

$$\tilde{\eta}_i(\theta) \equiv \arg\max_{\eta_i} \left[\sum_{t=1}^{T} \ell_{it}(\theta, \eta_i) - \beta_i(\theta, \eta) \right].$$

Under general conditions, $\tilde{\eta}_i(\theta)$ is asymptotically equivalent to $\hat{\eta}_i(\theta)$, so $\tilde{\theta}^{BC}$ will have similar bias-reducing properties as $\hat{\theta}^{BC}$.

The previous methods and the method that I use in this paper are all asymptotically equivalent, so that there are not known theoretical reasons to prefer one to another. A particular method may still be preferable for convenience of implementation. In our context, with multiple fixed effects, dynamics and expected derivatives with no closed form, trace-based bias corrections at the level of the likelihood seem the most convenient, although other alternatives are possible.

4 Monte Carlo Study

The practical importance of the bias corrections depends on how much bias is removed for the small T that is often relevant in econometric applications. In this section, the small sample performance of the bias-corrected estimator BCE, $\hat{\theta}^{BC}$, is explored relative to the fixed effects MLE, $\hat{\theta}$, in an AR(1)-EARCH(1) specification broadly calibrated to the one used in the empirical application, in terms of the sample size, the panel dimensions, and the variability across individuals, as detailed below.

The model design is the following

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \eta_{1i} + h_{it}^{1/2}\epsilon_{it}, \quad (t = 1, ..., T;\ i = 1, ..., n) \\ h_{it} &= \exp\left(\eta_{2i} + \beta\left[\sqrt{\epsilon_{it-1}^2 + \Lambda} - \sqrt{2/\pi}\right]\right), \\ \epsilon_{it} &\sim N(0,1), \end{aligned}$$

where Λ is a small positive number used to approximate the absolute value function by means of a rotated hyperbola, and $\sqrt{2/\pi}$ is an approximation for $E(|\epsilon_{it-1}|)$ given that $\epsilon_{it} \sim N(0,1)$.

The process was started at $y_{i0} = 0$, then 700 time periods are generated before the sample actually starts. The data were generated with $\eta_{1i} \sim N(0,1)$, and $\eta_{2i} \sim N(-2,1)$. Other model parameters are set as follows: $T = \{8; 16\}$, $n = \{200; 2,000\}$, and $\theta = (\alpha = \{0.5; 0.0\}, \beta = \{0.5; 0.0\})'$. Then, 100 Monte Carlo replications are used for each design, with just ϵ_{it} redrawn in each replication, and I draw y_i 50 times, at each stage, to obtain the simulated data for the individual block-bootstrap.[23]

[23] As explained in Section 5, in the empirical analysis I use data on 2,066 individuals for the period 1968-1993 of the PSID. It is a very unbalanced panel with, on average, 16 time observations per individual. In addition, the sample is restricted to individuals with at least nine years of usable wages data. This means that, conditional of the initial observations y_{i0}, T_i would be at least 8. For these reasons I initially set $T = \{8; 16\}$, and $n = 2,000$. Eventually, I have also simulated the model for $n = 200$, because I expect the bias corrected estimators to improve much more with T than with n, whereas a smaller n speeds up computation. The values chosen for the parameters that determine the distributions of the unobserved individual effects try to mimic the behaviour of similar moments in the data. In the case of η_{1i}, the values 0 and 1 approximate, respectively, the mean and variance of the sample distribution of individual means on logwage data. Analogously for η_{2i}, -2 and 1 approximate the mean and variance of the sample distribution of individual logvariances.

The vector of common parameters, θ, is estimated by maximum likelihood, $\hat{\theta}$, and applying the analytically bias-corrected estimator with $m = 2$, $\hat{\theta}^{BC}$, defined in equations (1) and (2), respectively. Given the complexity of the design, I cannot get a closed form solution for the estimator of the individual fixed parameters as a explicit function of θ. Therefore, to maximize the likelihood function, I use a double Quasi-Newton's method algorithm. In each step of the algorithm, $\widehat{\eta}_i(\theta)$ is computed such that, for that given value of θ, the individual likelihood is maximized with respect to η_i.[24] The same procedure applies for the bias corrected likelihood.[25,26]

I present results for two different data generating processes: one with scalar individual fixed effects (one in which η_{1i} is omitted), and another with multiple fixed effects (the one described above). Tables 1 and 2 report the median bias, the median absolute error and the sample standard deviation, along with the standard error and the corresponding 0.95 coverage rates for each design.[27] Failure refers to the fraction of cases of divergence or failure to converge in the nonlinear solution over the 100 Monte Carlo replications.

Table 1 reports the results corresponding to the DGP with scalar fixed effects for $n = 200$.[28] Given that the design does not included individual effects in the mean, the estimate of α is almost not biased. On the contrary, the MLE of β is seriously downward biased, even for $T = 16$. After applying the correction, the estimate for β is closer to the true value of the parameter, specially when $T = 16$. In addition, we can see that the standard errors estimated by individual block-bootstrap represent a good approximation to the Monte Carlo standard deviation.

Table 2 reports the results corresponding to the DGP with multiple fixed effects for $n = 200$. Once again, the bias-corrected estimator can remove a substantial part of that bias, now both in the estimate of α and β, when T is relatively small.

5 Estimation Results

In this section, I apply the analytically bias-corrected likelihood methodology to estimate an empirical model for the conditional mean and the conditional variance of prime-age male earnings. As Meghir and Pistafferi (2004), I use data on 2,066 individuals for the period 1968-1993 of the PSID. It is an unbalanced panel with 32,066 observations. I select male heads aged 25 to 55 with at least nine years of usable wages data. Step-by-step details on sample selection are reported in Appendix 7.2. Sample composition by year and demographic characteristics are presented in Appendix 7.3.

The dependent variable is annual real wages of the heads.[29] Figures 1 and 2 plot the mean and the

[24] Strictly speaking, I compute n individual maximizations inside the one for θ.

[25] In addition, for computing the analytical bias expression I need to calculate numerical derivatives.

[26] The residuals used in approximating the initial variance h_{i1} are updated within the ML estimation of the common parameters and individual effects.

[27] The coverage rate reports the fraction of times the .95 confidence interval contained the true value, with the confidence interval obtained by the 0.025 -0.975 quantiles of the distribution of the 50 bootstrap-sample estimates.

[28] I do not report here the results for $n = 2,000$, because - as expected - increasing the number of individuals from $n = 200$ to $n = 2,000$ has little effect on the magnitude of the estimated bias.

[29] The earnings variable in the PSID includes the labour portion of money income from all sources. That is apart, from wages,

variance of log real wages against time for education group and for the whole sample. These figures look very similar to the ones in Meghir and Pistaferri (2004, pp. 4-5) and, as they say, reproduce well known facts about the distribution of male earnings in the U.S. (Levy and Murnane, 1992).

5.1 Estimation of the Model

The dependent variable that I use in the estimation, y_{it}, is logwage residuals from first stage regressions on year dummies, education, a quadratic in age, dummies for race, region of residence, and residence in a SMSA.[30] In this version of the model, I deal with aggregate effects in the variance by regarding y_{it} as standardized wages.[31]

The equation estimated is

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \eta_{1i} + h_{it}^{1/2}\epsilon_{it}, \;\; (t = 1, ..., T_i;\; i = 1, ..., n) \\ h_{it} &= \exp\left(\eta_{2i} + \beta\left[\sqrt{\epsilon_{it-1}^2 + \Lambda} - \sqrt{2/\pi}\right]\right). \end{aligned} \tag{3}$$

Given the results in the previous section, I estimate the vector of common parameters, θ, by maximum likelihood and using the analytically bias-corrected estimator. The first two columns in Table 3 summarize the estimation results. As in the Monte Carlo exercise, I obtain that the maximum likelihood estimate is below the corrected one. In fact, after applying the bias correction, I obtain estimates for both parameters over 0.5. Not only the persistence in the mean is significant. Also the state dependence effects in the volatility of wages seem important.

Correlations between unobserved heterogeneity and observed outcomes. The main advantage of adopting a fixed effects perspective is that we can capture unobserved permanent heterogeneity among individuals in a very robust way. If we were able to observe that individual heterogeneity, including observed measures would be much better in terms of estimation. However, many individual characteristics are usually unobserved for the econometrician, as in the regression summarized below.

Another nice feature of this approach is that we obtain estimates of the individual fixed effects and, therefore, we can evaluate the relation between those effects and measurable outcomes.[32]

it includes the labour part of farm income and/or business income, bonuses, overtime, commissions, professional practice, and the labour part of income from roomers and boarders. As noted by Gottschalk and Moffitt (1993), some of the components of this measure of labour income may reflect capital income, such as the labour part of farm income. In the paper I use annual wages to the exclusion of other components of money income from labour.

[30] In earnings dynamics research it is standard to adopt a two step procedure. In the first stage regression, the log of real wages is regressed on control variables and year dummies to eliminate group heterogeneities and aggregate time effects. Then, in the second stage, the unobserved heterogeneity and dynamics of the residuals are modelled. Given the large samples that are used to form the residuals, the fact that the estimation is performed in two stages is of little consequence.

[31] I define standardized logwages y_{it} as the individual logwages net of aggregate effects both in the mean, μ_t, and in the variance, σ_t^2. Formally,

$$\log w_{it} = \hat{\mu}_t + \hat{\sigma}_t y_{it}.$$

where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are calculated each year as the sample mean and the sample variance, respectively.

[32] As shown by Arellano and Bonhomme (2009b) the coefficients estimates obtained when regressing fixed effects estimates, $\hat{\eta}_{2i}$, on a set of strictly exogenous regressors $\mathbf{F}_i$, yields consistent estimates for the coefficients of the projection of the population individual effects η_{2i} on the regressors $\mathbf{F}_i$. However, the standard errors of the estimates of the projection coefficients in the regression of $\hat{\eta}_{2i}$ on $\mathbf{F}_i$ are inconsistent and need to be corrected. Instead, I obtain the standard errors by bootstrap.

Table 4 reports projection coefficients and bootstrap standard errors from linear regressions of the estimated fixed effects in the volatilities of logwages on some observed outcomes for the individuals in the sample.[33,34] Results show that being married, older, and white, are negatively associated with individual fixed effects in the variance. Also, being a technical worker, a manager, or having large values of tenure. On the other hand, being a sales or a services worker, having moved from one job to other at least once, or having a low educational degree, are associated with higher volatility. The direction of the association is the one that one would expect.

Risk Tolerance. The 1996 wave of the PSID includes a module on risk preferences data. The index of risk tolerance (inverse of risk aversion) is obtained from answers to hypothetical questions about lotteries, as designed by Barsky et al. (1997). For those individuals with information available on the risk tolerance index (around 54 per cent of the original sample), we run a simple regression of the estimated individual fixed effects over dummies for each index category.[35] Figure 3 shows the estimated coefficients (solid line) along with the corresponding 0.95 confidence intervals (dotted lines) obtained by bootstrap. As shown in the figure, risk-tolerant individuals seem to be associated with higher values of the individual fixed effects in the variance, whereas the pattern is less clear for the fixed effects in the mean. However, the differences are not statistically significant.

5.2 Checking for Non-normality

In this section, I apply deconvolution techniques as in Horowitz and Markatou (1996) to estimate the distribution of the errors and check the normality assumption. Although the assumption of normality is not necessary for the validity of the analytically bias-corrected estimator, checking this distributional assumption turns out useful for other purposes as illustrated below.

The technique that I use is a normal probability plot of residuals in first-differences, as shown in Figure 4. The differenced residuals are represented in the x-axis, whereas the values in the y-axis represent the inverse normal of the cumulative distribution of the empirical distribution of the data. If the data are approximately normally distributed, the points should form an approximate straight line.[36]

The top part of Figure 4 shows the normal probability plot of the residuals in first-differences.[37] The figure indicates that the tails of the distribution of errors are thicker than those of the normal distribution. The bottom part of Figure 4 reports the corresponding plot for the standardized residuals in first-differences. Now, we obtain almost a straight line meaning no departure from normality. Similar conclusions may be

[33] For variables that change over time, we take as a reference point the last observation for each individual in the sample.

[34] We also compute the sample correlation between the effects in levels and in the variances. We find a negative correlation, meaning that higher levels of earnings are related with lower volatilities.

[35] The index can take values .15, .28, .35 and .57.

[36] The figure also contains the corresponding pointwise confidence intervals, displayed as dotted lines.

[37] Estimated residuals and estimated standardized residuals respectively defined as $\hat{e}_{it} = y_{it} - \hat{\alpha}^{BC} y_{it-1} - \hat{\eta}_{1i}$, and $\hat{\epsilon}_{it} = \frac{y_{it} - \hat{\alpha}^{BC} y_{it-1} - \hat{\eta}_{1i}}{h_{it}^{1/2}(\hat{\eta}_{2i}, \hat{\epsilon}_{it-1})}$, where $h_{it}(\hat{\eta}_{2i}, \hat{\epsilon}_{it-1}) = \exp\left\{\hat{\eta}_{2i} + \hat{\beta}^{BC}\left[|\hat{\epsilon}_{it-1}| - \sqrt{2/\pi}\right]\right\}$.

reached in terms of the kurtosis of those distributions, as reported in Table 5, given that the kurtosis for the standardized residuals in first-differences is closer to 3.

Fit of the model. Given the distributional assumption, the initial conditions and the parameter estimates, $\hat{\theta}^{BC}$ and $\hat{\eta}_i$, now I simulate an unbalanced panel of standardized logwage observations with the same dimension as the PSID sample, and - with this simulated panel - I evaluate the fit of the model. Figure 5 shows the kernel densities of logwages in the data and according to the model. The main conclusion is that the model does a good job fitting the data.

5.3 Individual Heterogeneity

Similarly to the previous exercise, in this section I conduct counterfactuals to evaluate the existence of individual heterogeneity on the data.

The first one - Counterfactual 1 - is obtained using the model, the observed initial conditions and the parameter estimates, $\hat{\theta}^{BC}$ and $\hat{\eta}_{2i}$, but now with $\eta_{1i} = \bar{\eta}_1, \forall i$, where $\bar{\eta}_1 = N^{-1}\sum_{i=1}^{N} \hat{\eta}_{1i}$. The second - Counterfactual 2 - is generated using the model, the initial conditions and the parameter estimates, $\hat{\theta}^{BC}$ and $\hat{\eta}_{1i}$, but now with $\eta_{2i} = \bar{\eta}_2, \forall i$, where $\bar{\eta}_2 = N^{-1}\sum_{i=1}^{N} \hat{\eta}_{2i}$.

Figure 6 reports the distribution of individual sample means on real logwage data, on simulated data and for the counterfactual 1 ($\eta_{1i} = \bar{\eta}_1, \forall i$). Analogously, Figure 7 shows the distribution of individual logvariances on real data, simulated data and counterfactual 2 ($\eta_{2i} = \bar{\eta}_2, \forall i$). In the figures, we can see that there exists significant variation across individuals not only in the means but also in the variances, and this variation is successfully captured by the model.

Moreover, using these counterfactuals we can say how much of the variance in logwages is due to individual heterogeneity in the mean and how much due to individual heterogeneity in the variance. In particular, for the counterfactual 1, the sample variance of logwages is equal to 0.7345, whereas for the counterfactual 2 the corresponding sample variance is 0.8718. That is, variation in $\hat{\eta}_{1i}$ accounts for by 26 per cent of the total variation in logwages, whereas variation in $\hat{\eta}_{2i}$ accounts for by 13 per cent.

5.4 Quantiles of log normal wages

Regarding dynamics, given the model specification, we can calculate the effects of the propagation of past shocks at different parts of the wage distribution. As derived in Appendix 7.4, I estimate mean marginal effects of past shocks over different quantiles of the logwage distribution as

$$\widehat{E}\left(\frac{\partial Q_\tau\left(y_{it}\right)}{\partial \epsilon_{it-s}}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{\partial Q_\tau\left(y_{it}\right)}{\partial \epsilon_{it-s}}\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{\partial \mu_{it}}{\partial \epsilon_{it-s}} + q_\tau \frac{\partial h_{it}^{1/2}}{\partial \epsilon_{it-s}}\right],$$

with

$$\begin{aligned} \mu_{it} &= \alpha y_{it-1} + \eta_{1i}, \\ h_{it} &= \exp\left(\eta_{2i} + \beta\left[\left(\epsilon_{it-1}^2 + \Lambda\right)^{1/2} - (2/\pi)^{1/2}\right]\right), \end{aligned}$$

where $Q_\tau(y_{it})$ is the τth quantile of the logwage distribution and q_τ the τth quantile of the $N(0,1)$.

The first row in Table 6 reports the estimates with respect to ϵ_{it-1}, and the second row the corresponding estimates for ϵ_{it-2}. The main result is that past shocks seem to have some effect over logwages even two periods later. Moreover, it seems that the effects with respect to ϵ_{it-1} increase slightly with the quantiles, although - as shown in Figure 8 - these differences are not statistically significant.

5.5 Job changes

The model abstracts for specific reasons for shocks as, in particular, job changes. Modelling job mobility is out of the scope of the paper but, to evaluate in an informal way if state dependent effects are related to job changes, I consider a sample in which the same individual in different jobs is treated as different individuals.[38] I apply the same sample selection as before, male heads aged 25 to 55 with at least nine years of usable wages data in each job spell, and obtain a panel with 1,346 individuals and 17,485 observations.

Estimation results are reported in the second block of two columns in Table 3. We can see that the significant ARCH effects in the variance disappear as soon as we consider a sample without job changes. It seems that the significant dynamics effects in the conditional variance are mostly due to the dynamics between jobs and not within jobs.

5.6 Measurement Issues

As stated in the Introduction, in this model there is not adjustment for measurement error although it is known that reporting errors may be important for PSID earnings data. Error components models for the covariance structure of earnings can easily accommodate a measurement error component. On the contrary, including a measurement component in the type of models like the one considered here, given the non-linearities, is essentially more complicated, and beyond the limits of this paper. Nevertheless, my model would be still consistent with a model in which measurement error would be interpreted as a fixed effect, in line with some recent findings of the literature on measurement error (Bound *et al.*, 2001; Gottschalk and Moffitt, 2009).

Another relevant issue is the extent to which attrition from the PSID may have affected the results. In this paper, I assume that attrition is all accounted for by the permanent characteristics in the individual fixed effects. To provide some evidence for this I compare previous estimates in section 5.1 to those obtained using only individuals who are 16 or more years in the sample (921 individuals). This kind of selection mimics attrition bias since it eliminates individuals observed for a shorter time period. Looking at the

[38] A model in which individual heterogeneity is treated as fixed would work worse in a sample with many job changes. Instead, I am considering

$$\begin{aligned} y_{ijt} &= \alpha y_{ijt-1} + \eta_{1ij} + e_{ijt};\ \text{individual } i \text{ in job } j, \\ y_{ij't} &= \alpha y_{ij't-1} + \eta_{1ij'} + e_{ijt};\ \text{individual } i \text{ in job } j', \end{aligned}$$

as two different individuals.

bias-corrected estimates in the third block of two columns in Table 3, the main conclusion is that they are not very different to those reported in the first two columns.[39]

5.7 Other Model Specification Options and Generality of the Estimation Method

Alternative Model Specifications. The model in this paper turns out to be a good description of the data used in the empirical analysis for the idiosyncratic part of the variation in earnings, net of aggregate shocks. Apart from that, there are other reasons for preferring this particular specification.

First, I choose a single-shock specification and model the conditional variance of the observed variable to describe how shocks propagate along the earnings distribution. This is in contrast to models in the tradition of Hall and Mishkin (1982), which distinguish between a permanent and a transitory shock. The model in Meghir and Pistafferi (2004) belongs to this category. Their innovation is to make the variance of each shock a function of individual effects and dynamics. A two-shock model can be mapped into a one-shock model, but the mapping is not straightforward (Arellano, 2003b). Given that there is just one observable variable, the identification of the two-shock model critically hinges on the unit root assumption for the permanent component. However, when the autoregressive roots are estimated no evidence is found of a unit root, at least not in my sample (see also Alvarez and Arellano, 2004, and Guvenen, 2009). In the Hall and Mishkin (1982) approach a permanent income shock was interpreted as a common factor in a bivariate model of consumption and income. However, from the point of view of developing a descriptive model of distributional income dynamics using only income data, the one-shock approach seems natural. This strategy connects directly with the perspective adopted in Chamberlain and Hirano (1999), who consider a decision maker who estimates the distribution of future income on the basis of her own income and the income of others using a panel such as PSID, but who lacks information on two latent shocks concerning herself or other agents.

Second, the order of the autoregressive process can be determined empirically. I select a first-order process, but I also tried a second lag and the corresponding estimated coefficient was not significant. Concerning the ARCH function, I select an exponential specification because it implies a conditional variance that is always nonnegative regardless of the parameter values and, in addition, it has a known steady-state distribution (Nelson, 1992).

The bias-corrected approach adopted in this paper works for general likelihood models, therefore I am not restricted to consider functional forms that deliver moment conditions usable in a GMM method, such as the linear specification used in Meghir and Pistafferi (2004). The bias corrected estimator used here has a bias of order $1/T^2$, whereas the bias of the within-group GMM estimator used by Meghir and Pistafferi (2004) is of order $1/T$ (although their method could be bias corrected).

Despite the differences, both in terms of the model and the estimation method, the qualitative results

[39] Fitzgerald *et al.* (1998) show that the PSID data set has experienced a significant amount of attrition, although they do not find indications that this causes noticeable bias.

in this paper are similar to the ones in Meghir and Pistafferi (2004), as they also find strong evidence of state dependence effects as well as evidence of unobserved heterogeneity in the variances. However, the interpretation of these findings would be different for the two models (they have two shocks and I only have one), as would be the implications, for example, in a model of consumption. Even so, those differences might be less of a concern when using the model as a description of the conditional distribution of wages.

Generality of the Estimation Method. As mentioned before, one of the main advantages of the bias-corrected approach over other methods is its generality. To illustrate this, I have also estimated the following alternative model proposed by Meghir and Windmeijer (1999) using the analytically bias-corrected likelihood methodology:

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \eta_{1i} + g_{it}^{1/2}\epsilon_{it}, \quad (t = 1, ..., T_i; \; i = 1, ..., n) \\ g_{it} &= \exp\left(\eta_{2i} + \beta\left[\sqrt{y_{it-1}^2 + \Lambda}\right]\right). \end{aligned} \tag{4}$$

This is a convenient specification, in terms of building the moment conditions, but it is more difficult to interpret because the conditional variance of e_{it}, g_{it}, it is a function of the past values of the dependent variable instead of the past values of the error.

The last two columns in Table 3 report the corresponding estimation results for the MLE, $\hat{\theta}$, and the analytically bias-corrected estimator, $\hat{\theta}^{BC}$. Although the estimates of β are a bit different, the main results do not change.

6 Conclusions

In this paper I propose a model for the conditional mean and the conditional variance of individual wages. It is a non linear dynamic panel data model with multiple individual fixed effects. For estimating the parameters of the model, I assume a distribution for the shocks and apply bias corrections to the concentrated likelihood. This corrects the bias of the estimated parameters from $O\left(T^{-1}\right)$ to $O\left(T^{-2}\right)$, so the estimator has a good finite sample performance and a reasonable asymptotic approximation for moderate T. In fact, Monte Carlo results show that the bias of the MLE is substantially corrected for samples designs that are broadly calibrated to the PSID dataset.

The main advantage of this approach is its generality. As we have seen, the method is generally applicable to take into account dynamics and multiple fixed effects. Another advantage is that the fixed effects are estimated as part of the estimation process.

The empirical analysis is conducted on data drawn from the 1968-1993 PSID dataset. In line with some previous references, I find a corrected estimate for the autoregressive coefficient in the mean less than one (Alvarez and Arellano, 2004; Guvenen, 2009), and positive ARCH effects for the variance (Meghir and Pistafferri, 2004). Job changes seem to be driving this dynamics in the variance. I also find important permanent differences across individuals in the variance. In addition, it turns out that this location-scale model explains the non-normality observed in logwage data.

Finally there are two issues, at least, that require further research: a more comprehensive model that includes job changes, and the comparison with female workers in terms of wage profiles.

7 Appendix

7.1 First Order Bias of the Concentrated Likelihood at an arbitrary value of the common parameter θ

Following Arellano and Hahn (2006, 2007), let me obtain the expression for the first order bias of the concentrated likelihood at an arbitrary value of the common parameter θ.

Let $\ell_i(\theta,\eta_i)=\sum_{t=1}^{T}\ell_{it}(\theta,\eta_i)/T$ where $\ell_{it}(\theta,\eta_i)=\ln f(y_{it}|y_{it-1},\theta,\eta_i)$ denotes the log likelihood of one observation. Let

$$\bar{\eta}_i(\theta)=\arg\max_{\eta_i}\operatorname{plim}_{T\to\infty}\ell_i(\theta,\eta_i),$$

and

$$\hat{\eta}_i(\theta)=\arg\max_{\eta_i}\ell_i(\theta,\eta_i),$$

so that under regularity conditions $\bar{\eta}_i(\theta_0)=\eta_{i0}$.

Following Severini (2002) and Pace and Salvan (2006), the concentrated likelihood for unit i

$$\hat{\ell}_i(\theta)=\ell_i(\theta,\hat{\eta}_i(\theta)),$$

can be regarded as an estimate of the unfeasible concentrated log likelihood

$$\bar{\ell}_i(\theta)=\ell_i(\theta,\bar{\eta}_i(\theta)).$$

Now, define

$$\begin{aligned}
u_{it}(\theta,\eta_i) &= \frac{\partial \ell_{it}(\theta,\eta_i)}{\partial\theta},\ v_{it}(\theta,\eta_i)=\frac{\partial \ell_{it}(\theta,\eta_i)}{\partial\eta_i},\\
u_i(\theta,\eta_i) &= \frac{1}{T}\sum_{t=1}^{T}u_{it}(\theta,\eta_i),\ v_i(\theta,\eta_i)=\frac{1}{T}\sum_{t=1}^{T}v_{it}(\theta,\eta_i),\\
H_i(\theta) &= -\lim_{T\to\infty}E\left[\frac{\partial v_i(\theta,\bar{\eta}_i(\theta))}{\partial\eta_i'}\right].
\end{aligned}$$

When η_{i0} is a vector of fixed effects, the Nagar expansion for $\hat{\eta}_i(\theta)-\bar{\eta}_i(\theta)$ takes the form

$$\hat{\eta}_i(\theta)-\bar{\eta}_i(\theta)=H_i^{-1}(\theta)\,v_i(\theta,\bar{\eta}_i(\theta))+\frac{1}{T}B_i(\theta)+O_p\left(\frac{1}{T^{3/2}}\right), \quad \text{(A.1)}$$

where

$$\begin{aligned}
B_i(\theta) &= H_i^{-1}(\theta)\Big[\Xi_i(\theta)\,vec\left(H_i^{-1}(\theta)\right)\\
&\quad +\frac{1}{2}E\left(\frac{\partial}{\partial\eta'}vec\frac{\partial v_i(\theta,\bar{\eta}_i(\theta))}{\partial\eta'}\right)'\left(H_i^{-1}(\eta)\otimes H_i^{-1}(\eta)\right)vec\left(\Upsilon_i(\eta)\right)\Big],\\
\Upsilon_i(\theta) &= \Upsilon_i(\theta;\theta_0,\eta_{i0})=\lim_{T\to\infty}TE\left[v_i(\theta,\bar{\eta}_i(\theta))\,v_i(\theta,\bar{\eta}_i(\theta))'\right],\\
\Xi_i(\theta) &= \Xi_i(\theta;\theta_0,\eta_{i0})=\lim_{T\to\infty}TE\left[\frac{\partial v_i(\theta,\bar{\eta}_i(\theta))}{\partial\eta'}\otimes v_i(\theta,\bar{\eta}_i(\theta))'\right].
\end{aligned}$$

Next, expanding $\ell_i(\theta, \hat{\eta}_i(\theta))$ around $\bar{\eta}_i(\theta)$ for fixed θ,

$$\begin{aligned}
& \ell_i(\theta, \hat{\eta}_i(\theta)) - \ell_i(\theta, \bar{\eta}_i(\theta)) = \frac{\partial \ell_i(\theta, \bar{\eta}_i(\theta))}{\partial \eta'}(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta)) \\
& +\frac{1}{2}(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta))' \frac{\partial^2 \ell_i(\theta, \bar{\eta}_i(\theta))}{\partial \eta \partial \eta'}(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta)) + O_p\left(\frac{1}{T^{3/2}}\right) \\
= \quad & \frac{\partial \ell_i(\theta, \bar{\eta}_i(\theta))}{\partial \Theta'}(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta)) \\
& +\frac{1}{2}(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta))' E\left(\frac{\partial^2 \ell_i(\theta, \bar{\eta}_i(\theta))}{\partial \eta \partial \eta'}\right)(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta)) + O_p\left(\frac{1}{T^{3/2}}\right) \\
= \quad & v_i(\theta, \bar{\eta}_i(\theta))'(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta)) \\
& -\frac{1}{2}(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta))' H_i(\theta)(\hat{\eta}_i(\theta) - \bar{\eta}_i(\theta)) + O_p\left(\frac{1}{T^{3/2}}\right).
\end{aligned}$$

Substituting (A.1)

$$\ell_i(\theta, \hat{\eta}_i(\theta)) - \ell_i(\theta, \bar{\eta}_i(\theta)) = \frac{1}{2} v_i(\theta, \bar{\eta}_i(\theta))' H_i^{-1}(\theta) v_i(\theta, \bar{\eta}_i(\theta)) + O_p\left(\frac{1}{T^{3/2}}\right).$$

Taking expectations

$$E[\ell_i(\theta, \hat{\eta}_i(\theta)) - \ell_i(\theta, \bar{\eta}_i(\theta))] = \frac{1}{2T}\text{trace}\left(H_i^{-1}(\theta)\Upsilon_i(\theta)\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

So the bias in the expected concentrated likelihood at an arbitrary θ is

$$b_i(\theta) = \frac{1}{2}\text{trace}\left(H_i^{-1}(\theta)\Upsilon_i(\theta)\right).$$

Thus,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\ell_{it}(\theta, \hat{\eta}_i(\theta)) - \sum_{i=1}^{N}\widehat{b}_i(\theta),$$

is expected to be a closer approximation to the target likelihood than $\sum_{i=1}^{N}\sum_{t=1}^{T}\ell_{it}(\theta, \hat{\eta}_i(\theta))$.

7.2 Sample Selection

Starting point: PSID 1968-1993 Family and Individual - merged files (53,005 individuals).

1. Drop members of the Latino sample (10,022 individuals) and those who are never heads of their households (26,945 individuals) = Sample (16,038 individuals).
2. Keep only those who are continuously heads of their households, keep only those who are in the sample for 9 years or more, and keep only those aged 25 to 55 over the period = Sample (5,247 individuals).
3. Drop female heads = Sample (4,036 individuals).
4. Drop those with a spell of self-employment, drop those with missing earnings, and drop those with zero or top-coded earnings data = Sample (2,205 individuals).
5. Drop those with missing education and race records, and those with inconsistent education records = Sample (2,148 individuals).
6. Drop those with outlying earnings records, that is, a change in log earnings greater than 5 or less than -3 and those with non continuous data = FINAL SAMPLE (2,066 individuals and 32,066 observations).

Table A.1. Sample selection

Number of individuals	Meghir & Pistaferri (2004)		Hospido (2010)		Difference
Starting point		53,013		53,005	8
Latino subsample	(10,022)	42,991	(10,022)	42,983	8
Never Heads	(26,962)	16,029	(26,945)	16,038	-9
Heads, Age, T_i¿9	(11,490)	4,539	(10,791)	5,247	-708
Female	(876)	3,663	(1,211)	4,036	-373
Self-employment, missing wages	(1,323)	2,340	(1,831)	2,205	135
Missing education and race	(187)	2,153	(57)	2,148	5
Outlying wages	(84)	2,069	(82)	2,066	3
FINAL SAMPLE: Individuals		2,069		2,066	
FINAL SAMPLE: Observations		31,631		32,066	

7.3 Sample composition and descriptive statistics

Table A.2. Distribution of observations by year

Year	Number of observations	Year	Number of observations	Year	Number of observations
1968	655	1977	1229	1986	1583
1969	694	1978	1263	1987	1536
1970	738	1979	1310	1988	1486
1971	780	1980	1380	1989	1434
1972	856	1981	1419	1990	1392
1973	943	1982	1464	1991	1348
1974	1018	1983	1506	1992	1315
1975	1098	1984	1559	1993	1256
1976	1178	1985	1626		

Table A.3. Descriptive Statistics

	1968	1980	1993
Age	36.99	36.61	41.45
	(6.58)	(9.22)	(5.74)
HS Dropout	0.44	0.25	0.12
HS Graduate	0.41	0.55	0.60
Hours	2272	2153	2135
	(573)	(525)	(560)
Married	0.84	0.83	0.83
White	0.68	0.66	0.69
Children	2.80	1.39	1.36
	(2.06)	(1.28)	(1.23)
Family Size	4.90	3.53	3.51
	(2.01)	(1.58)	(1.45)
North-East	0.18	0.16	0.16
North-Central	0.27	0.25	0.23
South	0.39	0.42	0.44
SMSA	0.68	0.67	0.53

Note: Standard deviations of non-binary variables in parentheses.

7.4 Quantiles of log normal wages

Let logwages $y = \log(w) \sim N(\mu, \sigma^2)$ with *cdf*

$$\Pr(\log w \leq r) = \Phi\left(\frac{r-\mu}{\sigma}\right).$$

The τth quantile of the logwage distribution, $Q_\tau(\log w)$, is the value of r such that

$$\Phi\left(\frac{Q_\tau(\log w)-\mu}{\sigma}\right) = \tau,$$

so that

$$\frac{Q_\tau(\log w)-\mu}{\sigma} = \Phi^{-1}(\tau) \equiv q_\tau,$$

where q_τ is the τth quantile of the $N(0,1)$ distribution. Given that

$$\Pr(\log w \leq r) = \Pr(w \leq \exp r),$$

so that

$$\Pr(\log w \leq Q_\tau(\log w)) = \Pr(w \leq \exp Q_\tau(\log w)) = \tau,$$

and, therefore, the τth quantile of the wage distribution, $Q_\tau(w)$, is

$$Q_\tau(w) = \exp Q_\tau(\log w) = \exp(\mu + q_\tau \sigma).$$

In the conditional case, regarding μ and σ as functions of $\log w_{it-1}$, we would write

$$\frac{\partial \log Q_\tau(w_{it})}{\partial \log w_{it-1}} = \frac{\partial \mu_{it}}{\partial \log w_{it-1}} + q_\tau \frac{\partial \sigma_{it}}{\partial \log w_{it-1}},$$

and estimate mean elasticities at different parts of the wage distribution as

$$\widehat{E}\left(\frac{\partial \log Q_\tau(w_{it})}{\partial \log w_{it-1}}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{\partial \log Q_\tau(w_{it})}{\partial \log w_{it-1}}\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{\partial \mu_{it}}{\partial y_{it-1}} + q_\tau \frac{\partial \sigma_{it}}{\partial y_{it-1}}\right].$$

Alternatively, regarding μ and σ as functions of ϵ_{it-1}, we can write

$$\frac{\partial Q_\tau(y_{it})}{\partial \epsilon_{it-1}} = \frac{\partial \mu_{it}}{\partial \epsilon_{it-1}} + q_\tau \frac{\partial \sigma_{it}}{\partial \epsilon_{it-1}},$$

and estimate mean marginal effects at different parts of the logwage distribution as

$$\widehat{E}\left(\frac{\partial Q_\tau(y_{it})}{\partial \epsilon_{it-1}}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{\partial Q_\tau(y_{it})}{\partial \epsilon_{it-1}}\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{\partial \mu_{it}}{\partial \epsilon_{it-1}} + q_\tau \frac{\partial \sigma_{it}}{\partial \epsilon_{it-1}}\right].$$

In general, we would obtain a impulse-response function considering different periods,

$$\frac{\partial Q_\tau(y_{it})}{\partial \epsilon_{it-s}} = \frac{\partial \mu_{it}}{\partial \epsilon_{it-s}} + q_\tau \frac{\partial \sigma_{it}}{\partial \epsilon_{it-s}}, \text{ for } s > 1.$$

In particular, for my model I have

$$\begin{aligned}
\mu_{it} &= \alpha y_{it-1} + \eta_{1i}, \\
\sigma_{it} &= h_{it}^{1/2} = \exp\left\{\frac{1}{2}\left(\eta_{2i} + \beta\left[\left(\epsilon_{it-1}^2 + \Lambda\right)^{1/2} - (2/\pi)^{1/2}\right]\right)\right\},
\end{aligned}$$

so that

$$\begin{aligned}
\frac{\partial \mu_{it}}{\partial y_{it-1}} &= \alpha, \\
\frac{\partial \sigma_{it}}{\partial y_{it-1}} &= \sigma_{it} \times \frac{\beta}{2} \times \frac{\epsilon_{it-1}}{\left[\epsilon_{it-1}^2 + \Lambda\right]^{1/2}} \times \frac{1}{h_{it-1}^{1/2}}, \\
\frac{\partial \mu_{it}}{\partial \epsilon_{it-1}} &= \alpha h_{it-1}^{1/2}, \\
\frac{\partial \sigma_{it}}{\partial \epsilon_{it-1}} &= \sigma_{it} \times \frac{\beta}{2} \times \frac{\epsilon_{it-1}}{\left[\epsilon_{it-1}^2 + \Lambda\right]^{1/2}}, \\
\frac{\partial \mu_{it}}{\partial \epsilon_{it-2}} &= \alpha^2 h_{it-2}^{1/2} + \frac{1}{2}\alpha\beta h_{it-1}^{1/2} \times \frac{\epsilon_{it-1}\epsilon_{it-2}}{\sqrt{\epsilon_{it-2}^2 + \Lambda}}, \\
\frac{\partial \sigma_{it}}{\partial \epsilon_{it-2}} &= \sigma_{it} \times \frac{\beta}{2} \times \frac{\epsilon_{it-1}}{\left[\epsilon_{it-1}^2 + \Lambda\right]^{1/2}} \times \sigma_{it-1} \times \frac{\beta}{2} \times \frac{\epsilon_{it-2}}{\left[\epsilon_{it-2}^2 + \Lambda\right]^{1/2}},
\end{aligned}$$

and so on.

References

[1] Abowd J, Card D. 1989. On the Covariance Structure of Earnings and Hours Changes. *Econometrica* **57**: 441-445.

[2] Albarrán P. 2004. Modelos Econométricos de Varianza Condicional para Datos de Panel, Universidad Complutense de Madrid, Doctoral Thesis, Chapter 3.

[3] Alvarez J, Arellano M. 2004. Robust likelihood estimation of dynamic panel data models. *CEMFI Working Paper* 0421, CEMFI, Madrid.

[4] Arellano M. 2003a. Discrete Choices with Panel Data. *Investigaciones Económicas* **27**: 423-458.

[5] Arellano M. 2003b. *Panel Data Econometrics*. Oxford University Press: Advanced Texts in Econometrics, Oxford.

[6] Arellano M, Bonhomme s. 2009a. Robust Priors in Nonlinear Panel Data Models. *Econometrica* **77**: 489-536.

[7] Arellano M, Bonhomme s. 2009a. Identifying Distributional Characteristics in Random Coefficients Panel Data Models, unpublished manuscript.

[8] Arellano M, Hahn J. 2006. A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects, unpublished manuscript.

[9] Arellano M, Hahn J. 2007. Understanding Bias in Nonlinear Panel Models: Some Recent Developments. In *Advances in Economics and Econometrics*, Blundell R, Newey W, Persson T (eds). Cambridge University Press: Cambridge.

[10] Arellano M, Honoré B. 2001. Panel Data Models. Some Recent Developments. In *Handbook of Econometrics*, Heckman J, Leamer E (eds). North Holland: Amsterdam.

[11] Autor DH, Katz LF, Kearney MS. 2008. Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics* **90**: 300-323.

[12] Barsky B, Juster FT, Kimball M, Shapiro M. 1997. Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. *Quarterly Journal of Economics* **112**: 537-579.

[13] Bester A, Hansen C. 2007. Frequentist Bias Reduction via Bayesian Priors, unpublished manuscript.

[14] Bester A, Hansen C. 2009. A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects. *Journal of Business and Economic Statistics* **27**: 131-148.

[15] Bollerslev T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31**: 307-327.

[16] Bound J, Brown C, Mathiowitz N. 2001. Measurement Error in Survey Data. In *Handbook of Econometrics*, Heckman J, Leamer E (eds). North Holland: Amsterdam.

[17] Carro J. 2007. Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects. *Journal of Econometrics* **140**: 503-528.

[18] Chamberlain G. 1984. Panel Data. In *Handbook of Econometrics*, Griliches Z, Intriligator MD (eds). North Holland: Amsterdam.

[19] Chamberlain G, Hirano K. 1999. Predictive Distributions Based on Longitudinal Earnings Data. *Annales d'Economie et de Statistique* **55-56**: 211-242.

[20] Cox DR, Reid N. 1987. Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society, Series B* **49**: 1-39.

[21] Dhaene G, Jochmans K. 2009. Split-panel jackknife estimation of fixed effects models, unpublished manuscript.

[22] DiCiccio TJ, Stern SE. 1993. An adjustment to Profile Likelihood Based on Observed Information. Technical Report, Department of Statistic, Stanford University.

[23] Fernández-Val I. 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* **150**: 71-85.

[24] Fitzgerald J, Gottschalk P, Moffitt R. 1998. An Analysis of Sample Attrition in Panel Data: The Michigan Panel of Income Dynamics. *Journal of Human Resources* **33**: 251-299.

[25] Gottschalk P, Moffitt R. 1993. The Growth of Earnings Instability in the U.S. Labor Market. *Brooking Papers on Economic Activity* **2**: 217-272.

[26] Gottschalk P, Moffitt R. 2009. The Rising Instability of U.S. Earnings. *Journal of Economic Perspectives* **23**: 3-24.

[27] Guvenen F. 2009. An Empirical Investigation of Labor Income Processes. *Review of Economic Dynamics* **12**: 58-79.

[28] Hahn J, Kuersteiner G. 2004. Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects, unpublished manuscript.

[29] Hahn J, Newey W. 2004. Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica* **72**: 1295-1319.

[30] Hall RE, Mishkin FS. 1982. The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households. *Econometrica* **50**: 461-81.

[31] Horowitz JL. 2003. Bootstrap Methods for Markov Processes. *Econometrica* **71**: 1049-1082.

[32] Horowitz JL, Markatou M. 1996. Semiparametric Estimation of Regression Models for Panel Data. *Review of Economic Studies* **63**: 145-168.

[33] Jensen ST, Shore SH. 2008. Changes in the Distribution of Income Volatility, unpublished manuscript.

[34] Lancaster T. 2002. Orthogonal Parameters and Panel Data. *Review of Economic Studies* **69**: 647-666.

[35] Levy F, Murnane RJ. 1992. U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations. *Journal of Economic Literature* **30**: 1333-1381.

[36] Lillard LA, Willis RJ. 1978. Dynamic Aspects of Earnings Mobility. *Econometrica* **46**: 985-1012.

[37] Lin CC. 2005. Estimation and Testing of the ARCH(1) effect in Panel Data when N and T are Large, unpublished manuscript.

[38] MaCurdy TE. 1982. The Use of Time-Series Processes to Model the Error Structure of Earnings in Longitudinal Data Analysis. *Journal of Econometrics* **18**: 83-114.

[39] Meghir C, Windmeijer F. 1999. Moment conditions for dynamic panel data models with multiplicative individual effects in the conditional variance. *Annales d'Economie et de Statistique* **55-56**: 317-330.

[40] Meghir C, Pistaferri L. 2004. Income variance dynamics and heterogeneity. *Econometrica* **72**: 1-32.

[41] Nelson DB. 1992. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **59**: 347-370.

[42] Neyman J, Scott EL. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16**: 1-32.

[43] Pace L, Salvan A. 2006. Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference* **136**: 3554-3564.

[44] Severini TA. 1998. An Approximation to the Modified Profile Likelihood Function. *Biometrika* **85**: 403-411.

[45] Severini TA. 2002. *Likelihood Methods in Statistics*, Oxford University Press: Oxford.

[46] Woutersen T. 2002. Robustness against Incidental Parameters, unpublished manuscript.

TABLES

Table 1. AR(1)-EARCH(1) model with scalar fixed effects.

Properties of		T=8						T=16					
$\hat{\theta}=\left(\hat{\alpha},\hat{\beta}\right)$		Bias	MAE	SD	SE	$CR_{0.95}$	Failure	Bias	MAE	SD	SE	$CR_{0.95}$	Failure
$\alpha=0.5$	MLE	-0.002	0.021	0.030	0.029	0.91	0.00	0.000	0.012	0.017	0.018	0.92	0.00
	BCE	0.003	0.024	0.035	0.036	0.97	0.08	0.000	0.010	0.017	0.018	0.94	0.02
$\beta=0.5$	MLE	-0.638	0.638	0.162	0.165	0.01		-0.138	0.138	0.066	0.062	0.28	
	BCE	-0.421	0.421	0.228	0.244	0.63		-0.040	0.047	0.066	0.064	0.96	
$\alpha=0.5$	MLE	-0.022	0.030	0.036	0.035	0.91	0.00	-0.001	0.011	0.015	0.017	0.91	0.00
	BCE	-0.002	0.022	0.029	0.034	0.96	0.07	-0.003	0.010	0.017	0.020	0.93	0.01
$\beta=0.0$	MLE	-0.722	0.722	0.139	0.166	0.00		-0.214	0.214	0.064	0.063	0.05	
	BCE	-0.111	0.111	0.268	0.295	0.94		-0.012	0.031	0.090	0.086	0.97	
$\alpha=0.0$	MLE	0.005	0.023	0.037	0.035	0.93	0.00	0.001	0.015	0.020	0.020	0.90	0.00
	BCE	0.018	0.041	0.051	0.049	0.99	0.03	0.013	0.032	0.034	0.029	0.92	0.01
$\beta=0.5$	MLE	-0.633	0.633	0.163	0.161	0.01		-0.138	0.138	0.064	0.059	0.28	
	BCE	-0.202	0.202	0.352	0.278	0.99		0.036	0.050	0.063	0.064	0.96	

Note: Bias=median bias, MAE=median absolute error, SD =sample standard deviation, SE=bootstrap standard error, CR=fraction of times the .95 bootstrap CI contained the true value, Failure=fraction of times of divergence or failure to converge.

DGP: The process was started at $y_{i0}=0$, then 700 time periods are generated before the sample actually starts.

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1} + h_{it}^{1/2}\epsilon_{it}, \ \ (t=1,...,T;\ i=1,...,n) \\
h_{it} &= \exp\left(\eta_{2i} + \beta\left[\sqrt{\epsilon_{it-1}^2+\Lambda} - \sqrt{2/\pi}\right]\right), \\
\epsilon_{it} &\sim N(0,1), \eta_{2i} \sim N\left(-2,1\right), n=200.
\end{aligned}
$$

100 Monte Carlo replications are used for each design, with just ϵ_{it} redrawn in each replication.

y_i was drawn 50 times, at each stage, to obtain the simulated data for the individual block-bootstrap.

Table 2. AR(1)-EARCH(1) model with multiple fixed effects.

Properties of		T=8						T=16					
$\hat{\theta}=\left(\hat{\alpha},\hat{\beta}\right)$		Bias	MAE	SD	SE	$CR_{0.95}$	Failure	Bias	MAE	SD	SE	$CR_{0.95}$	Failure
$\alpha=0.5$	MLE	-0.257	0.257	0.061	0.060	0.01	0.00	-0.103	0.103	0.023	0.020	0.01	0.00
	BCE	-0.121	0.121	0.111	0.111	0.83	0.10	-0.040	0.051	0.059	0.054	0.87	0.07
$\beta=0.5$	MLE	-0.568	0.568	0.240	0.235	0.82		-0.044	0.059	0.097	0.098	0.92	
	BCE	-0.050	0.105	0.317	0.276	0.94		0.054	0.055	0.097	0.094	0.70	
$\alpha=0.5$	MLE	-0.263	0.263	0.036	0.033	0.00	0.00	-0.108	0.108	0.026	0.021	0.01	0.00
	BCE	-0.084	0.084	0.134	0.126	0.96	0.09	-0.042	0.050	0.053	0.064	0.98	0.07
$\beta=0.0$	MLE	-0.079	0.079	0.014	0.016	0.18		-0.090	0.090	0.020	0.018	0.17	
	BCE	0.048	0.082	0.289	0.238	0.93		-0.021	0.068	0.107	0.104	0.93	
$\alpha=0.0$	MLE	-0.167	0.167	0.060	0.057	0.07	0.00	-0.059	0.059	0.021	0.020	0.18	0.00
	BCE	-0.052	0.052	0.079	0.082	0.94	0.13	0.008	0.015	0.039	0.039	0.98	0.11
$\beta=0.5$	MLE	-0.608	0.608	0.266	0.222	0.68		-0.057	0.071	0.101	0.094	0.84	
	BCE	0.050	0.050	0.257	0.265	0.95		-0.049	0.050	0.134	0.098	0.98	

Note: Bias=median bias, MAE=median absolute error, SD =sample standard deviation, SE=bootstrap standard error, CR=fraction of times the .95 bootstrap CI contained the true value, Failure=fraction of times of divergence or failure to converge.

DGP: The process was started at $y_{i0}=0$, then 700 time periods are generated before the sample actually starts.

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1}+\eta_{1i}+h_{it}^{1/2}\epsilon_{it}, \ \ (t=1,...,T;\ i=1,...,n)\\
h_{it} &= \exp\left(\eta_{2i}+\beta\left[\sqrt{\epsilon_{it-1}^{2}+\Lambda}-\sqrt{2/\pi}\right]\right),\\
\epsilon_{it} &\sim N(0,1), \eta_{1i}\sim N\left(0,1\right), \eta_{2i}\sim N\left(-2,1\right), n=200.
\end{aligned}
$$

100 Monte Carlo replications are used for each design, with just ϵ_{it} redrawn in each replication.

y_i was drawn 50 times, at each stage, to obtain the simulated data for the individual block-bootstrap.

Table 3. α and β estimates.

	[1] Model Eq.(3)		[2] Job Changes		[3] Attrition		[4] Model Eq.(4)	
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$
MLE	0.4822	0.4832	0.3768	0.0642	0.5659	0.5245	0.4904	0.3713
	(0.0114)	(0.0495)	(0.0146)	(0.0985)	(0.0159)	(0.0439)	(0.0091)	(0.0306)
BCE	0.5690	0.5790	0.4569	0.0757	0.6056	0.5693	0.5432	0.4145
(m=2)	(0.0388)	(0.0974)	(0.0360)	(0.1531)	(0.0310)	(0.0908)	(0.0095)	(0.0321)
Sample	Individuals	Observations	Individuals	Observations	Individuals	Observations	Individuals	Observations
size	2,066	32,066	1,346	17,485	921	18,645	2,066	32,066

Note: Bootstrap SE in parentheses.

Table 4. Correlations with observed variables.

Dependent variable: $\hat{\eta}_{2i}$	[1]	[2]	[3]	[4]
Birth Cohort	0.0088	0 .0106	0.0031	0.0074
	(0.0025)	(0.0026)	(0.0027)	(0.0029)
Married	-0.5122	-0.4677	-0.4708	-0.3703
	(0.0673)	(0.0673)	(0.0673)	(0.0674)
White	-0.6189	-0.4385	-0.4587	-0.4463
	(0.0820)	(0.0863)	(0.0864)	(0.0865)
Technical Workers		-0.4582	-0.4951	-0.4426
		(0.0863)	(0.0867)	(0.0986)
Administrators		-0.4374	-0.4938	-0.4781
		(0.1037)	(0.1036)	(0.1083)
Sales workers		0.2175	0.2311	0.1636
		(0.1125)	(0.1123)	(0.1125)
Services workers		0.3259	0.2871	0.1859
		(0.1095)	(0.1100)	(0.1088)
Operatives workers		0.0821	0.0872	0.0245
		(0.0935)	(0.0935)	(0.0932)
Movers			0.8019	0.5504
			(0.0702)	(0.0761)
Dropout				0.2158
				(0.1181)
Graduate				-0.0446
				(0.0748)
Tenure: 1-2 years				0.0079
				(0.1677)
Tenure: 2-3 years				-0.1614
				(0.1171)
Tenure: 4-9 years				-0.4318
				(0.1053)
Tenure: 9-19 years				-0.8226
				(0.0988)
Tenure: 20 years or more				-0.8870
				(0.1002)
Constant	-18.3945	-21.9209	-7.8407	-15.6994
	(4.9083)	(5.0164)	(5.2528)	(5.6893)

Note: Bootstrap SE in parentheses. Number of observations=2066 individuals.
Omitted group: Craftsman workers, Stayers, College, Tenure<1 year.

Table 5. Distribution of Residuals and Standarized Residuals in First Differences.

	Kurtosis	
Residuals in First Differences	21.3237	(1.0024)
Standarized Residuals in First Differences	3.4598	(0.0977)

Note: Bootstrap SE in parentheses.

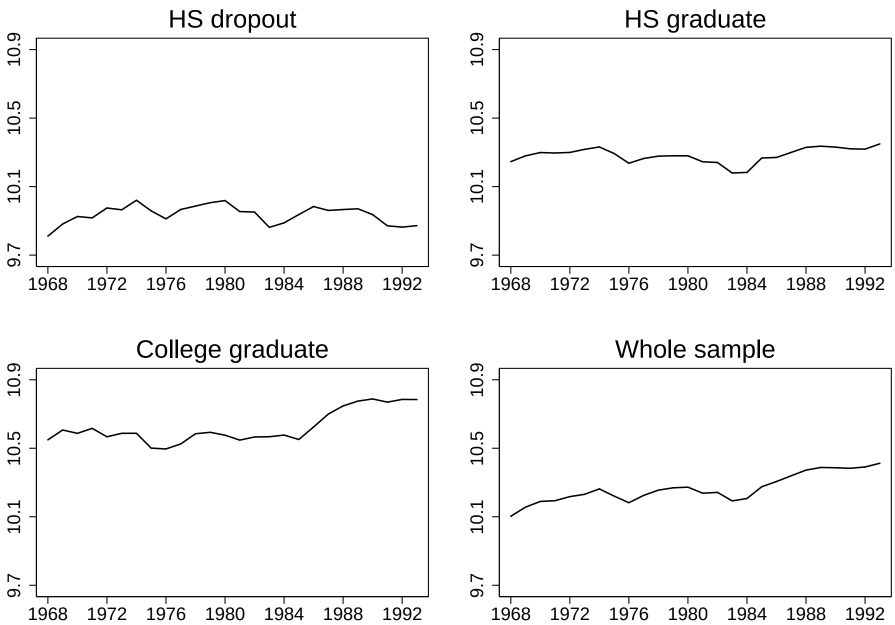
Table 6. Mean Marginal Effects with respect to past shocks at different quantiles.

τ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
ϵ_{it-1}	0.2543	0.2589	0.2622	0.2650	0.2677	0.2703	0.2732	0.2765	0.2811
	(0.0207)	(0.0206)	(0.0205)	(0.0205)	(0.0205)	(0.0205)	(0.0205)	(0.0205)	(0.0205)
ϵ_{it-2}	0.1455	0.1452	0.1451	0.1449	0.1448	0.1447	0.1445	0.1444	0.1441
	(0.0193)	(0.0190)	(0.0188)	(0.0184	(0.0184)	(0.0182)	(0.0180)	(0.0178)	(0.0175)

Note: Bootstrap SE in parentheses.

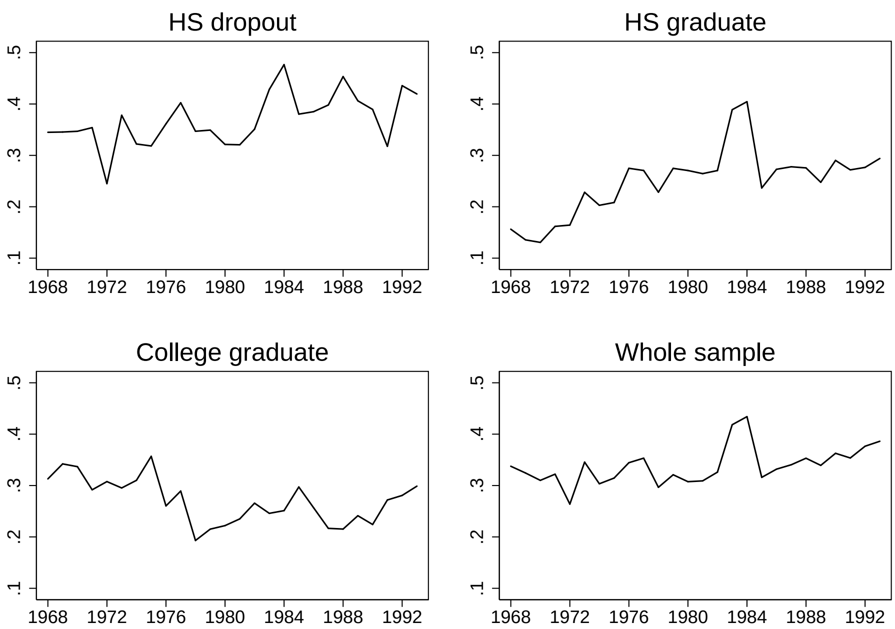
FIGURES

Figure 1. The mean of log wages.



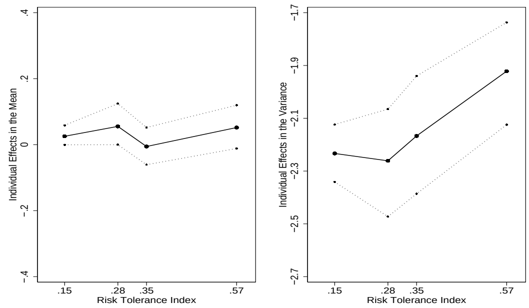
Note: PSID 1968-1993. Whole sample size: 2,066 individuals and 32,066 observations.

Figure 2. The variance of log wages.



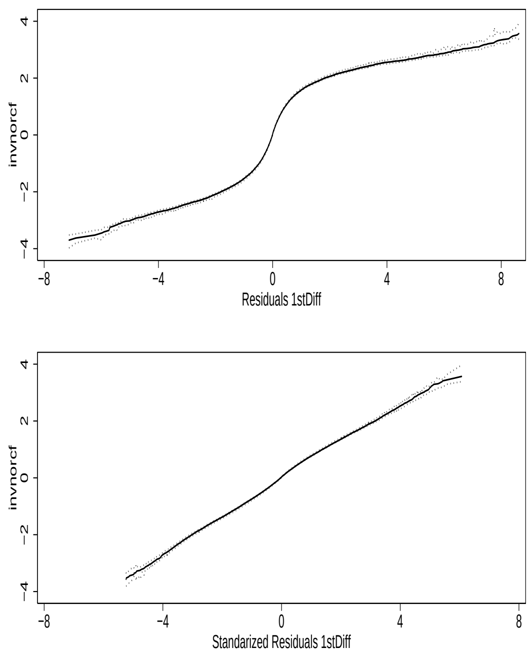
Note: PSID 1968-1993. Whole sample size: 2,066 individuals and 32,066 observations.

Figure 3. Correlations with Risk Tolerance.



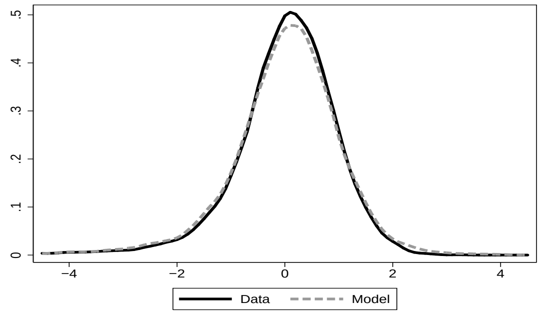
Note: PSID 1996, module on risk preferences data. Sample size: 1,535 individuals. Dotted lines represent 95% bootstrap confidence intervals.

Figure 4. Distribution of Residuals and Standarized Residuals in First Differences.



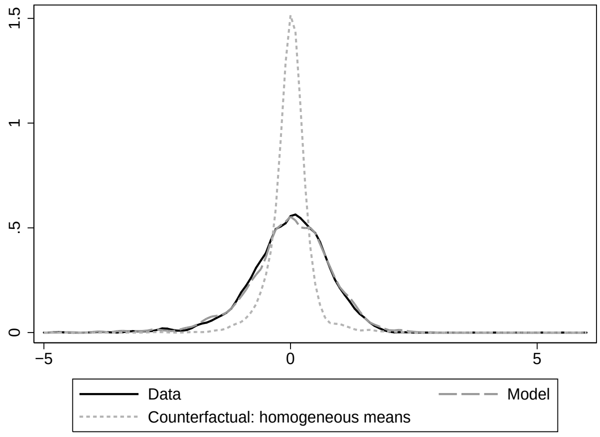
Note: PSID 1970-1993. Sample size: 27,934 observations. Dotted lines represent 95% pointwise bootstrap confidence intervals. See footnote 33 for the definition of estimated residuals and estimated standardized residuals.

Figure 5. Kernel densities of logwages and simulated logwages.



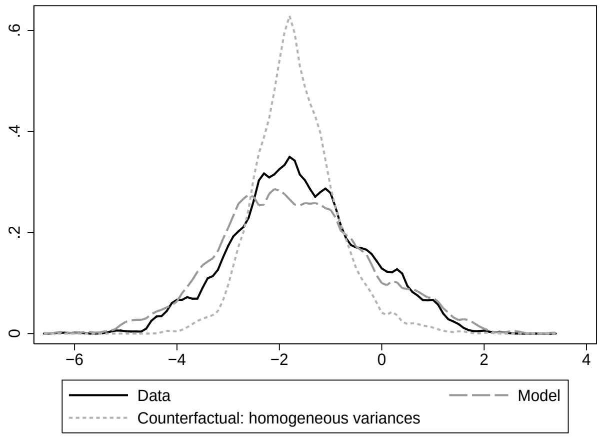
Note: PSID 1968-1993. Sample size: 32,066 observations. Bandwidth 0.10.

Figure 6. Kernel density of individual means.



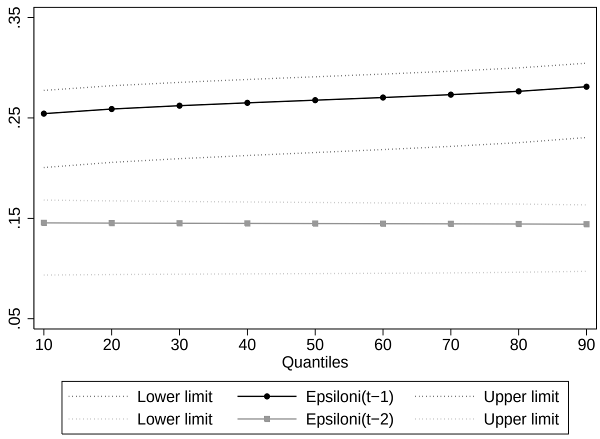
Note: PSID 1968-1993. Sample size: 32,066 observations. Bandwidth 0.10.

Figure 7. Kernel density of individual logvariances.



Note: PSID 1968-1993. Sample size: 32,066 observations. Bandwidth 0.10.

Figure 8. Mean Marginal Effects of Past Shocks over the Distribution of Wages.



Note: PSID 1968-1993. Sample size: 32,066 observations. Dotted lines represent 95% bootstrap confidence intervals.