MODELLING HETEROGENEITY AND DYNAMICS IN THE VOLATILITY OF INDIVIDUAL WAGES

Laura Hospido

CEMFI Working Paper No. 0717

December 2007

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551. Fax (34) 914 291 056
Internet: www.cemfi.es

I wish to thank Manuel Arellano, and Stéphane Bonhomme for encouragement and advice. My special thanks are also due to Richard Blundell for his suggestions and time. For comments and support, I am grateful to Pedro Albarrán, Jesús Carro, Jinyong Hahn, Alfredo Martín, Costas Meghir, Javier Mencía, Pedro Mira, Enrique Sentana, Gema Zamarro, and the seminar participants at CEMFI, University College London, Universidad Pablo de Olavide, Universitat Autònoma de Barcelona, Universidad Carlos III de Madrid – Statistics department, Universidad Jaume I de Castellón, Université Catholique de Louvain, Maastricht University, Bank of Spain, Stockholm School of Economics, the 9th IZA European 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 XXXI Simposio de Análisis Económico in Oviedo, the Royal Economic Society Job Market Conference in London, and the EEA-ESEM2007 meeting in Budapest. All errors are mine. I gratefully acknowledge financial support from Fundación Ramón Areces. The opinions and analyses are the responsibility of the author and, therefore, do not necessarily coincide with those of the Bank of Spain or the Eurosystem.

CEMFI Working Paper 0717
December 2007

MODELLING HETEROGENEITY AND DYNAMICS IN THE VOLATILITY OF INDIVIDUAL WAGES

Abstract

In this paper I consider a model for the heterogeneity and dynamics of the conditional mean and the conditional variance of standarized individual wages. In particular, I propose a dynamic panel data model with individual effects both in the mean and in a conditional ARCH type variance function. I posit a distribution for earning shocks and I build a modified likelihood function for estimation and inference in a fixed-T context. Using a newly developed bias-corrected likelihood approach makes it possible to reduce the estimation bias to a term of order $1/T^2$. The small sample performance of bias corrected estimators is investigated in a Monte Carlo simulation study. The simulation results show that the bias of the maximum likelihood estimator is substantially corrected for designs that are broadly calibrated to the PSID. The empirical analysis is conducted on data drawn from the 1968-1993 PSID. I find that it is important to account for individual unobserved heterogeneity and dynamics in the variance, and that the latter is driven by job mobility. I also find that the model 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.

Laura Hospido
Bank of Spain
laura.hospido@bde.es

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 distributions, building predictive earnings distributions, or calibrating consumption and saving models.

While several papers have focused on modelling the heterogeneity and time series properties of the conditional mean of earnings given its past (Lillard and Willis, 1978; MaCurdy, 1982; Abowd and Card, 1982, among others), the modelling of the conditional variance has been mostly neglected. However, in many applications it is important to understand the behavior of higher order moments of the process. This would be 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 uncertainty, she will be interested in forecasting not only the level of earnings but also its variance. The properties of the variance would be important for describing wage profiles over time and for better understanding what drives fluctuations in them. A richer specification can contribute also to modelling choices in models that use the earnings process as an input. In fact, 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).

There are also many papers that study the increase in the cross-sectional variance of earnings in the American economy since the 70's until today (see Katz and Autor, 1999). This growth in the aggregate variance is associated with an increase in inequality. Much less is known about the behaviour of the conditional variance given observed and unobserved individual characteristics.

In this paper, I propose a likelihood-based panel data model for the heterogeneity and dynamics of the conditional mean and the conditional variance of individual wages. In particular, I build a dynamic panel data model with linear individual effects in the mean and multiplicative individual effects in the conditional ARCH type variance function. Therefore, with this model, we can say to what extent the time evolution of the variance is determined by permanent individual heterogeneity or by state dependence effects. This distinction would be crucial, for instance, in the case of precautionary savings as the consumer would behave 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.

It is well known that failure to control for individual unobserved 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 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, which biases estimates of model parameters. In particular, the bias of the maximum likelihood estimator is of order $1/T$. The number of periods available for many panel data sets is such that it is not less natural to talk of time-series finite sample bias than of fixed-T inconsistency or underidentification. In this light, an alternative 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, 2006a, for a review). There are automatic methods based on simulation (Hahn and Newey, 2004), bias correction based on orthogonalization (Cox and Reid, 1987; Lancaster, 2002) and their extensions (Woutersen, 2002; Arellano, 2003), analytical bias correction of estimators (Hahn and Newey, 2004; Hahn and Kuersteiner, 2004), bias correction of the moment equation (Carro, 2006; Fernández-Val, 2005) and bias corrections for the concentrated likelihood (DiCiccio and Stern, 1993; Severini, 1998a; Pace and Salvan, 2005).

Following this perspective, I build 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. 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 contributions of the paper are twofold. First, I develop several versions of the modified likelihood based on DiCiccio and Stern (1993), Severini (1998a), Pace and Salvan (2005), and Arellano and Hahn (2006b) adapted to a dynamic conditional variance model. Second, I show how this approach works in practice for a specific empirical setting. The small sample performance of bias corrected estimators is investigated in a Monte Carlo study. The simulation results show that the bias of the maximum likelihood estimator is substantially corrected for samples designs that are broadly calibrated to the one used in the empirical application. The empirical analysis is conducted on data drawn from the 1968-1993 Panel Study of Income Dynamics (PSID). These models and data are interesting because we do not know much how the volatilities of individual wages behave in a period of increasing aggregate inequality. I find that it is important to account for individual unobserved heterogeneity and dynamics in the variance, and that the latter is driven by job mobility. I also find that the model explains the non-normality observed in logwage data.

In a similar sample for male earnings, Meghir and Pistafferi (2004) find strong evidence of state dependence effects as well as evidence of unobserved heterogeneity in the variances[1]. They also propose an autoregressive conditional heteroskedasticity panel data model of earnings dynamics, but they separate into a permanent component and a transitory component of earnings shocks. This can be appropriate in models where the author makes assumptions about the nature of the different shocks that affect the income process. Nevertheless, a model with a permanent component $I(1)$ imposes a unit root, i.e., a value for the autoregressive coefficient in the mean equal to one, whereas recent evidence suggests a value for this coefficient around $0.4 - 0.5$ (Alvarez and Arellano, 2004). I use a single-shock, multiple effects model instead[2]. This parsimonious specification would be useful for describing and estimating wage distributions (Chamberlain and Hirano, 1999). Meghir and Pistaferri recover orthogonality conditions for the estimation. Their method depends critically on the linear specification for the variance. But even in this case, they recognize that they cannot do fixed-T consistent GMM estimation because they have weak instruments. So, they implement a WG-GMM estimator which is only consistent when $T \to \infty$. What is specially worrying about this is that they have a bias of order $1/T$ as opposed to my estimator

[1]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-lineal" setting. Here we consider a different econometric framework, which let us handle models with multiple effects and estimators without being constrained to the availability of differencing schemes.

[2]Meghir and Windmeijer (1999) and Albarrán (2004) use single-shock models as well but they do not have an application to data.

which has a bias of order $1/T^2$. This difference is very important, as we will see in the simulations with respect to the MLE which also has a bias of order $1/T$. Even worse, because the WG-GMM estimator use arbitrary moment conditions it is thus less efficient than MLE. I choose an exponential specification that 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). What is interesting is that the estimation method does not depend on tricks applicable to the particular specification. It could also be used without major changes considering a quadratic specification as the one of Meghir and Pistaferri, or in other models.

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 is important for PSID wages 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 organised as follows. Section 2 presents the panel nonlinear dynamic model and the likelihood function. Section 3 reviews the alternative approaches for correcting the concentrated likelihood adapted to this particular setting. Section 4 shows some simulations to study the finite sample performance of the bias corrections for the concentrated likelihood. In Section 5, I present the empirical application on individual wages and in Section 6 the implications of the model for consumption growth. Section 7 concludes with some remarks on a future research agenda.

2 The Model and the Likelihood Function

2.1 The Model

I consider the following model of standardized logwages where i and t index individuals and time, respectively:[3]

$$y_{it} = \alpha y_{it-1} + \eta_i + e_{it} = \alpha y_{it-1} + \eta_i + h_{it}^{1/2} \epsilon_{it};\ \ (i = 1, ..., N; t = 1, ..., T)$$

with

$$E\left(y_{it} | y_i^{t-1}, \Theta_i\right) = \alpha y_{it-1} + \eta_i,$$

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

and

$$
\begin{aligned}
h_{it} &= Var\left(y_{it}|y_i^{t-1},\Theta_i\right) = E\left(e_{it}^2|y_i^{t-1},\Theta_i\right) \\
&= \exp\left(\psi_i + \beta\left[|\epsilon_{it-1}| - E\left(|\epsilon_{it-1}|\right)\right]\right) \\
&= h\left(\epsilon_{it-1},\psi_i\right).
\end{aligned}
$$

In these expressions, $\{y_{i0},...,y_{iT}\}_{i=1}^{N}$ are the observed data[4], $\Theta_i = (\eta_i,\psi_i)'$ are the individual unobserved fixed effects, e_{it} is an ARCH process, and $\{\epsilon_{it}\}$ is an *i.i.d.* sequence with zero mean and unit variance[5]. The log formulation implies that h_{it} is always nonnegative, regardless of the parameter values (Nelson, 1992). Finally, I denote the vector of common parameters as $\Gamma = (\alpha,\beta)'$.

For the conditional mean, I consider an autoregressive specification where the parameter α measures the persistence on the level of wages to shocks, η_i describe permanent unobserved heterogeneity and e_{it} reflects shocks that individuals receive every period[6]. Departing for the classical AR(1) process, I permit that the variances, given past observations, change over time and across individuals. This particular ARCH type specification allows me to capture two patterns of wage volatility. The first one is individual heterogeneity, ψ_i: wage volatilities of different individuals can vary differently. For instance, there can be different variances 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, reflecting the persistence on the volatility of wages to shocks.

2.2 The Likelihood Function

Under the assumption that $\epsilon_{it} \sim N(0,1)$, that is, $\epsilon_{it}|y_i^{t-1},\Theta_i \sim N(0,1)$ then, conditional on the past, the model is normal heteroscedastic

$$
y_{it}|y_i^{t-1},\Theta_i \sim N(\alpha y_{it-1} + \eta_i, h_{it}),
$$

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

$$
f\left(y_{i1},...,y_{iT}|y_{i0},\Theta_{i0}\right) = \prod_{t=1}^{T} f\left(y_{it}|y_{it-1},\Theta_{i0},\Gamma_0\right).
$$

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

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

[6] 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.

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 $t > 1$, we can write

$$\ln f\left(y_{it}|y_{it-1},\Theta_i,\Gamma\right)=\ell_{it}\left(\Gamma,\Theta_i\right)\propto -\frac{1}{2}\ln\left(h\left(\epsilon_{it-1},\psi_i\right)\right)-\frac{1}{2}\frac{\left(y_{it}-\alpha y_{it-1}-\eta_i\right)^2}{h\left(\epsilon_{it-1},\psi_i\right)}.$$

Initial conditions. Evaluation of the likelihood at $t = 1$ requires pre-sample values for ϵ_{it}^2 and h_{it}. For $t = 1$,

$$y_{i1}=\alpha y_{i0}+\eta_i+\left[h\left(\epsilon_{i0},\psi_i\right)\right]^{1/2}\epsilon_{i1},$$

where $h\left(\epsilon_{i0},\psi_i\right) = h\left(y_{i0},y_{i,-1},y_{i,-2},...\right)$. This is a model for $f\left(y_{i1}|y_{i0},y_{i(-1)},y_{i(-2)},...,\Theta_{i0}\right)$ or for $f\left(y_{i1}|y_{i0},\epsilon_{i0},\Theta_{i0}\right)$ where ϵ_{i0} resumes all the past values of y_{it}, but what we would need is $f\left(y_{i1}|y_{i0},\Theta_{i0}\right).$ Since,

$$E\left(y_{i1}|y_{i0},\Theta_{i0}\right)=E\left(y_{i1}|y_{i0},\epsilon_{i0},\Theta_{i0}\right)=\alpha y_{i0}+\eta_i,$$

and

$$\begin{aligned}
Var\left(y_{i1}|y_{i0},\Theta_{i0}\right) &= E\left(h\left(\epsilon_{i0},\psi_i\right)|y_{i0},\Theta_{i0}\right)+Var\left(\alpha y_{i0}+\eta_i|y_{i0},\Theta_{i0}\right)\\
&= E\left(h\left(\epsilon_{i0},\psi_i\right)|y_{i0},\Theta_{i0}\right)+Var\left(\eta_i|y_{i0},\Theta_{i0}\right)\\
&= \varphi\left(\eta_i,\psi_i,\Gamma\right).
\end{aligned}$$

Thus, $f\left(y_{i1}|y_{i0},\Theta_{i0}\right)$ would be a mixture given that:

$$f\left(y_{i1}|y_{i0},\Theta_{i0}\right)=\int f\left(y_{i1}|y_{i0},\epsilon_{i0},\Theta_{i0}\right)dG\left(\epsilon_{i0}|y_{i0},\Theta_{i0}\right).$$

For simplicity, I consider an approximate model where $y_{i1}|y_{i0},\Theta_{i0}\sim N\left(\alpha y_{i0}+\eta_i,h_{i1}\right)$ and, as suggested by Bollerslev (1986), I use the mean of the squared residuals as an estimate for $h_{i1}=\frac{1}{T}\sum\limits_{t=1}^{T}e_{it}^2$. [7] As $T\to\infty$, h_{i1} is the steady-state unconditional variance of e_{it} given fixed effects, that is,

$$\varphi\left(\eta_i,\psi_i,\Gamma\right)=p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(y_{it}-\alpha y_{it-1}-\eta_i\right)^2.$$

Let the individual likelihood function be

$$\pounds_i\left(\Gamma,\Theta_i\right)=\prod_{t=2}^{T}\frac{1}{\left[h\left(\epsilon_{it-1},\psi_i\right)\right]^{1/2}}\phi\left(\frac{y_{it}-\alpha y_{it-1}-\eta_i}{\left[h\left(\epsilon_{it-1},\psi_i\right)\right]^{1/2}}\right)\cdot\frac{1}{\left[h_{i1}\right]^{1/2}}\phi\left(\frac{y_{i1}-\alpha y_{i0}-\eta_i}{\left[h_{i1}\right]^{1/2}}\right),$$

[7] Another alternative would be adding the missing variances as parameters to be estimated.

and the log-likelihood of each observation

$$\ell_{it}(\Gamma,\Theta_i) = -\frac{1}{2}\ln(h_{it}) - \frac{1}{2}\frac{(y_{it}-\alpha y_{it-1}-\eta_i)^2}{h_{it}},$$

where

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

3 Correcting the Likelihood Function

In this section, I adopt a likelihood-based approach that allows me to deal with dynamics and multiple fixed effects in the estimation. The MLE of Γ, concentrating out the Θ_i, is the solution to

$$\widehat{\Gamma} \equiv \arg\max_{\Gamma} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\ell_{it}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right); \ \widehat{\Theta}_i(\Gamma) \equiv \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T}\ell_{it}(\Gamma,\Theta).$$

Incidental Parameters Problem. In this context, 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. The problem arises because the unobserved individual effects Θ_i are replaced by sample estimates $\widehat{\Theta}_i(\Gamma)$: as only a finite number T of observations are available to estimate each Θ_i, the estimation error of $\widehat{\Theta}_i(\Gamma)$ does not vanish as the sample size N grows, and this error contaminates the estimates of common parameters in nonlinear models. Let

$$L(\Gamma) \equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left[\sum_{t=1}^{T}\ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)\right].$$

Then, from the usual maximum likelihood properties, for $N\to\infty$ with T fixed, $\widehat{\Gamma}_T = \Gamma_T + o_p(1)$, where $\Gamma_T \equiv \arg\max_{\Gamma} L(\Gamma)$. In general, $\Gamma_T \neq \Gamma_0$, but $\Gamma_T \to \Gamma_0$ as $T\to\infty$.

Due to the noise in estimating $\widehat{\Theta}_i(\Gamma)$, the expectation of the concentrated likelihood is not maximized at the true value of the parameter. This problem can be avoided by correcting the concentrated likelihood.

The bias in the expected concentrated likelihood at an arbitrary Γ can be expanded in orders of magnitude of T

$$E\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right) - \frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\Gamma,\bar{\Theta}_i(\Gamma)\right)\right] = \frac{b_i(\Gamma)}{T} + o\left(\frac{1}{T}\right),$$

where $\bar{\Theta}_i(\Gamma)$ maximizes $\lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^{T}\ell_{it}(\Gamma,\Theta)\right]$. As it is shown in Appendix A, the form of the approximate bias of the concentrated likelihood is:

$$\frac{b_i(\Gamma)}{T} \approx \frac{1}{2}tr\left(H_i(\Gamma)\,Var\left[\widehat{\Theta}_i(\Gamma)\right]\right) = \frac{1}{2T}tr\left(H_i^{-1}(\Gamma)\,\Upsilon_i(\Gamma)\right),$$

where

$$H_i(\Gamma) = -\lim_{T\to\infty} E\left[\frac{\partial^2 \ell_i\left(\Gamma, \overline{\Theta}_i(\Gamma)\right)}{\partial\Theta_i \partial\Theta_i'}\right],$$

$$\Upsilon_i(\Gamma) = \lim_{T\to\infty} TE\left[\frac{\partial \ell_i\left(\Gamma, \overline{\Theta}_i(\Gamma)\right)}{\partial\Theta_i} \cdot \frac{\partial \ell_i\left(\Gamma, \overline{\Theta}_i(\Gamma)\right)}{\partial\Theta_i'}\right],$$

and

$$\ell_i(\Gamma, \Theta_i) = \frac{1}{T}\sum_{t=1}^{T} \ell_{it}(\Gamma, \Theta).$$

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 (2006b).

In this paper, I consider three alternative estimators of Γ which maximize a bias-corrected concentrated likelihood function like:

$$\begin{aligned}\widetilde{\Gamma} &= \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N} \ell_{mi}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right) \\ &= \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} \ell_{it}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right) - \frac{1}{T}\hat{b}_i(\Gamma)\right].\end{aligned}$$

Letting $\hat{b}_i(\Gamma)$ be an estimated bias, $\widetilde{\Gamma}$ is expected to be less biased than the MLE $\widehat{\Gamma}$.

Trace Based Approach. An estimator for the bias term in the modified likelihood would be

$$\widehat{b}_i(\Gamma) = \frac{1}{2} tr\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right).$$

Determinant Based Approach. In a likelihood context, it is appropriate to consider a local version of the estimated bias using that at the truth $H_i^{-1}(\Gamma_0)\,\Upsilon_i(\Gamma_0) = 1$ (Pace and Salvan, 2005). As it is shown at the end of Appendix A, this local version of $\widehat{b}_i(\Gamma)$ gives

$$\widehat{b}_i(\Gamma) = -\frac{1}{2}\ln\det \widehat{H}_i(\Gamma) + \frac{1}{2}\ln\det \widehat{\Upsilon}_i(\Gamma),$$

or

$$\frac{\widehat{b}_i(\Gamma)}{T} = \frac{1}{2}\ln\det \widehat{H}_i(\Gamma) + \frac{1}{2}\ln\det \widehat{Var}\left[\widehat{\Theta}_i(\Gamma)\right].$$

In practice, for estimating the bias I need to estimate the hessian term, $H_i(\Gamma)$, and the expected outer product term, $\Upsilon_i(\Gamma)$. For estimating the first one I use its sample counterpart:

$$\widehat{H}_i(\Gamma) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_{it}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right)}{\partial\Theta_i\partial\Theta_i'}.$$

With regard to $\Upsilon_i(\Gamma)$, note that since

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\frac{\partial \ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial \Theta_i}=0,$$

also

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\sum\nolimits_{s=1}^{T}\frac{\partial \ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial \Theta_i}\cdot\frac{\partial \ell_{is}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial \Theta_i'}=0,$$

so that using the observed quantities evaluated at $\widehat{\Theta}_i(\Gamma)$ will not work. The three different corrections, presented below, are based on three different estimators for this second term of the bias.

3.1 Trace Based Approach for Pseudo Likelihoods

Since $\Upsilon_i\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)=0$, a trimmed version of $\Upsilon_i(\Gamma)$ might work. That is,

$$\widehat{\Upsilon}_i(\Gamma)=\Omega_0+\sum_{l=1}^{r}\left(\Omega_l+\Omega_l'\right),$$

$$\Omega_l=\frac{1}{T-l}\sum_{t=l+1}^{T}\left(1-\frac{l}{r+1}\right)\frac{\partial \ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial \Theta}\cdot\frac{\partial \ell_{it-l}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial \Theta'}.$$

In principle r 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 r at both ends of the admissible range ($r=0$ and $r=T-1$), in practice r will be chosen equal to 2 or 3.

The trace based approach can be regarded as an objective-function counterpart to the bias-corrected estimator in Hahn and Kuersteiner (2004).

3.2 Determinant Based Approach Using Expected Quantities

This approach is based on the expectation

$$\begin{aligned}&\bar{\Upsilon}_i\left(\Gamma,\Theta_i;\Gamma_0,\Theta_{i0}\right)\\ \equiv\;& TE_{\{\Gamma_0,\Theta_{i0}\}}\left[\left[\frac{\partial \ell_i(\Gamma,\Theta_i)}{\partial\Theta_i}-E\left(\frac{\partial \ell_i(\Gamma,\Theta_i)}{\partial\Theta_i}\right)\right]\cdot\left[\frac{\partial \ell_i(\Gamma,\Theta_i)}{\partial\Theta_i'}-E\left(\frac{\partial \ell_i(\Gamma,\Theta_i)}{\partial\Theta_i'}\right)\right]\right]\end{aligned}$$

obtained using the true density $f\left(y_i|y_{i0},\Gamma_0,\Theta_{i0}\right)$. Notice that in this case (for an arbitrary (Γ,Θ_i)), centering the expected outer product term is crucial because only for $E\left(\frac{\partial \ell_i(\Gamma,\overline{\Theta}_i(\Gamma))}{\partial\Theta_i}\right)$ this expectation is equal to zero. Also it is important to note that this expected quantity can be obtained for given values of (Γ,Θ_i) and (Γ_0,Θ_{i0}), analytically or numerically, because in the likelihood context the density of the data is available. However, it cannot be calculated at (Γ_0,Θ_{i0}) because true values are unknown. The estimator solves this problem replacing (Γ_0, Θ_{i0}) by their ML estimates $\left(\widehat{\Gamma}, \widehat{\Theta}_i\right)$. This give us the useful quantity: $\widetilde{\Upsilon}_i \left(\Gamma, \widehat{\Theta}_i(\Gamma); \widehat{\Gamma}, \widehat{\Theta}_i\right)$.

This approach can be regarded as a dynamic version of Severini (1998a) or DiCiccio and Stern (1993) approximations to the modified profile likelihood.

Iterated Bias-Corrected Likelihood Estimation. An undesirable feature of this approach is its dependence on $\widehat{\Gamma}$, which may have a large bias. This problem can be avoided by considering an iterative procedure. That is, once a first corrected estimate is available,

$$\widetilde{\Gamma}_I = \arg\max_{\Gamma} \frac{1}{N} \sum_{i=1}^{N} \ell_{mi} \left(\Gamma, \widehat{\Theta}_i(\Gamma); \widehat{\Gamma}, \widehat{\Theta}_i\right),$$

I could use it to calculate a second one:

$$\widetilde{\Gamma}_{II} = \arg\max_{\Gamma} \frac{1}{N} \sum_{i=1}^{N} \ell_{mi} \left(\Gamma, \widehat{\Theta}_i(\Gamma); \widetilde{\Gamma}_I, \widehat{\Theta}_i\left(\widetilde{\Gamma}_I\right)\right).$$

Pursuing the iteration,

$$\widetilde{\Gamma}_K = \arg\max_{\Gamma} \frac{1}{N} \sum_{i=1}^{N} \ell_{mi} \left(\Gamma, \widehat{\Theta}_i(\Gamma); \widetilde{\Gamma}_{K-1}, \widehat{\Theta}_i\left(\widetilde{\Gamma}_{K-1}\right)\right),$$

until convergence, it would be possible to obtain an estimator $\widetilde{\Gamma}_\infty$ that solves

$$\sum_{i=1}^{N} q_{mi} \left(\Gamma, \widehat{\Theta}_i(\Gamma); \Gamma, \widehat{\Theta}_i(\Gamma)\right) = 0,$$

where $q_{mi}(\Gamma, \Theta_i; \Gamma_0, \Theta_{i0})$ denotes the score of $\ell_{mi}(\Gamma, \Theta_i; \Gamma_0, \Theta_{i0})$ for fixed Γ_0 and Θ_{i0}.

3.3 Determinant Based Approach Using Bootstrap

The first step consists in generating parametric bootstrap samples $\{y_{i1}^m, ..., y_{iT}^m\}_{i=1}^N$ with $(m = 1, ..., M)$ from the model $\left\{\prod_{t=1}^T f\left(y_{it}|y_{i0}, \widehat{\Gamma}, \widehat{\Theta}_i\right)\right\}_{i=1}^N$ and, then, in obtaining the corresponding fixed effects estimates $\left\{\widehat{\Theta}_i^m(\Gamma)\right\}_{m=1}^M$. This approach, close to Pace and Salvan (2005), is based on using a bootstrap estimate of $Var\left[\widehat{\Theta}_i(\Gamma)\right]$ given by

$$\widehat{Var}\left[\widehat{\Theta}_i(\Gamma)\right] = \frac{1}{M} \sum_{m=1}^{M} \left[\widehat{\Theta}_i^m(\Gamma) - \widehat{\Theta}_i(\Gamma)\right]^2.$$

4 Monte Carlo Evidence

The practical importance of these bias corrections depends on how much bias is removed for the relatively small T that is often relevant in econometric applications.

In this section, I provide some simple versions of the model showing that these corrections can remove a large part of the bias even with small T.

4.1 The linear dynamic panel model with fixed effects

Consistent estimates of α for fixed T are available in the AR(1) case. I consider this model first to compare the bias correcting estimators described above with the one proposed by Lancaster (2002).

The model design is

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1} + \eta_i + \epsilon_{it}, \ \ (t = 1, ..., T; \ i = 1, ..., N) \\
\epsilon_{it} &\sim N(0,1), \ \ \eta_i \sim N(0,1), \\
y_{i0} &\sim N\left(\frac{\eta_i}{(1-\alpha)}, \frac{1}{(1-\alpha^2)}\right).
\end{aligned}
$$

The data are generated for $T = 8$ and 16, $N = 500$ and 1000, and for $\alpha = 0.5$, and 0.8. I have simulated samples for different samples sizes because I expect the modified MLE to improve much more with T than with N. And I have also simulated samples for different values of α because the larger the α the greater the serial correlation of y_{it}, thus I expect that the estimator performs worse.

Here the MLE of α is

$$\hat{\alpha} \equiv \arg\max_{\alpha} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \sum_{t=1}^{T} \ell_{it} \left(\alpha, \hat{\eta}_i (\alpha) \right) \right] = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{it} \tilde{y}_{it-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{it-1}^2},$$

where

$$\hat{\eta}_i (\alpha) \equiv \arg\max_{\eta} \frac{1}{T} \sum_{t=1}^{T} \ell_{it} (\alpha, \eta) = \bar{y}_i - \alpha \bar{y}_{i(-1)},$$

and $\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}$, $\bar{y}_{i(-1)} = \frac{1}{T} \sum_{t=1}^{T} y_{it-1}$, $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{y}_{it-1} = y_{it-1} - \bar{y}_{i(-1)}$. I also consider several bias-correcting estimators of α that are obtained by maximization of a modified concentrated log likelihood like

$$\widetilde{\alpha} \equiv \arg\max_{\alpha} \frac{1}{N} \sum_{i=1}^{N} \ell_{mi} \left(\alpha, \widehat{\eta}_i (\alpha) \right).$$

- Trace Based Approach with Trimming: this approach uses a trimmed version of $\Upsilon_i(\alpha)$, that is,

$$\widehat{\Upsilon}_i(\alpha) = \Omega_0 + 2\sum_{l=1}^{r}\Omega_l,$$

where

$$\Omega_l = \frac{1}{T-l}\sum_{t=l+1}^{T}\left(1 - \frac{l}{r+1}\right)\frac{\partial \ell_{it}}{\partial \eta_i}\cdot\frac{\partial \ell_{it-l}}{\partial \eta_i},$$

for r small. So,

$$\ell_{mi}(\alpha, \hat{\eta}_i(\alpha)) = -\frac{1}{2T}\sum_{t=1}^{T}(y_{it} - \alpha y_{it-1} - \hat{\eta}_i(\alpha))^2 - \frac{1}{2T}\left(\widehat{H}_i^{-1}(\alpha)\,\widehat{\Upsilon}_i(\alpha)\right).$$

- Determinant Based Approach Using Expected Quantities: in this case,

$$\widehat{H}_i(\alpha) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_{it}(\alpha, \widehat{\eta}_i(\alpha))}{\partial \eta^2} = 1,$$

$$\begin{aligned}\bar{\Upsilon}_i(\alpha, \eta_i; \alpha_0, \eta_{i0}) &= TE_0\left[\left[\frac{\partial \ell_i(\alpha,\eta)}{\partial \eta_i} - E\left(\frac{\partial \ell_i(\alpha,\eta)}{\partial \eta_i}\right)\right]^2 \middle| y_{i0}\right] \\ &= TVar_0\left[\frac{\partial \ell_i(\alpha,\eta)}{\partial \eta_i}\middle| y_{i0}\right] = TVar\left[\bar{v}_i | y_{i0}\right].\end{aligned}$$

where $\bar{v}_i = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ell_{it}(\alpha,\eta)}{\partial \eta}$, [8] and as it is shown in Appendix B

$$\bar{\Upsilon}_i(\alpha, \eta; \alpha_0, \eta_0) = 1 + T(\alpha_0 - \alpha)^2\omega_T(\alpha_0) + 2T(\alpha_0 - \alpha)\psi_T(\alpha_0),$$

with

$$\begin{aligned}\omega_T(\alpha_0) &= \frac{1}{T^2}\left[1 + (1+\alpha_0)^2 + \left(1 + \alpha_0 + \alpha_0^2\right)^2 + \ldots + \left(1 + \alpha_0 + \ldots + \alpha_0^{T-2}\right)^2\right], \\ \psi_T(\alpha_0) &= \frac{1}{T^2}\left[\left(1 + \alpha_0 + \ldots + \alpha_0^{T-2}\right) + \left(1 + \alpha_0 + \ldots + \alpha_0^{T-3}\right) + \ldots + 1\right].\end{aligned}$$

Thus

$$\bar{\Upsilon}_i(\alpha, \widehat{\eta}_i(\alpha); \hat{\alpha}, \hat{\eta}_i) = 1 + T(\hat{\alpha} - \alpha)^2\omega_T(\hat{\alpha}) + 2T(\hat{\alpha} - \alpha)\psi_T(\hat{\alpha}).$$

It follows that in this case

$$\ell_{mi}(\alpha, \hat{\eta}_i(\alpha); \hat{\alpha}, \hat{\eta}_i) = -\frac{1}{2T}\sum_{t=1}^{T}(y_{it} - \alpha y_{it-1} - \hat{\eta}_i(\alpha))^2 - \frac{1}{2T}\ln\bar{\Upsilon}_i(\alpha, \hat{\eta}_i(\alpha); \hat{\alpha}, \hat{\eta}_i).$$

- Determinant Based Approach Using a Parametric Bootstrap Estimate of $Var[\hat{\eta}_i(\alpha)]$: now

$$\ell_{mi}(\alpha, \hat{\eta}_i(\alpha)) = -\frac{1}{2T}\sum_{t=1}^{T}(y_{it} - \alpha y_{it-1} - \hat{\eta}_i(\alpha))^2 - \frac{1}{2}\ln\widehat{Var}[\hat{\eta}_i(\alpha)],$$

[8] In what follows I omit the argument in ℓ_{it} for notational simplicity.

where

$$\widehat{Var}\left[\hat{\eta}_i(\alpha)\right] = \frac{1}{M}\sum_{m=1}^{M}\left[\hat{\eta}_i^m(\alpha) - \hat{\eta}_i(\alpha)\right]^2,$$

and m indexes the simulated samples by parametric bootstrap.

- Following Lancaster (2002), I consider the Approximate Conditional Likelihood:

$$\ell_{mi}(\alpha, \hat{\eta}_i(\alpha)) = -\frac{1}{2T}\sum_{t=1}^{T}\left(y_{it} - \alpha y_{it-1} - \hat{\eta}_i(\alpha)\right)^2 + \frac{b_T(\alpha)}{T},$$

where

$$b_T(\alpha) = \frac{1}{T}\left[\sum_{t=1}^{T-1}\left(\frac{T-t}{t}\right)\alpha^t\right].$$

Before presenting the results I want to mention that I use *Individual Block-Bootstrap,* that is, *fixed-T large-N non parametric bootstrap* for calculating the standard errors of the estimates. The assumption of independence across individual allows me to draw complete time series for each individual to capture the time series dependence, that is, I draw $y_i = (y_{i1}, ..., y_{iT})'$ S times to obtain the simulated data $\left\{y_i^{(s)}, y_{i(-1)}^{(s)}\right\}_{s=1}^{S}$. For each sample I obtain the corresponding estimates of α_0, $\left(\hat{\alpha}^{(1)}, ..., \hat{\alpha}^{(S)}\right)$, and the empirical distribution as an approximation of the distribution of $\hat{\alpha}$.[9]

Table 1 reports estimates, based on 300 Monte Carlo runs, for $T = 8$ and $N = 500$.

Table 1. Properties of $\hat{\alpha}$ ($T = 8$)

	$\alpha = 0.5$			$\alpha = 0.8$		
Estimator of α	Mean	SD	Mean SE	Mean	SD	Mean SE
MLE	0.2947	0.0173	0.0160	0.5263	0.0163	0.0156
Trimming	0.3782	0.0177	0.0197	0.5986	0.0158	0.0165
Expected Quantities	0.4365	0.0149	0.0151	0.6541	0.0146	0.0143
Bootstrap Variance	0.4745	0.0213	0.0193	0.7158	0.0182	0.0170
Lancaster	0.5006	0.0205	0.0197	0.7989	0.0240	0.0240

Note: N=500; simulations=300; parametric bootstrap samples=300; non parametric bootstrap samples=100; trimming=2. SD: Sample standard deviation. Mean SE: Mean of estimated standard errors by individual block-bootstrap.

I find some differences in the performance between these four types of bias corrections. I have also found that iterating bias correction, in the case of the first two corrections, improves a bit the estimation but for brevity I do not report here these results. An example of that is included in the next subsection. We see in the table that the fixed effects MLE is downward biased by around 35-40 percent in both cases. Bias corrections, except the one proposed by Lancaster (2002) that is consistent for fixed T, all

[9] Notice that, opposite to the block bootstrap procedure used in time-series literature (Hall and Horowitz, 1996; Horowitz, 2003), here I do not need to choose any bandwidth.

perform better when $\alpha = 0.5$. In this latter case, the corrections reduce the bias for at least a half. In addition, we can see that the mean of the standard errors estimated by individual block-bootstrap is a good approximation to the Monte Carlo standard deviation.

Table 2 presents estimates for $T = 16$ and $N = 500$.

Table 2. Properties of $\hat{\alpha}$ $(T = 16)$

	$\alpha = 0.5$			$\alpha = 0.8$		
Estimator of α	Mean	SD	Mean SE	Mean	SD	Mean SE
MLE	0.4008	0.0109	0.0106	0.6653	0.0097	0.0093
Expected Quantities	0.4589	0.0109	0.0111	0.7093	0.0096	0.0093
Bootstrap Variance	0.4962	0.0119	0.0115	0.7781	0.0106	0.0104
Trimming	0.4577	0.0106	0.0101	0.7093	0.0092	0.0089
Lancaster	0.4999	0.0119	0.0117	0.7993	0.0124	0.0119

Note: N=500; simulations=300; parametric bootstrap samples=200; non parametric bootstrap samples=100; trimming=2. SD: Sample standard deviation. Mean SE: Mean of estimated standard errors by individual block-bootstrap.

We can see that for $\alpha = 0.5$, the MLE has still an important bias, but the modified MLEs are closer to the true value. As before, corrections perform worse when $\alpha = 0.8$.

I do not report here the results for $N = 1000$, because increasing the number of individuals from $N = 500$ to $N = 1000$ has little effect on the magnitude of the estimated bias (much less effect that increasing T).

4.2 The linear dynamic panel model with multiple fixed effects

One of the advantages of the bias-correcting estimators with respect to the estimator proposed by Lancaster is their generality. With only a slight modification of the previous expressions it is possible to deal with a more complex model, as an AR(1) model with fixed effects in the conditional mean, η_i, and in the conditional variance, σ_i^2.

Now the model design is

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \eta_i + e_{it} = \alpha y_{it-1} + \eta_i + \sigma_i \epsilon_{it}, \ \ (t = 1, ..., T;\ i = 1, ..., N) \\ \epsilon_{it} &\sim N(0,1), \ \ \eta_i \sim N(0,1), \ \psi_i = \log \sigma_i^2 \sim N(-3.0, 0.8), \\ y_{i0} &\sim N\left(\frac{\eta_i}{(1-\alpha)}, \frac{\sigma_i^2}{(1-\alpha^2)}\right). \end{aligned}$$

The data are generated for $T = 8$ and 16, $N = 500$, and for $\alpha = 0.5$. I denote as $\Theta_i = \left(\eta_i, \sigma_i^2\right)'$ the vector of fixed effects. The MLE of α is

$$\begin{aligned}\hat{\alpha} &\equiv \arg\max_{\alpha} \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\alpha,\widehat{\Theta}_i(\alpha)\right)\right] \\ &= \arg\max_{\alpha} \frac{1}{N}\sum_{i=1}^{N}\left[-\frac{1}{2}\ln\hat{\sigma}_i^2(\alpha) - \frac{1}{2T}\sum_{t=1}^{T}\frac{(y_{it}-\alpha y_{it-1}-\hat{\eta}_i(\alpha))^2}{\hat{\sigma}_i^2(\alpha)}\right],\end{aligned}$$

where

$$\widehat{\Theta}_i(\alpha) = \begin{pmatrix}\hat{\eta}_i(\alpha) \\ \hat{\sigma}_i^2(\alpha)\end{pmatrix} = \begin{pmatrix}\bar{y}_i - \alpha\bar{y}_{i(-1)} \\ \frac{1}{T}\sum_{t=1}^{T}(y_{it}-\alpha y_{it-1}-(\bar{y}_i-\alpha\bar{x}_i))^2\end{pmatrix},$$

and $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{y}_{i(-1)} = \frac{1}{T}\sum_{t=1}^{T} y_{it-1}$, $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{y}_{it-1} = y_{it-1} - \bar{y}_{i(-1)}$.

Again, I consider several bias-correcting estimators of α that are obtained by maximization of a modified concentrated log likelihood like

$$\widetilde{\alpha} \equiv \arg\max_{\alpha}\frac{1}{N}\sum_{i=1}^{N}\ell_{mi}\left(\alpha,\widehat{\Theta}_i(\alpha)\right).$$

- Trace Based Approach with Trimming: this approach uses a trimmed version of $\Upsilon_i(\alpha)$, that is,

$$\widehat{\Upsilon}_i(\alpha) = \Omega_0 + \sum_{l=1}^{r}(\Omega_l + \Omega_l'),$$

where

$$\Omega_l = \frac{1}{T-l}\sum_{t=l+1}^{T}\left(1-\frac{l}{r+1}\right)\frac{\partial\ell_{it}}{\partial\Theta_i}\cdot\frac{\partial\ell_{it-l}}{\partial\Theta_i'},$$

for r small. So,

$$\ell_{mi}\left(\alpha,\widehat{\Theta}_i(\alpha)\right) = \frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\alpha,\widehat{\Theta}_i(\alpha)\right) - \frac{1}{2T}tr\left(\widehat{H}_i^{-1}(\alpha)\,\widehat{\Upsilon}_i(\alpha)\right).$$

- Determinant Based Approach Using Expected Quantities: now

$$\begin{aligned}H_i(\alpha) &= -\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\frac{\partial^2\ell_{it}}{\partial\eta^2} & \frac{\partial^2\ell_{it}}{\partial\eta\partial\sigma^2} \\ \frac{\partial^2\ell_{it}}{\partial\sigma^2\partial\eta} & \frac{\partial^2\ell_{it}}{\partial(\sigma^2)^2}\end{pmatrix} \\ &= \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\frac{1}{\sigma_i^2} & \frac{(y_{it}-\alpha y_{it-1}-\eta_i)}{\sigma_i^4} \\ \frac{(y_{it}-\alpha y_{it-1}-\eta_i)}{\sigma_i^4} & \left(\frac{(y_{it}-\alpha y_{it-1}-\eta_i)^2}{\sigma_i^6}\right) - \frac{1}{2\sigma_i^4}\end{pmatrix},\end{aligned}$$

and

$$\begin{aligned}&\bar{\Upsilon}_i(\alpha,\Theta_i;\alpha_0,\Theta_{i0}) \\ &= TE_0\left\{\left[\frac{\partial\ell_i(\alpha,\Theta_i)}{\partial\Theta_i} - E\left(\frac{\partial\ell_i(\alpha,\Theta_i)}{\partial\Theta_i}\right)\right]\left[\frac{\partial\ell_i(\alpha,\Theta_i)}{\partial\Theta_i'} - E\left(\frac{\partial\ell_i(\alpha,\Theta_i)}{\partial\Theta_i'}\right)\right]\middle| y_{i0}\right\}.\end{aligned}$$

Now, I obtain $\bar{\Upsilon}_i\left(\alpha,\widehat{\Theta}_i(\alpha);\hat{\alpha},\widehat{\Theta}_i\right)$ as a mean of $\{\Upsilon_i^m(\alpha)\}_{m=1}^M$ calculated in data simulated as $\left\{\prod_{t=1}^T f\left(y_{it}|y_{i0},\hat{\alpha},\widehat{\Theta}_i\right)\right\}_{i=1}^N$. That is,

$$\bar{\Upsilon}_i\left(\alpha,\widehat{\Theta}_i(\alpha);\hat{\alpha},\widehat{\Theta}_i\right)=\frac{1}{M}\sum_{m=1}^M\Upsilon_i^m(\alpha),$$

where

$$\Upsilon_i^m(\alpha)=\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left\{\left[\frac{\partial\ell_{it}}{\partial\Theta_i}-\left(\frac{1}{T}\sum_{r=1}^T\frac{\partial\ell_{ir}}{\partial\Theta_i}\right)\right]\cdot\left[\frac{\partial\ell_{is}}{\partial\Theta_i'}-\left(\frac{1}{T}\sum_{r=1}^T\frac{\partial\ell_{ir}}{\partial\Theta_i'}\right)\right]\right\},$$

and

$$\frac{\partial\ell_{it}}{\partial\Theta_i}=\begin{pmatrix}\frac{\partial\ell_{it}}{\partial\eta}\\ \frac{\partial\ell_{it}}{\partial\sigma^2}\end{pmatrix}=\begin{pmatrix}\frac{(y_{it}-\alpha y_{it-1}-\hat{\eta}_i(\alpha))}{\hat{\sigma}_i^2(\alpha)}\\ \frac{(y_{it}-\alpha y_{it-1}-\hat{\eta}_i(\alpha))^2-\hat{\sigma}_i^2(\alpha)}{2\left(\hat{\sigma}_i^2(\alpha)\right)^2}\end{pmatrix}.$$

This leads to

$$\begin{aligned}\ell_{mi}\left(\alpha,\widehat{\Theta}_i(\alpha);\hat{\alpha},\widehat{\Theta}\right) &= \frac{1}{T}\sum_{t=1}^T\ell_{it}\left(\alpha,\widehat{\Theta}_i(\alpha)\right)+\frac{1}{2T}\ln\det\widehat{H}_i(\alpha)\\ &\quad -\frac{1}{2T}\ln\det\bar{\Upsilon}_i\left(\alpha,\widehat{\Theta}_i(\alpha);\hat{\alpha},\widehat{\Theta}_i\right).\end{aligned}$$

- Determinant Based Approach Using a Bootstrap Estimate of $Var\left[\widehat{\Theta}_i(\alpha)\right]$: this approach is based on using the bootstrap estimate

$$\widehat{Var}\left[\widehat{\Theta}_i(\alpha)\right]=\frac{1}{M}\sum_{m=1}^M\left[\widehat{\Theta}_i^m(\alpha)-\widehat{\Theta}_i(\alpha)\right]\left[\widehat{\Theta}_i^m(\alpha)-\widehat{\Theta}_i(\alpha)\right]',$$

which leads to

$$\ell_{mi}\left(\alpha,\widehat{\Theta}_i(\alpha)\right)=\frac{1}{T}\sum_{t=1}^T\ell_{it}\left(\alpha,\widehat{\Theta}_i(\alpha)\right)-\frac{1}{2}\ln\det\left(\widehat{H}_i(\alpha)\widehat{Var}\left[\widehat{\Theta}_i(\alpha)\right]\right).$$

Table 3 reports estimates for $T=8$ and 16, and $N=500$.

Table 3. Properties of $\hat{\alpha}$ for $\alpha=0.5$

	$T=8$		$T=16$	
Estimator of α	Mean	SD	Mean	SD
MLE	0.2575	0.0169	0.3904	0.0113
Expected Quantities (1st)	0.4214	0.0225	0.4862	0.0160
Expected Quantities (2nd)	0.5115	0.0243	0.5119	0.0157
Bootstrap Variance (1st)	0.3753	0.0442	0.4707	0.0167
Bootstrap Variance (2nd)	0.4336	0.0515	0.4925	0.0172
Trimming	0.3105	0.0467	0.4444	0.0121

Note: N=500; simulations=300; parametric bootstrap samples=300; trimming=2. SD: Sample standard deviation.

We see in the table that the fixed effects MLE is downward biased in both cases. Here we can see that iterating bias correction improves substantially the estimation. In fact, bias corrections reduce the bias for at least a half and this bias practically disappears when I iterate the corrections.

4.3 The AR(1)-EARCH(1) panel model with fixed effects

Now the model design is

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + e_{it} = \alpha y_{it-1} + h_{it}^{1/2}\epsilon_{it}, \;\; (t=1,...,T; \; i=1,...,N) \\ h_{it} &= \exp\left(\psi_i + \beta\left[\sqrt{\epsilon_{it-1}^2 + \Lambda} - \sqrt{2/\pi}\right]\right) = h\left(\epsilon_{it-1}, \psi_i\right), \\ \epsilon_{it} &\sim N(0,1), \;\; \psi_i \sim N\left(-3.0, 0.8\right). \end{aligned}$$

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 $E\left[\sqrt{\epsilon_{it-1}^2 + \Lambda}\right]$ given that $\epsilon_{it-1} \sim N(0,1)$. The process is started at $y_{i0} = 0$, then 700 time periods are generated before the sample is generated. I denote as $\Gamma = (\alpha, \beta)$. The data are generated for $T = 8$ and 16, $N = 1000$, $\alpha = 0.5$, and $\beta = 0.5$. For each sample I have estimated Γ by maximum likelihood and, at the moment, by the trimming modified maximum likelihood.

The MLE of Γ is

$$\widehat{\Gamma} \equiv \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\Gamma, \widehat{\psi}_i\left(\Gamma\right)\right)\right],$$

where

$$\widehat{\psi}_i\left(\Gamma\right) \equiv \arg\max_{\psi} \frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\Gamma, \psi\right).$$

Since here I can not get a explicit expression of the fixed effects estimators as functions of α and β, I do a double maximization, strictly speaking N maximizations inside the one for Γ. I use a Quasi-Newton's Method algorithm to maximize the log likelihood function with respect to Γ, and in each step $\widehat{\psi}_i\left(\Gamma\right)$ is computed such that, for this given value of Γ, the individual log likelihood is maximized with respect to ψ.

The MMLE is

$$\begin{aligned} \widetilde{\Gamma} &= \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N}\ell_{mi}\left(\Gamma, \widehat{\psi}_i\left(\Gamma\right)\right) \\ &= \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\Gamma, \widehat{\psi}_i\left(\Gamma\right)\right) - \frac{\hat{b}_i\left(\Gamma\right)}{T}\right], \end{aligned}$$

where

$$\hat{b}_i\left(\Gamma\right) = \frac{1}{2}\left[\widehat{H}_i^{-1}\left(\Gamma\right)\widehat{\Upsilon}_i\left(\Gamma\right)\right],$$

for

$$\widehat{H}_i(\Gamma) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_{it}}{\partial \psi^2},$$

and a trimmed version of $\Upsilon_i(\Gamma)$ with r small

$$\widehat{\Upsilon}_i(\Gamma) = \Omega_0 + 2\sum_{l=1}^{r}\Omega_l,$$

$$\Omega_l = \frac{1}{T-l}\sum_{t=l+1}^{T}\left(1-\frac{l}{r+1}\right)\frac{\partial \ell_{it}}{\partial \psi_i}\cdot\frac{\partial \ell_{it-l}}{\partial \psi_i}.$$

In this case I calculate numerical first and second derivatives.

Table 4 reports estimates for $T = 8$ and 16, and $N = 1000$. In this case $\hat{\alpha}$ is not biased, and with the trimming correction I correct an otherwise seriously biased MLE of β.

Table 4. Properties of $\hat{\alpha}, \hat{\beta}$ for $\alpha = 0.5, \beta = 0.5$

Estimator of	$T = 8$				$T = 16$			
$(\alpha, \beta)'$	Mean $\hat{\alpha}$	SD $\hat{\alpha}$	Mean $\hat{\beta}$	SD $\hat{\beta}$	Mean $\hat{\alpha}$	SD $\hat{\alpha}$	Mean $\hat{\beta}$	SD $\hat{\beta}$
MLE	0.4994	0.0126	−0.1022	0.0845	0.5009	0.0069	0.3670	0.0284
Trimming	0.5012	0.0136	−0.0252	0.0973	0.5009	0.0070	0.4596	0.0284

Note: N=1000; simulations=100; trimming=2. SD: Sample standard deviation. T=8: trimming: 95% successful convergence. T=16: trimming: 100% successful convergence.

4.4 The AR(1)-EARCH(1) panel model with multiple fixed effects

Here the model design is

$$\begin{aligned}
y_{it} &= \alpha y_{it-1} + \eta_i + e_{it} = \alpha y_{it-1} + \eta_i + h_{it}^{1/2}\epsilon_{it}, \ \ (t = 1, ..., T;\ i = 1, ..., N) \\
h_{it} &= \exp\left(\psi_i + \beta\left[\sqrt{\epsilon_{it-1}^2 + \Lambda} - \sqrt{2/\pi}\right]\right) = h\left(\epsilon_{it-1}, \psi_i\right), \\
\epsilon_{it} &\sim N(0,1);\ \eta_i \sim N(0,1);\ \psi_i \sim N(-3.0, 0.8).
\end{aligned}$$

The process is started at $y_{i0} = 0$, then 700 time periods are generated before the sample is generated. I denote as $\Gamma = (\alpha, \beta)$. The data are generated for $T = 16$, $N = 1000$, $\alpha = 0.5$, and $\beta = 0$, and 0.5. For each sample I have estimated Γ by maximum likelihood and, at the moment, by the trimming modified maximum likelihood.

The MLE of Γ is

$$\widehat{\Gamma} \equiv \arg\max_{\Gamma}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{it}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right)\right],$$

where

$$\widehat{\Theta}_i(\Gamma) \equiv \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T} \ell_{it}(\Gamma, \Theta),$$

and the MMLE is

$$\begin{aligned}
\widetilde{\Gamma} &= \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N} \ell_{mi}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right) \\
&= \arg\max_{\Gamma} \frac{1}{N}\sum_{i=1}^{N} \left[\frac{1}{T}\sum_{t=1}^{T} \ell_{it}\left(\Gamma, \widehat{\Theta}_i(\Gamma)\right) - \frac{\hat{b}_i(\Gamma)}{T}\right],
\end{aligned}$$

where

$$\hat{b}_i(\Gamma) = \frac{1}{2} tr\left[\widehat{H}_i^{-1}(\Gamma)\, \widehat{\Upsilon}_i(\Gamma)\right],$$

for

$$\widehat{H}_i(\Gamma) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \ell_{it}}{\partial\Theta\partial\Theta'},$$

and a trimmed version of $\Upsilon_i(\Gamma, \Theta)$

$$\widehat{\Upsilon}_i(\Gamma, \Theta) = \Omega_0 + \sum_{l=1}^{r} (\Omega_l + \Omega_l'),$$

with

$$\Omega_l = \frac{1}{T-l}\sum_{t=l+1}^{T} \left(1 - \frac{l}{r+1}\right) \frac{\partial \ell_{it}}{\partial\Theta} \cdot \frac{\partial \ell_{it-l}}{\partial\Theta'}.$$

Also in this case I calculate numerical first, second and cross derivatives.

Table 5 reports estimates for $T = 16$ and $N = 1000$. Again, I obtain estimates with less bias when I use the modified maximum likelihood estimator.

Table 5. Properties of $\hat{\alpha}, \hat{\beta}$ for $\alpha = 0.5$ ($T = 16$)

Estimator of $(\alpha, \beta)'$	Mean $\hat{\alpha}$	SD $\hat{\alpha}$	Mean $\hat{\beta}$	SD $\hat{\beta}$
		$\beta = 0.5$		
MLE	0.3958	0.0092	0.4308	0.0388
Trimming	0.4308	0.0388	0.4819	0.0643
		$\beta = 0.0$		
MLE	0.3823	0.0175	-0.0465	0.0077
Trimming	0.4426	0.0210	-0.0286	0.0477

Note: N=1000; simulations=20; trimming=2; trimming $\beta = 0.5$: 85% successful convergence; trimming $\beta = 0.0$: 70% successful convergence. SD: Sample standard deviation.

5 Estimation Results

In this section I use the modified maximum likelihood method to estimate an empirical model for the conditional mean and the conditional variance of male wages. 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 C. Sample composition by year and by education, and demographic characteristics are presented in Appendix D.1.

The dependent variable is annual real wages of the heads, so I exclude other components of money income for labour as labour part of farm income, business income, overtime, commissions, etc. Figures 1 and 2, at the end of the document, plot the mean and the 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 log wages residuals from first stage regressions on year dummies, education, a quadratic in age, dummies for race (white), region of residence, and residence in a SMSA[10]. In this version of the model, I deal with aggregate effects in the variance by regarding y_{it} as standardized wages[11].

The equation estimated is

$$y_{it} = \alpha y_{it-1} + \eta_i + e_{it} = \alpha y_{it-1} + \eta_i + \sqrt{h_{it}}\epsilon_{it},\ (i = 1, ..., N; t = 0, ..., T)$$

with

$$h_{it} = \exp\left(\psi_i + \beta\left[\sqrt{\epsilon^2_{it-1} + \Lambda} - \sqrt{2/\pi}\right]\right) = h\left(\epsilon_{it-1}, \psi_i\right).$$

Table 6 presents the estimation results by MLE and by maximization of the trimmed corrected concentrated likelihood.

[10]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.

[11]For each year I calculate the sample wage variance and I take $\left(\log w_{it} - \hat{\mu}_t\right)/\hat{\sigma}_t$.

Table 6. α **and** β **estimates**

Estimator of $(\alpha, \beta)'$	$\widehat{\alpha}$	$\widehat{\beta}$
MLE	0.4822	0.4832
	(0.0114)	(0.0541)
Trimming ($r = 2$)	0.5690	0.5790
	(0.0397)	(0.0915)

Note: Mean of estimated standard errors by individual block-bootstrap in brackets.

As expected, I obtain that the maximum likelihood estimate is below the trimming estimate. In fact, after applying the bias correction, I obtain estimates of both parameters above 0.5. Not only the persistence in the mean is significant. Also the state dependence effects in the volatility of wages seem important.

5.1.1 Correlations between unobserved individual heterogeneity and observed outcomes

One important advantage of the *fixed effects perspective* adopted here is that I also obtain estimates of the unobserved individual heterogeneity and, therefore, I can evaluate the relation between those individual effects in the volatilities of wages and measurable outcomes.

Table 7 shows 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, moving 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 we could expect.

Table 7. Correlations with observed variables

Dependent variable: $\hat{\psi}_i$	[1]	[2]	[3]	[4]	[5]
Constant	−0.6933	−0.7242	−1.6997	−1.2663	−1.0868
	(0.1784)	(0.1834)	(0.1935)	(0.2142)	(0.2211)
Married	−0.4657	−0.4168	−0.4415	−0.3634	−0.3476
	(0.0683)	(0.0673)	(0.0649)	(0.0640)	(0.0632)
Age	−0.0138	−0.0146	−0.0054	−0.0052	−0.0042
	(0.0038)	(0.0037)	(0.0037)	(0.0040)	(0.0039)
White	−0.5984	−0.4237	−0.4409	−0.5229	−0.4337
	(0.0632)	(0.0651)	(0.0631)	(0.0609)	(0.0617)
Technical Workers		−0.4394	−0.4905		−0.4467
		(0.0912)	(0.0881)		(0.0974)
Administrators		−0.4222	−0.4751		−0.4743
		(0.0943)	(0.0911)		(0.0932)
Sales workers		0.2137	0.2325		0.1712
		(0.1076)	(0.1038)		(0.1015)
Services workers		0.3212	0.2837		0.1761
		(0.0983)	(0.0949)		(0.0921)
Operatives workers		0.0812	0.0919		0.0357
		(0.0886)	(0.0854)		(0.0828)
Movers			0.8177	0.5579	0.5734
			(0.0658)	(0.0686)	(0.0678)
Dropout				0.5319	0.1707
				(0.0872)	(0.1019)
Graduate				0.2260	−0.0511
				(0.0687)	(0.0790)
Tenure: 1-2 years				0.0132	−0.0009
				(0.1493)	(0.1474)
Tenure: 2-3 years				−0.1695	−0.1364
				(0.1175)	(0.1163)
Tenure: 4-9 years				−0.4308	−0.3991
				(0.0978)	(0.0969)
Tenure: 9-19 years				−0.8309	−0.7984
				(0.0929)	(0.0919)
Tenure: 20 years or more				−0.9530	−0.9001
				(0.1008)	(0.1002)
R^2	0.0822	0.1192	0.1808	0.2155	0.2389

Note: Number of observations=2066 individuals. Standard errors in brackets. Omitted group: Craftsman workers, Stayers, Education College, Tenure less than a year.

The $\hat{\psi}_i' s$ capture the unobserved heterogeneity in a very robust way. If we were able to observe the individual heterogeneity this would be much better but, if we look at the R^2 of the regression, we can see that with only the observed covariates we can not explain much of the variation across individuals.

5.1.2 Generality of the estimation method

I have also estimated a version of the model similar to Meghir and Windmeijer (1999). It is a convenient specification but 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 model is the following

$$y_{it} = \alpha y_{it-1} + \eta_i + e_{it} = \alpha y_{it-1} + \eta_i + \sqrt{g_{it}}\epsilon_{it};\ (i = 1, ..., N; t = 1, ..., T)$$

with

$$g_{it} = \exp\left(\psi_i + \beta\left[\sqrt{y_{it-1}^2 + \Lambda}\right]\right) = g\left(y_{it-1}, \psi_i\right).$$

Table 8 presents the corresponding results of the estimation of this model by MLE and by maximization of the trimmed corrected concentrated likelihood. Although the estimates of β are a bit different, the main results do not change.

Table 8. α **and** β **estimates**

Estimator of $(\alpha, \beta)'$	$\widehat{\alpha}$	$\widehat{\beta}$
MLE	0.4904	0.3713
	(0.0099)	(0.0313)
Trimming ($r = 2$)	0.5432	0.4145
	(0.0095)	(0.0337)

Note: Mean of estimated standard errors by individual block-bootstrap in brackets.

5.2 Checking for Nonnormality

The assumption of normality is not necessary for the validity of the estimation method used on the empirical application, but checking this distributional assumption can be useful for other purposes. The distribution of the errors are nonparametrically identified and can be estimated using deconvolution tecniques as in Horowitz and Markatou (1996). A normal probability plot of residuals in first-differences (Figure 3) indicates that the tails of the distribution of errors are thicker than those of the normal distribution. However the same plot but for the standardized residuals in first-differences (Figure 4) is almost a straight line, meaning no deviation from normality[12].

[12] Estimated residuals and estimated standardized residuals respectively defined as

$$\hat{e}_{it} = y_{it} - \hat{\alpha} y_{it-1} - \hat{\eta}_i.$$

and

$$\hat{\epsilon}_{it} = \frac{y_{it} - \hat{\alpha} y_{it-1} - \hat{\eta}_i}{h_{it}^{1/2}\left(\hat{\psi}_i, \hat{\epsilon}_{it-1}\right)},$$

where

$$h_{it}\left(\hat{\psi}_i, \hat{\epsilon}_{it-1}\right) = \exp\left\{\hat{\psi}_i + \hat{\beta}\left[\left|\hat{\epsilon}_{it-1}\right| - \sqrt{2/\pi}\right]\right\}.$$

5.2.1 Fit of the model

Given the distributional assumption, parameter estimates, $\hat{\alpha}_T$, $\hat{\beta}_T$, $\hat{\eta}_i$, $\hat{\psi}_i$, and initial conditions, y_{i0}, $\widehat{h_{i1}}$, I simulate an unbalanced panel of standardized logwages observations with the same dimensions as the PSID sample. The first thing that I evaluate with this simulated panel is the fit of the model. Figure 5 shows the kernel densities of logwages in the data and according to the model[13]. It seems that the model performs well.

5.2.2 Individual Heterogeneity

Then, for evaluating the existence of individual heterogeneity on the data, I calculate several counterfactuals in an analogous way. Counterfactual 1 is obtained using the model, the parameter estimates, $\hat{\alpha}_T, \hat{\beta}_T, \hat{\psi}_i$, and initial conditions, $y_{i0}, \widehat{h_{i1}}$, but now $\eta_i = \bar{\eta}, \forall i$, where $\bar{\eta} = N^{-1} \sum_{i=1}^{N} \hat{\eta}_i$. Similarly, counterfactual 2 is obtained using the model, the parameter estimates, $\hat{\alpha}_T, \hat{\beta}_T, \hat{\eta}_i$, and initial conditions, $y_{i0}, \widehat{h_{i1}}$, but now $\psi_i = \bar{\psi}, \forall i$, where $\bar{\psi} = N^{-1} \sum_{i=1}^{N} \hat{\psi}_i$. When I plot the individual means and individual logvariances of logwages (Figures 6 and 7, and Table 9 for some descriptive statistics of those distributions) we can see that there is variation across individuals not only in the means but also in the variances. In addition we can see that the model captures this variation successfully.

Table 9. Descriptive Statistics

Individual means	Data	Model	Counterfactual 1
Mean	-0.0180	-0.0059	-0.0086
Standard deviation	0.7848	0.8404	0.3876
Individual logvariances	**Data**	**Model**	**Counterfactual 2**
Mean	-1.5980	-1.7054	-1.6703
Standard deviation	1.2762	1.4346	0.7890

Using these counterfactuals I 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 according to the model. In particular, for the counterfactual 2, the sample variance of logwages is equal to 0.8581. That is, variation in $\hat{\psi}_i$ accounts for by 14 per cent of the total variation in log wages.

5.2.3 Dynamics: Quantiles of log normal wages

Regarding the dynamics, with a model like the one considered in this paper I can say how is the effect of lagged values at different parts of the wage distribution. In a general setting, let logwages $y = \log(w) \sim$

[13] The bandwidth is equal to 0.10 for all kernels in this section.

$N(\mu, \sigma^2)$ with *cdf*

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

The τth quantile of $\log w$, $Q_\tau(\log w)$, 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 quantileof the $N(0,1)$ distribution, and

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

To get quantiles for w, as opposed to $\log w$, note that

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

so that

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

Therefore,

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

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

$$\frac{\partial \log Q_\tau(w_{it})}{\partial \log w_{it-1}} = \frac{\partial \mu}{\partial \log w_{it-1}} + q_\tau \frac{\partial \sigma}{\partial \log w_{it-1}}.$$

In particular, for the model considered here

$$\mu_{it} = \alpha y_{it-1} + \eta_i,$$

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

and

$$\frac{\partial \mu_{it}}{\partial y_{it-1}} = \alpha,$$

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

Thus I can calculate a mean elasticity at different parts of the wage distribution as

$$
\varepsilon_\tau \left(\log w_{it-1}\right) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\frac{\partial \log Q_\tau \left(w_{it}\right)}{\partial \log w_{it-1}} \right].
$$

The first column in Table 10 shows that those elasticities increase with the quantiles. That is, there are different elasticities below and above the median, where the mean elasticity is just equal to the corrected estimate of alpha, $\hat{\alpha}_T$. In Table 10, and graphically in Figure 8, we can see that this pattern is very different for individuals with low (second column) or high (third column) values of the estimated fixed effects in the variance.

Table 10. Mean elasticities with respect to y_{it-1} at different quantiles

τ	All individuals	$\hat{\psi}_i$ under the mean	$\hat{\psi}_i$ above the mean
0.10	0.5595	0.5752	0.5421
0.20	0.5628	0.5731	0.5513
0.30	0.5651	0.5715	0.5580
0.40	0.5671	0.5702	0.5637
0.50	0.5690	0.5690	0.5690
0.60	0.5709	0.5678	0.5743
0.70	0.5729	0.5665	0.5800
0.80	0.5752	0.5649	0.5867
0.90	0.5785	0.5628	0.5959

Impulse-response function: functions of ϵ_{it-s} Now,

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

and in the conditional case, regarding μ and σ as functions of ϵ_{it-1},

$$
\frac{\partial Q_\tau \left(\log w_{it}\right)}{\partial \epsilon_{it-1}} = \frac{\partial \mu}{\partial \epsilon_{it-1}} + q_\tau \frac{\partial \sigma}{\partial \epsilon_{it-1}}.
$$

In particular, for the model considered here

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

and

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

Thus I can calculate a mean marginal effect at different parts of the logwage distribution as

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

Notice that

$$
\frac{\partial Q_\tau\left(\log w_{it}\right)}{\partial \epsilon_{it-1}}=\frac{\partial \log Q_\tau\left(w_{it}\right)}{\partial \log w_{it-1}} \times\left[h_{it-1}\left(\psi_i, \epsilon_{it-2}\right)^{1/2}\right].
$$

Now, for

$$
\frac{\partial Q_\tau\left(\log w_{it}\right)}{\partial \epsilon_{it-2}}=\frac{\partial \mu}{\partial \epsilon_{it-2}}+q_\tau \frac{\partial \sigma}{\partial \epsilon_{it-2}}.
$$

In particular, for the model considered here and

$$
\begin{aligned}
\mu_{it} &= \alpha^2\left(\alpha y_{it-3}+\eta_i+h_{it-2}\left(\psi_i, \epsilon_{it-3}\right)^{1/2} \epsilon_{it-2}\right)+(1+\alpha)\eta_i+\alpha h_{it-1}\left(\psi_i, \epsilon_{it-2}\right)^{1/2} \epsilon_{it-1} \\
&= \alpha^3 y_{it-3}+\left(1+\alpha+\alpha^2\right)\eta_i+\alpha h_{it-1}\left(\psi_i, \epsilon_{it-2}\right)^{1/2} \epsilon_{it-1}+\alpha^2 h_{it-2}\left(\psi_i, \epsilon_{it-3}\right)^{1/2} \epsilon_{it-2}, \\
\sigma_{it} &= \exp\left(\frac{\psi_i}{2}+\frac{\beta}{2}\left[\sqrt{\epsilon_{it-1}^2+\Lambda}-\sqrt{2/\pi}\right]\right) \\
&= \exp\left(\frac{\psi_i}{2}+\frac{\beta}{2}\left[\sqrt{\left(\frac{y_{it-1}-\alpha y_{it-2}-\eta_i}{h_{it-1}\left(\psi_i, \epsilon_{it-2}\right)^{1/2}}\right)^2+\Lambda}-\sqrt{2/\pi}\right]\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \mu_{it}}{\partial \epsilon_{it-2}} &= \alpha^2 h_{it-2}\left(\psi_i, \epsilon_{it-3}\right)^{1/2}+\frac{1}{2}\alpha\beta h_{it-1}\left(\psi_i, \epsilon_{it-2}\right)^{1/2} \frac{\epsilon_{it-1}\epsilon_{it-2}}{\sqrt{\epsilon_{it-2}^2+\Lambda}}, \\
\frac{\partial \sigma_{it}}{\partial \epsilon_{it-2}} &= \sigma_{it}\frac{\beta}{2}\left(\frac{\epsilon_{it-1}}{\left[\epsilon_{it-1}^2+\Lambda\right]^{1/2}}\right) \times \sigma_{it-1}\frac{\beta}{2}\left(\frac{\epsilon_{it-2}}{\left[\epsilon_{it-2}^2+\Lambda\right]^{1/2}}\right).
\end{aligned}
$$

The first panel in Table 11 shows the mean marginal effects with respect to ϵ_{it-1} over different quantiles of the logwage distribution and the second panel, the case with respect to ϵ_{it-2}. In Figure 9 we can see that past shocks seem to have effect over logwages even two periods apart.

Table 11. Mean marginal effects with respect to past shocks at different quantiles

τ	All individuals	$\hat{\psi}_i$ under the mean	$\hat{\psi}_i$ above the mean
With respect to ϵ_{it-1}			
0.10	0.2543	0.1349	0.3866
0.20	0.2589	0.1346	0.3966
0.30	0.2622	0.1344	0.4039
0.40	0.2650	0.1342	0.4101
0.50	0.2677	0.1340	0.4158
0.60	0.2703	0.1338	0.4216
0.70	0.2732	0.1336	0.4278
0.80	0.2765	0.1334	0.4351
0.90	0.2811	0.1331	0.4451
With respect to ϵ_{it-2}			
0.10	0.1455	0.0726	0.2262
0.20	0.1452	0.0725	0.2258
0.30	0.1451	0.0725	0.2255
0.40	0.1449	0.0724	0.2253
0.50	0.1448	0.0724	0.2251
0.60	0.1447	0.0723	0.2248
0.70	0.1445	0.0723	0.2246
0.80	0.1444	0.0722	0.2243
0.90	0.1441	0.0721	0.2239

5.2.4 Dynamics: Job changes

It is important taking into account that in a model where individual heterogeneity is treated as fixed effects we abstract for job changes. A specification like this

$$y_{it} = \alpha y_{it-1} + \eta_i + e_{it},$$

works worse if there are many job changes in the sample because η_i is fixed. In order to evaluate this concern, I consider a sample where individuals in different jobs are treated as different individuals. That is, for each individual

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \eta_{i1} + e_{it}; \text{ individual } i \text{ in job 1}, \\ y_{it} &= \alpha y_{it-1} + \eta_{i2} + e_{it}; \text{ individual } i \text{ in job 2}. \end{aligned}$$

I use data on 1,346 and 17,485 observations. I do the same sample selection as before. Sample composition by year and by education, and demographic characteristics are presented in Appendix D.2.

Results are reported in Table 12. We can see that the significant ARCH effects in the variance disappears as soon as we consider a sample without job changes.

Table 12. α **and** β **estimates**

Estimator of $(\alpha, \beta)'$	$\widehat{\alpha}$	$\widehat{\beta}$
MLE	0.3768	0.0642
	(0.0158)	(0.0846)
Trimming ($r = 2$)	0.4569	0.0757
	(0.0361)	(0.0592)

Note: Mean of estimated standard errors by individual block-bootstrap in brackets.

5.3 Attrition

A final issue is the extent to which attrition from the PSID has biased 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 the estimates in my sample 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. The estimates based on this sample are included in Table 13. The main conclusion is that the corrected estimates are not very different to those reported in Table 6.

Table 13. α **and** β **estimates**

Estimator of $(\alpha, \beta)'$	$\widehat{\alpha}$	$\widehat{\beta}$
MLE	0.5659	0.5245
	(0.0114)	(0.0412)
Trimming ($r = 2$)	0.6056	0.5693
	(0.0347)	(0.0717)

Note: Mean of estimated standard errors by individual block-bootstrap in brackets.

6 Implications for Consumption Growth

Given the results above I provide now an example that illustrates the effects that individual risk can have in explaining precautionary saving, that is, additional saving that results from the knowledge that the future is uncertain. Here, I follow most of the literature and I consider that additional saving is achieved by consuming less.

Over the last 30 years there has been a well-documented increase in cross-sectional income inequality in the US, and some authors have suggested that households are now exposed to more earnings instability than they were (Gottschalk and Moffitt, 1994). This figure suggests that precautionary saving motives associated with an increase in income risk could have become more important.

In the presence of complete insurance, either formal or informal, it should only be the component of risk that is common to all individuals in an economy that affects consumption. Banks, Blundell, and Brugiavini (2001) find that it is not the common component of risk, but instead the cohort-specific risks which dominate consumption growth. Their results corroborate the notion that if income uncertainty has been growing over the recent past then the failure of insurance between agents makes the precautionary motive for saving an increasingly important self-insurance mechanism. They use series of repeated cross sections of British households data, but they can not consider individual-specific risk due to the lack of panel data. Here, I evaluate the independent role of individual wage risk in consumption growth.

6.1 Consumption Model

Let us consider the following intertemporal consumption model[14] (Browning and Lusardi, 1996), where individuals choose consumption so as to maximize an intertemporal utility function subject to the intertemporal budget constraint:

$$\max_{\{C_{t+k}\}_{k=0}^{T-t}} E_t \sum_{k=0}^{T-t} \left[(1+\delta)^{-k} U(C_{t+k}, D_{t+k}) \right]$$

$$s.t. \quad A_{t+1+k} = (1+r_{t+k}) \cdot (A_{t+k} + Y_{t+k} - C_{t+k})$$

$$A_{T+1} \geq 0 \qquad\qquad (k = 0, ..., T-t)$$

where, for each period s, C_s is consumption, Y_s labour income or earnings, r_s real interest rate, A_s financial wealth (at the beginning of the period), δ subjective intertemporal rate, and D_s demographic characteristics. I assume the date of death is known and there are not explicit liquidity constraints.

The optimal intertemporal allocation of consumption verifies the Euler equation, that is,

$$E_t \left[\frac{1+r_t}{1+\delta} \cdot \frac{U_C(C_{t+1}, D_{t+1})}{U_C(C_t, D_t)} \right] = 1$$

where $U_c(\cdot)$ denotes the first derivative of the utility function with respect to consumption.

I assume a CRRA utility function:

$$U(C_t, D_t) = \frac{1}{1-\rho} \exp(\varphi' D_t) \cdot C_t^{1-\rho}$$

where $\rho > 0$ is the relative risk aversion coefficient. So,

$$\frac{1+r_t}{1+\delta} \cdot \exp(\varphi' \Delta D_{t+1}) \cdot \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} = 1 + \xi_{t+1},$$

[14] I omit the individual index for simplicity.

where $E_t\left[\xi_{t+1}\right]=0$. Taking logs and using the usual approximation for logs I obtain the linearized Euler equation:

$$\Delta \ln C_{t+1}=\frac{1}{\rho}\ln\left(r_t-\delta\right)+\frac{1}{\rho}\varphi'\Delta D_{t+1}+\frac{1}{2\rho}Var_t\left[\xi_{t+1}\right]+v_{t+1}.$$

The first term on the RHS of the equation takes into account the intertemporal substitution effect: an increase in r_t, opportunity cost of current consumption, implies a higher growth of future consumption. The second term considers how different stages of the life cycle are reflected on the consumption profile, by changes in circumstances implicit in demographic variables. Finally, the third term on the RHS of the equation captures precautionary saving. A rise in the expected variance of earnings innovations represents an increase in earnings risk and should depress period t consumption hence increasing the growth of consumption between t and $t+1$. In other words, a positive parameter implies that risk induces a delay in spending and current consumption is therefore reduced.

Notice that $Var_t\left[\xi_{t+1}\right]$ reflects uncertainty regarding future realizations of any uninsurable variable relevant for consumption. Thus, it is not sufficient to enter the wage risk term alone. A scaling term is required by which "poorer" individuals are more responsive to changes in earnings risk, $\pi_t=\left(\frac{Y_{t-1}}{C_{t-1}}\right)^2$. In consequence,

$$\Delta \ln C_{t+1}=\frac{1}{\rho}\ln\left(r_t-\delta\right)+\frac{1}{\rho}\varphi'\Delta D_{t+1}+\gamma\pi_t\sigma^2_{t+1}+v_{t+1}$$

where σ^2_{t+1} is a measure of the conditional variance of the wage shock.

6.2 Estimation and results

I use food consumption data from the PSID (1974-1987). In my sample[15], I estimate by OLS[16] the following empirical equation:

$$\Delta \ln C_{it+1}=\delta_t+\beta'\Delta D_{it+1}+\gamma\pi_{it}\sigma^2_{it+1}+v_{it+1},$$

where σ^2_{it+1} is replaced by

$$\hat{\sigma}^2_{it+1}=h_{it+1}\left(\hat{\epsilon}_{it};\hat{\Gamma},\hat{\Theta}_i,\text{initial conditions}\right).$$

[15] The sample includes 1,191 individuals and 15,192 observations.

[16] It would be interesting to follow the same approach as before considering a complete likelihood function:

$$L_{now}=L_{before}+\sum_{i,t}\left\{-\frac{1}{2}\ln\sigma^2_v-\frac{1}{2\sigma^2_v}\left[\Delta\ln\hat{C}_{t+1}-\gamma\pi_t\sigma^2_{t+1}\right]^2\right\}$$

where $\Delta\ln\hat{C}_{t+1}$ is obtained from first stage regressions of $\Delta\ln C_{t+1}$ on δ_t and ΔD_{t+1}.

Looking at the estimate for the γ parameter in Table 14, column 2, I obtain a significant and positive effect of this term on the consumption growth. As stated above, an increase in individual risk induces a reduction in current consumption and, therefore, an increase in the growth of consumption between t and $t+1$.

Table 14. Consumption Growth Equation

	[1]	[2]	[3]	[4] Total effect
Age	-0.0004	-0.0003	-0.0003	
	(0.0008)	(0.0008)	(0.0008)	
ΔChildren	0.1513	0.1527	0.1527	
	(0.0100)	(0.0101)	(.0101049)	
ΔAdults	0.1593	0.1594	0.1592	
	(0.0119)	(0.0119)	(0.0119)	Dropout
$\pi_{it}^2\sigma_{it+1}^2$		0.0801	0.1267	0.127 [0.009]
		(0.0308)	(0.0485)	Graduate
$\pi_{it}^2\sigma_{it+1}^2\times$			-0.0714	0.055 [0.082]
Graduate			(0.0597)	College
$\pi_{it}^2\sigma_{it+1}^2\times$			-0.1089	0.018 [0.905]
College			(0.1562)	
# Obs.		13,723		

Note: clustered standard errors in brackets. Time and cohort dummies included. t-ratios in squared brackets.

Regarding the interactions with education (columns 3 and 4), we can see that this positive effect is more important for the less educated people, slightly significant for the graduate and insignificant for the college educated. This result goes in line with the idea that there are more insurance possibilities for these latter.

7 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 previous literature, I find a corrected estimate for the autoregressive coefficient in the mean around 0.5 (Alvarez and Arellano, 2004), and positive ARCH effects for the variance (Meghir and Pistafferri, 2004). Job changes are driving this dynamics in the variance. I also find important fixed differences across individuals in the variance. In addition, it turns out that this located-scaled model explains the non-normality observed in logwage data. I then illustrate some implications that ARCH effects may have in the field of savings.

Finally there are three issues, at least, that require further research: measurement error in PSID wages, a more comprehensive model that include job changes, and the comparison with female workers in terms of wage profiles.

A First Order Bias of the Concentrated Likelihood at an arbitrary value of the common parameter Γ

Following Arellano and Hahn (2006a, 2006b), let us obtain the expression for the First Order Bias of the Concentrated Likelihood at an arbitrary value of the common parameter Γ. Let $\ell_i(\Gamma,\Theta_i) = \sum_{t=1}^{T} \ell_{it}(\Gamma,\Theta_i)/T$ where $\ell_{it}(\Gamma,\Theta_i) = \ln f(y_{it}|y_{it-1},\Gamma,\Theta_i)$ denotes the log likelihood of one observation. Let

$$\overline{\Theta}_i(\Gamma) = \arg\max_{\Theta_i} \operatorname{plim}_{T\to\infty} \ell_i(\Gamma,\Theta_i),$$

and

$$\widehat{\Theta}_i(\Gamma) = \arg\max_{\Theta_i} \ell_i(\Gamma,\Theta_i),$$

so that under regularity conditions $\overline{\Theta}_i(\Gamma_0) = \Theta_{i0}$.

Following Severini (2000) and Pace and Salvan (2005), the concentrated likelihood for unit i

$$\hat{\ell}_i(\Gamma) = \ell_i\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right),$$

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

$$\bar{\ell}_i(\Gamma) = \ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right).$$

Now, define

$$u_{it}(\Gamma,\Theta_i) = \frac{\partial \ell_{it}(\Gamma,\Theta_i)}{\partial \Gamma},\ v_{it}(\Gamma,\Theta_i) = \frac{\partial \ell_{it}(\Gamma,\Theta_i)}{\partial \Theta_i},$$
$$u_i(\Gamma,\Theta_i) = \frac{1}{T}\sum_{t=1}^{T} u_{it}(\Gamma,\Theta_i),\ v_i(\Gamma,\Theta_i) = \frac{1}{T}\sum_{t=1}^{T} v_{it}(\Gamma,\Theta_i),$$
$$H_i(\Gamma) = -\lim_{T\to\infty} E\left[\frac{\partial v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial \Theta_i'}\right].$$

When Θ_{i0} is a vector of fixed effects, the Nagar expansion for $\widehat{\Theta}_i(\Gamma) - \overline{\Theta}_i(\Gamma)$ takes the form

$$\widehat{\Theta}_i(\Gamma) - \overline{\Theta}_i(\Gamma) = H_i^{-1}(\Gamma)\, v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right) + \frac{1}{T} B_i(\Gamma) + O_p\left(\frac{1}{T^{3/2}}\right), \tag{A.1}$$

where

$$\begin{aligned} B_i(\Gamma) &= H_i^{-1}(\Gamma)\left[\Xi_i(\Gamma)\, vec\left(H_i^{-1}(\Gamma)\right)\right. \\ &\quad \left. + \frac{1}{2} E\left(\frac{\partial}{\partial \Theta'} vec \frac{\partial v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial \Theta'}\right)' \left(H_i^{-1}(\Gamma)\otimes H_i^{-1}(\Gamma)\right) vec\left(\Upsilon_i(\Gamma)\right)\right], \end{aligned}$$

and

$$\Upsilon_i(\Gamma) = \Upsilon_i(\Gamma;\Gamma_0,\Theta_{i0}) = \lim_{T\to\infty} TE\left[v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)'\right],$$

$$\Xi_i(\Gamma) = \Xi_i(\Gamma;\Gamma_0,\Theta_{i0}) = \lim_{T\to\infty} TE\left[\frac{\partial v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial\Theta'}\otimes v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)'\right].$$

Next, expanding $\ell_i\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)$ around $\overline{\Theta}_i(\Gamma)$ for fixed Γ,

$$\begin{aligned}
& \ell_i\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right) - \ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right) \\
= & \frac{\partial\ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial\Theta'}\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right) \\
& +\frac{1}{2}\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right)'\frac{\partial^2\ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial\Theta\partial\Theta'}\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right)+O_p\left(\frac{1}{T^{3/2}}\right)
\end{aligned}$$

$$\begin{aligned}
= & \frac{\partial\ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial\Theta'}\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right) \\
& +\frac{1}{2}\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right)'E\left(\frac{\partial^2\ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)}{\partial\Theta\partial\Theta'}\right)\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right)+O_p\left(\frac{1}{T^{3/2}}\right)
\end{aligned}$$

$$\begin{aligned}
= & v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)'\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right) \\
& -\frac{1}{2}\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right)'H_i(\Gamma)\left(\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right)+O_p\left(\frac{1}{T^{3/2}}\right).
\end{aligned}$$

Substituting (A.1)

$$\ell_i\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right) - \ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right) = \frac{1}{2}v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)'H_i^{-1}(\Gamma)\,v_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)+O_p\left(\frac{1}{T^{3/2}}\right).$$

Taking expectations

$$E\left[\ell_i\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right) - \ell_i\left(\Gamma,\overline{\Theta}_i(\Gamma)\right)\right] = \frac{1}{2T}tr\left(H_i^{-1}(\Gamma)\,\Upsilon_i(\Gamma)\right)+O_p\left(\frac{1}{T^{3/2}}\right).$$

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

$$b_i(\Gamma) = \frac{1}{2}tr\left(H_i^{-1}(\Gamma)\,\Upsilon_i(\Gamma)\right) = \frac{1}{2}tr\left(H_i(\Gamma)\,Var\left(\sqrt{T}\left[\widehat{\Theta}_i(\Gamma)-\overline{\Theta}_i(\Gamma)\right]\right)\right).$$

Thus,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)-\sum_{i=1}^{N}\widehat{b}_i(\Gamma),$$

is expected to be a closer approximation to the target likelihood than $\sum_{i=1}^{N}\sum_{t=1}^{T}\ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)$.

Moreover, in the likelihood context, it is appropriate to consider a local version of the estimated bias (Pace and Salvan 2005) constructed as an expansion of $\widehat{b}_i(\Gamma)$ at Γ_0 using that at the truth

$$H_i^{-1}(\Gamma_0)\,\Upsilon_i(\Gamma_0) = 1.$$

Taking $\widehat{b}_i(\Gamma) = \frac{1}{2} tr\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right)$ also

$$\widehat{b}_i(\Gamma) = \frac{1}{2}p + \frac{1}{2}\sum_{j=1}^{p}\left[\lambda_j\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right) - 1\right],$$

where $\lambda_j\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right)$ denotes the j-th eigenvalue of $\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)$ and p is the dimension of Γ. Thus a local version of $\widehat{b}_i(\Gamma)$ gives

$$\widehat{b}_i(\Gamma) = \frac{1}{2}p + \frac{1}{2}\sum_{j=1}^{p}\left[\lambda_j\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right)\right] + O_p\left(\frac{1}{T}\right).$$

Moreover

$$\begin{aligned}
\frac{1}{2}\sum_{j=1}^{p}\left[\lambda_j\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right)\right] &= \frac{1}{2}\ln\det\left(\widehat{H}_i^{-1}(\Gamma)\,\widehat{\Upsilon}_i(\Gamma)\right) \\
&= -\frac{1}{2}\ln\det\widehat{H}_i(\Gamma) + \frac{1}{2}\ln\det\widehat{\Upsilon}_i(\Gamma),
\end{aligned}$$

which provided justification for the bias-corrected concentrated that I have used.

B Analytical expression for $\bar{\Upsilon}_i(\alpha, \widehat{\eta}_i(\alpha); \hat{\alpha}, \hat{\eta}_i)$ in the AR(1) model

Let us obtain an expression for $\bar{\Upsilon}_i(\alpha, \widehat{\eta}_i(\alpha); \hat{\alpha}, \hat{\eta}_i)$ in the dynamic panel example:

$$y_{it} = \alpha y_{it-1} + \eta_i + \epsilon_{it},$$

where $\epsilon_{it} \sim iidN(0,1)$. Then

$$\begin{aligned}
\ell_{it}(\alpha,\eta) &= C - \frac{1}{2}(y_{it} - \alpha y_{it-1} - \eta_i)^2, \\
\frac{\partial \ell_{it}(\alpha,\eta)}{\partial \eta} &= y_{it} - \alpha y_{it-1} - \eta_i \equiv v_{it}(\alpha,\eta) \equiv v_{it}, \\
\bar{v}_i &= \frac{1}{T}\sum_{t=1}^{T} v_{it},
\end{aligned}$$

and

$$\bar{\Upsilon}_i(\alpha, \eta; \alpha_0, \eta_0) = T Var_0(\overline{v}_i | y_{i0}).$$

Note that

$$\begin{aligned} v_{it} &= \epsilon_{it} + (\alpha_0 - \alpha)\, y_{it-1} + (\eta_{i0} - \eta_i)\,, \\ \bar{v}_i &= \bar{\epsilon}_i + (\alpha_0 - \alpha)\, \bar{y}_{i(-1)} + (\eta_{i0} - \eta_i)\,, \end{aligned}$$

$$Var_0\left(\bar{v}_i | y_{i0}\right) = \frac{1}{T} + (\alpha_0 - \alpha)^2 Var_0\left(\bar{y}_{i(-1)} | y_{i0}\right) + 2\left(\alpha_0 - \alpha\right) Cov_0\left(\bar{y}_{i(-1)}, \bar{\epsilon}_i | y_{i0}\right),$$

where $\bar{y}_{i(-1)} = \frac{1}{T}\sum_{t=1}^{T} y_{it-1}$. Since

$$\begin{aligned} \bar{y}_{i(-1)} &= h_T(\alpha_0)\,\eta_{i0} + c_T(\alpha_0)\, y_{i0} + \\ & \frac{1}{T}\left[\left(1 + \alpha_0 + \ldots + \alpha_0^{T-2}\right)\epsilon_{i1} + \left(1 + \alpha_0 + \ldots + \alpha_0^{T-3}\right)\epsilon_{i2} + \ldots + \epsilon_{iT-1}\right], \end{aligned}$$

where

$$\begin{aligned} h_T(\alpha_0) &= \frac{1}{T}\left[1 + (1 + \alpha_0) + \left(1 + \alpha_0 + \alpha_0^2\right) + \ldots + \left(1 + \alpha_0 + \ldots + \alpha_0^{T-2}\right)\right], \\ c_T(\alpha_0) &= \frac{1}{T}\left(1 + \alpha_0 + \ldots + \alpha_0^{T-1}\right). \end{aligned}$$

Thus

$$Var_0\left(\bar{y}_{i(-1)} | y_{i0}\right) = \frac{1}{T^2}\left[1 + (1 + \alpha_0)^2 + \ldots + \left(1 + \alpha_0 + \ldots + \alpha_0^{T-2}\right)^2\right] \equiv \omega_T(\alpha_0)\,,$$

$$Cov_0\left(\bar{y}_{i(-1)}, \bar{\epsilon}_i | y_{i0}\right) = \frac{1}{T^2}\left[\left(1 + \alpha_0 + \ldots + \alpha_0^{T-2}\right) + \ldots + 1\right] \equiv \psi_T(\alpha_0)\,,$$

$$Var_0\left(\bar{v}_i | y_{i0}\right) = \frac{1}{T} + (\alpha_0 - \alpha)^2 \omega_T(\alpha_0) + 2\left(\alpha_0 - \alpha\right)\psi_T(\alpha_0)\,,$$

and

$$\bar{\Upsilon}_i\left(\alpha, \eta; \alpha_0, \eta_0\right) = 1 + T\left(\alpha_0 - \alpha\right)^2 \omega_T(\alpha_0) + 2T\left(\alpha_0 - \alpha\right)\psi_T(\alpha_0)\,.$$

Thus

$$\bar{\Upsilon}_i\left(\alpha, \widehat{\eta}_i(\alpha); \hat{\alpha}, \hat{\eta}_i\right) = 1 + T\left(\hat{\alpha} - \alpha\right)^2 \omega_T(\hat{\alpha}) + 2T\left(\hat{\alpha} - \alpha\right)\psi_T(\hat{\alpha})\,.$$

C 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 noncontinuous data.

   = FINAL SAMPLE (2,066 individuals and 32,066 observations).

Table C1. My sample vs. Meghir and Pistaferri (2004)

Number of individuals	Meghir & Pistaferri (2004)		Hospido (2006)		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, N>9	(11,490)	4,539	(10,791)	5,247	−708
Female	(876)	3,663	(1,211)	4,036	−373
Self-employment, missing wages	(1323)	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	

D Sample composition and descriptive statistics

D.1 Sample 1

Table D1.1. Distribution of observations by year

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

Table D1.2. Distribution of observations by education

	Number of Individuals			
Number of Years	Whole sample	High School Dropout	High School Graduate	College Graduate
9	212	52	128	32
10	200	43	122	35
11	155	43	82	30
12	143	36	81	26
13	143	34	87	22
14	147	35	86	26
15	145	38	82	25
16	118	26	71	21
17	127	30	76	21
18	87	20	48	19
19	97	21	57	19
20	91	19	54	18
21	91	25	48	18
22	78	19	44	15
23	52	12	33	7
24	46	15	19	12
25	42	12	27	3
26	52	26	46	20

Table D1.3. 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.

D.2 Sample 2

Table D2.1. Distribution of observations by year

Year	Number of observations	Year	Number of observations
1968	366	1981	708
1969	414	1982	767
1970	446	1983	809
1971	475	1984	858
1972	509	1985	921
1973	543	1986	894
1974	580	1987	866
1975	613	1988	837
1976	645	1989	808
1977	630	1990	766
1978	627	1991	734
1979	644	1992	696
1980	676	1993	653

Table D2.2. Distribution of observations by education

Number of Years	Number of Individuals: Whole sample	High School Dropout	High School Graduate	College Graduate
9	264	78	133	53
10	182	42	103	37
11	150	31	87	32
12	150	33	88	29
13	131	44	69	18
14	97	29	56	12
15	85	27	43	15
16	64	18	34	12
17	54	13	31	10
18	25	6	13	6
19	38	9	19	10
20	21	4	14	3
21	18	7	8	3
22	20	4	15	1
23	14	4	7	3
24	6	2	3	1
25	17	5	10	2
26	10	3	5	2

Table D2.3. Descriptive Statistics

	1968	1980	1993
Age	38.18	39.34	42.60
	(6.35)	(9.24)	(5.65)
HS Dropout	0.43	0.31	0.13
HS Graduate	0.41	0.51	0.62
Hours	2252	2146	2130
	(514)	(483)	(521)
Married	0.83	0.84	0.86
White	0.69	0.66	0.67
Children	2.88	1.39	1.37
	(2.06)	(1.28)	(1.28)
Family Size	5.03	3.65	3.60
	(2.00)	(1.64)	(1.47)
North-East	0.17	0.16	0.16
North-Central	0.29	0.27	0.24
South	0.38	0.45	0.45
SMSA	0.68	0.64	0.52

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

References

[1] Abowd, J. and D. Card (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", cap. 3, Tesis Doctoral, Universidad Complutense.

[3] Alvarez, J. and M. Arellano (2004): "Robust likelihood estimation of dynamic panel data models", CEMFI Working Paper No. 0421.

[4] Alvarez, J., M. Browning, and M. Ejrnæs (2001): "Modelling Income Processes with Lots of Heterogeneity", unpublished manuscript.

[5] Arellano, M. (2003): "Discrete Choices with Panel Data", *Investigaciones Económicas*, 27, 423-458.

[6] Arellano, M. and J. Hahn (2006a): "Understanding bias in nonlinear panel models: Some recent developments", In: R. Blundell, W. Newey, and T. Persson (eds.): *Advances in Economics and Econometrics, Ninth World Congress*, Cambridge University Press, forthcoming.

[7] Arellano, M. and J. Hahn (2006b): "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects", unpublished manuscript.

[8] Arellano, M. and B. Honoré (2001): "Panel Data Models. Some Recent Developments", in J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Vol. 5, Ch. 53.

[9] Banks, J., R. Blundell, and A. Brugiavini (2001): "Risk Pooling, Precautionar Saving and Consumption Growth", *Review of Economic Studies,* 68 (4), 757-779.

[10] Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

[11] Browning, M. and A. Lusardi (1996): "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature*, Vol. XXXIV, 1797-1855.

[12] Carro, J. (2006): "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects", *Journal of Econometrics*, in press.

[13] Chamberlain, G. (1984): "Panel Data", in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, Elsevier Science.

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

[15] Cox, D. R. and N. Reid (1987): "Parameter Orthogonality and Approximate Conditional Inference" (with discussion), *Journal of the Royal Statistical Society*, Series B, 49, 1-39.

[16] DiCiccio, T. J. and S. E. Stern (1993): "An adjustment to Profile Likelihood Based on Observed Information", Technical Report, Department of Statistic, Stanford University.

[17] Fernández-Val, I. (2005): "Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects", unpublished manuscript.

[18] Gottschalk, P. and R. Moffitt (1994): "The Growth of Earnings Instability in the US Labor Market", *Brookings Paper on Economic Activity*, 2, 217-272.

[19] Hahn, J. and G. Kuersteiner (2004): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", unpublished manuscript.

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

[21] Horowitz, J. L. (2003): "Bootstrap Methods for Markov Processes", *Econometrica*, 71, 1049-1082.

[22] Horowitz, J. L. and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data", *Review of Economic Studies*, 63, 145-168.

[23] Katz, L.F. and D. H. Autor (1999): "Changes in the wage structure and earnings inequality", in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3, Chapter 25, pp. 1436-1555. New York: North-Holland.

[24] Lancaster, T. (2002): "Orthogonal Parameters and Panel Data", *Review of Economic Studies*, 69, 647-666.

[25] Levy, F. and R. J. Murnane (1992): "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations", *Journal of Economic Literature*, 30, 1333-1381.

[26] Lillard, L. A. and R. J. Willis (1978): "Dynamic Aspects of Earnings Mobility", *Econometrica*, 46, 985-1012.

[27] Lin, C.-C. (2005): "Estimation and Testing of the ARCH(1) effect in Panel Data when N and T are Large", unpublished manuscript.

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

[29] Meghir, C. and F. Windmeijer (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.

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

[31] Nelson, D. B. (1992): "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, 59, 347-370.

[32] Neyman, J. and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16, 1-32.

[33] Pace, L. and A. Salvan (2005): "Adjustments of the profile likelihood from a new perspective", *Journal of Statistical Planning and Inference*, forthcoming.

[34] Severini , T. A. (1998): "An Approximation to the Modified Profile Likelihood Function", *Biometrika*, 85, 403-411.

[35] Severini, T. A. (2002): *Likelihood Methods in Statistics*, Oxford University Press.

[36] Woutersen, T. (2002): "Robustness against Incidental Parameters", unpublished manuscript.

FIGURES

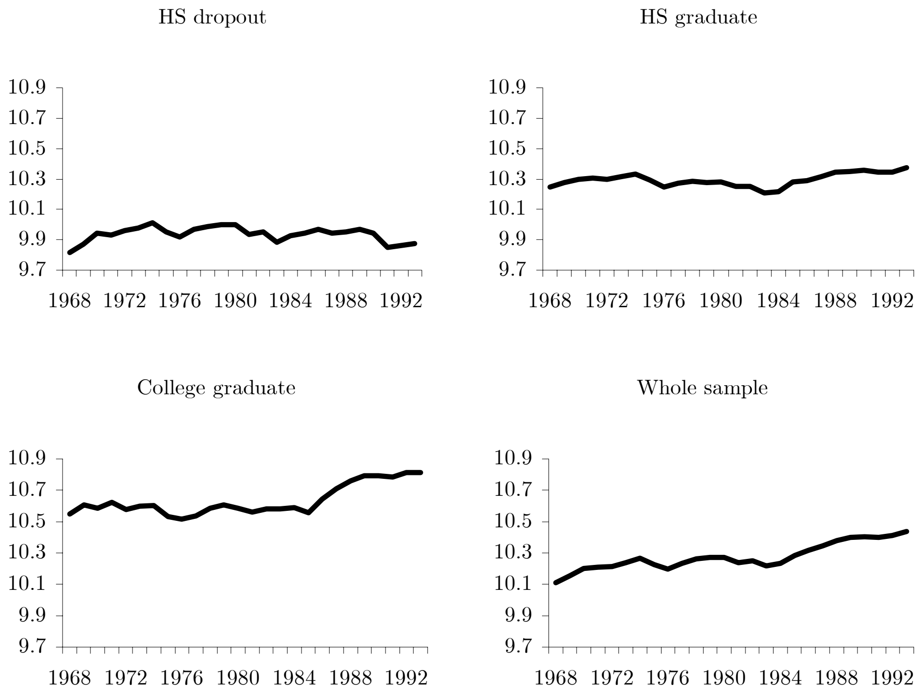

Figure 1. The mean of log wages.

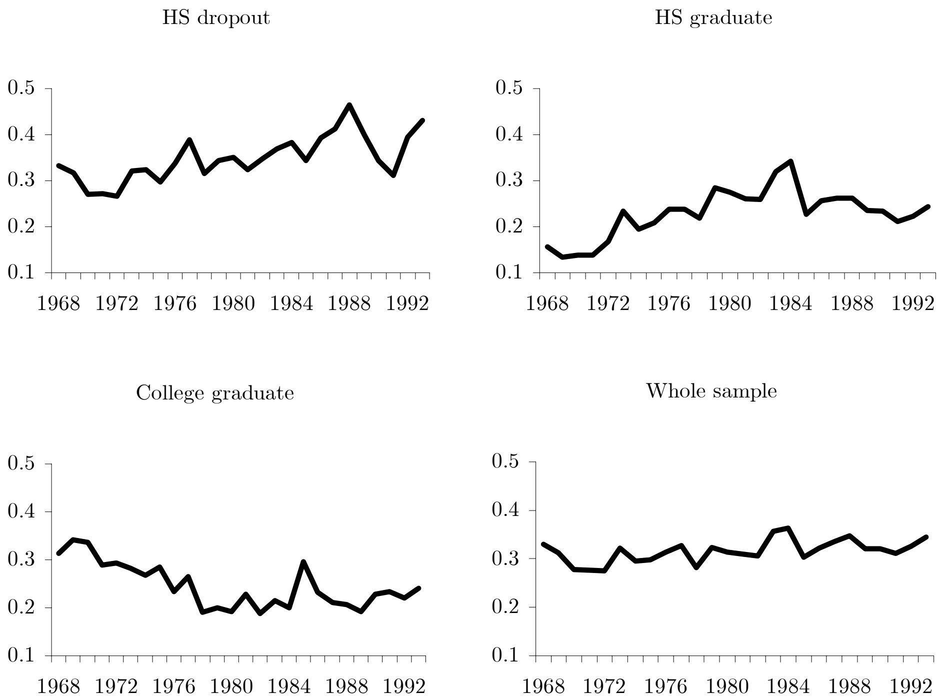

Figure 2. The variance of log wages.

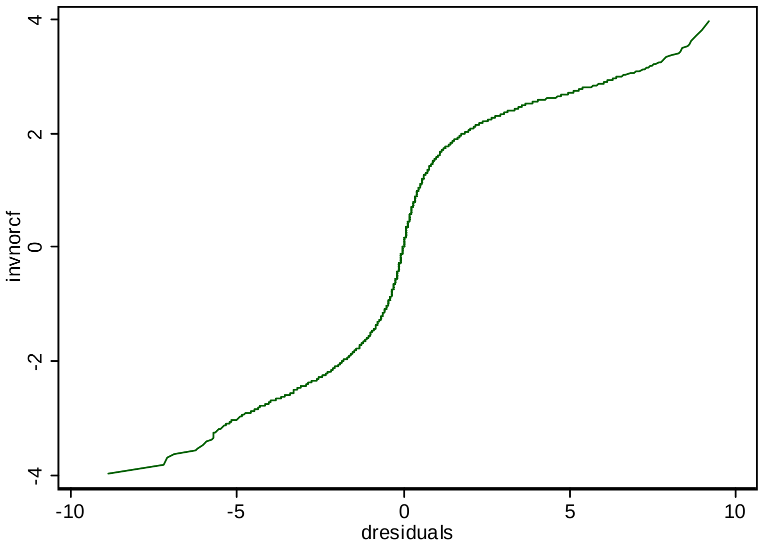

Figure 3. Distribution of Residuals in First Differences.

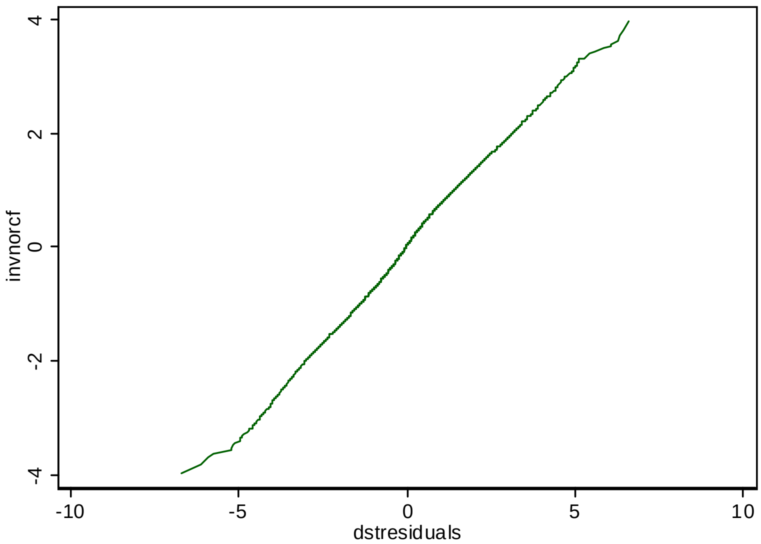

Figure 4. Distribution of Standarized Residuals in First Differences.

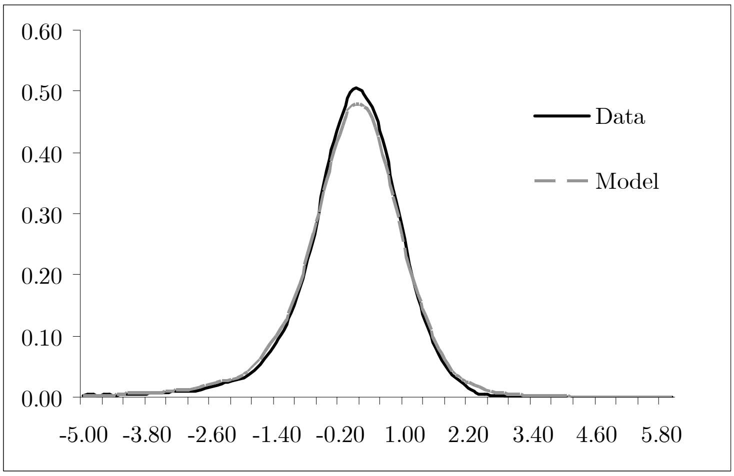

Figure 5. Kernel densities of logwages and simulated logwages.

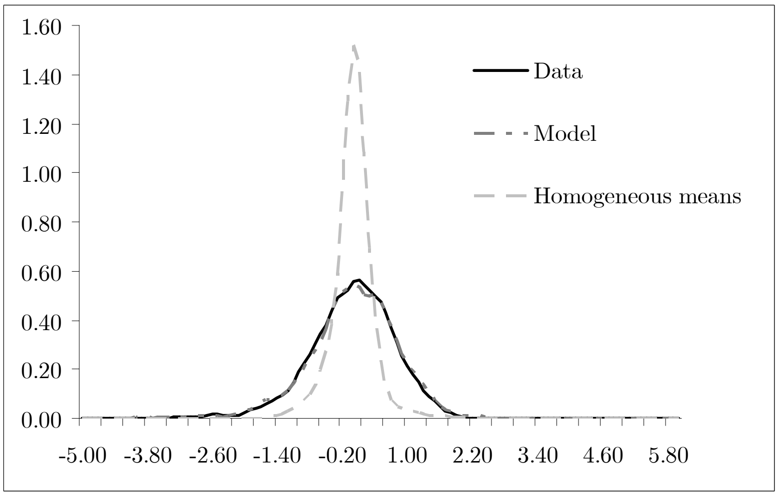

Figure 6. Kernel density of individual means.

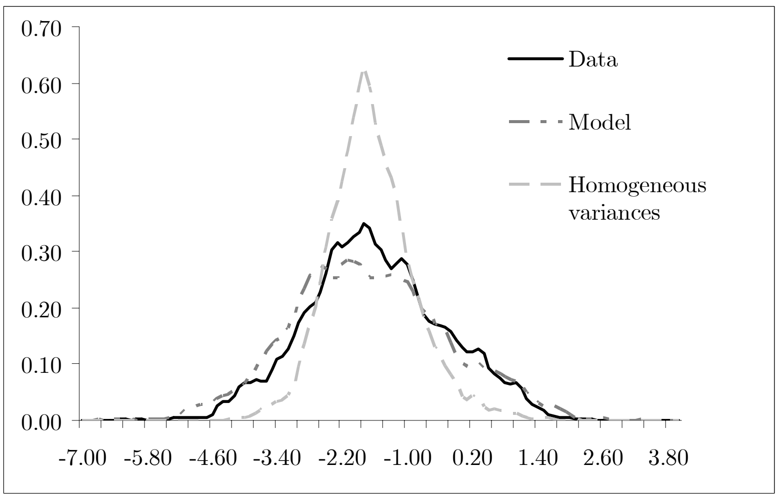

Figure 7. Kernel density of individual logvariances.

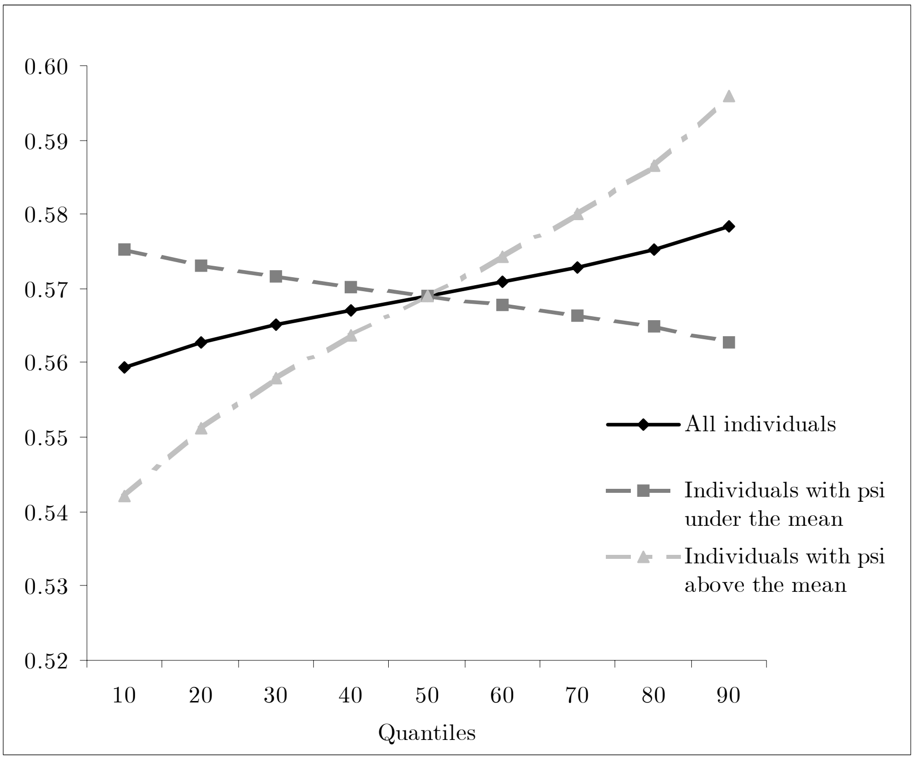

Figure 8. Mean Elasticities.

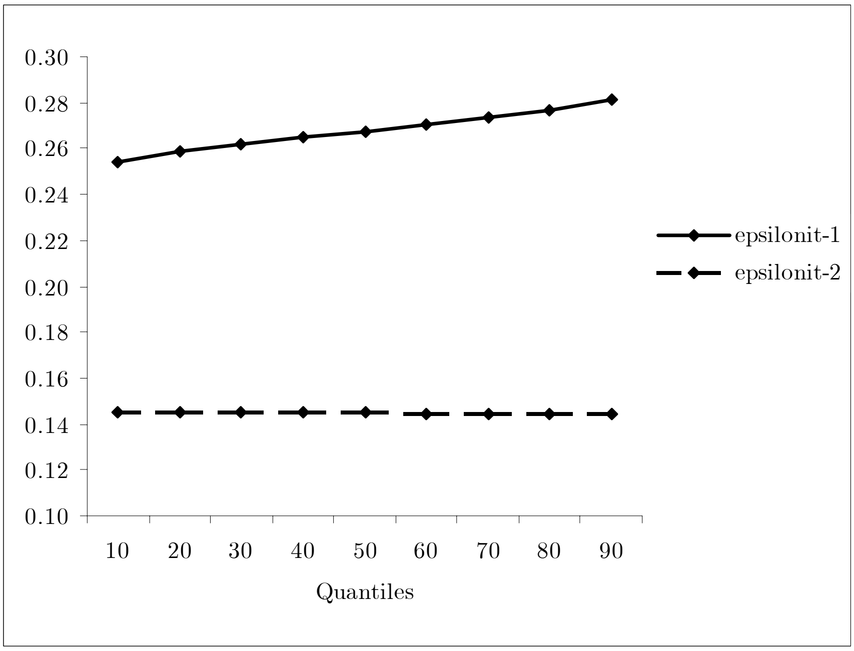

Figure 9. Mean Marginal Effects of Past Shocks over the Distribution of Wages.

CEMFI WORKING PAPERS

0501 C*laudio Michelacci and Vincenzo Quadrini*: "Borrowing from employees: Wage dynamics with financial constraints".

0502 *Gerard Llobet and Javier Suarez*: "Financing and the protection of innovators".

0503 *Juan-José Ganuza and José S. Penalva*: "On information and competition in private value auctions".

0504 *Rafael Repullo*: "Liquidity, risk-taking, and the lender of last resort".

0505 *Marta González and Josep Pijoan-Mas*: "The flat tax reform: A general equilibrium evaluation for Spain".

0506 *Claudio Michelacci and Olmo Silva*: "Why so many local entrepreneurs?".

0507 *Manuel Arellano and Jinyong Hahn*: "Understanding bias in nonlinear panel models: Some recent developments".

0508 *Aleix Calveras, Juan-José Ganuza and Gerard Llobet*: "Regulation and opportunism: How much activism do we need?".

0509 *Ángel León, Javier Mencía and Enrique Sentana*: "Parametric properties of semi-nonparametric distributions, with applications to option valuation".

0601 *Beatriz Domínguez, Juan José Ganuza and Gerard Llobet*: "R&D in the pharmaceutical industry: a world of small innovations".

0602 *Guillermo Caruana and Liran Einav*: "Production targets".

0603 *Jose Ceron and Javier Suarez*: "Hot and cold housing markets: International evidence".

0604 *Gerard Llobet and Michael Manove*: "Network size and network capture".

0605 *Abel Elizalde*: "Credit risk models I: Default correlation in intensity models".

0606 *Abel Elizalde*: "Credit risk models II: Structural models".

0607 *Abel Elizalde*: "Credit risk models III: Reconciliation reduced – structural models".

0608 *Abel Elizalde*: "Credit risk models IV: Understanding and pricing CDOs".

0609 *Gema Zamarro*: "Accounting for heterogeneous returns in sequential schooling decisions".

0610 *Max Bruche*: "Estimating structural models of corporate bond prices".

0611 *Javier Díaz-Giménez and Josep Pijoan-Mas*: "Flat tax reforms in the U.S.: A boon for the income poor".

0612 *Max Bruche and Carlos González-Aguado*: "Recovery rates, default probabilities and the credit cycle".

0613 *Manuel Arellano and Jinyong Hahn*: "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects".

0614 *Manuel Arellano and Stéphane Bonhomme*: "Robust priors in nonlinear panel data models".

0615 *Laura Crespo*: "Caring for parents and employment status of European mid-life women".

0701 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: "Royalty stacking in high tech industries: separating myth from reality".

0702 *Anne Layne-Farrar, A. Jorge Padilla and Richard Schmalensee*: "Pricing patents for licensing in standard setting organizations: Making sense of *FRAND* commitments".

0703 *Damien Geradin, Anne Layne-Farrar and A. Jorge Padilla*: "The *ex ante* auction model for the control of market power in standard setting organizations".

0704 *Abel Elizalde*: "From Basel I to Basel II: An analysis of the three pillars".

0705 *Claudio Michelacci and Josep Pijoan-Mas*: "The effects of labor market conditions on working time: the US-UE experience".

0706 *Robert J. Aumann and Roberto Serrano*: "An economic index of *riskiness*".

0707 *Roberto Serrano*: "El uso de sistemas dinámicos estocásticos en la Teoría de Juegos y la Economía".

0708 *Antonio Cabrales and Roberto Serrano*: "Implementation in adaptive better-response dynamics".

0709 *Roberto Serrano*: "Cooperative games: Core and Shapley value".

0710 *Allan M. Feldman and Roberto Serrano*: "Arrow's impossibility theorem: Preference diversity in a single-profile world".

0711 *Victor Aguirregabiria and Pedro Mira*: "Dynamic discrete choice structural models: A Survey".

0712 *Rene Saran and Roberto Serrano*: "The evolution of bidding behaviour in private-values auctions and double auctions".

0713 *Gabriele Fiorentini and Enrique Sentana:* "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models".

0714 *Antonio Díaz de los Ríos and Enrique Sentana:* "Testing uncovered interest parity: A continuous-time approach".

0715 *Francisco Peñaranda and Enrique Sentana:* "Duality in mean-variance frontiers with conditioning information".

0716 *Stefano Gagliarducci, Tommaso Nannicini and Paolo Naticchioni:* "Electoral rules and politicians' behavior: A micro test".

0717 *Laura Hospido*: "Modelling heterogeneity and dynamics in the volatility of individual wages".