Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages

Laura Hospido*

CEMFI and USC

Preliminary and incomplete version
Please, do not quote

March 31, 2006

Abstract

In this paper I consider a model for the heterogeneity and dynamics of the conditional mean and the conditional variance of individual wages. In particular, I propose a dynamic panel data model with linear individual effects in the mean and multiplicative individual effects in the conditional ARCH type variance function. I posit a distribution for earning shocks and I build a modified likelihood function for estimation and inference. Using a newly developed 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. 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 PSID samples. The empirical analysis is conducted on data drawn from the 1968-1993 PSID. I find a significant estimate for the AR coefficient in the mean and for the ARCH effects in the variance. However, the latter dissapear when there are no job changes in the sample.

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

*I wish to thank Manuel Arellano, and Stéphane Bonhomme for encouragement and advice. For comments and support, I am grateful to Jesús Carro, Alfredo Martín, Javier Mencía, Enrique Sentana, Gema Zamarro, and the seminar participants at CEMFI. All errors are mine. I gratefully acknowledge financial support from Fundación Ramón Areces. Please send comments or suggestions to hospido@cemfi.es.

1 Introduction

Estimates of 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 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 his/her future earnings, in order to guide savings or other decisions. As the individual faces various sorts of uncertainty, we shall be interested in forecasting not only the level of earnings but also its variance. The properties of the variance will be important for describing wage profiles over time and for better understanding what drives fluctuations in them. A richer specification can contribute also to modeling choices in models that use the earnings process as an input. In fact, recent studies stress the relevance of considering a variance that varies with 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 since the 70's until today. This growth in the aggregate variance is associated with an increase in inequality. Much less it 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. In the model people at different levels of the income distribution face a different variance of their time-income profile.

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. In this paper 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, under this problem, 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 (2005) 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 (2004), Fernández-Val (2005)) and bias corrections for the concentrated likelihood (DiCiccio and Stern (1993), Severini (1998), 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.

I develop, for the first time, several versions of the modified likelihood based on DiCiccio and Stern (1993), Severini (1998), and Pace and Salvan (2005) adapted to a nonlinear dynamic 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 aplication.

The empirical analysis is conducted on data drawn from the 1968-1993 Panel Study of Income Dynamics (PSID). These model 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 a significant estimate for the AR coefficient in the mean and for the ARCH effects in the variance. However, the latter dissapear when there are no job changes in the sample.

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. 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. This parsimonious specification would be useful for describing and estimating wage distributions (Chamberlain and Hirano (1999)). Meghir and Windmeijer (1999) and Albarrán (2004) use single-shock models as well. They recover orthogonality conditions for the estimation of ARCH process but, in real data, GMM estimators perform bad.

Two limitations of the model 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 wages variance may be due to job mobility, so these issues need to be addresed 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 preliminary empirical application on individual wages. Section 6 concludes with a future research agenda.

2 The Model and the Likelihood Function

2.1 The Model

I consider the following model of individual wages where i and t index individuals and time, respectively:[1]

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

with

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

and

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

In these expressions, $\{y_{i0}, ..., y_{iT}\}_{i=1}^N$ are the observed data, $\Theta_i = (\eta_i, \psi_i)'$ are the individual unobserved fixed effects, $\Upsilon^T = (\omega_0, ..., \omega_T)'$ are time effects, e_{it} is an ARCH process, $\{\epsilon_{it}\}$ is an *i.i.d.* sequence with zero mean and unit variance, and Λ is a small positive number used to approximate the absolute value function by means of a rotated hyperbola, so that h_{it} is everywhere differentiable. The log formulation implies that h_{it} is always nonnegative, regardless of the parameter values (Nelson, 1992). Finally, we denote the vector of common parameters as $\Gamma = (\alpha, \beta, \omega_0, \omega_1, \ldots \omega_T)'$.

For the conditional mean, I consider an autorregresive 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. 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 three 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 feature captures the fact that variances at each period may differ as a result of aggregate effects, ω_t. The last one is dynamics, β, reflecting that

[1] 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.

periods of high volatility in wages tend to be consecutive and viceversa. This feature would be noticiable not only for sellers, but also for funds managers or, in general, for workers that receive bonuses.

2.2 The Likelihood Function

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

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

and the individual likelihood, conditioned on initial observations, fixed effects, strictly exogenous variables (note that I assume strict exogeneity in time effects), is

$$f\left(y_{i1}, ..., y_{iT}|y_{i0}, \omega_0, ..., \omega_T, \Theta_{i0}\right) = \prod_{t=1}^{T} f\left(y_{it}|y_{it-1}, \Upsilon^t, \Theta_{i0}, \Gamma_0\right).$$

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 write

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

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

$$y_{i1} = \alpha y_{i0} + \eta_i + \left[h\left(\epsilon_{i0}, \psi_i, \omega_1\right)\right]^{1/2}\epsilon_{i1},$$

where $h\left(\epsilon_{i0}, \psi_i, \omega_1\right) = h\left(y_{i0}, y_{i,-1}, y_{i,-2}, ...\right)$. We have a model for $f\left(y_{i1}|y_{i0}, y_{i,-1}, y_{i,-2}, ..., \omega^1, \Theta_{i0}\right)$ or for $f\left(y_{i1}|y_{i0}, \epsilon_{i0}, \omega^1, \Theta_{i0}\right)$ where ϵ_{i0} resumes all the past values of y_{it}, but what we need is $f\left(y_{i1}|y_{i0}, \omega^1, \Theta_{i0}\right)$. We know that

$$E\left(y_{i1}|y_{i0}, \omega^1, \Theta_{i0}\right) = E\left(y_{i1}|y_{i0}, \epsilon_{i0}, \omega^1, \Theta_{i0}\right) = \alpha y_{i0} + \eta_i,$$

and

$$\begin{aligned} Var\left(y_{i1}|y_{i0}, \omega^1, \Theta_{i0}\right) &= E\left(h\left(\epsilon_{i0}, \psi_i, \omega_1\right)|y_{i0}, \omega^1, \Theta_{i0}\right) + Var\left(\alpha y_{i0} + \eta_i|y_{i0}, \omega^1, \Theta_{i0}\right) \\ &= E\left(h\left(\epsilon_{i0}, \psi_i, \omega_1\right)|y_{i0}, \omega^1\omega_1, \Theta_{i0}\right) + Var\left(\eta_i|y_{i0}, \omega^1, \Theta_{i0}\right) \\ &= \varphi\left(\eta_i, \psi_i, \Gamma\right). \end{aligned}$$

Thus, $f\left(y_{i1}|y_{i0},\omega^1,\Theta_{i0}\right)$ would be a mixture given that:

$$f\left(y_{i1}|y_{i0},\omega^1,\Theta_{i0}\right)=\int f\left(y_{i1}|y_{i0},\epsilon_{i0},\omega^1,\Theta_{i0}\right)dG\left(\epsilon_{i0}|y_{i0},\omega^1,\Theta_{i0}\right).$$

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

Let the individual likelihood function be

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

and the log-likelihood of each observation

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

where

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

3 A likelihood-based solution to the incidental parameters problem in dynamic nonlinear models with multiple effects

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

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

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 $\widehat{\Theta}_i\left(\Gamma\right)$: as only a finite number T of observations are available to estimate each Θ_i, the estimation error of $\widehat{\Theta}_i\left(\Gamma\right)$ does not vanish as the sample size N grows, and this error contamines the estimates of common parameters in nonlinear models. Let

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

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

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

$$\begin{aligned}
H_i(\Gamma) &= -E\left[\frac{\partial^2\ell_{it}(\Gamma,\Theta_i)}{\partial\Theta_i\partial\Theta_i'}\right],\\
\Upsilon_i(\Gamma) &= E\left[\frac{\partial\ell_{it}(\Gamma,\Theta_i)}{\partial\Theta_i}\cdot\frac{\partial\ell_{is}(\Gamma,\Theta_i)}{\partial\Theta_i'}\right].
\end{aligned}$$

I will consider three alternative estimators of Γ which maximize a bias-corrected concentrated likelihood function:

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

Letting $\hat{b}_i(\Gamma)$ be an estimated bias, $\widetilde{\Gamma}$ is expected to be less biased than the MLE $\widehat{\Gamma}$. Moreover, in a likelihood context, we can 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

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

This is why, in a likelihood context, I use a *Determinant Based Approach* and, for Pseudo Likelihoods, I use a *Trace Based Approach*.

For estimating the bias we need to estimate the hessian term, $H_i(\Gamma)$, and the expected outer product term, $\Upsilon_i(\Gamma)$. For estimating the first one we can 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 $T^{-1}\sum_{t=1}^{T}\frac{\partial \ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial\Theta_i} = 0$, also $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\partial \ell_{it}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial\Theta_i} \cdot \frac{\partial \ell_{is}\left(\Gamma,\widehat{\Theta}_i(\Gamma)\right)}{\partial\Theta_i'} = 0$, 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 Determinant Based Approach Using Expected Quantities

This approach is based on the expectation

$$\bar{\Upsilon}_i(\Gamma, \Theta_i; \Gamma_0, \Theta_{i0}) = E_0\left[\Upsilon_i(\Gamma)|y_{i0}\right].$$

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: $\bar{\Upsilon}_i\left(\Gamma, \widehat{\Theta}_i(\Gamma); \widehat{\Gamma}, \widehat{\Theta}_i\right)$. It can be regarded as a dynamic version of Severini (1998) 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 we have a first corrected estimate

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

we calculate

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

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, we shall 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}.

3.2 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,

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

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

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.

3.3 Determinant Based Approach Using a Bootstrap Estimate of $Var\left[\widehat{\Theta}_i(\Gamma)\right]$

The first step consists in generating parametric bootstrap samples $\{y_{i1}^m, ..., y_{iT}^m\}_{i=1}^N$ $(m = 1, ..., M)$ 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, calculating 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. For each sample I have estimated α by maximum likelihood and by modified maximum likelihood. I have simulated samples for different samples sizes because I expect the modified MLE's 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 estimators perform worse.

Here the MLE of α is

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

where

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

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

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

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

$$\begin{aligned} \widehat{H}_i\left(\alpha\right) &= -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\ell_{it}\left(\alpha,\widehat{\eta}_i\left(\alpha\right)\right)}{\partial\eta^2} = 1, \\ \Upsilon_i\left(\alpha\right) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\partial\ell_{it}\left(\alpha,\eta\right)}{\partial\eta}\cdot\frac{\partial\ell_{is}\left(\alpha,\eta\right)}{\partial\eta} = T\overline{v}_i^2, \end{aligned}$$

where $\bar{v}_i = \frac{1}{T}\sum\limits_{t=1}^{T} v_{it}$, $v_{it} = \frac{\partial\ell_{it}(\alpha,\eta)}{\partial\eta}$,[3] and as it is shown in Appendix B

$$\begin{aligned} \bar{\Upsilon}_i\left(\alpha,\eta;\alpha_0,\eta_0\right) &= TE\left(\overline{v}_i^2|y_{i0}\right) = 1 + T\left(\alpha_0-\alpha\right)^2\omega_T\left(\alpha_0\right) + 2T\left(\alpha_0-\alpha\right)\psi_T\left(\alpha_0\right) \\ &\quad + T\left\{\left(\alpha_0-\alpha\right)\left[h_T\left(\alpha_0\right)\eta_{i0} + c_T\left(\alpha_0\right)y_{i0}\right] + \left(\eta_{i0}-\eta_i\right)\right\}^2, \end{aligned}$$

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

with

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

Thus

$$
\begin{aligned}
\bar{\Upsilon}_i\left(\alpha,\widehat{\eta}_i(\alpha);\hat{\alpha},\hat{\eta}_i\right) &= 1+T(\hat{\alpha}-\alpha)^2\omega_T(\hat{\alpha})+2T(\hat{\alpha}-\alpha)\psi_T(\hat{\alpha})\\
&\quad +T\left\{(\hat{\alpha}-\alpha)\left[h_T(\hat{\alpha})\hat{\eta}_i+c_T(\hat{\alpha})y_{i0}\right]+\left(\hat{\eta}_i-\widehat{\eta}_i(\alpha)\right)\right\}^2,
\end{aligned}
$$

or using $\hat{\eta}_i-\widehat{\eta}_i(\alpha)=-(\hat{\alpha}-\alpha)\,\bar{y}_{i(-1)}$, also:

$$
\begin{aligned}
\bar{\Upsilon}_i\left(\alpha,\widehat{\eta}_i(\alpha);\hat{\alpha},\hat{\eta}_i\right) &= 1+T(\hat{\alpha}-\alpha)^2\omega_T(\hat{\alpha})+2T(\hat{\alpha}-\alpha)\psi_T(\hat{\alpha})\\
&\quad +T(\hat{\alpha}-\alpha)^2\left\{h_T(\hat{\alpha})\hat{\eta}_i+c_T(\hat{\alpha})y_{i0}-\bar{y}_{i(-1)}\right\}^2.
\end{aligned}
$$

It follows that in this case

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

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

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

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 indexs the simulated samples by parametric bootstrap.

- 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,

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

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

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

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* 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 $\hat{\alpha}$.[4]

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
Expected Quantities	0.4077	0.0172	0.0184	0.5702	0.0156	0.0156
Bootstrap Variance	0.4745	0.0213	0.0193	0.7158	0.0182	0.0170
Trimming	0.3726	0.0168	0.0154	0.5845	0.0155	0.0150
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=1. SD: Sample standard deviation. Mean SE: Mean of estimated standard errors by individual block-bootstrap.

I find some difference 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. We will see an example in the next subsection. We see in the table that the fixed effects MLE is biased downward by around 35-40 percent in both cases. The bias corrections, except the one proposed by Lancaster (2002) that is consistent for fixed T, all

[4] Notice that, opposite to the block bootstrap procedure used in time-series literature (Hall and Horrowitz (1996), Horrowitz (2002)), 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	0.4412	0.0109	0.0111	0.6766	0.0096	0.0093
Bootstrap Variance	0.4962	0.0119	0.0115	0.7781	0.0106	0.0104
Trimming	0.4442	0.0106	0.0101	0.6949	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=200; trimming=1. SD: Sample standard deviation. Mean SE: Mean of estimated standard errors by individual block-bootstrap.

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 mentioned of the bias-correcting estimators with respect to the estimator proposed by Lancaster is its generality. With only a slight modification of the previous expressions we can deal with a more complex model. Let us see how the modified MLE's work in finite sample for a 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 $\alpha = 0.5$. We denote as $\Theta_i = \left(\eta_i, \sigma_i^2\right)'$ the vector of fixed effects. The MLE of α is

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

where

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

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

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

- Determinant Based Approach Using Expected Quantities: now

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

$$\begin{aligned}
\Upsilon_i(\alpha) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\begin{pmatrix}\frac{\partial\ell_{it}}{\partial\eta}\cdot\frac{\partial\ell_{is}}{\partial\eta} & \frac{\partial\ell_{it}}{\partial\eta}\cdot\frac{\partial\ell_{is}}{\partial\sigma^2}\\ \frac{\partial\ell_{it}}{\partial\sigma^2}\cdot\frac{\partial\ell_{is}}{\partial\eta} & \frac{\partial\ell_{it}}{\partial\sigma^2}\cdot\frac{\partial\ell_{is}}{\partial\sigma^2}\end{pmatrix} \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\begin{pmatrix}\frac{(y_{it}-\alpha y_{it-1}-\eta_i)}{\sigma_i^2}\cdot\frac{(y_{is}-\alpha y_{is-1}-\eta_i)}{\sigma_i^2} & \frac{(y_{it}-\alpha y_{it-1}-\eta_i)}{\sigma_i^2}\cdot\frac{(y_{is}-\alpha y_{is-1}-\eta_i)^2-\sigma_i^2}{2\sigma_i^4}\\ \frac{(y_{it}-\alpha y_{it-1}-\eta_i)^2-\sigma_i^2}{2\sigma_i^4}\cdot\frac{(y_{is}-\alpha y_{is-1}-\eta_i)}{\sigma_i^2} & \frac{(y_{it}-\alpha y_{it-1}-\eta_i)^2-\sigma_i^2}{2\sigma_i^4}\cdot\frac{(y_{is}-\alpha y_{is-1}-\eta_i)^2-\sigma_i^2}{2\sigma_i^4}\end{pmatrix}.
\end{aligned}$$

And 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 $\left\{\prod_{t=1}^{T}f\left(y_{it}|y_{i0},\hat{\alpha},\widehat{\Theta}_i\right)\right.$ That is,

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

which leads to

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

- 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

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

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

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

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,

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

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)	0.3900	0.0346	0.4739	0.0160
Expected Quantities (2nd)	0.4720	0.0424	0.5040	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	0.3205	0.0349	0.4333	0.0117

Note: N=500; simulations=300; parametric bootstrap samples=300; trimming=1. SD: Sample standard deviation.

We see in the table that the fixed effects MLE is biased downward 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 we iterate the corrections.

I have also found conclusions very similar in an alternative specification with time effects in the variance in addition to the individual heterogeneity.

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

Now the model design is

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

where I use $\sqrt{2/\pi}$ as aproximation for $E\left[\sqrt{\epsilon_{it}^2+\Lambda}\right]$ given that $\epsilon_{it} \sim N(0,1)$. 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

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

where

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

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 $\widehat{\psi}_i(\Gamma)$ is computed such that, for this given value of Γ, the individual log likelihood is maximized with respect to ψ.

The MMLE is

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

where

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

for

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

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 we calculate numerical first and second derivatives.

Table 4 reports estimates for $T=8$ and 16, and $N=1000$.

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	0.4978	0.0117	−0.1023	0.0777	0.4989	0.0077	0.3603	0.0245
Trimming	0.4991	0.0127	0.0297	0.0764	0.4990	0.0077	0.4647	0.0251

Note: N=1000; trimming=1. SD: Sample standard deviation. T=8: simulations=100; trimming: 95% successful convergence. T=16: simulations=50; trimming: 100% successful convergence.

In this case $\hat{\alpha}$ is not biased, and with the trimming correction we can correct an otherwise seriously biased MLE of β.

4.4 The AR(1)-EARCH(1) panel model with both fixed effects in the mean and in the variance

Here the model design is

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

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 $T = 16$, $N = 1000$, $\alpha_0 = 0.5$, and $\beta_0 = 0.5$. 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

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

where

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

and the MMLE is

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

where

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

for

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

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 we calculate numerical first, second and cross derivatives. Table 5 reports estimates for $T = 16$ and $N = 1000$.

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

Estimator of $(\alpha, \beta)'$	Mean $\hat{\alpha}$	SD $\hat{\alpha}$	Mean $\hat{\beta}$	SD $\hat{\beta}$
MLE	0.3984	0.0072	0.4174	0.0377
Trimming	0.4383	0.0206	0.4782	0.0748

Note: N=1000; simulations=20; trimming=1. SD: Sample standard deviation.

Again, I obtain estimates with less bias when I use the modified maximum likelihood estimator.

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 Panel Study of Income Dynamics. 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 labor as labor part of farm income, business income, overtime, comissions, etc. Figures 1 and 2, at the end of the paper, 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)).

The variable that I use in the estimation 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.

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

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

In this preliminar version of the model, I deal with additive aggregate effects in the variance by regarding y_{it} as standarized wages[5].

First, I estimate in the sample by MLE the following four specifications of the model: (i) model A consists on a specification AR(1)-EARCH(1) without fixed effects in the mean nor in the variance; (ii) model B is a specification AR(1)-EARCH(1) only with fixed effects in the mean; (iii) model C, a specification AR(1)-EARCH(1) only with fixed effects in the variance; and finally (iv) model D consists on a specification AR(1)-EARCH(1) with both fixed effects in the mean and in the variance. The results are the following

Table 6. α **and** β **estimates for different specifications**

	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{E(\eta_i)}$	$\widehat{Var(\eta_i)}$	$\widehat{E(\psi_i)}$	$\widehat{Var(\psi_i)}$
Model A	0.8448	0.5391				
Model B	0.5071	0.8829	0.0201	0.1682		
Model C	0.9027	0.3910			-1.7429	1.8725
Model D	0.4822	0.4832	0.0025	0.1754	-2.0284	1.9080

Note: $\widehat{E(\cdot)}$: mean of estimated individual fixed effects; $\widehat{Var(\cdot)}$: variance of estimated individual fixed effects.

The idea behind is that if we do not take into account unobserved heterogeneity, both in the mean and in the variance, we obtain much bigger estimates for α and β. Also when only one of the two types of effects is considered, $\hat{\alpha}$ or $\hat{\beta}$, one in each case, capture the effect of the unobserved heterogeneity.

Table 7 presents the estimation results by MLE and by maximization of the trimmed corrected concentrated likelihood. Before comenting the values in the table, only to mention that AR(1) GMM estimates for these sample give us $\hat{\alpha}$'s around 0.40 ($\hat{\alpha}_{WG} = 0.4048$, $\hat{\alpha}_{GMM1} = 0.4138$, $\hat{\alpha}_{GMM2} = 0.4092$,

[5] For each year I calculate the sample wage variance, $\hat{\sigma}_t^2 = \frac{1}{N}\sum_i (y_{it} - \bar{y}_t)^2$, and I take $(y_{it} - \bar{y}_t)/\hat{\sigma}_t$.

$\hat{\alpha}_{GMM-SYSTEM} = 0.4140$). Keep in mind that in a model with time heteroskedasticity (Alvarez and Arellano (2004)) GMM estimates were very small.

Table 7. α **and** β **estimates**

Estimator of $(\alpha, \beta)'$	$\widehat{\alpha}$	$\widehat{\beta}$
MLE	0.4822	0.4832
	(0.0114)	(0.0541)
Trimming $(r = 2)$	0.5370	0.5546
	(0.0397)	(0.0915)

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

In table 7, as expected, we can see that the MLE is underestimating the value of α and β. 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 true state dependence effects in the volatiliy of wages seem important.

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 maximixation 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.

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 asses 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 1,363 and 17,729 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.

The results are reported in Table 9. We can see that the significant ARCH effects in the variance disappears as soon as we consider a sample without job changes.

Table 9. 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	0.0758
	(0.0361)	(0.0592)

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

6 Conclusions

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

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

The empirical analysis is conducted on data drawn from the 1968-1993 PSID dataset. Estimates of different specifications for the wage model point to the importance of taking unobserved heterogeneity into account. In line with previous literature, we find a corrected estimate for the autorregresive coefficient in the mean around 0.5, and positive ARCH effects for the variance. However the latter disappear when there are not job changes in the sample.

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 (2005), 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

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

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

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

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

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

and

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

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

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

Substituting (A.1) we get

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

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, we expect that

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

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

Moreover, in the likelihood context, we can consider a local version of the estimated bias constructed as an expansion of $\widehat{b}_i(\Gamma)$ at Γ_0 using that at the truth $H_i^{-1}(\Gamma_0)\, \Upsilon_i(\Gamma_0) = 1$ (Pace and Salvan 2005). If we consider $\widehat{b}_i(\Gamma) = \frac{1}{2} tr\left(\widehat{H}_i^{-1}(\Gamma)\, \widehat{\Upsilon}_i(\Gamma)\right)$ 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

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

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

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

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

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

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

$$\begin{aligned}\ell_{it}(\alpha,\eta)&=C-\frac{1}{2}\left(y_{it}-\alpha y_{it-1}-\eta_i\right)^2,\\ \frac{\partial\ell_{it}(\alpha,\eta)}{\partial\eta}&=y_{it}-\alpha y_{it-1}-\eta_i\equiv v_{it}(\alpha,\eta)\equiv v_{it},\end{aligned}$$

and

$$\Upsilon_i(\alpha,\eta)=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\partial\ell_{it}}{\partial\eta}\cdot\frac{\partial\ell_{is}}{\partial\eta'}=T\overline{v}_i^2,$$

where $\bar{v}_i=\frac{1}{T}\sum_{t=1}^{T}v_{it}$, and

$$\bar{\Upsilon}_i(\alpha,\eta;\alpha_0,\eta_0)=TE\left(\overline{v}_i^2|y_{i0}\right).$$

Note that

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

$$\begin{aligned}E_0\left(\bar{v}_i|y_{i0},\eta_{i0}\right)&=(\alpha_0-\alpha)\,E_0\left(\bar{y}_{i(-1)}|y_{i0}\right)+(\eta_{i0}-\eta_i),\\ Var_0\left(\bar{v}_i|y_{i0},\eta_{i0}\right)&=\frac{1}{T}+(\alpha_0-\alpha)^2\,Var_0\left(\bar{y}_{i(-1)}|y_{i0}\right)+2(\alpha_0-\alpha)\,Cov_0\left(\bar{y}_{i(-1)},\bar{\epsilon}_i|y_{i0}\right),\end{aligned}$$

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

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

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

$$
\begin{aligned}
E_0\left(\bar{y}_{i(-1)}|y_{i0}\right) &= h_T(\alpha_0)\,\eta_{i0} + c_T(\alpha_0)\,y_{i0}, \\
Var_0\left(\bar{y}_{i(-1)}|y_{i0}\right) &= \frac{1}{T^2}\left[1 + (1+\alpha_0)^2 + \left(1+\alpha_0+\alpha_0^2\right)^2 + \ldots + \left(1+\alpha_0+\ldots+\alpha_0^{T-2}\right)^2\right] \equiv \omega_T(\alpha_0), \\
Cov_0\left(\bar{y}_{i(-1)}, \bar{\epsilon}_i|y_{i0}\right) &= \frac{1}{T^2}\left[\left(1+\alpha_0+\ldots+\alpha_0^{T-2}\right) + \left(1+\alpha_0+\ldots+\alpha_0^{T-3}\right) + \ldots + 1\right] \equiv \psi_T(\alpha_0),
\end{aligned}
$$

$$
\begin{aligned}
E\left(\overline{v}_i^2|y_{i0}\right) &= Var_0\left(\bar{v}_i|y_{i0}\right) + E_0^2\left(\bar{v}_i|y_{i0},\eta_{i0}\right) \\
&= \frac{1}{T} + (\alpha_0-\alpha)^2\,\omega_T(\alpha_0) + 2(\alpha_0-\alpha)\,\psi_T(\alpha_0) \\
&\quad + \left[(\alpha_0-\alpha)\,E_0\left(\bar{y}_{(-1)}|y_{i0}\right) + (\eta_{i0}-\eta_i)\right]^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\bar{\Upsilon}_i(\alpha,\eta;\alpha_0,\eta_0) &= 1 + T(\alpha_0-\alpha)^2\,\omega_T(\alpha_0) + 2T(\alpha_0-\alpha)\,\psi_T(\alpha_0) \\
&\quad + T\left\{(\alpha_0-\alpha)\left[h_T(\alpha_0)\,\eta_{i0} + c_T(\alpha_0)\,y_{i0}\right] + (\eta_{i0}-\eta_i)\right\}^2.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\bar{\Upsilon}_i\left(\alpha,\widehat{\eta}_i(\alpha);\hat{\alpha},\hat{\eta}_i\right) &= 1 + T(\hat{\alpha}-\alpha)^2\,\omega_T(\hat{\alpha}) + 2T(\hat{\alpha}-\alpha)\,\psi_T(\hat{\alpha}) \\
&\quad + T\left\{(\hat{\alpha}-\alpha)\left[h_T(\hat{\alpha})\,\hat{\eta}_i + c_T(\hat{\alpha})\,y_{i0}\right] + (\hat{\eta}_i-\eta_i(\alpha))\right\}^2.
\end{aligned}
$$

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	380	1981	721
1969	427	1982	777
1970	461	1983	817
1971	491	1984	861
1972	524	1985	924
1973	559	1986	896
1974	593	1987	868
1975	645	1988	839
1976	659	1989	810
1977	643	1990	768
1978	638	1991	736
1979	652	1992	698
1980	688	1993	654

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	266	80	134	52
10	185	43	103	39
11	152	33	86	33
12	149	33	87	29
13	133	45	69	19
14	100	29	58	13
15	86	28	42	16
16	63	17	34	12
17	56	13	32	11
18	20	6	9	5
19	44	10	23	11
20	22	4	15	3
21	20	9	9	2
22	20	4	15	1
23	14	4	7	3
24	7	2	3	2
25	16	4	10	2
26	10	3	5	2

Table D2.3. Descriptive Statistics

	1968	1980	1993
Age	38.09	39.41	42.59
	(6.33)	(9.26)	(5.65)
HS Dropout	0.43	0.31	0.13
HS Graduate	0.40	0.51	0.62
Hours	2256	2148	2129
	(517)	(483)	(521)
Married	0.83	0.84	0.86
White	0.70	0.66	0.67
Children	2.87	1.39	1.37
	(2.07)	(1.28)	(1.28)
Family Size	5.02	3.65	3.60
	(1.99)	(1.63)	(1.47)
North-East	0.18	0.16	0.16
North-Central	0.29	0.27	0.23
South	0.37	0.45	0.45
SMSA	0.69	0.65	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 (2005), "Understanding bias in nonlinear panel models: Some recent developments", CEMFI Working Paper No. 0507.

[7] 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.

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

[9] Carro, J. (2004), "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects", unpublished manuscript.

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

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

[12] 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.

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

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

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

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

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

[18] 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.

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

[20] 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.

[21] 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.

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

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

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

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

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

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

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

FIGURES

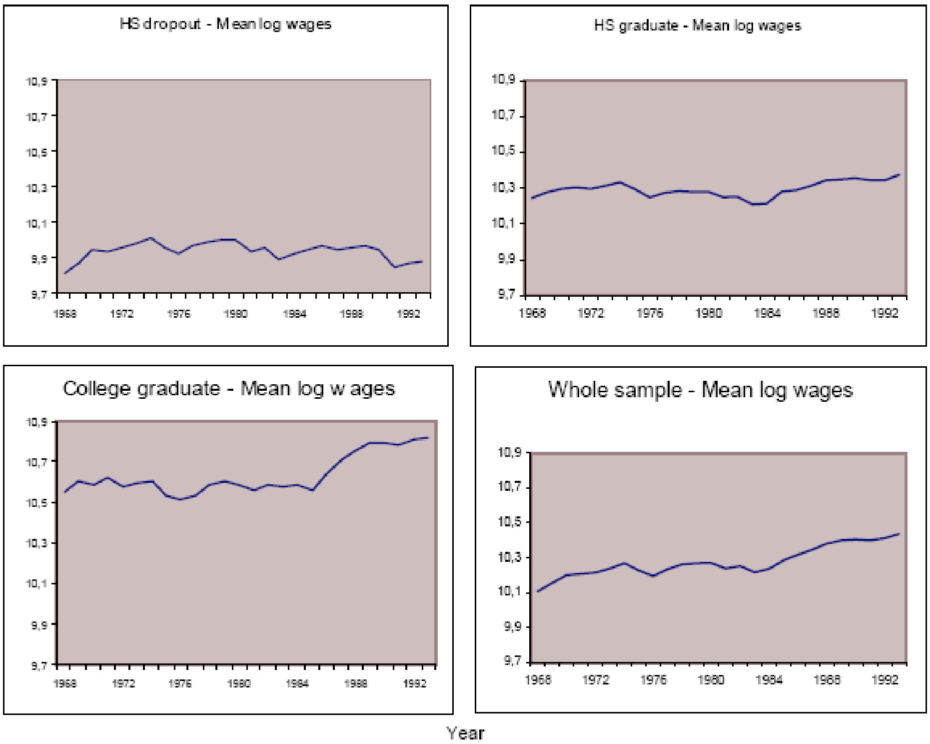

Figure 1. The mean of log wages

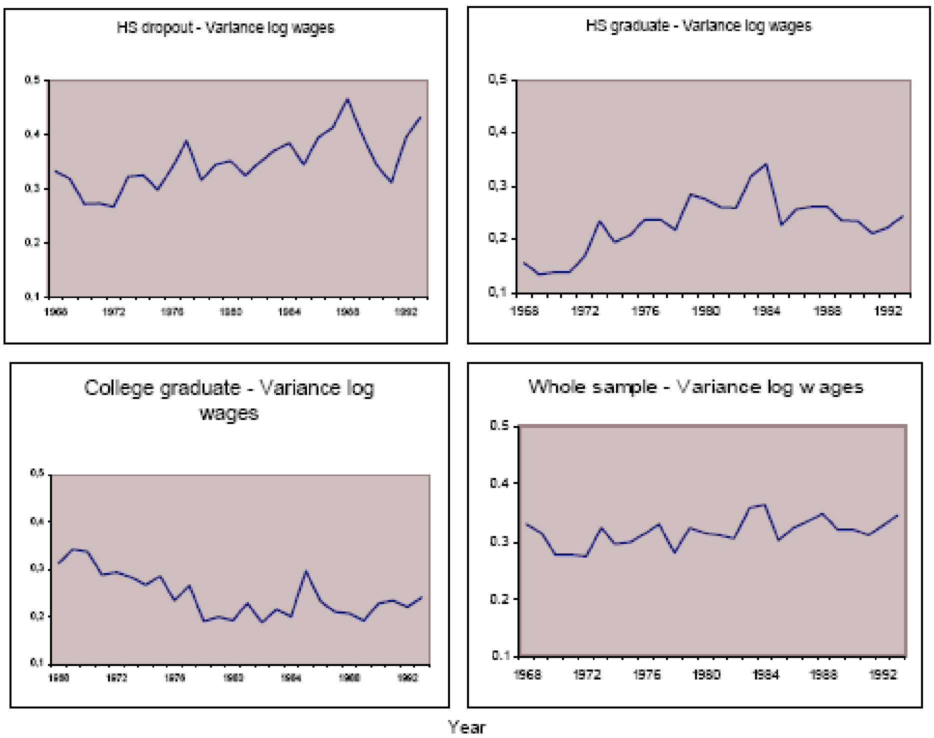

Figure 2. The variance of log wages