Estimation of Dynamic Panel Data Models using Particle Filters

Wen Xu*
Department of Economics & Oxford-Man Institute
University of Oxford

PRELIMINARY, COMMENTS WELCOME

June 14, 2014

Abstract

In this paper we study dynamic panel data models with stochastic volatility in a macroeconomic context. The models are represented in state space forms and estimated using particle filter techniques in both a frequentist framework (maximization of approximate simulated-likelihood) and in a Bayesian framework (particle Metropolis-Hastings sampler). A two-step LSDV-QML estimator is also proposed. Ignoring the existence of stochastic volatility might induce systematically false rejection in panel unit root tests. Monte Carlo studies show that particle-filter based estimators are more precise than other estimators on average in finite samples in the presence of stochastic volatility and even in the case of homoscedasticity especially when T is large. Our methodology is straightforwardly applied to panel VAR models.

Keywords: Dynamic Panel Data Models, Stochastic Volatility, Particle Filters, State Space Modeling, Least Squares Dummy Variable, Quasi-Maximum Likelihood, Panel Unit Root Tests

1 Introduction

It has been well documented that there is time-varying volatility in macroeconomic time series data. Many studies focused on the moderated volatility in the growth rate of U.S. real GDP. For example, Kim and Nelson (1999) and McConnell and Perez-Quiros (2000) independently identified a structural decline in the volatility in the first quarter of 1984. Blanchard and Simon (2001) argued however that the substantial reduction in volatility commenced

*This version is preliminary. Email: wen.xu@economics.ox.ac.uk. We thank Neil Shephard, Kevin Sheppard and Arnaud Doucet for helpful advice.

in 1950s, was interrupted during 1970s and reverted to the long-run trend in late 1980s. They concluded that the moderation is possibly caused by the decrease in the volatility of consumption, investment and government spending and there exists a strong relationship between the trend in output volatility and inflation volatility. Weiss (1984) analysed 16 U.S. macroeconomic time series and showed the evidence of conditional heteroscedasticity. Stock and Watson (2003) found the decline in volatility was common among many U.S. macroeconomic time series. One of their conclusions is the moderation is associated more with a decrease in the magnitude of unforecastable disturbances than with the propagation mechanism of those disturbances. Fernández-Villaverde and Rubio-Ramírez (2013) provided an updated documentation of the great moderation in the U.S. economy. Besides, they showed the presence of time-varying volatility of the Emerging Markets Bond Index+ spread reported by J.P. Morgan. Blanchard and Simon (2001) and Stock and Watson (2003) found the time variation in volatility also happened in other developed countries.

Traditionally homoscedasticity is assumed for the innovations of macroeconomic time series. Recently, motivated by the studies mentioned above, several papers relaxed the assumption and incorporated time-varying volatility in the models. Sims and Zha (2006) studied a structural vector autoregression with regime switching. They found that allowing time variation across regimes in the variance of the disturbances only rather than also in coefficients would fit the data best. The best fit however cannot account for the movement of U.S. inflation of the 1970s and 1980s. Fernández-Villaverde and Rubio-Ramírez (2007) and Justiniano and Primiceri (2008) both estimated DSGE models with stochastic volatility on the structural shocks. The former used particle filters for estimation to study the effect of nonlinearity, while the latter estimated the linearized models using Bayesian Markov chain Monte Carlo methods (MCMC). Koop and Korobilis (2010) discussed the time-varying parameter vector autoregression (VAR) models with multivariate stochastic volatility using MCMC. These papers all support the necessity for considering time-varying volatility in macroeconomic modelling. Hamilton (2010) argued that time-varying volatility should be considered even when the conditional mean is the direct object of interest. One reason is that hypothesis tests on the mean might be invalid if the dynamic of variance is misspecified. Another is that statistical efficiency gains can be obtained by incorporating appropriate features of time-varying volatility into the estimation of the conditional mean.

Dynamic panel data models have become increasingly popular in macroeconomics to study common relationships across countries or regions, such as growth convergence (e.g. Islam, 1995), purchasing power parity (e.g. Frankel and Rose, 1996) and mean reversion of interest rates (Wu and Chen, 2001). Compared to the microeconomic panel, the time dimension T is relatively large ($>$20) and the cross-sectional dimension N is relatively small ($<$100) in a typical macroeconomic panel dataset. Judson and Owen (1999) compared the finite sample performance of different estimation techniques including the least squares dummy variable (LSDV) and generalized method of moments (GMM) in a macro panel setting. Panel unit root tests have been often used in macro panels since Levin and Lin (1992) as nonstationarity deserves more attention when T grows large.

In this paper, we study the estimation and inference of dynamic panel data models with stochastic volatility in a macroeconomic context. We propose a class of parameter estimators using particle filters (also known as sequential Monte Carlo methods) which are simulation-based filtering techniques to estimate the posterior density of the state-space for nonlinear and non-Gaussian state space models (see Creal 2012 for a survey of the methods for economic applications). There are various particle filtering algorithms including the bootstrap filter (Gordon et al., 1993), auxiliary particle filters (Pitt and Shephard, 1999), mixed Kalman filters (Chen and Liu, 2000) among others. Different algorithms differ mainly in the choices of incremental importance distributions and resampling algorithms which are aimed to improve the level of statistical efficiency in terms of Monte Carlo variation. After estimating the posterior density, the likelihoods can be computed for the models where there is no analytical solution. The simulated likelihood can be used in parameter estimation either in a frequentist way (e.g. maximization of the log-likelihood) or in a Bayesian framework such as particle MCMC (Andrieu et al., 2010). We also propose a two-step estimator LSDV-QML (Quasi-Maximum Likelihood) and investigate its asymptotic properties. We find that asymptotic variance of the LSDV estimators can be much larger when the stochastic volatility exists and naively ignoring stochastic volatility would lead to systematically false rejection in Harris and Tzavalis (1999)'s panel unit root tests.

The paper is organized as follows. Section 2 introduces dynamic panel data models with stochastic volatility. Section 3 discusses the particle-filter based estimators. Section 4 proposes the two-step LSDV-QML estimator and investigates its asymptotic properties. Section 5 presents the Monte Carlo simulation to study the finite sample properties of our estimators. Section 6 concludes the paper.

2 Model Specifications

In general, we study a linear dynamic panel data model with the following specification

$$y_{it} = B(L)y_{i,t-1} + \gamma x_{it} + \mu_i + \varepsilon_{it} \tag{1}$$

where $i = 1, ..., N$ and $t = 1, ..., T$; $B(L)$ denotes a polynomial of the lag operator; x_{it} is a $K \times 1$ vector of additional regressors and γ is the associated coefficient vector; μ_i is the individual effect; ε_{it} is the disturbance specified as martingale difference sequences, i.e. $E(\varepsilon_{it}|\mathcal{Y}_{i,t-1}) = 0$ in which $\mathcal{Y}_{i,t-1}$ is the information set including all observations of y_{it} up to $t-1$.

One popular way of parameterizing martingale differences is through stochastic volatility, which assumes two error processes so it is more flexible than GARCH to model conditional heteroscedasticity. The standard specification is given by

$$\varepsilon_{it} = \sigma_{it}\epsilon_{it}$$

$$\log(\sigma_{it}^2) = \mu + \phi(\log(\sigma_{i,t-1}^2) - \mu) + \eta_{it}$$

where $\log(\sigma_{it}^2)$ is log volatility whose evolution follows an AR(1) process with mean vector μ and coefficient ϕ; ϵ_{it} is independent of σ_{it}^2; ϵ_{it} and η_{it} follow a joint normal distribution

$$\begin{pmatrix} \epsilon_{it} \\ \eta_{it} \end{pmatrix} \sim N(0, \begin{bmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \theta^2 \end{bmatrix}).$$

in which σ_ϵ^2 is set to 1 for indentifiability reasons. In this setting, the conditional variance of disturbances ε_{it} is σ_{it}^2 which is time varying. Various extensions of the basic specification can be made. For example, it is possible to incorporate dependence between ϵ_{it} and η_{it}, which is called the leverage effect suggested by the evidence of stock returns; it is also possible to model error terms using a fat-tailed distribution. When $|\phi|$ is less than 1, the log-volatility process is strictly stationary given σ_{i1}^2, which implies stationarity of ε_{it}.

We consider the fixed effect specifications under which no restrictions are imposed on the data generating process of μ_i. Many possible specifications can be allowed such as cross-sectional dependence, heteroskedasticity and correlation between individual effects and disturbances.

The focus of the paper is on estimation of the unknown coefficients in $B(L)$ and parameters in the distribution of disturbances. We consider AR(1) dynamic panel models with no other regressors as the benchmark specifications, although we will briefly investigate the finite sample performance of our estimators in the models with an additional regressor in Section 5. Specifically, we discuss

$$y_{it} = \beta y_{i,t-1} + \mu_i + \varepsilon_{it} \tag{2}$$

As for the existence of stochastic volatility, the tractable expressions for exact likelihood functions are not known. State space models are useful means in time series analysis to study dynamics of the system through latent variables named state variables (Durbin and Koopman, 2012). It consists of an observation equation and a transition equation. It is commonly assumed that the series of state variables is a Markov chain and observation variables are conditionally independent given the state variables. The first difference of the equation (2) can be put into state space forms. One form, which is useful in studying statistical properties of the model, is written as

$$\begin{aligned} \Delta y_{it} &= \begin{bmatrix} 1 & \epsilon_{it} \end{bmatrix} \alpha_{it} \\ \alpha_{it} &= \begin{bmatrix} \beta\alpha_{i,t-1,1} + (\beta - 1)\epsilon_{i,t-1}\alpha_{i,t-1,2} \\ e^{\frac{\mu(1-\phi)+\eta_t}{2}} \alpha_{i,t-1,2}^{\phi} \end{bmatrix} \end{aligned}$$

where the state vector $\alpha_{it} = \begin{bmatrix}\alpha_{it,1}\\ \alpha_{it,2}\end{bmatrix} = \begin{bmatrix}\beta\Delta y_{i,t-1} - \sigma_{i,t-1}\epsilon_{i,t-1}\\ \sigma_{it}\end{bmatrix}$. The conditional density of the observation variable and the transition density are respectively given by

$$g(\Delta y_{it}|\alpha_{it}) = \frac{1}{\sqrt{2\pi}\alpha_{it,2}} e^{-\frac{(\Delta y_{it}-\alpha_{it,1})^2}{2\alpha_{it,2}^2}}$$

$$q(\alpha_{it}|\alpha_{i,t-1}) = \frac{1}{\sqrt{2\pi}\alpha_{it,2}\theta/2} e^{-\frac{2(\log(\alpha_{it,2}/\alpha_{i,t-1,2}^{\phi}) - \frac{\mu(1-\phi)}{2})^2}{\theta^2}} \frac{1}{\sqrt{2\pi}(\beta-1)\alpha_{i,t-1,2}} e^{-\frac{(\alpha_{it,1}-\beta\alpha_{i,t-1,1})^2}{2(\beta-1)^2\alpha_{i,t-1,2}^2}}$$

Another form is

$$\Delta y_{it} = \begin{bmatrix}1 & 1 & 0\end{bmatrix} \alpha_{it} \tag{3}$$

$$\alpha_{it} = \begin{bmatrix}\beta & \beta & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix} \alpha_{i,t-1} + \begin{bmatrix}0\\ 1\\ -1\end{bmatrix} \sigma_{it}\epsilon_{it} \tag{4}$$

where the state vector $\alpha_{it} = \begin{bmatrix}\beta\Delta y_{i,t-1}\\ \sigma_{it}\epsilon_{it} - \sigma_{i,t-1}\epsilon_{i,t-1}\\ -\sigma_{it}\epsilon_{it}\end{bmatrix}$. We will use the latter state space form throughout the remaining paper for the reason explained in the next section.

Although it is not the focus of this paper, our methodology is straightforwardly applied to panel VAR models which is

$$y_{it} = B(L)y_{i,t-1} + \mu_i + \varepsilon_{it} \tag{5}$$

where y_{it} is a $\tilde{K} \times 1$ vector of variables of interest; $B(L)$ denotes a $\tilde{K} \times \tilde{K}$ matrix of polynomials of the lag operator; μ_i is a $\tilde{K} \times 1$ vector of the individual effects; ε_{it} is a $\tilde{K} \times 1$ vector of the disturbances specified as martingale difference sequences. In the multivariate case, there are several ways stochastic volatility can be modelled. The basic specification (Harvey et al., 1994) is given by

$$\varepsilon_{it} = V_{it}^{1/2}\epsilon_{it}$$

$$h_{it} = \mu_i + \Phi_i(h_{i,t-1} - \mu_i) + \eta_{it}$$

where $V_{it}^{1/2} = \text{diag}(\sigma_{it,1}, \ldots, \sigma_{it,\tilde{K}})$ and $h_{it} = (\log(\sigma_{it,1}^2), \ldots, \log(\sigma_{it,\tilde{K}}^2))$ is a vector of the series-specific log volatilities whose evolution follows a first order stationary VAR with the mean vector μ_i and the coefficient matrix Φ_i; ϵ_{it} and η_{it} follow a joint normal distribution

$$\begin{pmatrix}\epsilon_{it}\\ \eta_{it}\end{pmatrix} \sim N(0, \begin{bmatrix}\Sigma_\epsilon & 0\\ 0 & \Sigma_\eta\end{bmatrix}).$$

in which the diagonal elements of Σ_ϵ are set to 1 for indentifiability reasons. In this setting, the conditional covariance matrix of disturbances ε_{it} is $V_{it}^{1/2}\Sigma_\epsilon V_{it}^{1/2}$ which is time varying, while the conditional correlation matrix is Σ_ϵ which is constant. To simplify the model, the log volatilities are often assumed to become conditionally independent, i.e. Φ_i and Σ_η are diagonal matrices.

A panel VAR(1) model is given by

$$\mathbf{y}_{it} = B\mathbf{y}_{i,t-1} + \mu_i + \varepsilon_{it}$$

For instance, $\tilde{K} = 2$. Let $\epsilon_{it} = (\epsilon_{it,1}, \epsilon_{it,2})'$. Similar to single equation panel data models, one state-space form is given by

$$\Delta\mathbf{y}_{it} = \begin{bmatrix} 1 & 0 & \epsilon_{it,1} \\ 0 & 1 & \epsilon_{it,2} \end{bmatrix} \alpha_{it}$$

$$\alpha_{it} = \begin{bmatrix} \beta_{11}\alpha_{i,t-1,1} + \beta_{12}\alpha_{i,t-1,2} + ((\beta_{11}-1)\epsilon_{i,t-1,1} + \beta_{12}\epsilon_{i,t-1,2})\alpha_{i,t-1,3} \\ \beta_{21}\alpha_{i,t-1,1} + \beta_{22}\alpha_{i,t-1,2} + (\beta_{21}\epsilon_{i,t-1,1} + (\beta_{22}-1)\epsilon_{i,t-1,2})\alpha_{i,t-1,3} \\ e^{\frac{\mu(1-\phi)+\eta_t}{2}}\alpha_{i,t-1,3}^{\phi} \end{bmatrix}$$

where the state vector $B = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$, $\alpha_{it} = \begin{bmatrix} \alpha_{it,1} \\ \alpha_{it,2} \\ \alpha_{it,3} \end{bmatrix} = \begin{bmatrix} \beta_{11}\Delta y_{i,t-1,1} + \beta_{12}\Delta y_{i,t-1,2} - \sigma_{i,t-1}\epsilon_{i,t-1,1} \\ \beta_{21}\Delta y_{i,t-1,1} + \beta_{22}\Delta y_{i,t-1,2} - \sigma_{i,t-1}\epsilon_{i,t-1,2} \\ \sigma_{it} \end{bmatrix}$.

Another form is

$$\Delta\mathbf{y}_{it} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \alpha_{it} \tag{6}$$

$$\alpha_{it} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{11} & 0 & \beta_{12} & 0 \\ \beta_{21} & \beta_{22} & \beta_{21} & 0 & \beta_{22} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \alpha_{i,t-1} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \sigma_{it} \begin{bmatrix} \epsilon_{it,1} \\ \epsilon_{it,2} \end{bmatrix} \tag{7}$$

where the state vector $\alpha_{it} = \begin{bmatrix} \beta_{11}\Delta y_{i,t-1,1} + \beta_{12}\Delta y_{i,t-1,2} \\ \beta_{21}\Delta y_{i,t-1,1} + \beta_{22}\Delta y_{i,t-1,2} \\ \sigma_{it}\epsilon_{it,1} - \sigma_{it-1}\epsilon_{i,t-1,1} \\ -\sigma_{it}\epsilon_{it,1} \\ \sigma_{it}\epsilon_{it,2} - \sigma_{it-1}\epsilon_{i,t-1,2} \\ -\sigma_{it}\epsilon_{it,2} \end{bmatrix}$.

It is straightforward to extend the state space forms to the model with an arbitrary $\tilde{K}$.

3 Particle-Filter-Based Estimation and Inference

In this section we estimate the panel models using particle-filter-based methods. Particle filters can be regarded as the extension of Kalman filters to address non-linear and non-Gaussian state space models. It is a simulation-based technique to obtain filtered and smoothed estimates of the states as well as the unbiased estimate of the likelihood given some regularity conditions. In nonlinear and non-Gaussian state space models, posterior density of state variables seldom has the closed-form expression so it is approximated by a discrete distribution made of weighted draws called particles. Particle filters can be implemented in several ways, varying with choices of incremental importance distributions and resampling algorithms leading to different levels of statistical efficiency. In this paper, the special structure of our state space models allows for the use of mixture Kalman filters (Chen and Liu, 2000) also called Rao–Blackwellization. Specifically, a class of state space models suitable for mixture Kalman filters can be written as

$$y_t = Z(\alpha_{t,1})\alpha_{t,2} + \varepsilon_t$$

$$\alpha_{t,2} = T(\alpha_{t,1})\alpha_{t-1,2} + \eta_t$$

where $\varepsilon_t \sim N(0, H(\alpha_{t,1}))$, $\eta_t \sim N(0, Q(\alpha_{t,1}))$ and $\alpha_{t,1}$ follows a first order Markov process; the state vector $\alpha_t = (\alpha'_{t,1}, \alpha'_{t,2})'$. The parameter matrices Z, T, H and Q depend on part of the state vector $\alpha_{t,1}$. The structure is special in that it is a linear normal state space model conditional on $\alpha_{t,1}$. Mixture Kalman filters integrate out the subset of the state variable $\alpha_{t,2}$ in order to reduce the Monte Carlo variation of the simulation-based estimators and improve the statistical efficiency. The state space forms of panel data models can be written as this structure such as (3)-(4) or (6)-(7) in which σ_{it} and α_{it} correspond to $\alpha_{t,1}$ and $\alpha_{t,2}$, respectively. Next we will show the specific algorithm of mixture Kalman filters. For simplicity we restrict our exposition to the model (2), while it is straightforward to adapt it for other specifications.

3.1 Mixture Kalman Filters

As the model is a linear normal state space model conditional on part of the state variables, the resulting system can be addressed by Kalman filters. The particle state variable in our model is $\sigma_{it}^{2(j)}$ where the superscript j indexes the particle. Let $m_{it|t-1}^{(j)} = \mathrm{E}(\alpha_{it}|\mathcal{Y}_{i,t-1}, \sigma_{it}^{2(j)})$ and $\Sigma_{it|t-1}^{(j)} = \mathrm{Var}(\alpha_{it}|\mathcal{Y}_{i,t-1}, \sigma_{it}^{2(j)})$ denote prediction mean and variance for particle j which can be recursively derived in Kalman filters. At the end of each iteration over time, the algorithm produces M simulated state variables and the corresponding weight $\{\sigma_{it}^{2(j)}, m_{it|t-1}^{(j)}, \Sigma_{it|t-1}^{(j)}, \hat{w}_{it|t-1}^{(j)}\}_{j=1}^{M}$.

Given the values of parameters β, μ, ϕ and θ, the algorithm of mixture Kalman filters is as follows.

For $i = 1, ..., N$,

1) Set the starting values. For the particle $j = 1, ..., M$, draw

$$\log(\sigma_{i2}^{2(j)}) \sim N(\log(\mu), \frac{\theta^2}{1-\phi^2})$$

and set the

$$m_{i2|1}^{(j)} = \begin{bmatrix} \beta \Delta y_{i1} \\ 0 \\ 0 \end{bmatrix}$$

$$\Sigma_{i2|1}^{(j)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2\sigma_{i2}^{2(j)} & -\sigma_{i2}^{2(j)} \\ 0 & -\sigma_{i2}^{2(j)} & \sigma_{i2}^{2(j)} \end{bmatrix}$$

It should be noted that because of the special structure, only the entry in the 2rd row of $m_{i2|1}^{(j)}$ matters for inference. Similarly, only the entry in the 2rd row and 2rd column of $\Sigma_{i2|1}^{(j)}$ matters.

The log importance weight $w_{i2|1}^{(j)} = 0$ and the normalized importance weight $\hat{w}_{i2|1}^{(j)} = \frac{1}{M}$.

For $t = 2, ..., T$,
2) Draw

$$\log(\sigma_{i,t+1}^{2(j)}) = (1-\phi)\log(\mu) + \phi\log(\sigma_{it}^{2(j)}) + \theta\eta_{i,t+1}^{(j)}$$

for each $j = 1, ..., M$ where $\eta_{i,t+1}^{(j)}$ follows the standard normal distribution.
3) Compute the conditional likelihood for each particle j. That is,

$$l_{it}^{(j)} = -0.5\log|V_{it}^{(j)}| - 0.5v_{it}^{(j)}(V_{it}^{(j)})^{-1}v_{it}^{(j)}$$

where the forecast error $v_{it}^{(j)} = \Delta y_{it} - \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} m_{it|t-1}^{(j)}$ and forecast variance $V_{it}^{(j)} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \Sigma_{it|t-1}^{(j)} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

Update the $w_{i,t+1|t}^{(j)} = w_{it|t-1}^{(j)} + l_{it}^{(j)}$ and $\hat{w}_{i,t+1|t}^{(j)} = \frac{\exp(w_{i,t+1|t}^{(j)})}{\sum_{j=1}^{M}\exp(w_{i,t+1|t}^{(j)})}$.

4) Resample with replacement M particles $\sigma_{i,t+1}^{2(j)}$, $m_{it|t-1}^{(j)}$ and $\Sigma_{it|t-1}^{(j)}$ with the weight $\hat{w}_{i,t+1|t}^{(j)}$ every three increments[1]. After doing this, reset $w_{it|t-1}^{(j)} = 0$ and $\hat{w}_{it|t-1}^{(j)} = \frac{1}{M}$.

5) Update Kalman filter estimates

$$m_{i,t+1|t}^{(j)} = \begin{bmatrix} \beta & \beta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} m_{it|t-1}^{(j)} + K_{it}^{(j)} v_{it}^{(j)}$$

$$\Sigma_{i,t+1|t}^{(j)} = \begin{bmatrix} \beta & \beta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \Sigma_{it|t-1}^{(j)} L_{it}^{(j)\prime} + \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \sigma_{i,t+1}^{2(j)} \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}$$

[1]It is an ad-hoc choice for stability of the algorithm, see Shephard (2013).

where $K_{it}^{(j)} = \begin{bmatrix} \beta & \beta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \Sigma_{it|t-1}^{(j)} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} (V_{it}^{(j)})^{-1}$ and $L_{it}^{(j)} = \begin{bmatrix} \beta & \beta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} - K_{it}^{(j)} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$.

6) Go to 2).

The particle estimate of conditional log-likelihood is recorded

$$\log[\hat{L}(\Delta y_{it}|\mathcal{Y}_{i,t-1}, \beta, \mu, \phi, \theta)] = \log[\sum_{j=1}^{M} \hat{w}_{it|t-1}^{(j)} \exp(l_{it}^{(j)})]$$

so the simulation-based estimate of the joint log-likelihood is

$$\log[\hat{L}(\Delta \mathbf{y}_1, \dots, \Delta \mathbf{y}_N|\beta, \mu, \phi, \theta)] = \sum_{i=1}^{N} \sum_{t=1}^{T} \log[\sum_{j=1}^{M} \hat{w}_{it|t-1}^{(j)} \exp(l_{it}^{(j)})] \tag{8}$$

where $\Delta \mathbf{y}_i = (y_{i1}, \dots, y_{iT})$.

Resampling is used in the literature of particle filters to alleviate the weight degeneracy problem (Creal, 2012). Without resampling, as the series grows over time, one particle's normalized importance weight converges to one while the others converge to zero. In other words, the discrete distribution made of weighted draws would become degenerate. Resampling is a crucial means to make the algorithm stable by eliminating the particles which have low importance weights and multiplying the heavily weighted particles. The simplest resampling algorithm is multinomial resampling introduced in Gordon et al. (1993). It draws new particles $\{\tilde{\sigma}_{it}^{2(j)}, \tilde{m}_{it|t-1}^{(j)}, \tilde{\Sigma}_{it|t-1}^{(j)}\}_{j=1}^{M}$ from the point mass distribution $\{\sigma_{it}^{2(j)}, m_{it|t-1}^{(j)}, \Sigma_{it|t-1}^{(j)}, \hat{w}_{it|t-1}^{(j)}\}_{j=1}^{M}$. Specifically,

1) Draw M uniform numbers $\{U^{(j)}\}_{j=1}^{M}$on the interval $[0, 1]$.

2) Generate M multinomial variables $i_j = F^{-1}(U^{(j)})$ where the generalized inverse function $F^{-1}(u) = i$ if $\sum_{j=1}^{i-1} \hat{w}_{it|t-1}^{(j)} < u \leq \sum_{j=1}^{i} \hat{w}_{it|t-1}^{(j)}$.

3) New particles $\{\tilde{\sigma}_{it}^{2(j)}, \tilde{m}_{it|t-1}^{(j)}, \tilde{\Sigma}_{it|t-1}^{(j)}\} = \{\sigma_{it}^{2(i_j)}, m_{it|t-1}^{(i_j)}, \Sigma_{it|t-1}^{(i_j)}\}$ and corresponding log weights $\tilde{w}_{it|t-1}^{(j)} = 0$ for $j = 1, ..., M$.

Let $N_{it|t-1}^{(j)}$ denote the number of times $\{\sigma_{it}^{2(j)}, m_{it|t-1}^{(j)}, \Sigma_{it|t-1}^{(j)}\}$ is drawn with replacement. Multinomial resampling is unbiased in the sense that expected $N^{(j)}$ given all weights is proportional to its normalized weight, i.e. $\mathrm{E}(N_{it|t-1}^{(j)}|\hat{w}_{it|t-1}^{(1)}, \dots, \hat{w}_{it|t-1}^{(M)}) = M\hat{w}_{it|t-1}^{(j)}$.

Other resampling methods in the literature consist of stratified resampling (Kitagawa, 1996), residual resampling (Liu and Chen, 1998) and systematic resampling (Carpenter et al., 1999), all of which are unbiased algorithms but can be more efficient than multinomial resampling under some circumstances. We choose multinomial resampling because it appears to be a requirement for good asymptotic performance of the estimation method in Olsson and Rydén (2008) we will use later.

Next we will estimate parameters using the simulated estimate of likelihood (8) in both a frequentist and a Bayesian framework. Reviews on the existing methods for parameter estimation using particle filters can be found in Kantas et al. (2009) and Creal (2012).

3.2 Maximization of Approximate Simulated-Likelihood

Although (8) is an unbiased estimator of the exact likelihood under some regularity conditions (Del Moral, 2004), direct maximization of the simulated likelihood suffers from discontinuity induced from the generalized inverse operation at the resampling stage, which makes invalid the common gradient-based optimization methods. Pitt (2002) overcame the problem of non-smoothness by developing a new resampling method, but his method is valid only when the dimension of state space is one, which is too restricted in practical applications. Several papers performed maximum likelihood estimation via the Expectation Maximization algorithm, e.g. Olsson et al. (2008), which is numerically stable and computationally cheap but only guaranteed to be locally optimal. Olsson and Rydén (2008) approximated the likelihood by means of step functions or B-spline interpolation, and showed consistency and asymptotic normality of the estimators which maximize the approximate likelihood under some assumptions. This is the only work we are aware of which studies asymptotic properties of parameter estimators in the particle filter literature. One of their assumptions is a compact state space which is obviously not the case in our model. We will however still use their method for parameter estimation for two reasons. One is that the compactness assumption can be potentially released given new results of uniform convergence properties in time dimension (Douc et al. (2012), Whiteley (2011))[2], although the full proof of the extension is beyond the scope of this paper; the other reason is the good performance of Monte Carlo studies shown below.

Olsson and Rydén (2008) discretized the parameter space Ω by a grid $\bar{\Omega} \triangleq \{\omega_g\}_{g=1}^{G} \subseteq \Omega$. Let $[\omega]$ denote the closest point in the grid to $\omega \in \Omega$[3]. The grid-based particle approximation of the likelihood using piecewise constant functions is given as

$$\log[\hat{L}(\Delta\mathbf{y}_1, \ldots, \Delta\mathbf{y}_N|\omega)] \approx \log[\hat{L}(\Delta\mathbf{y}_1, \ldots, \Delta\mathbf{y}_N|[\omega])]$$

The approximation can also be made via spline interpolation, which is more efficient than piecewise constant functions but suffers from higher computation costs as the dimension of parameter space grows. Although this particle filter based method is pretty slower than other frequentist type of approaches such as OLS and GMM. it is much faster than some Bayesian methods like particle Metropolis-Hastings sampler described in the next section.

[2]Thanks Professor Arnaud Doucet for indicating this point in a personal correspondence.

[3]If there is more than one point having the smallest distance from ω, the point with lowest index g will be chosen.

3.3 Particle Metropolis-Hastings Sampler

Andrieu et al. (2010) combined particle filters with standard MCMC algorithms such as independent Metropolis–Hastings sampler, marginal Metropolis–Hastings sampler or Gibbs sampler. Specifically, they showed the asymptotic convergence and good performance of a MCMC algorithm when using an unbiased particle-based estimator of the likelihood. An attractive feature of particle MCMC methods is minimal tuning: we only need to design a proposal distribution for parameters. The disadvantage is the high computation cost with $O(NTM)$ operations per MCMC step in panel models. In the paper we use particle independent Metropolis-Hastings sampler to estimate parameters. A random walk proposal is used for $\log(\beta)$, $\log(\mu)$, $\log(\phi)$ and $\log(\theta)$ with standard deviations varing with N and T in order to achieve the best acceptance probability. Let β^*, μ^*, ϕ^* and θ^* denote the candidate draws; β , μ , ϕ and θ denote the existing values. Using the noninformative prior $f(\beta,\mu,\phi,\theta)=1$, the probability of accepting β^* , μ^* , ϕ^* and θ^* can be written as

$$\min\left\{1, \frac{\hat{L}(\Delta\mathbf{y}_1,\ldots,\Delta\mathbf{y}_N|\beta^*,\mu^*,\phi^*,\theta^*)\beta^*\mu^*\phi^*\theta^*}{\hat{L}(\Delta\mathbf{y}_1,\ldots,\Delta\mathbf{y}_N|\beta,\mu,\phi,\theta)\beta\mu\phi\theta}\right\}.$$

4 LSDV-QML Estimation and Inference

In this section, we propose a consistent two-step estimator which is much faster than particle-filter based estimators by virtue of efficiency losses. In the first step, we estimate the AR coefficients using LSDV by ignoring the existence of stochastic volatility. In the second step, the parameters in stochastic volatility are estimated using QML with the residuals obtained in the first step estimation.

LSDV eliminates the fixed effect by within transformation $\tilde{y}_{it}=y_{it}-\bar{y}_i$ where $\bar{y}_i=\sum_{t=1}^T y_{it}$. The estimate of β is then obtained using OLS on the transformed equation $\tilde{y}_{it}=\beta\tilde{y}_{i,t-1}+\tilde{\varepsilon}_{it}$. It can be written as

$$\hat{\beta}_{LSDV}=\frac{\sum_{i=1}^N \mathbf{y}_{i,(-1)}Q\mathbf{y}_i}{\sum_{i=1}^N \mathbf{y}_{i,(-1)}Q\mathbf{y}_{i,(-1)}} \tag{9}$$

where $Q=I-\iota\iota'/T$ and ι is a T-dimension vector of ones. The transformation induces endogeneity due to the correlation between $\tilde{y}_{it-1}$ and $\tilde{\varepsilon}_{it}$. As a result, finite sample biases are non-negligible for a fixed T (Nickell, 1981).

Now we explore the asymptotic properties of the first-step estimator when the series is stationary or has a unit root. Basically we will extend the theorems in Harris and Tzavalis (1999) and Alvarez and Arellano (2003).

Assumptions

1. $|\beta|<1$.

2. The initial values of y_{it} follow the steady state distribution $y_{i0}=\frac{\mu_i}{1-\beta}+\sum_{t=0}^{\infty}\beta^t\varepsilon_{i,-t}$.

3. The stochastic volatility is stationary:$|\phi|<1$.

4. The initial values of $\log(\sigma_{it}^2)$ follow the steady state distribution $\log(\sigma_{i0}^2) = \mu + \sum_{t=0}^{\infty}\beta^t\eta_{i,-t}$.

5. $\{\epsilon_{it}\}$ $(i = 1, \ldots, N;\ t = 1, \ldots T)$ are i.i.d. variables across both time and individuals with $\mathrm{E}(\epsilon_{it}) = 0$ and $\mathrm{Var}(\epsilon_{it}) = 1$ and $\mathrm{E}(\epsilon_{it}^4) = \kappa < \infty$ and independent of μ_i and η_{it} for all i and t.

6. $\{\eta_{it}\}$ $(i = 1, \ldots, N;\ t = 1, \ldots T)$ are i.i.d. variables normally distributed across both time and individuals with $\mathrm{E}(\eta_{it}) = 0$ and $\mathrm{Var}(\eta_{it}) = \theta^2 < \infty$.

Theorem 1. *(LSDV Estimators: Stationary)*
Under Assumptions 1-6, as $N \to \infty$ and $T \to \infty$,

$$\sqrt{NT}[\hat{\beta}_{LSDV} - (\beta - \frac{1}{T}(1+\beta))] \xrightarrow{d} N(0, B(\beta, \phi, \theta))$$

where $B(\beta, \phi, \theta) = (1-\beta^2)^2\sum_{t=1}^{\infty}[\exp(\frac{\theta^2\phi^t}{1-\phi^2})\beta^{2t-2}]$.

If $\phi > 0$ and $\theta > 0$, $B(\beta, \phi, \theta) > B(\beta, 0, 0)$ which is equal to $1-\beta^2$ same as in Alvarez and Arellano (2003). Stochastic volatility changes asymptotic variances of the estimator while keeps unchanged asymptotic means. When $\beta = 0.7$, for example, asymptotic variance is approximately 0.51 for the model without stochastic volatility, but can be 1.52 when $\phi = 0.9$ and $\theta = 0.5$.

Assumptions

7. The data generation process has a unit root: $\beta = 1$.

8. The initial values y_{i0} are fixed.

Theorem 2. *(LSDV Estimators: Unit Root)*
Under Assumptions 3-8, as $N \to \infty$ and T is fixed,

$$\sqrt{N}(\hat{\beta}_{LSDV} - (1 - \frac{3}{T+1})) \xrightarrow{d} N(0, B_u(\phi, \theta))$$

where $B_u(\phi, \theta) = \frac{36\left(2-5T+2T^2\right)}{5(-1+T)T(1+T)^3}\exp(\frac{\theta^2}{1-\phi^2})\kappa + \sum_{t=1}^{T-1}[\exp(\frac{\theta^2\phi^t}{1-\phi^2})C(t)]$ *in which*
$C(t) = \frac{36\left(-9t^5+30t^4T-5t^3T(2+11T)+5t^2T\left(1+2T+13T^2\right)-2t\left(-2+5T+5T^2+20T^4\right)+T\left(-4+10T+5T^2+9T^4\right)\right)}{5(-1+T)^2T^2(1+T)^4}$.
Additionally, as $N \to \infty$ and $T \to \infty$,

$$\sqrt{N}T(\hat{\beta}_{LSDV} - 1 + \frac{3}{T+1}) \xrightarrow{d} N(0, \frac{51}{5})$$

Even if it is always consistent, there is an asymptotic bias term in the asymptotic distribution. If $\phi > 0$ and $\theta > 0$, $B_u(\phi, \theta) > B_u(0, 0)$ which is equal to $\frac{3(17T^2-20T+17)}{5(T-1)(T+1)^3}$ as in Harris and Tzavalis (1999). Stochastic volatility changes asymptotic variance of the estimator when T is fixed but keep it asymptotically same when $T \to \infty$. When $T = 20$, for example, asymptotic variance is approximately 0.02 for the model without stochastic volatility, but can be 0.07 when $\phi = 0.9$ and $\theta = 0.5$. $B_u(\phi, \theta) > B_u(0, 0)$ implies that naively ignoring stochastic volatility would systematically reject Harris and Tzavalis (1999)'s panel unit root test more frequently than it should be, although the estimates lose only a little statistical efficiency.

In the second step, we first estimate the error terms using the first-step estimator, i.e.,

$$\hat{\varepsilon}_{it} = y_{it} - \hat{\beta}_{LSDV} y_{i,t-1} - \hat{\mu}_i$$

where $\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\beta}_{LSDV} y_{i,t-1})$ is a consistent estimator of the individual effect. Next, we use QML (Harvey and Shephard, 1996) to estimate the parameters in stochastic volatility. Specifically, we calculate the quasi-likelihood of $\hat{\varepsilon}_{it}$ via Kalman filters in a state space form

$$\log(\hat{\varepsilon}_{it}^2) = \log(\sigma_{it}^2) + \mathrm{E}(\log(\epsilon_{it}^2)) + \xi_{it}$$

$$\log(\sigma_{it}^2) = \mu + \phi(\log(\sigma_{i,t-1}^2) - \mu) + \eta_{it}$$

where $\mathrm{E}(\log(\epsilon_{it}^2)) = 1.27$ and ξ_{it} is a normal variable with mean zero and variance 4.93. QML estimators of ϕ, θ and μ are consistent and asymptotically normal.

5 Monte Carlo Studies

In this section, we investigate the finite-sample performance of particle-filter-based estimators and LSDV-QML estimators. For the moment, we study the frequentist approach in Section 3.2 rather than the Bayesian approach in Section 3.3 because the former is much faster than the latter. When estimating $\hat{\beta}$, we compare our estimators with some standard estimators in the literature including instrumental variables (IV) (Anderson and Hsiao, 1982), standard GMM (GMM) (Arellano and Bond, 1991) and system GMM (SGMM) (Blundell and Bond, 1998). The quality of the estimators are evaluated by biases and root mean square errors (RMSE). Our study complements previous Monte Carlo studies for panel data models in a macroeconomic context such as Judson and Owen (1999). Next we will briefly review each candidate estimator other than LSDV and particle filters.

1) IV overcomes endogeneity by applying to the first-differenced equation the two-stage least squares method with lagged levels or lagged differences being instrumental variables, such as $y_{i,t-2}$ or $\triangle y_{i,t-2}$. The Anderson-Hsiao IV estimators are consistent for a fixed T as $N \to \infty$. We will use $y_{i,t-2}$ as the instrument, since Arellano (1989) showed the estimators using lagged differences as instruments have very large variance. IV estimators have very poor finite sample properties when the instrument is only weakly correlated with the endogenous variable. The problem arises when β is sufficiently close to one as shown in our results.

2) GMM exploits a set of linear orthogonality conditions aimed for improving efficiency, i.e. $\mathrm{E}(y_{i,t-s}\triangle\varepsilon_{it}) = 0$ for $t \geqslant 3$ and $s \geqslant 2$ given predetermined initial conditions $\mathrm{E}(y_{i1}\varepsilon_{it}) = 0$ for $t \geqslant 2$, serially uncorrelated shocks $\mathrm{E}(\varepsilon_{it}\varepsilon_{is}) = 0$ for $t \neq s$ and random effect $\mathrm{E}(\mu_i\varepsilon_{is}) = 0$. GMM estimators are all consistent but the relative efficiency depends on the weight matrix. We use one-step GMM instead of two-step GMM for its better finite sample performance. Similar to

IV estimators, standard first-differenced GMM estimators in Arellano and Bond (1991) are also subject to the problem of weak instruments in case $\beta \to 1$.

3) System GMM uses extra moment conditions including the lagged differences as instruments for equations in levels given an additional initial condition $\mathrm{E}(\triangle y_{i2}\mu_i) = 0$. A large improvement in finite sample properties is expected particularly when $\beta \to 1$ provided that the extra initial condition is valid. The validity of initial conditions depends on the proper design of data generation process. We employ the two-step version of system GMM estimators, where the second step estimators are based on the residuals from the one-step estimators.

The baseline data generation process in our Monte Carlo studies is

$$y_{it} = \beta y_{i,t-1} + (1-\beta)\mu_i + \varepsilon_{it}$$

$$\mu_i = \sqrt{\tau}(\frac{q_i - 1}{\sqrt{2}})\varsigma_i$$

$$\varepsilon_{it} = \sigma_{it}\epsilon_{it}$$

$$\log(\sigma_{it}^2) = \mu + \phi(\log(\sigma_{i,t-1}^2) - \mu) + \eta_{it}$$

where $q_i \sim \chi_1^2$; $\epsilon_{it}, \varsigma_i \sim N(0,1)$ and $\eta_{it} \sim N(0,\theta^2)$; q_i, ϵ_{it}, ς_i and η_{it} are all i.i.d. within series and also independent of each others. For β we consider three types, i.e. $\beta = 0.5$ for a series which is not very persistent, $\beta = 0.9$ for a very persistent series and $\beta = 1$ which is a unit root process. Binder et al. (2005) showed that τ, which measures the degree of cross-section to time-series variation, can influence the finite sample performance of GMM-type estimators. We set $\tau = 1$. μ represents the long-run mean of the log volatility and is set equal to $\log(0.04)$. We want to check the performance of particle filter based estimators both when stochastic volatility exists and not, so for ϕ and θ we consider two cases: $\phi = \theta = 0$ is a panel model without stochastic volatility; $\phi = 0.9$ and $\theta = 0.5$ represents an example of persistent stochastic volatility. For initialization, we set $y_{i0} = 0$ and discard the first 100 observations of the simulated data in order that the series is long enough to eliminate the initial effect. As the particle filter based estimation is relatively computationally intensive, we can only carry out 100 replications for all experiments in this section.

Tables 1-3 list the basic simulation results. Our findings suggest that particle-filter-based estimators are more precise than other estimators on average, although no estimator is the best choice in all circumstances, which varies with the size of the panel. As T grows, all estimators other than system GMM become more precise, since the burnin peroid of 100 observations is perhaps not long enough to eliminate the initial effect for system GMM estimators. When $T = 50$, particle-filter-based estimators are more precise than the others in the presence of stochastic volatility and even in the case of homoscedasticity, except that in the unit root case without stochastic volatility where LSDV is slightly better. When $T = 20$, system GMM shows advantage over the others in many cases, while particle filters are best when $N = 20$ and standard GMM is best

T	N	ϕ	LSDV	IV	GMM	SGMM	PF
20	20	0	-0.082	0.005	-0.093	0.059	-0.079
			(0.095)	(0.193)	(0.110)	(0.190)	**(0.093)**
		0.9	-0.077	0.022	-0.093	0.050	-0.075
			(0.102)	(0.176)	(0.123)	(0.130)	**(0.095)**
	50	0	-0.087	0.011	-0.048	0.100	-0.085
			(0.090)	(0.110)	**(0.062)**	(0.131)	(0.088)
		0.9	-0.093	0.012	-0.070	0.047	-0.094
			(0.322)	(0.121)	**(0.095)**	(0.100)	(0.101)
50	20	0	-0.030	0.005	-0.035	-0.057	-0.025
			(0.044)	(0.075)	(0.048)	(0.249)	**(0.044)**
		0.9	-0.033	0.007	-0.039	-0.032	-0.028
			(0.053)	(0.094)	(0.058)	(0.177)	**(0.048)**
	50	0	-0.030	-0.001	-0.032	0.048	-0.025
			(0.035)	(0.047)	(0.037)	(0.106)	**(0.035)**
		0.9	-0.030	0.000	-0.036	0.031	-0.026
			(0.045)	(0.071)	(0.050)	(0.076)	**(0.041)**

Table 1: Summary of simulation results of $\hat{\beta}$ when $\beta = 0.5$. In each circumstance, the first row gives the biases and the second row gives RMSE in brackets. Numbers in bold font indicate the estimator with the smallest RMSE.

when $N = 50$ if $\beta = 0.5$. Table 4 studies the robustness of particle-filter-based estimators and LSDV under the fat-tail distribution (student t distribution with degree of freedom 5). Comparing Tables 1 and 4, one can see that particle filters are almost unaffected by the fat tails but LSDV estimators are affected in the presence of stochastic volatility. Table 5 lists the summary of simulation results of $\hat{\phi}$, $\hat{\theta}$ and $\hat{\mu}$. The estimates in the presence of stochastic volatility are much more precise than those without stochastic volatility, which suffers from the boundary issue we will solve soon.

We also study an AR(1) panel model with an additional regressor. Specifically, the data generation processes for the dependent variable and the additional regressor are

$$y_{it} = \beta y_{i,t-1} + \alpha x_{it} + (1-\beta)\mu_i + \varepsilon_{it}$$

$$x_{it} = q x_{i,t-1} + \zeta_{it}$$

where the additional regressor x_{it} follows an AR(1) process with coefficient $q = 0.9$ and $N(0, 0.01)$ error terms. α is set equal to 0.7. The data generation processes for μ_i and ε_{it} are same as before. Table 6 lists the simulation result when $N = 50$, $T = 50$, $\beta = 0.5$, $\phi = 0.9$, $\theta = 0.5$ and $\mu = \log(0.04)$. The results show that the estimates of α are acceptable although they are a little less precise than those of β.

T	N	ϕ	LSDV	IV	GMM	SGMM	PF
20	20	0	-0.127	-0.227	-0.166	-0.110	-0.127
			(0.134)	(2.111)	(0.178)	(0.168)	**(0.133)**
		0.9	-0.129	-0.035	-0.168	-0.092	-0.117
			(0.141)	(0.803)	(0.185)	**(0.125)**	(0.128)
	50	0	-0.130	0.013	-0.127	-0.034	-0.126
			(0.131)	(0.160)	(0.138)	**(0.053)**	(0.127)
		0.9	-0.135	0.019	-0.150	-0.064	-0.124
			(0.138)	(0.164)	(0.163)	**(0.084)**	(0.127)
50	20	0	-0.046	0.007	-0.053	-0.295	-0.038
			(0.050)	(0.102)	(0.058)	(0.383)	**(0.045)**
		0.9	-0.049	0.010	-0.057	-0.182	-0.031
			(0.057)	(0.119)	(0.065)	(0.244)	**(0.040)**
	50	0	-0.045	-0.001	-0.054	-0.087	-0.039
			(0.046)	(0.058)	(0.056)	(0.108)	**(0.043)**
		0.9	-0.046	0.001	-0.060	-0.065	-0.027
			(0.049)	(0.088)	(0.063)	(0.082)	**(0.034)**

Table 2: Summary of simulation results of $\hat{\beta}$ when $\beta = 0.9$. In each circumstance, the first row gives the biases and the second row gives RMSE in brackets. Numbers in bold font indicate the estimator with the smallest RMSE.

T	N	ϕ	LSDV	GMM	SGMM	PF
20	20	0	-0.152	-0.221	-0.088	-0.148
			(0.155)	(0.230)	**(0.119)**	(0.152)
		0.9	-0.156	-0.233	-0.045	-0.152
			(0.163)	(0.247)	**(0.067)**	(0.158)
	50	0	-0.153	-0.223	-0.062	-0.147
			(0.154)	(0.233)	**(0.080)**	(0.149)
		0.9	-0.160	-0.261	-0.046	-0.157
			(0.164)	(0.273)	**(0.055)**	(0.159)
50	20	0	-0.063	-0.076	-0.235	-0.066
			(0.065)	(0.078)	(0.281)	(0.068)
		0.9	-0.066	-0.082	-0.138	-0.053
			(0.069)	(0.087)	(0.180)	**(0.055)**
	50	0	-0.061	-0.088	-0.082	-0.064
			(0.061)	(0.090)	(0.094)	(0.065)
		0.9	-0.061	-0.098	-0.053	-0.049
			(0.063)	(0.100)	(0.062)	**(0.051)**

Table 3: Summary of simulation results of $\hat{\beta}$ when $\beta = 1$. In each circumstance, the first row gives the biases and the second row gives RMSE in brackets. Numbers in bold font indicate the estimator with the smallest RMSE.

T	N	ϕ	LSDV	PF
50	20	0	-0.032	-0.023
			(0.042)	(0.039)
		0.9	-0.038	-0.030
			(0.078)	(0.045)
	50	0	-0.032	-0.023
			(0.036)	(0.034)
		0.9	-0.043	-0.027
			(0.073)	(0.041)

Table 4: Summary of simulation results of $\hat{\beta}$ when $\beta = 0.5$ and the shocks have t-distribution with degree of freedom 5. In each circumstance, the first row gives the biases and the second row gives RMSE in brackets.

6 Conclusions

Motivated by the evidence of time-varying volatility in the residuals of many estimated dynamic regression models in macroeconomics, we study dynamic panel data models with stochastic volatility in a macroeconomic context. We propose two classes of parameter estimators: one is based on the particle filter technique and the other is a two-step LSDV-QML estimator. For particle filters, we mainly study the frequentist approach which is the maximization of approximate simulated-likelihood, since it is much faster than the particle Metropolis-Hastings sampler. The main simulation result shows particle-filter-based estimators are more precise than the others on average in the presence of stochastic volatility and even in the case of homoscedasticity especially when T is large. We derive the asymptotic distribution of LSDV estimators in the presence of stochastic volatility. It implies that naively ignoring stochastic volatility would systematically reject Harris and Tzavalis (1999)'s panel unit root test more frequently than it should be. Our methodology can also be applied to panel VAR models.

Appendices

Proof of Theorems

Lemma 1. *(Moments of Stochastic Volatility)*
Under Assumptions 3-4,

$$\mathrm{E}(\sigma_{it}^2) = \exp(\mu + \frac{\theta^2}{2(1-\phi^2)})$$

$$\mathrm{E}(\sigma_{it}^4) = \exp(2\mu + \frac{2\theta^2}{1-\phi^2})$$

$$\mathrm{Cov}_{s-t} = \mathrm{E}(\sigma_{it}^2\sigma_{is}^2) = \exp(2\mu + \frac{\theta^2(1+\phi^{|t-s|})}{1-\phi^2})$$

Lemma 1 collects some useful moments of stochastic volatility process which will be applied in the proof of theorems.

Lemma 2. *Under Assumptions 1-6,*

$$\mathrm{Var}(\frac{1}{T}\sum_{t=1}^{T} w_{it}^2) \to 0$$

where $w_{i,t-1} = y_{i,t-1} - \frac{\mu_i}{1-\beta}$.

Proof of Lemma 2

Firstly, note that w_{it} is simply an AR(1) process with the error term being a martingale difference, i.e. $w_{it} = \beta w_{i,t-1} + \varepsilon_{it}$.

Next, we have

$$(1-\beta^2)\frac{1}{T}\sum_{t=1}^{T} w_{it}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 - \beta^2\frac{1}{T}(w_{iT}^2 - w_{i0}^2) + 2\beta\frac{1}{T}\sum_{t=1}^{T} w_{i,t-1}\varepsilon_{it}$$

Since $\frac{1}{T}(w_{iT}^2 - w_{i0}^2) \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{t=1}^{T} w_{i,t-1}\varepsilon_{it} \xrightarrow{p} 0$ given that $\{w_{i,t-1}\varepsilon_{it}\}$ is a martingale difference series, we only need to prove $\mathrm{Var}(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2) \to 0$.

Because of Lemma 1, we know that $\{\varepsilon_{it}^2\}$ is stationary and asymptotically uncorrelated with finite moments up to the second order. Therefore, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2 \xrightarrow{p} \mathrm{E}(\varepsilon_{it}^2)$ (e.g. White (2001)). As a result $\mathrm{Var}(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}^2) \to 0$ and $\mathrm{Var}(\frac{1}{T}\sum_{t=1}^{T} w_{it}^2) \to 0$.

Proof of Theorem 1

Let $\mathbf{x_i} = \mathbf{y}_{i,(-1)}$. First, note that

$$\mathrm{E}(\sum_{i=1}^{N}\mathbf{x}_i^{'}Q\varepsilon_i) = N\mathrm{E}(\mathbf{x}_i^{'}Q\varepsilon_i) = -N\frac{\exp(\mu + \frac{\theta^2}{2(1-\phi^2)})}{1-\beta}(1 - \frac{1-\beta^T}{T(1-\beta)})$$

since

$$
\begin{aligned}
\mathrm{E}(\mathbf{x}_i^{'}Q\varepsilon_i) &= \mathrm{E}(\mathbf{x}_i^{'}\varepsilon_i) - \frac{1}{T}\iota^{'}\mathrm{E}(\varepsilon_i x_i^{'})\iota \\
&= 0 - \frac{\exp(\mu + \frac{\theta^2}{2(1-\phi^2)})}{1-\beta}(1 - \frac{1-\beta^T}{T(1-\beta)}) \\
&= -\frac{\exp(\mu + \frac{\theta^2}{2(1-\phi^2)})}{1-\beta}(1 - \frac{1-\beta^T}{T(1-\beta)})
\end{aligned}
$$

Next, we have

$$
\frac{\sum_{i=1}^N \mathbf{x}_i^{'}Q\varepsilon_i}{\sqrt{NT}} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T w_{it-1}\varepsilon_{it} - \sqrt{\frac{T}{N}}\sum_{i=1}^N \bar{w}_{i(-1)}\bar{\varepsilon}_i
$$

For the first part on the right-hand side,

$$
\begin{aligned}
\mathrm{Var}(\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T w_{it-1}\varepsilon_{it}) &= \mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^T w_{it-1}\varepsilon_{it}) \\
&= \mathrm{E}(w_{it-1}^2\varepsilon_{it}^2) \\
&= \exp(2\mu + \frac{\theta^2}{1-\phi^2})\sum_{t=1}^{\infty}[\exp(\frac{\theta^2\phi^t}{1-\phi^2})\beta^{2t-2}]
\end{aligned}
$$

For the second part on the right-hand side,

$$
\mathrm{Var}(\sqrt{\frac{T}{N}}\sum_{i=1}^N \bar{w}_{i(-1)}\bar{\varepsilon}_i) = T\mathrm{Var}(\bar{w}_{i(-1)}\bar{\varepsilon}_i) \to O(T^{-1})
$$

Therefore,

$$
\mathrm{Var}(\frac{\sum_{i=1}^N \mathbf{x}_i^{'}Q\varepsilon_i}{\sqrt{NT}}) \to \exp(2\mu + \frac{\theta^2}{1-\phi^2})\sum_{t=1}^{\infty}[\exp(\frac{\theta^2\phi^t}{1-\phi^2})\beta^{2t-2}]
$$

Moreover,

$$
\frac{\sum_{i=1}^N \mathbf{x}_i^{'}Q\mathbf{x}_i}{NT} = \frac{\sum_{i=1}^N w_i^{'}Qw_i}{NT} = \frac{1}{N}\sum_{i=1}^N(\frac{1}{T}\sum_{t=1}^T w_{i,t-1}^2 - \bar{w}_{i,(-1)}^2)
$$

$$
\begin{aligned}
\mathrm{E}(\frac{\sum_{i=1}^N \mathbf{x}_i^{'}Q\mathbf{x}_i}{NT}) &= \mathrm{E}(w_{i,t-1}^2) - \mathrm{E}(\bar{w}_{i,(-1)}^2) \\
&= \frac{\exp(\mu + \frac{\theta^2}{2(1-\phi^2)})}{1-\beta^2} - O(T^{-1})
\end{aligned}
$$

Because of Lemma 2,

$$\mathrm{Var}(\frac{\sum_{i=1}^{N}\mathbf{x}_i^{'}Q\mathbf{x}_i}{NT}) = \frac{1}{N}\mathrm{Var}(\frac{1}{T}\sum_{t=1}^{T} w_{i,t-1}^2 - \bar{w}_{i,(-1)}^2) \to 0$$

As a result,

$$\frac{\sum_{i=1}^{N}\mathbf{x}_i^{'}Qx_i}{NT} \xrightarrow{p} \frac{\exp(\mu + \frac{\theta^2}{2(1-\phi^2)})}{1-\beta^2}$$

The remaining proof follows exactly that of Theorem 1 in Alvarez and Arellano (2003).

Proof of Theorem 2

Under Assumptions 7-8,

$$\mathbf{y}_{i,(-1)} = \iota y_{i0} + C\varepsilon_i \tag{10}$$

where

$$C = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix}$$

By substituting (10) into (9), the bias term in LSDV estimators can be written as

$$\hat{\beta} - 1 = \frac{\sum_{i=1}^{N}\{\sum_{t=1}^{T} A_1\varepsilon_{it}^2 + \sum_{t=1}^{T}\sum_{s=t+1}^{T} B_1\varepsilon_{it}\varepsilon_{is}\}}{\sum_{i=1}^{N}\{\sum_{t=1}^{T} A_2\varepsilon_{it}^2 + \sum_{t=1}^{T}\sum_{s=t+1}^{T} B_2\varepsilon_{it}\varepsilon_{is}\}}$$

where $A_1 = -\frac{T-t}{T}$, $B_1 = -\frac{T-s-t}{T}$, $A_2 = \frac{t(T-t)}{T}$ and $B_2 = \frac{2t(T-s)}{T}$.

The probability limits of the numerator and denominator are $-\frac{T-1}{2}\mathrm{E}(\sigma_{it}^2)$ and $\frac{T^2-1}{6}\mathrm{E}(\sigma_{it}^2)$, respectively. Therefore,

$$\hat{\beta} - 1 \xrightarrow{p} -\frac{3}{T+1}$$

Now we need to derive the asymptotic variance. Notice that

$$\hat{\beta}-1+\frac{3}{T+1} = \frac{\sum_{i=1}^{N}\{\sum_{t=1}^{T}(A_1 + A_2\frac{3}{T+1})\varepsilon_{it}^2 + \sum_{t=1}^{T}\sum_{s=t+1}^{T}(B_1 + B_2\frac{3}{T+1})\varepsilon_{it}\varepsilon_{is}\}}{\sum_{i=1}^{N}\{\sum_{t=1}^{T} A_2\varepsilon_{it}^2 + \sum_{t=1}^{T}\sum_{s=t+1}^{T} B_2\varepsilon_{it}\varepsilon_{is}\}}$$

The numerator has zero mean, so its variance is

$$N(A_3\mathrm{E}(\sigma_{it}^4)\kappa + \sum_{t=1}^{T-1} B(t)\mathrm{Cov}_t)$$

where

$$A_3 = \frac{(-2+T)(-1+T)(-1+2T)}{15T(1+T)}$$

$$B(t) = \frac{-9t^5+30t^4T-5t^3T(2+11T)+5t^2T\left(1+2T+13T^2\right)-2t\left(-2+5T+5T^2+20T^4\right)+T\left(-4+10T+5T^2+9T^4\right)}{5T^2(1+T)^2}.$$

As a result,

$$\mathrm{Var}(\sqrt{N}(\hat{\beta} - 1 + \frac{3}{T+1})) \to B_u(\phi, \theta)$$

in which $B_u(\phi, \theta)$ is given in Theorem 2. Theorem 2 is proved.

References

Alvarez, Javier, and Manuel Arellano, 2003, The time series and cross-section asymptotics of dynamic panel data estimators, Econometrica 71(4), 1121–1159.

Anderson, Theodore Wilbur, and Cheng Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of econometrics 18(1), 47–82.

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein, 2010, Particle markov chain monte carlo methods, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(3), 269–342.

Arellano, Manuel, 1989, A note on the anderson-hsiao estimator for panel data, Economics Letters 31(4), 337–341.

Arellano, Manuel, and Stephen Bond, 1991, Some tests of specification for panel data: Monte carlo evidence and an application to employment equations, The Review of Economic Studies 58(2), 277–297.

Binder, Michael, Cheng Hsiao, and M Hashem Pesaran, 2005, Estimation and inference in short panel vector autoregressions with unit roots and cointegration, Econometric Theory 21(4), 795–837.

Blanchard, Olivier, and John Simon, 2001, The long and large decline in us output volatility, Brookings Papers on Economic Activity 2001(1), 135–174.

Blundell, Richard, and Stephen Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics 87(1), 115–143.

Carpenter, James, Peter Clifford, and Paul Fearnhead, 1999, Improved particle filter for nonlinear problems, IEE Proceedings-Radar, Sonar and Navigation 146(1), 2–7.

Chen, Rong, and Jun S Liu, 2000, Mixture kalman filters, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62(3), 493–508.

Creal, Drew, 2012, A survey of sequential monte carlo methods for economics and finance, Econometric Reviews 31(3), 245–296.

Del Moral, Pierre, 2004, Feynman-kac formulae, genealogical and interacting particle systems with applications. New York: Springer-Verlag.

Douc, Randal, Eric Moulines, and Jimmy Olsson, 2012, Long-term stability of sequential monte carlo methods under verifiable conditions, arXiv:1203.6898.

Durbin, James, and Siem Jan Koopman, 2012, Time series analysis by state space methods. Oxford University Press.

Fernández-Villaverde, Jesús, and Juan F Rubio-Ramírez, 2007, Estimating macroeconomic models: A likelihood approach, The Review of Economic Studies 74(4), 1059–1087.

_ 2013, Macroeconomics and volatility: Data, models, and estimation. in Daron Acemoglu, Manuel Arellano and Eddie Dekel, eds, Advances in Economics and Econometrics: Tenth World Congress, Vol. 3, Cambridge University Press, 137-183.

Frankel, Jeffrey A, and Andrew K Rose, 1996, A panel project on purchasing power parity: mean reversion within and between countries, Journal of International Economics 40(1), 209–224.

Gordon, Neil J, David J Salmond, and Adrian FM Smith, 1993, Novel approach to nonlinear/non-gaussian bayesian state estimation, in: IEE Proceedings F (Radar and Signal Processing), vol. 140 IET 107–113.

Hamilton, James D, 2010, Macroeconomics and arch. in Tim Bollerslev, Jeffry R. Russell and Mark Watson, eds, Volatility and Time Series Econometrics: Essays in Honor of Robert Engle, Oxford University Press, 79-96.

Harris, Richard DF, and Elias Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed, Journal of Econometrics 91(2), 201–226.

Harvey, Andrew C, and Neil Shephard, 1996, Estimation of an asymmetric stochastic volatility model for asset returns, Journal of Business & Economic Statistics 14(4), 429–434.

Harvey, Andrew, Esther Ruiz, and Neil Shephard, 1994, Multivariate stochastic variance models, The Review of Economic Studies 61(2), 247–264.

Islam, Nazrul, 1995, Growth empirics: a panel data approach, The Quarterly Journal of Economics 110(4), 1127–1170.

Judson, Ruth A, and Ann L Owen, 1999, Estimating dynamic panel data models: a guide for macroeconomists, Economics Letters 65(1), 9–15.

Justiniano, Alejandro, and Giorgio E Primiceri, 2008, The time-varying volatility of macroeconomic fluctuations, American Economic Review 98(3), 604–641.

Kantas, Nicholas, Arnaud Doucet, Sumeetpal Sindhu Singh, and Jan Marian Maciejowski, 2009, An overview of sequential monte carlo methods for parameter estimation in general state-space models, 15th IFAC Symposium on System Identification 15, 774–785.

Kim, Chang-Jin, and Charles R Nelson, 1999, Has the us economy become more stable? a bayesian approach based on a markov-switching model of the business cycle, Review of Economics and Statistics 81(4), 608–616.

Kitagawa, Genshiro, 1996, Monte carlo filter and smoother for non-gaussian nonlinear state space models, Journal of Computational and Graphical Statistics 5(1), 1–25.

Koop, G, and D Korobilis, 2010, Bayesian multivariate time series methods for empirical macroeconomics, Foundations and Trends in Econometrics 3(4), 267–358.

Levin, Andrew, and Chien-Fu Lin, 1992, Unit root tests in panel data: Asymptotic and finite-sample properties, University of California at San Diego, Economics Working Paper Series.

Liu, Jun S, and Rong Chen, 1998, Sequential monte carlo methods for dynamic systems, Journal of the American Statistical Association 93(443), 1032–1044.

McConnell, Margaret M., and Gabriel Perez-Quiros, 2000, Output fluctuations in the united states: What has changed since the early 1980's?, American Economic Review 90(5), 1464–1476.

Nickell, Stephen John, 1981, Biases in dynamic models with fixed effects, Econometrica 49(6), 1417–26.

Olsson, Jimmy, and Tobias Rydén, 2008, Asymptotic properties of particle filter-based maximum likelihood estimators for state space models, Stochastic Processes and their Applications 118(4), 649–680.

Olsson, Jimmy, Olivier Cappé, Randal Douc, and Eric Moulines, 2008, Sequential monte carlo smoothing with application to parameter estimation in nonlinear state space models, Bernoulli 14(1), 155–179.

Pitt, Michael K, 2002, Smooth particle filters for likelihood evaluation and maximisation. Unpublished Working Paper.

Pitt, Michael K, and Neil Shephard, 1999, Filtering via simulation: Auxiliary particle filters, Journal of the American Statistical Association 94(446), 590–599.

Shephard, Neil, 2013, Martingale unobserved component models. Unpublished Working Paper.

Sims, Christopher A, and Tao Zha, 2006, Were there regime switches in us monetary policy?, The American Economic Review 96(1), 54–81.

Stock, James H, and Mark W Watson, 2003, Has the business cycle changed and why?, NBER: Macroeconomics Annual 2002, 159–230.

Weiss, Andrew A, 1984, Arma models with arch errors, Journal of Time Series Analysis 5(2), 129–143.

White, Halbert, 2001, Asymptotic theory for econometricians. Academic press New York.

Whiteley, Nick, 2011, Stability properties of some particle filters, arXiv:1109.6779.

Wu, Jhy-Lin, and Show-Lin Chen, 2001, Mean reversion of interest rates in the eurocurrency market, Oxford Bulletin of Economics and Statistics 63(4), 459–473.

				LSDV-QML			PF		
β	T	N	ϕ	$\hat{\phi}$	$\hat{\theta}$	$\hat{\mu}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\mu}$
0.5	20	20	0	0.261	0.290	-0.002	0.323	0.132	-0.001
				(0.440)	(0.454)	(0.005)	(0.450)	(0.195)	(0.004)
			0.9	0.009	-0.087	0.002	-0.055	0.116	0.001
				(0.085)	(0.246)	(0.010)	(0.085)	(0.181)	(0.010)
		50	0	0.217	0.243	-0.002	0.380	0.093	-0.001
				(0.398)	(0.392)	(0.004)	(0.513)	(0.128)	(0.003)
			0.9	0.019	-0.087	0.001	-0.026	0.068	0.001
				(0.045)	(0.153)	(0.006)	(0.050)	(0.116)	(0.007)
	50	20	0	0.221	0.231	-0.001	0.271	0.116	-0.001
				(0.402)	(0.360)	(0.003)	(0.410)	(0.165)	(0.003)
			0.9	-0.004	-0.025	0.002	-0.027	0.072	0.003
				(0.042)	(0.133)	(0.008)	(0.050)	(0.121)	(0.009)
		50	0	0.251	0.200	-0.001	0.283	0.064	-0.001
				(0.413)	(0.302)	(0.002)	(0.412)	(0.089)	(0.002)
			0.9	0.004	-0.040	0.001	-0.013	0.037	0.001
				(0.030)	(0.087)	(0.004)	(0.028)	(0.066)	(0.006)
0.9	20	20	0	0.276	0.250	-0.003	0.326	0.154	-0.003
				(0.451)	(0.421)	(0.005)	(0.444)	(0.227)	(0.004)
			0.9	0.000	-0.096	0.001	-0.048	0.107	0.000
				(0.126)	(0.225)	(0.010)	(0.080)	(0.179)	(0.010)
		50	0	0.205	0.271	-0.004	0.375	0.107	-0.003
				(0.395)	(0.406)	(0.004)	(0.516)	(0.145)	(0.003)
			0.9	0.022	-0.102	0.000	-0.025	0.074	0.001
				(0.053)	(0.171)	(0.005)	(0.051)	(0.130)	(0.007)
	50	20	0	0.215	0.254	-0.001	0.305	0.109	-0.001
				(0.380)	(0.377)	(0.003)	(0.433)	(0.153)	(0.003)
			0.9	-0.003	-0.033	0.003	-0.016	0.055	0.003
				(0.045)	(0.130)	(0.008)	(0.041)	(0.099)	(0.008)
		50	0	0.211	0.219	-0.001	0.355	0.067	-0.001
				(0.368)	(0.324)	(0.002)	(0.478)	(0.090)	(0.002)
			0.9	0.004	-0.044	0.001	-0.011	0.028	0.001
				(0.027)	(0.085)	(0.004)	(0.031)	(0.066)	(0.005)
1	20	20	0	0.235	0.272	-0.005	0.325	0.165	-0.004
				(0.420)	(0.438)	(0.006)	(0.453)	(0.224)	(0.005)
			0.9	0.012	-0.089	-0.001	-0.033	0.080	-0.001
				(0.074)	(0.220)	(0.009)	(0.067)	(0.151)	(0.010)
		50	0	0.251	0.203	-0.005	0.365	0.111	-0.004
				(0.425)	(0.325)	(0.005)	(0.501)	(0.163)	(0.004)
			0.9	0.024	-0.106	-0.001	-0.020	0.064	-0.001
				(0.053)	(0.169)	(0.005)	(0.044)	(0.114)	(0.007)
	50	20	0	0.237	0.270	-0.002	0.289	0.123	-0.002
				(0.413)	(0.389)	(0.004)	(0.416)	(0.176)	(0.003)
			0.9	0.001	-0.041	0.002	-0.024	0.070	0.000
				(0.038)	(0.123)	(0.008)	(0.049)	(0.128)	(0.008)
		50	0	0.264	0.22726	-0.002	0.325	0.076	-0.002
				(0.428)	(0.338)	(0.003)	(0.466)	(0.112)	(0.002)
			0.9	0.009	-0.054	0.001	-0.006	0.023	0.001
				(0.029)	(0.094)	(0.004)	(0.029)	(0.065)	(0.005)

Table 5: Summary of simulation results of $\hat{\phi}$, $\hat{\theta}$ and $\hat{\mu}$. In each circumstance, the first row gives the biases and the second row gives RMSE in brackets.

	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\theta}$	$\hat{\mu}$
LSDV-QML	-0.027	0.025	-0.004	-0.020	0.001
	(0.038)	(0.045)	(0.030)	(0.087)	(0.005)
PF	-0.030	-0.062	-0.025	-0.004	0.008
	(0.042)	(0.071)	(0.046)	(0.070)	(0.009)

Table 6: Summary of simulation results of the AR(1) panel models with an additional regressor when $N = 50$, $T = 50$, $\beta = 0.5$, $\alpha = 0.7$, $\phi = 0.9$, $\theta = 0.5$ and $\mu = \log(0.04)$. For each estimator, the first row gives the biases and the second row gives RMSE in brackets.