Estimating Vector Autoregressions with Long Heterogeneous Panels

Alessandro Rebucci*
International Monetary Fund

Initial draft: August 1998
This draft: January 2000

Abstract

This paper studies cross sectional slope heterogeneity in stationary, exactly identified panel VARs. I show that (i) the heterogeneity bias of standard pooled estimators is generally different from zero in large samples, but its theoretical magnitude is not necessarily large and depends on the moments of the cross sectional distribution of the slope parameters and the error terms; (ii) slope heterogeneity must be relatively high for the bias of pooled estimators to be substantial in finite samples; (iii) the time dimension of the panel must be longer than previously thought for the small T bias of the mean group estimator to become negligible if the error terms are contemporaneously correlated.

JEL: C33, C13, C15.

Keywords: dynamic panel data, slope heterogeneity, VAR.

.

*The views expressed in this paper are exclusively those of the author and not those of the International Monetary Fund. The first draft of this paper was written at the Department of Economics of Queen Mary and Westfield College (QMW), University of London. Financial support from QMW College is gratefully acknowledged. I thank my Ph. D. advisors Christopher Gilbert and Valentina Corradi, and Manuel Arellano, Ron Smith, Elias Tzavalis and Giovanni Urga for helpful suggestions and discussions. I also thank Kyung So Im for providing me some of his Gauss codes. The usual disclaimer applies.

1 Introduction

Vector autoregressive systems (VARs) are a useful device to summarise and analyse the dynamic interaction of a given set of variables of interest. When the are several decision units to be considered (i.e., several agents, countries, sectors, etc.), the possibility of pooling them in a single system emerges. Pooling different decision units is attractive because it increases the number of degree of freedom available and, potentially, the efficiency of the estimates so obtained. On the other hand, pooling different decision units poses inferential problems with regards to the representative or typical unit.

For the purpose of the analysis in this paper, we can think of a panel VAR (PVAR) as a standard dynamic panel data model (DPM) where no regressor is *strongly exogenous*. Much of the existing literature on DPMs is focused on the problem of pooling heterogeneous units with respect to the unconditional mean (the intercept of the regression equation), or the unconditional variance (the variance of the error term in the regression equation), of the variables of interest.[1] The problem of pooling heterogeneous units with respect to the time series correlations of the variables of interest (the slope parameters of the regression equation) has started to be investigated only more recently (see Robertson and Symons [1992] and Pesaran and Smith [1995]). Pesaran and Smith [1995], in particular, have shown that if the slope parameters of a standard dynamic panel model differ across individual units, then a number of commonly used pooled estimators give rise to *inconsistent* estimates of the true cross sectional mean of the parameters of interest, even when both the number of individual units and time periods are large. To solve this problem, they propose an *arithmetic average* of the time series estimates of the parameters of interest, and indeed they show that this estimator, called *mean group estimator* (MG), is consistent. Furthermore, Pesaran, Smith and Im [1996] give monte Carlo simulation evidence showing that the bias in conventional estimates induced by the presence of slope heterogeneity may be substantial in finite samples.

This paper is a first attempt at investigating slope heterogeneity in PVARs. This issue is particularly relevant as applied works using VARs estimated with panel data are appearing in growing numbers, while so far little attention has been devoted to the statistical problems involved in moving from a

[1] See Baltagi [1995], Hisiao [1986], and Matyas and Sevestre [1996] for standard results on dynamic panel data models.

standard, dynamic heterogeneous panel specification to a VAR one.[2] On the other hand, the range of statistical problems involved is wide, and therefore the analysis carried out in this paper is necessarily limited in scope. First, I shall restrict my attention to exactly identified VARs in time series sense. Thus, I focus only on the reduced form estimation of the model.[3] Second, I shall assume that slope parameters are constant over time, and I consider only I(0) systems.[4] Motivated by typical macro applications such as those using the Heatson and Summers [1991] data set, I shall consider only long panels (sometime called random fields in the literature), paying particular attention to extremely unfavorable panel dimensions.[5]

Within the boundaries of these limitations, in this paper, I discuss alternative strategies for estimating VARs with panel data sets, I study the determinants of the asymptotic bias of the *fixed effect* estimator (FE), and I study the finite sample performance of the MG estimator by means of Monte Carlo simulation techniques. The main findings of the paper are that, in VAR specifications, (i) the bias introduced in pooled estimators by cross sectional heterogeneity of the slope parameters is generally different from zero, but its sign and magnitude cannot be predicted accurately; (ii) slope heterogeneity must be high to be a source of concern in finite samples; (iii) when this happens, the panel must be longer than a typical macro data set for the MG estimator to represent a viable solution to the problem.

[2]Applied works using VARs estimated with panel data sets include, among others, Attanasio et al. [1998], Holtz-Eakin, Newy, and Rosen [1988] and [1989], Carroll and Weil [1994], Mohapatra *et al.* [1996], Hoffmaister *et al.* [1997a] and [1997b], Andersen *et al.* [1997], Rebucci [1998], and Ciccarelli and Rebucci [2000] . Theoretical problems involved in moving from a standard DPM specification to a DPM one in which explanatory variables are *weakly* rather than *strongly* exogenous are discussed, for instance, in Kiviet [1998].

[3]See Hoogstrate [1998] for an analysis of identification issues in VARs estimated with panel data. See Krishnakumar [199?] for a discussion of identification through restrictions on the variance covariance matrix in simultaneous equation models estimated with panel data.

[4]See Holtz-Eakin, Newy, and Rosen [1988] for a Classical framework for estimation and inference in VARs with slope parameters that are constant across individual units, but varying across time periods. See Canova and Ciccarelli [1998] for a tentative, Bayesian framework of estimation and inference in VARs with slope parameters varying both across time periods and individual units. See Banerjee [1999] and the references quoted therein for a survey of the literature on testing for unit root and cointegration with panel data.

[5]See Judson and Owen [1999] for Monte Carlo evidence on standard DPMs estimated with long panels.

The paper is organised as follows: section 2 spells out the model and discusses alternative estimators; section 3 proves the consistency of the MG estimator and studies the bias of the FE estimator; section 4 sets up the Monte Carlo experiment and gives the finite sample results; section 5 concludes. Technical appendices and the GAUSS code for the simulation exercise follow.

2 Alternative estimators for PVARs

I consider the following heterogeneous panel of covariance stationary, mean square ergodic VARs of order 1:[6]

$$Y'_{i,t} = A'_i Y'_{i,t-1} + \alpha'_i + \varepsilon'_{i,t}, \qquad i = 1, \cdots N; \quad t = 1, \cdots T; \tag{1}$$

where $Y'_{i,t}$ is an $Mx1$ vector of variables of interest, $\varepsilon'_{i,t}$ is a $Mx1$ vector of serially and sectionally uncorrelated innovations with variance-covariance matrix (VCM) Σ_i (i.e., $\varepsilon'_{i,t} \sim iid(0, \Sigma_i)$), α'_i is an $Mx1$ vector of individual specific fixed or random effects, A'_i is an MxM matrix of individual specific slope coefficients, N is the number of cross sectional units and T is the number of time periods. Following Pesaran and Smith [1995], I assume that the A'_i matrices vary across individuals according to the following, simple random coefficient specification:

$$A'_i = A' + \eta'_i, \tag{2}$$

where A' is a MxM constant matrix, η'_i is a MxM random matrix distributed independently of $\varepsilon'_{i,t}$ with zero mean and constant VCM equals to Ω (i.e., $vec(\eta'_i) \sim iid(0, \Omega)$). Given the stationarity assumption, this specification allows the maximum degree of heterogenity. With this specification, the dynamic relationship among the variables of interest can differ across sectional units in the level of the $Y's$, the variability of the $Y's$, and the time series correlation pattern among the $Y's$. In other words, the PVAR in equation (1) allows each individual unit to be diffirent in all these three dimensions.

Suppose one is interested in estimating the typical or representative propagation mechanism of shocks in the sample, A', the cross sectional mean of A'_i. When T is large enough to estimate individual time series regression separately, this can be obtained in three different ways: first, by stacking

[6] The analysis in the paper can be easily generalised to a VAR of any order.

the data and using pooled estimators such as the fixed (FE) or random effects (RE) estimators, corrected for cross section heteroskedasticity in the variance of the innovations $\varepsilon'_{i,t}$ if necessary; second, by averaging data across sectional units and estimating aggregate time series regressions (ATS); third, by estimating individual time series regressions and averaging these estimates across sectional units (MG). If the panel is not only *long*, but also *homogeneous* in the slope parameters, i.e., $\eta'_i = 0$ for all i, then all these three estimation procedures give rise to consistent estimates of the matrix of parameters of interest, A', for large N . In this case, the chioce among alternative estimators ought to be dictated by efficiency considerations based on assumptions on the nature of the individual specific effects, α'_i. If the panel is *long* but *heterogeneous* in the slope parameters, as pointed out by Pesaran and Smith [1995] for standard DPM, pooled estimators and the ATS estimator yield inconsistent estimates of A', regardless of the sectional dimension of the panel. Instead, in this case, the MG estimator yields a consistent estimate of A' for both N and T large.[7]

I shall study the asymptotic bias of the FE estimator and prove the consistency of the MG estimator under slope heterogenity in the next section of the paper, however, in order to see why pooled estimator cannot be consistent in this case, substitute equation 2 in 1, then the model becomes:

$$Y'_{i,t} = A'Y'_{i,t-1} + \alpha'_i + \nu'_{i,t}, \qquad \nu'_{i,t} = \varepsilon'_{i,t} + \eta'_i Y'_{i,t-1}. \tag{3}$$

It is now clear from equation 3 that the new vector of error terms, $\nu'_{i,t}$, is correlated with the vector of regressors, $Y'_{i,t-1}$, thus rendering pooled estimators inconsistent.[8] Moreover, in a standard DPM specifications, as shown by Pesaran and Smith [1995] and Pesaran, Smith and Im [1996], there are special cases in which instrumental variables type of estimators (IV) provide a solution to this problem, while in a PVAR specification there are no such special cases; a point already noted implicitely by Holtz-Eakin *et al.* [1988].

To see why the ATS estimator cannot be consistent under slope heterogeneity, simply take the arithmetic average of equation (3) accross i: the correlation between the new aggregate disturbance terms and the regressors in the new aggregate equations is now evident, thus rendering also ordinary least square estimates of this equation inconsistent. This estimator however

[7] See Hsiao et *al.* [1997] and Andersen *et al.* [1997] for alternative, consistent estimation procedures when the panel is short and heterogeneous in the slope parameters.

[8] Note the analogy with an error in variable type of model.

is unattractive also under slope homogeneity because it does not increase the number of degree of freedom available, which is often a critical issue in estimating VARs. For these reasons, I shall not pursue this alternative estimation procedure further in the paper.[9]

Finally, note that the FE estimator is asymptotically equivalent to the RE estimator if the panel is homogeneous in the slope parameters and the individual specific effects are random, but uncorrelated with the regressor; the class of IV estimators for $DPMs$ homogeneous in the slope parameters is wide, ranging from the simple first difference estimator of Anderson and Hisiao [1986] to alternative generalised method of moment estimators, but there is no consensus yet in the literature on which is the most appropriate choice when the panel is long.[10] In the next section, therefore, I shall compare only the asymptotic properties of the FE and the MG estimators. However, I will consider a simple IV alternative in the Monte Carlo simulation exercise in the third section of the paper.

3 Asymptotic properties

In order to derive the MG and FE estimators and their properties some notation is needed. So, let us transpose equation 1 and 3 to obtain,

$$Y_{i,t} = Y_{i,t-1}A_i + \alpha_i + \varepsilon_{i,t}, \qquad i = 1, \cdots N \quad t = 1, \cdots T; \tag{4}$$

$$Y_{i,t} = Y_{i,t-1}A + \alpha_i + \nu_{i,t}, \qquad \nu_{i,t} = \varepsilon_{i,t} + Y_{i,t-1}\eta_i; \tag{5}$$

where $Y_{i,t} = [y_{i,1,t}, \cdots, y_{i,M,t}]$, $Y_{i,t-1} = [y_{i,1,t-1}, \cdots, y_{i,M,t-1}]$, $\alpha_i = [\alpha_{i,1}, \cdots, \alpha_{i,M}]$, $\varepsilon_{i,t} = [\varepsilon_{i,1,t}, \cdots, \varepsilon_{i,M,t}]$, $\nu_{i,t} = [\nu_{i,1,t}, \cdots, \nu_{i,M,t}]$, all of them of dimension $1xM$. Then define the following variables:

$$Y = \begin{bmatrix} Y_{1,1} \\ \vdots \\ Y_{1,T} \\ \vdots \\ Y_{N,T} \end{bmatrix}, Y_{-1} = \begin{bmatrix} Y_{1,0} \\ \vdots \\ Y_{1,T-1} \\ \vdots \\ Y_{N,T-1} \end{bmatrix}, \overline{\nu} = \begin{bmatrix} \nu_{1,1} \\ \vdots \\ \nu_{1,T} \\ \vdots \\ \nu_{N,T} \end{bmatrix},$$

[9] See Pesaran and Smith [1995] and the references quoted therein for further details on the inconsistency of the ATS estimator under slope heterogenity in a standard DPM context.

[10] See Arellano and Alvarez [1998] on this point.

$$\overline{\varepsilon} = \begin{bmatrix} \varepsilon_{1,0} \\ \vdots \\ \varepsilon_{1,T} \\ \vdots \\ \varepsilon_{N,T} \end{bmatrix}, \overline{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \otimes i_T = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix};$$

where Y, Y_{-1}, $\overline{\nu}$, $\overline{\varepsilon}$, and $\overline{\alpha}$ are all of dimension $NTxM$, i_T denotes a $Tx1$ vector of ones, and $\otimes$ denotes the Kronecker product. Finally, let us define the following matrix operators:

$$D = I_N \otimes i_T;$$
$$P_D = D(D'D)^{-1}D' = I_N \otimes i_T i_T'/T = I_N \otimes i_T(i_T' i_T)^{-1} i_T';$$
$$Q_D = I_{NT} - P_D = I_{NT} - D(D'D)^{-1}D' = I_N \otimes [I_T - i_T(i_T' i_T)^{-1} i_T'] = I_N \otimes H_T;$$
$$P = I_M \otimes P_D;$$
$$Q = I_M \otimes Q_D;$$

where D is the usual matrix of individual dummies, P_D is the usual between operator, Q_D is the usual within operator, P and Q generalise the latter two operators to a SUR system of equations (see Cornwell et al. [1992]), $H_T = I_T - i_T(i_T' i_T)^{-1} i_T'$, and $i_T' i_T = T$, with I denoting the identity matrix.

3.1 The mean group estimator

Grouping all time observations for each individual unit i, the model in equation 4 becomes

$$Y_i = Y_{i,-1}A_i + \alpha_i \otimes i_T + \overline{\varepsilon}_i \qquad i = 1, \cdots N; \tag{6}$$

where Y_i, $Y_{i,-1}$, and $\overline{\varepsilon}_i$ are the ith, TxM dimensional elements of Y, Y_{-1}, and $\overline{\varepsilon}$ respectively. If we apply the vec operator to this equation, and define $y_i = vec(Y_i)$, $X_i = (I_M \otimes Y_{i,-1})$, $a_i = vec(A_i)$, and $\underline{\varepsilon}_i = vec(\overline{\varepsilon}_i)$, then the model can be represented as

$$y_i = X_i a_i + \alpha_i \otimes i_T + \underline{\varepsilon}_i, \qquad i = 1, \cdots N; \tag{7}$$

where y_i, $(\alpha_i \otimes i_T)$, and $\underline{\varepsilon}_i$ are of dimension $TMx1$, X_i is of dimension $TMxM^2$, and a_i is of dimension M^2x1. Premultiplaying this equation by $(I_M \otimes H_T)$, i.e., taking deviations from the time mean, we obtain

$$\widetilde{y}_i = \widetilde{X}_i a_i + \widetilde{\underline{\varepsilon}}_i,$$

where $\widetilde{y}_i = (I_M \otimes H_T)y_i$, $\widetilde{X}_i = (I_M \otimes H_T)X_i$, $\widetilde{\underline{\varepsilon}_i} = (I_M \otimes H_T)$, and $(I_M \otimes H_T)(\alpha_i \otimes i_T) = 0$.

The *ordinary least square (OLS)* estimator of a_i for each individual unit i is given by

$$\widehat{a}_i = \left(\widetilde{X}_i{}'\widetilde{X}_i\right)^{-1}\left(\widetilde{X}_i{}'\widetilde{y}_i\right), \qquad i = 1, \cdots N;$$

and hence the *mean group* estimator (MG) of $a = vec(A)$ is given by

$$\widehat{a}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{a}_i.$$

To prove the consistency of the MG estimator, notice then that

$$\begin{aligned}
\widehat{a}_i &= \left(\widetilde{X}_i{}'\widetilde{X}_i\right)^{-1}\left(\widetilde{X}_i{}'\widetilde{y}_i\right) \\
&= \left(X_i'\left(I_M \otimes H_T\right)X_i\right)^{-1}\left(X_i'\left(I_M \otimes H_T\right)y_i\right) \\
&= \left(\left(I_M \otimes Y_{i,-1}\right)'\left(I_M \otimes H_T\right)\left(I_M \otimes Y_{i,-1}\right)\right)^{-1} \\
&\quad \times \left(\left(I_M \otimes Y_{i,-1}\right)'\left(I_M \otimes H_T\right)y_i\right) \\
&= \left(I_M \otimes Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(I_M \otimes Y_{i,-1}'H_Ty_i\right) \\
&= \begin{bmatrix} \left(Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_Ty_{i,1}\right) \\ \vdots \\ \left(Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_Ty_{i,M}\right) \end{bmatrix} \\
&= a_i + \begin{bmatrix} \left(Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_T\right)\underline{\varepsilon}_{i,1} \\ \vdots \\ \left(Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_T\right)\underline{\varepsilon}_{i,M} \end{bmatrix}
\end{aligned}$$

where $y_{i,j}$ and $\underline{\varepsilon}_{i,j}$ are the jth, $Tx1$ dimensional elements of y_i and $\underline{\varepsilon}_i$ respectively. Therefore, defining $\overline{a} = \frac{1}{N}\sum_{i=1}^{N} a_i$, for $j = 1, \cdots M$, we also have

$$\widehat{a}_{MG} = \overline{a} + \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} \left(Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_T\underline{\varepsilon}_{i,1}\right) \\ \vdots \\ \left(Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_T\underline{\varepsilon}_{i,M}\right) \end{bmatrix}.$$

Now, under usual regularity conditions for stationary dynamic models, for fixed N and $T \to \infty$, we have

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty}\left(\widehat{a}_{MG}\right) &= \overline{a} + \operatorname*{plim}_{T\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}\left(Y_{i,-1}'H_T Y_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_T\underline{\varepsilon}_{i,1}\right)\\ \vdots \\ \left(Y_{i,-1}'H_T Y_{i,-1}\right)^{-1}\left(Y_{i,-1}'H_T\underline{\varepsilon}_{i,M}\right)\end{bmatrix}\right)\\
&= \overline{a} + \frac{1}{N}\sum_{i=1}^{N}\left(\begin{bmatrix}\operatorname{plim}_{T\to\infty}\left(\frac{Y_{i,-1}'H_T Y_{i,-1}}{T}\right)^{-1}\operatorname{plim}_{T\to\infty}\left(\frac{Y_{i,-1}'H_T\underline{\varepsilon}_{i,1}}{T}\right)\\ \vdots \\ \operatorname{plim}_{T\to\infty}\left(\frac{Y_{i,-1}'H_T Y_{i,-1}}{T}\right)^{-1}\operatorname{plim}_{T\to\infty}\left(\frac{Y_{i,-1}'H_T\underline{\varepsilon}_{i,M}}{T}\right)\end{bmatrix}\right)\\
&= \overline{a}.
\end{aligned}$$

But as $N \to \infty$, by the low of large numbers, we also have

$$\overline{a}\,\underrightarrow{P}\,a,$$

as a_i is *iid* across individuals, thus establishing the consistency of the MG estimator[11].

3.2 The fixed effect estimator

Stacking all observations in panel format, the model in equation 5 can be represented as

$$Y = Y_{-1}A + \overline{\alpha} + \overline{\nu} \qquad \overline{\nu} = \overline{\varepsilon} + Y_{-1}\eta_i.$$

Applying the *vec* operator to this equation and defining $y = vec(Y)$, $X = (I_M \otimes Y_{-1})$, $a = vec(A)$, $\alpha = vec(\overline{\alpha})$, $\nu = vec(\overline{\nu})$, $\varepsilon = vec(\overline{\varepsilon})$, this can be rewritten in SUR format as

$$y = Xa + \alpha + \nu \qquad \nu = \varepsilon + Xvec(\eta_i); \tag{8}$$

where y, α, ν, and ε are of dimension $NTMx1$, X is of dimension $NTMxM^2$, and a and $vec(\eta_i)$ are of dimension M^2x1. Then, applying the within operator Q to the SUR system 8, i.e., taking deviations from the time means, we obtain

$$\widetilde{y} = \widetilde{X}a + \widetilde{\nu},$$

[11] Note that this result, following Pesaran and Smith [1995], is obtained fixing N and letting T pass to infinity and subsequently allowing N to tend to infinity. See Phillips and Moon [1999] for alternative approaches to the asymptotic theory of random fields.

where $\widetilde{y} = Qy$, $\widetilde{X} = QX$, $\widetilde{\nu} = Q\nu$, and $Q\alpha = 0$. Hence, the *fixed effect* (FE) estimate of a is given by

$$\widehat{a}_{FE} = \left(\widetilde{X}'\widetilde{X}\right)^{-1}\left(\widetilde{X}'\widetilde{y}\right). \tag{9}$$

Proving the inconsistency of the FE estimator is a bit more tedious. In the appendix at the end of the paper, I show that the asymptotic bias of the FE estimator of a is given by the following expression:

$$\underset{N\to\infty,T\to\infty}{\text{plim}} \left(\widehat{a}_{FE} - a\right) = \tag{10}$$

$$\begin{bmatrix} \left(E\left[vec^{-1}\left(\left(I - A_i' \otimes A_i'\right)^{-1} vec(\Sigma_i)\right)\right]\right)^{-1} \\ \times\left(E\left[vec^{-1}\left(\left(I - A_i' \otimes A_i'\right)^{-1} vec(\Sigma_i)\right)\eta_{i,1}\right]\right) \\ \vdots \\ \left(E\left[vec^{-1}\left(\left(I - A_i' \otimes A_i'\right)^{-1} vec(\Sigma_i)\right)\right]\right)^{-1} \\ \times\left(E\left[vec^{-1}\left(\left(I - A_i' \otimes A_i'\right)^{-1} vec(\Sigma_i)\right)\eta_{i,M}\right]\right) \end{bmatrix},$$

where $\eta_{i,j}$ is the jth, $Mx1$ dimensional element of $vec(\eta_i)$, vec^{-1} undoes the vec operation, E denotes expectations with respect to the joint distribution of A_i and Σ_i, and $j = 1, \cdots M$ indexes equations. Under the stationarity assumption, the expectations in this equation are well defined and the asymptotic bias turns out to be generally different from zero.

In principle, an explicit solution for the asymptotic bias of the FE estimator can be obtained computing these expectations under suitable distributional assumptions for A_i and Σ_i. In practice, even the simplest, general VAR specification has no closed-form solution. Consider, for instance, the following bivariate VAR with only one source of slope heterogeneity and without heteroskedaticity:

$$\begin{aligned} z_{i,t} &= \lambda_i z_{i,t-1} + \beta_i x_{i,t-1} + \alpha_i + u_{i,t}, \\ x_{i,t} &= \gamma_i z_{i,t-1} + \rho_i x_{i,t-1} + \mu_i + v_{i,t}; \end{aligned} \tag{11}$$

where

$$\begin{aligned} Y_{i,t}' &= \begin{bmatrix} z_{i,t} \\ x_{i,t} \end{bmatrix}, A_i' = \begin{bmatrix} \lambda_i & \beta_i \\ \gamma_i & \rho_i \end{bmatrix}, A' \begin{bmatrix} \lambda & \beta \\ \gamma & \rho \end{bmatrix}, \eta_i' = \begin{bmatrix} 0 & \xi_i \\ 0 & 0 \end{bmatrix} \\ \alpha_i' &= \begin{bmatrix} \alpha_i \\ \mu_i \end{bmatrix}, \varepsilon_{i,t}' = \begin{bmatrix} u_{i,t} \\ v_{i,t} \end{bmatrix}, \text{ and } \Sigma_i = \Sigma = \begin{bmatrix} \sigma^2 & \phi \\ \phi & \tau^2 \end{bmatrix}. \end{aligned}$$

It is easily seen that

$$(I - A_i' \otimes A_i') = \begin{bmatrix} 1-\lambda^2 & -\lambda\beta_i & -\lambda\beta_i & -\beta_i^2 \\ -\lambda\gamma & 1-\lambda\rho & -\beta_i\gamma & -\beta_i\rho \\ -\lambda\gamma & -\beta_i\gamma & 1-\lambda\rho & -\beta_i\rho \\ -\gamma^2 & -\gamma\rho & -\gamma\rho & 1-\rho^2 \end{bmatrix}.$$

It can also be shown that the inverse of this matirx given by

$$(I - A_i' \otimes A_i')^{-1} = \frac{1}{\Upsilon_{i,0}} \begin{bmatrix} \Upsilon_{i,1,1} & \Upsilon_{i,1}\beta_i & \Upsilon_{i,1}\beta_i & \Upsilon_{i,2}\beta_i^2 \\ \Upsilon_{i,1}\gamma & \Upsilon_{i,2,2} & \Upsilon_{i,2}\beta_i\gamma & \Upsilon_{i,3}\beta_i \\ \Upsilon_{i,1}\gamma & \Upsilon_{i,2}\beta_i\gamma & \Upsilon_{i,2,2} & \Upsilon_{i,3}\beta_i \\ \Upsilon_{i,2}\gamma^2 & \Upsilon_{i,3}\gamma & \Upsilon_{i,3}\gamma & \Upsilon_{i,3,3} \end{bmatrix};$$

where:

$$\begin{aligned} \Upsilon_{i,0} &= (1-\beta_i\gamma-\lambda-\rho+\lambda\rho)(1-\beta_i\gamma+\lambda+\rho+\lambda\rho)(\lambda\rho-\beta_i\gamma-1); \\ \Upsilon_{i,1} &= \left(\lambda\rho^2-\rho\beta_i\gamma-\lambda\right); \\ \Upsilon_{i,2} &= -(\lambda\rho-\beta_i\gamma+1); \\ \Upsilon_{i,3} &= \left(\lambda^2\rho-\lambda\beta_i\gamma-\rho\right); \\ \Upsilon_{i,1,1} &= \left(-\lambda\rho^3+\rho^2+\rho^2\gamma\beta_i+\lambda\rho+\beta_i\gamma-1\right); \\ \Upsilon_{i,2,2} &= \left(-\lambda^2\rho^2+\lambda\beta_i\rho\gamma+\rho^2+\lambda^2+\beta_i\gamma-1\right); \\ \Upsilon_{i,3,3} &= \left(-\lambda^3\rho+\lambda^2+\lambda^2\gamma\beta_i+\lambda\rho+\beta_i\gamma-1\right). \end{aligned}$$

Then, considering only the first equation with $\eta_{i,1}' = \begin{bmatrix} 0 & \xi_i \end{bmatrix}$, it is easily seen that

$$vec^{-1}\left((I - A_i' \otimes A_i')^{-1} vec(\Sigma_i)\right) =$$

$$\frac{1}{\Upsilon_{i,0}} \begin{bmatrix} (\Upsilon_{i,1,1}\sigma^2 + 2\Upsilon_{i,1}\beta\phi + \Upsilon_{i,2}\beta^2\tau^2) & (\Upsilon_{i,1}\gamma\sigma^2 + \Upsilon_{i,2,2}\phi + \Upsilon_{i,2}\beta\gamma\phi + \Upsilon_{i,3}\beta\tau^2) \\ (\Upsilon_{i,1}\gamma\sigma^2 + \Upsilon_{i,2,2}\phi + \Upsilon_{i,2}\beta\gamma\phi + \Upsilon_{i,3}\beta\tau^2) & (\Upsilon_{i,2}\gamma^2\sigma^2 + 2\Upsilon_{i,3}\gamma\phi + \Upsilon_{i,3,3}\tau^2) \end{bmatrix}$$

and that

$$vec^{-1}\left((I - A_i' \otimes A_i')^{-1} vec(\Sigma_i)\right)\eta_{i,1} =$$

$$\frac{1}{\Upsilon_{i,0}} \begin{bmatrix} (\Upsilon_{i,1}\gamma\sigma^2 + \Upsilon_{i,2,2}\phi + \Upsilon_{i,2}\beta\gamma\phi + \Upsilon_{i,3}\beta\tau^2)\,\xi_i \\ (\Upsilon_{i,2}\gamma^2\sigma^2 + 2\Upsilon_{i,3}\gamma\phi + \Upsilon_{i,3,3}\tau^2)\,\xi_i \end{bmatrix}.$$

Hence, in this simple case, the asymptotic bias of the FE estimator of $\begin{bmatrix} \lambda & \beta \end{bmatrix}'$ in eqaution 11 is given by

$$\operatorname*{plim}_{N\to\infty,T\to\infty} \begin{pmatrix} \widehat{\lambda}_{FE}-\lambda \\ \widehat{\beta}_{FE}-\beta \end{pmatrix} =$$

$$\left(E\left\{ \frac{1}{\Upsilon_{i,0}} \begin{bmatrix} \Upsilon_{i,1,1}\sigma^2+2\Upsilon_{i,1}\beta_i\phi+\Upsilon_{i,2}\beta_i^2\tau^2 & \Upsilon_{i,1}\gamma\sigma^2+\Upsilon_{i,2,2}\phi+\Upsilon_{i,2}\beta_i\gamma\phi+\Upsilon_{i,3}\beta_i\tau^2 \\ \Upsilon_{i,1}\gamma\sigma^2+\Upsilon_{i,2,2}\phi+\Upsilon_{i,2}\beta_i\gamma\phi+\Upsilon_{i,3}\beta_i\tau^2 & \Upsilon_{i,2}\gamma^2\sigma^2+2\Upsilon_{i,3}\gamma\phi+\Upsilon_{i,3,3}\tau^2 \end{bmatrix} \right\} \right)^{-1} \tag{12}$$

$$\times \left(E\left\{ \frac{1}{\Upsilon_{i,0}} \begin{bmatrix} \left(\Upsilon_{i,1}\gamma\sigma^2+\Upsilon_{i,2,2}\phi+\Upsilon_{i,2}\beta_i\gamma\phi+\Upsilon_{i,3}\beta_i\tau^2\right)\xi_i \\ \left(\Upsilon_{i,2}\gamma^2\sigma^2+2\Upsilon_{i,3}\gamma\phi+\Upsilon_{i,3,3}\tau^2\right)\xi_i \end{bmatrix} \right\} \right).$$

But this equation cannot be simplified further without additional assumptions because it involves highly non-linear functions of the random variable β_i. In the case of a generic VAR specification, therefore, it is impossible to predict the precise sign and magnitude of the asyntotic bias of the FE estimator and to study its determinants.

An explicit solution for the asymptotic bias of the FE estimator, however, can be obtained (at least) in two special cases of interest: first, in order to study the role of ϕ, by assuming that $\gamma = 0$ for any distribution of ξ_i —i.e., assuming that z *does not* cause x in Granger sense and thus that x is (still) strongly exogenous for the estimation of λ and β; and second, in order to see the consequences of relaxing the assumption of strong exogenity of x for the estimation of β, by assuming that $\lambda = \rho = 0$, $\phi = 0$, and ξ_i is uniformly distributed.

3.2.1 Case 1

Assuming $\gamma = 0$ in 11, the model becomes

$$\begin{aligned} z_{i,t} &= \lambda z_{i,t-1}+\beta_i x_{i,t-1}+\alpha_i+u_{i,t}, \\ x_{i,t} &= \rho x_{i,t-1}+\mu_i+v_{i,t}. \end{aligned} \tag{13}$$

In the appendix at the end of the paper, I show that, in this case,

$$\operatorname*{plim}_{N\to\infty,T\to\infty} \begin{pmatrix} \widehat{\lambda}_{FE}-\lambda \\ \widehat{\beta}_{FE}-\beta \end{pmatrix} = \begin{bmatrix} \frac{\rho(1-\lambda\rho)\left(1-\lambda^2\right)\omega}{\Psi_1+\Psi_2} \\ -\frac{\beta\rho^2\left(1-\lambda^2\right)\omega+\Psi_3}{\Psi_1+\Psi_2} \end{bmatrix} \tag{14}$$

where:

$$\begin{aligned}
\Psi_1 &= (\sigma^2/\tau^2)\left(1-\rho^2\right)(1-\lambda\rho)^2 + \left(1-\lambda^2\rho^2\right)\omega + \left(1-\rho^2\right)\beta^2; \\
\Psi_2 &= -(\phi^2/\tau^2)\left(1-\rho^2\right)\left(1-\lambda^2\right) - 2(\phi/\tau)\left(1-\rho^2\right)(1-\lambda)\beta; \\
\Psi_3 &= (\phi/\tau)\left(1-\rho^2\right)\left(1-\lambda^2\right)\rho\omega.
\end{aligned}$$

The size of the asymptotic bias of the FE estimator, in this case, depends not only upon the mean coefficients (λ, β, ρ), the variance of β_i (ω) and the ratio (σ^2/τ^2), as in the standard DPM case analysed by Pesaran and Smith [1995], but also on the sign and the magnitude of ϕ. Moreover, it can be shown (see appendix) that both $\widehat{\beta}_{FE}$ and $\widehat{\lambda}_{FE}$ may over or underestimate the true values of β and λ depending on the sign of ρ and ϕ, and the magnitude of ϕ relative to the absolute value of $(2\beta\tau/(1+\lambda))$.[12]

3.2.2 Case 2

Suppose that $\lambda = \rho = 0$, i.e., z *does* cause x in Granger sense and hence x is only weakly exogenous for the estimation of β. Under this hypothesis, the model becomes:

$$\begin{aligned}
z_{i,t} &= \beta_i x_{i,t-1} + \alpha_i + u_{i,t;} \\
x_{i,t} &= \gamma z_{i,t-1} + \mu_i + v_{i,t}.
\end{aligned} \tag{15}$$

Substituting $\lambda = \rho = 0$ in equation 12, further assuming without loss of generality that $\sigma^2 = \tau^2 = 1$ and simplifying the resulting expression, it is easily seen that:

$$\operatorname*{plim}_{N\to\infty,T\to\infty} \begin{pmatrix} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\beta}_{FE} - \beta \end{pmatrix} =$$

$$\frac{1}{\Delta'} \begin{bmatrix} E\left(\frac{(1+\gamma^2)}{\left(1-\beta_i^2\gamma^2\right)}\right) E\left(\frac{\phi\xi_i}{(1-\beta_i\gamma)}\right) - E\left(\frac{\phi}{(1-\beta_i\gamma)}\right) E\left(\frac{\left(1+\gamma^2\right)\xi_i}{\left(1-\beta_i^2\gamma^2\right)}\right) \\ E\left(\frac{(1+\beta_i^2)}{\left(1-\beta_i^2\gamma^2\right)}\right) E\left(\frac{\left(1+\gamma^2\right)\xi_i}{\left(1-\beta_i^2\gamma^2\right)}\right) - E\left(\frac{\phi}{(1-\beta_i\gamma)}\right) E\left(\frac{\phi\xi_i}{(1-\beta_i\gamma)}\right) \end{bmatrix} \tag{16}$$

[12] Note that, further assuming $\phi = 0$, we obtain the standard DPM result of Pesaran and Smith [1995]; thus, that the size of the asymptotic bias of the FE estimator depends upon: (i) on the mean coefficients λ, β, ρ; (ii) the variance of β_i, denoted $\omega_{2,2}$; (iii) and the ratio (σ^2/τ^2), with $\widehat{\beta}_{FE}$ always underestimating β and $\widehat{\lambda}_{FE}$ over or underestimatimating λ depending on wether ρ is positive or negative.

where

$$\Delta' = E\left(\frac{(1+\beta_i^2)}{(1-\beta_i^2\gamma^2)}\right)E\left(\frac{(1+\gamma^2)}{(1-\beta_i^2\gamma^2)}\right) - \left\{E\left(\frac{\phi}{(1-\beta_i\gamma)}\right)\right\}^2.$$

When $\lambda = \rho = 0$, in a standard DPM specification, the asymptotic bias of the FE estimator disappears, as can it be seen substituting this assumption in equation 14 above. In a VAR specification, instead, it does not. Under stationarity, which requires $\left|\sqrt{\beta_i\gamma}\right| < 1$, the expectations in equation 16 are well defined and generally different from zero, unless $\xi_i = 0$ for all i. Moreover, further assuming that $\gamma = 1$, $\phi = 0$, and that ξ_i is uniformly distributed, it can be shown (see appendix) that the bias of $\widehat{\lambda}_{FE}$ vanishes, while that of $\widehat{\beta}_{FE}$ is always positive and increasing in the variance of ξ_i (and increasing in β) for given β (for given variance of ξ_i), vanishing only if $\beta = 0$ or $\xi_i = 0$ for all i.

4 Small sample properties

This section looks at Monte Carlo simulation evidence for the special case 2 discussed at the end of the previous section. This case is particularly interesting because it helps us analysing both distinguishing features of a VAR specification, namely the contemporaneous correlation between the variables of interest and their lagged interdependence, while mantaining full control over the Monte Carlo experiment. Richer VAR specifications (e.g., λ and $\rho \neq 0$ and/or multiple sources of heterogeneity) would be more realistic, but the validity of the Monte Carlo results would diminish because it would be practically unworkable to control for all the features of the model potentially affecting the outcomes of the experiment. (Cf. Hendry [1984] and Devidson and MacKinnon [1993, Chap. 21].)

4.1 Experiment design

I consider the following particular parametrization of the model in equation 15:

$$\beta_i = \beta + \xi_i;$$

$$\xi_i \sim Uniform[\pm\omega(1-\beta)], \qquad 0 \leq \omega < 1 \qquad 0 \leq \beta < 1;$$

$$\gamma = 1;$$

$$\beta = \{0.2; 0.8\};$$

$$\omega = \{0; 0.2; 0.8\};$$

$$(N, T) = \{(50, 50); (20, 50); (50, 20); (20, 20); (10, 50)\};$$

$$\begin{bmatrix} u_{i,t} \\ v_{i,t} \end{bmatrix} \sim NIID(0, \Sigma), \qquad \Sigma = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}, \qquad \phi = \{0; \pm 0.9\};$$

$$\begin{bmatrix} z_{i,0} \\ x_{i,0} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$\begin{array}{c} \alpha_i \sim NIID(1,1) \\ \mu_i \sim NIID(1,1) \end{array}.$$

Following Pesaran, Smith, and Im [1996], I introduce only one source of slope heterogeneity in the model, i.e. $\xi_i \sim Uniform[\pm\omega(1-\beta)]$, but unlike them I use the uniform rather then the normal distribution to characterise the cross sectional distribution of ξ_i. The uniform distribution allows control of the degree of slope heterogeneity introduced in the model, through the scale parameter ω, while guaranteeing that no individual unit violates the stationarity assumption as long as ω and β are both less than 1.[13] If $\xi_i \sim Uniform[\pm\omega(1-\beta)]$, then β_i is also uniformly distributed with mean β and variance $\frac{\omega^2(1-\beta)^2}{12}$ respectively. Therefore, ω controls the dispersion of the cross sectional distribution around a given β, and thus the degree of slope heterogeneity introduced in the model for given average persistence, which is is minimal for $\omega = 0$ and maximal for $\omega = 1$, always ensuring that both individual eigenvalues are less than 1 in absolute value.

If $\lambda = \rho = 0$, the eigenvalues of the individual VAR systems are given by $\pm\sqrt{\gamma\beta_i}$. Stationarity requires that $\left|\pm\sqrt{\gamma\beta_i}\right| < 1$ and constraints the range of variation of β_i for given γ. Somewhat arbitrarily, I maintain $\gamma = 1$ throughout the experiment and let β_i vary in the interval $[\pm 1]$.

From the asymptotic analysis we know that the heterogenity bias of pooled estimator depends on the average degree of persistence in the sample and the degree of heteorgeneity of the slope parameters. Choosing $\beta = \{0.2; 0.8\}$, means considering average charachteristic roots equal to

[13] Hsiao, Pesaran, and Tahmiscioglu [1997] use the truncated normal distribution rather than the normal in their Monte Carlo experiment in order to avoid explosive (or unstable) simulated series. Given that I used the uniform distribution in the asymptotic analysis, it seems preferable to conduct the Monte Carlo experiment with the same distributional assumption.

± 0.45 and ± 0.89 respectively: a relatively low and relatively high degree of persistence. Choosing $\omega = \{0; 0.2; 0.8\}$, means considering the homogeneity case, a case of relatively low slope heterogeneity, and a case of relatively high slope heterogeneity.[14]

I assume an homogeneous VCM of the error terms, and I set $\sigma^2 = \tau^2 = 1$. The choice of an homogeneous VCM is dictated by the need to asses the influence of ϕ on the finite sample bias of the estimates in insulation from the possible role of heteroskedasticity. I set $\sigma^2 = \tau^2 = 1$, so that ϕ doeas not determine only the covariance between x and z, but also their correlation which is bounded to lie between -1 and 1. Looking at $\phi = \{0; \pm 0.9\}$, I consider the case of uncorrelated error terms and the cases of highly correlated error terms, either positively or negatively, to highlight clearly the potential effects of this feature of the model of finite sample properties of the estimates.

I examine typical dimensions of a macro panel data set (i.e., $(N, T) = \{(50, 50); (20, 20); (20, 50); (20, 50)\}$) and one extreme case (i.e., $(N, T) = \{(10, 50)\}$) in order to control for situations in which there are very few individual units likely to arise working with subgroups of individuals.

Finally, the vector of error terms, $\left[\begin{array}{cc} u_{i,t} & v_{i,t} \end{array} \right]'$, is generated from a bivariate normal distribution with VCM Σ. The initial conditions are set to zero, and a standard assumption is made to generate the individual effects α_i and μ_i. For each iteration, $50 + T$ observations are generated and than only T are used to compute the estimates. Each experimental run is based on 1000 replications and different runs start from the same seed.[15]

4.2 Monte Carlo results

Tables 1 through 5 report the results of the Monte Carlo experiment. The experiment consists of 90 runs or different cases (5 panel dimensions, times 2 degrees of persistence, times 3 degrees of heterogeneity, times 3 degrees of contemporaneous covariance). Each table reports the results for a different panel dimension (Table 1, $(N, T) = (50, 50)$; Table 2, $(N, T) = (20, 50)$; Table 3, $(N, T) = (10, 50)$; Table 4, $(N, T) = (50, 20)$; Table 5, $(N, T) = (20, 20)$). Heterogeneity increases from left to right ($\omega = 0, 0.2, 0.8$), and the contemporaneous correlation of the error terms varies from top to botton ($\phi = 0, 0.9, -0.9$). Persistence is relatively low ($\beta = 0.2$) in the upper part of

[14] See Table 6 for a summary of the implications of these assumptions.

[15] Gauss code available upon request.

the tables and is relatively high in the lower part ($\beta = 0.8$). For each run of the experiment, the tables report the estimated parameters (λ and β), their estimated standard errors (*s.e.*), the absolute value of the finite sample bias (*bias*), which is equal to the estimated parameter value in the case of λ, their experimental standard deviations (*s.d.*), and, for β only, the finite sample bias as a percentage of the true value of β (*Fbias as % of true value*). When applicable, the asymptotic bias of β as a percentage of the true value is also reported (*Abias* as % of true value).

4.2.1 Homogeneous panels

In the benchmark case of a homgeneous, large and long panel data set with relatively low persistence and no correlation between the error terms (see upper left corner of Table 1), the IV estimator does quite well with very small finite sample bias and standard errors that, though considerably higher, are of the same order of magnitude than those of the FE and the MG estimates. The FE estimator too performs well in this benchmark case, even though, as expected, the finite sample biases of β and λ are of one and two orders of magnitude larger than those of the IV estimator respectively. The MG estimator, in this case, scores as well as the FE estimator in terms of efficiency and finite sample bias of the estimate of λ, but it clearly underperforms the FE estimator in terms of bias of β: the MG underestimates the true value of β by more than 14 percent even when T=50, while the downward bias of the FE is only about 7 percent in this case.

Decreasing $N = 20$ for fixed $T = 50$ does not affect these results (see Tables 2 and 3); while decreasing $T = 20$ for fixed $N = 50$ has a much larger impact (see Table 4): the FE's bias of β increases to more than 15 percent of the true value and that of λ moves from -0.02 to -0.06 in absolute value; the MG's bias of β shoots up to more than 30 percent of the true value and that of λ rises similarly from -0.03 to -0.06 in absolute value.

Somehow surprisingly, the introduction of a correlation between the error terms in the benchmark, hogeneous case above ($N, T = 50, 50$) considerably affects the MG and the FE estimates of both λ and β, albeit in a different way: the bias of λ is smaller (larger) in absolute value than the case in which $\phi = 0$ if $\phi > 0$ ($\phi < 0$); the bias of β is always larger than the case in which $\phi = 0$, and even more so when $\phi < 0$. The IV estimates of λ and β are also affected by $\phi \neq 0$ in a similar way, but the magnitude of this effect is much smaller and practically insignificant. Note that these are fairly robust results that do not change if we increase persistence (see below) and/or change the dimension of the panel. Experimenting with larger time dimensions everything else equal, i.e., $T = 100$ and $T = 200$, it was possible to establish that we would need at least 70-75 time observations to bring the MG's bias down to below 10 percent of the true value of β with $\phi = 0$, and more than 100 observations to bring it below 10 percent with $\phi = -0.9$. Only 60-70 time observations would be needed, instead, to get the bias of the FE estimator of β down to below 10 percent of the true value even with $\phi = -0.9$.[16]

Increasing persistence, rising β from 0.2 to 0.8 (see lower part of Table 1), reduces considerably the bias of the FE and the MG estimators without affecting their efficiency. The standard errors of the IV estimates, instead, increase dramatically with persistence. Decreasing N=20 for fixed T=50, with relatively high persistence (see lower part of Table 2), does not affect the results for FE and the MG estimators, but exacerbates the inefficiency of the IV estimator, rendering the estimated β insignificant; while reducing T=20 for fixed N=50 (see lowerr part of Table 4) pushes the biases of the FE and MG estimators back to their benchmark values under low persistence and renders the IV estimator not only inefficient but also as inconsistent as the MG.

In summary, these Monte Carlo results bear out a well known conclusion in the literature on DPMs, and help qualifying this in the case of a PVAR specification: there is a trade-off between consistency and efficiency in estimating homogeneous dynamic models with panel data which suggests of using IV type of estimators when the panel is relatively short and FE or RE type of estimators when the panel is relatively long (e.g., T>20-30)[17]. When working with a VAR specification, however, one should not disregard the samll sample bias on the coefficients apart than that on the lagged dependent variable as negligible; second, and most importantly, the number of time observations needed to get rid of the small sample bias of the FE or RE estimates is probably larger that generally reccommended, given that the VCM of the error terms is unlikely to be diagonal; third, the number of time observations required to be able to neglect the small sample bias of these estimators appears to depend on the degree of persistance at system

[16] Results not reported, but available upon request.

[17] See Judson and Owen [1999] for a Monte Carlo study of alternative estimation procedures for long, homogeneous DPMs.

level rather than only on the average value of the coefficient of the lagged dependent variable. By pushing up the inefficiency of IV type of estimators and pushing down the inconsistency of FE and RE type of estimators, for a given T, higher persistance may tilt the balance in favor of the latter and hence reduce the minimum number of time observations needed.

4.2.2 Heterogeneous panels

Interestingly, under relatively low heterogeneity, the results are generally very close to those under homogeneity (see the second three columns of each table). Thus, suggesting that heterogeneity must be high to be a serious source of concern in finite samples. Under relatively high heterogeneity, instead, as expected, the small sample bias of pooled estimators of both λ and β may be sizable (see the last three columns of each table).

In the benchmark case of a heterogeneous, large and long panel data set with relatively low persistence and uncorrelated error terms (see upper part of Table 1), the IV estimator does particularly badly: its biases are larger than those of the FE estimator and its standard errors are bigger than those of the MG estimator. The MG estimator does quite well in this case with biases less than half those of the FE estimator and higher standard errors than the FE only for λ. Note however that, even when the number of individual units is small (N=20 for given T=50), the MG estimates are precise enough to distinguish between the significance of β and the insignificance of λ (see Table 2).The FE estimator lies between the MG and the IV estimator, with a finite sample bias of approximately 30 percent of the true β. The asymptotic theory predicts a positive heterogeneity bias of 48 percent of the true β in this case; while the small sample bias is only about 60 percent of the large sample bias. This is partly because of the small T bias which is sizable and of opposite sign even for T=50 as we saw above. The benchmark FE estimate of λ is equal to -0.05 as compared with a theoretical heterogeneity bias equal to zero and a finite T bias under homogeneity equal to -0.025. The presence of slope heterogeneity, therefore, appears to exacerbate the (negative) small T bias of the FE estimate of λ and possibly β.

Similar results are obtained introducing correlation between the error terms, with the MG estimator performing better than the FE estimator, which in turn improves upon the IV estimator. We have no theoretical benchmark values for the heterogeneity bias of the FE estimator when $\phi \neq 0$, however, it appears that introducing correlation between the error terms compounds the effects of the small T bias on both λ and β. Thus, suggesting that the heterogeneity bias has the same (negative) sign of the small T bias in this case.

A smaller N =20 for fixed $T = 50$, as already noted, does little difference to the performance of the MG estimator. However, a very small N (e.g., equal 10) does affect the efficiency of the estimates obtained considerably (cf. Table 2 and 3). A shorter T =20 for fixed N creates more serious problems, especially for the estimation of β (see Table 4). The small T bias of the MG estimator is about 30 percent of the true value of β when $\phi = 0$, and exceeds 60 percent when $\phi = -0.90$. On the other hand, when $\phi = 0$, the small T bias of the FE estimator is large enough to offset most of the heterogeneity bias, yielding an overall small sample bias that is less than 10 percent of the true β. The performance of the FE estimator, however, deteriorates sharply once correlation between the error terms is introduced, possibly because of a switch in the heterogeneity bias when $\phi \neq 0$. The performance of the IV estimator does not deteriorate further–as compared with the benchmark heterogenous case–by shortening the time dimension of the panel, with or without correlated error terms; but it does not improve either. The IV estimator therefore is still of no help in this case. In brief, none of the estimators considered give satisfactory results if the panel is heterogeneous and relatively short.

All three estimation procedures show lower finite sample biases when persistence is higher (see bottom right part of Table 1). The FE and the MG estimators also have somewhat lower standard errors in this case; while the efficiency of the IV estimator deteriorates further, compared to the case of low persistence, yielding a misleading estimate of β. Interestingly, in this case, the FE estimator behaves slightly better than the MG estimator even in terms of consistency. For instance, the FE bias of β is about 4 percent of the true value when $\phi = 0$–compared with a theoretical value of 6.2 percent, and about 2 percent of the true value when $\phi = 0.9$: roughly 40 and 80 percent less respectively than the MG's small T bias. This is due in part to the fact that, when $\beta = \omega = 0.8$, the absolute value of the variance of β_i is one order magnitude smaller than that implied by $\beta = 0.2$ and $\omega = 0.8$ (see summary table in appendix). But also to the fact that the heterogeneity bias of the FE estimator becomes positive when β increases from 0.2 to 0.8, and hence offsetting rather than compounding the effect of the FE's small T bias.

As in the case of low persistence, a smaller N=20 for fixed T=50 decreases the efficiency of the FE and the MG estimates, but leaves their biases almost unchanged (see bottom right of Table 2). Decreasing T=20 for fixed N=50 increases their small T bias enough to offset completely the (arguably positive) heterogeneity bias of the FE estimator and to push the bias of the MG estimator well above 10 percent of true value of β regardless of the value of ϕ (see bottom right part of Table 4). As a result, the FE estimator does remarkably better than the MG in this case notwithstanding the relatively high degree of slope heterogeneity. If either of the two panel dimensions is decreased with high persistence, the IV estimates of both λ and β become misleading (see bottom part of Table 2 and 4), and they turn out to be clearly implausible if both panel dimensions are relatively small (see table 5).

In summary, IV type of estimators can yield very misleading results if the panel is heterogeneous: they are not only inefficient, but also badly inconsistent. The performance of fixed and random effects estimators depend on the time dimension of the panel, the degree of average persistence, and the properties of the VCM of the error terms. Therefore, it is difficult to formulate recommendations with general validity. Nonetheless, they may produce better estimates than the MG in some points of the parameter space, even under relatively high heterogeneity. The MG is a safe bet when heterogeneity is high and T is very large. However, if T is not long enough, the MG risks solving one problem by creating another one of equal magnitude and opposite sign. When the panel is heterogeneous and short, there is no obvious solution to the problem posed by slope heterogeneity; a Bayesian approach to slope heterogeneity, as pursued for instance by Hisiao, Pesaran, and Tahmiscioglu [1997] for DPMs and Canova and Ciccarelli [1999] for PVARs, seems to be the only way forward. Because of computational constraints, trying to correct the MG estimator for the small T bias, as done by Pesaran and Zhao [1997] and advocated by Judson and Owen [1999], does not sound to be a viable solution dealing with VAR specifications.

5 Conclusions

Applied researchers estimate VARs with panel data relying on known asymptotic and finite sample results for DPMs. This paper shows that with a VAR specification things are more complicated: the choice of the right estimation technique to use depends crucially on the time dimension of the data set, the dispersion of the cross sectional distribution of the slope parameters around their mean, the average degree of persistence in the system, and the properties of the variance covariance matrix of the error terms.

The asymptotic analysis suggests that: (i) there are no meaningful special cases in which the heterogeneity bias of standard pooled estimators vanishes; (ii) the covariance of the error terms may add or subtract to the magnitude of the heterogeneity bias depending on its own sign and magnitude. The Monte Carlo experiment indicates that: (iii) the dispersion of the slope parameters around their sample mean must be high in absolute terms for the heterogeneity bias of pooled estimators to be substantial in finite samples; (iv) on the other hand, the time dimension of the panel must be longer than generally thought for the small T bias of the mean group estimator to be negligible when the covariance of the error terms is different from zero. The Monte Carlo experiment has shown also that: (v) a few individual units are sufficient to obtain relatively efficient MG estimates, and that (vi) IV type of estimators are particularly vulnerable to slope heterogeneity and/or high persistence (i.e., characteristic roots close to the unit circle), but they perform very well if the panel is relatively homogeneous and persistence is low.

These results suggest using the MG estimator only when slope heterogeneity is relatively high and the time dimension of the panel is very long. However, how heterogeneous a panel data set must be to become a source of concern, and how long the panel must be for the mean group estimator to represent a valid solution, remain an empirical questions given that the actual size of the small sample biases will depend on the non linear interaction of a large number of parameters.

A Inconsistency of the fixed effect estimator

A.1 VAR: The general case

From equation 9 in the text, we know that the *fixed effect* estimate of a is given by

$$\begin{aligned} \widehat{a}_{FE} &= \left(\widetilde{X}'\widetilde{X}\right)^{-1}\left(\widetilde{X}'\widetilde{y}\right) \\ &= \left(X'QX\right)^{-1}\left(X'Qy\right), \end{aligned}$$

as Q is symmetric and idempotent. Substituting for y from 8 in this equation, we obtain:

$$\left(\widehat{a}_{FE} - a\right) = \left(X'QX\right)^{-1}\left(X'Q\nu\right);$$

where

$$\begin{aligned} X'QX &= \left(I_M \otimes Y_{-1}'\right)\left(I_M \otimes Q_D\right)\left(I_M \otimes Y_{-1}'\right) \\ &= \left(I_M \otimes Y_{-1}'Q_D Y_{-1}\right) \\ &= \left(I_M \otimes Y_{-1}'\left(I_N \otimes H_T\right)Y_{-1}\right) \\ &= I_M \otimes \sum_{i=1}^{N} Y_{i,-1}' H_T Y_{i,-1}, \end{aligned}$$

with $Y_{i,-1}$being the ith, TxM-dimensional element of $Y_{-1;}$ and

$$\begin{aligned} X'Q\nu &= \left(I_M \otimes Y_{-1}'\right)\left(I_M \otimes Q_D\right)\nu \\ &= \left(I_M \otimes Y_{-1}'Q_D\right)\nu \\ &= \left(I_M \otimes Y_{-1}'Q_D\right)\begin{bmatrix} \nu_1 \\ \vdots \\ \nu_M \end{bmatrix} \\ &= \begin{bmatrix} Y_{-1}'Q_D\nu_1 \\ \vdots \\ Y_{-1}'Q_D\nu_M \end{bmatrix} = \begin{bmatrix} Y_{-1}'\left(I_N \otimes H_T\right)\nu_1 \\ \vdots \\ Y_{-1}'\left(I_N \otimes H_T\right)\nu_M \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^{N} Y_{i,-1}' H_T \nu_{i,1} \\ \vdots \\ \sum_{i=1}^{N} Y_{i,-1}' H_T \nu_{i,M} \end{bmatrix}, \end{aligned}$$

with $\nu_j = \begin{bmatrix} \nu_{1,j,1} & \cdots & \nu_{1,j,T} & \cdots & \nu_{N,j,T} \end{bmatrix}'$ being the jth, $NTx1$-dimensional element of ν and $\nu_{i,j} = \begin{bmatrix} \nu_{i,j,1} & \cdots & \nu_{i,j,T} \end{bmatrix}'$ being the ith, $Tx1$-dimensional element of ν_j for $j = 1, \cdots M$.[18][19]. Therefore,

$$(\widehat{a}_{FE} - a) = \begin{bmatrix} \left(\sum_{i=1}^{N} Y_{i,-1}' H_T Y_{i,-1}\right)^{-1} \left(\sum_{i=1}^{N} Y_{i,-1}' H_T \nu_{i,1}\right) \\ \vdots \\ \left(\sum_{i=1}^{N} Y_{i,-1}' H_T Y_{i,-1}\right)^{-1} \left(\sum_{i=1}^{N} Y_{i,-1}' H_T \nu_{i,M}\right) \end{bmatrix}. \tag{17}$$

In order to obtain the probability limit of the various terms in equation 17, we first group all time observations for each $i = 1, \cdots N$, as in equation 6 in the text here reported,

$$Y_i = Y_{i,-1} A_i + \alpha_i \otimes i_T + \overline{\varepsilon}_i, \tag{18}$$

where Y_i, $Y_{i,-1}$,and $\overline{\varepsilon}_i$ are the ith -TxM dimensional- elements of Y, Y_{-1}, and $\overline{\varepsilon}$ respectively. Given that each VAR is stationary, assuming further that the

[18] Suppose $N = 2$, then

$$\begin{aligned} Y_{-1}' \left(I_N \otimes H_T\right) Y_{-1} &= \begin{bmatrix} Y_{1,-1}' & Y_{2,-1}' \end{bmatrix} \begin{bmatrix} H_T & 0 \\ 0 & H_T \end{bmatrix} \begin{bmatrix} Y_{1,-1} \\ Y_{2,-1} \end{bmatrix} \\ &= Y_{1,-1}' H_T Y_{1,-1} + Y_{2,-1}' H_T Y_{2,-1}. \end{aligned}$$

[19] Note that $\nu_{i,j} = \varepsilon_{i,j} + Y_{i,-1}\eta_{i,j}$, where $\eta_{i,j}$ is the jth, $Mx1$ dimensional element of $vec(\eta_i)$, with $\varepsilon_j = \begin{bmatrix} \varepsilon_{1,j,1} & \cdots & \varepsilon_{1,j,T} & \cdots & \varepsilon_{N,j,T} \end{bmatrix}'$ being the jth, $NTx1$-dimensional element of ε and $\varepsilon_{i,j} = \begin{bmatrix} \varepsilon_{i,j,1} & \cdots & \varepsilon_{i,j,T} \end{bmatrix}'$ being the ith, $Tx1$-dimensional element of ε_j for $j = 1, \cdots M$.

process in equation 4 started a long time ago[20], one has that

$$Y_i = (\alpha_i \otimes i_T)(I - A_i)^{-1} + \sum_{s=0}^{\infty} \overline{\varepsilon}_{i,-s} A_i^s, \tag{19}$$

where $\overline{\varepsilon}_{i,-s}$ are TxM martices of observations on the sth-order lags of $\overline{\varepsilon}_i$, and thus that

$$\begin{aligned} Y_{i,-1} &= (\alpha_i \otimes i_T)(I - A_i)^{-1} + \sum_{s=0}^{\infty} \overline{\varepsilon}_{i,-s-1} A_i^s && (20) \\ Y'_{i,-1} &= (I - A_i)^{-1\prime} (\alpha_i \otimes i_T)' + \sum_{s=0}^{\infty} A_i'^s \overline{\varepsilon}'_{i,-s-1}. && (21) \end{aligned}$$

Then we state the following proposition that can be proven by results in Appendix C of Pesaran and Smith (1995),

$$\operatorname*{plim}_{T \to \infty} \left(\frac{\overline{\varepsilon}'_{i,-s} H_T \overline{\varepsilon}_{i,-\tau}}{T} \right) = \begin{array}{ll} \Sigma_i & for \quad s = \tau \\ 0 & for \quad s \neq \tau \end{array}. \tag{22}$$

Finally, we notice that

$$\operatorname*{plim}_{N \to \infty, T \to \infty} \left(\frac{\sum_{i=1}^N Y'_{i,-1} H_T Y_{i,-1}}{NT} \right) = \operatorname*{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \operatorname*{plim}_{T \to \infty} \left(\frac{Y'_{i,-1} H_T Y_{i,-1}}{T} \right) \tag{23}$$

$$\operatorname*{plim}_{N \to \infty, T \to \infty} \left(\frac{\sum_{i=1}^N Y'_{i,-1} H_T \nu_{i,j}}{NT} \right) = \operatorname*{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \operatorname*{plim}_{T \to \infty} \left(\frac{Y'_{i,-1} H_T \nu_{i,j}}{T} \right) \tag{24}$$

for $j = 1, \cdots M$.

[20] In fact:

$$\begin{aligned} Y_i &= Y_{i,-1} A_i + \alpha_i \otimes i_T + \varepsilon_i \\ &= (Y_{i,-2} A_i + \alpha_i \otimes i_T + \varepsilon_{i-1}) A_i + \alpha_i \otimes i_T + \varepsilon_i \\ &= Y_{i,-2} A_i^2 + (\alpha_i \otimes i_T) + (\alpha_i \otimes i_T) A_i + \varepsilon_i + \varepsilon_{i-1} A_i \\ &= Y_{i,-3} A_i^3 + (\alpha_i \otimes i_T) + (\alpha_i \otimes i_T) A_i + (\alpha_i \otimes i_T) A_i^2 + \varepsilon_i + \varepsilon_{i-1} A_i + \varepsilon_{i-2} A_i^2 \\ &= \vdots \\ &= \lim_{s \to \infty} Y_{i,-s} A_i^s + (\alpha_i \otimes i_T)(I - A_i)^{-1} + \sum_{s=0}^{\infty} \varepsilon_{i,-s} A_i^s \end{aligned}$$

Consider first the probability limit of $Y'_{i,-1}H_T Y_{i,-1}/T$. Substituting equation 20 for $Y'_{i,-1}$ and $Y_{i,-1}$ respectively we have

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T Y_{i,-1}}{T}\right) &= \underset{T\to\infty}{\text{plim}}\left(\frac{(I-A_i)^{-1\prime}(\alpha_i\otimes i_T)' H_T (\alpha_i\otimes i_T)(I-A_i)^{-1}}{T}\right) \quad (25)\\
&\quad + \underset{T\to\infty}{\text{plim}}\left(\frac{\left(\sum_{s=0}^{\infty} A_i^{s\prime}\overline{\varepsilon}'_{i,-s-1}\right) H_T \left(\sum_{\tau=0}^{\infty}\overline{\varepsilon}_{i,-\tau-1}A_i\tau\right)}{T}\right) \quad (26)\\
&= \sum_{s=0}^{\infty} A_i^{s\prime}\Sigma_i A_i^s\\
&= vec^{-1}\left((I-A'_i\otimes A'_i)^{-1} vec(\Sigma_i)\right),
\end{aligned}$$

as $(\alpha_i\otimes i_T)' H_T = 0$, $\text{plim}_{T\to\infty}\left(\frac{\overline{\varepsilon}'_{i,-s}H_T\overline{\varepsilon}_{i,-\tau}}{T}\right) = 0$ for $s\neq\tau$, and $vec\left(\sum_{s=0}^{\infty}A_i^{s\prime}\Sigma_i A_i^s\right) = \sum_{s=0}^{\infty} vec\left(A_i^{s\prime}\Sigma_i A_i^s\right)$, with vec^{-1} undoing the vec operation. Consider then the probability limit of $Y'_{i,-1}H_T\nu_{i,j}/T$ for $j=1,\cdots M$. Noting that $\nu_{i,j} = \varepsilon_{i,j}+Y_{i,-1}\eta_{i,j}$, where $\eta_{i,j}$ is the jth, $Mx1$ dimensional element of $vec(\eta_i)$ (with $\varepsilon_j = \begin{bmatrix}\varepsilon_{1,j,1} & \cdots & \varepsilon_{1,j,T} & \cdots & \varepsilon_{N,j,T}\end{bmatrix}'$ being the jth, $NTx1$-dimensional element of ε and $\varepsilon_{i,j} = \begin{bmatrix}\varepsilon_{i,j,1} & \cdots & \varepsilon_{i,j,T}\end{bmatrix}'$ being the ith, $Tx1$-dimensional element of ε_j for $j=1,\cdots M$), and substituting we obtain:

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T\nu_{i,j}}{T}\right) &= \underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T\varepsilon_{i,j}}{T}\right) + \underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T Y_{i,-1}}{T}\right)\eta_{i,j} \quad (27)\\
&= \left(\sum_{s=0}^{\infty}A_i^{s\prime}\Sigma_i A_i^s\right)\eta_{i,j} \quad (28)\\
&= vec^{-1}\left((I-A'_i\otimes A'_i)^{-1} vec(\Sigma_i)\right)\eta_{i,j}
\end{aligned}$$

for $j=1,\cdots M$. In fact,

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T\varepsilon_{i,j}}{T}\right) &= \underset{T\to\infty}{\text{plim}}\left(\frac{\left[(I-A_i)^{-1\prime}(\alpha_i\otimes i_T)' + \sum_{s=0}^{\infty}A_i^{\prime s}\overline{\varepsilon}'_{i,-s-1}\right]H_T\varepsilon_{i,j}}{T}\right)\\
&= \underset{T\to\infty}{\text{plim}}\left(\frac{(I-A_i)^{-1\prime}(\alpha_i\otimes i_T)' H_T\varepsilon_{i,j}}{T}\right)\\
&\quad + \sum_{s=0}^{\infty}A_i^{s\prime}\underset{T\to\infty}{\text{plim}}\left(\frac{\overline{\varepsilon}'_{i,-s-1}H_T\varepsilon_{i,j}}{T}\right)\\
&= 0
\end{aligned}$$

because $(\alpha_i \otimes i_T)' H_T = 0$ and $\text{plim}_{T\to\infty}\left(\frac{\overline{\varepsilon}'_{i,-s-1}H_T\varepsilon_{i,j}}{T}\right) = 0$ for each $j = 1, \cdots M$ [**proof on my notes named Multivariate A**].

Now, substituting 25 and 27 in 23 and 24 we have

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y'_{i,-1}H_T Y_{i,-1}}{NT}\right) = \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)$$

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y'_{i,-1}H_T \nu_{i,j}}{NT}\right) = \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N} vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\eta_{i,j}$$

for $j = 1, \cdots M$; and since A_i and Σ_i are *iid* across i, by the law of large numbers, we also have

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y'_{i,-1}H_T Y_{i,-1}}{NT}\right) = E\left[vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\right]$$

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y'_{i,-1}H_T \nu_{i,j}}{NT}\right) = E\left[vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\eta_{i,j}\right]$$

for $j = 1, \cdots M$, where E denotes expectation with respect the joint distribution of A_i and Σ_i.Thus, substituting these last two equations in 12, we obtain equation **??** in the text.

$$\underset{N\to\infty,T\to\infty}{\text{plim}}(\widehat{a}_{FE} - a) = \begin{bmatrix} \left(E\left[vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\right]\right)^{-1} \\ \times\left(E\left[vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\eta_{i,1}\right]\right) \\ \vdots \\ \left(E\left[vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\right]\right)^{-1} \\ \times\left(E\left[vec^{-1}\left(\left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i)\right)\eta_{i,M}\right]\right) \end{bmatrix},$$

which is generally different from 0, unless $\eta_i = 0$ for all i.

A.2 VAR: Special cases

A.2.1 Case 1

Suppose that $\gamma = 0$ in the bivariate VAR in equation 11, then equation 12 in the text becomes

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\begin{pmatrix} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\beta}_{FE} - \beta \end{pmatrix}_{VAR} =$$

$$\left(E\left\{\frac{1}{\Upsilon_0'}\left[\begin{array}{ll}\Upsilon_{1,1}'\sigma+2\Upsilon_1'\beta_i\phi+\Upsilon_2'\beta_i^2\tau^2 & \Upsilon_{2,2}'\phi+\Upsilon_3'\beta_i\tau^2\\ \Upsilon_{2,2}'\phi+\Upsilon_3'\beta_i\tau^2 & \Upsilon_{3,3}'\tau^2\end{array}\right]\right\}\right)^{-1} \quad (29)$$
$$\times\left(E\left\{\frac{1}{\Upsilon_0'}\left[\begin{array}{c}\Upsilon_{2,2}'\phi+\Upsilon_3'\beta_i\tau^2\xi_i\\ \Upsilon_{3,3}'\tau^2\xi_i\end{array}\right]\right\}\right)$$

with:

$$\begin{aligned}
\Upsilon_0' &= (1-\lambda-\rho+\lambda\rho)(1+\lambda+\rho+\lambda\rho)(\lambda\rho-1)\\
&= -(1-\lambda^2)(1-\rho^2)(1-\lambda\rho)\\
\Upsilon_1' &= (\lambda\rho^2-\lambda);\\
\Upsilon_2' &= -(\lambda\rho+1);\\
\Upsilon_3' &= (\lambda^2\rho-\rho);\\
\Upsilon_{1,1}' &= (-\lambda\rho^3+\rho^2+\lambda\rho-1);\\
\Upsilon_{3,3}' &= (-\lambda^3\rho+\lambda^2+\lambda\rho-1).
\end{aligned}$$

Substituting for these expressions, which once divided by Υ_0' simplify considerably, and defining $\delta=\frac{(1+\lambda\rho)}{(1-\lambda^2)(1-\lambda\rho)}$ we obtain

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\left(\begin{array}{c}\widehat{\lambda}_{FE}-\lambda\\ \widehat{\beta}_{FE}-\beta\end{array}\right)=$$

$$\left(E\left\{\left[\begin{array}{ll}\frac{\sigma_i^2}{(1-\lambda^2)}+\frac{\tau_i^2\beta_i^2\delta}{(1-\rho^2)}+\frac{2\lambda\beta_i\phi_i}{(1-\lambda^2)(1-\lambda\rho)} & \frac{\rho\beta_i\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)}\\ \frac{\rho\beta_i\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)} & \frac{\tau_i^2}{1-\rho^2}\end{array}\right]\right\}\right)^{-1}$$
$$\times\left(E\left\{\left[\begin{array}{l}\frac{\phi_i}{(1-\lambda\rho)}+\frac{\left(\rho\beta_i\tau_i^2\right)\xi_i}{(1-\lambda\rho)(1-\rho^2)}\\ \frac{\tau_i^2\xi_i}{1-\rho^2}\end{array}\right]\right\}\right).$$

Taking then expectations with respect to the distribution of β_i and denoting $\omega_{2,2}$ the variance of ξ_i we also have

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\left(\begin{array}{c}\widehat{\lambda}_{FE}-\lambda\\ \widehat{\beta}_{FE}-\beta\end{array}\right)=$$

$$\left[\begin{array}{ll}\frac{\sigma_i^2}{(1-\lambda^2)}+\frac{\tau_i^2\delta\left(\beta^2+\omega_{2,2}\right)}{(1-\rho^2)}+\frac{2\lambda\beta\phi_i}{(1-\lambda^2)(1-\lambda\rho)} & \frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)}\\ \frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)} & \frac{\tau_i^2}{1-\rho^2}\end{array}\right]^{-1}$$

$$\times \begin{bmatrix} \frac{\rho\omega_{2,2}\tau_i^2}{(1-\lambda\rho)(1-\rho^2)} \\ 0 \end{bmatrix}.$$

Thus we get,

$$\operatorname*{plim}_{N\to\infty,T\to\infty} \begin{pmatrix} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\beta}_{FE} - \beta \end{pmatrix} =$$

$$= \frac{1}{\Delta} \begin{bmatrix} \frac{\tau_i^2}{1-\rho^2} & -\frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)} - \frac{\phi_i}{(1-\lambda\rho)} \\ -\frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)} - \frac{\phi_i}{(1-\lambda\rho)} & \frac{\sigma_i^2}{(1-\lambda^2)} + \frac{\tau_i^2\delta\left(\beta^2+\omega_{2,2}\right)}{(1-\rho^2)} + \frac{2\lambda\beta\phi_i}{(1-\lambda^2)(1-\lambda\rho)} \end{bmatrix}$$

$$\times \begin{bmatrix} \frac{\rho\omega_{2,2}\tau_i^2}{(1-\lambda\rho)(1-\rho^2)} \\ 0 \end{bmatrix}$$

$$= \frac{1}{\Delta} \begin{bmatrix} \left(\frac{\tau_i^2}{1-\rho^2}\right)\left(\frac{\rho\omega_{2,2}\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}\right) \\ -\left(\frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)} + \frac{\phi_i}{(1-\lambda\rho)}\right)\left(\frac{\rho\omega_{2,2}\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}\right) \end{bmatrix};$$

where

$$\begin{aligned} \Delta &= \left(\frac{\sigma_i^2}{(1-\lambda^2)} + \frac{\tau_i^2\delta\left(\beta^2+\omega_{2,2}\right)}{(1-\rho^2)} + \frac{2\lambda\beta\phi_i}{(1-\lambda^2)(1-\lambda\rho)}\right)\left(\frac{\tau_i^2}{1-\rho^2}\right) \\ &\quad -\left(\frac{\rho\beta_i\tau_i^2}{(1-\lambda\rho)(1-\rho^2)} + \frac{\phi_i}{(1-\lambda\rho)}\right)^2 \end{aligned}$$

After some algebra simplifications we finally obtain

$$\operatorname*{plim}_{N\to\infty,T\to\infty} \begin{pmatrix} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\beta}_{FE} - \beta \end{pmatrix} = \begin{bmatrix} \frac{\rho(1-\lambda\rho)\left(1-\lambda^2\right)\omega_{2,2}}{\Psi_1+\Psi_2} \\ -\frac{\beta\rho^2\left(1-\lambda^2\right)\omega_{2,2}+\Psi_3}{\Psi_1+\Psi_2} \end{bmatrix} \tag{30}$$

with:

$$\begin{aligned} \Psi_1 &= (\sigma^2/\tau^2)\left(1-\rho^2\right)(1-\lambda\rho)^2 + \left(1-\lambda^2\rho^2\right)\omega_{2,2} + \left(1-\rho^2\right)\beta^2; \\ \Psi_2 &= -(\phi^2/\tau^2)\left(1-\rho^2\right)\left(1-\lambda^2\right) - 2(\sigma/\tau)\left(1-\rho^2\right)(1-\lambda)\beta; \\ \Psi_3 &= (\phi/\tau)\left(1-\rho^2\right)\left(1-\lambda^2\right)\rho\omega_{2,2}. \end{aligned}$$

A.2.2 Case 2

Substituting $\gamma = 1$, $\phi = 0$[21] in equation BIASSVAR2, we have

$$\underset{N\to\infty,T\to\infty}{\text{plim}} \left(\widehat{\beta}_{FE} - \beta\right)$$

$$= \frac{E\left(\frac{\xi_i}{\left(1-\beta_i^2\gamma^2\right)}\right)}{E\left(\frac{1}{\left(1-\beta_i^2\gamma^2\right)}\right)}.$$

Assuming

$$\xi_i \sim Uniform[\pm\omega(1-\beta)] \qquad 0 \leq \omega < 1 \qquad 0 \leq \beta < 1$$

where ω is a scale parameter controlling the variance of ξ_i, and solving the integrals we get:

$$= \frac{\begin{array}{c} -(1+b)\ln(1+b+\omega(1-b)) - (1-b)\ln(1-b-\omega(1-b)) \\ +(1+b)\ln(1+b-\omega(1-b)) + (1-b)\ln(1-b+\omega(1-b)) \end{array}}{\begin{array}{c} \ln(1+b+\omega(1-b)) - \ln(1-b-\omega(1-b)) \\ -\ln(1+b-\omega(1-b)) + \ln(1-b+\omega(1-b)) \end{array}}.$$

Which is plotted below.

[21] Both these assumptions could be relaxed and I am working on $\phi = 0$.

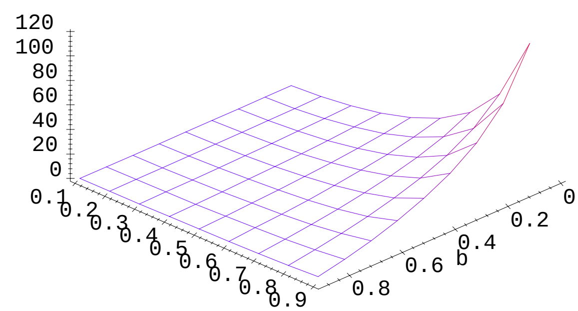

Figure 2

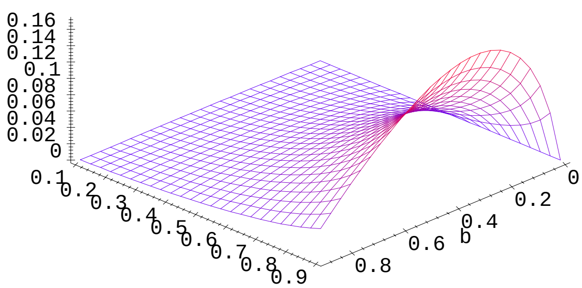

Figure 1:

Table 1 - Monte Carlo Results (N,T=50,50)

	MG	FE	IV	MG	FE	IV	MG	FE	IV
	Omeg=0			Omega=0.2			Omega=0.8		
Beta=0.2									
Fi=0									
Lambda									
S.e.	0.0196	0.0195	0.0861	0.0195	0.0193	0.0874	0.0178	0.0162	0.129
Bias	*-0.0246*	*-0.0241*	*0.0001*	*-0.0244*	*-0.0247*	*-0.0039*	*-0.0221*	*-0.0492*	-0.0827
S.d.	0.0198	0.0195	0.0351	0.0198	0.0199	0.0355	0.0183	0.0341	0.0533
Beta	0.1715	0.1866	0.1986	0.1718	0.1899	0.2017	0.1724	0.2614	0.2674
S.e	0.0142	0.0141	0.0487	0.0189	0.014	0.0496	0.0518	0.0127	0.0783
Bias	-0.0285	-0.0134	-0.0014	-0.0282	-0.0101	0.0017	-0.0276	0.0614	0.0674
Fbias as % of true value	*-14.3%*	*-6.7%*	*-0.7%*	*-14.1%*	*-5.1%*	*0.9%*	*-13.8%*	*30.7%*	*33.7%*
Abias as % of true value					*1.8%*			*48.8%*	
S.d.	0.0144	0.014	0.0243	0.0193	0.0196	0.0278	0.0516	0.0657	0.0658
Fi=0.9									
Lambda									
S.e.	0.0292	0.0291	0.0507	0.0291	0.0287	0.052	0.0265	0.0236	0.0897
Bias	*0.0154*	*-0.0075*	*0.0015*	*0.015*	*0.0072*	*0.007*	*0.0076*	*0.1933*	0.0635
S.d.	0.0301	0.0292	0.0325	0.0299	0.0298	0.0329	0.0275	0.0441	0.0456
Beta	0.1516	0.1794	0.1977	0.152	0.1743	0.1894	0.1564	0.1258	0.0867
S.e	0.0209	0.0211	0.0415	0.0243	0.021	0.0413	0.0534	0.0186	0.0416
Bias	-0.0484	-0.0206	-0.0023	-0.048	-0.0257	-0.0106	-0.0436	-0.0742	-0.1133
Fbias as % of true value	*-24.2%*	*-10.3%*	*-1.2%*	*-24.0%*	*-12.9%*	*-5.3%*	*-21.8%*	*-37.1%*	*-56.7%*
S.d.	0.0214	0.0209	0.0308	0.0247	0.0251	0.0337	0.0538	0.0597	0.0727
Fi=--0.9									
Lambda									
S.e.	0.0303	0.0299	0.3298	0.0302	0.0295	0.3239	0.0283	0.0241	0.3199
Bias	*-0.0741*	*-0.0493*	*-0.0018*	*-0.0734*	*-0.0658*	*-0.0242*	*-0.0624*	*-0.3*	-0.2722
S.d.	0.0301	0.0295	0.0883	0.0302	0.0302	0.0879	0.0292	0.0495	0.1084
Beta	0.1435	0.1715	0.1981	0.1437	0.1663	0.1936	0.1472	0.1134	0.1827
S.e	0.0213	0.0214	0.1776	0.0243	0.0212	0.1761	0.0523	0.0189	0.2039
Bias	-0.0565	-0.0285	-0.0019	-0.0563	-0.0337	-0.0064	-0.0528	-0.0866	-0.0173
Fbias as % of true value	*-28.3%*	*-14.3%*	*-1.0%*	*-28.2%*	*-16.9%*	*-3.2%*	*-26.4%*	*-43.3%*	*-8.7%*
S.d.	0.0212	0.0209	0.0512	0.0245	0.0249	0.0524	0.0522	0.0538	0.0665
Beta=0.8									
Fi=0									
Lambda									
S.e.	0.0119	0.0103	0.5832	0.0119	0.0103	0.726	0.0118	0.0086	0.9041
Bias	*-0.0282*	*-0.025*	*0.0032*	*-0.0283*	*-0.0256*	*-0.0028*	*-0.0263*	*-0.0177*	0.1019
S.d.	0.012	0.0106	0.1881	0.0121	0.0108	0.7508	0.012	0.0178	0.5751
Beta	0.7474	0.778	0.8015	0.7474	0.7798	0.7988	0.7501	0.8309	0.9428
S.e	0.0108	0.0094	0.522	0.0113	0.0093	0.6507	0.0168	0.008	0.8263
Bias	-0.0526	-0.022	0.0015	-0.0526	-0.0202	-0.0012	-0.0499	0.0309	0.1428
Fbias as % of true value	*-6.6%*	*-2.8%*	*0.2%*	*-6.6%*	*-2.5%*	*-0.2%*	*-6.2%*	*3.9%*	*17.9%*
Abias as % of true value					*0.3%*			*6.2%*	
S.d.	0.011	0.0096	0.1688	0.0116	0.0104	0.6722	0.0171	0.0234	0.5238
Fi=0.9									
Lambda									
S.e.	0.0238	0.0212	0.2685	0.0238	0.0211	0.2787	0.0235	0.0182	2.78
Bias	*-0.0089*	*-0.022*	*0.0022*	*-0.009*	*-0.0216*	*0.0008*	*-0.0083*	*-0.0071*	-0.1399
S.d.	0.0238	0.0207	0.0644	0.0239	0.0208	0.067	0.0237	0.0242	5.8661
Beta	0.7339	0.7773	0.7993	0.7341	0.7784	0.8004	0.7363	0.8101	0.7146
S.e	0.0216	0.0193	0.2413	0.0218	0.0192	0.2502	0.0252	0.0169	2.4414
Bias	-0.0661	-0.0227	-0.0007	-0.0659	-0.0216	0.0004	-0.0637	0.0101	-0.0854
Fbias as % of true value	*-8.3%*	*-2.8%*	*-0.1%*	*-8.2%*	*-2.7%*	*0.1%*	*-8.0%*	*1.3%*	*-10.7%*
S.d.	0.022	0.019	0.0594	0.0223	0.0194	0.0614	0.0256	0.0266	5.3667
Fi=--0.9									
Lambda									
S.e.	0.0282	0.0248	42.7986	0.0282	0.0247	43.1479	0.0269	0.0127	0.316
Bias	*-0.0682*	*-0.0462*	*2.1192*	*-0.0681*	*-0.0484*	*-1.7058*	*-0.0582*	*0.0091*	0.0712
S.d.	0.029	0.0259	94.3066	0.029	0.0262	28.8504	0.027	0.0264	0.1303
Beta	0.7128	0.7592	2.6971	0.7131	0.7596	-0.7272	0.7239	0.8576	0.9149
S.e	0.0254	0.0224	38.2932	0.0257	0.0224	38.7169	0.0284	0.0118	0.2899
Bias	-0.0872	-0.0408	1.8971	-0.0869	-0.0404	-1.5272	-0.0761	0.0576	0.1149
Fbias as % of true value	*-10.9%*	*-5.1%*	*237.1%*	*-10.9%*	*-5.1%*	*-191%*	*-9.5%*	*7.2%*	*14.4%*
S.d.	0.0261	0.0232	84.3605	0.0264	0.0237	25.8701	0.0289	0.0305	0.1166

Note - *S.e*.: estimated standard errors;

Bias: absolute value of the finite sample bias (equal to the estimated parameter value in the case of lambda).

S.d.: finite sample bias' experimental standard deviations.

Fbias as % of true value: finite sample bias as a percentage of the true value of beta.

Abias as % of true value: asymptotic bias as a percentage of the true value.

Table 2 - Monte Carlo Results (N,T=20,50)

	MG	FE	IV	MG	FE	IV	MG	FE	IV
	Omeg=0			Omega=0.2			Omega=0.8		
Beta=0.2									
Fi=0									
Lambda									
S.e.	0.0312	0.0309	0.1366	0.031	0.0306	0.1387	0.0283	0.0259	0.2038
Bias	-0.0245	-0.0241	-0.0007	-0.0243	-0.0247	-0.0047	-0.0227	-0.0468	-0.0805
S.d.	0.0308	0.0304	0.0531	0.0304	0.0312	0.053	0.0273	0.0508	0.0805
Beta	0.1731	0.1876	0.2001	0.1727	0.19	0.202	0.1713	0.2547	0.2585
S.e	0.0222	0.0223	0.0775	0.0294	0.0222	0.0788	0.0811	0.0202	0.1244
Bias	-0.0269	-0.0124	0.0001	-0.0273	-0.01	0.002	-0.0287	0.0547	0.0585
Fbias as % of true value	-13.5%	-6.2%	0.1%	-13.7%	-5.0%	1.0%	-14.4%	27.4%	29.3%
Abias as % of true value					1.8%			48.8%	
S.d.	0.0225	0.0227	0.0365	0.0295	0.031	0.0421	0.0811	0.102	0.1042
Fi=0.9									
Lambda									
S.e.	0.0464	0.0459	0.0804	0.0462	0.0454	0.0825	0.0419	0.0376	0.1403
Bias	0.0134	-0.0092	-0.0007	0.0131	0.0049	0.0044	0.0053	0.1817	0.0571
S.d.	0.0444	0.0446	0.0499	0.0442	0.0449	0.0496	0.0405	0.0677	0.0658
Beta	0.1539	0.1813	0.2014	0.1536	0.1755	0.1927	0.1557	0.1267	0.0918
S.e	0.0329	0.0334	0.0661	0.0379	0.0332	0.0657	0.0836	0.0294	0.0675
Bias	-0.0461	-0.0187	0.0014	-0.0464	-0.0245	-0.0073	-0.0443	-0.0733	-0.1082
Fbias as % of true value	-23.1%	-9.4%	0.7%	-23.2%	-12.3%	-3.7%	-22.2%	-36.7%	-54.1%
S.d.	0.0328	0.0336	0.0468	0.0373	0.0394	0.0514	0.0833	0.0914	0.1124
Fi=--0.9									
Lambda									
S.e.	0.0477	0.0472	0.5253	0.0477	0.0467	0.516	0.0446	0.0383	0.5204
Bias	-0.0718	-0.0478	0.0014	-0.0714	-0.0632	-0.0214	-0.0612	-0.2837	-0.2616
S.d.	0.0493	0.0489	0.1318	0.0495	0.0497	0.13	0.0463	0.0772	0.1666
Beta	0.146	0.1735	0.2015	0.1453	0.1678	0.1955	0.1465	0.1148	0.1765
S.e	0.0333	0.0338	0.2832	0.0379	0.0335	0.2808	0.082	0.0299	0.3336
Bias	-0.054	-0.0265	0.0015	-0.0547	-0.0322	-0.0045	-0.0535	-0.0852	-0.0235
Fbias as % of true value	-27.0%	-13.3%	0.8%	-27.4%	-16.1%	-2.3%	-26.8%	-42.6%	-11.8%
S.d.	0.0344	0.0348	0.077	0.0387	0.04	0.0788	0.0819	0.0846	0.1086
Beta=0.8									
Fi=0									
Lambda									
S.e.	0.0187	0.0164	1.6225	0.0187	0.0163	8.0411	0.0187	0.0139	6.0483
Bias	-0.0289	-0.0261	0.0263	-0.0289	-0.0266	1.4099	-0.0269	-0.0202	-0.6189
S.d.	0.019	0.0166	1.3371	0.019	0.0168	49.8178	0.019	0.0242	26.1986
Beta	0.7485	0.7782	0.8236	0.7485	0.7798	2.0226	0.7507	0.8259	0.2549
S.e	0.017	0.0148	1.452	0.0177	0.0148	7.0503	0.0263	0.0129	5.6312
Bias	-0.0515	-0.0218	0.0236	-0.0515	-0.0202	1.2226	-0.0493	0.0259	-0.5451
Fbias as % of true value	-6.4%	-2.7%	3.0%	-6.4%	-2.5%	152.8%	-6.2%	3.2%	-68.1%
Abias as % of true value					0.3%			6.2%	
S.d.	0.0172	0.0152	1.1904	0.0177	0.0161	43.2428	0.0261	0.0346	24.9263
Fi=0.9									
Lambda									
S.e.	0.0376	0.0336	0.4914	0.0376	0.0334	0.5352	0.0373	0.0292	6.9102
Bias	-0.0115	-0.0244	0.0041	-0.0116	-0.0241	0.0038	-0.0107	-0.0105	-0.7107
S.d.	0.0383	0.0327	0.1211	0.0382	0.0325	0.1628	0.0377	0.0358	30.7153
Beta	0.7361	0.7788	0.8051	0.7361	0.7797	0.8071	0.7379	0.8078	0.2452
S.e	0.0339	0.0305	0.4435	0.0343	0.0303	0.4819	0.0394	0.027	5.9619
Bias	-0.0639	-0.0212	0.0051	-0.0639	-0.0203	0.0071	-0.0621	0.0078	-0.5548
Fbias as % of true value	-8.0%	-2.7%	0.6%	-8.0%	-2.5%	0.9%	-7.8%	1.0%	-69.4%
S.d.	0.0343	0.0298	0.1107	0.0344	0.0299	0.1486	0.0391	0.0381	26.1735
Fi=--0.9									
Lambda									
S.e.	0.0441	0.0394	128.8833	0.044	0.0392	124.5329	0.0417	0.0223	7.3664
Bias	-0.0676	-0.0473	26.474	-0.0672	-0.0492	-26.8017	-0.0577	-0.0025	-0.0769
S.d.	0.0441	0.0396	799.2592	0.0439	0.0397	849.9118	0.0415	0.0426	4.156
Beta	0.7148	0.7588	24.4217	0.7151	0.7591	-22.9458	0.725	0.8432	0.7783
S.e	0.0398	0.0356	115.0735	0.04	0.0355	110.5616	0.0441	0.0207	6.555
Bias	-0.0852	-0.0412	23.6217	-0.0849	-0.0409	-23.7458	-0.075	0.0432	-0.0217
Fbias as % of true value	-10.7%	-5.2%	2952.7%	-10.6%	-5.1%	-2968%	-9.4%	5.4%	-2.7%
S.d.	0.0403	0.0361	713.181	0.0407	0.0366	753.1006	0.045	0.0515	3.7241

Note - *S.e* .: estimated standard errors;

Bias: absolute value of the finite sample bias (equal to the estimated parameter value in the case of lambda).

S.d.: finite sample bias' experimental standard deviations.

Fbias as % of true value: finite sample bias as a percentage of the true value of beta.

Abias as % of true value: asymptotic bias as a percentage of the true value.

Table 3 - Monte Carlo Results (N,T=10,50)

	MG	FE	IV	MG	FE	IV	MG	FE	IV
		Omeg=0			Omega=0.2			Omega=0.8	
Beta=0.2									
Fi=0									
Lambda									
S.e.	0.0424	0.0438	0.192	0.0421	0.0433	0.1953	0.0382	0.0368	0.2945
Bias	-0.0235	-0.0227	-0.0005	-0.0236	-0.0236	-0.005	-0.0221	-0.0444	-0.0804
S.d.	0.0437	0.0434	0.0751	0.0432	0.0444	0.0756	0.0397	0.0681	0.1267
Beta	0.1719	0.1859	0.1957	0.1717	0.1882	0.1974	0.1706	0.2449	0.2414
S.e	0.031	0.0316	0.1091	0.0418	0.0314	0.1113	0.115	0.0287	0.1841
Bias	-0.0281	-0.0141	-0.0043	-0.0283	-0.0118	-0.0026	-0.0294	0.0449	0.0414
Fbias as % of true value	-14.1%	-7.1%	-2.2%	-14.2%	-5.9%	-1.3%	-14.7%	22.5%	20.7%
Abias as % of true value					1.8%			48.8%	
S.d.	0.0315	0.0312	0.0535	0.0424	0.0438	0.0627	0.1158	0.1425	0.1504
Fi=0.9									
Lambda									
S.e.	0.0634	0.0651	0.1128	0.0629	0.0643	0.1159	0.0573	0.0534	0.1957
Bias	0.0162	-0.0047	0.0037	0.0154	0.0085	0.0087	0.0078	0.1707	0.0529
S.d.	0.066	0.0656	0.0721	0.0656	0.0662	0.072	0.0599	0.0926	0.0954
Beta	0.1518	0.1775	0.194	0.152	0.1722	0.1851	0.155	0.1256	0.0863
S.e	0.0457	0.0474	0.0932	0.0533	0.047	0.0929	0.1184	0.0418	0.0996
Bias	-0.0482	-0.0225	-0.006	-0.048	-0.0278	-0.0149	-0.045	-0.0744	-0.1137
Fbias as % of true value	-24.1%	-11.3%	-3.0%	-24.0%	-13.9%	-7.5%	-22.5%	-37.2%	-56.9%
S.d.	0.0468	0.0471	0.0683	0.0546	0.0558	0.0749	0.1193	0.131	0.1574
Fi=--0.9									
Lambda									
S.e.	0.0659	0.0669	0.741	0.0656	0.0661	0.7337	0.0609	0.0546	0.8846
Bias	-0.0721	-0.0491	-0.01	-0.0718	-0.0647	-0.0335	-0.0617	-0.2692	-0.2559
S.d.	0.0664	0.0658	0.1922	0.0661	0.0666	0.193	0.0621	0.1056	0.4019
Beta	0.1445	0.1704	0.1921	0.1442	0.1648	0.1859	0.1457	0.112	0.1624
S.e	0.0465	0.0479	0.3994	0.0534	0.0475	0.4002	0.1156	0.0426	0.5923
Bias	-0.0555	-0.0296	-0.0079	-0.0558	-0.0352	-0.0141	-0.0543	-0.088	-0.0376
Fbias as % of true value	-27.8%	-14.8%	-4.0%	-27.9%	-17.6%	-7.1%	-27.2%	-44.0%	-18.8%
S.d.	0.0465	0.0467	0.1135	0.0536	0.0553	0.1191	0.1165	0.122	0.3062
Beta=0.8									
Fi=0									
Lambda									
S.e.	0.0259	0.0234	572	0.026	0.0233	3.3038	0.0257	0.0203	3.3195
Bias	-0.0271	-0.0243	81.3346	-0.0271	-0.0249	0.1447	-0.0249	-0.0201	0.2187
S.d.	0.0277	0.0247	2577.9821	0.0276	0.025	3.6137	0.0269	0.0329	3.8278
Beta	0.7467	0.775	74.17	0.7465	0.7764	0.9271	0.7491	0.8161	1.0382
S.e	0.0238	0.0212	516.0427	0.0248	0.0211	2.9555	0.0371	0.0188	3.0166
Bias	-0.0533	-0.025	73.37	-0.0535	-0.0236	0.1271	-0.0509	0.0161	0.2382
Fbias as % of true value	-6.7%	-3.1%	9171.3%	-6.7%	-3.0%	15.9%	-6.4%	2.0%	29.8%
Abias as % of true value					0.3%			6.2%	
S.d.	0.0249	0.022	2325.9178	0.0257	0.0233	3.2442	0.0376	0.0475	3.5335
Fi=0.9									
Lambda									
S.e.	0.052	0.048	1.0914	0.052	0.0477	1.6528	0.0514	0.0422	3.2541
Bias	-0.005	-0.0187	0.0259	-0.005	-0.0185	0.0228	-0.004	-0.0078	0.2363
S.d.	0.0545	0.0493	0.6187	0.0543	0.0494	1.1718	0.0532	0.0513	6.361
Beta	0.7313	0.772	0.8133	0.7312	0.7728	0.8112	0.7331	0.7982	1.0633
S.e	0.0473	0.0435	0.979	0.048	0.0433	1.4716	0.0556	0.0389	2.9755
Bias	-0.0687	-0.028	0.0133	-0.0688	-0.0272	0.0112	-0.0669	-0.0018	0.2633
Fbias as % of true value	-8.6%	-3.5%	1.7%	-8.6%	-3.4%	1.4%	-8.4%	-0.2%	32.9%
S.d.	0.0498	0.0442	0.5826	0.0499	0.0448	1.0444	0.0565	0.0571	6.3401
Fi=--0.9									
Lambda									
S.e.	0.0615	0.0562	21.8603	0.0614	0.0559	43.93	0.0582	0.0359	5.0536
Bias	-0.0667	-0.0464	-0.2119	-0.0665	-0.0483	11.9799	-0.0571	-0.015	0.358
S.d.	0.0647	0.0574	7.0411	0.0644	0.0575	373.4841	0.0613	0.06	3.7825
Beta	0.7136	0.756	0.6051	0.7136	0.756	11.4596	0.7233	0.8221	1.1569
S.e	0.0558	0.0508	19.5658	0.0561	0.0506	39.2048	0.0619	0.0331	4.528
Bias	-0.0864	-0.044	-0.1949	-0.0864	-0.044	10.6596	-0.0767	0.0221	0.3569
Fbias as % of true value	-10.8%	-5.5%	-24.4%	-10.8%	-5.5%	1332%	-9.6%	2.8%	44.6%
S.d.	0.0582	0.0518	6.2987	0.0583	0.0523	332.3963	0.0637	0.0714	3.3687

Note - *S.e* .: estimated standard errors;

Bias: absolute value of the finite sample bias (equal to the estimated parameter value in the case of lambda).

S.d.: finite sample bias' experimental standard deviations.

Fbias as % of true value: finite sample bias as a percentage of the true value of beta.

Abias as % of true value: asymptotic bias as a percentage of the true value.

Table 4 - Monte Carlo Results (N,T=50,20)

	MG	FE	IV	MG	FE	IV	MG	FE	IV
		Omeg=0			**Omega=0.2**			**Omega=0.8**	
					Beta=0.2				
					Fi=0				
Lambda									
S.e.	0.0318	0.0317	0.1369	0.0316	0.0314	0.1391	0.0293	0.0269	0.2116
Bias	-0.0591	-0.0582	0.0061	-0.0585	-0.0596	0.0024	-0.0547	-0.1111	-0.0837
S.d.	0.0328	0.0322	0.0629	0.0326	0.0328	0.063	0.0301	0.0493	0.1108
Beta	0.133	0.1679	0.2014	0.133	0.17	0.2041	0.133	0.2175	0.2632
S.e	0.0235	0.0228	0.0775	0.0262	0.0227	0.0788	0.0527	0.021	0.128
Bias	-0.067	-0.0321	0.0014	-0.067	-0.03	0.0041	-0.067	0.0175	0.0632
Fbias as % of true value	-33.5%	-16.1%	0.7%	-33.5%	-15.0%	2.1%	-33.5%	8.8%	31.6%
Abias as % of true value					1.8%			48.8%	
S.d.	0.0235	0.0229	0.0434	0.0257	0.026	0.0461	0.0509	0.0665	0.0922
					Fi=0.9				
Lambda									
S.e.	0.0473	0.0462	0.0806	0.0471	0.0457	0.0829	0.0436	0.0381	0.1459
Bias	0.0349	-0.019	0.0057	0.034	-0.0059	0.0115	0.0148	0.1367	0.0624
S.d.	0.0488	0.0474	0.0539	0.0487	0.0475	0.0539	0.0455	0.0568	0.0874
Beta	0.0858	0.1504	0.2004	0.0861	0.1444	0.1915	0.0954	0.0843	0.0853
S.e	0.0336	0.034	0.0659	0.0354	0.0338	0.0655	0.057	0.0299	0.0665
Bias	-0.1142	-0.0496	0.0004	-0.1139	-0.0556	-0.0085	-0.1046	-0.1157	-0.1147
Fbias as % of true value	-57.1%	-24.8%	0.2%	-57.0%	-27.8%	-4.3%	-52.3%	-57.9%	-57.4%
S.d.	0.0338	0.0335	0.0529	0.035	0.0354	0.0545	0.055	0.0641	0.0871
					Fi=--0.9				
Lambda									
S.e.	0.0501	0.0493	0.5291	0.0502	0.0487	0.5191	0.0496	0.0397	0.5309
Bias	-0.1774	-0.1198	0.0089	-0.1763	-0.136	-0.0131	-0.1519	-0.374	-0.2682
S.d.	0.0505	0.0497	0.1961	0.0505	0.0506	0.1917	0.0506	0.0707	0.2345
Beta	0.0657	0.1308	0.203	0.0653	0.1249	0.1982	0.0721	0.0614	0.1821
S.e	0.0345	0.0349	0.285	0.0361	0.0346	0.2821	0.0565	0.0309	0.3381
Bias	-0.1343	-0.0692	0.003	-0.1347	-0.0751	-0.0018	-0.1279	-0.1386	-0.0179
Fbias as % of true value	-67.2%	-34.6%	1.5%	-67.4%	-37.6%	-0.9%	-64.0%	-69.3%	-9.0%
S.d.	0.0344	0.0343	0.1118	0.0355	0.0357	0.1116	0.0552	0.0548	0.1508
					Beta=0.8				
					Fi=0				
Lambda									
S.e.	0.0228	0.0188	3.7494	0.0228	0.0187	3.5867	0.0227	0.0166	3.9044
Bias	-0.0746	-0.0679	0.1404	-0.0746	-0.0687	-0.1721	-0.0704	-0.0663	-0.2768
S.d.	0.0233	0.0194	4.3626	0.0233	0.0196	5.015	0.0229	0.0249	8.2833
Beta	0.6795	0.7417	0.9232	0.6796	0.743	0.6472	0.6853	0.7864	0.5948
S.e	0.0211	0.017	3.3473	0.0213	0.0169	3.2269	0.0246	0.0153	3.5664
Bias	-0.1205	-0.0583	0.1232	-0.1204	-0.057	-0.1528	-0.1147	-0.0136	-0.2052
Fbias as % of true value	-15.1%	-7.3%	15.4%	-15.1%	-7.1%	-19.1%	-14.3%	-1.7%	-25.7%
Abias as % of true value					0.3%			6.2%	
S.d.	0.0207	0.0183	3.8959	0.021	0.0186	4.4779	0.0244	0.0278	7.5651
					Fi=0.9				
Lambda									
S.e.	0.0421	0.0343	1.0586	0.0421	0.0341	3.9655	0.0418	0.03	3.4717
Bias	-0.0353	-0.0615	0.0396	-0.0353	-0.0615	0.7848	-0.0324	-0.0579	0.6013
S.d.	0.042	0.0325	1.6292	0.0418	0.0325	24.2966	0.041	0.0338	17.0949
Beta	0.6531	0.7412	0.8357	0.6532	0.742	1.474	0.6574	0.7706	1.4006
S.e	0.0385	0.0312	0.9543	0.0386	0.0311	3.4341	0.0405	0.0277	3.1393
Bias	-0.1469	-0.0588	0.0357	-0.1468	-0.058	0.674	-0.1426	-0.0294	0.6006
Fbias as % of true value	-18.4%	-7.4%	4.5%	-18.4%	-7.3%	84.3%	-17.8%	-3.7%	75.1%
S.d.	0.0376	0.0302	1.5258	0.0375	0.0302	20.8421	0.0388	0.0324	15.9502
					Fi=--0.9				
Lambda									
S.e.	0.0583	0.0483	16.2352	0.0583	0.0482	14.7817	0.0552	0.0292	0.3055
Bias	-0.1699	-0.1231	0.5404	-0.1695	-0.1259	0.5422	-0.1471	-0.0454	0.06
S.d.	0.0589	0.0498	18.0415	0.059	0.05	17.8643	0.0564	0.0487	0.1114
Beta	0.5966	0.6923	1.2844	0.5972	0.6918	1.2877	0.6216	0.8078	0.9031
S.e	0.0528	0.0436	14.5323	0.0529	0.0435	13.2459	0.0527	0.0271	0.2802
Bias	-0.2034	-0.1077	0.4844	-0.2028	-0.1082	0.4877	-0.1784	0.0078	0.1031
Fbias as % of true value	-25.4%	-13.5%	60.6%	-25.4%	-13.5%	61%	-22.3%	1.0%	12.9%
S.d.	0.0541	0.0456	16.1269	0.0544	0.0458	16.0029	0.0547	0.0499	0.1011

Note - *S.e* .: estimated standard errors;

Bias: absolute value of the finite sample bias (equal to the estimated parameter value in the case of lambda).

S.d.: finite sample bias' experimental standard deviations.

Fbias as % of true value: finite sample bias as a percentage of the true value of beta.

Abias as % of true value: asymptotic bias as a percentage of the true value.

Table 5 - Monte Carlo Results (N,T=20,20)

	MG	FE	IV	MG	FE	IV	MG	FE	IV
		Omega=0			Omega=0.2			Omega=0.8	
Beta=0.2									
Fi=0									
Lambda									
S.e.	0.0502	0.0501	0.2154	0.0499	0.0497	0.2185	0.0463	0.0428	0.3428
Bias	*-0.0616*	*-0.06*	*-0.0005*	*-0.0614*	*-0.0617*	*-0.0044*	*-0.0578*	*-0.1105*	*-0.0924*
S.d.	0.0491	0.0491	0.0998	0.0487	0.0497	0.1012	0.0457	0.0729	0.191
Beta	0.1345	0.1679	0.1975	0.1345	0.1697	0.2	0.1337	0.2117	0.2481
S.e	0.0365	0.0361	0.1221	0.041	0.036	0.1241	0.083	0.0333	0.209
Bias	-0.0655	-0.0321	-0.0025	-0.0655	-0.0303	0	-0.0663	0.0117	0.0481
Fbias as % of true value	*-32.8%*	*-16.1%*	*-1.3%*	*-32.8%*	*-15.2%*	*0.0%*	*-33.2%*	*5.9%*	*24.1%*
Abias as % of true value					1.80%			48.76%	
S.d.	0.0365	0.0352	0.0676	0.0405	0.0409	0.0718	0.081	0.1015	0.1505
Fi=0.9									
Lambda									
S.e.	0.0737	0.073	0.1267	0.0734	0.0722	0.1298	0.068	0.0603	0.2248
Bias	*0.0296*	*-0.0215*	*0.0021*	*0.0283*	*-0.0095*	*0.007*	*0.0093*	*0.1252*	*0.0474*
S.d.	0.0713	0.0711	0.0832	0.0713	0.0718	0.0839	0.0674	0.0848	0.1435
Beta	0.0881	0.151	0.1967	0.0883	0.1454	0.1884	0.0962	0.0867	0.0827
S.e	0.0521	0.0538	0.1045	0.0552	0.0534	0.1039	0.0898	0.0474	0.1085
Bias	-0.1119	-0.049	-0.0033	-0.1117	-0.0546	-0.0116	-0.1038	-0.1133	-0.1173
Fbias as % of true value	*-56.0%*	*-24.5%*	*-1.7%*	*-55.9%*	*-27.3%*	*-5.8%*	*-51.9%*	*-56.7%*	*-58.7%*
S.d.	0.0522	0.0526	0.0802	0.0548	0.0569	0.0831	0.0882	0.1027	0.1351
Fi=--0.9									
Lambda									
S.e.	0.0791	0.0781	0.8462	0.0791	0.0771	0.83	0.0778	0.0633	0.9303
Bias	*-0.1788*	*-0.1215*	*-0.0066*	*-0.1773*	*-0.1371*	*-0.0276*	*-0.1534*	*-0.3637*	*-0.271*
S.d.	0.0801	0.0785	0.3038	0.0802	0.0791	0.3051	0.0787	0.1074	0.5034
Beta	0.0675	0.1312	0.1952	0.0674	0.125	0.19	0.0732	0.0607	0.1667
S.e	0.0539	0.0552	0.4559	0.0565	0.0548	0.4512	0.0886	0.049	0.5947
Bias	-0.1325	-0.0688	-0.0048	-0.1326	-0.075	-0.01	-0.1268	-0.1393	-0.0333
Fbias as % of true value	*-66.3%*	*-34.4%*	*-2.4%*	*-66.3%*	*-37.5%*	*-5.0%*	*-63.4%*	*-69.7%*	*-16.7%*
S.d.	0.0549	0.0546	0.1716	0.057	0.0572	0.175	0.087	0.0873	0.3248
Beta=0.8									
Fi=0									
Lambda									
S.e.	0.0355	0.0299	4.7085	0.0356	0.0298	5.0015	0.0353	0.0269	56.9449
Bias	*-0.0757*	*-0.0689*	*0.0044*	*-0.0758*	*-0.0699*	*-0.4241*	*-0.072*	*-0.0702*	*-15.8254*
S.d.	0.0354	0.0305	8.0247	0.0352	0.0306	9.5895	0.0344	0.036	505.1086
Beta	0.6807	0.7396	0.8012	0.6806	0.7408	0.4154	0.6853	0.7784	-13.822
S.e	0.0328	0.027	4.1968	0.0331	0.027	4.4947	0.0382	0.0248	52.6981
Bias	-0.1193	-0.0604	0.0012	-0.1194	-0.0592	-0.3846	-0.1147	-0.0216	-14.622
Fbias as % of true value	*-14.9%*	*-7.6%*	*0.2%*	*-14.9%*	*-7.4%*	*-48.1%*	*-14.3%*	*-2.7%*	*-1827.8%*
Abias as % of true value					0.3%			6.20%	
S.d.	0.034	0.0277	7.0965	0.0342	0.0281	8.5703	0.0385	0.0408	467.964
Fi=0.9									
Lambda									
S.e.	0.0659	0.0547	1.7056	0.066	0.0545	2.3262	0.0653	0.0485	3.9622
Bias	*-0.0368*	*-0.0617*	*0.0171*	*-0.0365*	*-0.0619*	*-0.0947*	*-0.034*	*-0.0595*	*0.0016*
S.d.	0.0657	0.0502	1.5336	0.0653	0.0498	3.5297	0.0647	0.0503	4.6627
Beta	0.6541	0.7388	0.8136	0.6538	0.7396	0.6908	0.6564	0.7642	0.8488
S.e	0.0601	0.0499	1.5339	0.0603	0.0496	2.1583	0.0629	0.0447	3.296
Bias	-0.1459	-0.0612	0.0136	-0.1462	-0.0604	-0.1092	-0.1436	-0.0358	0.0488
Fbias as % of true value	*-18.2%*	*-7.7%*	*1.7%*	*-18.3%*	*-7.6%*	*-13.7%*	*-18.0%*	*-4.5%*	*6.1%*
S.d.	0.0636	0.0469	1.3586	0.0635	0.047	3.6605	0.0648	0.0514	3.5815
Fi=--0.9									
Lambda									
S.e.	0.0905	0.0770	21.6392	0.0904	0.0768	16.4314	0.0859	0.0508	1.0953
Bias	*-0.1724*	*-0.1264*	*-0.1505*	*-0.1719*	*-0.1293*	*0.7013*	*-0.1525*	*-0.0648*	*0.1337*
S.d.	0.0914	0.0789	33.8497	0.0909	0.0789	23.7062	0.0868	0.0774	2.6257
Beta	0.5961	0.6882	0.6630	0.5967	0.6877	1.4318	0.6183	0.7852	0.9633
S.e	0.0820	0.0695	19.3838	0.0820	0.0694	14.7180	0.0820	0.0469	0.9944
Bias	-0.2039	-0.1118	-0.1370	-0.2033	-0.1123	0.6318	-0.1817	-0.0148	0.1633
Fbias as % of true value	*-25.5%*	*-14.0%*	*-17.1%*	*-25.4%*	*-14.0%*	*79.0%*	*-22.7%*	*-1.9%*	*20.4%*
S.d.	0.0835	0.0710	30.3672	0.0832	0.0710	21.3443	0.0819	0.0778	2.3588

Note - *S.e* .: estimated standard errors;

Bias: absolute value of the finite sample bias (equal to the estimated parameter value in the case of lambda).

S.d.: finite sample bias' experimental standard deviations.

Fbias as % of true value: finite sample bias as a percentage of the true value of beta.

Abias as % of true value: asymptotic bias as a percentage of the true value.

The following table summarises the implications of the assumptions on β and ω when $\phi = 0$:

Table 6-Summary of Experiment Design

	$\beta = \omega = 0.2$	$\beta = 0.8$ $\omega = 0.2$	$\beta = 0.2$ $\omega = 0.8$	$\beta = \omega = 0.8$
Average roots	± 0.4	± 0.9	± 0.4	± 0.9
Range of β_i	$[\pm 0.36]$	$[\pm 0.84]$	$[\pm 0.84]$	$[\pm 0.96]$
Range of ξ_i	$[\pm 0.16]$	$[\pm 0.04]$	$[\pm 0.64]$	$[\pm 0.16]$
Variance of β_i	2.1×10^{-3}	1.3×10^{-4}	3.4×10^{-2}	2.1×10^{-3}
Asyn. bias*	1.8%	0.3%	48.8%	6.2%

* In percent of the true value of β.

References

[1] Alvarez J. and M. Arellano. The time series and cross section asymptotics of dynamic panel data estimators. Mimeo, CEMFI, Madrid (Spain), 1998.

[2] Attanasio O. Et al.. Investment, saving, and growth. *World Bank*, 1998.

[3] Andersen et al.. *Revista Brasilera De Econometria*, 1997.

[4] K. Hadri Abadir K. and E. Tzavalis. The influence of var dimensions on estimator biases. *Econometrica*, 1998.

[5] Canova F. and M. Ciccarelli. Bayesian estimation, inference, and forecasting in var models with panel data. Mimeo, Pompeu Fabra University, Barcellona (Spain), 1999.

[6] Carroll C. D. and D. N. Weil. Saving and growth: A reinterpretation. *Carnegie-Rochester Conference Series on Public Policy*, 40:133–192, 1994.

[7] Ciccarelli M. and A. Rebucci. The transmission mechanism of the european monetary policy: A dynamic panel data approach. Work in progress, 2000.

[8] P. Schmidt Cornwell C. and D. Wyhowski. Simultaneous equations and panel data. *Journal of Econometrics*, 51:151–181, 1992.

[9] Baltagi B. H.. *Econometric Analysis of Panel Data.* Jhon Wiley and Sons, New York, 1995.

[10] Banerjee A.. Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, November 1999.

[11] Hendry D.. Monte carlo experimentation in econometrics. In Griliches Z. and M. D. Intriligator, editors, *Handbook of Econometrics, Vol II*, chapter 16, pages 937–976. Elsevier Science Publishers, 1984.

[12] Hoffmaister A. W. and J. F. Roldos. Are business cycles different in asia and latin america? *IMF Working Paper No. 97/79*, 1997.

[13] Hsiao C.. *Analysis of Panel Data*. Cambridge University Press, Cambridge, 1986.

[14] M. H. Pesaran Hsiao C and A. K. Tahmiscioglu. Bayes estimation of short-run coefficients in dynamic panel data models. http//, 1997.

[15] W. Newey Holtz-Eakin D. and H. Rosen. Estimating vector autoregressions with panel data. *Econometrica*, 56(6):1371–1395, 1988.

[16] W. Newey Holtz-Eakin D. and H. Rosen. The revenues-expenditure nexus: Evidence from local government data. *International Economic Review*, 30(2):415–429, May 1989.

[17] Hoogstrate A. J.. *Dynamic Panel Data Models: Theory and Macroeconomic Applications*. PhD thesis, Maastricht University, 1998.

[18] Kiviet J. F.. Expectations of expansions for estimators in a dynamic panel data model; some results for weakly-egogenous regressors. In K. Lahiri Hsiao C. and M. H. Pesaran, editors, *Analysis of Panels and Limited Dependent Variables: A Volume in Honor of G. S. Maddala*. Cambridge Univerrsity Press, Cambridge, 1998.

[19] Krishnakumar J.. Simultanous equations. In Matyas L. and P. Sevestre, editors, *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, chapter 9, pages 196–234. Kluwer Academic Publishers, 1996.

[20] Judson R. A. and A. L. Owen. Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters*, 65:9–15, 1999.

[21] C. B. Barret Mohapatra S. and D. Snyder. The dynamic effects of us food aid on recipient country agriculture. November 1996.

[22] Phillips P. C. B. and H. Moon. Linear regression theory for nonstationary panel data. *Econometrica*, 1999.

[23] Pesaran M. H. and R. Smith. Estimating long run relationships from dynamic heterogeneous panels. *Juournal of Econometrics*, 68:79–113, 1995.

[24] R. Smith Pesaran M. H. and K. S. Im. Dynamic linear models for heterogeneous panels. In Matyas L. and P. Sevestre, editors, *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, chapter 8, pages 145–196. Kluwer Academic Publishers, Boston, MA, 1996.

[25] Pesaran M. H. and Z. Zhao. Bias reduction in estimating long-run relationship from dynamic heterogeneous panels. 1997. http
.

[26] Rebucci A.. External shocks, macroeconomic policy and growth: A panel var approach. *GEI Working Paper No. 40*, 1998.

[27] Robertson D. and J. Simons. Some strange propoerties of panel data estimators. *Journal of Applied Econometrics*, 7:175–189, 1992.

[28] J. F. Roldos Hoffmaister A. W. and P. Wickham. Macroeconomic fluctuations in sub-saharan africa. *IMF Staff Papers*, 1998.