WP/03/73

IMF Working Paper

On the Heterogeneity Bias of Pooled Estimators in Stationary VAR Specifications

Alessandro Rebucci

INTERNATIONAL MONETARY FUND

© 2003 International Monetary Fund

IMF Working Paper

Research Department

On the Heterogeneity Bias of Pooled Estimators in Stationary VAR Specifications

Prepared by Alessandro Rebucci [1]

Authorized for distribution by Tamim A. Bayoumi

April 2003

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper studies asymptotically the bias of the fixed effect (FE) estimator induced by cross-section heterogeneity in the slope parameters of stationary vector autoregressions (VARs). The paper also compares the FE, the mean group estimator (MG), and a simple instrumental variable alternative (IV) in Monte Carlo simulations. The main results are: (i) asymptotically, the heterogeneity bias of the FE may be more or less severe in VAR specifications than in standard dynamic panel data specifications; (ii) in Monte Carlo simulations, slope heterogeneity must be relatively high to be a source of concern for pooled estimators; (iii) when this happens, the panel must be longer than a typical macro dataset for the MG to be a viable solution.

JEL Classification Numbers: C33, C13, C15

Keywords: Dynamic Panel Data Models, Monte Carlo Simulation, Heterogeneity Bias, VARs

Author's E-Mail Address: ARebucci@imf.org

[1] This paper draws on chapter two of the author's Ph.D. dissertation at Queen Mary University of London (QMUL). It circulated previously as "Estimating VARs with Long Heterogeneous Panels." The author is grateful to his advisor, Christopher Gilbert, and to Manuel Arellano, Valentina Corradi, Kyung So Im, Ron Smith, Giovanni Urga, and seminar participants in at QMUL and the 2000 World Congress of the Econometric Society for suggestions, discussions, comments, and encouragement. Remaining errors are the responsibility of the author.

Contents Page

I. Introduction....3

II. The Model and Alternative Estimation Strategies....5
A. The Model....5
B. Alternative Estimation Strategies....7

III. Asymptotic Analysis....10
A. Notation....10
B. Results....12
The General Case....12
A First Special Case: Weak Exogenity Fails....15
A Second (Very) Special Case: Granger Non-Causality Fails....16

IV. Monte Carlo Analysis....18
A. Experiment Design....19
B. Results....21
Homogeneous Panels....21
Heterogeneous Panels....23

V. Conclusions....26

References....42

Figures
1. Heterogeneity Bias of $\hat{\beta}_{FE}$ for y=1 (in absolute value)....36
2. Heterogeneity Bias of $\hat{\beta}_{FE}$ for y=1 (in percent of true value)....36

Tables
1. Monte Carlo Results (N, T = 50, 50)....37
2. Monte Carlo Results (N, T = 20, 50)....38
3. Monte Carlo Results (N, T = 10, 50)....39
4. Monte Carlo Results (N, T = 50, 20)....40
5. Monte Carlo Results (N, T = 20, 20)....41

Appendix I: The Heterogeneity Bias of the Fixed Effects Estimator....28
A. The General Case....28
B. A Special Case: Weak Exogeneity Fails....32

I. Introduction

Vector autoregressive systems (VARs) are a useful device to summarise and analyse the dynamic interaction of a given set of variables of interest as originally proposed by Sims (1980). When there are several decision units to be considered (i.e., several agents, countries, or sectors)On the one hand, the possibility of pooling them in a single system emerges. Pooling different decision units is attractive because it increases the number of degrees of freedom available and, potentially, the efficiency of the estimates so obtained; thus, it potentially also reduces the risk of overfitting.[2] On the other hand, pooling different decision units poses inferential problems with regard to the representative or typical unit: it may introduce an aggregation bias, if the slope parameters of individual regressions are heterogeneous, which is called "heterogeneity bias" in this literature.

We can think of a VAR estimated with panel data (a Panel VAR or $PVAR$) 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), and/or the unconditional variance (the variance of the error term in the regression equation), of the variables of interest. 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 by 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 DPM 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 the *mean group estimator*, 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 extends some of the results for heterogeneous $DPMs$ of Pesaran and Smith (1995); Pesaran, Smith and Im (1996) and Hsiao, Pesaran, and Tahmiscioglu (1997) to a $PVAR$ specification.

In the broader context of the existing theoretical literature on $PVARs$, the analysis carried out in the paper is limited in scope. First, consistent with the rest of the literature, I shall restrict my attention to exactly identified $VARs$ in the time series sense and hence focus on the estimation of the reduced-form of the model.[3] Second, I shall assume that slope parameters are constant

---

[2] The risk of overfitting is underlined by proponents of a Bayesian approach to estimation of $VARs$ such as Doan, Litterman, and Sims (1984).

[3] On the difficulties arising from the interaction between estimation and identification issues in dynamic panel simultaneous equation models, see Krishnakumar (1996).

over time; consider only stationary systems; and, unlike most of the existing literature, focus only on the estimation of the short-run dynamics of the system as Hsiao, Pesaran, and Tahmiscioglu (1997) do.[4] Third, I neglect in part interdependence between individual units by assuming that this can be satisfactorily modelled through the inclusion of common, exogenous, and observable variables in each individual VAR as assumed.[5] Finally, motivated by typical macroeconomic applications such as those using the Heatson and Summers (1991) dataset, as is most of the literature, I consider only long panels and pay particular attention to unfavorable panel dimensions in the Monte Carlo simulations. Nonetheless, the reduced-form model studied in the paper may be applied to the analysis of the dynamic impact and the relative importance of different shocks—as for instance done by Rebucci (1998)—or the analysis of Granger causality issues—as for instance done by Carrol and Weil (1994)—when either economic theory or prior analysis of individual time series indicates that stationarity is assured.[6]

Within the boundaries of these limitations, this paper studies the determinants of the heterogeneity bias of the fixed effect estimator (FE) in a model in which the regressors are not strongly exogenous, because either weak exogeneity or Granger causality fails, and studies the finite sample properties of the FE, MG, and a simple instrumental variable estimator (IV), by means of Monte Carlo simulations, in a model in which both weak endogeneity and Granger causality fail.

The main results of the analysis are that (i) asymptotically, the heterogeneity bias of the FE may be more or less severe in VAR specifications than in standard DPM specifications; (ii) in Monte Carlo simulations, slope heterogeneity must be relatively high to be a source of concern for pooled estimators; (iii) when this happens, the panel must be longer than a typical macro dataset for the MG to be a viable solution. The main implication of the analysis is that empirical Bayesian estimators such as those proposed by Hsiao, Pesaran and Tahmiscioglu (1997) and Canova and Ciccarelli (2000) seem more promising alternatives to estimate VARs with heterogeneous panel data.

---

[4] For surveys on the now large literature on nonstationary DPMs and testing for unit root and cointegration in panel data, see Banerjee (1999), Phillips and Moon (2000), and Simith (2000). For extensions of some of these results to a general VAR specification, see Larsson et al. (1998 and 1999) and Banerjee et al. (2000). Note, however, that these latter contributions bring the analysis back to a pure time series dimension, thus essentially defeating the purpose of using panel data estimators to improve efficiency and hence reduce the risk of overfitting. See Holtz-Eakin et al. (1988) for a framework in which parameters may change over time, and hence stationarity is not required, but must be homogeneous across-section.

[5] Both seminal contributions of Pesaran and Smith (1995) and Phillips and Moon (1999) assume cross-section independence. To my knowledge, Robertson and Symons (2000) were the first in this literature to develop a seemingly unrelated regression model allowing for some, limited cross-section interdependence for panel data sets of non trivial sectional dimension. Alternative approaches to modeling cross-section interdependence typical of macro data sets include dynamic factor analysis pioneered by Forni et al. (2001) and numerical Bayesian estimation of large time series VARs proposed by Canova and Ciccarelli (2000). But these contributions use rather different technologies than those used in this paper.

[6] See Boyd and Smith (2000) and Attanasio and others (1999) for comparisons of the estimators analysed in this paper with actual data.

The paper is organised as follows. Section II spells out the model and discusses alternative estimation strategies. Section III studies the bias of the FE estimator asymptotically. Section IV sets up the Monte Carlo experiment and reports the finite sample results. Section V concludes. The derivation of the asymptotic bias of the FE estimator in the most general case considered and the analysis of one of the two special cases considered are reported in the appendix. The GAUSS code for the Monte Carlo exercise is available on request.

II. The Model and Alternative Estimation Strategies

A. The Model

Consider the following general VAR describing the behavior of the i^{th} individual unit:

$$Y'_{i,t} = \sum_{j=1}^{N} A^{j\prime}_{i,s}(L)Y'_{j,t-1} + \chi'_{i,s}(L)d'_t + \alpha'_i + \varepsilon'_{i,t}, \tag{1}$$

with

$$i = 1, \cdots N; \quad t = 1, \cdots T; \quad s = 1, \cdots S \quad \text{and } S \leq T.$$

Here, $Y'_{i,t}$ and d'_t denote, respectively, a $(Mx1)$ and $(Kx1)$ vector of individual and time specific and common-across-individuals observable variables of interest; $A^{j\prime}_{i,s}(L)$ and $\chi'_{i,s}(L)$ are (MxM) and (MxK) time-varying matrix polynomials in the lag operator L (e.g., $LY'_{i,t} = Y'_{i,t-1}$), of order p and q, respectively; α'_i is a $(Mx1)$ vector of individual specific fixed or random effects; $\varepsilon'_{i,t}$ is a $(Mx1)$ vector of error terms with $\varepsilon'_{i,t} \sim iid(0, \Sigma_{i,s})$; and S denotes the number of sub-samples.

This is a general heterogeneous $PVAR$ in that, in addition to unconstrained contemporaneous and lagged individual units' interdependence, it allows for the maximum degree of parameter heterogeneity, places few restrictions on the data as far as stationarity and exogeneity is concerned, and is potentially suitable for forecasting as well as for inference and policy analysis. Unfortunately, however, this model cannot be estimated in most commonly encountered contexts without imposing additional restrictions.

In the rest of the paper, as anticipated in the introduction, I shall make the following assumptions, for all i:

(i)The VAR in (1) is a covariance-stationary, mean square ergodic process and its parameters are constant over time.

(ii)Individual units are not interrelated except for common exogenous factors; thus, $\sum_{j=1}^{N} A^{j\prime}_{i,}(L) = A^{i\prime}_{i}(L)$ with $E(\varepsilon'_t\varepsilon_t) = (I \otimes \Sigma_i)$, where $\varepsilon'_t = \begin{bmatrix} \varepsilon'_{1,t} & \cdots & \varepsilon'_{N,t} \end{bmatrix}'$, E denotes the expectation operator with respect to the distribution of ε'_t, I is an identity matrix of conforming dimension, and $\otimes$ denotes the Kronecker product.

(iii)In addition, following Pesaran and Smith (1995), I assume that α'_i is a vector of constants to be estimated (a vector of pure fixed effects) and A'_i varies across individual units according to the random coefficient specification:

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

where A' is a (MxM) constant matrix and η'_i is a (MxM) random matrix distributed independently of $\varepsilon'_{i,t}$ and $Y'_{i,t}$, with zero mean and constant variance-covariance matrix equal to Ω—i.e., $vec(\eta'_i) \sim iid(0, \Omega)$. Thus, individual specific effects are fixed while the slope parameters of the VAR vary randomly across section and are distributed independently of the regressors and the error terms.

I assume further and without loss of generality that $p = 1$ and $M = 2$ and $\chi'_i(L) = 0$ for all i.[7] Then, model (1) becomes the following stationary, bivariate heterogenous $PVAR$ of first order:

$$\begin{bmatrix} z_{i,t} \\ x_{i,t} \end{bmatrix} = \begin{bmatrix} \lambda_i & \beta_i \\ \gamma_i & \rho_i \end{bmatrix} \begin{bmatrix} z_{i,t-1} \\ x_{i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_i^z \\ \alpha_i^x \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,t}^z \\ \varepsilon_{i,t}^x \end{bmatrix}, \tag{3}$$

$$i = 1, \cdots N; \quad t = 1, \cdots T;$$

where

$$Y'_{i,t} = \begin{bmatrix} z_{i,t} & x_{i,t} \end{bmatrix}', \alpha'_i = \begin{bmatrix} \alpha_i^z & \alpha_i^x \end{bmatrix}', \varepsilon'_{i,t} = \begin{bmatrix} \varepsilon_{i,t}^z & \varepsilon_{i,t}^x \end{bmatrix}',$$

$$A'_i = \begin{bmatrix} \lambda_i & \beta_i \\ \gamma_i & \rho_i \end{bmatrix}, \text{ and } E(\varepsilon'_{i,t}\varepsilon_{i,t}) = \Sigma_i = \begin{bmatrix} \sigma_i & \phi_i \\ \phi_i & \tau_i \end{bmatrix}.$$

It is now easily seen that Pesaran and Smith's (1995) DPM may be interpreted as a *restricted* heterogenous $PVAR$ in which, in addition to the hypothesis (i)-(iii) above, $\gamma_i = 0$ and the "correlation" between the variables considered has been organized as "economic causation" from $x_{i,t}$ to $z_{i,t}$. Pesaran and Smith (1995) specify the following heterogeneous dynamic panel data model (DMP):

$$z_{i,t} = \widetilde{\lambda}_i z_{i,t-1} + \widetilde{\varphi}_i x_{i,t} + \widetilde{\alpha}_i^z + u_{i,t}^z, \tag{4}$$

$$x_{i,t} = \widetilde{\rho} x_{i,t-1} + \widetilde{\alpha}_i^x (1 - \widetilde{\rho}) + u_{i,t}^x,$$

with

$$\lambda_i = \lambda + \eta_i^\lambda \quad \varphi_i = \varphi + \eta_i^\varphi \tag{5}$$

where $\eta'_i = \begin{bmatrix} \eta_i^\lambda & \eta_i^\varphi \end{bmatrix}'$ has zero mean and constant variance-covariance matrix (Θ) and is distributed independently of $Y'_{i,t} = \begin{bmatrix} z_{i,t} & x_{i,t} \end{bmatrix}'$ and $u'_{i,t} = \begin{bmatrix} u_{i,t}^z & u_{i,t}^x \end{bmatrix}'$, $E(u_{i,t}^z u_{i,t}^z) = v_i^z$, $E(u_{i,t}^x u_{i,t}^x) = v_i^x$, and $E(u_{i,t}^z u_{i,t}^x) = E(u_{i,t}^x u_{i,t}^z) = 0$.[8] If we multiply (3) by the inverse of the

[7] Assuming $p = 1$ and $M = 2$ is certainly not restrictive for the purpose of comparing results with the DPM literature. In addition, a VAR of order p may always be represented in companion form as a VAR of order one. As shown by Abadir et a. (1999), however, the number of variables entering a VAR may not only affect the efficiency but also the biases of the estimator used.

[8] Here, the process for $x_{i,t}$ does not need to be univariate and its lagged values can be included in the equation for $z_{i,t}$ without affecting the properties of the parameter estimates.

unique matrix Φ_i such that $\Sigma_i = \Phi_i V_i \Phi_i'$ and drop (without loss of generality) the term $\widetilde{\beta}_i x_{i,t-1}$ in the equation for $z_{i,t}$ above, where

$$\Phi_i = \begin{bmatrix} 1 & \widetilde{\varphi}_i \\ 0 & 1 \end{bmatrix}, \quad V_i = \begin{bmatrix} v_i^z & 0 \\ 0 & v_i^x \end{bmatrix},$$

with $\widetilde{\lambda}_i = (\lambda_i - \widetilde{\varphi}_i \gamma_i) = \lambda_i$, $\widetilde{\beta}_i = (\beta_i - \widetilde{\varphi}_i \rho_i)$, $\widetilde{\alpha}_i^z = (\alpha_i^z - \widetilde{\varphi}_i \alpha_i^x)$, $u_{i,t}^z = (\varepsilon_{i,t}^z - \widetilde{\varphi}_i \varepsilon_{i,t}^x)$, $u_{i,t}^x = \varepsilon_{i,t}^x$, and $E(u_{i,t}' u_{i,t}) = E(\Phi^{-1'} \varepsilon_{i,t}' \varepsilon_{i,t} \Phi^{-1}) = V_i$, we then find that

$$\begin{bmatrix} 1 & -\widetilde{\varphi}_i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_{i,t} \\ x_{i,t} \end{bmatrix} = \begin{bmatrix} \widetilde{\lambda}_i & \widetilde{\beta}_i \\ 0 & \rho_i \end{bmatrix} \begin{bmatrix} z_{i,t-1} \\ x_{i,t-1} \end{bmatrix} + \begin{bmatrix} \widetilde{\alpha}_i^z \\ \alpha_i^x \end{bmatrix} + \begin{bmatrix} u_{i,t}^z \\ u_{i,t}^x \end{bmatrix},$$

which is exactly the same as (4).[9] Thus, the key assumption distinguishing (3) from (4) is that, in the latter, $x_{i,t}$ is *weakly exogenous* for the estimation of $\widetilde{\lambda}_i$, $\widetilde{\varphi}_i$ *and* $z_{i,t}$ *does not Granger-cause* $x_{i,t}$—i.e., $x_{i,t}$ is *strongly exogenous* for the estimation of $\widetilde{\varphi}_i$ and $\widetilde{\lambda}_i$ implying that the process of $z_{i,t}$ and $x_{i,t}$ can be estimated separately.[10]

B. Alternative Estimation Strategies

Suppose one is interested in estimating A (or λ and φ) the cross-sectional average of A_i (or λ_i and φ_i, respectively). When T is large enough to estimate individual time series regressions separately, this can be obtained in three different ways.[11] First, by stacking the data and using pooled estimators such as the FE estimator (sometime called also least squares dummy estimator, within estimator, or covariance estimator), the random effect estimator (RE), or instrumental variables-type estimators (IV), possibly correcting for cross section heteroschedasticity in the variance of the innovations $U_{i,t}'$ if necessary. Second, by averaging data across section and estimating an aggregate time series regression (ATS). Third, by estimating individual time series regressions and averaging these estimates across section or groups, a procedure called mean group (MG) estimation by Pesaran and Smith (1995).[12]

If the panel is not only *long* but also *homogeneous* in the slope parameters—i.e., $\eta_i' = 0$ for all

---

[9] As known, this decomposition exists always but is not unique and depends on the variables' order. The complications involved in moving from a standard DPM specification to a one in which explanatory variables are only *weakly* rather than *strongly* exogenous are discussed also by Kiviet (1998).

[10] Pesaran and Smith (1995) actually assume that $x_{i,t}$ is *strictly* exogenous (i.e., is independent of $u_{i,t}^z$ at all lads and lags) for the estimation of λ_i and φ_i in the equation for $z_{i,t}$. This implies that parametrizing the equation for $z_{i,t}$ differently, as for instance by inverting the transformation discussed in the previous sub-section, weak exogeneity of $x_{i,t}$ for the estimation of λ_i and φ_i in the equation for $z_{i,t}$ could fail.

[11] There is also a fourth method that is averaging the data over time and estimating an aggregate cross section regression. While this estimator (sometime called between estimator) has better asymptotic properties than pooled or aggregate time series estimators when the panel is heterogeneous, it does not allows for estimation of the model's short run dynamic, and thus is not considered here.

[12] The average of time series estimates may be weighted or unweighted, in principle. In the paper, I used only unweighted averages. The weighted average of the time series estimates is sometime called 'Swamy estimator', as was originally proposed by Swamy (1970) for the estimation of static models with randomly varying slope parameters, or empirical-Bayes estimator, as it can be interpreted as a 'mixed estimator' in the sense of Theil (1971).

i—then all three estimation procedures yield *consistent* estimates of the parameters of interest for large T and fixed N, even though they are all biased in finite samples because of the presence of the lagged dependent variable.[13] In this case, the choice among alternative estimators ought to be dictated by efficiency considerations based on assumptions on the nature of the individual specific effects (α_i') the initial condition of the data ($z_{i,0}$ and $x_{i,0}$), and the particular dimension of the dataset at hand.

The FE estimator is asymptotically equivalent to the RE iestimator in terms of efficiency, but since the latter is inconsistent when the individual specific effects are correlated with the regressors even for large T, the former is generally preferable. Simple IV-type estimators and generalized method of moments-type estimators (GMM) are consistent also for large N and fixed T.[14] In this case, GMM-type estimators are more efficient than simple IV estimators, but they have been shown to perform worse when T is relatively large because of overfitting problems. Therefore, the question of how large should be T relative to N to prefer the FE estimator to IV-type of estimators in long, homogeneous $DPMs$ remains open.[15] In addition, the FE estimator and IV-type estimators have recently been shown to be asymptotically equivalent in terms of efficiency when both N and T are large, but their asymptotic biases depend on the rates at which N and T increase in this case.[16]

If the panel is *long* and *heterogeneous* in the slope parameters, Pesaran and Smith (1995) have shown that pooled estimators (the FE estimator as well as IV-type estimators) and the ATS estimator generally yield inconsistent estimates of λ and φ, regardless of the time dimension of the panel, while the MG estimator is consistent for both N and T large.[17]

To see why pooled estimators cannot be consistent, substitute (5) in the first equation of (4). The model becomes:

$$z_{i,t} = \lambda z_{i,t-1} + \varphi x_{i,t} + \alpha_i^z + w_{i,t}^z, \qquad w_{i,t}^z = u_{i,t}^z + \eta_i^{\lambda} z_{i,t-1} + \eta_i^{\varphi} x_{i,t}. \tag{6}$$

It is now evident that the new error term, $w_{i,t}^z$, is contemporaneously correlated with the regressors and also autocorrelated to extent to which the regressors are autocorrelated. Similarly, averaging (4) across section (and denoting simple averages with over-bars), shows that the new aggregate

---

[13] See Nickel (1981) and Anderson and Hisiao (1981 and 1982) on the FE and simple IV-type estimators; see Pesaran and Smith (1995) and the references quoted therein on the ATS estimator.

[14] The literature on *short, homogeneous* $DPMs$ is vast and reviewed in any textbook on panel data analysis.

[15] See Judson and Owen (1999) for Monte Carlo simulation evidence on the relative performance of FE and IV-type estimators in relatively long, homogeneous panels.

[16] See Arellano and Alvarez, (1998) on this point.

[17] See Hsiao, Pesaran, and Tahmiscioglu (1999) for alternative Bayesian estimators when the panel is not only *heterogeneous* in the slope parameters, but also *short*.

error term is not independent of the aggregate regressors:

$$\overline{z}_t = \lambda \overline{z}_{t-1} + \varphi \overline{x}_t + \overline{w}_t^z \qquad \overline{w}_t^z = \overline{u}_t^z + \sum_{i=1}^{N} (\eta_i^\lambda z_{i,t-1} + \eta_i^\varphi x_{i,t}).$$

Pesaran and Smith (1995) argue also that standard corrections for error autocorrelation are unlikely to solve this problem given the structure of the composite error terms ($w_{i,t}^z$ and $\overline{w}_t^z$). Similarly, they show that IV estimation can work only in very special cases.

More specifically, they study the heterogeenity bias of the FE estimator and show that, when there is only one source of slope heterogeneity, the probability limit of $\widehat{\lambda}_{FE}$ and $\widehat{\varphi}_{FE}$ is:

$$\underset{N\to\infty,T\to\infty}{\text{plim}} \begin{pmatrix} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\varphi}_{FE} - \delta \end{pmatrix} = \begin{pmatrix} \frac{\rho(1-\lambda\rho)(1-\lambda^2)\theta}{\Psi} \\ -\frac{\varphi\rho^2(1-\lambda^2)\theta}{\Psi} \end{pmatrix} \tag{7}$$

where

$$\Psi = (v_i^z/v_i^x)\left(1-\rho^2\right)(1-\lambda\rho)^2 + \left(1-\lambda^2\rho^2\right)\theta + \left(1-\rho^2\right)\beta.$$

The size of this bias depends upon: (i) on the mean coefficients λ, φ, ρ; (ii) the variance of φ_i, denoted θ; (iii) and the ratio (v_i^z/v_i^x), with $\widehat{\varphi}_{FE}$ always underestimating φ, and $\widehat{\lambda}_{FE}$ over or underestimatimating λ depending on wether ρ is positive or negative. The bias disappears only if $\rho = 0$ or $\theta = 0$, or if λ approaches one from below when $\rho \neq 0$ and $\theta \neq 0$. Moreover,

$$\underset{\rho\to 1}{\text{plim}}(\widehat{\lambda}_{FE}) = 1 \quad \underset{\rho\to 1}{\text{plim}}(\widehat{\varphi}_{FE}) = 0,$$

irrespective of the true values of λ and φ.

In the case of a VAR specification, the MG estimator is the natural benchmark because it is consistent under both heterogeneity and homogeneity, even though it could be less efficient than the FE under homogeneity.[18] In addition, as noted, the RE estimator is asymptotically equivalent to the FE estimator in DMP specifications if the individual effects are uncorrelated with the regressors, but is not consistent when this assumption is violated. In order to avoid making specific assumptions on the properties of the individual effects, I consider only the EF estimator in the rest of the paper. At the same time, the ATS estimator is unattractive even under slope homogeneity in VAR specifications because it does not increase the number of degrees of freedom available, which is often a critical issue in this context. Therefore, I shall not pursue this

---

[18] The consistency of $\widehat{\lambda}_{MG}$, $\widehat{\theta}_{MG}$, and their estimated varaince-covariance matrix,

$$\widehat{\Omega}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\left[(\widehat{\lambda}_{iOLS} - \widehat{\lambda}_{MG}), (\widehat{\varphi}_{iOLS} - \widehat{\varphi}_{MG})\right]' \left[(\widehat{\lambda}_{iOLS} - \widehat{\lambda}_{MG}), (\widehat{\varphi}_{iOLS} - \widehat{\varphi}_{MG})\right],$$

is proven by Pesaran, Smith, and Im (1996). Hisiao, Pesaran, and Tahmiscioglu (1999) show also that the MG estimator is asymptotically normal for large N and T as long as $\sqrt{N}/T \to 0$ as both N and $T \to \infty$. These results may be easily generalized to a VAR specificatiuon. The proof of the consistency of the MG estimator and a discussion of its asymptotic properties is available on request.

alternative estimation procedure further here.

Unlike in DPM specifications, as noted implicitly by Holtz-Eakin et al. (1988), there are no special cases in which one can find valid instruments for consistent estimation of the parameters of a heterogenous $PVAR$. This is because both lagged $z_{i,t}$ and lagged $x_{i,t}$ depend on η_i^λ and η_i^φ in a VAR specification, and hence contemporaneous and lagged $x_{i,t}$ are correlated with the composite error term $w_{i,t}$ even if the there is only one source of slope heterogeneity (i.e., $\eta_i^\lambda = 0$ for all i).[19] More generally, exogenous variables that are uncorrelated with $w_{i,t}$ will also be uncorrelated with the regressors. Furthermore, as noted before, the class of IV-type estimators for homogeneous $DPMs$ is wide, ranging from simple first (or quasi-first) difference estimators to computationally more demanding GMM-type estimators, but there is no consensus yet in the literature on which is the most appropriate choice when the panel is long. Therefore, I shall not investigate the heterogeneity bias of IV-type estimators asymptotically and will consider only a simple IV estimator that has been shown to preform well when T is relatively large (by Judson and Owen, 1999) in the Monte Carlo experiments.

In the next section, therefore, I study the heterogeneity bias of the FE estimator, asymptotically. In the following one, I compare the performance of the MG estimator with that of the FE estimator and a simple IV alternative by means of Monte Carlo simulations.

III. Asymptotic Analysis

A. Notation

In order to derive the FE estimator and its properties we need to establish some notation.

Let us transpose (1) and (2) and substitute the latter in the former to obtain:

$$Y_{i,t} = Y_{i,t-1}A_i + \alpha_i + \varepsilon_{i,t}, \qquad A_i = A + \eta_i \tag{8}$$

$$Y_{i,t} = Y_{i,t-1}A + \alpha_i + \nu_{i,t}, \qquad \nu_{i,t} = \varepsilon_{i,t} + Y_{i,t-1}\eta_i, \tag{9}$$

$$i = 1, \cdots N; \quad t = 1, \cdots T,$$

where $Y_{i,t} = [y_{i,t}^1, \cdots, y_{i,t}^M]$, $Y_{i,t-1} = [y_{i,t-1}^1, \cdots, y_{i,t-1}^M]$, $\alpha_i = [\alpha_i^1, \cdots, \alpha_i^M]$, $\varepsilon_{i,t} = [\varepsilon_{i,t}^1, \cdots, \varepsilon_{i,t}^M]$, $\nu_{i,t} = [\nu_{i,t}^1, \cdots, \nu_{i,t}^M]$ have all dimension $(1xM)$.

Collect all T time observations for each individual unit i in the (TxM) vectors

$$Y_i = \begin{bmatrix} Y_{i,1} \\ \vdots \\ Y_{i,T} \end{bmatrix}, Y_{i,-1} = \begin{bmatrix} Y_{i,0} \\ \vdots \\ Y_{1,T-1} \end{bmatrix}, \nu_i = \begin{bmatrix} \nu_{1,1} \\ \vdots \\ \nu_{i,T} \end{bmatrix}, \overline{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i,0} \\ \vdots \\ \varepsilon_{i,T} \end{bmatrix},$$

[19] See Pesaran, Smith, and Im, 1996, pp. 149-150.

to write (8) as:

$$Y_i = Y_{i,-1}A_i + (\alpha_i \otimes i_T) + \overline{\varepsilon}_i \qquad i = 1, \cdots N. \tag{10}$$

Applying the vec operator to both sides of (10) and defining $y_i = vec(Y_i)$, $X_i = (I_M \otimes Y_{i,-1})$, $a_i = vec(A_i)$, $\underline{\varepsilon}_i = vec(\overline{\varepsilon}_i)$, for all T time observations and each individual unit i, the model can be represented also in SUR format as:

$$y_i = X_i a_i + (\alpha_i \otimes i_T) + \underline{\varepsilon}_i, \qquad i = 1, \cdots N; \tag{11}$$

where y_i, $(\alpha_i \otimes i_T)$, and $\underline{\varepsilon}_i$ have dimension $TMx1$, X_i has dimension $TMxM^2$, and a_i has dimension M^2x1.

Similarly, stack all N time series in the $(NTxM)$ vectors:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}, Y_{-1} = \begin{bmatrix} Y_{1,-1} \\ \vdots \\ Y_{N,-1} \end{bmatrix}, \overline{\nu} = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_N \end{bmatrix} \overline{\varepsilon} = \begin{bmatrix} \overline{\varepsilon}_1 \\ \vdots \\ \overline{\varepsilon}_N \end{bmatrix},$$

$$\overline{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \otimes i_T = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix},$$

with i_T denoting a $(Tx1)$ vector of ones, to write (9) as:

$$Y = Y_{-1}A + \overline{\alpha} + \overline{\nu} \qquad \overline{\nu} = \overline{\varepsilon} + Y_{-1}\eta_i. \tag{12}$$

Applying the vec operator to both sides of (12) and defining $y = vec(Y)$, $X = (I_M \otimes Y_{-1})$, $a = vec(A)$, $\alpha = vec(\overline{\alpha})$, $\nu = vec(\overline{\nu})$, $\varepsilon = vec(\overline{\varepsilon})$, the model can be rewritten in SUR format, for all T time observations and N individual units, as

$$y = Xa + \alpha + \nu \qquad \nu = \varepsilon + Xvec(\eta_i); \tag{13}$$

where y, α, ν, and ε have dimension $NTMx1$, X has dimension $NTMxM^2$, and a and $vec(\eta_i)$ have dimension M^2x1.

Now define the following matrix operators:

$$D = I_N \otimes i_T;$$

$$P_D = D(D'D)^{-1}D' = I_N \otimes i_T i_T'/T = I_N \otimes i_T(i_T' i_T)^{-1} i_T';$$

$$Q_D = I_{NT} - P_D = I_{NT} - D(D'D)^{-1}D' = I_N \otimes [I_T - i_T(i_T' i_T)^{-1} i_T'];$$

$$P = I_M \otimes P_D;$$

$$Q = I_M \otimes Q_D;$$

where D is the usual matrix of individual dummies, P_D and Q_D are the usual (symmetric and idempotent) 'between' and 'within' operator, respectively (e.g., Baltagi, 1995), with I_N, I_T, I_M, I_{NT} denoting identity matrices of conforming dimension. P and Q generalise the latter two operators to a system of seemingly unrelated regressions (SUR) (see Cornwell, Schmidt, and Wyhowski, 1992).

Finally, define $H_T = [I_T - i_T(i_T' i_T)^{-1} i_T']$ so that $Q_D = I_N \otimes H_T$. Noting that $i_T' i_T = T$, we can see by direct inspection that H_T transforms any row vector of T elements in deviations from their average. It follows from this that the operator $(I_M \otimes H_T)$ transforms all M components (of dimension $Tx1$) of vectors like y_i (of dimension $TMx1$) in deviations from their time averages; and hence $(I_M \otimes H_T)$ transforms the stacked vector of all T time observations on all M variable of system (11) in deviations from time averages, for each individual unit i. It is easily seen that $(I_M \otimes H_T)$ is also symmetric and idempotent as it has the same matrix structure as Q_D.

B. Results

The General Case

Take deviations from time averages for all individual units N by applying the generalized within operator Q to (13), to obtain:

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

where $\widetilde{y} = Qy$, $\widetilde{X} = QX$, $\widetilde{\nu} = Q\nu$, and $Q\alpha = 0$. The FE estimate of a therefore is:

$$\widehat{a}_{FE} = \left(\widetilde{X}'\widetilde{X}\right)^{-1}\left(\widetilde{X}'\widetilde{y}\right). \tag{14}$$

Proving the inconsistency of the FE estimator is a bit more tedious. In the appendix at the end of the chapter, I show that:

$$\underset{N\to\infty, T\to\infty}{\text{plim}} \left(\widehat{a}_{FE} - a\right) = \begin{bmatrix} \left(E\left[\Lambda_i\right]\right)^{-1}\left(E\left[\Lambda_i \eta_i^1\right]\right) \\ \vdots \\ \left(E\left[\Lambda_i\right]\right)^{-1}\left(E\left[\Lambda_i \eta_i^M\right]\right) \end{bmatrix} \tag{15}$$

where

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

with Λ_i denoting the unconditional variance-covariance matrix of the data, η_i^j the j^{th} $(Mx1)$ column of $vec(\eta_i)$ for $j = 1, \cdots M$, and E the expectation with respect to the joint distribution of A_i and Σ_i.

Under stationarity, the expectations in this equation are well defined and generally different from zero. Equation (15), therefore, shows that the heterogeneity bias of the FE estimator is asymptotically different from zero in general. In principle, an explicit solution for the heterogeneity bias of the FE estimator can be obtained computing these expectations under suitable distributional assumptions for A_i and Σ_i. In practice, however, even the simplest, heterogeneous VAR specification has no closed-form solution.

Consider, for instance, (3) with only one source of slope heterogeneity and homoskedastic error terms:

$$\begin{aligned} z_{i,t} &= \lambda z_{i,t-1} + \beta_i x_{i,t-1} + \alpha_i^z + \varepsilon_{i,t}^z, \\ x_{i,t} &= \gamma z_{i,t-1} + \rho x_{i,t-1} + \alpha_i^x + \varepsilon_{i,t}^x; \end{aligned} \tag{16}$$

where

$$A' = \begin{bmatrix} \lambda & \beta \\ \gamma & \rho \end{bmatrix}, \eta_i' = \begin{bmatrix} 0 & \xi_i \\ 0 & 0 \end{bmatrix} \text{ and } \Sigma_i = \Sigma = \begin{bmatrix} \sigma & \phi \\ \phi & \tau \end{bmatrix}.$$

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 matrix is 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:

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

By direct inspection, we can also see that Λ_i must equal the $(2x2)$ matrix,

$$\frac{1}{\Upsilon_{i,0}} \begin{bmatrix} \Xi_{i,1} & \Xi_{i,3} \\ \Xi_{i,3} & \Xi_{i,2} \end{bmatrix}, \tag{17}$$

where

$$\Xi_{i,1} \equiv \left(\Upsilon_{i,1,1}\sigma^2+2\Upsilon_{i,1}\beta_i\phi+\Upsilon_{i,2}\beta_i^2\tau^2\right),$$

$$\Xi_{i,2} \equiv \left(\Upsilon_{i,2}\gamma^2\sigma^2+2\Upsilon_{i,3}\gamma\phi+\Upsilon_{i,3,3}\tau^2\right),$$

$$\Xi_{i,3} \equiv \left(\Upsilon_{i,1}\gamma\sigma^2+\Upsilon_{i,2,2}\phi+\Upsilon_{i,2}\beta_i\gamma\phi+\Upsilon_{i,3}\beta_i\tau^2\right).$$

In fact, its vectorised form is given by

$$\begin{aligned} vec(\Lambda_i) &= (I - A_i' \otimes A_i')^{-1} vec(\Sigma_i) \\ &= \frac{1}{\Upsilon_{i,0}} \left[\begin{array}{cccc} \Xi_{i,1} & \Xi_{i,3} & \Xi_{i,3} & \Xi_{i,2} \end{array} \right]' \\ &= \frac{1}{\Upsilon_{i,0}} \left[\begin{array}{c} \left(\Upsilon_{i,1,1}\sigma^2 + 2\Upsilon_{i,1}\beta_i\phi + \Upsilon_{i,2}\beta_i^2\tau^2\right) \\ \left(\Upsilon_{i,1}\gamma\sigma^2 + \Upsilon_{i,2,2}\phi + \Upsilon_{i,2}\beta_i\gamma\phi + \Upsilon_{i,3}\beta_i\tau^2\right) \\ \left(\Upsilon_{i,1}\gamma\sigma^2 + \Upsilon_{i,2,2}\phi + \Upsilon_{i,2}\beta_i\gamma\phi + \Upsilon_{i,3}\beta_i\tau^2\right) \\ \left(\Upsilon_{i,2}\gamma^2\sigma^2 + 2\Upsilon_{i,3}\gamma\phi + \Upsilon_{i,3,3}\tau^2\right) \end{array} \right]. \end{aligned}$$

Consider now the first equation of (3), with $\eta_i^{1\prime} = \left[\begin{array}{cc} 0 & \xi_i \end{array} \right]$ denoting the first column of η_i'. Substituting for Λ_i in (15) we find that, even in this simple case,

$$\operatorname*{plim}_{N\to\infty, T\to\infty} \left(\begin{array}{c} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\beta}_{FE} - \beta \end{array} \right) = \tag{18}$$

$$\left(E\left\{ \frac{1}{\Upsilon_{i,0}} \left[\begin{array}{c} \Xi_{i,4}\xi_i \\ \Xi_{i,5}\xi_i \end{array} \right] \right\} \right) \left(E\left\{ \frac{1}{\Upsilon_{i,0}} \left[\begin{array}{cc} \Xi_{i,1} & \Xi_{i,3} \\ \Xi_{i,3} & \Xi_{i,2} \end{array} \right] \right\} \right)^{-1},$$

where

$$\Xi_{i,4} = \left(\Upsilon_{i,1}\gamma\sigma^2 + \Upsilon_{i,2,2}\phi + \Upsilon_{i,2}\beta_i\gamma\phi + \Upsilon_{i,3}\beta_i\tau^2\right),$$

$$\Xi_{i,5} = \left(\Upsilon_{i,2}\gamma^2\sigma^2 + 2\Upsilon_{i,3}\gamma\phi + \Upsilon_{i,3,3}\tau^2\right).$$

This equation cannot be simplified further without additional assumptions because it involves non-linear functions of the random variable β_i. In the case of a general $PVAR$ specification, therefore, it is not possible to predict the sign and analyse the determinants of the heterogeneity bias of the FE estimator.

An explicit solution for the heterogeneity bias of the FE estimator, however, can be obtained in two special cases of interest. First, a *close form* solution can be obtained by assuming that $\gamma = 0$—i.e., assuming that weak exogeneity of x for the estimation of λ and β fails, but Granger non-causality of z for x continues to hold. This case allow us to study the role of ϕ in (18). Second, an *approximate* solution for $\gamma \neq 0$ can be obtained by assuming that $\phi = 0$ and that $\lambda = \rho = 0$—i.e., assuming that weak exogeneity of x for the estimation of λ and β holds, but Granger non-causality of z for x fails. This second case allows us to examine the role γ in (18). The next two subsections look at each of these two special cases in turn. A third special case, in which both ϕ and γ are different from zero but $\lambda = \rho = 0$ and $\gamma = 1$—i.e., a case in which both distinguishing features of a $PVAR$ specification are present—is analyzed by means of Monte Carlo simulation in the final section of the chapter.

A First Special Case: Weak Exogenity Fails

Let's assume that $\phi \neq 0$ but $\gamma = 0$, then (16) becomes:

$$\begin{aligned} z_{i,t} &= \lambda z_{i,t-1} + \beta_i x_{i,t-1} + \alpha_i^z + \varepsilon_{i,t}^z , \\ x_{i,t} &= \rho x_{i,t-1} + \alpha_i^x + \varepsilon_{i,t}^x . \end{aligned} \tag{19}$$

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

$$\underset{N\to\infty,T\to\infty}{\text{plim}} \begin{pmatrix} \widehat{\lambda}_{FE} - \lambda \\ \widehat{\beta}_{FE} - \beta \end{pmatrix} = \begin{bmatrix} \frac{\rho(1-\lambda\rho)\left(1-\lambda^2\right)\omega}{\Psi_1+\Psi_2} \\ -\frac{\beta\rho^2\left(1-\lambda^2\right)\omega+\Psi_3}{\Psi_1-\Psi_2} \end{bmatrix}, \tag{20}$$

where:

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

In this case, the size of the asymptotic bias of the FE estimator 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.[20] Moreover, in the appendix, I show that in this case 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 the absolute value of ϕ relative to the absolute value of $(2\beta\tau/(1+\lambda))$. If $\phi \neq 0$, therefore, it is possible that the heterogeneity bias compounds instead of offsetting the small T bias of the FE estimator, thereby yielding estimation results potentially more distorted than in standard DPM specifications. Nonetheless, the bias disappears if $\rho = 0$ or $\omega = 0$, or if λ approaches one from below when $\rho \neq 0$ and $\omega \neq 0$, as in the case analysed by Pesaran and Smith. The result that

$$\underset{\rho\to 1}{\text{plim}}(\widehat{\lambda}_{FE}) = 1 \quad \underset{\rho\to 1}{\text{plim}}(\widehat{\beta}_{FE}) = 0$$

irrespective of the true values of λ and φ also continues to hold, as Ψ_2 and Ψ_3 tend to zero as ρ approaches unity.

In summary, the main difference compared to the result of Pesaran and Smith is that, when $\phi \neq 0$, it becomes more difficult to predict the sign of the heterogeneity bias of the FE. In particular, it is possible that the heterogeneity bias of both $\widehat{\lambda}_{FE}$ and $\widehat{\beta}_{FE}$ compounds instead of offsetting the small T bias if the correlation between the error terms is sufficiently enough.

---

20 Indeed, it is easy to see that, further assuming that $\phi = 0$ in (19), we obtain Pesaran and Smith's result previously reported in (7).

A Second (Very) Special Case: Granger Non-Causality Fails

Suppose $\gamma \neq 0$ but $\lambda = \rho = 0$, then the model becomes:

$$\begin{aligned} z_{i,t} &= \beta_i x_{i,t-1} + \alpha_i^z + \varepsilon_{i,t}^z , \\ x_{i,t} &= \gamma x_{i,t-1} + \alpha_i^x + \varepsilon_{i,t}^x . \end{aligned} \tag{21}$$

Assume also, without loss of generality, that $\sigma^2 = \tau^2 = 1$. Substituting these hypotheses in equation (18) and simplifying the resulting expression, it is easily seen that:

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

$$\frac{1}{\Delta'} \begin{bmatrix} E\left(\frac{(1-\gamma^2)}{(1-\beta_i^2\gamma^2)}\right) E\left(\frac{\phi\xi_i}{(1-\beta_i\gamma)}\right) - E\left(\frac{\phi}{(1-\beta_i\gamma)}\right) E\left(\frac{(1+\gamma^2)\xi_i}{(1-\beta_i^2\gamma^2)}\right) \\ E\left(\frac{(1+\beta_i^2)}{(1-\beta_i^2\gamma^2)}\right) E\left(\frac{(1-\gamma^2)\xi_i}{(1-\beta_i^2\gamma^2)}\right) - E\left(\frac{\phi}{(1-\beta_i\gamma)}\right) E\left(\frac{\phi\xi_i}{(1-\beta_i\gamma)}\right) \end{bmatrix} \tag{22}$$

where

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

If we now assume $\phi = 0$ in (22), we can see that the heterogeneity bias of $\widehat{\lambda}_{FE}$ vanishes, while that of $\widehat{\beta}_{FE}$ is given by:

$$\operatorname*{plim}_{N\to\infty,T\to\infty} \left(\widehat{\beta}_{FE} - \beta\right) = \frac{E\left(\frac{\xi_i}{(1-\beta_i^2\gamma^2)}\right)}{E\left(\frac{1}{(1-\beta_i^2\gamma^2)}\right)}. \tag{23}$$

By taking a second-order Taylor expansion around the cross-sectional mean of ξ_i (which is zero) of the two non-linear functions of ξ_i inside the brackets of the numerator and the denominator of this expression and calculating the expectations with respect to the distribution of ξ_i, the asymptotic bias of $\widehat{\beta}_{FE}$ can be approximated as follows:

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

$$\cong \frac{E\left(\frac{1}{1-\gamma^2\beta^2}+\frac{2\gamma^2\beta}{(1-\gamma^2\beta^2)^2}\xi_i+\left(\frac{8\gamma^4\beta^2}{(1-\gamma^2\beta^2)^3}+\frac{2\gamma^2}{(1-\gamma^2\beta^2)^2}\right)\xi_i^2\right)}{E\left(0+\frac{1}{1-\gamma^2\beta^2}\xi_i+\frac{4\gamma^2\beta}{(1-\gamma^2\beta^2)^2}\xi_i^2\right)}$$
$$= \frac{\frac{4\gamma^2\beta}{(1-\gamma^2\beta^2)^2}\omega}{\frac{1}{1-\gamma^2\beta^2}+\left(\frac{8\gamma^4\beta^2}{(1-\gamma^2\beta^2)^3}+\frac{2\gamma^2}{(1-\gamma^2\beta^2)^2}\right)\omega},$$
$$= \frac{(1-\gamma^2\beta^2)(4\gamma^2\beta)\omega}{(1-\gamma^2\beta^2)^2+(6\gamma^4\beta^2+2\gamma^2)\omega}.$$

According to this approximation, $\widehat{\beta}_{FE}$, always underestimates β. This bias vanishes only if $\omega = 0$, or if either γ or β are equal 0, or if $|\gamma\beta|$ approaches one from below when ω, γ, and β are different from 0. For a given value of γ its size depends on the average value of the roots of the system (equal to $\pm(\gamma\beta)^{1/2}$ in this case) and the variance of β_i (ω). Moreover, it is possible to show that, for given values of γ and β, the bias is always increasing in ω, by noting the first derivative of the expression approximating the asymptotic bias of $\widehat{\beta}_{FE}$ with respect to ω is positive if $\left(1-(1-\gamma^2\beta^2)^2\right) > 0$, which is always satisfied under stationarity. The relation between average persistence (measured by the average absolute value of the roots of the system) and the size of the bias for given variance of β_i, instead, does not seem to be monotonic.

This last point may be seen clearly in a very special case (a case analysed also by means of Monte Carlo simulations in the next section) assuming that $\gamma = 1$. If $\lambda = \rho = 0$, stationarity requires that $\left|\pm(\gamma\beta)^{1/2}\right| < 1$ and constrains the range of variation of β_i for given γ, and vice-versa. If $\gamma = 1$, then stationarity requires $|\beta_i| < 1$ for all i and average persistence increases one-to-one with β. Further, assume that ξ_i is uniformly distributed over the interval $[\pm\omega(1-\beta)]$ with $0 \leq \omega \leq 1$ (where ω now denotes a scale parameter that controls the dispersion of ξ_i around a given β), and $0 < \beta < 1$ for simplicity.[21] In this very special case, the asymptotic bias of $\widehat{\beta}_{FE}$ is given by

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\left(\widehat{\beta}_{FE}-\beta\right) \cong \frac{4(1-\beta^2)\beta\left(\frac{\omega^2(1-\beta)^2}{12}\right)}{(1-\beta^2)^2+(6\beta^2+2)\left(\frac{\omega^2(1-\beta)^2}{12}\right)}. \tag{24}$$

Figures 1 and 2 plot this expression (for $0 \leq \omega \leq 1$ and $0 < \beta < 1$) in absolute value and in percent of the true value of β, respectively. As we can see from these plots, the absolute value of the bias increases with β initially, peacks around $\beta = 0.5$, and then decreases toward zero as β approaches one. In percent of the true value of β, instead, the bias is monotonically decreasing in β for any given value of ω. The intuition is simple: slope heterogeneity induces correlations between the error term and the regressors, and autocorrelation in the error term to the extent to

---

21 If ξ_i is distributed uniformly over the interval $[\pm\omega(1-\beta)]$, then β_i is also uniformly distributed with mean β and variance $\frac{\omega^2(1-\beta)^2}{12}$. Increasing ω for given β therefore implies increasing the variance of β_i.

which the regressors are autocorrelated. Higher persistence, induces stronger autocorrelation in the error term, and hence a larger bias. However, since we have assumed that all VAR systems are stationary, above a certain level of persistence, the scope for heterogeneity decreases. In the limit, when average persistence in the system is maximal, all individual units must have very similar parameter values, and hence the heterogeneity bias disappears.

In summary, the main difference compared to the result of Pesaran and Smith for standard $DPMs$ is that, in the presence of a feedback from $z_{i,t}$ to $x_{i,t}$ (i.e., when $\gamma \neq 0$), the bias of $\widehat{\beta}_{FE}$ does not vanish even if the process for $x_{i,t}$ is serially uncorrelated. This confirms what previously noted discussing the IV-type estimators, and thus that in VAR specifications there are fewer special cases in which the heterogeneity bias of pooled estimators disappears. The magnitude of this bias, however, could be small in percent of the true value of the parameters of interest if average persistence in the system is sufficiently high. This further suggests that slope heterogeneity should be a more serious source of concern for $VARs$ estimated in first differences rather than levels of the variables of interest.

The study of this second special case concludes the analysis of the large sample properties of the FE estimator. In the next section, I shall study the performance of the FE, the MG and an IV alternative in a model in which both distinguishing features of a VAR specification are present. However, before proceeding, it is opportune to summarize the conclusions of the asymptotic analysis.

When $\lambda = \rho = 0$, the heterogeneity bias of the FE estimator disappears in a standard DPM specification. In a VAR specification, instead, it does not. Under stationarity, the expectations in equation (22) are well defined and generally different from zero, and the bias depends on both $\beta_i \gamma$ and ϕ, unless $\xi_i = 0$ for all i. Thus, predicting the magnitude and the sign of this bias theoretically is difficult in a reasonably general case. However, we have shown that, first, the heterogeneity bias of the FE estimator could change sign for a given average value of the parameters of interest if the correlation between the error terms is sufficiently strong, possibly compounding rather than offsetting its small T bias. Second, the magnitude of heterogeneity bias of the FE estimator may be small relative to the true value of the parameters of interest if persistence in the system is relatively high. In a VAR specification, therefore, the heterogeneity bias of pooled estimators could be more or less severe than in standard DPM specifications.

IV. Monte Carlo Analysis

This section looks at Monte Carlo simulation evidence in a specification in which both weak exogeneity and Granger non-causality fail (i.e., both ϕ and γ are different from zero) while $\lambda = \rho = 0$. The model studied is (21) and the implicit form of the heterogeneity bias of the FE estimator is given by (22). This simple case is interesting because it helps us analyse the interaction of the two distinguishing features of a VAR specification discussed above, the contemporaneous correlation between the variables of interest and their lagged interdependence, while maintaining full control over the Monte Carlo experiment. Richer VAR specifications (e.g., with $\lambda \neq 0$ and/or $\rho \neq 0$, or with multiple sources of heterogeneity) would be more realistic, but the generalizability of the Monte Carlo results (in the sense of Hendry, 1984) would diminish because it would be practically unworkable to control for all the features of the model potentially affecting the outcomes of the experiment. The model is also interesting to analyse because the short run effects of $x_{i,t}$ on $z_{i,t}$ and $z_{i,t}$ on $x_{i,t}$ coincide with their long run effects under the assumptions made.

In the rest of this section, I will compare the performance of alternative estimators under different assumptions on the size of the panel, and degree of heterogeneity and average persistence across section. The next subsection describes the set up of the experiment. The following one reports and discusses the simulation results.

A. Experiment Design

Following Pesaran, Smith, and Im (1996), and consistently with the analysis in the previous section, I consider only one source of slope heterogeneity (i.e., $\beta_i = \beta + \xi_i$ with ξ_i uniformly distributed over the interval $[\pm\omega(1-\beta)]$ with $0 \leq \omega \leq 1$ and $0 < \beta < 1$). Unlike Persaran et a. (1996), I use the uniform rather then the normal distribution to characterise the cross-sectional distribution of ξ_i because this allows me to control for the degree of slope heterogeneity introduced in the model through a single scale parameter (ω), while guaranteeing that no individual unit violates the stationarity assumption as long as $|\gamma\beta| < 1$.[22]

Somewhat arbitrarily, I maintain $\gamma = 1$ throughout the experiment and let β_i vary in the open interval (± 1). If $\gamma = 1$, the absolute value of the true cross-sectional mean of β_i ($|\beta|$) controls the average degree of persistence in the model. This is minimal for $|\beta| = 0$ and maximal as $|\beta|$ approaches one. As the variance of β_i is $\frac{\omega^2(1-\beta)^2}{12}$, for given persistence, ω controls the dispersion of the cross-sectional distribution of β_i around β (i.e., the degree of slope heterogeneity introduced in the model), which is minimal for $\omega = 0$ and maximal for $\omega = 1$, always ensuring that both individual eigenvalues are less than one in absolute value.[23]

I consider the specific values $\beta = \{0.2; 0.8\}$ and $\omega = \{0; 0.2; 0.8\}$, which represent six points in the parameter space plotted in Figures 1 and 2 and characterised in the table below for $\omega \neq 0$. Choosing $\beta = \{0.2; 0.8\}$ implies average characteristic roots equal to ± 0.45 and ± 0.89 respectively: a relatively low and relatively high degree of average persistence. Choosing

---

[22] Hsiao, Pesaran, and Tahmiscioglu (1997) use the truncated normal distribution rather than the uniform in their Monte Carlo experiment to avoid explosive (or unstable) simulated series. There are two reasons why I prefer to use the uniform distribution. First, under this assumption, I can derive the *exact* asymptotic value of the heterogeneity bias of the FE estimator in the special case section in which $\phi = \lambda = \rho = 0$ and $\gamma = 1$ by integrating analytically the numerator and the denominator of (23). Second, assuming that slope heterogeneity is uniformly distributed within some theoretically determined bounds does not seem a bad assumption in practice: it is not immediately evident that one could have strong a-priori reasons to assume a hump-shaped distribution across sections for the short run parameters of interest.

[23] As already noted, if $\lambda = \rho = 0$, the eigenvalues of individual VAR systems are given by $\pm\sqrt{\gamma\beta_i}$. Stationarity requires that $|\pm\sqrt{\gamma\beta_i}| < 1$ and constrains the range of variation of β_i for given γ, and vice-versa. Therefore, if ξ_i is distributed uniformly over the interval $[\pm\omega(1-\beta)]$, stationarity is assured for all i for given β.

$\omega = \{0; 0.2; 0.8\}$, means considering the homogeneity case, a case of low heterogeneity, and a case of high heterogeneity, relatively to a given level of average persistence. In fact, we can see from the table below that, under the assumption made, the range of β_i and ξ_i and the variance of β_i, for given absolute of the average roots in the system, increases monotonically.

Characterising four points of the parameter space

	$\beta = \omega = 0.2$	$\beta = 0.8\, \omega = 0.2$	$\beta = 0.2\, \omega = 0.8$	$\beta = \omega = 0.8$
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}

I assume an homogeneous variance-covariance matrix of the error terms and set $\sigma^2 = \tau^2 = 1$. The choice of a homoskedastic specification is dictated by the desire to assess the influence of ϕ on the finite sample properties of the estimators considered in insulation from the possible role of its heteroskedasticity. By setting $\sigma^2 = \tau^2 = 1$, ϕ does not only determine the covariance between x and z, but also their correlation which is bounded to lie between -1 and 1. I consider $\phi = \{0; \pm 0.9\}$, the case of uncorrelated error terms and the cases of, either positively or negatively, highly correlated error terms to highlight the potential effects of this feature of the model on the finite sample properties of the estimators.

I examine typical dimensions of a macro panel dataset and, in addition, one case to control for situations in which there are very few individual units, likely to arise in working with subgroups of individuals, as for instance in the next chapter of the thesis:

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

Finally, the vector of error terms is generated from a bivariate normal distribution with variance-covariance matrix Σ and the initial conditions equal zero, while a standard assumption is made to generate the individual effects α_i^z and α_i^x. Thus:

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

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

$$\begin{matrix} \alpha_i^z \sim NIID(1, 1) \\ \alpha_i^x \sim NIID(1, 1) \end{matrix}.$$

Each experimental run is based on 1000 replications and different runs start from the same seed so that the results can be easily replicated, and different experimental runs are based on the same set of randomly generated numbers. For each replication, $50 + T$ observations are generated with the final T observations used to compute the estimates.[24]

B. 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 values of ϕ).

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, $(N, T) = (20, 20)$).

In these tables, 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 the tables and is relatively high in the lower part ($\beta = 0.8$). In all runs, $\lambda = \rho = 0$ and $\gamma = 1$.

For each run of the experiment, the tables report the estimated parameters (λ and β, denoted 'Lambda' and 'Beta' respectively in the tables), their estimated standard errors (denoted 'S.e.'), the absolute value of the finite sample bias (denoted 'Bias'), which equals the estimated parameter value in the case of λ, their experimental standard deviations (denoted 'S.d.'), and, for β only, the finite sample bias as a percentage of the true value of β (denoted as 'Fbias as % of true value').

If applicable, the exact asymptotic bias of β as a percentage of its true value is also reported (denoted as "Abias as % of true value"), where the latter is computed by integrating analytically the two expectations in (23) with respect to the distribution of ξ_i under the assumptions made in the experiment and described in the previous subsection.

Homogeneous Panels

In the benchmark case of a homogeneous, large and long panel dataset 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, which, although considerably higher, are of the same order of magnitude than those of the FE and the MG estimators. The FE estimator performs well too 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. However, it clearly underperforms the FE estimator in terms of bias, underestimating the true value of β by more than 14 percent even when $T = 50$. The downward bias of the FE, instead, is only about 7 percent in this case.

Decreasing N to 20 for fixed $T = 50$ does not affect these results (see Tables 2 and 3), while decreasing T to 20 for fixed $N = 50$ has strong impact (see Table 4): in this case, the bias of the

[24] The Monte Carlo experiment is programmed in Gauss and the code is available on request.

FE estimator 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 from -0.03 to -0.06 in absolute value.

Interestingly, the introduction of a correlation between the error terms in the benchmark, homogeneous case above (i.e., $N, T = 50, 50$ and $\phi \neq 0$) affects considerably 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 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 practically insignificant.[25]

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 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 $\phi = -0.9$. Instead, 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 $\phi = -0.9$ (Results not reported).

Increasing persistence by rising β from 0.2 to 0.8 (see lower part of Table 1) reduces the bias of the FE and the MG estimators considerably without affecting their efficiency. The standard errors of the IV estimates, instead, increase dramatically with persistence. Decreasing N to 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, which yield a standard error of the estimate larger than the estimate itself in this case, and hence render this estimate of β insignificant. Instead, reducing T to 20 for fixed $N = 50$ (see lower 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 biased as much as the MG.

In summary, this first set of Monte Carlo results in the absence of slope heterogeneity bears out a well known conclusion in the dynamic panel data literature and help to qualify it in the case of a VAR specification: there is a trade-off between consistency and efficiency in estimating homogeneous models suggesting to use IV-type estimators when the panel is relatively short and FE or RE-type estimators when the panel is relatively long—say $T > 20 - 30$, as recommended by Judson and Owen (1999). However, one should not disregard the small sample bias on coefficients other than that on the lagged dependent variable as negligible when working with a VAR specification because, as we saw, their small T bias may be substantial. In addition, and more importantly, in a VAR specification, the number of time observations needed to reduce the small T bias of FE and RE-type estimators is probably larger than 20-30 as generally recommended for standard DPMs because the variance-covariance matrix of the error terms is unlikely to be diagonal in practice. In a VAR specification, the time dimension needed to neglect the small T bias of these estimators appears to depend upon the degree of persistence at system

[25] Note that these results are fairly robust to increased persistence and/or changed panel dimensions (see below).

level rather than only upon the average value of the coefficient of the lagged dependent variable as in a standard DPM. By pushing up the estimated standard errors of IV-type estimators and pushing down the bias of FE and RE-type estimators, for a given T, higher persistence may actually tilt the balance in favor of the latter.

Heterogeneous Panels

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

In the benchmark case of a large and long panel dataset with relatively low persistence and uncorrelated error terms (see upper right part of Table 1), the IV estimator does particularly badly. Its biases and standard errors are larger than those of the FE estimator and the MG estimator, respectively. Instead, the MG estimator does quite well in this case, with biases less than half those of the FE estimator in absolute value and standard errors considerably higher than those of the FE estimator only for β.[26] The FE estimator lies between the MG and the IV estimator, with a bias (of approximately 30 percent of the true value for β) comparable to that of the IV estimator and the lowest standard errors.

Two more facts are worth noting from the results in this benchmark, high heterogeneity case. First, the presence of slope heterogeneity appears to *exacerbate* the (negative) small T bias of pooled estimates of λ: the FE and IV estimates of λ equal -0.05 and -0.08, respectively, compared to a true values of zero and estimates under homogeneity equal to -0.025 and 0.0001, respectively (see upper left part of Table 1). Second, the heterogeneity bias of the FE estimator of β appears to approach its asymptotic value rather quickly. The overall finite sample bias of the FE estimator of β, in fact, turn out about 60 percent of that predicted by asymptotic theory (i.e., 48 percent of the true β in this case—see "Abias as % of true value" in the upper right part of Table 1). This despite a small T bias of opposite sign partially *offsetting* it. At the same time, decreasing T for fixed N, the heterogeneity bias of the FE estimator of β drops to less than 20 percent of its theoretical value (see upper right part of Table 4), while decreasing N for fixed T (see upper right part of Table 2 and 3) leaves it above 50 percent of its asymptotic value.

The results found by introducing correlation between the error terms are similar (see middle right part of Table 1). In this case too, the MG estimator performs better than the FE estimator, which in turn improves upon the IV estimator. We have no theoretical benchmark for the asymptotic value of the heterogeneity bias of the FE estimator when both $\phi \neq 0$ and $\gamma \neq 0$. Nonetheless, it appears that introducing correlation between the error terms *compounds* rather then offsetting the small T bias of *both* λ and β, when this correlation is large. This result suggests that the

[26] Note however that, even if the number of individual units is relatively small ($N = 20$ for $T = 50$), the MG estimates are still precise enough to distinguish between the significance of β and the insignificance of λ (see upper right part of Table 2). This ceases to hold for a very small N, say $N = 10$ (see Table 3 and the text below).

heterogeneity bias of the FE estimator of β has the same (negative) sign as the small T bias when the error terms are strongly correlated, regardless of the sign of this correlation, thereby giving rise to potentially more distorted estimates that in a standard DPM.

A smaller $N = 20$ for fixed $T = 50$ does little difference to the performance of the MG estimator. But a very small N (say equal 10) does affect the efficiency of the estimates obtained considerably (cfr. Table 2 and 3).

None of the estimators considered give satisfactory results if the panel is heterogeneous and relatively short. A shor $T = 20$ for fixed $N = 50$ causes much more serious problems, especially for the estimation of β (see Table 4). The small T bias of the MG estimator increases sharply to about 30 percent of the true value of β when $\phi = 0$, and exceeds 60 percent when $\phi = -0.90$. On the other hand, the small T bias of the FE estimator is large enough to offset the heterogeneity bias almost completely when $\phi = 0$, yielding an overall finite sample bias that is less than 10 percent of the true β in this case. But, as already noted, its performance deteriorates sharply once correlation between the error terms is introduced (with a bias equal to almost 70 percent of the true value of β if $\phi = -0.9$). This is because of the strong compounding effect of the heterogeneity bias of β. If the time dimension of the panels is reduced from $T = 50$ to $T = 20$ for fixed $N = 50$, with or without correlated error terms, the performance of the IV estimator does not deteriorate further (as compared to the benchmark heterogenous case in which $T = 50$ and $N = 50$), but it does not improve either; the IV estimator is still of no help in this case.

All three estimation procedures show lower finite sample biases when persistence is higher (see bottom right part of Table 1), and hence better performance. 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 in which persistence is low; thus, yielding a misleading estimate of β. Interestingly, in this case, the FE estimator behaves better than the MG estimator even in terms of bias: the bias of the FE estimator in percent of the true value of β is about 4 percent when $\phi = 0$ (compared to a theoretical value of 6.2 percent), about 2 percent when $\phi = +0.9$, and about 7 percent when $\phi = -0.9$, while the biases of the MG estimator are -6.2, -8.0, and -9.5 percent, respectively.

The asymptotic analysis in the previous section suggests two reasons for this result. First, as shown by Figures 1 and 2, higher persistence reduces the scope for heterogeneity under the ("homogeneity") assumptions that all individual VAR systems are stationary; the asymptotic value of the heterogeneity bias of pooled estimators should be relatively smaller in these cases. For instance, when $\phi = 0$ and $\gamma = 1$ and $\beta = \omega = 0.8$, the variance of β_i is one order of magnitude smaller than that implied by $\phi = 0$ and $\gamma = 1$ and $\beta = 0.2$ and $\omega = 0.8$ (see summary table in the text above), and the asymptotic bias of the FE estimator decreases from 48 percent to just over 6 percent of the true value of β in this case (compare "Abias as % of true value" in the upper and lower right part of every table). Second, the heterogeneity bias of the FE estimator seems to be become positive when β increases from 0.2 to 0.8, and hence offsets rather than compounds the effect of the FE's small T bias; a possibility which we had identified clearly in the case (not considered in the Monte Carlo experiments) in which $\phi \neq 0$, but $\gamma = 0$.

With relatively high persistence, as in the case of low persistence, a smaller $N = 20$ for fixed $T = 50$ affects negatively the efficiency of the FE and the MG estimates, but leaves their biases almost unchanged (see bottom right of Table 2). Instead, decreasing T to 20 for fixed $N = 50$, increases their small T biases enough to offset completely the 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, in this case, the FE estimator does remarkably better than the MG notwithstanding a 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 break down completely when both panel dimensions are relatively small (see Table 5).

In summary, in a model in which both distinguishing features of a VAR specifications are present, IV-type estimators can yield very misleading results if the panel is heterogeneous: they are not only inefficient, but also badly biased. The performance of FE and RE-type estimators depends on the time dimension of the panel, the degree of average persistence, the degree of slope heterogeneity, and also the strength of the correlation of the error terms in a VAR specification. The presence of strongly correlated error terms, in particular, may induce shifts in the sign of the heterogeneity bias of these estimators regardless of the degree of persistence. It is thus difficult to formulate recommendations that have general validity.

Nonetheless, we learned that FE and RE-type estimators may produce better estimates than the MG in some points of the parameter space, even under relatively high heterogeneity, and particularly so in the presence of high persistence and contemporaneous correlation among the error terms.[27] This is because the small T bias and the heterogeneity bias of these estimators have opposite sign in some points of the parameter space. By the same token, in those points of the parameter space in which the heterogeneity bias has the same sign as the small T bias, the FE estimator may perform particularly badly. The MG turns out 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 relatively short—say as short as $T = 20$, which would be regarded rather long in the traditional literature on $DPMs$—there is no obvious solution to the problem posed by slope heterogeneity.

In this latter case, a Bayesian estimation approach, as pursued by Hsiao, Pesaran, and Tahmiscioglu (1997) for $DPMs$ and Canova and Ciccarelli (2000) for $PVARs$, seems a viable solution, as long as the cross-sectional dimension of the panel is moderate. In fact, computational costs are likely to limit the applicability of the estimation procedure proposed by Canova and Ciccarelli (2000) to very large cross sections of multivariate time series.[28] Alternatively, one could

---

[27] Persistance and contemporaneous correlation might explain why Attanasio et. al (1999) do not encounter significant differences between FE and MG estimates when applied to a VAR for saving, investment, and growth in a large sample of industrial and developing countries despite the evidence of relatively high heterogeneity, at least among developing countries, provided by Boyd and Smith (2000).

[28] One additional advantage of a Bayesian estimation approach to panel VARs is that the problem posed by non-stationarity can be solved in much more simple and direct way in this framework by designing appropriate priors. (See

try to correct the MG estimator for its small T bias by using expansions similar to those derived by Pesaran and Zhao (1997) for standard $DPMs$, or to develop a correction for the FE estimator based on approximations of its heterogeneity bias similar to one developed in the previous section of this paper in the special case in which $\gamma \neq 1$ but $\phi = 0$ and advocated by Judson and Owen (1999) for standard $DPMs$.

V. Conclusions

Applied researchers sometimes estimate $VARs$ with panel data relying on known asymptotic and finite sample results for $DPMs$. In this paper, I have shown that estimating a VAR with a macro panel dataset may be more complicated than that: the choice of the right technique depends on the time dimension of the dataset, the dispersion of the cross-sectional distribution of the slope parameters, the average degree of persistence in the system, and the variance-covariance matrix of the error terms, including particularly the strength of the contemporaneous correlations that are usually different from zero in most applications.

The asymptotic analysis suggests that (i), in a model in which strong exogeneity fails because of contemporaneous correlation between the error terms, the covariance term may add or subtract to the magnitude of the heterogeneity bias of pooled estimators, depending on its own sign and magnitude, and may induce changes of sign in the bias as compared with the case in which the error terms are uncorrelated; (ii) in a model in which strong exogeneity fails because of a lagged feedback from the endogenous variable to the weakly exogenous variable, the heterogeneity bias in relation to the true value of the parameters of interest is always positive, increasing in the degree of heterogeneity for given persistence, and decreasing in the level of persistence for given heterogeneity in the system.

These results suggest that it is more difficult to predict the sign and the order of magnitude of the heterogeneity bias of pooled estimators in a general VAR specification than in a standard DPM, and warrant particular caution when the VAR is estimated in first differences (and persistence is usually lower), or when the estimated elements of the correlation matrix of the reduced-form residuals are relatively large (as often happens when estimating VAR in levels).

The Monte Carlo experiment indicates that (i) the finite sample value of the heterogeneity bias of pooled estimators converges rather quickly to its asymptotic value, at least in the very special case for which we have both asymptotic and small sample results; (ii) the dispersion of the slope parameters around their mean must be high in absolute terms for the heterogeneity bias of pooled estimators to be substantial; (iii) the FE estimator may perform worse than in standard DPM specifications in those points of the parameter space in which the heterogeneity bias has the same sing as the small T bias, but could perform better than the MG estimator in others, and particularly so when persistence is relatively higher; (iv) on the other hand, the time dimension of the panel must be longer than generally thought for the small T bias of the MG estimator to be negligible when the covariance of the error terms is different from zero. The Monte Carlo

a special issue of the Journal of Applied Econometrics (1991, Vol. 6.) for more details on this issue.

experiment has shown also that (v) a few individual units are sufficient to obtain relatively efficient MG estimates, and (vi) IV-type estimators are particularly vulnerable to slope heterogeneity and/or high persistence, 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 dataset 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, remains an empirical question given that the actual size of the overall biases will depend on the nonlinear interaction of a large number of parameters. More generally, these difficulties suggest that other approaches, such as those proposed by Pesaran et al. (1997) and Canova and Ciccarelli (2001), could be more successfully applied to the estimation of VARs with heterogeneous macro panel data.

I. The Heterogeneity Bias of the Fixed Effects Estimator

A. The General Case

From (14) and (13) in the text, we know that

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

as $\widetilde{y}=\widetilde{X}a+\widetilde{\nu}$. Hence, as Q is symmetric and idempotent,

$$\begin{aligned}\left(\widehat{a}_{FE}-a\right) &= \left(\widetilde{X}'\widetilde{X}\right)^{-1}\left(\widetilde{X}'\widetilde{\nu}\right)\\ & \quad \left(X'QX\right)^{-1}\left(X'Q\nu\right).\end{aligned}$$

In order to derive $\operatorname{plim}_{N\to\infty,T\to\infty}\left(\widehat{a}_{FE}-a\right)$ we need to take a few intermediate steps.

First, note that

$$Y'_{-1}\left(I_N\otimes H_T\right)Y_{-1}=\sum_{i=1}^{N}Y'_{i,-1}H_TY_{i,-1}.$$

In fact, suppose $N=2$,

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

Then, because of the definition of X, Q, and ν, and the properties of the Kronecker product, we have

$$\begin{aligned}X'QX &= \left(I_M\otimes Y'_{-1}\right)\left(I_M\otimes Q_D\right)\left(I_M\otimes Y_{-1}\right)\\ &= \left(I_M\otimes Y'_{-1}Q_DY_{-1}\right)\\ &= \left(I_M\otimes Y'_{-1}\left(I_N\otimes H_T\right)Y_{-1}\right)\\ &= I_M\otimes\sum_{i=1}^{N}Y'_{i,-1}H_TY_{i,-1},\\ &= \begin{bmatrix}\sum_{i=1}^{N}Y'_{i,-1}H_TY_{i,-1} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \sum_{i=1}^{N}Y'_{i,-1}H_TY_{i,-1}\end{bmatrix}\end{aligned}$$

and

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

where $Y_{i,-1}$is the $i^{th}(TxM)$ element of $Y_{-1,}$ $\nu^j = \begin{bmatrix} \nu_{1,1}^j & \cdots & \nu_{1,T}^j & \cdots & \nu_{N,T}^j \end{bmatrix}'$ is the j^{th} $(NTx1)$ element of ν, and $\nu_i^j = \begin{bmatrix} \nu_{i,1}^j & \cdots & \nu_{i,T}^j \end{bmatrix}'$ is the i^{th} $(Tx1)$ element of ν^j for $j = 1, \cdots M$. Note also that $\nu_i^j = \varepsilon_i^j + Y_{i,-1}\eta_i^j$, where η_i^j is the j^{th} $(Mx1)$ element of $vec(\eta_i)$, with $\varepsilon^j = \begin{bmatrix} \varepsilon_{1,1}^j & \cdots & \varepsilon_{1,T}^j & \cdots & \varepsilon_{N,T}^j \end{bmatrix}'$ being the j^{th} $(NTx1)$ element of ε and $\varepsilon_i^j = \begin{bmatrix} \varepsilon_{i,1}^j & \cdots & \varepsilon_{i,T}^j \end{bmatrix}'$ being the i^{th} $(Tx1)$ element of ε^j for $j = 1, \cdots M$. Therefore,

$$(\widehat{a}_{FE} - a) = \begin{bmatrix} \left(\sum_{i=1}^N Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(\sum_{i=1}^N Y_{i,-1}'H_T\nu_i^1\right) \\ \vdots \\ \left(\sum_{i=1}^N Y_{i,-1}'H_TY_{i,-1}\right)^{-1}\left(\sum_{i=1}^N Y_{i,-1}'H_T\nu_i^M\right) \end{bmatrix}. \tag{25}$$

Second, since (8) is covariance-stationary, assuming further that the process started a long time ago (i.e., $\lim_{s\to\infty} Y_{i,-s}A_i^s = 0$), we have:

$$Y_i = (\alpha_i \otimes i_T)(I - A_i)^{-1} + \sum_{s=0}^{\infty} \overline{\varepsilon}_{i,-s}A_i^s, \tag{26}$$

and thus

$$Y_{i,-1} = (\alpha_i \otimes i_T)(I - A_i)^{-1} + \sum_{s=0}^{\infty} \overline{\varepsilon}_{i,-s-1}A_i^s, \tag{27}$$

$$Y_{i,-1}' = (I - A_i)^{-1\prime}(\alpha_i \otimes i_T)' + \sum_{s=0}^{\infty} A_i'^s\overline{\varepsilon}_{i,-s-1}', \tag{28}$$

where $\overline{\varepsilon}_{i,-s}$ ($\overline{\varepsilon}'_{i,-s}$) are the (TxM) martices of observations on the s^{th}-order lags of $\overline{\varepsilon}_i$. In fact,

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

Third, following Appendix C of Pesaran and Smith (1995), it is possible to show that:

$$\operatorname*{plim}_{T\to\infty}\left(\frac{\overline{\varepsilon}'_{i,-s}H_T\overline{\varepsilon}_{i,-\tau}}{T}\right) = \begin{matrix} \Sigma_i & for & s = \tau \\ 0 & for & s \neq \tau \end{matrix} . \tag{29}$$

Finally, note that

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\left(\frac{\sum_{i=1}^{N} Y'_{i,-1}H_TY_{i,-1}}{NT}\right) = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\operatorname*{plim}_{T\to\infty}\left(\frac{Y'_{i,-1}H_TY_{i,-1}}{T}\right) \tag{30}$$

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\left(\frac{\sum_{i=1}^{N} Y'_{i,-1}H_T\nu_i^{,j}}{NT}\right) = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\operatorname*{plim}_{T\to\infty}\left(\frac{Y'_{i,-1}H_T\nu_i^{,j}}{T}\right) \tag{31}$$

for $j = 1, \cdots M$.

Consider first $\operatorname{plim}_{T\to\infty}\left(Y'_{i,-1}H_TY_{i,-1}/T\right)$ in (30). Substituting (27) and (28) for $Y_{i,-1}$ and $Y'_{i,-1}$ we have:

$$\operatorname*{plim}_{T\to\infty}\left(\frac{Y'_{i,-1}H_TY_{i,-1}}{T}\right) \tag{32}$$

$$\begin{aligned}
&= \operatorname*{plim}_{T\to\infty}\left(\frac{\left(I - A_i\right)^{-1\prime}\left(\alpha_i \otimes i_T\right)' H_T\left(\alpha_i \otimes i_T\right)\left(I - A_i\right)^{-1}}{T}\right) \\
&\quad + \operatorname*{plim}_{T\to\infty}\left(\frac{\left(\sum_{s=0}^{\infty} A_i^{s\prime}\overline{\varepsilon}'_{i,-s-1}\right)H_T\left(\sum_{\tau=0}^{\infty}\overline{\varepsilon}_{i,-\tau-1}A_i^{\tau}\right)}{T}\right) \\
&= \sum_{s=0}^{\infty} A_i^{s\prime}\Sigma_iA_i^s = \Lambda_i = E(Y'_{i,-1}Y_{i,-1})
\end{aligned}$$

where

$$vec(\Lambda_i) = \left(I - A'_i \otimes A'_i\right)^{-1} vec(\Sigma_i),$$

with E denoting the unconditional expectation with respect to the distribution of $\overline{\varepsilon}_i$. In fact,

$(\alpha_i \otimes i_T)' H_T = 0$, $\text{plim}_{T\to\infty}\left(\frac{\bar{\varepsilon}'_{i,-s} H_T \bar{\varepsilon}_{i,-\tau}}{T}\right) = 0$ for $s \neq \tau$ because of (29), and

$$\begin{aligned} vec\left(\sum_{s=0}^{\infty} A_i^{s\prime}\Sigma_i A_i^s\right) &= \sum_{s=0}^{\infty} vec\left(A_i^{s\prime}\Sigma_i A_i^s\right) \\ &= \sum_{s=0}^{\infty}\left(A_i^{s\prime}\otimes A_i^s\right) vec\left(\Sigma_i\right) \\ &= \sum_{s=0}^{\infty}\left(A_i'\otimes A_i\right)^s vec\left(\Sigma_i\right) \\ &= \left(I - A_i'\otimes A_i'\right)^{-1} vec\left(\Sigma_i\right), \end{aligned}$$

where $vec(A+B) = vec(A)+vec(B)$, $vec\,(A\Sigma B) = (A\otimes B)\, vec\Sigma$, and $(A^s \otimes B^s) = (A\otimes B)^s$ for any suitable matrix A and B, because of the properties of the vec operator and the Kronecker product, with the last equality of this expression deriving from a standard multivariate generalisation of the convergence of an infinite geometric series with argument less than one in absolute value, under stationarity.[28]

Consider then $\text{plim}_{T\to\infty}(Y'_{i,-1}H_T\nu_i^j/T)$ in (31) for $j = 1,\cdots M$. Recalling that $\nu_i^j = \varepsilon_i^j + Y_{i,-1}\eta_i^j$, where η_i^j and ε_i^j were defined above and substituting this in $(\text{plim}_{T\to\infty}(Y'_{i,-1}H_T\nu_i^j/T))$, for $j = 1,\cdots M$, we obtain:

$$\underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T\nu_{i,j}}{T}\right) \tag{33}$$

$$\begin{aligned} &= \underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T\varepsilon_{i,j}}{T}\right) + \underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_TY_{i,-1}}{T}\right)\eta_{i,j} \\ &= \left(\sum_{s=0}^{\infty} A_i^{s\prime}\Sigma_i A_i^s\right)\eta_i^j \\ &= \Lambda_i\eta_i^j, \end{aligned}$$

where $vec(\Lambda_i)$ was defined above. In fact, for each $j = 1,\cdots M$, we have:

$$\underset{T\to\infty}{\text{plim}}\left(\frac{Y'_{i,-1}H_T\varepsilon_{i,j}}{T}\right)$$

[28] See Hendry (1995, page. 112) and Hemilton (1994, page 264-266 and page 298-300) and their mathematical appendices for more details.

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

as $(\alpha_i \otimes i_T)' H_T = 0$ and $\text{plim}_{T\to\infty}\left(\frac{\overline{\varepsilon}_{i,-s-1}' H_T\varepsilon_{i,j}}{T}\right) = 0$ because equal to the j^{th} column of $\text{plim}_{T\to\infty}\left(\frac{\overline{\varepsilon}_{i,-s}' H_T\overline{\varepsilon}_{i,-\tau}}{T}\right)$ in (29).

Now, substituting (32) and (33) in (30) and (31) we have, for $j = 1, \cdots M$,

$$
\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y_{i,-1}' H_T Y_{i,-1}}{NT}\right) = \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\Lambda_i,
$$

$$
\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y_{i,-1}' H_T \nu_i^j}{NT}\right) = \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\eta_{i,j}
$$

and since A_i and Σ_i are iid across i, by the law of large numbers, we also have

$$
\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y_{i,-1}' H_T Y_{i,-1}}{NT}\right) = E\left[\Lambda_i\right]
$$

$$
\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\frac{\sum_{i=1}^{N} Y_{i,-1}' H_T \nu_i^j}{NT}\right) = E\left[\Lambda_i\eta_{i,j}\right]
$$

for $j = 1, \cdots M$, where E denotes expectation with respect the joint distribution of A_i and Σ_i.

Therefore, substituting these last two expressions in (25), we obtain equation (15) in the text,

$$
\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\widehat{a}_{FE} - a\right) = \begin{bmatrix} \left(E\left[\Lambda_i\right]\right)^{-1}\left(E\left[\Lambda_i\eta_i^1\right]\right) \\ \vdots \\ \left(E\left[\Lambda_i\right]\right)^{-1}\left(E\left[\Lambda_i\eta_i^M\right]\right) \end{bmatrix},
$$

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

B. A Special Case: Weak Exogeneity Fails

Suppose that $\gamma = 0$ in the bivariate VAR in equation (16), then (18) in the text becomes:

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

$$\left(E\left\{\frac{1}{\Upsilon_0'}\left[\begin{array}{ll}\Upsilon_{1,1}'\sigma+2\Upsilon_1'\beta_i\phi+\Upsilon_2'\beta_i^2\tau^2 & \Upsilon_{2,2}'\phi+\Upsilon_3'\beta_i\tau^2\\ \Upsilon_{2,2}'\phi+\Upsilon_3'\beta_i\tau^2 & \Upsilon_{3,3}'\tau^2\end{array}\right]\right\}\right)^{-1} \tag{34}$$
$$\times\left(E\left\{\frac{1}{\Upsilon_0'}\left[\begin{array}{c}\Upsilon_{2,2}'\phi+\Upsilon_3'\beta_i\tau^2\xi_i\\ \Upsilon_{3,3}'\tau^2\xi_i\end{array}\right]\right\}\right)$$

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 have:

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\begin{array}{c}\widehat{\lambda}_{FE}-\lambda\\ \widehat{\beta}_{FE}-\beta\end{array}\right)=$$
$$\left(E\left\{\left[\begin{array}{ll}\frac{\sigma_i^2}{(1-\lambda^2)}+\frac{\tau_i^2\beta_i^2\delta}{(1-\rho^2)}+\frac{2\lambda\beta_i\phi_i}{(1-\lambda^2)(1-\lambda\rho)} & \frac{\rho\beta_i\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)}\\ \frac{\rho\beta_i\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)} & \frac{\tau_i^2}{1-\rho^2}\end{array}\right]\right\}\right)^{-1}$$
$$\times\left(E\left\{\left[\begin{array}{l}\frac{\phi_i}{(1-\lambda\rho)}+\frac{(\rho\beta_i\tau_i^2)\xi_i}{(1-\lambda\rho)(1-\rho^2)}\\ \frac{\tau_i^2\xi_i}{1-\rho^2}\end{array}\right]\right\}\right).$$

Taking expectations with respect to the distribution of β_i and denoting ω the variance of ξ_i we also have

$$\underset{N\to\infty,T\to\infty}{\text{plim}}\left(\begin{array}{c}\widehat{\lambda}_{FE}-\lambda\\ \widehat{\beta}_{FE}-\beta\end{array}\right)=$$
$$\left[\begin{array}{ll}\frac{\sigma_i^2}{(1-\lambda^2)}+\frac{\tau_i^2\delta(\beta^2+\omega)}{(1-\rho^2)}+\frac{2\lambda\beta\phi_i}{(1-\lambda^2)(1-\lambda\rho)} & \frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)}\\ \frac{\rho\beta\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}+\frac{\phi_i}{(1-\lambda\rho)} & \frac{\tau_i^2}{1-\rho^2}\end{array}\right]^{-1}$$
$$\times\left[\begin{array}{l}\frac{\rho\omega\tau_i^2}{(1-\lambda\rho)(1-\rho^2)}\\ 0\end{array}\right].$$

Thus we get,

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

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

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

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

where

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

After some algebric simplifications we finally obtain

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\begin{pmatrix}\widehat{\lambda}_{FE}-\lambda\\ \widehat{\beta}_{FE}-\beta\end{pmatrix}=\begin{bmatrix}\frac{\rho(1-\lambda\rho)(1-\lambda^2)\omega}{\Psi_1+\Psi_2}\\ -\frac{\beta\rho^2(1-\lambda^2)\omega+\Psi_3}{\Psi_1+\Psi_2}\end{bmatrix} \tag{35}$$

with:

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

To study the sign of (35), write it as:

$$\operatorname*{plim}_{N\to\infty,T\to\infty}\begin{pmatrix}\widehat{\lambda}_{FE}-\lambda\\ \widehat{\beta}_{FE}-\beta\end{pmatrix}=\begin{bmatrix}\frac{\Psi_4}{\Psi_1+\Psi_2}\\ -\frac{\Psi_5+\Psi_3}{\Psi_1+\Psi_2}\end{bmatrix}, \tag{36}$$

where,

$$\Psi_4=\rho\,(1-\lambda\rho)\left(1-\lambda^2\right)\omega_{2,2},$$

and

$$\Psi_5 = \beta\rho^2 \left(1 - \lambda^2\right) \omega_{2,2}.$$

For $|\lambda| < 1$ and $|\rho| < 1$, Ψ_1 is always positive because sum of positive terms. Noting that Ψ_2 is an incomplete linear equation of the second order in ϕ, it is easy to show that:

$$\begin{array}{ll} \Psi_2 > 0 & \text{if } \beta > 0 \text{ and } \frac{-2\beta\tau}{1+\lambda} < \phi < 0, \text{ or if } \beta < 0 \text{ and } 0 < \phi < \frac{-2\beta\tau}{1+\lambda}, \\ \Psi_2 < 0 & \text{if } \beta > 0 \text{ and } \phi > 0 \text{ or } \phi < \frac{-2\beta\tau}{1+\lambda}, \text{ or if } \beta < 0 \text{ and } \phi < 0 \text{ or } \phi > \frac{-2\beta\tau}{1+\lambda}, \end{array} \tag{37}$$

and hence that, for $|\phi| > \left|\frac{2\beta\tau}{1+\lambda}\right|$, Ψ_2 is negative regardless of the sign of β. Note also that $\Psi_3 > 0$ for $\phi\rho > 0$, and hence:

$$\begin{array}{ll} \Psi_3 > 0 & \text{if } \phi > 0 \text{ and } \rho > 0, \text{ or } \phi < 0 \text{ and } \rho < 0, \\ \Psi_3 < 0 & \text{if } \phi < 0 \text{ and } \rho > 0, \text{ or } \phi > 0 \text{ and } \rho < 0. \end{array} \tag{38}$$

Finally, it is evident that $\Psi_4 > 0$ for $\rho > 0$, and that $\Psi_5 > 0$ for $\beta > 0$.

We can now see that, unlike the case in which $\phi = 0$, for any given value of ρ (which is the only determinant of the sign of Ψ_4 under stationarity), the sign of the asymptotic bias of $\widehat{\lambda}_{FE}$ (determined by the sign of the term $\Psi_4/(\Psi_1 + \Psi_2)$) will change for a sufficiently large absolute value of the covariance term. This is because Ψ_2 is negative for any $|\phi| > \left|\frac{2\beta\tau}{1+\lambda}\right|$ and offsets Ψ_1 (which is always positive) for a sufficiently large value of $|\phi|$, while the sign of Ψ_4 is affected only by ρ. Suppose for instance that $\rho > 0$, $\lambda = 0.6$, $\tau = 1$, and $|\beta| = 0.4$, than for any $|\phi| > 0.125$ the large sample bias of $\widehat{\lambda}_{FE}$ becomes negative, compounding rather than offsetting the small T bias of this estimator in finite samples.

Similarly, the sign of the large sample bias of $\widehat{\beta}_{FE}$, in the case in which $\phi \neq 0$, may be positive or negative, depending on the sign of ρ and ϕ and the magnitude of ϕ. To see this, consider the term $(\Psi_5 + \Psi_3) / (\Psi_1 + \Psi_2)$, which determines the sign of this bias. If $\beta > 0$ and both ϕ and ρ are either positive or negative, then $(\Psi_5 + \Psi_3)$ is always positive because both Ψ_5 and Ψ_3 are always positive (see equations 37 and 38). However, if $\beta > 0$ and $\phi > 0$, Ψ_2 is negative (see equation 37), and, for a sufficiently large value of ϕ, Ψ_2 will offset Ψ_1 (which is always positive), thereby causing $(\Psi_5 + \Psi_3) / (\Psi_1 + \Psi_2)$ to change sign. It is straightforward to see that the same result holds also in the case in which $\beta > 0$ and both ϕ and ρ are negative, for ϕ sufficiently more negative than $\frac{-2\beta\tau}{1+\lambda}$. If $\beta < 0$ and ϕ and ρ have opposite signs, $(\Psi_5 + \Psi_3)$ is always negative because both Ψ_5 and Ψ_3 are always negative (see equations 37 and 38). However, $(\Psi_1 + \Psi_2)$ will be negative only for ϕ sufficiently smaller than 0 or more positive than $\frac{-2\beta\tau}{1+\lambda}$, thereby causing $(\Psi_5 + \Psi_3) / (\Psi_1 + \Psi_2)$ to change sign at some point.

Figure 1: Heterogeneity Bias of $\widehat{\beta}_{FE}$ for $\gamma = 1$ (in absolute value)

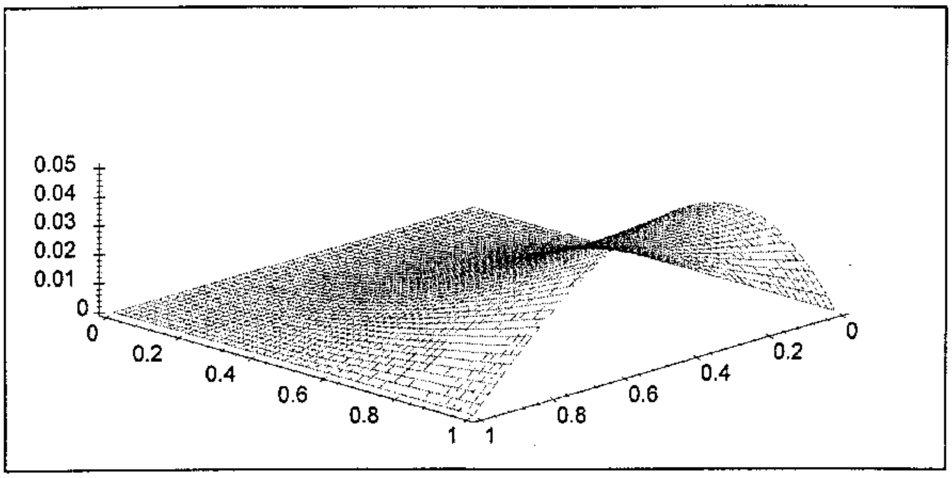

Figure 2: Heterogeneity Bias of $\widehat{\beta}_{FE}$ for $\gamma = 1$ (in percent of true value)

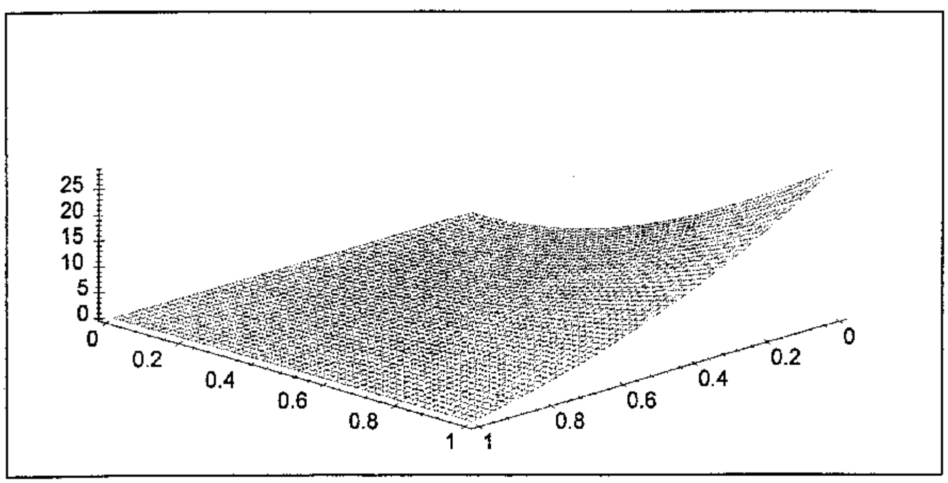

Table 1. Monte Carlo Results (N,T=50,50)

	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*	*-0.0095*	*-0.0614*	*-0.0617*	*-0.0044*	*-0.0378*	*-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*	*-0.0363*	*-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*	*-0.1503*	*-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.

References

Abadir, K. M., K. Hadri, and E. Tzavalis, 1999, "The Influence of VAR Dimensions on Estimator Biases," *Econometrica*, Vol. 67, pp. 163–81.

Alvarez, J., and M. Arellano, 1998, "The Time Series and Cross Section Asymptotics of Dynamic Panel Data Estimators" (unpublished; Mardrid: CEMFI).

Andersen, L. E., C.W.J. Granger, and E. J. Reis, 1997, "A Random Coeficient VAR Transition Model of the Changes in Land Use in The Brazilian Amazon," *Revista Brasilera de Econometria*, Vol. 17, No. 1, pp. 1–13.

Anderson, T. W., and C. Hsiao, 1981, "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, Vol. 76, pp. 598–606.

———, 1982, "Formulation and Estimation of Dynamic Using Panel Data," *Journal of Econometrics*, Vol. 18, pp. 47–82.

Attanasio, O., L. Picci, and A. Scorcu, 1998, "Saving, Growth and Investment" (unpublished; Washington: World Bank).

Banerjee, A., 1999, "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Economics and Statistics*, Vol. 61, No. 4, pp. 603–29.

———, M. Marcellino, and C. Osbat, 2000, "Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data," mimeo, (Florence: European University Institute).

Binder, M., C. Hsiao, and M. H. Pesaran, 2000, "Estimation and Inference in Short Panel Vector Autogressions with Unit Roots and Cointegration" (unpublished; College Park, Maryland: University of Maryland).

Boyd D., and R. Smith, 2000, "Some Econometric Issues in Measuring the Monetary Transmission Mechanism with an Application to Developing Countries" (unpublished; London: University of East London).

Canova, F., 1995, "Vector Autoregressive Models: Specification, Estimation, Inference, and Forecasting," in *Handbook of Applied Econometrics Macroeconomics*, ed. by H. Pesaran and M. Wickens (Oxford, England: Blackwell).

———, and M. Ciccarelli, 2000, "Forecasting and Turning Point Prediction in a Bayesian Panel VAR Model," UPF WP No AD 2000–05.

Carroll, C. D., and D. Weil, 1984, "Saving and Growth: A Reinterpretation," Carnegie-Rochester Conference Series on Public Policy, Vol. 40, pp. 133–92.

Cornwell, C., P. Schmidt, and D. Wyhowski, 1992, "Simultaneous Equations and Panel Data," *Journal of Econometrics*, Vol. 51, pp. 151–81.

Doan, T., R. Litterman, and C. Sims, 1984, "Forecasting and Conditional Projections Using Realist Prior Distributions," *Econometric Review*, Vol. 3, No. 1, pp. 1-100.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2002, "The Generalized Factor Model: Identification and Estimation," *Review of Economics and statistics*, forthcoming.

Hall, S., and G. Urga, 1997, "Stochastic Common Trends and Long-Run Relationships in Heterogeneous Panels," Centre for Economics Forecasting, London Business School, London, Discussion Paper, No. DP 27–95.

Hendry, D., 1984, "Monte Carlo Experimentation in Econometrics," in *Handbook of Econometrics*, Vol. II, Chapter 16, pp. 937-976, ed. by Z. Griliches and M. D. Intriligator.

Holtz-Eakin, D., W. Newey, and H. Rosen, 1988, "Estimating Vector Autiregressions with Panel Data," *Econometrica*, Vol. 56, No. 6, pp. 1371–95.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu, 1997, "Bayes Estimation of Short Run Coefficients in Dynamic Panel Data Models" (unpublished; Cambridge: Cambridge University Press).

Judson, R. A., and A. L. Owen, 1999, "Estimating Dynamic Panel Data Models: A Guide for Macro-econometricians," Economic letters, Vol. 65, No. 9, pp. 9–15.

Kiviet, J. F., 1995, "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 68, pp. 53–78.

———, "Expectations of Expansions for Estimators in a Dynamic Panel Data Model: Some Results for Weakly-Exogenous Regressors," (unpublished; Amsterdam: Amsterdam University).

Krishnakumar, J., 1996, "Simultaneous Equations," in *The Econometrics of Panel Data*, ed. by L. Mátyás and P. Sevestre (London: Kluwer Academic Publishers).

Larsson, R., and J. Lyhagen, 1999, "Likelihood-based Inference in Multivariate Panel Cointegration Models" (unpublished; Stockholm: Department of Statistics, Stockholm University).

———, and M. Löthgren, 1998, "Likelihood-based Cointegration Tests in Heterogeneous Panels," Stockholm School of Economics WP No. 250.

Pesaran, M. H., and R. Smith, 1995, "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, Vol. 68, pp. 79–113.

———, and K. Im, 1996, "Dynamic Linear Models for Heterogeneous Panels," in *The Econometrics of Panel Data*, ed. by L. Mátyás and P. Sevestre (London: Kluwer Academic Publishers).

Pesaran, M. H., and Z. Zhao, 1997, "Bias Reduction in Estimating Long-run Relationships from Dynamic Heterogeneous Panels," in *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala*, ed. by C. Hsiao, K. Lahiri, L-F Lee, and M.H. Pesaran (Cambridge: Cambridge University Press).

Phillips, P. C., and H. Moon, 1999, "Linear regression Theory for Non-stationary Panel Data," *Econometrica.*

———, 2000, "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Review*, Vol. 19, No. 3, pp. 263–86.

Rebucci, A., 1998, "External Shocks, Macroeconomic Policy, and Growth: A Panel VAR Approach," ESRC/GEI WP No. 40.

Robertson, D., and J. Symons, 1992, "Some Strange Properties of Panel Data Estimators," *Journal of Applied Econometrics*, Vol. 7, pp. 175–89.

———, 2000, "Factor Residual in Sur Regressions: Estimating Panels Allowing for Cross Sectional Correlation," (unpublished).

Simith, R. P., 2000, "Estimation and Inference with Non-stationary Panel Time-series Data," (unpublished; London: Department of Economics, Birbeck College).

Sims, C., 1980, "Macroeconomics and Reality," *Econometrica*, Vol. 48, pp. 1–48.

Swamy, P. A., 1970, "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, Vol. 38, pp. 311–23.

Theil, H., 1971, Principles of Econometrics, (New York: Wiley).