

Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix

Maurice J.G. Bun[□]
University of Amsterdam

preliminary version: 11 January, 2000

JEL-code: C13; C23; E41

Keywords: asymptotic expansions, bias correction, bootstrap, dynamic panel data model, heteroscedasticity, money demand

Abstract

Approximation formulae are developed for the bias of ordinary and generalized Least Squares Dummy Variable (LSDV) estimators in dynamic panel data models. Results from Kiviet (1995, 1999) are extended to higher-order dynamic panel data models with general covariance structure. The focus is on estimation of both short- and long-run coefficients. The results show that proper modelling of the disturbance covariance structure is indispensable. The bias approximations are used to construct bias corrected estimators which are then applied to quarterly data from 14 European Union countries. Money demand functions for M1, M2 and M3 are estimated for the EU area as a whole for the period 1991:I-1995:IV. The empirical results show that in general plausible long-run effects are obtained by the bias corrected estimators. Moreover, bias correction can be substantial underlining the importance of more refined estimation techniques. Also the efficiency gains by exploiting the heteroscedasticity and cross-correlation patterns between countries are considerable.

1. Introduction

In this study we analyse various least squares based estimation procedures for the dynamic panel data model with fixed individual effects and a nonscalar covari-

[□] Tinbergen Institute & Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (MBUN@FEE.UvA.NL). I want to thank Jan Kiviet for his constructive comments.

ance matrix. Both the ordinary and generalized Least Squares Dummy Variables (LSDV) estimators are considered. The choice of the model and estimators is based mainly on the typical empirical study at hand, i.e. estimation of money demand functions in the area of the European Union (EU). The data are a cross-section of times series for 14 EU countries and the number of cross-section units N in the dataset is relatively small compared with the time dimension T . In an earlier simulation study (Bun and Kiviet, 1999), we found that in panel data models with a scalar covariance matrix the bias of least squares based techniques is relatively small compared to instrumental variables based methods when T is larger than N . Based on a mean squared error criterion, least squares methods are to be preferred in this case.

Notwithstanding the superior performance of least squares methods, they are still biased in dynamic models and require T large for consistency. Kiviet (1995, 1999) derives an approximation formula for the bias of the ordinary LSDV estimator in the first-order stable dynamic panel data model with normal disturbances and a scalar covariance matrix. We use these and other results on bias correction in higher-order dynamic regression models (Kiviet and Phillips, 1994) to develop bias expressions for higher-order dynamic panel data models with general covariance structure. Both extensions are necessary to apply bias corrected estimators to the data used in the empirical study of money demand. It turns out that the first-order dynamic model is not general enough to capture all the dynamic features in the data. Apart from cross-sectional heteroscedasticity, also interdependencies between countries are likely to be present, i.e. individual disturbance terms may be correlated. To the extent that there is in fact cross-sectional heteroscedasticity and also dependencies between the cross-section units in the dataset, one should explicitly take them into account in any inference procedure exploiting the panel nature of the data.

In the case of the money demand relationship the long-run effects are important for policymakers. Hence, a clear distinction is made between estimation of short- and long-run parameter vectors. The direct bias correction on long-run coefficients, proposed by Pesaran and Zhao (1999) in the context of the dynamic random coefficients model, is applicable here also. In the type of model analysed in this study, the estimation of the variances of the coefficient estimators by conventional asymptotic expressions can be dramatically inaccurate (Freedman and Peters, 1984). Hence, we make use of bootstrap procedures to estimate standard errors.

The bias expressions developed here are used to construct bias corrected estimators and they are applied in the empirical study on money demand. Various authors have estimated a money demand function based on aggregated time series for the whole EU area and tested the stability of this function through time

(Kremers and Lane, 1990; Monticelli and Papi, 1996; Fase and Winder, 1998). All those studies use time series techniques, but considering the EU countries as a cross-section one can possibly use panel data techniques.

As compared with the aggregate time series approach the use of panel data techniques is different at least in two respects. First of all, it is not necessary to convert money stock and income measures for the different countries into one common currency as is the case for the aggregated time series approach. As long as a suitable conversion measure and functional form are chosen, the individual constants in the panel data model will absorb the effects of this conversion. Second, as in dynamic panel data models the individual effects are typically filtered out before estimation, the cross-section dimension in the panel implies extra data to estimate the same number of unknown parameters. Hence, it seems possible to use fewer time observations as in the aggregate time series approach. To the extent that one is primarily interested in a description of the very near past this is convenient, because especially the short-run parameters of the money demand relationship may not have been constant over the last few decades.

Section 2 gives an outline of the model. In evaluating the bias terms of the estimators, a detailed knowledge of the stochastic structure of the model is needed and this is described in this section. In section 3 bias expressions for ordinary and generalized LSDV estimators are developed. Estimators of both short- and long-run parameters are analysed. In section 4 the limiting distributions of the various estimators are given and an outline is given of the bootstrap procedure for estimating the standard errors. In section 5 the estimation techniques are applied to estimate EU wide demand functions for M1, M2 and M3. The emphasis is on the plausibility of coefficient estimates, the magnitude of bias correction terms and efficiency gains through taking into account the more general covariance structure in the estimation procedure. Section 6 concludes.

2. Model

We consider the higher-order dynamic panel data model

$$y_{it} = \sum_{p=1}^P \rho_p y_{i;t-p} + \alpha_0 x_{it} + \alpha_i + \varepsilon_{it}; \quad i = 1; \dots; N; \quad t = 1; \dots; T; \quad (2.1)$$

In this model the dependent variable y_{it} is regressed on a $K \times 1$ vector of explanatory variables x_{it} ; P lagged values of the dependent variable and an individual specific constant α_i . The explanatory variables in x_{it} are assumed to be strictly exogenous, i.e.

$$E[x_{it} \varepsilon_{js}] = 0; \quad \delta_{ij}; \quad t; s;$$

and the individual effects α_i are assumed fixed, but unknown. Note that both the univariate processes for y_{it} and the elements of x_{it} may contain unit roots. However, the relationship (2.1) between y_{it} and x_{it} is assumed to be stable. For $p = 1$ this implies $\rho_{11} < 1$, but in higher order models more complicated restrictions on the autoregressive coefficients are required for stability. The disturbances ε_{it} are uncorrelated through time, but we allow for heteroscedasticity across cross-section units and non-zero contemporaneous cross-correlations, i.e.

$$\begin{aligned} E[\varepsilon_{it}] &= 0; & \forall i; t; \\ E[\varepsilon_{it}\varepsilon_{jt}] &= \Sigma_{ij}; & \forall i; j; t; \\ E[\varepsilon_{it}\varepsilon_{js}] &= 0; & \forall i; j; t \neq s; \end{aligned}$$

Stacking the observations over time we got

$$\begin{aligned} y_i &= \sum_{p=1}^P \rho_p y_{i;1:p} + X_i \alpha_i + \alpha_i \mathbf{1}_T + \varepsilon_i & i = 1; \dots; N \quad (2.2) \\ &= W_i \alpha_i + \alpha_i \mathbf{1}_T + \varepsilon_i \end{aligned}$$

where $y_{i;1:p} = (y_{i;1:p}; \dots; y_{i;T:p})'$; $X_i = (x_{i1}; \dots; x_{iT})'$; $\mathbf{1}_T = (1; \dots; 1)'$ a $T \times 1$ vector of ones, $\alpha_i = (\alpha_0; \dots; \alpha_p)'$; $\rho = (\rho_1; \dots; \rho_p)'$ and $W_i = [y_{i;1:p}; \dots; y_{i;T:p}; X_i]$:

Like Kiviet (1999) we decompose y into a relevant random component, denoted by a tilde, and irrelevant random plus deterministic components, denoted by a bar. The relevant random component is in some way related to the disturbance term ε_{it} ; while the irrelevant component is not, i.e.

$$\begin{aligned} \bar{y}_i &= \sum_{p=1}^P \rho_p \bar{y}_{i;1:p} + \varepsilon_i & i = 1; \dots; N; \quad (2.3) \\ \tilde{y}_i &= \sum_{p=1}^P \rho_p \tilde{y}_{i;1:p} + X_i \alpha_i + \alpha_i \mathbf{1}_T; \end{aligned}$$

where we use the assumption that we have fixed individual effects and only strict exogenous explanatory variables, i.e. $X_i = 0$ and $\alpha_i = 0$. For the initial values we assume

$$\begin{aligned} \bar{y}_{i;1:p} &= 0; & p = 1; \dots; P; \quad (2.4) \\ \tilde{y}_{i;1:p} &= y_{i;1:p}; \end{aligned}$$

so we condition on p fixed starting values. Introducing a $T \times T$ matrix L_T with ones on the first subdiagonal and zeros elsewhere and defining

$$L_T = \begin{pmatrix} 0 & & & & \mathbf{1}_{i,1} \\ \mathbf{1}_{i,1} & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \rho_p L_T^p \mathbf{A} \end{pmatrix}; \quad (2.5)$$

we write for the relevant random components from (2.3)

$$y_i = \beta_T' \epsilon_i; \quad i = 1; \dots; N: \quad (2.6)$$

To analyse the estimators in the next section we need a decomposition of the matrix $A_T W_i$ with $A_T = (I_T - \beta_T \beta_T')$: We write

$$\begin{aligned} A_T W_i &= A_T [y_{i,1} \dots y_{i,P} : X_i] \\ &= \sum_{p=1}^P A_T L_T^p \epsilon_p^0; \end{aligned} \quad (2.7)$$

because $X_i = 0$ and $A_T y_{i,p} = A_T L_T^p y_i$ and where e_p is $(P + K) \times 1$ unit vector with its p th element equal to one.

Stacking the observations also across individuals one gets

$$y = W\beta + D\gamma + \epsilon; \quad (2.8)$$

where $\gamma = (\gamma_1; \dots; \gamma_N)'$ is a $N \times 1$ vector, $W = [W_1^0; \dots; W_N^0]'$ and $D = (I_N - \beta_T)$ are $NT \times (K + P)$ and $NT \times N$ matrices of stacked observations respectively whereas y and ϵ are $N \times T$ vectors. The assumptions about ϵ can be written as

$$\begin{aligned} E[\epsilon] &= 0; \\ E[\epsilon \epsilon'] &= \Sigma = S - I_T; \end{aligned} \quad (2.9)$$

with S a $N \times N$ matrix with typical element s_{ij} : For the relevant stochastic components in AW we find from (2.7)

$$\begin{aligned} AW &= \sum_{p=1}^P AL^p \epsilon_p^0 \\ &= \sum_{p=1}^P \lambda_p \epsilon_p^0; \end{aligned} \quad (2.10)$$

where $A = I_N - A_T$, $L = I_N - L_T$; $\lambda = I_N - \beta_T$ and $\lambda_p = AL^p$.

Model (2.8) with (2.9) is a generalized dynamic regression model and in the next section efficient estimation of both short- and long-run coefficients will be considered. The elements of the parameter vector β are called short-run coefficients and $\mu = (\beta_1 \dots \beta_P)$ is called the long-run coefficient vector.

3. Coefficient estimators

3.1. short-run coefficients

The ordinary least squares estimator for β in (2.8) is the familiar Least Squares Dummy Variables (LSDV) or fixed effect estimator. Using partitioned regression results it can be expressed as

$$\begin{aligned}\hat{\beta}_{\text{LSDV}} &= (W^0 M_D W)^{-1} W^0 M_D Y \\ &= (W^0 A W)^{-1} W^0 A y;\end{aligned}\quad (3.1)$$

where $M_D = I_{NT} - D(D^0 D)^{-1} D^0 = A$. Note that $A = I_N - A_T$ is the well-known within transformation which wipes out the individual effects α_i .

The estimation error is

$$\hat{\beta}_{\text{LSDV}} - \beta = (W^0 A W)^{-1} W^0 A u;\quad (3.2)$$

which depends in a non-linear way on the stochastic term u because of (2.10). Defining $Q = E[W^0 A W]$ using the same approach as in Kiviet (1995, 1999), but now for fixed N ; we expand

$$\begin{aligned}(W^0 A W)^{-1} &= Q^{-1} - Q^{-1} (W^0 A W - Q) Q^{-1} \\ &\quad + Q^{-1} (W^0 A W - Q) Q^{-1} (W^0 A W - Q) Q^{-1} \\ &\quad + o_p(T^{-2});\end{aligned}\quad (3.3)$$

and we find for the estimation error

$$(W^0 A W)^{-1} W^0 A u = 2Q^{-1} W^0 A u - Q^{-1} W^0 A W Q^{-1} W^0 A u + o_p(T^{-1});\quad (3.4)$$

Hence, for the bias of the LSDV estimator we find

$$E[\hat{\beta}_{\text{LSDV}} - \beta] = 2Q^{-1} E[W^0 A u] - Q^{-1} E[W^0 A W Q^{-1} W^0 A u] + o(T^{-1});\quad (3.5)$$

In Appendix A it is shown that the approximation for the bias in the LSDV estimator equals

$$\begin{aligned}E[\hat{\beta}_{\text{LSDV}} - \beta] &= B_{\text{LSDV}}(T^{-1}) + o(T^{-1}) \\ &= \sum_{p=1}^P \text{tr}(I_p) Q^{-1} e_p \\ &\quad - \sum_{i=1}^P Q^{-1} W^0 I_p - A W Q^{-1} e_p - \sum_{i=1}^P \text{tr}(Q^{-1} W^0 I_p - A W) Q^{-1} e_p \\ &\quad - \sum_{p=1}^P \sum_{r=1}^P \sum_{s=1}^P q_{rs} \text{tr}(-I_p I_r I_s) Q^{-1} e_p + o(T^{-1});\end{aligned}\quad (3.6)$$

with

$$\begin{aligned}\hat{W} &= E(W); \\ Q &= \hat{W}^0 A \hat{W} + \frac{1}{p} \frac{1}{r} \text{tr}(e_p e_r^0); \\ q_{rs} &= e_r^0 Q^{-1} e_s;\end{aligned}$$

Using this result, we can construct a bias corrected estimator, denoted by LSDV_c, as

$$\hat{\pm}_{\text{LSDV}_c} = \hat{\pm}_{\text{LSDV}} - \hat{B}_{\text{LSDV}}(T^{-1}); \quad (3.7)$$

using any consistent preliminary estimators for \pm and B in $\hat{B}(T^{-1})$, i.e. based on the ordinary LSDV estimator. The corrected LSDV estimator will be unbiased upto order $O(T^{-1})$, i.e.

$$E \hat{\pm}_{\text{LSDV}_c} = \pm + o(T^{-1}); \quad (3.8)$$

The ordinary LSDV estimator does not take the covariance structure of ϵ into account. Hence, we analyse also the generalized LSDV estimator of \pm ; denoted by $\hat{\pm}_{\text{GLSDV}}$, i.e.

$$\begin{aligned}\hat{\pm}_{\text{GLSDV}} &= (W^0 A^{-1} A W)^{-1} W^0 A^{-1} A y \\ &= (W^0 A^{-1} W)^{-1} W^0 A^{-1} y;\end{aligned} \quad (3.9)$$

with

$$\begin{aligned}A^{-k} A &= (I_N - A_T)(S^k - I_T)(I_N - A_T) \\ &= S^k - A_T \\ &= A^{-k} \\ &= -^k A;\end{aligned} \quad (3.10)$$

for every real k :

The estimation error is

$$\hat{\pm}_{\text{GLSDV}} - \pm = (W^0 A^{-1} A W)^{-1} W^0 A^{-1} A \epsilon; \quad (3.11)$$

Defining $A^a = A^{-1} A$ and $Q^a = E[W^0 A^a W]$ we expand

$$\begin{aligned}(W^0 A^a W)^{-1} &= Q^{a-1} - Q^{a-1} (W^0 A^a W - Q^a) Q^{a-1} \\ &\quad + Q^{a-1} (W^0 A^a W - Q^a) Q^{a-1} (W^0 A^a W - Q^a) Q^{a-1} \\ &\quad + o_p(T^{-2});\end{aligned} \quad (3.12)$$

and we find for the estimation error

$$(W^0 A^a W)^{-1} W^0 A^a \epsilon = 2Q^{a-1} W^0 A^a \epsilon - Q^{a-1} W^0 A^a W Q^{a-1} W^0 A^a \epsilon + o_p(T^{-1}); \quad (3.13)$$

Hence, for the bias of the GLSDV estimator we ...nd

$$E \hat{\beta}_{GLSDV}^h = 2Q^{\alpha} \mathbb{1}_p^{-1} E[W^0 A^{\alpha}] \mathbb{1}_p; Q^{\alpha} \mathbb{1}_p^{-1} E[W^0 A^{\alpha} W Q^{\alpha} \mathbb{1}_p^{-1} W^0 A^{\alpha}] \mathbb{1}_p + o(T^{-1}); \quad (3.14)$$

In Appendix A we derive the following approximation for the bias in the GLSDV estimator

$$\begin{aligned} E \hat{\beta}_{GLSDV}^h &= B_{GLSDV}(T^{-1}) + o(T^{-1}) \\ &= \sum_{p} \text{tr}(\mathbb{1}_p) Q^{\alpha} \mathbb{1}_p^{-1} e_p \\ &\quad + \sum_{p} Q^{\alpha} \mathbb{1}_p^{-1} W^0 \mathbb{1}_p^{-1} \mathbb{1}_p A W Q^{\alpha} \mathbb{1}_p^{-1} e_p \\ &\quad + \sum_{p} \text{tr} \left(Q^{\alpha} \mathbb{1}_p^{-1} W^0 \mathbb{1}_p^{-1} \mathbb{1}_p A W \right) Q^{\alpha} \mathbb{1}_p^{-1} e_p \\ &\quad + \sum_{p, r, s} q_{rs}^{\alpha} \text{tr}(\mathbb{1}_p^0 \mathbb{1}_r \mathbb{1}_s) Q^{\alpha} \mathbb{1}_p^{-1} e_p + o(T^{-1}); \end{aligned} \quad (3.15)$$

with

$$\begin{aligned} Q^{\alpha} &= W^0 A^{-1} A W + \sum_{p, r} \text{tr}(\mathbb{1}_p^0 \mathbb{1}_r) e_p e_r^0; \\ q_{rs}^{\alpha} &= e_r^0 Q^{\alpha} \mathbb{1}_p^{-1} e_s; \end{aligned}$$

In practice the GLSDV estimator cannot be calculated because α is unknown. We therefore analyse also the two-step feasible GLSDV estimator

$$\hat{\beta}_{FGLSDV}^h = (W^0 \hat{A}^{-1} A W)^{-1} W^0 \hat{A}^{-1} A y; \quad (3.16)$$

where any consistent preliminary estimator, i.e. the LSDV estimator, for the covariance matrix is used

$$\begin{aligned} \hat{A} &= \hat{S} - I_T; \\ \mathbb{A}_{ij} &= \frac{(y_i - W_i \hat{\beta}_{LSDV}^h)^0 A_T (y_j - W_j \hat{\beta}_{LSDV}^h)}{T}; \end{aligned} \quad (3.17)$$

It can be shown that the bias to order $O(T^{-1})$ in the FGLSDV estimator is the same as for the GLSDV. To prove this we follow the lines of Kiviet et al. (1995). The estimation error of the FGLSDV estimator is

$$\hat{\beta}_{FGLSDV}^h - \beta = (W^0 A (\hat{S}^{-1} - I_T) A W)^{-1} W^0 A (\hat{S}^{-1} - I_T) A \epsilon; \quad (3.18)$$

and in Appendix B it is shown that

$$E \hat{\beta}_{FGLSDV}^h = E \hat{\beta}_{GLSDV}^h + o(T^{-1}); \quad (3.19)$$

i.e. the bias approximation to order $O(T^{-1})$; denoted by $B_{FGLSDV}(T^{-1})$; is the same as for the GLSDV estimator. Hence, we can construct a bias corrected estimator, denoted by FGLSDVc, as

$$\hat{\pm}_{FGLSDVc} = \hat{\pm}_{FGLSDV} - B_{FGLSDV}(T^{-1}); \quad (3.20)$$

using any consistent preliminary estimators for \pm and $-$ in $B_{FGLSDV}(T^{-1})$. The corrected FGLSDV estimator will be unbiased upto order $O(T^{-1})$, i.e.

$$E \hat{\pm}_{FGLSDVc} = \pm + o(T^{-1});$$

3.2. long-run coefficients

The estimators in the previous subsection can be used to construct estimators for the long-run coefficients μ by

$$\hat{\mu} = \Delta (1_i \mathbb{1}_P^{\Delta}); \quad (3.21)$$

where Δ and $\mathbb{1}_P^{\Delta}$ are any of the estimators considered before and $\mathbb{1}_P$ is a $P \times 1$ vector of ones. If bias corrected estimators like (3.7) or (3.20) are used the resulting long-run estimator is called "naive" by Pesaran and Zhao (1999), which analyse several estimators of the long-run coefficients in the context of the dynamic random coefficient model. The "naive" or indirect way of bias correction in (3.21) does not lead to an estimator unbiased to order $O(T^{-1})$. Note that $\hat{\pm} = \pm + O_p(T^{-\frac{1}{2}})$ irrespective of the estimator used. Hence, we can write

$$\begin{aligned} \hat{\mu} &= \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} \mathbb{1}_P^{\Delta} (\mathbb{1}_P^{\Delta}) \\ &= \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} \mathbb{1}_P^{\Delta} \frac{\mathbb{1}_P^{\Delta} (\mathbb{1}_P^{\Delta})}{(1_i \mathbb{1}_P^{\Delta})} \\ &= \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} + \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} \frac{\mathbb{1}_P^{\Delta} (\mathbb{1}_P^{\Delta})}{(1_i \mathbb{1}_P^{\Delta})} - \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} \frac{(\mathbb{1}_P^{\Delta})^2}{(1_i \mathbb{1}_P^{\Delta})^2} + o_p(T^{-1}) \\ &= \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} + \frac{1}{(1_i \mathbb{1}_P^{\Delta})} (\Delta \mathbb{1}_P^{\Delta}) + \frac{1}{(1_i \mathbb{1}_P^{\Delta})^2} (\mathbb{1}_P^{\Delta} \mathbb{1}_P^{\Delta}) \\ &\quad + \frac{1}{(1_i \mathbb{1}_P^{\Delta})^2} (\Delta \mathbb{1}_P^{\Delta}) (\mathbb{1}_P^{\Delta}) - \frac{1}{(1_i \mathbb{1}_P^{\Delta})^3} (\mathbb{1}_P^{\Delta})^2 + o_p(T^{-1}); \end{aligned}$$

Therefore, we find for the expectation of $\hat{\mu}$ the following

$$\begin{aligned} E \hat{\mu} &= \frac{\Delta}{(1_i \mathbb{1}_P^{\Delta})} + \frac{1}{(1_i \mathbb{1}_P^{\Delta})} E(\Delta \mathbb{1}_P^{\Delta}) + \frac{1}{(1_i \mathbb{1}_P^{\Delta})^2} E(\mathbb{1}_P^{\Delta} \mathbb{1}_P^{\Delta}) \\ &\quad + \frac{1}{(1_i \mathbb{1}_P^{\Delta})^2} E(\Delta \mathbb{1}_P^{\Delta}) (\mathbb{1}_P^{\Delta}) - \frac{1}{(1_i \mathbb{1}_P^{\Delta})^3} E(\mathbb{1}_P^{\Delta})^2 + o(T^{-1}) \\ &= \mu + O(T^{-1}); \end{aligned} \quad (3.22)$$

because all remaining terms are in general non-zero and of order $O(T^{-1})$.

Pesaran and Zhao (1999) propose a direct way of bias correction. Using original uncorrected estimators as in (3.1) or (3.9), rearranging (3.22) we find for the bias in the long-run coefficient vector μ

$$\begin{aligned} E(\hat{\mu}_i^h - \mu_i) &= B_{\mu} + o(T^{-1}) \\ &= \frac{1}{(1 - \rho_P^0)^2} \mathbf{h}' (1 - \rho_P^0) (B_{-} + \mu \rho_P^0 B_{\circ}) + \text{Cov}(\hat{\rho}_P^0; \rho_P^0) \mu \text{Var}(\hat{\rho}_P^0) \\ &\quad + o(T^{-1}); \end{aligned} \tag{3.23}$$

where $B_{-} = E(\hat{\rho}_P^0 - \rho_P^0)$ and $B_{\circ} = E(\hat{\rho}_P^0 \rho_P^0)$: This can be used to construct corrected estimators of the long-run coefficients, which are unbiased upto order $O(T^{-1})$.

4. Asymptotic distributions and estimation of standard errors

In this section we concentrate on the asymptotic distributions of the various coefficient estimators. We will focus first on the limiting distributions of estimators for the short- and long-run coefficient vectors. Note that bias corrected estimators have the same limiting behaviour as their uncorrected counterparts. Next, we consider the estimation of standard errors of the coefficient estimators by either using asymptotic variance expressions following from the limiting distributions or applying bootstrap procedures.

4.1. Limiting distributions

Let us define

$$\begin{aligned} R_{WAW} &= \text{plim}_{T \rightarrow \infty} \frac{1}{T} W^0 A W \\ R_{WA-W} &= \text{plim}_{T \rightarrow \infty} \frac{1}{T} W^0 A - W \\ R_{WAW}^a &= \text{plim}_{T \rightarrow \infty} \frac{1}{T} W^0 A - i^{-1} W; \end{aligned} \tag{4.1}$$

From the usual asymptotic reasoning it follows that

$$\rho_{T \pm \text{LSDV}}^3 \hat{\rho}_P^0 \pm \frac{1}{T} \gg N^h(0); R_{WAW}^{-1} R_{WA-W} R_{WAW}^{-1}; \tag{4.2}$$

and

$$\rho_{T \pm \text{GLSDV}}^3 \hat{\rho}_P^0 \pm \frac{1}{T} \gg N^h(0); R_{WAW}^{a-1}; \tag{4.3}$$

The limiting distribution of the FGLSDV estimator can be derived from the limiting behaviour of the GLSDV estimator. Assuming

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \frac{1}{T} W^0 A (\hat{\beta}_i - \beta_i) W &= 0 \\ \text{plim}_{T \rightarrow \infty} \frac{1}{T} W^0 A (\hat{\beta}_i - \beta_i) &= 0; \end{aligned}$$

the FGLSDV estimator has the same limiting distribution as the GLSDV estimator, i.e.

$$\sqrt{T} (\hat{\beta}_{\text{FGLSDV}} - \beta) \xrightarrow{D} N(0; R_{WAW}^{-1}) \quad (4.4)$$

The limiting distributions for the estimators of μ readily follow. We can write $\hat{\mu} = g(\hat{\beta})$ with $\hat{\beta}$ any of the estimators analysed above. Denoting the first derivative with respect to β by the $(K + P) \times K$ matrix G we have

$$\sqrt{T} (\hat{\mu} - \mu) \xrightarrow{D} N(0; GV G')$$

where V is the asymptotic covariance matrix of $\hat{\beta}$:

4.2. Variance estimators

In principle the limiting results of the previous subsection can be used to estimate the finite sample covariance matrices of $\hat{\beta}$ and $\hat{\mu}$. The asymptotic covariance matrix for $\hat{\beta}$ can be estimated consistently with the following expressions

$$\begin{aligned} \hat{V}(\hat{\beta}_{\text{LSDV}}) &= (W^0 A W)^{-1} W^0 A \hat{W} (W^0 A W)^{-1} \\ \hat{V}(\hat{\beta}_{\text{LSDVc}}) &= (W^0 A W)^{-1} W^0 A \hat{W}_c (W^0 A W)^{-1} \\ \hat{V}(\hat{\beta}_{\text{FGLSDV}}) &= (W^0 A \hat{W})^{-1} \\ \hat{V}(\hat{\beta}_{\text{FGLSDVc}}) &= (W^0 A \hat{W}_c)^{-1} \end{aligned} \quad (4.5)$$

The elements of $\hat{W} = \hat{S} - I_T$ and $\hat{W}_c = \hat{S}_c - I_T$ can be consistently estimated using the LSDV or LSDVc residuals, i.e.

$$\begin{aligned} \hat{\mathcal{W}}_{ij} &= \frac{(y_i - W_i \hat{\beta}_{\text{LSDV}})' A_T (y_j - W_j \hat{\beta}_{\text{LSDV}})}{T} \\ \hat{\mathcal{W}}_{ij}^c &= \frac{(y_i - W_i \hat{\beta}_{\text{LSDVc}})' A_T (y_j - W_j \hat{\beta}_{\text{LSDVc}})}{T} \end{aligned} \quad (4.6)$$

The asymptotic covariance matrix for $\hat{\mu}$ can be estimated according to

$$\text{EstAsyVar}(\hat{\mu}) = G \hat{V}(\hat{\beta}) G'$$

where $\hat{V}(\hat{\pm})$ is any of the expressions in (4.5).

Freedman and Peters (1984) note that for the FGLSDV estimator the expression in (4.5) is very inaccurate in finite samples, i.e. the true standard deviations are underestimated using conventional first-order asymptotic approximations. An alternative approach is using bootstrap procedures to estimate standard errors. Following their lines of reasoning we propose the following parametric resampling scheme, i.e.

- 2 Obtain the estimators $\hat{\pm}$, $\hat{\Sigma}$ and $\hat{\mu}$ (LSDV, LSDVc, FGLSDV or FGLSDVc)
- 2 Take a random sample $E^m = (\epsilon_1^m; \dots; \epsilon_T^m)$ from $N(0; \hat{\Sigma})$ with $\epsilon_t^m = (\epsilon_{1t}^m; \dots; \epsilon_{Nt}^m)'$; $t = 1; \dots; T$; and construct $\epsilon^m = \text{vec}(E^{m0})$
- 2 Calculate $Ay^m = AW^m \hat{\pm} + A\epsilon^m$
- 2 Estimate the model and calculate bootstrap estimators $\hat{\pm}^m$ and $\hat{\mu}^m$ with the resampled data $(y^m; W^m)$

Remark that due to the presence of lagged values of y in the regressor matrix W there is an asterisk in the resampled data matrix W^m , i.e. a recursive sampling scheme is used. Because the normality assumption is used in the derivation of the bias expressions we employ this assumption here and use a parametric bootstrap procedure contrary to Freedman and Peters (1984).

It can be shown that the same limiting distribution theory holds for the bootstrapped estimators $\hat{\pm}^m$ and $\hat{\mu}^m$, so asymptotically there is no difference with the original estimators. However, the finite sample distribution of $\hat{\pm}^m$ and $\hat{\mu}^m$ can be simulated on the computer, i.e. by repeating the steps above B times, B realisations of $\hat{\pm}^m$ and $\hat{\mu}^m$ are created and

$$\frac{1}{B} \sum_{b=1}^B \hat{\pm}_b^m \quad \frac{1}{B} \sum_{b=1}^B \hat{\mu}_b^m \quad \frac{1}{B} \sum_{b=1}^B \hat{\Sigma}_b^m$$

$$\frac{1}{B} \sum_{b=1}^B \hat{\pm}_b^m \quad \frac{1}{B} \sum_{b=1}^B \hat{\mu}_b^m \quad \frac{1}{B} \sum_{b=1}^B \hat{\Sigma}_b^m$$

are the bootstrap estimators of the unknown covariance matrices of $\hat{\pm}$ and $\hat{\mu}$. The results in Freedman and Peters (1984) show that for the FGLSDV estimator this bootstrap variance estimator of $\hat{\pm}$ underestimates the true covariance matrix of $\hat{\pm}$ much less than the conventional formula in (4.5). In the next section we will use this bootstrap procedure also for the other estimators.

5. The demand for money in the European Union

In this section the performance of the various estimators is examined in an empirical application. Money demand in the European Union is analysed by panel data techniques. The dataset used is from Fase and Winder (1998) and contains time series on several variables for Belgium (BE), Denmark (DK), Germany (GE), United Kingdom (UK), Finland (FIN), France (FR), Greece (GR), Ireland (IE), Italy (IT), The Netherlands (NL), Austria (AT), Portugal (PT), Spain (SP) and Sweden (SWE). Together with Luxembourg, which is not included in the dataset, these countries currently form the European Union. The time series of the variables have quarterly frequency, are not seasonally adjusted and are collected over the period 1970-1995. The variables in the dataset are M1, M2, M3, real GNP, GNP deflator, short- and long-term interest rates.

For each of the definitions of money stock specification (2.1) is estimated using

$$x_{it} = (\ln gnp_{it}; \ln gnp_{i;t_1-1}; rs_{it}; rs_{i;t_1-1}; rl_{it}; rl_{i;t_1-1}; ir_{it}; ir_{i;t_1-1}; s_{1;t}; s_{2;t}; s_{3;t})^0 \quad (5.1)$$

and where the dependent variable y_{it} is the logarithm of real money stock, i.e. $\ln(M1=P)_{it}$; $\ln(M2=P)_{it}$ or $\ln(M3=P)_{it}$. The explanatory variables are contemporaneous and one-period lagged values of real income (gnp), short- (rs) and long-term (rl) interest rates and the inflation rate (ir): To account for seasonal patterns a set of seasonal dummy variables (s_1 ; s_2 and s_3) is included. Furthermore, lagged values of the dependent variable are incorporated to model autoregressive dynamic adjustments. Separate regressions for the individual countries, which are not reported here, suggest to include one lagged value for the M1 specification and to use two lagged values for the M2 and M3 specifications. Hence, the dimension of the parameter vector \pm is $K + 1$ for the M1 specification and $K + 2$ for M2 and M3 with $K = 11$:

In order to make valid inference with panel data techniques both parameter constancy through time and over countries must hold to some extent. To avoid parameter variability through time, we have chosen to analyse a relatively short time span, i.e. only the years after the German reunification in 1990 are considered and the sample period is 1991:I-1995:IV. As far as parameter constancy over countries is concerned, it is reasonable to assume that by taking a recent period the problem of parameter heterogeneity across countries is mitigated. We are therefore reasonably confident to impose common slope vectors, but allow for individual specific effects.

The number of countries analysed is $N = 14$. For M1 one period is lost in constructing the lagged value of the dependent variable, so for this specification the first estimation period is 1991:II and $T = 19$: The estimation period for M2 and M3 is 1991:III-1995:IV, so one extra period is lost in constructing the two-period lagged value of the dependent variable and $T = 18$:

For M1 the estimation results of the short-run coefficients are in Table 1. Table 2 gives the estimates for the long-run coefficients. In these and other tables only the bootstrapped standard errors are given, because, as argued before, standard first-order asymptotic approximations are inaccurate here. The number of bootstrap replications used is 100. Using the LSDV residuals the LMF 1 and LMF 2 test statistics for first and second order residual autocorrelation are 0.040 and 2.424 (p-values are 0.841 and 0.096). Regarding the short-run coefficient estimators the bias corrected estimators produce in general a higher autoregressive coefficient than original estimators, while the bias correction in the other coefficients seems to be small. Considering the variance estimators the decrease in variance is apparent when using the FGLSDV estimator compared with the LSDV estimator. The table with long-run coefficients reasserts these efficiency gains. Note that the difference between "naive" and "direct" bias correction of long-run coefficients is negligible. The long-term coefficients have plausible values commonly found in empirical studies on the demand for M1:

The results for M2 are in Tables 3 and 4. The LMF 1 and LMF 2 statistics are 0.099 and 0.249 (p-values are 0.754 and 0.781). As noted before, a two-period lagged value of the dependent variable is included also, so the bias corrections according to (3.6) and (3.15) have been applied for $P = 2$: The sum of the autoregressive coefficients is considerably higher than the same parameter for M1 implying more persistence in the demand for M2 as compared with M1: Again the difference in accuracy of ordinary and generalized LSDV is apparent, i.e. estimated standard deviations are considerably lower in case of generalized LSDV. The long-run coefficients in Table 4 are again plausible in most cases. However, the bias corrected LSDV estimates take somewhat other values than the other estimators.

Tables 5 and 6 contain the estimation results for M3. The LMF 1 and LMF 2 statistics are 0.000 and 2.369 (p-values are 0.993 and 0.106). The same remarks on the specification of M2 can be made for M3 too. However, for M3 the long-run effect of inflation has the a priori expected negative sign although it is still poorly determined. Also the coefficient of the two-period lagged dependent variable is small and not significant despite its significance in some of the individual regressions. Note the difference in point estimates of the various corrected long-run estimators: while for the short-run estimators the LSDV and GLSDV results are more or less the same, different results are found for the long-run estimators.

We compare the general pattern of the long-term estimates with the (semi) elasticities found in earlier research based on the aggregated time series approach. Overviews of these results can be found in Fase and Winder (1993) and Monticelli and Papi (1996). If not restricted to one, the income elasticity is always close to unity for M1 and larger than one for both M2 and M3. In this study, the long-

run estimates of the various corrected estimators reflect this pattern in general. However, the long-run income estimates of the uncorrected estimators are always below one and likely underestimating the true long-run income effect for M2 and M3. As far as the interest rate semi-elasticities are concerned, in general they are close to the estimates found in earlier studies. The only exception is the long-term interest rate effect for M2; which is found to be particularly strong as compared with other studies. Considering inflation in this study no significant long-run effects for the inflation rate apart from small effects for M3 have been found, whereas other studies do find significant effects also for M1 or M2.

6. Concluding Remarks

With panel data techniques money demand functions for M1; M2 and M3 are estimated for the EU area as a whole. As far as we know, until now only aggregate time series studies have been undertaken in this area. Because of the typical dimensions of the panel at hand, which is dominated by its time dimension, least squared based methods are used instead of instrumental variables techniques. The latter are commonly used in the typical small T, large N panel.

Despite its superior performance in this type of panel, least squares estimators are substantially biased in dynamic models. Hence, approximation formulae for the bias of the various estimators are developed upto order $O(T^{-1})$ using results of Kiviet (1995, 1999) and related work. These approximations are then used to construct bias corrected estimators. From the bias approximations it is seen that falsely assuming a scalar covariance matrix will lead to corrected estimators, which still contain a bias term of order $O(T^{-1})$. This result underlines the importance of taking into account the true covariance structure of the disturbances.

The empirical results show that in general the bias corrected estimators produce plausible long-run effects commonly found in empirical studies on money demand. As such, the panel data approach is a valuable alternative to the aggregate time series approach. Moreover, due to the typical nature of panel data less time observations are needed compared with the aggregated time series approach. As is shown by the empirical results the bias terms can be substantial in this type of data reasserting the importance of more refined estimation techniques. Also the efficiency gains of exploiting the heteroscedasticity and cross-correlation patterns between countries are considerable.

References

Bun, M.J.G., and J.F. Kiviet (1999), On the small sample accuracy of various inference procedures in dynamic panel data models, mimeo.

Fase, M.M.G. and C.C.A. Winder (1993), The demand for money in The Netherlands and the other EC countries, *De Economist* 141, 471-495.

Fase, M.M.G. and C.C.A. Winder (1998), Wealth and the demand for money in the European Union, *Empirical Economics* 23, 507-524.

Freedman, D.A., and S.C. Peters (1984), Bootstrapping a regression equation: some empirical results, *Journal of the American Statistical Association* 79, 97-106.

Kiviet, J.F. (1995), On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics* 68, 53-78.

Kiviet, J.F. (1999), Expectations of expansions for estimators in a dynamic panel data model; some results for weakly exogenous regressors, in C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variables*, Cambridge University Press, Cambridge.

Kiviet, J.F. and G.D.A. Phillips (1994), Bias assesment and reduction in linear error-correction models, *Journal of Econometrics* 63, 215-243.

Kiviet, J.F., Phillips G.D.A., and B. Schipp (1995), The bias of OLS, GLS and ZEF estimators in dynamic seemingly unrelated regression models, *Journal of Econometrics* 68, 241-266.

Kremers, J.J.M. and T.D. Lane (1992), Economic and monetary integration and the aggregate demand for money in the EMS, *IMF sta^x papers vol37*, 777-805.

Monticelli, C. and L. Papi (1996), European integration, monetary co-ordination, and the demand for money, *The Clarendon Press, Oxford*.

Pesaran, M.H., and Z. Zhao (1999), Bias reduction in estimating long-run relationships from dynamic heterogenous panels, in C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variables*, Cambridge University Press, Cambridge.

A. Bias in the ordinary and generalized LSDV estimators

In this appendix the expressions (3.6) and (3.15) are derived. Using the decomposition of W into a irrelevant and relevant stochastic part, i.e. $W = \hat{W} + \mathcal{W}$; we have

$$\begin{aligned} E[W^0 A^i] &= E[\hat{W}^0 A^i] + E[\mathcal{W}^0 A^i] \\ &= \sum_{p=1}^P \text{tr}(\hat{W}_{i p}^0) e_p; \end{aligned} \quad (\text{A.1})$$

using $A\mathcal{W} = \sum_{p=1}^P \hat{W}_{i p}^0 e_p$. For Q we write

$$\begin{aligned} Q &= E[W^0 A W] \\ &= \hat{W}^0 A \hat{W} + E[\mathcal{W}^0 A \mathcal{W}] \\ &= \hat{W}^0 A \hat{W} + \sum_p \sum_r \text{tr}(\hat{W}_{i p}^0 \hat{W}_{i r}^0) e_p e_r^0; \end{aligned} \quad (\text{A.2})$$

Omitting terms with zero moments we have also

$$\begin{aligned} E[\hat{W}^0 A \mathcal{W} Q_i^{-1} W^0 A^i] &= \hat{W}^0 A \hat{W} Q_i^{-1} E[\mathcal{W}^0 A^i] + \hat{W}^0 E[A \mathcal{W} Q_i^{-1} \hat{W}^0 A^i] \\ &\quad + E[\hat{W}^0 A \hat{W} Q_i^{-1} \hat{W}^0 A^i] + E[\mathcal{W}^0 A \mathcal{W} Q_i^{-1} W^0 A^i] \\ &= \hat{W}^0 A \hat{W} Q_i^{-1} \sum_p E[\hat{W}_{i p}^0] e_p + \hat{W}^0 \sum_p \text{tr}(\hat{W}_{i p}^0) A \hat{W} Q_i^{-1} e_p \\ &\quad + \sum_p e_p E[\hat{W}_{i p}^0 \hat{W} Q_i^{-1} \hat{W}^0 A^i] + \sum_p \sum_r \sum_s q_{rs} E[\hat{W}_{i p}^0 \hat{W}_{i r}^0 \hat{W}_{i s}^0] e_p \\ &= \hat{W}^0 A \hat{W} Q_i^{-1} \sum_p \text{tr}(\hat{W}_{i p}^0) e_p \\ &\quad + \sum_p \hat{W}_{i p}^0 A \hat{W} Q_i^{-1} e_p + \sum_p \text{tr}(\hat{W}_{i p}^0 Q_i^{-1} \hat{W}_{i p}^0 A \hat{W}^0 A^i) e_p \\ &\quad + \sum_p \sum_r \sum_s q_{rs} \text{tr}(\hat{W}_{i p}^0 \hat{W}_{i r}^0) \text{tr}(\hat{W}_{i s}^0) + 2 \text{tr}(-\hat{W}_{i p}^0 \hat{W}_{i r}^0 \hat{W}_{i s}^0) e_p \\ &= \sum_p \text{tr}(\hat{W}_{i p}^0) e_p + \sum_p \hat{W}_{i p}^0 A \hat{W} Q_i^{-1} e_p \\ &\quad + \sum_p \text{tr}(\hat{W}_{i p}^0 Q_i^{-1} \hat{W}_{i p}^0 A \hat{W}^0 A^i) e_p \\ &\quad + 2 \sum_p \sum_r \sum_s q_{rs} \text{tr}(-\hat{W}_{i p}^0 \hat{W}_{i r}^0 \hat{W}_{i s}^0) e_p \end{aligned} \quad (\text{A.3})$$

where $q_{rs} = e_r^0 Q_i^{-1} e_s$ and we have used $\hat{W}^0 A \hat{W} = Q_i \sum_p \sum_r \text{tr}(\hat{W}_{i p}^0 \hat{W}_{i r}^0) e_p e_r^0$. Using (A.1) and (A.3) the result in (3.6) now readily follows.

For the bias in the GLSDV estimator we have to evaluate the expectations in (3.14). Now

$$\begin{aligned}
E[W^0 A^{(i)}] &= E[W^0 A^{-i} A^{(i)}] + E[W^0 A^{-i} A^{(i)}] \\
&= \sum_{p=1}^P E[W^0 A^{-i} A^{(i)}] e_p \\
&= \sum_{p=1}^P \text{tr}(A_p) e_p
\end{aligned} \tag{A.4}$$

and for Q^{α} we write

$$\begin{aligned}
Q^{\alpha} &= W^0 A^{-i} A W^0 + E[W^0 A^{-i} A W^0] \\
&= W^0 A^{-i} A W^0 + \sum_{p,r} \text{tr}(A_p A_r) e_p e_r^0
\end{aligned} \tag{A.5}$$

Also we write omitting terms with zero moments

$$\begin{aligned}
E[W^0 A^{(i)} W Q^{\alpha} W^0 A^{(i)}] &= W^0 A^{(i)} W Q^{\alpha} W^0 E[W^0 A^{(i)}] + W^0 E[W^0 A^{(i)} W Q^{\alpha} W^0] W^0 A^{(i)} \\
&\quad + E[W^0 A^{(i)} W Q^{\alpha} W^0] W^0 A^{(i)} + E[W^0 A^{(i)} W Q^{\alpha} W^0] W^0 A^{(i)} \\
&= W^0 A^{(i)} W Q^{\alpha} W^0 \sum_{p=1}^P E[W^0 A^{(i)}] e_p \\
&\quad + \sum_{p=1}^P W^0 A^{(i)} W Q^{\alpha} W^0 E[W^0 A^{(i)}] e_p \\
&\quad + \sum_{p=1}^P E[W^0 A^{(i)} W Q^{\alpha} W^0] W^0 A^{(i)} e_p \\
&\quad + \sum_{p,r,s} q_{rs}^{\alpha} E[W^0 A^{(i)} W Q^{\alpha} W^0] e_p \\
&= W^0 A^{(i)} W Q^{\alpha} W^0 \sum_{p=1}^P \text{tr}(A_p) e_p \\
&\quad + \sum_{p=1}^P W^0 A^{(i)} W Q^{\alpha} W^0 e_p + \sum_{p=1}^P \text{tr}(Q^{\alpha} W^0 A^{(i)} W^0) e_p \\
&\quad + \sum_{p,r,s} q_{rs}^{\alpha} \text{tr}(A_p A_r) \text{tr}(A_s) + 2 \text{tr}(A_p A_r A_s) e_p \\
&= \sum_{p,r,s} \text{tr}(A_p) e_p + \sum_{p=1}^P W^0 A^{(i)} W Q^{\alpha} W^0 e_p \\
&\quad + \sum_{p=1}^P \text{tr}(Q^{\alpha} W^0 A^{(i)} W^0) e_p \\
&\quad + 2 \sum_{p,r,s} q_{rs}^{\alpha} \text{tr}(A_p A_r A_s) e_p
\end{aligned} \tag{A.6}$$

where $q_{rs}^{\alpha} = e_r^0 Q^{\alpha} e_s^0$ and we have used $W^0 A^{(i)} W = Q^{\alpha} \sum_{p,r} \text{tr}(A_p A_r) e_p e_r^0$. Hence, using (A.4) and (A.6) the bias expression in (3.15) follows.

B. Bias in the feasible generalized LSDV estimator

We give in this appendix a proof of (3.19), i.e. the bias approximations of the GLSDV and FGLSDV estimators are the same upto order $O(T^{-1})$: We will need the following

$$\begin{aligned}\hat{S} &= E^0 A_T E = T + O_p(T^{-1}) \\ E &= (\epsilon_1; \dots; \epsilon_N); \end{aligned} \quad (B.1)$$

where ϵ_i is the $T \times 1$ disturbance vector belonging to individual i : A proof of a similar result is given in Kiviet et al. (1993). As a consequence $\hat{S}_i - S$ can be replaced by $E^0 A_T E = T + S$ without changing the order of the approximation, i.e.

$$E^0 A_T E = T + S + O_p(T^{-\frac{1}{2}}); \quad (B.2)$$

and

$$\hat{S}^{i-1} = S^{i-1} + S^{i-1} \frac{E^0 A_T E}{T} + O_p(T^{-1}); \quad (B.3)$$

The estimation error of the FGLSDV estimator (3.18) consists of two factors. The first factor in (3.18) can be expressed as

$$W^0 A (\hat{S}^{i-1} - I_T) A W = W^0 A (S^{i-1} - I_T) A W + W^0 A ((\hat{S}^{i-1} - S^{i-1}) - I_T) A W; \quad (B.4)$$

with

$$\begin{aligned} W^0 A (S^{i-1} - I_T) A W &= W^0 A (S^{i-1} - I_T) A W + W^0 A (S^{i-1} - I_T) A W \\ &\quad + W^0 A (S^{i-1} - I_T) A W + W^0 A (S^{i-1} - I_T) A W \\ &= Q^a + W^0 A (S^{i-1} - I_T) A W + W^0 A (S^{i-1} - I_T) A W \\ &\quad + W^0 A (S^{i-1} - I_T) A W + E W^0 A (S^{i-1} - I_T) A W \\ &= Q^a + A_1 + A_2 + A_3; \end{aligned} \quad (B.5)$$

where $Q^a = O(T)$ and A_1, A_2 and A_3 are $O_p(T^{-\frac{1}{2}})$ and

$$\begin{aligned} W^0 A ((\hat{S}^{i-1} - S^{i-1}) - I_T) A W &= W^0 A \frac{\bar{A}}{T} \frac{E^0 A_T E}{T} (S^{i-1} - I_T) A W + O_p(1) \\ &= W^0 A \frac{\bar{A}}{T} \frac{E^0 A_T E}{T} (S^{i-1} - I_T) A W \\ &\quad + W^0 A \frac{\bar{A}}{T} \frac{E^0 A_T E}{T} (S^{i-1} - I_T) A W \\ &\quad + W^0 A \frac{\bar{A}}{T} \frac{E^0 A_T E}{T} (S^{i-1} - I_T) A W \\ &\quad + W^0 A \frac{\bar{A}}{T} \frac{E^0 A_T E}{T} (S^{i-1} - I_T) A W \\ &\quad + W^0 A \frac{\bar{A}}{T} \frac{E^0 A_T E}{T} (S^{i-1} - I_T) A W + O_p(1); \end{aligned}$$

The second and third term in this expression are actually of order $O_p(1)$, so

$$\begin{aligned} W^0 A((\hat{S}^{i-1} - S^{i-1}) - I_T) A W &= \sum_i W^0 A \frac{\tilde{A}}{T} E^0 A_T E \sum_i S^{i-1} - I_T A W \\ &= \sum_i W^0 A \frac{\tilde{A}}{T} E^0 A_T E \sum_i S^{i-1} - I_T A W + O_p(1) \\ &= A_4 + A_5 + O_p(1); \end{aligned} \quad (B.6)$$

where A_4 and A_5 are $O_p(T^{\frac{1}{2}})$. Hence, from (B.4), (B.5) and (B.6) we ...nd

$$\begin{aligned} W^0 A(\hat{S}^{i-1} - I_T) A W &= Q^{\alpha} + \sum_{i=1}^{\infty} A_i + O_p(1) \\ &= I + \sum_{i=1}^{\infty} A_i + O_p(1) Q^{\alpha i-1} Q^{\alpha}; \end{aligned} \quad (B.7)$$

and

$$\begin{aligned} W^0 A(\hat{S}^{i-1} - I_T) A W^{i-1} &= Q^{\alpha i-1} I + \sum_{i=1}^{\infty} A_i + O_p(1) Q^{\alpha i-1} \\ &= Q^{\alpha i-1} \sum_{i=1}^{\infty} A_i + O_p(1) Q^{\alpha i-1} + O_p(T^{i-2}) \\ &= Q^{\alpha i-1} \sum_{i=1}^{\infty} A_i Q^{\alpha i-1} + O_p(T^{i-2}); \end{aligned} \quad (B.8)$$

We now consider the second factor in (3.18)

$$\begin{aligned} W^0 A(\hat{S}^{i-1} - I_T) A'' &= W^0 A(S^{i-1} - I_T) A'' + W^0 A((\hat{S}^{i-1} - S^{i-1}) - I_T) A'' \\ &= W^0 A(S^{i-1} - I_T) A'' + W^0 A \frac{\tilde{A}}{T} E^0 A_T E \sum_i S^{i-1} - I_T A'' \\ &= \sum_i W^0 A \frac{\tilde{A}}{T} E^0 A_T E \sum_i S^{i-1} - I_T A'' + O_p(T^{i-\frac{1}{2}}) \\ &= A_6 + A_7 + A_8 + A_9 + O_p(T^{i-\frac{1}{2}}); \end{aligned} \quad (B.9)$$

where A_6, A_7 are $O_p(T^{\frac{1}{2}})$ and A_8 and A_9 are $O_p(1)$. From (B.8) and (B.9) the estimation error of the FGLSDV estimator is

$$\hat{\pm}_{\text{FGLSDV}} \pm = \sum_{i=1}^{\infty} Q^{\alpha i-1} \sum_{i=1}^{\infty} A_i Q^{\alpha i-1} \sum_{i=6}^{\infty} A_i + O_p(T^{i-\frac{3}{2}})$$

$$\begin{aligned}
&= Q^{\alpha_i-1} \bar{A}^{\alpha_i-1} A_i + Q^{\alpha_i-1} \bar{A}^{\alpha_i-1} A_i + Q^{\alpha_i-1} \bar{A}^{\alpha_i-1} A_i + O_p(T^{i-\frac{3}{2}}) \\
&= Q^{\alpha_i-1} \bar{A}^{\alpha_i-1} A_i + Q^{\alpha_i-1} \bar{A}^{\alpha_i-1} A_i + O_p(T^{i-\frac{3}{2}}); \tag{B.10}
\end{aligned}$$

using the fact that A_8 and A_9 are $O_p(1)$ and Q^{α_i-1} is $O(T^{i-1})$. Evaluating the expectation of the estimation error in (B.10) we got many terms. Noting that

$$E \hat{\pm}_{\text{GLSDV}}^{\mathbf{h}} \pm^{\mathbf{i}} = Q^{\alpha_i-1} E[A_6 + A_7] + Q^{\alpha_i-1} E^{\mathbf{h}}(A_1 + A_2 + A_3)Q^{\alpha_i-1}(A_6 + A_7) + o(T^{i-1}); \tag{B.11}$$

we have the following

$$E \hat{\pm}_{\text{FGLSDV}}^{\mathbf{h}} \pm^{\mathbf{i}} = Q^{\alpha_i-1} E[A_8 + A_9] + Q^{\alpha_i-1} E^{\mathbf{h}}(A_4 + A_5)Q^{\alpha_i-1}(A_6 + A_7) + o(T^{i-1}); \tag{B.12}$$

We have to evaluate the expectations of the six remaining terms on the right hand side in (B.12). It is easily seen that $E[A_8] = 0$ and $E[A_4Q^{\alpha_i-1}A_6] = 0$: We will sketch the proof that the other four expectations are all of order $O(T^{i-1})$, so premultiplied by Q^{α_i-1} their contribution is $o(T^{i-1})$. In the following all summations run from 1 to N except the index p ; which runs from 1 to P . Consider ...rst

$$\begin{aligned}
E[A_9] &= \sum_i E[W^0 A \bar{S}^{i-1} \frac{E^0 A_T E}{T} \bar{S}^{i-1} - I_T A] \\
&= \sum_i E[W^0 A \bar{S}^{i-1} \frac{E^0 A_T E}{T} \bar{S}^{i-1} - I_T A] \\
&\quad + E[W^0 A (\bar{S}^{i-1} - I_T) A] : \tag{B.13}
\end{aligned}$$

Defining \mathbf{w}_j as the j th column of \bar{S}^{i-1} and \mathbf{w}_{ij} as its ij th element we write for the ...rst term in (B.13)

$$E[W^0 A \bar{S}^{i-1} \frac{E^0 A_T E}{T} \bar{S}^{i-1} - I_T A] = \sum_i \sum_j E[\mathbf{w}_i^0 \frac{E^0 A_T E}{T} \mathbf{w}_j W_i^0 A_T] : \tag{B.14}$$

Evaluating a particular term in (B.14) we got

$$E[\mathbf{w}_i^0 \frac{E^0 A_T E}{T} \mathbf{w}_j W_i^0 A_T] = \sum_r \sum_s E[\mathbf{w}_i^0 \frac{E^0 A_T E}{T} \mathbf{w}_j W_i^0 A_T]$$

$$\begin{aligned}
&= \sum_{r,s} \sum_{p} E \frac{1}{T} \sum_{i,j} \frac{A_{rs}^0}{T} \sum_{i,p} \sum_{j} e_p \\
&= \frac{1}{T} \sum_{r,s,p} \sum_{i,j} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \\
&= \frac{1}{T} \sum_{r,s,p} \sum_{i,j} \frac{1}{T} \sum_{i,p} \sum_{j} e_p (\text{tr}(A_T) \text{tr}(I_{pT}) + 2 \text{tr}(I_{pT})) \\
&= \sum_{r,s,p} \sum_{i,j} \frac{1}{T} \sum_{i,p} \sum_{j} e_p + O(T^{-1}) \\
&= E \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p + O(T^{-1}) \tag{B.15}
\end{aligned}$$

Hence, substituting (B.15) into (B.14)

$$E \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p = E \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p + O(T^{-1}) \tag{B.16}$$

and now it is easily seen from (B.13) and (B.16) that $E[A_9]$ is of order $O(T^{-1})$:

Next consider

$$\begin{aligned}
E \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p &= \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \\
&= \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \\
&\quad + \sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \tag{B.17}
\end{aligned}$$

For the ...rst term in (B.17) we write

$$\begin{aligned}
&\sum_{i,j} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \\
&= \sum_{i,j,k,l} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \tag{B.18}
\end{aligned}$$

with

$$\begin{aligned}
&\sum_{i,j,k,l} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \\
&= \sum_{p} \sum_{i,j,k,l} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \\
&= \sum_{p} \sum_{r,s} \sum_{p} \frac{1}{T} \sum_{i,p} \sum_{j} e_p \tag{B.19}
\end{aligned}$$

The expectation of a particular term in (B.19) is

$$\begin{aligned}
& E \left[\frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^p \sum_{l=1}^p \sum_{r=1}^p \sum_{s=1}^p \sum_{t=1}^p \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \sum_{x=1}^p \sum_{y=1}^p \sum_{z=1}^p \sum_{\substack{i \\ j \\ k \\ l \\ p \\ T}} \mathbf{A}_T \mathbf{W}_i^0 \mathbf{A}_T \mathbf{W}_j \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{e}_p \right] \\
&= \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^p \sum_{l=1}^p \sum_{r=1}^p \sum_{s=1}^p \sum_{t=1}^p \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \sum_{x=1}^p \sum_{y=1}^p \sum_{z=1}^p \sum_{\substack{i \\ j \\ k \\ l \\ p \\ T}} \mathbf{A}_T \mathbf{W}_i^0 \mathbf{A}_T \mathbf{W}_j \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{e}_p \\
&= \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^p \sum_{l=1}^p \sum_{r=1}^p \sum_{s=1}^p \sum_{t=1}^p \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \sum_{x=1}^p \sum_{y=1}^p \sum_{z=1}^p (\text{tr}(\mathbf{A}_T) \text{tr}(\mathbf{I}_{p \times T}) + 2 \text{tr}(\mathbf{I}_{p \times T})) \mathbf{W}_i^0 \mathbf{A}_T \mathbf{W}_j \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{e}_p \\
&= \sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^p \sum_{l=1}^p \sum_{r=1}^p \sum_{s=1}^p \sum_{t=1}^p \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \sum_{x=1}^p \sum_{y=1}^p \sum_{z=1}^p \text{tr}(\mathbf{I}_{p \times T}) \mathbf{W}_i^0 \mathbf{A}_T \mathbf{W}_j \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{e}_p + O(T^{-1}); \tag{B.20}
\end{aligned}$$

Using this result and follow the same steps back it follows that $E[\mathbf{A}_4 \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{A}_7]$ is of order $O(T^{-1})$: In the same fashion the expectations of the remaining two terms $\mathbf{A}_5 \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{A}_6$ and $\mathbf{A}_5 \mathbf{Q}^{\mathbf{i}^{-1}} \mathbf{A}_7$ can be shown to be of order $O(T^{-1})$ too.

Having derived the order of magnitude of the expectations of the several terms on the right hand side of (B.12) and noting that $\mathbf{Q}^{\mathbf{i}^{-1}}$ is of order $O(T^{-1})$; it is straightforward to see that

$$E \left[\hat{\mathbf{h}}_{\pm \text{FGLSDV } \mathbf{i} \pm \mathbf{i}} \right] = E \left[\hat{\mathbf{h}}_{\pm \text{GLSDV } \mathbf{i} \pm \mathbf{i}} \right] = o(T^{-1}); \tag{B.21}$$

so the magnitude of the bias upto order $O(T^{-1})$ is the same for the GLSDV and FGLSDV estimators.

Table 1:
Estimation results of the short-run coefficients for M1*

	LSDV	LSDV _c	FGLSDV	FGLSDV _c
$\ln(M1=P)_{i;tj-1}$	0.631 (0.059)	0.733 (0.061)	0.645 (0.045)	0.682 (0.042)
$\ln gnp_{it}$	0.278 (0.045)	0.248 (0.042)	0.277 (0.033)	0.269 (0.032)
$\ln gnp_{i;tj-1}$	0.020 (0.041)	-0.005 (0.039)	0.018 (0.030)	0.012 (0.030)
rs_{it}	-0.006 (0.002)	-0.006 (0.002)	-0.005 (0.002)	-0.004 (0.002)
$rs_{i;tj-1}$	0.002 (0.002)	0.003 (0.002)	0.001 (0.002)	0.001 (0.002)
rl_{it}	-0.006 (0.003)	-0.005 (0.003)	-0.004 (0.003)	-0.004 (0.003)
$rl_{i;tj-1}$	0.004 (0.004)	0.003 (0.003)	0.003 (0.003)	0.001 (0.003)
ir_{it}	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)
$ir_{i;tj-1}$	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.002 (0.002)

* N=14, T=19, P+K=12

* Figures in parentheses are standard errors

Table 2:
Long-run estimates for M1*

	LSDV	LSDV c	FGLSDV	FGLSDV c		
gnp	0.809 (0.157)	0.910 (0.194)	0.935 (0.192)	0.829 (0.119)	0.881 (0.120)	0.880 (0.120)
rs	-0.012 (0.004)	-0.012 (0.005)	-0.012 (0.005)	-0.011 (0.003)	-0.011 (0.003)	-0.012 (0.003)
rl	-0.005 (0.007)	-0.007 (0.008)	-0.007 (0.008)	-0.007 (0.006)	-0.007 (0.006)	-0.008 (0.006)
ir	-0.000 (0.005)	0.000 (0.005)	-0.000 (0.005)	0.000 (0.004)	0.000 (0.004)	0.000 (0.004)

* Figures in parentheses are standard errors

Table 3:
Estimation results of the short-run coefficients for M2*

	LSDV	LSDV c	FGLSDV	FGLSDV c
$\ln(M2=P)_{i;t_i-1}$	0.906 (0.046)	0.979 (0.047)	0.912 (0.036)	0.972 (0.036)
$\ln(M2=P)_{i;t_i-2}$	-0.085 (0.043)	-0.071 (0.044)	-0.110 (0.034)	-0.139 (0.034)
$\ln gnp_{it}$	0.210 (0.037)	0.203 (0.037)	0.199 (0.028)	0.200 (0.028)
$\ln gnp_{i;t_i-1}$	-0.041 (0.041)	-0.061 (0.041)	-0.043 (0.032)	-0.055 (0.032)
rs_{it}	-0.001 (0.002)	-0.000 (0.002)	-0.000 (0.001)	-0.000 (0.001)
$rs_{i;t_i-1}$	0.004 (0.001)	0.004 (0.001)	0.003 (0.001)	0.003 (0.001)
rl_{it}	-0.005 (0.002)	-0.006 (0.002)	-0.004 (0.002)	-0.004 (0.002)
$rl_{i;t_i-1}$	-0.000 (0.002)	0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
ir_{it}	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)
$ir_{i;t_i-1}$	-0.002 (0.002)	-0.002 (0.002)	-0.003 (0.002)	-0.003 (0.002)

* N=14, T=18, P+K=13

* Figures in parentheses are standard errors

Table 4:
Long-run estimates for M2*

	LSDV	LSDV c	FGLSDV	FGLSDV c
gnp	0.945 (0.250)	1.540 (0.257)	1.383 (0.257)	0.781 (0.203)
rs	0.017 (0.005)	0.035 (0.005)	0.029 (0.005)	0.016 (0.004)
rl	-0.032 (0.008)	-0.053 (0.008)	-0.048 (0.008)	-0.030 (0.007)
ir	0.007 (0.009)	0.013 (0.009)	0.010 (0.009)	0.001 (0.007)

* Figures in parentheses are standard errors

Table 5:
Estimation results of the short-run coefficients for M3*

	LSDV	LSDV c	FGLSDV	FGLSDV c
$\ln(M3=P)_{i;t_1}$	0.840 (0.090)	0.918 (0.094)	0.822 (0.074)	0.847 (0.075)
$\ln(M3=P)_{i;t_2}$	0.018 (0.087)	0.027 (0.089)	0.030 (0.070)	0.014 (0.071)
$\ln gnp_{it}$	0.200 (0.039)	0.190 (0.039)	0.199 (0.030)	0.198 (0.030)
$\ln gnp_{i;t_1}$	-0.068 (0.037)	-0.092 (0.038)	-0.056 (0.028)	-0.059 (0.028)
rs_{it}	-0.000 (0.001)	-0.000 (0.002)	-0.001 (0.001)	-0.001 (0.001)
$rs_{i;t_1}$	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)	0.003 (0.001)
rl_{it}	-0.004 (0.002)	-0.004 (0.002)	-0.003 (0.002)	-0.003 (0.002)
$rl_{i;t_1}$	0.002 (0.002)	0.002 (0.002)	0.000 (0.002)	0.000 (0.002)
ir_{it}	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.001)	-0.002 (0.001)
$ir_{i;t_1}$	0.002 (0.002)	0.002 (0.002)	0.001 (0.001)	0.001 (0.001)

* N=14, T=18, P+K=13

* Figures in parentheses are standard errors

Table 6:
Long-run estimates for M3*

	LSDV	LSDV c		FGLSDV	FGLSDV c	
gnp	0.928 (0.264)	1.772 (0.289)	1.506 (0.295)	0.972 (0.219)	0.999 (0.219)	1.335 (0.237)
rs	0.011 (0.007)	0.035 (0.008)	0.025 (0.008)	0.014 (0.006)	0.015 (0.006)	0.025 (0.007)
rl	-0.018 (0.011)	-0.037 (0.012)	-0.031 (0.012)	-0.019 (0.009)	-0.019 (0.009)	-0.027 (0.010)
ir	-0.001 (0.010)	-0.000 (0.010)	-0.002 (0.011)	-0.003 (0.007)	-0.003 (0.007)	-0.004 (0.008)

* Figures in parentheses are standard errors