Is Robust Inference with OLS Sensible in Time Series Regressions? Investigating Bias and MSE Trade-offs with Feasible GLS and VAR Approaches.

Richard T. Baillie

Department of Economics, Michigan State University, USA

School of Economics and Finance, Queen Mary University of London, UK

Rimini Center for Economic Analysis, Italy

Kun Ho Kim

Department of Economics, Hanyang University, Seoul, Korea

February 15, 2016

Abstract

It has become commonplace in applied time series econometric work to estimate regressions with consistent, but asymptotically inefficient OLS and to base inference of conditional mean parameters on robust standard errors. This approach seems mainly to have occurred due to concern at the possible violation of strict exogeneity conditions from applying GLS. We first show that even in the case of the violation of contemporaneous exogeneity, that the asymptotic bias associated with GLS will generally be less than that of OLS. This result extends to Feasible GLS where the error process is approximated by a sieve autoregression. The paper also examines the trade offs between asymptotic bias and efficiency related to OLS, feasible GLS and inference based on full system VAR. We also provide simulation evidence and several examples including tests of efficient markets, orange juice futures and weather and a control engineering application of furnace data. The evidence and general conclusion is that the widespread use of OLS with robust standard errors is generally not a good research strategy. Conversely, there is much to recommend $FGLS$ and VAR system based estimation.

JEL Classification: C22; F31; G01; G15.

Keywords: OLS, GLS, bias, robust inference, strict exogeneity.

1 Introduction

Traditional econometric textbooks usually devote a chapter on the topic of Generalized Least Squares (GLS), and describe its desirable properties in terms of efficiency. Extensions to Feasible (GLS), or $FGLS$, where the covariance matrix of disturbances are replaced with estimated quantites are also described in this literature. However, the application of $FGLS$ in applied time series econometric work has now become relatively rare and the standard technique has become to use consistent, but asymptotically inefficient, OLS estimation of regression parameters and to then use robust standard errors for subsequent inference on the regression parameters which are considered the primary parameters of interest. For the rest of this article we refer to this as the Ordinary Least Squares Robust Standard Error, or $OLSRSE$ approach. We question the desirability of this approach and provide favorable evidence for the use of $FGLS$. First,we show that even in the case where there is contemporaneous correlation between a regression error term and the explanatory variables used within that regression, that GLS can have lower bias as well as MSE compared with OLS. This result is found to extend to more general situations where $FGLS$ is based on using a sieve AR approximation to the true unknown regression error process.

Pierce (1971) provided a detailed treatment of the MLE of regressions with $ARMA$ disturbances, exogenous variables and independently identically distributed Guassian innovations. Many other subsequent articles used related dynamic models such as the Autoregressive Lag (ADL) Model and then $VARs$ were used to model the dynamics in the conditional means of time series data. Correspondingly there are many studies estimating using GLS or $FGLS$ to estimate regressions with $ARMA$ errors. More recently, Kapetanios and Psaradakis (2013) have shown how $FGLS$ can be implemented through the use of sieve ARs on regression residuals. They find that the subequent $FGLS$ estimator dominates in terms of coverage rates, many of the currently known methods for calculating $OLSRSE$. Despite the extensive amount of past work on parameteric modeling of dynamic relationships with time series data, there has been significant recent use of the $OLSRSE$ methodology where the only model being estimated is a static linear regression and all the attention focused on inference involving robust standard errors of the corresponding regression parameters.

There appears to be two main justifications provided by applied researchers for the use of the $OLSRSE$ approach. The first reason that has been stated is based around the claim that it is difficult to specify an appropriate error structure for regression disturbance covariance matrix. However, this objection is largely irrelevant given the complicated parametric time series models being routinely estimated and also by the findings of Kapetanios and Psaradakis (2013) who find that the $FGLS$ estimator dominates in terms of coverage rates, many of the currently known methods for calculating $OLSRSE$[1].

The second reason frequently given for the use of $OLSRSE$ procedures is the concern that GLS requires filtering of observed time series which can lead to a violation of the exogeniety requirements and hence lead to inconsistent estimates of the regression parameters. The corollary suggested by several authors is to always using consistent, but asymptotically inefficient OLS estimates and robustifying to obtain the $OLSRSE$ inferential method. These arguments go back at least to Hansen and Hodrick (1980), Hsieh (1983) and Hayashi and Sims (1983); and originally arose in regression model based tests of rational expectations and market efficiency. Subsequently, the very influential article by Newey and West (1987) led to a standard method for estimating regression residuals covariance matrix and hence of computing $OLSRSE$ method.

Despite the widespead use of this methodology, no previous study to our knowledge, has attempted to investigate the trade offs between bias and MSE in the choice of using $OLSRSE$, or $FGLS$ or the obvious alternative of the estimation of a full system of equations which has recently been advocated by Sims (2010). As noted by Sims (2010), "to simply push the NW button in STATA" may not always be the optimal strategy.

While the application of HAC may be desirable in virtually situations in cross section and some panel data studies; we argue that it is far from clear in time series applications. In particualr this arises from the relative ease of modeling time dependency in the conditional first moment. It is important to distinguish this from the application of $QMLE$ derived by Bollerslev and Wooldridge (1991), which provides robust standard errors for a wide variety of departures from Gaussianity. Hence the $QMLE$ approach uses MLE assuming Gaussianity to provide consistent estimation of the parameters in the first and second conditional moments and then robustifies for possible departures from normality. This method is then valid given the correct specification of the first two conditional moments.

Our results suggest that $FGLS$ is generally to be the preferred single equation estimation

[1] The comments in this paper are confined to time series data. In cross section work there seems eminent justification for robustifying for unobserved heterogeneity and to use White standard errors, or HAC.

and inferential procedure compared with $OLSRSE$ even when the assumption of contemporaneous exogeneity has been violated. It also seems that given contemporaneous exogeneity that the breakdown of strict exogeneity is relatively unusual. It generally requires explanatory variables to include returns or rational expectations forecast errors. Simple LM testss based on the cross covariance function of OLS residuals and lagged and leading explanatory variables can be an appropriate form of diagnostic test. However, the ideal and preferred strategy will generally involve inference in a system such as a VAR. At a more fundamental level this paper questions the whole motivation of using simple regression estimation procedures on time series data. We know from simple VAR analysis that relationships between variables over time are generally hightly complicated dynamic systems with lagged variables. The estimation of a static regression in such a situation is very unclear in terms of its usefulness or interpretation. In the final analysis, Impulse Response (IR) seems the most generic and robust form of inferenece in multivariate time series data. Some simple examples later in this article illustrate the perils of estimating single equation static models.

The plan of the rest of the paper is as follows: the next section analyzes the properties involving bias and MSE of OLS. GLS and $FGLS$ in the single equation case where there is contemporaneous exogeniety. The deree of bias of GLS estimator are surprisingly good compared with those of OLS. Section 3 investigates the properties of $OLSRSE$, $FGLS$ and $QMLE$ in the context of a VAR data generating process. The presence of joint endogeneity tends to increase the degree of bias and MSE likely to occur in applied work. The next section discusses the concept of strict endogeneity and concludes that it is of relatively little importance in applied work, which suggests that for single equation inference in most cases $FGLS$ should be used instead of $OLSRSE$.

2 OLS and GLS without Exogeneity

Many of the basic issues in this paper can be ilustrated with the very simple single variable regression model with $AR(1)$ errors and an explanatory variable which is also an autoregressive process; so that

$$y_t = \beta x_t + u_t \tag{1}$$

and

$$u_t = \phi u_{t-1} + \varepsilon_{u,t} \tag{2}$$

where $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t \varepsilon_s) = 0$ for $s \neq t$. The consistency of the inefficient OLS estimator of the regression parameter in equation (1) depends only upon the condition of contemporaneous exogeneity that $E(x_t u_t) = 0$. This is the essential condition for the consistency of the OLS estimator. The application of GLS will be asymptotically equivalent to the Cochrane Orcutt estimator, which is

$$(y_t - \phi y_{t-1}) = \beta(x_t - \phi x_{t-1}) + (u_t - \phi u_{t-1}) \tag{3}$$

The consistency of GLS will now require that

$$E(x_t - \phi x_{t-1})(u_t - \phi u_{t-1}) = 0 \tag{4}$$

which implies the following four conditions; (i) $E(x_t u_t) = 0$, (ii) $E(x_{t-1} u_{t-1}) = 0$, (iii) $E(x_{t-1} u_t) = 0$ and (iv) $E(x_t u_{t-1}) = 0$. The first two conditions imply that the contemporaneous error is uncorrelated with the contemporaneous explanatory variable, as in the standard assumption for the consistency of OLS. However the third and fourth conditions requite strictly exogenous regressors, where the errors are uncorrelated with both past and future regressors. All of these conditions are then necesary for the consistency of GLS.

Before dealing with the impact of the issue of strict exogeneity, we first consider properties of OLS and GLS in the above simple model when there is the additional $AR(1)$ equation to generate the explanatory variable

$$x_t = \rho x_{t-1} + \varepsilon_{x,t} \tag{5}$$

and where there are the additional assumptions that $E(\varepsilon_{x,t}) = 0$, $E(\varepsilon_{x,t}^2) = \sigma_x^2$ and $E(\varepsilon_{x,t,}\varepsilon_{x,s}) = 0$ for $s \neq t$. Hence we now assume that the error disturbances are contemporaneously correlated with the disturbances of the explanatory variable, so that $E(\varepsilon_{u,t}\varepsilon_{x,t}) = \sigma_{ux}^2 \neq 0$. Hence the standard assumption of contemporaneous exogeneity is violated and hence both OLS and GLS will be inconsistent estimators of the slope parameter, β.

For the time being we maintain the assumption of strict exogeneity so that $E(\varepsilon_{u,t} x_{t-k}) = 0$, for all integer values of k except when $k = 0$ so that only contemporaneous exogeneity is being violated. Initially we focus on the properties of the inconsistent OLS and GLS estimators. It is relatively straightforward to show that

$$\widehat{\beta_{OLS}} = \left(\sum x_t^2\right)^{-1} \sum x_t y_t$$

and

$$p\lim\left(\widehat{\beta_{OLS}} - \beta\right) = \left(\frac{1-\rho^2}{\sigma_x^2}\right)\left(\frac{\sigma_{ux}^2}{1-\rho\phi}\right) \tag{6}$$

Correspondingly the GLS estimator is

$$\widehat{\beta}_{GLS} = \left\{\sum (x_t - \phi x_{t-1})^2\right\}^{-1} \left\{\sum (x_t - \phi x_{t-1})(y_t - \phi y_{t-1})\right\}$$

and we can show that

$$p\lim\left(\widehat{\beta_{GLS}} - \beta\right) = \left(\frac{1-\rho^2}{\sigma_x^2}\right)\left(\frac{\sigma_{ux}^2}{1+\rho^2-2\rho\phi}\right) \tag{7}$$

The relative bias of the estimators can be seen from the ratio

$$\frac{p\lim\left(\widehat{\beta_{OLS}} - \beta\right)}{p\lim\left(\widehat{\beta_{GLS}} - \beta\right)} = \left(\frac{1-\rho\phi}{1-\rho\phi+\phi(\phi-\rho)}\right) \tag{8}$$

In a typical economic time series the autocorrelation is generally positive so that it is reasonable to examine bias and MSE for the intervals of $0 < \rho < 1$ and $0 < \phi < 1$. Then the bias of the OLS estimator will exceed that of the GLS estimator if $\phi(\phi - \rho) > 0$, which implies that the bias of OLS is greater than GLS when the persistence of the autoregressive component of the error term is greater than that of the explanatory variable. This interesting result has also to be seen in the important context that MSE of the OLS estimator considerably exceeding that of the GLS estimator. The simulation results in Table 1 highlight this result for the above process. The simulation results also include results for the $FGLS$ estimator which is defined as

$$\widehat{\beta}_{FGLS} = \left\{\sum \left(x_t - \widehat{\phi} x_{t-1}\right)^2\right\}^{-1} \left\{\sum \left(x_t - \widehat{\phi} x_{t-1}\right)\left(y_t - \widehat{\phi} y_{t-1}\right)\right\} \tag{9}$$

where $\widehat{\phi} = \left(\sum \widehat{u_t^2}\right)^{-1} \left(\sum \widehat{u}_t \widehat{u_{t-1}}\right)$, and $\widehat{u}_t = y_t - \widehat{\beta_{OLS}} x_t$. It should be noted that this "feasible"

GLS estimator assumes that the order of the error process is known and that it is only the error process parameter(s) that are unknown. In practice, the investigator will typically know considerably less than this, and in the next section of the paper we make our investigation more realistic by considering a sieve AR approximation to the unknown error process of $\widehat{u_t}$.

3 Robust Inference in Single Equation Regressions

The above results in section 2 would seem to imply that parameteric modeling of a regression error and subsequent estimation of a regression with say $ARMA$ errors may be a good strategy. Indeed the methodology and properties for this has been provided many years ago by Pierce (1971) and others. In particular, Pierce (1971) provided a detailed treatment of the MLE of regressions with $ARMA$ disturbances with independent and identically Guassian innovations. However, both the supposed difficulty of modeling regressoin errors with an $ARMA$ or other model and concern about the possible breakdown of strict exogeneity has led many researchers to ignore the time series literature of jointly modeling structural and error proces dynamics simualtaneously and to revert back to using OLS estimation supplemented with robust standard errors as provided by Newey and West (1987) for the implementation of robust inference. Hence they abandon striving for efficiency and their research strategy is purely determined by using consistent OLS estimators. The possibility of the violation of strict exogeneity is generally assumed rather than tested for, and the researcher then settles for using inefficient estimation. In this context we now consider estiamtion of the instantaneous regression of the form,

$$y_t = \beta^{/} x_t + u_t \tag{10}$$

where $\beta^{/} = (\beta_1 ... \beta_k)$ and is a k dimensional vector of parameters and $x_t^{/} = (x_{it}, x_{2t}, ...x_{kt})$ is a k dimensional vector of explanatory variables. Also, u_t is a weakly stationary process such that $E(u_t) = 0$, and $E(u_t^2) < \infty$ and $E(u_t u_{t-r}) \to 0$ as $r \to \infty$. It is assumed that $Q = p \lim \sum_{j=1} (x_j x_j^{/})$. To guard against the possible violation of the strict exogeneity requirement the $OLSRSE$ estimator is commonly used. Sometimes the phrase "instantaneous relationship" is used to justify estimating this static regression. However, it seems clearly true that multivariate time series data usually implies complex leading and lagging relationships between dynamic variables; in which case the static regression does not capture any impact, or interim nor long run multiplier, and seems a genuinely uninteresting quantity.

In general the OLS estimator has the limiting distribution of

$$T^{1/2}(\hat{\beta} - \beta_0) \rightarrow N(Q^{-1}VQ^{-1}) \tag{11}$$

The OLSRSE regression required assumptions follow from the conditions 5, page 217 in Hamilton (1994) which are that (i) all the x_t are stochastic and independent of all u_s for all s and t; (ii) $x_t u_t$ is an mds; (iii) is a psd matrix, with which is also psd and (iv) for all i, k, p, q and t.

Hence the "instantaneous regression" type regresion is meant to focus attention on the main parameter or effect of interest, namely the instantaneous effect of the explanatroy variables on the dependent variable. In order to guard against mis-specification, variable omission or any unknown dynamics, it is assumed to be approximated by a high order autoregression which does not affect the consistency of the estimator. Rather than risk possible inconsistency of GLS through correcting for the AR error the $OLSRSE$ methodology is adopted. The main issue then becomes of estimating the covariance matrix V,

$$Cov\left(\hat{\beta}\right) = \left(T^{-1}\sum_{t=1} x_t x_t'\right)^{-1} \left(T^{-1}\sum_{t=1} x_t x_t' \hat{u}_t^2 + \sum_{j=1}\{1 - j/(q+1)\}(\mathbf{\Gamma}_j + \mathbf{\Gamma}_{-j})\right) \left(T^{-1}\sum_{t=1} x_t x_t'\right)^{-1} \tag{12}$$

where

$$\mathbf{\Gamma}_j = T^{-1}\sum_{t=1} \hat{u}_t x_t x_{t-j}^{/} \hat{u}_{t-j}$$

there are several versions of the estimator of the error covariance matricx in the above robust, or HAC consistent estimator due to Newey and West (1987), and the resulting standard errors are said to be robust to both heteroskedasticity and autocorrelation. Note that the HAC depends upon thre choice of bandwidth q, and also the form of the kernel which in this case is the Barlett kernel with linearly decaying weights. The choice of bandwidth has been discussed by Andrews (1991), Newey and West (1994) who emphasize reduction in the MSE of the estimation of the covariance matrix and also by Sun, Phillips and Jun (2008) who discuss the choice of lag length. Kapetanios and Psaradakis (2013) analyze the practical impact of the choices of these quantities in the implementation of the semi parametric NW statistic and contrast this with $FGLS$.

With only cross section date the method reduces to White standard errors and are robust to heteroskedasticity of unknown form. Thus seems a generally sensible procedure in the context of cross section data; and is not at issue in this paper, which restricts attention to time series data alone.

4 Feasible GLS without Exogeneity

We consider the single equation with k explanatory variables and an autocorrelated disturbance.

$$y_t = \beta^{/} x_t + u_t \tag{13}$$

Then the application of GLS will require

$$\widehat{\beta} = (X^{/}\Omega^{-1}X)^{-1}X^{/}\Omega^{-1}Y \tag{14}$$

and when Ω is unknown, a $FGLS$ estimator can be used and following Amemiya (1973) an approximating high order $AR(p)$ model can be implemented to the error term z_t in the static regression

$$\widehat{z_t} = y_t - \widehat{b_{OLS}} x_t \tag{15}$$

For a positive integer p whuch is based on a funtion of the sample size T, then as $p \to \infty$ and $p/T \to 0$ as $T \to \infty$ and let $\widehat{\phi}(p) = \left(\widehat{\phi_1}, ...\widehat{\phi_p}\right)$ denote the OLS estimator of the scalar $AR(p)$ model which are obtaind by the minimization of

$$(T-p)^{-1} \sum \left(\widehat{z_t} - \widehat{\phi_1}\widehat{z_{t-1}} - .... - \widehat{\phi_p}\widehat{z_{t-p}}\right)^2$$

over the range $\phi(p) \in \mathbb{R}^p$. Then a convenient method for calculating the $FGLS$ is

$$\widehat{b_{FGLS}} = \left\{\sum \left(1 - \sum \widehat{\phi}_j x_{t-j}\right)\left(1 - \sum \widehat{\phi}_j x_{t-j}\right)\right\}^{-1} \left\{\sum \left(1 - \sum \widehat{\phi}_j x_{t-j}\right)\left(\sum \widehat{\phi}_j y_{t-j}\right)\right\}$$

It is necesary to define the asymptotic properties of the $FGLS$ which covers weak stationarity of the errors and regressors.

Assumption 1:(i) The error process z_t is α-mixing of size $-\eta$ for some $\eta > 1$, (ii) $sup_t E(|z|^{2\kappa}) < \infty$, for some $\kappa > 2$, (iii) $\sum j^c |\delta| < \infty$ for some $c > 0.5$, such that $\{c(\kappa-2)\} / \{2(\kappa-1)\} > 0.5$.

Assumption 2: (i) The regressor process $\{x_t\}$ is an L_{2r} bounded L_2 near epoch dependent

L_2 NED process of size $-c$ on a g dimensional α-mixing process of size $-\eta$ where$\eta > 1$, such that $\{c(\kappa-2)\}/\{2(\kappa-1)\} > 0.5$.for some $r > 2$; (ii) $\{x_t\}$ and $\{z_t\}$ are mutually independent and (iii) $E(x_t x_t^{/})$ is non singular.

Clearly, the application of $FGLS$ in the situation we are considering will violate the part (ii) of assumption 2. The $FGLS$ estimator is based on estimating a sieve $AR(p)$ model to the OLS residuals. Then

$$\left(\widehat{b_{FGLS}} - b\right) = \left\{\sum\left(1-\sum\widehat{\phi}_j x_{t-j}\right)\left(1-\sum\widehat{\phi}_j x_{t-j}\right)\right\}^{-1}\left\{\sum\left(1-\sum\widehat{\phi}_j x_{t-j}\right)\left(\sum\widehat{z}_j y_{t-j}\right)\right\}$$

and expresions for its asymptotic bias is derived in Appendix 2. One of the features of the limiting distribution of the $FGLS$ estimator is the dependence on choice of order of the AR for the OLS residuals which is based on the BIC.

5 Simulation for Single Equation Model

We consider the single eqThe relative bias and MSE of the different forms of estimators in the above situation can be illustrated by the single variable model with $k=1$. Hence the simulation results in Tables 1 and 2 are for the process design

$$y_t = \beta x_t + u_t \tag{16}$$

$$\phi(L)u_t = \epsilon_{u,t} \tag{17}$$

$$\rho(L)x_t = \epsilon_{x,t} \tag{18}$$

The interest in the results are derived from the non zero covariance $\sigma^2_{u,x}$ which provides for contemporaneous exogeneity. In all the simulation designs the innovations $\epsilon_{x,t}$ and $\epsilon_{u,t}$ are generated from a bivariate $NID(0,V)$ process where

$$V = \begin{pmatrix} \sigma_x^2 & \sigma_{x,\epsilon}^2 \\ \sigma_{x,\epsilon}^2 & \sigma_\epsilon^2 \end{pmatrix}.$$

For the results in Table 1 the parameters are set as $\beta = 2, \rho = 0.5, \phi = 0.9$; and $V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. Hence the results in Table 1 are essentially for the same situation as analyzed in section 3 with $AR(1)$ processes for both the explanatory variable and the regression error process. As expected, the results in Table 1 indicate the considerable increase in efficiency and hence reduction in MSE of the estimate of β from using GLS compared with OLS. However, the results also provide quite dramatic evidence of the reduction in bias associated with GLS compared to OLS. These results are in line with the theoretical result in section 3 and are to some extent more extreme than expected. Furthermore, the performance of the $FGLS$ and show that its properties are very close to those of the GLS estimator of β. The results tend to strongly support the application of GLS or $FGLS$ in this model regardless of whether or not cotemporaneous exogeneity is present.

The generality of these results is seen in the further simulations where the error process and generating mechanism for the explanatory variable are both from higher order AR processes. and in particular $AR(3)$ processes. The designs used in Table 2 are for $\phi(L) = (1 - 0.50L + .56L^2 - .08L^3)$ and $\rho(L) = (1 - 0.8800L + .8385L^2 - .7220L^3)$ which has complex conjugate roots of 1 ± 1 and 0.9. These results suggest that in the single equation context where past dependent variable (y) does not Granger-cause the explanatory variable (x), and y and x only have contemporaneous Granger-causality; then the breakdown of contemporaneous exogeneity may not be a significant problem in terms of inducing additional parameter estimation bias. However, the overall performance of GLS and $FGLS$ in terms of bias and MSE are considerably better than OLS. Section 4 of this paper explores a more general and to some extent more practically important situation where y and x are jointly endogenous, and it will be shown that the performanace of single equation OLS methodology deteriorates substantially, while GLS and $FGLS$ in general provides significant improvement.

For the general linear variable regression model with $AR(p)$ errors and an explanatory variable which is also an autoregressive process; so that

$$y_t = \beta^/ x_t + u_t \tag{19}$$

and

$$\phi(L)u_t = \varepsilon_{u,t} \tag{20}$$

where $\phi(L) = \left(1 - \phi_1 L - ... - \phi_p L^p\right)$, while $x_t^{/} = (x_{1,t}.....x_{k,t})$ and $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t \varepsilon_s) = 0$ for $s \neq t$. Then the $FGLS$ estimator will be

$$\widehat{\beta_{FGLS}} = \left[\sum \left\{\widehat{\phi}(L)x_t\right\}\left\{\widehat{\phi}(L)x_t\right\}^{/}\right]^{-1} \sum \left\{\widehat{\phi}(L)x_t\right\}\left\{\widehat{\phi}(L)y_t\right\} \tag{21}$$

and the bias of the $FGLS$ is shown in the appendix to be.

6 OLSRSE, GLS and FGLS in a VAR

When considering the properties of the single equation OLS, GLS and $FGLS$ methodology it is necessary to consider their application to a situation where the variables have been generated from a very general time series process. This will also allow for a detailed assessment of the efficacy of $OLSRSE$ methodology as opposed to $FGLS$ and alternative full information methods. Hence we consider the data generating process of the vector time series process, where $\mathbf{Y}_t$ which is defined from the Wold decomposition as an m dimensional multivariate stochastic process of the form

$$\mathbf{Y}_t = \sum_{j=0}^{\infty} \mathbf{\Psi}_j \boldsymbol{\epsilon}_{t-j},$$

where $\boldsymbol{\epsilon}_t$ is an unobserved process such that $E\left(\boldsymbol{\epsilon}_t\right) = \mathbf{0}$, $E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{/}\right) = \mathbf{\Omega}$ which is an m dimensional, positive semi definite, covariance matrix and $E\left(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s^{/}\right) = \mathbf{0}$ for $t \neq s$. The sequence of Impulse Response (IRs) or Wold Decomposition matrices are defined such that $\mathbf{\Psi}_0 = \mathbf{I}$, and $\mathbf{\Psi}_j$ is a sequence of $m \times m$ matrices of constants. On defining $\mathbf{\Psi}(\mathbf{L}) = \sum_{j=0}^{\infty} \mathbf{\Psi}_j L^j$, the square summability condition

$$\sum_{j=0}^{\infty} \mathbf{\Psi}_j \mathbf{\Omega} \mathbf{\Psi}_j^{/} < \infty$$

is assumed to be satisfied. It is assumed $\boldsymbol{\epsilon}_t$ *is an* m *dimensional ergodic martingale difference sequence, so that* $E(\boldsymbol{\epsilon}_t|\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, ...) = 0$*, and* $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{/}|\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, ...) = \mathbf{\Omega}$ *and its third and fourth moments matrices are finite constants.* The above process can be represented by the finite dimenional $VAR(p)$ system,

$$\mathbf{\Phi}(L)\mathbf{Y}_t = \boldsymbol{\epsilon}_t$$

Or,

$$\mathbf{Y}_t = \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{Y}_{t-j} + \boldsymbol{\epsilon}_t$$

which is clearly non-orthogonalized. In most practical applications of $VARs$ and of IR analysis it is desirable to base analysis on orthogonalized innovations[2]. There is also an issue of identification with the orthogonalized IRs, which are standardized in the sense that the covariance matrix of the innovations are equal to the identity matrix rather than $\mathbf{\Omega}$. Hence an investigator may wish to provide estimates of $\left\{\mathbf{\Psi}_j \mathbf{\Omega}^{-1/2}\right\}_{j=1}^{h}$ rather than $\{\mathbf{\Psi}_j\}_{j=1}^{h}$. Since $\mathbf{\Omega}^{1/2}$ is not unique, then for a given $\mathbf{\Omega}$, it is necessary to provide further identifying assumptions; for example see Inoue and Kilian (2013) and Chapter 4 of Canova (2007) for a discussion. Then $\mathbf{R\Omega R}^{/} = \mathbf{I}$,where $\mathbf{R}$; is an upper diagonal matrix that can be calculated form the eigenvalues of $\mathbf{\Omega}$. The corresponding orthogonalized VAR is then

$$\mathbf{R}\mathbf{Y}_t = \sum_{j=0}^{p} \mathbf{A}_j \mathbf{Y}_{t-j} + \mathbf{u}_t$$

where $\mathbf{R\Phi}_j = \mathbf{A}_j$, for $j = 1, 2, ...p$ and $\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t$. The orthogolalized VAR leads to the contemporaneous values of some of the variables appearing in each equation. For example, a bivariate $VAR(p)$ dgp would lead to the first equation of the orthogonalized VAR being an Autoregressive Distributed Lag (ADL) model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ...\alpha_r y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + ...\beta_q x_{t-p} + u_t \tag{22}$$

so that the presence of the contemporaneous explanatory x variable implies the fact that the disturbances in this ADL are contemporaneously uncorrelated with the disturbances of other equations in the VAR. Clearly, if the lag order p is sufficiently large, then the error term in the ADL will be white noise since it is simply the first equation of the VAR. However, quite commonly in applied work; see for example Stock and Watson (2007), pages 516 through 518; it is asumed that u_t could be autocorrelated so that HAC standard errors are required[3].

The basic situation can be simply illustrated by considering the orthogonalized two dimen-

[2] For simplicity we restrict ourselves to stationary processes; while formulations with I(1) variables could involve lagged error correction terms as in Engle and Granger (1987).

[3] Autoregressive error processes can arise from common factors in the ADL which is well known from Sargan (1963) and Hendry and Mizon (1978).

sional $VAR(1)$ model as the data generating process, where $\mathbf{Y}_t = \begin{pmatrix} y_t & x_t \end{pmatrix}$ and

$$\mathbf{RY}_t = \mathbf{AY}_{t-1} + \mathbf{u}_t$$

where $\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$ and $\mathbf{R} = \begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix}$, so that β is highly nonlinear and is derived from the equation $\mathbf{R\Omega R}^{/} = \mathbf{I}$,so that

$$(\mathbf{R} - \mathbf{A}L) = \begin{pmatrix} 1 - \alpha L & -\beta \\ -a_{21}L & \gamma - \alpha_{22}L \end{pmatrix}$$

Then the orthogonalized two dimensional $VAR(1)$ system is .

$$\begin{pmatrix} \alpha & -\beta \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{23}$$

The $b = \beta/\alpha$ parameter then represents the contemporaneous impact of x on y. Essentially this seems a rather uninteresting quantity in the context of a dynamic time series system; and the calculation of Impulse Responses (IRs) and possibly dynamic multipliers would be considerably more meaningful. Nevertheless, it has become fashionable to estimate the simple static regression

$$y_t = bx_t + z_t \tag{24}$$

and to interpret the slope coefficient as the "instantaneous impact". The limitations of this approach can be seen in the context of the bivariate VAR. Then in the $g = 2$ case it is straightforward to show that

$$\Omega = \begin{pmatrix} \gamma^2 + \beta^2 & \alpha^2\gamma^2 \\ \alpha^2\gamma^2 & \gamma^{-2} \end{pmatrix}.$$

Then the instantaneous impact parameter b is simply $b = \beta/\alpha$ and $b = \omega_{12}/\omega_{22}$ and the simple static regression is

$$y_t = bx_t + z_t \tag{25}$$

the regression disturbance term is

$$z_t = u_{1t} + \alpha_{11} y_{t-1} + \alpha_{12} x_{t-1} \tag{26}$$

Clearly the application of OLS will be inconsistent due to the contemporaneous correlation of the explanatory variable and disturbance term. In particular,

$$E(z_t x_t) = \alpha_{11} E(y_{t-1} x_t) + \alpha_{12} E(x_t x_{t-1}) = \alpha_{11} e_2^{/} \Gamma_1 e_1 + \alpha_{12} e_2^{/} \Gamma_1 e_2 \tag{27}$$

where $\Gamma_j = Cov(\mathbf{Y}_t \mathbf{Y}_{t-j}^{/})$, and $e_j^{/}$ is in general a g dimensional selection vector which is the null vector except for unity in the $j\prime th$ element. Hence the contemporaneous covariation between the error term and the explanatory variable can be expressed in terms of the VAR parameters and the elements of the processes autocovariance matrix; and therefore are easily computed from knowledge of the autocovariances of the VAR. The OLS estimator of b will therefore possess the following asymptotic bias

$$\left(\widehat{b_{OLS}} - b\right) = \left(\Sigma x_t^2\right)^{-1} (\Sigma x_t z_t) = [Var(x_t)]^{-1} E(x_t z_t) \tag{28}$$

and

$$p \lim \left(\widehat{b_{OLS}} - b\right) = \{p \lim [Var(x_t)]\}^{-1} \{p \lim E(x_t z_t)\} \tag{29}$$

which is easily calculated from the population quantities of the VAR for all simulation designs.

7 VAR Simulations

In terms of a simulation design we can set up a $VAR(1)$ data generating process and choose $\beta = 2$ which implies joint endogeneity of the two variables. The choice of the α parameters in the $A(L)$ matrix must imply a stationary $I(0)$ process and the main issue will be the degree of persistence implied for the u_t error process in the static regression. Although there seems to be no readily available interpretation of the β coefficient in a static regression of y on x when the variables are generated by dynamic interactions best summarized by a VAR; there is nevertheless consdierable popularity at reporting this type of regression result. Indeed the nature of many canned packages such as $STATA$ actively promotes and encourages such an approach to applied researchers. For this reason the properties of the estrimated β is one of the main focuses of this study. The $VAR(1)$ model used in the designs is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 1/3 & -1/6 \\ -1/3 & 1/2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{30}$$

and the roots of $|I - A|$ are (1/6) and (2/3), which indicate that the VAR is covariance stationary. Also, $\Omega = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$ and the corresponding $R = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$ which gives the orthogonalized system of

$$\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 1 & -7/6 \\ -1/3 & 1/2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{31}$$

In the following simulations 1,000 replications were generated of the above $VAR(1)$ design, and were implemented in the following estimation strategies:

(i) OLS of both equations of the original non orthogonalized VAR parameters and then calculate $\widehat{\Omega}$ from the estimated autrocovariiance matrix at lag zero of the residuals from both equations. Then estimate β from the estimated Ω parameters. Then compute the bias and standard error of the estimate of β.

(ii) Then use the OLS estimaton to calculate the robust standard errors from the NW, BVK and Hansen and Hodrick (1980) procedures.

(iii) compute $FGLS$ from the OLS residuals from the regression in equation (9) subsequent estimation of sieve ARs for the residuals and then estimation of the $FGLS$.

(iv) Estimation of the system based VAR parameters and determine their theoretical asymptotic MSE from their limiting distributon given in Appendix 1.

(v) Some empirical studies have estimated a static regression model in equation (.) with an $AR(p)$ error term and used $\widehat{b}[1-\widehat{\phi}(L)]^{-1}$.as an estimate of a long run multiplier. In this approach the adjustedments for autocorrelated errors are simply incorporated into the structural form and become Impulse Responses (IRs).

8 Strict Exogeneity

Given the ease of applying $FGLS$ through a sieve AR, the only remaining motivation for the use of the $OLSRSE$ approach would be due to concern about possible breakdowns of the struct exogeneity assumption. This would involve trhe situation where the error term in the regression model in uncorrelated with contemporaneous explanatory variables, but is instead correlated with either lagged and / or future explanatory variables. Hence the regression would be

$$y_t = bx_t + z_t \tag{32}$$

with the valid contemporaneous exogeneity assumption of $E(x_t z_t) = 0$, but a violation of strict exogeneity so that $E(x_{t-r} z_t) \neq 0$, for some ineger. This would be tested by a standard Lagrange Multiplier test which is equivalent to an omitted variable test, and is

$$z_t = c + \gamma_1 x_{t-1} + ... + \gamma_r x_{t-r} + \eta_t,$$

where $E(\eta_t) = 0$, $E(\eta_t^2) = \sigma_\eta^2$ and $E(\eta_t \eta_s) = 0$, for $s \neq t$. Then the Lagrange Multiplier (LM) test is simply $TR^2$of the regression residuals on the variables correlated with the lagged regresion error term. An alternative hypothesis we can take an linear ADL and do not require specifiying both equations of the system. Under strict exogeneity contemporaneous exogeneity is restricted to hold, Hence the hypothesis that x_{t-r} is correlated with z_t under the maintaind hypothesis of strict exogeneity, but is not correlated with the contemporaneous error term, would imply that x_t and x_{t-r} are uncorrelated. So that in the case of $r = 1$, the strict exogeneity assumption would only be appropriate if x_t was a martingale. This fact is quite revealing since the leading examples of violations of strict exogeneity have been due to forward market forecast errors in Hansen and Hodrick (1980), and orange juice futures and number of days of freezing weather in Orlando, due to Roll (1984). Both examples essentially rely on rational expectations forecast errors which are expected to be uncorrelated. However, virtually all other time series variables would be expected to be autocorrelated; which makes the hypothesis of strict exogeneity unrealistic in many situations in economics.

9 Example: Forward Market Efficiency Tests

The original applied problem that motivated the development of the $OLSRSE$ technique appears to have been the study by Hansen and Hodrick (1980) on testing for market efficiency in forward and futures currency markets. Hansen and Hodrick (1980) use a regression where the dependent variable is the difference between a spot exchange rate at time $t + k$ and a k period forward rate at time t. Hence on denoting s_{t+k} as the logarithm of the spot exchange rate at time $t+k$ and f_t as the forward exchange rate of a k period maturuty time at time t, then the theory of market efficiency implies that $E_t s_{t+k} = f_t$. The theory assumes rational expectations and also a constant risk premium. Hansen and Hodrick (1980) consider the case where $k > 1$ so that the frequency of observations exceeds the maturity time of the forward contract. Then $u_{t+k} = (s_{t+k} - f_t)$ will be an $MA(k)$ process under the null hypothesis. Filtered data to implement GLS may possibly violate the strict exogeneity condition since a future error term, or forecast error may be correlated with the regressors which could include laged forecast errors. This issue then prompted Hansen and Hodrick (1980) to use an $OLSRSE$ type procedure.

More recent tests of the unbiasedness theory have tended to use the regression $\Delta s_{t+k} = \alpha + \beta(f_t - s_t) + u_{t+k}$, and to test the theory that $\alpha = 0$, $\beta = 1$ and $E(u_t u_{t+j}) = 0$ for $j > k$. It is worth noting that an alternative literature to test this theory developed that used a system based VAR procedure with tests based on cross equation restrictions on the VAR coefficients; see Hakkio (1981), Baillie, Lippens and McMahon (1983) and Baillie (1989) . This approach essentially avoided the strict exogeneity issue by means of systems modeling. Baillie, Lippens and McMahon (1983) address the identical problem by using VAR and Wald tests of cross equation restrictions. It should also be noted that the "unbiasedness" test on forward markets has been overwhelmingly rejected and research has progressed to modeling time dependent risk premium and other explanations for the failure of uncovered interest rate parity. However, the situation and approach described by Hansen and Hodrick (1980) is an interesting problem in the context of this paper, especially since it was one of the leading original motivations for using the $OLSRSE$ approach. Accordingly, we have estimated the equation

$$(s_{t+k} - f_t) = \alpha + \beta(f_t - s_t) + u_{t+k}$$

for weekly data from January 1985 through October 2015, with the spot rates recorded on Thursdays and the 30 day forward rates recorded on Tuesdays for six major currencies of Australia, Canada, Japan, New Zealand. Switzerland and UK against the numeraire US dollar. The data are from Datastream and consist of $T = 1609$ observations. We accordingly estimate the above equation by four methods;

(i) OLS with regular standard errors,

(ii) by $OLSRSE$,

(iii) $FGLS$ with the error terms modeled by an $MA(4)$ process and

(iv) by $FGLS$ with the error term estimated by a sieve AR process.

The results are given in tables 7 and 8 with the usual result of negative slope coefficients and wide spread of the standard errors of the OLS and $OLSRSE$ methods compared with the $FGLS$ approach.

Since the data is measured weekly, the time of the expectation or prediction would be (22/5), or 4.40 weeks. In terms of daily data there are typically 22 daily innovations. Hence the first order autocorrelation would cover 17 of the 22 innovations in the forecast horizon and hence $\rho_1 = 17/22 = 0.77$. Similarly, $\rho_2 = 0.52$, $\rho_3 = 0.32$, $\rho_4 = 0.09$ and $\rho_k = 0$, for $k \geq 5$. As noted by Baillie and Bollerslev (1990) these autocorrelations are consistent with the following invertible $MA(4)$ process,

$$u_{t+k} = \varepsilon_t - 0.84\varepsilon_{t-1} - 0.77\varepsilon_{t-2} - 0.32\varepsilon_{t-3} - 0.09\varepsilon_{t-4}.$$

where ε_t is white noise. Hence it is also possible to impose these restrictions while estimating by $FGLS$ with the error terms modeled by an $MA(4)$ process and this is to be completed in future version of this paper.

10 Conclusion

It has become commonplace in applied time series econometric work to estimate regressions with consistent, but asymptotically inefficient OLS and to base inference of conditional mean parameters on robust standard errors. This approach seems mainly to have occurred due to concern at the possible violation of strict exogeneity conditions from applying GLS. We first show that even in the case of the violation of contemporaneous exogeneity, that the asymptotic bias associated with GLS will generally be less than that of OLS. This result extends to Feasible GLS where the error process is approximated by a sieve autoregression. The paper also examines the trade offs between asymptotic bias and efficiency related to OLS, feasible GLS and inference based on full system VAR. We also provide simulation evidence and several examples including tests of efficient markets, orange juice futures and weather and a control engineering application of furnace data. The evidence and general conclusion is that the widespread use of OLS with robust standard errors is generally not a good research strategy. Conversely, there is much to recommend $FGLS$ and VAR system based estimation.

11 Appendix 1: VAR Asymptotics

We consider the $VAR(p)$ model $\mathbf{U}_t = \mathbf{C}\mathbf{U}_{t-1} + \mathbf{v}_t$ where $\mathbf{U}_t^{/} = [\mathbf{y}_t, \mathbf{y}_{t-1}, ...\mathbf{y}_{t-r+1},]$. The lag order $r = max(p,q)$, while $\mathbf{v}_t^{/} = [\boldsymbol{\varepsilon}_t, 0,]$, and with null matrices of the appropriate dimension. Then,

$$\mathbf{C} = \begin{bmatrix} \mathbf{\Pi}_1\mathbf{\Pi}_2...\mathbf{\Pi}_{r-1} & \mathbf{\Pi}_r \\ I & 0 \end{bmatrix}$$

Then $\boldsymbol{\theta}^{/}$ =$[d_1, ...d_m, vec(\mathbf{N}^{/}\mathbf{C}), vech(\mathbf{\Omega})]$ and $\mathbf{N}^{/} = [\mathbf{I}, \mathbf{0}]$, which is of dimension m by $2mr$.Then on estimation of the structural parameters $\boldsymbol{\theta}$ by either approximate or full MLE,

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{L} N(\mathbf{0}, \mathbf{V}). \tag{33}$$

where $V = \begin{pmatrix} \Omega \otimes \Gamma_0^{-1} & 0 \\ 0 & 2J^{/}\left(\Omega^{-1} \otimes \Omega^{-1}\right) J \end{pmatrix}$ and $vec(\Omega) = J\omega$. Note that $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$, and the symbol $\xrightarrow{L}$ denotes convergence in distribution.

References

[1] Amemiya, T. 1973. Generalized least squares with an estimated autocovariance matrix. Econometrica, 41, 723-732.

[2] Andrews, D.W.K. 1991. Heteroscedasticity and autocorrelation consistent covariance matricx estimation. Econometrica 59, 817-858.

[3] Baillie, R.T. 1989. Econometric Tests of Rationality and Market Efficiency. Econometric Reviews, 8, 151-186.

[4] Baillie, R.T. and Bollerslev, T. 1990. A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. Journal of International Money and Finance, 9, 309-324.

[5] Baillie, R.T., Kiliç, R., 2006. Do asymmetric and nonlinear adjustments explain the forward premium anomaly? Journal of International Money and Finance 25, 22–47.

[6] Baillie, R.T., Lippens, R.E and P.C. McMahon, 1983. Testing rational expectations and efficiency in the foreign exchange market . Econometrica 51, 553-563.

[7] Baillie, R.T. and Kim, K.-H. 2015. Was it Risk? Or was it fundamentals? Explaining excess currency returns with kernel smoothed regressions. Journal of Empirical Finance, 34, 99-111.

[8] Bollerslev, T. and Wooldridge, J.M. 1992. Quasi-Maximum Likelihood Estimation of Dynamic Models with Time Varying Covariances. Econometric Reviews, ll, 143–172.

[9] Box, G.E.P. and Jenkins, G.M. 1970. Time Series Analysis, Forecasting and Control. San Francisco, Holden Day.

[10] Box, G.E.P. and Tiao, G. 1975. Intervention Analysis with Applications to Economic and Environmental Problems. Journal of the American Statistical Association 70, 70–79.

[11] Canova, F., 2007. Methods for Applied Macroeconomic Research. Princeton University Press.

[12] Davidson, J.E.H., 1994. Stochastic limit Theory, Oxford University Press, Oxford.

[13] den Haan, W.J. and A. Levin, 1997. A practitioners guide to robust covariance matrix estimation. Handbook of Statistics: Robust Inference, (edited by G.S. Maddala and C.R. Rao, 15, 299-342.

[14] Hakkio, C.S., 1981. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis, International Economic Review, ??, ??–853.

[15] Hansen, L.P., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis, Journal of Political Economy, 88, 829–853.

[16] Hayashi, F. and Sims, C.A. 1983. Nearly efficient estimation of time series models with predetermined, but not exogenous instruments. Econometrica, 51, 783-798.

[17] Hodrick, R.J., 1987. The empirical evidence on the efficiency of forward and futures foreign exchange markets. Harwood, London.

[18] Hsieh, D.A., 1983. A heteroskedasticity consistent covariance matrix estimator for tiime series regressions. Computation and analysis of multiple structural change models. Journal of Econometrics 22, 281–290.

[19] Inoue, A. and L. Kilian, 2013. Inference on impulse response functions in structural VAR models. Journal of Econometrics 177, 1-13.

[20] Kapetanios, G. and Z. Psaradakis, 2013. Semiparametric sieve type GLS inference. Econometric Reviews ??, ??.

[21] Kiefer, N.M., T.J. Vogelsang and H. Bunzel, 2000. Simple robust testing of regression hypotheses, Econometrica, 68, 695-714.

[22] Lothian, J.R., Wu, L., 2011. Uncovered interest-rate parity over the past two centuries. Journal of International Money and Finance 30, 448–473.

[23] Roll, R. 1984 Orange juice and weather. American Economic Review, 74, 861-880.

[24] Newey, W.K. and K.D. West, 1987. A simple positive semi definite heteroscedasticity and autocorrelation consistent covariance matix, Econometrica, 55, 703-708.

[25] Newey, W.K. and K.D. West, 1994. Automatic lag selection in covariance matrix estimation, Review of Economic Studies, 61, 634–654.

[26] Pierce, D.A., 1971. Least squares estimation in the regression model with autoregressive moving average errors. Biometrika, 58, 299-312..

[27] Sims, C.A., 2010. But economics is not an experimental science. Journal of Economic Perspectives 24, 59-68.

[28] Zhou, S., 2002. The forward premium anomaly and the trend behavior of the exchange rates. Economic Letters 76, 273–279.

Table 1: Bias and Mean Squared Error (MSE) for OLS, GLS and $FGLS$ for Single Equation Regression with $AR(1)$ Errors

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	0.2723	0.1232	0.2716	0.0993	0.2677	0.0831	0.2645	0.0820
FGLS	0.1656	0.0315	0.1652	0.0293	0.1644	0.0280	0.1629	0.0273
GLS	0.1649	0.0312	0.1648	0.0292	0.1642	0.0280	0.1628	0.0273

Key: This table reports results from a simulation with 1,000 replications. The data is generated from the model $y_t = \beta x_t + u_t$; where $x_t = \rho x_{t-1} + \epsilon_{x,t}$ and $u_t = \phi u_{t-1} + \epsilon_{u,t}$. The innovations $\epsilon_{x,t}$ and $\epsilon_{u,t}$ are generated from a bivariate $NID(0, V)$ process where $V = \begin{pmatrix} \sigma_x^2 & \sigma_{x,\epsilon}^2 \\ \sigma_{x,\epsilon}^2 & \sigma_\epsilon^2 \end{pmatrix}$. In all experiments $\beta = 2, \rho = 0.5, \phi = 0.9$; and $V = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$.

.

Table 2: Bias and Mean Squared Error (MSE) for OLS, GLS and $FGLS$ for Single Equation Regression with $AR(1)$ Errors

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	0.6838	0.4930	0.6828	0.4902	0.6805	0.4752	0.6707	0.4744
FGLS	0.4129	0.1736	0.4127	0.1710	0.4121	0.1709	0.4116	0.1706
GLS	0.4124	0.1722	0.4117	0.1707	0.4113	0.1703	0.4108	0.1703

Key: This table reports results from a simulation with 1,000 replications. The data is generated from the model $y_t = \beta x_t + u_t$; where $x_t = \rho x_{t-1} + \epsilon_{x,t}$ and $u_t = \phi u_{t-1} + \epsilon_{u,t}$. The innovations $\epsilon_{x,t}$ and $\epsilon_{u,t}$ are generated from a bivariate $NID(0,V)$ process where $V = \begin{pmatrix} \sigma_x^2 & \sigma_{x,\epsilon}^2 \\ \sigma_{x,\epsilon}^2 & \sigma_\epsilon^2 \end{pmatrix}$. In all experiments $\beta = 2, \rho = 0.5, \phi = 0.9$; and $V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

.

Table 3: Bias and Mean Squared Error (MSE) for OLS, GLS and $FGLS$ for Single Equation Regression with $AR(3)$ Errors.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	0.1621	0.0347	0.1596	0.0287	0.1582	0.0281	0.1561	0.0280
FGLS	0.1193	0.0173	0.1184	0.0152	0.1173	0.0145	0.1171	0.0144
GLS	0.1188	0.0172	0.1182	0.0151	0.1170	0.0145	0.1169	0.0144

Key: This table reports results from a simulation with 1,000 replications. The data is generated from the model $y_t = \beta x_t + u_t$; where $x_t = 0.50x_{t-1} - 0.56x_{t-2} - 0.08x_{t-3} + \epsilon_{x,t}$ and $u_t = 0.8800u_{t-1} - 0.8385u_{t-2} + 0.7220u_{t-3} + \epsilon_{u,t}$. The innovations $\epsilon_{x,t}$ and $\epsilon_{u,t}$ are generated from a bivariate $NID(0, V)$ process where $V = \begin{pmatrix} \sigma_x^2 & \sigma_{x,\epsilon}^2 \\ \sigma_{x,\epsilon}^2 & \sigma_\epsilon^2 \end{pmatrix}$. In all experiments $\beta = 2$ and $V = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$.

.

Table 4: Bias and Mean Squared Error (MSE) for OLS, GLS and $FGLS$ for Single Equation Regression with $AR(3)$ Errors.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	0.4099	0.1753	0.4082	0.1723	0.4053	0.1661	0.4005	0.1625
FGLS	0.3005	0.0927	0.2999	0.0911	0.2967	0.0885	0.2948	0.0875
GLS	0.2989	0.0917	0.2987	0.0905	0.2961	0.0881	0.2942	0.0871

Key: This table reports results from a simulation with 1,000 replications. The data is generated from the model $y_t = \beta x_t + u_t$; where $x_t = 0.50x_{t-1} - 0.56x_{t-2} - 0.08x_{t-3} + \epsilon_{x,t}$ and $u_t = 0.8800u_{t-1} - 0.8385u_{t-2} + 0.7220u_{t-3} + \epsilon_{u,t}$. The innovations $\epsilon_{x,t}$ and $\epsilon_{u,t}$ are generated from a bivariate $NID(0, V)$ process where $V = \begin{pmatrix} \sigma^2_x & \sigma^2_{x,\epsilon} \\ \sigma^2_{x,\epsilon} & \sigma^2_\epsilon \end{pmatrix}$. In all experiments $\beta = 2$ and $V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

.

Table 5: Bias and Mean Squared Error (MSE) for OLS, $FGLS$ and VAR for Single Equation Regression $y_t = \beta x_t + z_t$, where data is generated by a $VAR(1)$ process.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	-0.8402	0.7138	-0.8395	0.7085	-0.8384	0.7077	-0.8331	0.7046
OLS-NW	-0.8402	0.7117	-0.8395	0.7075	-0.8384	0.7063	-0.8331	0.7042
OLS-Andrews	-0.8402	0.7114	-0.8395	0.7075	-0.8384	0.7063	-0.8331	0.7042
FGLS	-0.4497	0.2047	-0.4437	0.1981	-0.4385	0.1929	-0.4371	0.1915
VAR(1)	-0.0014	0.0050	0.0008	0.0026	0.0007	0.0013	0.0003	0.0009

Key: See text in section for a full description of the VAR data generating process.

.

Table 6: Bias and Mean Squared Error (MSE) for OLS, $FGLS$ and VAR for Single Equation Regression $y_t = \beta x_t + z_t$, where data is generated by a $VAR(3)$ process.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	0.2542	0.0978	0.2535	0.0779	0.2519	0.0744	0.2465	0.0727
OLS-NW	0.2542	0.0894	0.2535	0.0777	0.2519	0.0718	0.2465	0.0708
OLS-Andrews	0.2542	0.0894	0.2535	0.0777	0.2519	0.0718	0.2465	0.0709
FGLS	-0.1453	0.0366	-0.1444	0.0311	-0.1442	0.0261	-0.1285	0.0250
VAR(3)	0.0017	0.0051	-0.0010	0.0024	0.0008	0.0013	0.0006	0.0009

Key: See text in section for a full description of the VAR data generating process.

.

Table 7: Estimates of the coefficient in the Hansen-Hodrick (1980) model.

	OLS	SE	SE-NW	SE-Andrews	FGLS	BIC-lag
AUD	-1.0245	0.0733	0.0860	0.0855	-0.8469	13
CAD	-1.0416	0.0699	0.1128	0.1131	-0.8975	13
CHF	-1.1161	0.0788	0.0922	0.0922	-0.8339	13
GBP	-1.1891	0.0809	0.1071	0.1057	-0.8702	14
JPY	-1.1752	0.0782	0.0870	0.0866	-0.8339	13
NZD	-1.0848	0.0687	0.0845	0.0842	-0.8823	14

.

Table 8: Estimates of the coefficient in the forward premium regression.

	OLS	SE	SE-NW	SE-Andrews	FGLS	BIC-lag
AUD	-0.0245	0.0733	0.0860	0.0855	0.1531	13
CAD	-0.0416	0.0699	0.1128	0.1131	0.1025	13
CHF	-0.1161	0.0788	0.0922	0.0922	0.1661	13
GBP	-0.1891	0.0809	0.1071	0.1057	0.1298	14
JPY	-0.1752	0.0782	0.0870	0.0866	0.1661	13
NZD	-0.0848	0.0687	0.0845	0.0842	0.1177	14