Is Robust Inference with OLS Sensible in Time Series Regressions? Comparisons 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

September 27, 2016

Abstract

The practice of estimating static conditional mean functions from time series data and then using "robust" covariance matrices for inference is examined against alternative approaches. We show that when contemporaneous exogeneity is violated, that the asymptotic bias associated with GLS can actually be less than that of OLS. This result extends to Feasible GLS where the error process is approximated by a sieve autoregression. The impact of violations of strict exogeneity is also examined by similar approaches and Feasible GLS again out performs robust OLS procedures. The paper provides three detailed empirical examples based on (i) multi period tests of rational expectations and market efficiency, (ii) estimates of the Taylor Rule for monetary policy and (iii) the effect of weather condtions on the OJ futures market returns. We conclude that Feasible GLS generally performs extremely well compared to OLS with robust standard errors; while there is also much to recommend Autoregressive Distributed Lag (ADL) approaches in situations where underlying exogeneity conditions are suspect.

JEL Classification: C22, C31. *Keywords*: OLS, GLS, Feasible GLS, asymptotic bias, robust inference, strict exogeneity.Taylor Rule, tests of rational expectations over multi horizons.

1 Introduction

Microeconometric studies frequently specify a conditional expectation function that has clear covariates relating to treatment effects and policy variables. This tends to give rise to relative confidence in the inclusion of explanatory variables and leads to a model specification such as

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

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. It is quite common for such studies to assume that

$$E(u_t|x_t) = 0 \tag{2}$$

and to then obtain OLS estimates with robust standard errors to adjust for model mis-specification. This often requires heroic assumptions such as $NIID$ disturbances.

The transference of this approach to situations with time series data would generally seem to be inappropriate. In particular, the impact of neglected variables is problematic and time series data typically involves dynamic dependencies between the variables that invalidates the assumption in equation (2). Also the specification of many economic models involves adjustment mechanisms or expectations of variables which cannot be handled in the static regression framework. Hence estimation of a static type regression, or "instantaneous" regression as equation (1) is sometimes known, will invariably result in autocorrelated residuals which often represent omission of important dynamics from the model.

It should be noted that the above static single equation approach is in stark contrast to traditional time series econometric approaches which generally involve some form of model building with diagnostic testing for mis-specifation. Ideally the model building strategy will involve statistical diagnostics alongside the use of economic theory. Also the traditional single equation ADL or multiple equation VAR type approach can deliver important economic information on dynamic multipliers and Impulse Responses (IR). Clearly the static single equation approach in equations 1 and 2 is incapable of being used for such policy analysis. Nor will the static instaneous regression be appropriate for either ex ante, or ex post prediction where modeling dynamics is essential for calculating minimum MSE predictors.

It is an interesting issue as to the interpretation of an estimated version of equation (1) when it is estimated by OLS and found to have autocorrelated errors. Many supposed static economic relationships have hidden dynamics and forms of mis specificaton which only become apparent when more complex models compared with equation (1) are considered. We consider these issues later in this paper. For example, on making the standard assumption for a vector of covariance stationary time series process $\{y_t : x_t\}$ having a vector Wold decomposition; then the coefficients in equation (1) will generally not have an economically meaningful interpretation.

However, despite the above comments the implementation of OLS and subsequent application of robust standard; i.e. to use $OLSRSE$ inference, has become more common in work with time series data. Clearly the use of $OLSRSE$ bypasses traditional model building methodology and use of diagnostic testing in the formulation of an econometric model. Hence an unfortunate divide appears to have arisen between time series econometric analysts and those wanting to specify static economic relationships between variables. We try to address these issues in this paper.

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.

It should be noted that the $OLSRSE$ approach is sometimes justified in terms of (i) the difficulty of applying GLS and (ii) possible violation of strict exogeneity when applying GLS. The first justification in terms of difficulties with GLS, seems not to be valid given the relative ease of applying various Feasible GLS, or $FGLS$ procedures following work of Amemiya (1973) and particularly sieve autoregressions of Grenander (1981) and recently extended by Kapetanios and Psaradakis (2015). Also, under stronger assumptions of independent, identically distributed Guassian innovations, the properties of estimating regressions with $ARMA$ disturbances has been derived by Pierce (1971) and frequently implemented in applied literature[1].

[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. Another form of robust procedure is the $QMLE$ of Bollerslev and Wooldridge (1991), which provides robust standard errors for certain types of departures from conditional Gaussianity when a standard Normal density has been maximized and it is assumed there is a correct specification of the first two conditional moments.

The plan of this paper is the following: sections 2 through 6 explore some of the technical issues associated with the $OLSRSE$ methodology. In particular the next section summarizes some of the basic statistical issues involved and the simplest case when all the exogeneity requirements are satisfed, so that $OLSRSE$ is generally inefficient compared with GLS and versions of $FGLS$. Section 3 analyzes the more interesting, and practically relevant case when the assumption of concerning the application of OLS and GLS when the condition of contemporaneous exogeneity does not hold; but when strict exogeneity is maintained. Then, clearly both OLS and $FGLS$ will be inconsistent. However, we show quite surprisingly that the asymptotic bias associated with GLS is substantially less than that of OLS for some regions of the parameter space. This result hinges on the relative persistence of the processes generating the explanatory variables vis a vis the error proces dynamics in the single equation process. These results suggest that $FGLS$ is often the preferred single equation estimation and inferential procedure compared with $OLSRSE$ even when the assumption of contemporaneous exogeneity has been violated.

Section 4 then examines the application of Feasible GLS, or $FGLS$ when sieve Autoregressive (AR) models are used to approximate the error process dynamics. Using results of Kapetanios and Psaradakis (2015) shows how the use of information criteria, particularly BIC can lead to consistent model selection and a simple automatic procedure for calculating $FGLS$ in practice. Section 5 then provides simulation evidence on the performance of OLS, GLS and $FGLS$ and also a comparison of different methods for computing the robust standard errors. Section 6 then considers the use of the same procedures when the data generating process is a more general covariance stationary multivariate time series having Wold decomposition, which allows an infinite order VAR representation. However, generally more model based strategies of estimating $ADLs$ and multiple equation $VARs$ are likely to deliver improved results in terms of more efficient parameter estimates which can then be used for deriving estimated dynamic multipliers and impulse responses.

The next section of the paper addresses the situation where either weak or strict exogeneity restrictions are violated. We compare the $OLSRSE$ and $FGLS$ approaches. Clearly a situation with a breakdown of strict exogeneity can be most easily handled within a simultaneous equation model, which requires specifying further equation(s). Interestingly, even in the presence of violations of strict exogeneity, $FGLS$ often dominates $OLSRSE$ in terms of asymptotic bias and efficiency within the single equation estimation framework. Although, if possible, full system estimation is generally the preferred strategy.

The next three sections of this paper with three major empirical examples in the asset pricing, financial economic and macroeconomic literature. Each one involves clear choice of modeling strategies The first one in section 7 concerns tests of rational expectations and market efficiency where the sampling interval of the data exceeds the period of the forward contract. The seminal paper in this area was by Hansen and Hodrick (1980) who also developed the first estimate of the OLS covariance matrix and hence the first $OLSRSE$ type procedure. We analyze three separate formulations of this hypothesis including the well known Fama regression where the spot returns over k periods are regressed on the contemporaneous forward premium. The model also implies a particular $MA(k)$ autocorrelation structure to the errors. We compare the previously outlined alternative econometric methodologies and find favorable evicedence for the $FGLS$ approach. Section 8 deals with the Taylor Rule which was originally specified as a static instantaneous type regression involving nominal interest rates, inflation around its target level and finally the output gap. The static Taylor Rule regression has considerable residual autocorrelation and we contrast $OLSRSE$, $FGLS$ and more complex ADL together with dynamic multipliers and IRs from a VAR. The results indciate the superiority of going to the more complex dynamic model. If the investigator insists in staying in the single eqaution mode, then $FGLS$ appears to have substantial gains in estimated parameter efficiency compared with $OLSRSE$. The final empirical example concerns the OJ futures market and weather conditions.

The paper finally has a brief conclusions section which summarizes the results and offers some recommendations for the development of future time series regression work. In particular, we advise that robust standard error procedures should be used with considerable caution and that standard computer packages should also include the easily computed $FGLS$ procedure based on sieve autoregressions and ADL and VAR modeling of the conditional mean function[2].

[2] 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.

2 OLSRSE with Contemporaneous Exogeneity

Consider the basic linear model

$$y_t = \beta^{'} x_t + u_t \tag{3}$$

where there are stricly exogenous variables so that $E(u_t|x_t) = 0$, $E(u_t|x_{t-j}) = 0$, for all $j = ... - 2, -1, 0, 1, 2....$. In matrix terms $Y = X\beta + u$, where $E(uu^{/}) = \Omega$, which is symmetric and positive definite. On making the additional assumptions that 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 $Q = p\lim T^{-1} \sum_{j=1}(x_j x_j^{/})$ and $u_t x_t$ is a martingale difference sequence. Then the OLS estimator will have the properties

$$T^{1/2}(\hat{\beta} - \beta_0) \to N(0, Q^{-1}\Omega Q^{-1}) \tag{4}$$

An important issue is to construct an estimate of Ω from using minimal assumptions. The application of GLS estimation requires filtering of the observed time series which can lead to a violation of the exogeniety requirements and hence lead to inconsistent estimates of the regression parameters. The corrollary suggested by several authors is to use consistent, but asymptotically inefficient OLS estimates and robustify 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) which uses a simple Bartlett window, or kernel, led to a standard method for estimating the error covariance matrix and hence of computing $OLSRSE$ method. The heteroskedastic and autocorrelation consistent, or HAC, estimator of its asymptotic covariance matrix Ω , is estimated by the method of Newey and West (1987), or Andrews (1991). The usual Newey West estimate of Ω is ,

$$\widehat{\Omega} = \left(T^{-1} \sum_{t=1} x_t x_t' \hat{u}_t^2 + \sum_{j=1} \{1 - j/(q+1)\}(\mathbf{\Gamma}_j + \mathbf{\Gamma}_{-j}) \right) \tag{5}$$

where

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

so that all forms of contemporaneous, weak and strict exogeneity hold. den Haan and Levin (1997) report large size distortions on subsequent inference. The above robust, or HAC consistent estimator of 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 the 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.

The clear alternative to the above method is to use some form of Generalized Least Squares (GLS) estimator which exploits forms of dependency in the error process to calculate a more efficient estimator of the regression parameters . There is a long history on the application of feasible GLS dating back to Hannan (1962) and Amemiya (1973). If an $ARMA$ structure is known for the errors then it is well known both how to compute and the asymptotic properties of GLS which will be approximate MLE under Gaussianity; see Pierce (1971). In the usual case of an unknown error process then Amemiya (1973) has proposed approximating the error process by an autoregression where the order of the process grows slowly with the sample size to compute a Feasible GLS estimator of the regression parameters β. Grenander (1981) refers to this type of non parametric method as a sieve autoregression and the whole procedure as sieve GLS. In the presence of non stochastic regressors and when the errors are generated by a linear process with i.i.d. innovations; then such an $FGLS$ type estimation procedure has been shown by Amemiya (1973) to be asymptotically Normal with asymptotic Gauss Markov efficient. Kapetanios and Psaradakis (2015) have extended Amemiya's results to situations where both the regressors and errors have short range dependency with absolutuely summable and finite autocovariances; and also for the possibility that the order of the approxomating sieve AR is data dependent. We build on these results in this paper.

3 OLS and GLS without Contemporaneous 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

$$y_t = \beta x_t + u_t \tag{6}$$

and

$$u_t = \phi u_{t-1} + \varepsilon_{u,t} \tag{7}$$

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

$$(y_t - \phi y_{t-1}) = \beta(x_t - \phi x_{t-1}) + (u_t - \phi u_{t-1}) \tag{8}$$

The consistency of GLS will now require that

$$E(x_t - \phi x_{t-1})(u_t - \phi u_{t-1}) = 0 \tag{9}$$

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 condition is most generally known as weak exogeneity and the fourth condition is a condition for strict exogeneity. Hence the last two conditions imply that the errors are uncorrelated with both past and future regressors. All four 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

$$x_t = \rho x_{t-1} + \varepsilon_{x,t} \tag{10}$$

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 straightforward to show that

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

and

$$p\lim\left(\widehat{\beta_{OLS}} - \beta\right) = \left(\frac{1-\rho^2}{\sigma_x^2}\right)\left(\frac{\sigma_{ux}^2}{1-\rho\phi}\right) \tag{11}$$

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

$$p\lim\left(\widehat{\beta_{GLS}} - \beta\right) = \left(\frac{1-\rho^2}{\sigma_x^2}\right)\left(\frac{\sigma_{ux}^2}{1+\phi^2-2\rho\phi}\right) \tag{12}$$

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

$$\frac{p\lim\left(\widehat{\beta_{OLS}} - \beta\right)}{p\lim\left(\widehat{\beta_{GLS}} - \beta\right)} = \left(\frac{1-\rho\phi+\phi(\phi-\rho)}{1-\rho\phi}\right) \tag{13}$$

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. Figure 1 shows the ratio of biases for $0 < \rho < 1$ and $0 < \phi < 1$. 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.

It can also be shown that

$$MSE\left(\widehat{\beta_{OLS}}\right) = \left(\frac{\sigma_u^2}{1-\phi^2}\right)\left(\frac{1+\rho\phi}{1-\rho\phi}\right) \tag{14}$$

while

$$MSE\left(\widehat{\beta_{GLS}}\right) = \left(\frac{\sigma_x^2}{1-\rho^2}\right)\left(1+\phi^2-2\rho\phi\right) \tag{15}$$

is somewhat harder to obtain analytically. The model can be readily generalized to the following case with $AR(k)$ errors:

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

and the explanatory variable

$$x_t = \rho x_{t-1} + \varepsilon_{x,t} \tag{17}$$

and

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_k u_{t-k} + \epsilon_{u,t} \tag{18}$$

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 $s \neq t$. It is shown in Appendix 1 that

$$plim\left(\hat{\beta}_{OLS} - \beta\right) = \frac{\mathbb{E}(x_t u_t)}{\mathbb{E}(x_t^2)} = \frac{\sigma_{ux}^2}{1 - \phi_1\rho - \phi_2\rho^2 - \cdots - \phi_k\rho^k}\left(\frac{1-\rho^2}{\sigma_x^2}\right) \tag{19}$$

and the corresponding bias for GLS is

$$plim\left(\hat{\beta}_{GLS} - \beta\right) = \left(\frac{1-\rho^2}{\sigma_x^2}\right)\frac{\sigma_{ux}^2}{1 + \sum_{i=1}^k \phi_i^2 - 2\sum_{i=1}^k \phi_i\rho^i + 2\sum_{i<j}^k \sum^k \phi_i\phi_j\rho^{j-i}} \tag{20}$$

and hence the relative bias of OLS vis a vis GLS is

$$\frac{plim\left(\hat{\beta}_{OLS} - \beta\right)}{plim\left(\hat{\beta}_{GLS} - \beta\right)} = \frac{1 + \sum_{i=1}^k \phi_i^2 - 2\sum_{i=1}^k \phi_i\rho^i + 2\sum_{i<j}^k \sum^k \phi_i\phi_j\rho^{j-i}}{1 - \phi_1\rho - \phi_2\rho^2 - \cdots - \phi_k\rho^k} \tag{21}$$

Tables 1 through 4 provide simulation results for the biases and MSE for some particular cases of the above model. The results for the sample size of $T = 1,000$ are also identical within two decimal places to the theoretical asymptotic results and are discussed in detail in section 5.

4 Feasible GLS with Minimal Assumptions

In our earlier example with only first order dynamics in the explanatory variable and error process; i.e. equations (6), (7) and (10); then $FGLS$ amounts to being the simple Cochrane

Orcut method and the $FGLS$ estimator ist

$$\widehat{\beta}_{FGLS} = \left\{ \sum \left(x_t - \widehat{\phi} x_{t-1} \right)^2 \right\}^{-1} \left\{ \sum \left(x_t - \widehat{\phi} x_{t-1} \right) \left(y_t - \widehat{\phi} y_{t-1} \right) \right\} \tag{22}$$

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 we now consider the more realistic situation of using a sieve AR approximation to the unknown error process of $\widehat{u}_t$.

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

$$y_t = \beta' x_t + u_t \tag{23}$$

The GLS estimator is

$$\widehat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \tag{24}$$

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

$$\widehat{z}_t = y_t - \widehat{b_{OLS}} x_t \tag{25}$$

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

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

It is necesary to define the asymptotic properties of the $FGLS$ which covers weak stationarity of the errors and regressors. Following Kapetanios and Psaradakis (2013), we make the following assumptions:

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.

Assumption 3: The preliminary estimator of the regression coefficients $\widetilde{\beta} = \beta + O_p(T^{-1/2})$ as $T \to \infty$, which is very mild and is satisfied by OLS and other estimators.

Assumption 4: $h = h_T \to \infty$ and $h_T = o_p[T/\ln(T)]^{1/4}$ as $T \to \infty$ and is consistent with u_t being represented by a finite dimensional AR process. Then from Kapetanios and Psaradakis (2013), we have

Theorem: Under the above assumptions; $T^{1/2}(\widehat{\beta} - \widetilde{\beta}) = o_p(1)$ as $T \to \infty$, and

$$T^{1/2}(\widehat{\beta} - \beta) \to N(0, V)$$

where $V = plim(T^{-1}X^{/}\Omega X)^{-1}$. The $FGLS$ estimator is therefore based on fitting a sieve $AR(p)$ model to the OLS residuals and several possible information criteria can be used to determine the order of the AR process. In general, the information criteria are of the form

$$ln(\widehat{\sigma}^2_{u,h}) + hC_T T^{-1}$$

where C_T is a particular information criteria and in the case of this paper, we use the Schwarz, BIC which is $C_T = \ln(T)$. Then the form of the $FGLS$ is

$$\widehat{\beta_{FGLS}} = \left[\sum \left\{\widehat{\phi}(L)x_t\right\}\left\{\widehat{\phi}(L)x_t\right\}^{/}\right]^{-1} \sum \left\{\widehat{\phi}(L)x_t\right\}\left\{\widehat{\phi}(L)y_t\right\} \quad (26)$$

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 Comparisons of OLSRSE and FGLS for the Single Equation Model

This section discusses the results of a simulation study to evaluate the bias and MSE of different forms of the estimators in the regression model with $AR(p)$ errors and $AR(p)$ process generating the explanatory variable. We have one explanatory variable; so the the simulation design is

$$y_t = \beta x_t + u_t \tag{27}$$

$$\phi(L) u_t = \epsilon_{u,t} \tag{28}$$

$$\rho(L) x_t = \epsilon_{x,t} \tag{29}$$

where $\phi(L) = \left(1 - \phi_1 L - ... - \phi_p L^p\right)$, with all its roots outside the unit circle, and $E(\varepsilon_{u,t}) = 0$, $E(\varepsilon^2_{u,t}) = \sigma^2_u$ and $E(\varepsilon_{u,t}\varepsilon_{u,s}) = 0$ for $s \neq t$. Furthermore $\rho(L) = \left(1 - \rho_1 L - ... - \rho_q L^q\right)$, also with all its roots outside the unit circle and $E(\varepsilon_{x,t}) = 0$, $E(\varepsilon^2_{x,t}) = \sigma^2_x$ and $E(\varepsilon_{x,t}\varepsilon_{x,s}) = 0$ 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

$$V = \begin{pmatrix} \sigma^2_x & \sigma^2_{x,\epsilon} \\ \sigma^2_{x,\epsilon} & \sigma^2_\epsilon \end{pmatrix}.$$

The results in Table 1 are for the design with $AR(1)$ processes for both the error process and explanatory variable with the parameters being $\beta = 2, \rho = 0.5, \phi = 0.9$; so that there is considerably more persistence in the error process than the explanatory variable. The first panel has the contemporaneous covariance of $\sigma^2_{x,\epsilon} = 0.2$. The results in Table 1 confirm the theoretical results in section 2 with a considerable increase in efficiency and hence reduction in MSE of the estimate of β from using GLS compared with OLS. 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. Table 2 reports results for the situation with $\sigma^2_{x,\epsilon} = 0.5$; and as expected as the degree of violation of contemporaneous exogeneity increaes, then the asymptotic bias of both OLS and GLS increases. However, the reduction in MSE from using GLS and $FGLS$ as opposed to OLS is still dramatic, but not as much as before.

Further simulations in table 3 are for when explanatory variable process and the error process are both $AR(3)$ processes. The designs used are $\phi(L) = (1 - 0.50L + .56L^2 - .08L^3)$ 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.

6 OLSRSE, GLS and FGLS when data are from a VAR

It is instructive to ask the question as to the origin of the desire to estimate the “instantaneous" regression in equation (1) from time series data. For example, an investigator may be interested in the estimation of a single firm’s Cobb Douglas production function from time series data. Simple economic theory posits a static relationship when the log of output is regressed on the log of labor input and the log of capital input. There can be many different interpretations if the estimation of the instantaneous regression finds autocorrelated errors. The obvious possibilities are due to the presence of important omitted variables and / or the impact of dynamic adjustments from previous time periods. There is also the possibility that the regression parameters measuring the output elasticities with respect to labor and capital may conceivably change over time; which makes the appropriate model specification considerably more challenging. Furthermore, there are the distinct possibility that there can be violations of strict exogeneity with labor and capital possibly being related to lagged output. Such a situation could arguably be most easily handled by system estimation through a VAR.

Many other examples similar to the above can easily be given. However, it seems clear that the most appropriate maintained hypothesis is that for a covariance stationary process $\{y_t : x_t\}$, satisfies a multivariate Wold decomposition which can be approximated by a VAR. This will generally involve complex leading and lagging relationships between dynamic variables; in which case the static regression does not capture any impact, interim, or long run multiplier, and seems a genuinely uninteresting quantity.

As an illustration of some of the issues involved we consider the unemployment - minimum wage rate type equation. In the following y_t denotes employment or unemployment and x_t represents some measurement of wage rates. We are interested in estimating the conditional expectation $E(y_t \mid x_t, x_{t-1}, ...)$. One method has been to estimate the regression

$$y_t = \sum_{j=0}^{k} \beta_j x_{t-j} + u_t, \tag{30}$$

where u_t is $I(0)$ but is probably autocorrelated. The β_j can be interpreted as dynamic multipliers; regardless of whether or not u_t is autocorrelated. However, there is the possible issue of violation of $E(u_t x_t) = 0$, so that contemporaneous exogeneity is called into question. Although it can be argued that changes in the "policy variable" of minimum wage rate, should be causally prior to changes in unemployment provided the data sampling frequency is sufficiently fine. There is also the possibility that lagged unemployment should be included in the equation. The more general approach would be to jointly model the unemployment - wage rate relationship. Usually unemployment and wage rates are considered to be $I(0)$ processes. Then we can consider the vector time series process $\mathbf{Y}_t$ where $\mathbf{Y}_t = \{y_t : x_t\}$ and is defined from the Wold decomposition as a two dimensional multivariate stochastic process. In the general case $\mathbf{Y}_t$ is defined from the Wold decomposition as an m dimensional multivariate stochastic process of the form

$$\mathbf{Y}_t = \sum_{j=0}^{\infty} \mathbf{\Psi}_j \boldsymbol{\epsilon}_{t-j}, \tag{31}$$

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,

$$\mathbf{\Phi}(L)\mathbf{Y}_t = \boldsymbol{\epsilon}_t, \tag{32}$$

Or,

$$\mathbf{Y}_t = \sum_{j=1}^{p} \mathbf{\Phi}_j \mathbf{Y}_{t-j} + \boldsymbol{\epsilon}_t, \tag{33}$$

which is clearly non-orthogonalized. In most practical applications of $VARs$ and of IR analysis it is desirable to base analysis on orthogonalized innovations[3]. 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 $\mathbf{R}\mathbf{\Omega}\mathbf{R}^{/} = \mathbf{I}$,where $\mathbf{R}$; is an upper diagonal matrix that can be calculated form the eigenvalues of $\mathbf{\Omega}$. The corresponding orthogonalized VAR is then

$$\mathbf{R}\mathbf{Y}_t = \sum_{j=0}^{p} \mathbf{A}_j \mathbf{Y}_{t-j} + \mathbf{u}_t, \tag{34}$$

where $\mathbf{R}\mathbf{\Phi}_j = \mathbf{A}_j$, for $j = 1, 2, ...p$ and $\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t$. The orthogonalized 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

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ...\alpha_r y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + ...\beta_q x_{t-p} + \zeta_t \tag{35}$$

[3] 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).

so that the presence of the contemporaneous explanatory x variable implies the fact that the disturbances ζ_t in this ADL are contemporaneously uncorrelated with the disturbances of other equations in the VAR.

To return to the previous illustration of the unemployment - wage rate equation; the probable causal priority of wages vis a vis unemployment then restricts the ordering of the variables in $\mathbf{Y}_t$ for impulse response analysis. The orthogonalized VAR will then have current or contemporaneous wages in the unemployment equation. On writing the ADL as

$$\alpha(L)y_t = \beta(L)x_t + \zeta_t, \tag{36}$$

which can be interpreted as the regression of unemployment on lagged wages and also lagged unemployment, but with white noise errors ζ_t in contrast to autocorrelated errors u_t in the original unemployment - wage rate model. The solved out solution of the ADL is

$$y_t = \alpha(L)^{-1}\beta(L)x_t + \alpha(L)^{-1}\zeta_t, \tag{37}$$

hence $\gamma(L) = \alpha(L)^{-1}\beta(L)$ are the dynamic multipliers from current and lagged wages to unemployment and the signal or structural part of the model. While $\alpha(L)^{-1}\zeta_t$ are the autocorrelated noise terms which represent the effect of current and lagged innovations of unemployment conditioned on current and past wage rates. Practical applications are generally either concerned with the above dynamic multipliers and / or the regular pure IRs from the orthogonalized VAR[4].

The basic situation can be simply illustrated by considering the orthogonalized two dimensional $VAR(1)$ model as the data generating process, where $\mathbf{Y}_t = \{y_t : x_t\}$, and

$$\mathbf{R}\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{u}_t$$

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

[4]Sometimes in applied work it is assumed that ζ_t could be autocorrelated; and $AR(p)$ error processes can arise from common factors in the ADL. However, this modeling aspect is generally neglected in modern applied work and simply replaced with higher order lags in ADL. See Stock and Watson (2007), pages 516 through 518 for an alternative viewpoint.

the equation $\mathbf{R}\Omega\mathbf{R}^{/}=\mathbf{I}$,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 .

$$\begin{pmatrix} \alpha & -\beta \\ 0 & \gamma \end{pmatrix}\begin{pmatrix} y_t \\ x_t \end{pmatrix}=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix}+\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{38}$$

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

$$y_t=bx_t+z_t \tag{39}$$

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

$$y_t=bx_t+z_t \tag{40}$$

the regression disturbance term is

$$z_t=u_{1t}+\alpha_{11}y_{t-1}+\alpha_{12}x_{t-1} \tag{41}$$

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

$$E(z_tx_t)=\alpha_{11}E(y_{t-1}x_t)+\alpha_{12}E(x_tx_{t-1})=\alpha_{11}e_2^{/}\Gamma_1e_1+\alpha_{12}e_2^{/}\Gamma_1e_2 \tag{42}$$

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

$$\left(\widehat{b_{OLS}} - b\right) = \left(\Sigma x_t^2\right)^{-1}\left(\Sigma x_t z_t\right) = \left[Var(x_t)\right]^{-1}E(x_t z_t) \tag{43}$$

and

$$p\lim\left(\widehat{b_{OLS}} - b\right) = \left\{p\lim\left[Var(x_t)\right]\right\}^{-1}\left\{p\lim E(x_t z_t)\right\} \tag{44}$$

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 1/3 & -1/6 \\ -1/3 & 1/2 \end{pmatrix}\begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{45}$$

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

$$\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 1 & -7/6 \\ -1/3 & 1/2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{46}$$

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 was applied to both equations of the original non orthogonalized VAR and the matrix $\widehat{\Omega}$ was estimated from the autocovariance matrix at lag zero of the residuals from both equations. The "instantaneous regression" parameter β was then estimated from the estimate of Ω. The bias and standard error of the estimate of β.

(ii) The above OLS estimates were used to calculate the robust standard errors from the NW, BVK and Hansen and Hodrick (1980) procedures.

(iii) The $FGLS$ estimator was computed from the OLS residuals with the sieve $AR(p)$ models estimated from the residuals with the order p determined by BIC.

(iv) Estimation of the system based VAR parameters and determine their theoretical asymptotic MSE from their limiting distribution given in Appendix 2.

Simulation results for the estimation of the instantaneous parameter β from the above $VAR(1)$ are given in Table 5 and corresponding results for a dgp of a $VAR(3)$ are presented in Table 6.

8 Weak and Strict Exogeneity

This section briefly notes some issues involving the concept of weak or strict exogeneity in a time series context. In general concerns about possible violation of weak or strict exogeneity are most easily addressed by systems estimation as in a VAR. In situations where the investigator is for some reason limited to single equation analysis, the error term in the regression model is assumed uncorrelated with contemporaneous explanatory variables, but a concern is possible correlation of the error term with either lagged and / or future explanatory variables. Hence the regression will be

$$y_t = \beta' x_t + u_t \tag{47}$$

with the valid contemporaneous exogeneity assumption of $E(x_t u_t) = 0$, but a violation of weak exogeneity so that $E(x_{t-r} u_t) \neq 0$, for some integer $r \neq 0$. Under strict exogeneity contemporaneous exogeneity is restricted to hold. Hence the hypothesis that x_{t-r} is correlated with u_t under the maintained 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) described by Stock and Watson (2011). 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 [5]. An LM or Score test for strict exogeneity can be formed from the residuals of the regression $\widehat{u_t} = y_t = \widehat{\beta'_{OLS}} x_t$ and the cross correlation function

$$r_j = \left(\sum \widehat{u_t} \left(x_{t-j} - \overline{x_{t-j}} \right) \right) / \left(\sum \widehat{u_t}^2 \right)^{-1/2} \left(\sum \left(x_{t-j} - \overline{x_{t-j}} \right)^2 \right)^{-1/2}$$

Then under the H_0 of strict exogeneity it follows that $r_j = 0$ for $j = 1, 2, ...m$ the statistic

$$Q = T \sum r_j^2$$

will be asympotically χ^2_m under the null hypothesis. The alternative hypothesis in this case is an ADL from a VAR. An alternative hypothesis we can take an linear ADL and do not require specifiying both equations of the system.

In order to assess the impact of violations of weak and / or strict exogeneity on the application of OLS or GLS we can consider the bias and MSE of estimating a single equation when the true model is from a system or possibly VAR. The impact of neglecting weak exogeneity will be to imply a particular vector $ARMA$ structure; so that the effect of applying OLS will be very similar to the cases considered in section 6. A simple illustration of this can be from considering the simulation based on the single variable regression model with $AR(1)$ errors and $AR(1)$ explanatory variable:

$$y_t = \beta x_t + u_t \tag{48}$$

[5] The possible testing for weak exogeneity or systems modeling to avoid strict exogeneity issues are far more complicated in panel data and are not addressed in this article.

$$u_t = \phi u_{t-1} + \varepsilon_{u,t} \tag{49}$$

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

Usually $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t \varepsilon_s) = 0$ for $s \neq t$; and that $E(\varepsilon_{x,t}) = 0$, $E(\varepsilon_{x,t}^2) = \sigma_x^2$ and $E(\varepsilon_{x,t,}\varepsilon_{x,s}) = 0$ for $s \neq t$. We previously found some interesting results when $E(\varepsilon_{u,t}\varepsilon_{x,t}) = \sigma_{ux}^2 \neq 0$.On defining $\varepsilon_t^{/} = \begin{pmatrix} \varepsilon_{u,t} & \varepsilon_{x,t} \end{pmatrix}$ and for the simulation design, $\varepsilon_t \sim NID(0, \Omega)$, with Ω diagonal. The effects of the breakdowns of weak and strict exogeneity from assuming

$$E(\varepsilon_{u,t}\varepsilon_{x,t-1}) = \gamma_1 \neq 0.$$

and

$$E(\varepsilon_{u,t-1}\varepsilon_{x,t}) = \gamma_2 \neq 0.$$

$$\varepsilon_t = \eta_t - \Theta\eta_{t-1},$$

where $\eta_t \sim NID(0, V)$ and $\Theta = \begin{pmatrix} 0 & \gamma_1 \\ \gamma_2 & 0 \end{pmatrix}$.which implies a $VARMA(1,1)$ process for $\begin{pmatrix} u_t & x_t \end{pmatrix}$; and the simulation results are presented in Tables 17, 18 and 19. The results are indicative of the effects of neglecting weak exogeneity, strict exogeneity and both forms simultaneously. In the time series context; this is fundamentally equivaent to neglecting the complexity of the dynamics in a VAR or $VARMA$ system of equation. As before, $FGLS$ compares extremely favorably with OLS and the $OLSRSE$ approach. Although of course, it is still preferable to estimate the system VAR or $VARMA$ if possible.

Neglecting weak or strict exogeneity can be seen from writing the structural part of the system as an orthogonalized VAR which is

$$\begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi & -\beta\phi \\ 0 & \phi \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{u,t} \\ \varepsilon_{x,t} \end{pmatrix}.$$

and

$$\begin{pmatrix} \varepsilon_{u,t} \\ \varepsilon_{x,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & \gamma_1 \\ \gamma_2 & 0 \end{pmatrix} \begin{pmatrix} \eta_{1,t-1} \\ \eta_{2,t-1} \end{pmatrix}.$$

Hence the system can be expressed as $RY_t = AY_{t-1} + \varepsilon_t$ and $\varepsilon_t = \eta_t - \Theta\eta_{t-1}$; which implies that the structural equations are the $VARMA(1,1)$ model given by $RY_t = AY_{t-1} + \eta_t - \Theta\eta_{t-1}$, where η_t is a vector white noise process. The system can therefore be approximated by a VAR and hence dynamic multipliers and impulse responses computed in the regular manner. While this example is merely illustrative the same analysis will hold for higher order systems in both the structural equations and vector error processes; and the orthogonalized system will reduce to a finite dimensional $VARMA$ process.

9 Tests of Rational Expectations and Market Efficiency

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 rational expectations and constant risk premia in forward and futures currency markets. This type of dual null hypothesis was sometimes known as a test for market efficiency. In the following denote the logarithm of the spot exchange rate at time t as s_t and the logarithm of the k period maturity time forward exchange rate at time t as f_t.Then under the hypothesis of rational expectations and time invariant risk premium, $E_t s_{t+k} = f_t$, where E_t denotes the expectations operator conditioned on a sigma field of information. 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. A standard single equation parameterization of the problem is to base a test of the unbiasedness hypothesis on the regression

$$(s_{t+k} - s_t) = \alpha + \beta(f_t - s_t) + u_{t+k}, \tag{50}$$

where $E(u_t u_{t+j}) = 0$ for $j > k$ and to test the theory that $\alpha = 0, \beta = 1$. On initial consideration it might be considered appropriate to estimate the above by GLS with u_{t+k} folowing the invertible $MA(k)$ process

$$u_{t+k} = \varepsilon_{t+k} - \theta_1\varepsilon_{t+k-1} - .... - \theta_k\varepsilon_t \tag{51}$$

and with ε_{t+k} being white noise and possibly Gaussian. However, while the imposition of contemporaneous exogeneity restriction $E\{(f_t - s_t), \varepsilon_t\} = 0$ may be regarded acceptable; while the strict exogeneity condition that $E\{(f_t - s_t), \varepsilon_{t-j}\} = 0$ for integer valued j where $j \neq 0$ is not considered appropriate by Hansen and Hodrick (1980). Hence the use of "filtered data" which is required in the implementation of GLS may possibly lead to inconsistent estimates of the regression parameters in equation (46). This issue motivated Hansen and Hodrick (1980) to use an $OLSRSE$ type procedure. It is worth noting that an alternative literature to test this theory developed that used a system based VAR procedure with Wald type tests based on cross equation restrictions on the VAR coefficients has been developed by Baillie, Lippens and McMahon (1983). See Hodrick (1987) and Baillie (1989) for a discussion of these alternative methodologies.

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 on the methodology surrounding the $OLSRSE$ approach. Accordingly, we estimated equation (46) 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 $T = 1,609$ weekly observations. Inference on the model was achieved by (i) estimation of equation (45) by OLS with regular, but inappropriate standard errors, (ii) estimation of equation (45) by $OLSRSE$, where Newey West standard errors and also using standard errors from the adjusted estimated covariance matrix by Andrews (1991) and also by Kiefer, Vogelsang and Bunzel (2000); and (iii) estimation of equation (45) by $FGLS$ where the error term is estimated by a sieve $AR(p)$ process with order of the autoregression chosen by BIC.

The results are given in table 7. Both OLS and $FGLS$ find the standard negative slope coefficient estimates which is typical in this literature. The OLS estimates are further from unity than the $FGLS$ estimates; and the standard errors from $FGLS$ are generally considerably smaller than those based on NW from OLS. The results in table 7 are not particularly relevant except for indicating that data set has fairly standard properties. However, the more relevant results are displayed in table 8 and are generated by simulation. The aim of the simulation is to compare the properties of OLS and $FGLS$ under the null hypothesis of rational expectations and a time invariant risk premium. In order to be more precise with the theory we can impose additional restrictions on the time series prosess for $u_{t+k} = (s_{t+k} - f_t)$. In particular, since the forward rates are for a maturity time of four weeks so that $k = 4$ and since the data is measured weekly, the time of the expectation or prediction would be (22/5), or 4.40 weeks. Usually there are 22 daily innovations between the forward rate and the corressponding future spot rate; although the number of innovations can vary between 19 and 26 in extreme months. On taking 22 innovations as the appropriate number, then 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,

$$u_{t+k} = \varepsilon_t - 0.84\varepsilon_{t-1} - 0.77\varepsilon_{t-2} - 0.32\varepsilon_{t-3} - 0.09\varepsilon_{t-4} \tag{52}$$

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. For the purpose of the simulation we use six different data generating processes based on weekly data for each of the six currencies used in table 7. For each particular currency the spot exchange rate is s_t, and is the observed natural logarithm of the nominal exchange rate. Then the artificial, or constructed forward rate f_t is based on the equations

$$f_t = s_{t+4} + u_{t+4}$$

$$u_t = \varepsilon_t - 0.84\varepsilon_{t-1} - 0.77\varepsilon_{t-2} - 0.32\varepsilon_{t-3} - 0.09\varepsilon_{t-4}$$

where $\varepsilon_t \sim NID(0, \sigma^2)$. The only parameter to calibrate in the simulation is σ^2. Since

$$Var(u_t) = \sigma^2(1 + 0.84^2 + 0.77^2 + 0.32^2 + 0.09^2\ ) = 2.41\sigma^2$$

the $Var(u_t)$ is therefore backed out from the sample estimates and the realized value of $\widehat{\sigma^2} = \widehat{Var(u_t)}/2.41$. The properties of the different estimation methods and their corresponding bias and MSE are then simulated from the regression of

$$(s_{t+4} - s_t) = \alpha + \beta(f_t - s_t) + u_{t+k}$$

where the null hypothesis will be that $\beta = 1$. Table 8 summarizes the results of the simulations of the above regressions with the $OLSRSE$ directly compared with the $FGLS$ estimates. One possible strategy for implementing the $FGLS$ estimator would be to have imposed an $MA(4)$ structure on the disturbances. However, this could be interpreted as giving an undue advantage of the $FGLS$ procedure. For this reason we also maintained the approach of estimating a sieve AR on the residuals of each OLS regression. The order of the AR chosen by the BIC was as high as 12 lags for some currencies. In summary the results are quite conclusive in showing a significant reduction in both the bias and MSE of the beta in the Fama regression for all six currencies. The reduction in bias and MSE was generally in the range of 22% and 26% respectively for all six currencies.

10 Orange Juice Futures and Weather

In a well known study, Roll (1984) examined orange juice, or OJ, futures prices and their relationship to weather as measured by the number of freezing days in each month in the Orlando area of Florida; where over 90% of OJ concentrate are produced. This example is particularly interesting given that Stock and Watson (2000) use Roll's analysis in context of derivation of dynamic multipliers and the concept of strict exogeneity. On face value it would be expected that OJ returns would have little autocorrelation so that OLS estimation of a regression of OJ returns on past weather would appear reasonable. Indeed this is the aproach used by Stock and Watson (2000). However, examination of the autocorrelation function of OJ returns reveals quite strong and statistically significant autocorrelation at the seasonal lags 12 and 13. When applying $FGLS$ a sieve $AR(12)$ is chosen by BIC and the instantaneous effect of weather on OJ returns is 0.51 with no significant effects at higher order lags. Hence the reported estimated model in Table 11 is essentially the distributed lag regression of $y_t = \beta(L)FDD_t + u_t$. The presence of seasonal autocorrelation in returns implies the necessity of estimating the ADL model of $\alpha(L)y_t = \beta(L)FDD_t + u_t$; so that the estimated dynamic multipliers will be $y_t = \alpha(L)^{-1}\beta(L)FDD_t$.

The literature on the OJ futures market contains concerns about the possible breakdown of strict exogeneity in the returns - lagged weather relationship. Roll (1984) used daily data and is concerned with the issue that non synchronous data on futures prices and weather which could cause possible violations of strict exogeneity. Another reason would be if traders have access to more information than the US metereological services and possibly obtain relevant information from local citrus farmers which is additional to the meterological services. However, a regression of the weather variable regressed on lagged returns and lagged weather reveals no evidence that past OJ returns have predictable content on future weather. Hence with monthly data there is no evidence for the violation of strict exogeneity. However, the analysis of daily data might reveal a different result and would hence require estimation of a VAR and calculation of impulse responses for satisfactory, or robust analysis. It should also be noted that an article by Boudoukh et al (2007) derives interesting results using nonlinear measures of weather that include the addition of increasing the importance of the weather variable as temperatures fall close to the threshold of freezing at 32^0 Fahrenheit.

11 Taylor Rule

Another interesting example of the choice of $OLSRSE$ versus $FGLS$ methodology, and which also indicates the potential inappropriate use of instantaneous regressions, is provided by the well known Taylor Rule. The original article by Taylor (1993), proposed a monetary policy rule for central banks to set the nominal interest rate, or Federal Funds Rate (FFR), from the equation

$$i_t = \pi_t + r_t^* + \lambda_1(\pi_t - \pi_t^*) + \lambda_2\widetilde{y_t},$$

where i_t is the short term nominal interest rate (FFR), r_t^* is the equilibrium rate of interest, π_t is the CPI inflation rate, and π_t^* is the target rate of inflation. Also, $\widetilde{y_t}$ is the output gap, which is defined as $\widetilde{y_t} = 100\left\{(y_t - y_t^*)/y_t^*\right\}$,where y_t is the log of GDP and y_t^* is potential, or trend output. The original work of Taylor (1993) deterministically set $\lambda_1 = \lambda_2 = 0.5$, and $r_t^* = \pi_t^* = 0.02$. Taylor's original article did not involve any estimation work and was purely concerned with the ex post tracking of the policy rule to evaluate its success. However, subsequent research has been more focused on treating the Taylor Rule has a relationship to be estimated by econometric methods. In the estimation of a model such as

$$i_t = \pi_t + r_t^* + \lambda_1(\pi_t - \pi_t^*) + \lambda_2\widetilde{y_t} + \varepsilon_t,$$

it is assumed that ε_t is a white noise error term, which is asociated with being a monetary policy shock; while the predicted value of i_t has the interpretation of being the recommended short term interest rate. Hence there is some ambiguity as to whether the above equation is an econometric depiction of reality, or alternatively is purely a rule for conducting the ideal monetary policy. The static nature of the relationship is reminiscent of the "instantanous regression" issue described earlier.

In our empirical work we modeled the inflation rate by the $US\ GDP$ deflator; and potential or trend output by the trend component from the Hodrick Prescott filter. Given the literature on the perceived importance of regime, we report estimates of the Taylor rule equation in Tables 9 and 10. The results in Table 9 are for the era when Alan Greenspan was chairman of the Federal Reserve from $1987\ Q1$ through $2005\ Q4$; while the results in Table 10 are for the separate regime of when Ben Bernanke was chairman of the Federal Reserve from $2006\ Q1$ through $2014\ Q1$. This follows from articles such as Judd and Rudebusch (1998) which emphasize the changes in policy rules across different chairmen of the Federal Reserve.

The $OLSRSE$ estimates of the policy parameters λ_1 and λ_2 in the Greenspan regime are both 0.90, with Newey West standard errors of 0.45 and 0.23 respectively. However, there is evidence of substantial autocorrelation in the residuals of this estimated model. The $FGLS$ estimates and their asymptotic standard errors, which are given in the penultimate column of Table 9, are derived from an $AR(3)$ sieve autoregression chosen by BIC on the residuals of the $OLSRSE$ equation. The $FGLS$ estimated Taylor rule has the policy parameters λ_1 and λ_2 being 0.74 and 0.24 respectively; with asymptotic standard errors of 0.23 and 0.10 respectively. Hence the $FGLS$ estimated sum of the policy parameters is less than unity; in contrast to those obtained by OLS. Furthermore the reduction in standard errors comparing those of the $FGLS$ to $OLSRSE$ are approximately 50% for both parameter estimates and the residual variance of the equation estimated by $FGLS$ is 5% of that of OLS.

The estimates of the Taylor rule for the regime when Ben Bernanke was chairman of the Federal Reserve from 2006Q1 through 2014Q1 arguably shows even sharper distinction between the two inferential approaches. The estimated policy parameters λ_1 and λ_2 are 0.66 and 0.82 respectively from OLS, and 0.37 and 0.38 respcetively from $FGLS$. The reduction in asymptotic standard errors from using $FGLS$ are again around 50% of those of OLS.

An important question concerns the reasons and impact of the substantial autocorrelation in the Taylor rule type equations that makes $FGLS$ appear so preferable over OLS. Most likely the extent of the autocorrelation in the error term is due to neglecting the dynamics that occur in the relationships between interest rates, inflation and the output gap. In particular, a Taylor rule equation estimated over the whole sample appears to be relatively unstable with substantial time variation in estimated parameters. Apart from Judd and Rudebusch (1998) noting that the policy rule and hence parameters vary depending on the chair person of the Federal Reserve; a number of other authors have commonly included lagged interest rates in the specification of the Taylor rule equation. See Clarida, Gali and Gertler (2000) and Castelnuovo (2003) who consider the central bank is attempting to adjust nominal interest rates in a gradual fashion, or by interest rate smoothing.

Also, Coibion and Goldsmith (2012) introduced further explanatory variables involving published Federal Reserve system forecasts of future output growth and the unemployment rate. These published forecasts could conceivably be replaced with linear projections on current and lagged interest rates, output gaps, inflation, etc. Such a specification includes a much larger set of lagged information on standard macro variables. Hence the econometric specification of a relationship known as the Taylor Rule becomes a dynamic specification like an ADL, or one equation of a VAR. Such a system would then be used to derive dynamic multipliers or impulse responses.

12 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. Modern applied time series econometrics tends to estimate regressions with consistent, but asymptotically inefficient OLS and to base inference of conditional mean parameters on robust standard errors. This approach appears to be implemented 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 can actually 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 properties of estimating distributed lag regressions rather than basing inference on a full system VAR.

We provide a detailed example of the issues from the application of to the testing of rational expectations and risk neutrality in the theory of efficient markets. Further research is currently under way to consider the impact of the above ideas on inference in the estimated Taylor rulemacro model.

The paper also shows that even when strict exogeneity is violated the $FGLS$ approach dominates $OLSRSE$ in terms of bias and MSE. The general conclusions of the paper 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.

13 Appendix 1: Relative Biases in Regression with AR(p) Errors

From the model in equations (9), (10) and (11) it can be shown that for the OLS estimator

$$\hat{\beta}_{OLS} = \beta + \frac{\frac{1}{T}\sum x_t u_t}{\frac{1}{T}\sum x_t^2} = \beta + \frac{\frac{1}{T}\sum x_t(\phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_k u_{t-k} + \epsilon_{u,t})}{\frac{1}{T}\sum x_t^2}$$

Hence

$$plim\left(\hat{\beta}_{OLS} - \beta\right) = \frac{\phi_1 \mathbb{E}(x_t u_{t-1}) + \phi_2 \mathbb{E}(x_t u_{t-2}) + \cdots + \phi_k \mathbb{E}(x_t u_{t-k}) + \mathbb{E}(x_t \epsilon_{u,t})}{\mathbb{E}(x_t^2)}$$

where $\mathbb{E}(x_t^2) = \frac{\sigma_x^2}{1-\rho^2}$. On noting that,

$$\mathbb{E}(x_{t+j} u_t) = \rho^j \mathbb{E}(x_t u_t), \qquad j = 0, 1, 2, \cdots$$

and from $\mathbb{E}(x_t \epsilon_{u,t}) = \sigma_{ux}^2$ and $\mathbb{E}(x_t u_t) = \phi_1 \mathbb{E}(x_t u_{t-1}) + \phi_2 \mathbb{E}(x_t u_{t-2}) + \cdots + \phi_k \mathbb{E}(x_t u_{t-k}) + \mathbb{E}(x_t \epsilon_{u,t})$,

$$\mathbb{E}(x_t u_t) = \frac{\sigma_{ux}^2}{1 - \phi_1 \rho - \phi_2 \rho^2 - \cdots - \phi_k \rho^k}$$

Therefore,

$$plim\left(\hat{\beta}_{OLS} - \beta\right) = \frac{\mathbb{E}(x_t u_t)}{\mathbb{E}(x_t^2)} = \frac{\sigma_{ux}^2}{1 - \phi_1 \rho - \phi_2 \rho^2 - \cdots - \phi_k \rho^k}\left(\frac{1-\rho^2}{\sigma_x^2}\right)$$

For the GLS estimator

$$\begin{aligned}\hat{\beta}_{GLS} &= \frac{\frac{1}{T}\sum_t (x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})(y_t - \phi_1 y_{t-1} - \cdots - \phi_k y_{t-k})}{\frac{1}{T}\sum_t (x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})^2} \\ &= \beta + \frac{\frac{1}{T}\sum_t (x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})(u_t - \phi_1 u_{t-1} - \cdots - \phi_k u_{t-k})}{\frac{1}{T}\sum_t (x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})^2}\end{aligned}$$

That is,

$$\hat{\beta}_{GLS} - \beta = \frac{\frac{1}{T}\sum_t (x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})(u_t - \phi_1 u_{t-1} - \cdots - \phi_k u_{t-k})}{\frac{1}{T}\sum_t (x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})^2}$$

Hence

$$plim\left(\hat{\beta}_{GLS} - \beta\right) = \frac{\mathbb{E}\left[(x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})(u_t - \phi_1 u_{t-1} - \cdots - \phi_k u_{t-k})\right]}{\mathbb{E}\left[(x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})^2\right]}$$

To derive the denominator of the above we note that $\mathbb{E}\left(x_t^2\right) = \rho^2 \mathbb{E}\left(x_{t-1}^2\right) + \sigma_x^2$ and hence $\mathbb{E}\left(x_t x_{t+k}\right) = \rho^k \mathbb{E}\left(x_t^2\right) = \frac{\sigma_x^2 \rho^k}{1-\rho^2}$Then, the denominator becomes:

$$\mathbb{E}\left\{\left[x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k}\right]^2\right\} = \frac{\sigma_x^2}{1-\rho^2}\left(1 + \sum_{i=1}^{k}\phi_i^2 - 2\sum_{i=1}^{k}\phi_i \rho^i + 2\sum_{i<j}^{k}\sum^{k}\phi_i\phi_j\rho^{j-i}\right)$$

To obtain the numerator,

$$\begin{aligned} & \mathbb{E}\left[(x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})(u_t - \phi_1 u_{t-1} - \cdots - \phi_k u_{t-k})\right] \\ = & \ \mathbb{E}\left[(x_t - \phi_1 x_{t-1} - \cdots - \phi_k x_{t-k})\epsilon_{u,t}\right] \\ = & \ \mathbb{E}(x_t \epsilon_{u,t}) = \rho \mathbb{E}(\epsilon_{u,t} x_{t-1}) + \mathbb{E}(\epsilon_{u,t}\epsilon_{x,t}) = \sigma_{ux}^2 \end{aligned} \quad (53)$$

Thus, by applying the previous results,

$$plim\left(\hat{\beta}_{GLS}-\beta\right)=\left(\frac{1-\rho^2}{\sigma_x^2}\right)\frac{\sigma_{ux}^2}{1+\sum_{i=1}^{k}\phi_i^2-2\sum_{i=1}^{k}\phi_i\rho^i+2\sum_{i<j}^{k}\sum^{k}\phi_i\phi_j\rho^{j-i}}$$

Therefore,

$$\frac{plim\left(\hat{\beta}_{OLS}-\beta\right)}{plim\left(\hat{\beta}_{GLS}-\beta\right)}=\frac{1+\sum_{i=1}^{k}\phi_i^2-2\sum_{i=1}^{k}\phi_i\rho^i+2\sum_{i<j}^{k}\sum^{k}\phi_i\phi_j\rho^{j-i}}{1-\phi_1\rho-\phi_2\rho^2-\cdots-\phi_k\rho^k}$$

14 Appendix 2: 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,

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\xrightarrow{L}N(\mathbf{0},\mathbf{V}). \tag{54}$$

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 matrix 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] Boudoukh, J., M. Richardson, Y. Shen and Whitelaw, R.F. 2007. Do asset prices reflect fundamentals? Freshly squeezed evidence from the OJ market. Journal of Financial Economics, 83, 397–412.

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

[11] 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.

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

[13] Castelnouva, E., 2003. Taylor rules, omitted variables and interest rate smoothing in the US, Economics Letters, 81, 55–59.

[14] Clarida, R., J. Gali and M. Gertler, 2000. Monetary policy rules and macroeconomic stability: evidence from some theory, Quarterly Journal of Economics, 115, 147-180.

[15] Coibion, O. and D. Goldstein, 2012. One for some, or one for all? Taylor rules and interregional heterogeneity, Journal of Money, Credit and Banking, 44, 401–432.

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

[17] 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.

[18] Grenander, U., 1981. Indirect Inference, Wiley, New York.

[19] Hamilton, J., 1994. Time Series Econometics, Princeton University Press.

[20] 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.

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

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

[23] 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.

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

[25] Judd, J.P. and Rudebusch, G.D., 1998. Taylor's rules and the Fed: 1970-1997", Federal Reserve Bank of San Francisco, 1–17.

[26] Kapetanios, G. and Z. Psaradakis, 2015. Semiparametric sieve type GLS inference. Econometric Reviews ??, ??.

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

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

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

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

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

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

[33] Rudebusch, G.D. 2002. Term structure evidence on interest rate smoothing and monetary policy inertia, Journal of Monetary Economics, 49, 1161–1187.

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

[35] Stock, J.H. and M. Watson, 2011. Intoduction to Econometrics. Third edition, Pearson.

[36] Taylor, J.B., 1993. Discretion versus policy rules in practice, Carnegie-Rochester Conference Series on Public Policy, 39, 195–214.

Table 1: Bias and Mean Squared Error (MSE) of 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) of 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 $V = \begin{pmatrix} \sigma^2_x & \sigma^2_{x,\epsilon} \\ \sigma^2_{x,\epsilon} & \sigma^2_\epsilon \end{pmatrix}$. 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) of 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) of 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 $V = \begin{pmatrix} \sigma_x^2 & \sigma_{x,\epsilon}^2 \\ \sigma_{x,\epsilon}^2 & \sigma_\epsilon^2 \end{pmatrix}$. In all experiments $\beta = 2$ and $V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

.

Table 5: Bias and Mean Squared Error (MSE) of 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	0.7186	-0.8375	0.7076	-0.8367	0.7029	-0.8374	0.7035
OLS-NW	-0.8402	0.7184	-0.8375	0.7097	-0.8367	0.7026	-0.8374	0.7036
OLS-Andrews	-0.8402	0.7184	-0.8375	0.7097	-0.8367	0.7026	-0.8374	0.7037
OLS-KVB	-0.8402	0.7062	-0.8375	0.7034	-0.8367	0.7005	-0.8374	0.7018
FGLS	-0.4507	0.2056	-0.4412	0.1958	-0.4381	0.1925	-0.4363	0.1908
VAR(1)	0.0012	0.0050	0.0014	0.0026	-0.0010	0.0013	-0.0007	0.0010

Key: OLS-NW refers to standard errors calculated by the Newey West methodology; OLS-Andrews denotes standard errors from the method of Andrews (1991); while OLS-KVB denotes standard errors from the method of Kiefer, Vogelsang and Bunzel (2000). See the text in section for a full description of the VAR data generating process.

.

Table 6: Bias and Mean Squared Error (MSE) of 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	0.2629	0.1044	0.2539	0.0820	0.2515	0.0722	0.2469	0.0675
OLS-NW	0.2629	0.1258	0.2539	0.0809	0.2515	0.0716	0.2469	0.0694
OLS-Andrews	0.2629	0.1258	0.2539	0.0809	0.2515	0.0716	0.2469	0.0694
OLS-KVB	0.2629	0.0819	0.2539	0.0676	0.2515	0.0636	0.2469	0.0611
FGLS	-0.1395	0.0392	-0.1379	0.0286	-0.1411	0.0250	-0.1423	0.0251
VAR(3)	-0.0035	0.0053	0.0018	0.0025	0.0008	0.0013	0.0001	0.0010

Key: as for the key to table 5.

.

Table 7: Bias and Mean Squared Error (MSE) for OLS, $OLSRSE$, $FGLS$ and GLS for Single Equation Regression $y_t = \beta x_t + u_t$, where both the weak and strict exogeneity assumptions are violated.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	-0.0814	0.0131	-0.0807	0.0094	-0.0805	0.0081	-0.0794	0.0078
OLS-NW	-0.0814	0.0126	-0.0807	0.0091	-0.0805	0.0082	-0.0794	0.0082
OLS-Andrews	-0.0814	0.0126	-0.0807	0.0091	-0.0805	0.0082	-0.0794	0.0082
OLS-KVB	-0.0814	0.0088	-0.0807	0.0068	-0.0805	0.0068	-0.0794	0.0065
FGLS	-0.0591	0.0084	-0.0561	0.0054	-0.0553	0.0044	-0.0552	0.0042
GLS	-0.0556	0.0078	-0.0540	0.0050	-0.0537	0.0042	-0.0534	0.0040

Key: The parameter values are $\beta = 2$, $\phi = 0.4$, $\rho = 0.6$, $\gamma_1 = 0.3$, $\gamma_2 = -0.2$ and $\Omega = I$

.

Table 8: Bias and Mean Squared Error (MSE) for OLS, $OLSRSE$, $FGLS$ and GLS for Single Equation Regression $y_t = \beta x_t + u_t$, where only the strict exogeneity assumption is violated.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	0.0625	0.0098	0.0625	0.0069	0.0639	0.0055	0.0624	0.0050
OLS-NW	0.0625	0.0090	0.0625	0.0061	0.0639	0.0053	0.0624	0.0050
OLS-Andrews	0.0625	0.0090	0.0625	0.0061	0.0639	0.0053	0.0624	0.0050
OLS-KVB	0.0625	0.0061	0.0625	0.0041	0.0639	0.0042	0.0624	0.0042
FGLS	0.0067	0.0053	0.0055	0.0025	0.0031	0.0012	0.0019	0.0010
GLS	-0.0013	0.0046	-0.0007	0.0023	-0.0006	0.0011	-0.0014	0.0009

Key: The parameter values are $\beta = 2$, $\phi = 0.4$, $\rho = 0.6$, $\gamma_1 = 0$, $\gamma_2 = -0.2$ and $\Omega = I$

.

Table 9: Bias and Mean Squared Error (MSE) for OLS, $OLSRSE$, $FGLS$ and GLS for Single Equation Regression $y_t = \beta x_t + u_t$, where only the weak exogeneity assumption is violated.

	T=200		T=400		T=800		T=1000	
Estimator	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
OLS	-0.1506	0.0285	-0.1496	0.0255	-0.1505	0.0243	-0.1508	0.0240
OLS-NW	-0.1506	0.0330	-0.1496	0.0248	-0.1505	0.0239	-0.1508	0.0239
OLS-Andrews	-0.1506	0.0330	-0.1496	0.0248	-0.1505	0.0239	-0.1508	0.0239
OLS-KVB	-0.1506	0.0242	-0.1496	0.0226	-0.1505	0.0229	-0.1508	0.0228
FGLS	-0.0746	0.0114	-0.0701	0.0078	-0.0694	0.0063	-0.0683	0.0057
GLS	-0.0555	0.0078	-0.0556	0.0057	-0.0554	0.0044	-0.0552	0.0040

Key: The parameter values are $\beta = 2$, $\phi = 0.4$, $\rho = 0.6$, $\gamma_1 = 0.3$, $\gamma_2 = 0$ and $\Omega = I$

Table 10: Estimates of Beta in the Fama regression of spot returns on lagged forward premium.

	OLS	SE	SE-NW	SE-Andrews	SE-KVB	FGLS	FGLS-SE	BIC-lag
Australia	-1.0245	0.0733	0.0860	0.0855	0.0251	-0.8469	0.0213	13
Canada	-1.0416	0.0699	0.1128	0.1131	0.0295	-0.8975	0.0194	13
Switzerland	-1.1161	0.0788	0.0922	0.0922	0.0607	-0.8339	0.0244	13
UK	-1.1891	0.0809	0.1071	0.1057	0.0297	-0.8702	0.0247	14
Japan	-1.1752	0.0782	0.0870	0.0866	0.0623	-0.8339	0.0233	13
New Zealand	-1.0848	0.0687	0.0845	0.0842	0.0551	-0.8823	0.0195	14

.

Table 11: Bias and MSE of the OLS and $FGLS$ Estimates of Beta in the Fama regression.

	Australia			
	Model 1		Model 2	
	Bias	MSE	Bias	MSE
OLS	-0.5520	0.5948	-0.5443	0.5888
FGLS	-0.4514	0.4700	-0.4573	0.4708
	Canada			
	Model 1		Model 2	
	Bias	MSE	Bias	MSE
OLS	-0.2034	0.1530	-0.1989	0.1410
FGLS	-0.1722	0.1308	-0.1678	0.1200
	Switzerland			
	Model 1		Model 2	
	Bias	MSE	Bias	MSE
OLS	-0.5671	0.6278	-0.5381	0.5842
FGLS	-0.4584	0.4922	-0.4395	0.4584
	UK			
	Model 1		Model 2	
	Bias	MSE	Bias	MSE
OLS	-0.4115	0.3958	-0.4330	0.4165
FGLS	-0.3384	0.3227	-0.3532	0.3357
	Japan			
	Model 1		Model 2	
	Bias	MSE	Bias	MSE
OLS	-0.5071	0.5072	-0.4860	0.4739
FGLS	-0.4045	0.3959	-0.3951	0.3745
	New Zealand			
	Model 1		Model 2	
	Bias	MSE	Bias	MSE
OLS	-0.6461	0.7396	-0.6449	0.7338
FGLS	-0.5256	0.5684	-0.5276	0.5743

Key: Model 1 is $(s_{t+4} - s_t) = \alpha + \beta(f_t - s_t) + u_{t+k}$ and Model 2 is $(s_{t+4} - f_t) = \alpha + \beta(f_t - s_t) + u_{t+k}$. For both models, Bias is $\times 10^{-3}$ and MSE is $\times 10^{-6}$.

Table 12: Estimated Taylor rule during the “Greenspan era" (1987:Q4–2005:Q4).

	OLS	SE	SE-NW	SE-Andrews	SE-KVB	**FGLS**	FGLS-SE
Inflation	**0.8975**	0.2734	0.4462	0.4462	0.3654	**0.7408**	0.2259
Output Gap	**0.9000**	0.1826	0.2342	0.2342	0.0785	**0.2414**	0.0982
Constant	**2.6127**	0.6851	1.0179	1.0179	0.6343	**0.1365**	0.0522
Var(residual)	**2.8956**	N/A	N/A	N/A	N/A	**0.1302**	N/A
BIC-lag	N/A	N/A	N/A	N/A	N/A	**3**	N/A

Key: The US quarterly data during 1987:Q4–2005:Q4 are used here. Inflation is measured by the US GDP deflator, while the output gap is measured by the Hodrick-Prescott filtered US GDP.

Table 13: Estimated Taylor rule during the “Bernanke era" (2006:Q1–2014:Q1).

	OLS	SE	SE-NW	SE-Andrews	SE-KVB	**FGLS**	FGLS-SE
Inflation	**0.6636**	0.4735	0.6241	0.6084	0.2722	**0.3650**	0.2157
Output Gap	**0.8152**	0.2171	0.2381	0.2336	0.0608	**0.3789**	0.1044
Constant	**0.3358**	0.9240	1.2397	1.2155	0.4440	**0.1143**	0.1052
Var(residual)	**1.3312**	N/A	N/A	N/A	N/A	**0.1061**	N/A
BIC-lag	N/A	N/A	N/A	N/A	N/A	**3**	N/A

Key: The US quarterly data during 2006:Q1–2014:Q1 are used here. Inflation is measured by the US GDP deflator, while the output gap is measured by the Hodrick-Prescott filtered US GDP.

.

Table 14: Autoregressive distributed lag (ADL) of current interest rate on lagged interest rate, current/lagged inflation minus its target and current/lagged output gap variables.

	Coefficient	OLS std. error
Lag 1 of interest rate	1.5439	0.1278
Lag 2 of interest rate	-0.5785	0.2101
Lag 3 of interest rate	0.0251	0.1134
Lag 0 of inflation gap	0.4556	0.2250
Lag 1 of inflation gap	-0.3043	0.3130
Lag 2 of inflation gap	-0.1980	0.3157
Lag 3 of inflation gap	0.0423	0.2185
Lag 0 of output gap	0.2469	0.0850
Lag 1 of output gap	0.0073	0.1250
Lag 2 of output gap	-0.3540	0.1200
Lag 3 of output gap	0.1103	0.0959

Table 15: Dynamic multipliers of inflation and output gap variables during the “Greenspan" regime.

	Dynamic multiplier
Lag 0 of inflation gap	0.4556
Lag 1 of inflation gap	0.3991
Lag 2 of inflation gap	0.1546
Lag 3 of inflation gap	0.0616
Lag 0 of output gap	0.2469
Lag 1 of output gap	0.3885
Lag 2 of output gap	0.1029
Lag 3 of output gap	0.0507

.

Table 16: Autoregressive distributed lag (ADL) of current interest rate on lagged interest rate, current/lagged inflation minus its target and current/lagged output gap variables during the "Bernanke" regime.

	Coefficient	OLS std. error
Lag 1 of interest rate	1.1301	0.1815
Lag 2 of interest rate	0.1657	0.2910
Lag 3 of interest rate	-0.3593	0.1709
Lag 0 of inflation gap	0.3997	0.2006
Lag 1 of inflation gap	-0.1344	0.2830
Lag 2 of inflation gap	-0.2125	0.2814
Lag 3 of inflation gap	0.1007	0.1789
Lag 0 of output gap	0.1977	0.0912
Lag 1 of output gap	-0.2316	0.1429
Lag 2 of output gap	-0.3215	0.1585
Lag 3 of output gap	0.3409	0.1119

Key: The variance of residual is 0.0424.

Table 17: Dynamic multipliers of inflation and output gap variables during the "Bernanke" regime.

	Dynamic multiplier
Lag 0 of inflation gap	0.3997
Lag 1 of inflation gap	0.3173
Lag 2 of inflation gap	0.2123
Lag 3 of inflation gap	0.2496
Lag 0 of output gap	0.1977
Lag 1 of output gap	-0.0082
Lag 2 of output gap	-0.2980
Lag 3 of output gap	-0.0682

Table 18: OLS, OLSRSE and FGLS of OJ Returns regressed on the number of freezing degree days (i.e. weather).

	Coefficient	Std. error	Var. of residual	BIC-lag
OLS	0.4585	0.0575	22.9695	N/A
OLS-NW	0.4585	0.1320	22.9695	N/A
OLS-Andrews	0.4585	0.1320	22.9695	N/A
OLS-KVB	0.4585	0.0495	22.9695	N/A
FGLS	0.4372	0.0549	22.1270	12

.

Table 19: Autoregressive distributed lag (ADL) of current return on the BIC-chosen number of lagged return and lagged weather variables.

	Coefficient	OLS std. error
Lag 1 of return	0.0968	0.0401
Lag 2 of return	0.0162	0.0403
Lag 3 of return	0.0605	0.0402
Lag 4 of return	-0.0069	0.0402
Lag 5 of return	-0.0479	0.0402
Lag 6 of return	0.0553	0.0402
Lag 7 of return	-0.0614	0.0402
Lag 8 of return	-0.0504	0.0403
Lag 9 of return	0.0260	0.0403
Lag 10 of return	0.0279	0.0403
Lag 11 of return	0.0587	0.0403
Lag 12 of return	-0.1264	0.0397
Lag 0 of weather	0.4853	0.0575
Lag 1 of weather	0.0984	0.0596
Lag 2 of weather	0.0232	0.0592
Lag 3 of weather	0.0124	0.0592
Lag 4 of weather	-0.0039	0.0592
Lag 5 of weather	0.0411	0.0591
Lag 6 of weather	-0.0051	0.0599
Lag 7 of weather	0.0261	0.0599
Lag 8 of weather	-0.0012	0.0599
Lag 9 of weather	-0.0104	0.0599
Lag 10 of weather	-0.1359	0.0599
Lag 11 of weather	0.0907	0.0605
Lag 12 of weather	-0.0887	0.0612

Table 20: ADL of current return on 18 lagged return and 18 lagged weather variables.

	Coefficient	OLS std. error
Lag 1 of return	0.0666	0.0413
Lag 2 of return	0.0078	0.0414
Lag 3 of return	0.0523	0.0413
Lag 4 of return	-0.0087	0.0411
Lag 5 of return	-0.0594	0.0411
Lag 6 of return	0.0466	0.0410
Lag 7 of return	-0.0559	0.0406
Lag 8 of return	-0.0558	0.0406
Lag 9 of return	0.0229	0.0406
Lag 10 of return	0.0282	0.0406
Lag 11 of return	0.0617	0.0405
Lag 12 of return	-0.1090	0.0406
Lag 13 of return	-0.0806	0.0408
Lag 14 of return	-0.0078	0.0408
Lag 15 of return	-0.0407	0.0408
Lag 16 of return	-0.0231	0.0408
Lag 17 of return	-0.0423	0.0408
Lag 18 of return	-0.0482	0.0403
Lag 0 of weather	0.4952	0.0584
Lag 1 of weather	0.1293	0.0619
Lag 2 of weather	0.0477	0.0621
Lag 3 of weather	0.0219	0.0621
Lag 4 of weather	-0.0057	0.0621
Lag 5 of weather	0.0362	0.0620
Lag 6 of weather	-0.0029	0.0619
Lag 7 of weather	0.0247	0.0605
Lag 8 of weather	0.0022	0.0600
Lag 9 of weather	-0.0120	0.0599
Lag 10 of weather	-0.1376	0.0599
Lag 11 of weather	-0.0996	0.0606
Lag 12 of weather	-0.0981	0.0617
Lag 13 of weather	-0.0373	0.0629
Lag 14 of weather	-0.0396	0.0629
Lag 15 of weather	-0.0013	0.0629
Lag 16 of weather	0.0255	0.0628
Lag 17 of weather	0.0233	0.0627
Lag 18 of weather	0.0267	0.0622

Table 21: Stock and Watson: return variable on current and 18 lagged weather variables.

	Coefficient	OLS std. error
Lag 0 of weather	0.5115	0.0594
Lag 1 of weather	0.1681	0.0594
Lag 2 of weather	0.0673	0.0594
Lag 3 of weather	0.0647	0.0594
Lag 4 of weather	0.0206	0.0594
Lag 5 of weather	0.0345	0.0594
Lag 6 of weather	0.0258	0.0594
Lag 7 of weather	0.0036	0.0587
Lag 8 of weather	-0.0260	0.0585
Lag 9 of weather	-0.0033	0.0585
Lag 10 of weather	-0.1134	0.0585
Lag 11 of weather	-0.0721	0.0587
Lag 12 of weather	-0.1463	0.0593
Lag 13 of weather	-0.0982	0.0600
Lag 14 of weather	-0.0604	0.0601
Lag 15 of weather	-0.0272	0.0601
Lag 16 of weather	0.0011	0.0601
Lag 17 of weather	0.0009	0.0601
Lag 18 of weather	-0.0009	0.0600

.

Table 22: ADL of current weather on the BIC-chosen number of lagged weather and lagged return variables.

	Coefficient	OLS std. error
Lag 1 of weather	0.0230	0.0421
Lag 2 of weather	0.0349	0.0418
Lag 3 of weather	0.0048	0.0418
Lag 4 of weather	0.0052	0.0418
Lag 5 of weather	-0.0053	0.0418
Lag 6 of weather	-0.0091	0.0424
Lag 7 of weather	0.0003	0.0424
Lag 8 of weather	0.0147	0.0423
Lag 9 of weather	0.0080	0.0424
Lag 10 of weather	0.0078	0.0423
Lag 11 of weather	0.1304	0.0424
Lag 12 of weather	0.1817	0.0426
Lag 1 of return	0.0151	0.0284
Lag 2 of return	-0.0241	0.0284
Lag 3 of return	-0.0070	0.0284
Lag 4 of return	-0.0107	0.0284
Lag 5 of return	0.0151	0.0284
Lag 6 of return	0.0123	0.0284
Lag 7 of return	-0.0052	0.0284
Lag 8 of return	-0.0327	0.0284
Lag 9 of return	-0.0178	0.0285
Lag 10 of return	-0.0104	0.0285
Lag 11 of return	-0.0065	0.0284
Lag 12 of return	-0.0108	0.0281

Table 23: ADL of current weather on the BIC-chosen number of lagged weather variables.

	Coefficient	OLS std. error
Lag 1 of weather	0.0322	0.0396
Lag 2 of weather	0.0258	0.0394
Lag 3 of weather	-0.0024	0.0394
Lag 4 of weather	-0.0006	0.0394
Lag 5 of weather	0.0001	0.0394
Lag 6 of weather	0.0000	0.0399
Lag 7 of weather	0.0001	0.0399
Lag 8 of weather	0.0002	0.0399
Lag 9 of weather	-0.0038	0.0399
Lag 10 of weather	0.0011	0.0399
Lag 11 of weather	0.1229	0.0398
Lag 12 of weather	0.1760	0.0400

. .