Journal of Data Science 0 (0), 1–19
June 2023
DOI: 10.6339/23-JDS1108
Statistical Data Science

Revisiting the Use of Generalized Least Squares in Time Series Regression Models

Yue Fang[1,*], Sergio G. Koreisha[2], and Qi-man Shao[3]

[1]*Department of Economics and Decision Sciences, China International Business School, Shanghai, P.R. China*
[2]*Lundquist College of Business, University of Oregon, Eugene, OR, USA*
[3]*Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, P.R. China*

Abstract

Linear regression models are widely used in empirical studies. When serial correlation is present in the residuals, generalized least squares (GLS) estimation is commonly used to improve estimation efficiency. This paper proposes the use of an alternative estimator, the approximate generalized least squares estimators based on high-order AR(p) processes (GLS-AR). We show that GLS-AR estimators are asymptotically efficient as GLS estimators, as both the number of AR lag, p, and the number of observations, n, increase together so that $p = o(n^{1/4})$ in the limit. The proposed GLS-AR estimators do not require the identification of the residual serial autocorrelation structure and perform more robust in finite samples than the conventional FGLS-based tests. Finally, we illustrate the usefulness of GLS-AR method by applying it to the global warming data from 1850–2012.

Keywords *autocorrelation; efficient estimation; hypothesis testing; serial correlation; time domain*

1 Introduction

Linear regression models are widely used in empirical studies. When serial correlation is present in the residuals, generalized least squares (GLS) estimation is commonly used to obtain estimates for the regression coefficients and to generate forecasts. In this article we propose the use of an alternative estimator, the approximate generalized least squares estimators based on high-order AR processes (GLS-AR). We show that although GLS and GLS-AR estimators are asymptotically equivalent, they differ appreciably in terms of finite sample properties. For sample sizes typically found in empirical research, statistical inferences based on tests using GLS-AR are more robust and reliable than those based on GLS estimation as well as those based on generic, multipurpose procedures such as heteroscedastic autocorrelation consistent (HAC) estimation.

Although extensive research has been conducted over the last five decades on the properties of estimators and predictors obtained from serially correlated regression models since the seminal contributions of Cochrane and Orcutt (1949) and Prais and Winsten (1954), until recently, with few exceptions (noticeably Goldberger, 1962; Amemiya, 1973; Engle, 1974), most of the studies assumed that the error covariance matrix Σ_n from the regression model,

$$Y = X\boldsymbol{\beta} + u, \tag{1}$$

*Corresponding author. Email: fyue@ceibs.edu.

Received January 3, 2023; Accepted June 10, 2023

where Y is an $(n \times 1)$ vector of observations on a dependent variable, X is a $(n \times k)$ design matrix with full column rank and u is an $(n \times 1)$ random vector with $E(u|X) = 0$ and $E(u'u|X) = \Sigma_n$, was either known or could be estimated consistently from the data so that GLS could be used to estimate the regression model.

In practice, of course, the error terms are unobservable quantities and their covariance structures have to be identified from the data as part of the specification of the model. Identification of the structure of the covariance matrix Σ_n when GLS is used to estimate the regression model with autocorrelated disturbances, however, as noted by Walker (1967), Kadiyala (1970), King (1983), and Koreisha and Pukkila (1985) among others is difficult and challenging. Walker and King discussed the practical difficulties of testing for the autoregressive model of order one (AR(1)) against the moving average model of order one (MA(1)) disturbances. Kadiyala has shown that the null hypothesis that u follows $N(0, \Sigma_{n,1})$ cannot be tested against the alternative hypothesis that u follows $N(0, \Sigma_{n,2})$ under certain conditions on $\Sigma_{n,i}$ $(i = 1, 2)$ and the design matrix X. The test aside from a scalar factor may have the same distribution for both the null and the alternative hypotheses. Moreover, as pointed out by Thursby (1987), if the regression suffers from omitted variables, then it may not be possible to identify the form of the autocorrelation present in the regression model.

As an alternative to GLS, several more recent studies have considered the use of high-order AR models to approximate the more general linear model. Maddala (2001) (p. 250), for instance, argued that "... from the practical point of view, for most economic data, it is just sufficient to determine the order of the AR process. Thus, if a significant DW statistic is observed, the appropriate strategy would be to try see whether the errors are generated by a higher-order AR process than AR(1) and then undertake estimation." (He did not provide any theoretical justification for this approach). Koreisha and Fang (2001) compared the asymptotic and finite sample efficiencies of ordinary least squares (OLS) and GLS vis-à-vis GLS-AR. They showed that under some regularity conditions GLS-AR estimators are consistent. They found that the feasible GLS-AR estimators yielded more efficient estimates than OLS, and were comparable to feasible GLS particularly when the order of the autoregression was set near $[n^{1/2}]$.

In this study we will derive the asymptotic distributional properties of GLS-AR(p) estimators, and construct test statistics based on them for testing the significance of regression coefficients. In particular we will show that as p increases at an appropriate rate dependent on n, the GLS-AR estimators, will yield estimates for serially correlated regression parameters that are asymptotically efficient, and that have the same limiting distribution as those of GLS estimators based on correct residual autocovariance structures. Thus, since the GLS-AR estimators are best linear unbiased estimators (BLUE), there is not much to be gained in trying to identify the correct order and form of the serial correlation since estimation and inferences about the regression coefficients can be efficiently made based on AR corrections.

Our theoretical contributions, as we will show, extend previous findings in pure time series settings (and most often in the frequency domain, e.g. Berk (1974)'s classic paper) to time series regression models under very general structures of design matrices including time trend regressors.

More specifically we will consider the following estimators of β for the linear regression model specified by (1):

$$\hat{\beta} = (X'\Xi_n^{-1}X)^{-1}X'\Xi_n^{-1}Y, \tag{2}$$

where Ξ_n is a positive definite matrix. Setting $\Xi_n = I_n$ in (2) defines the OLS estimator, $\hat{\beta}_{\mathrm{OLS}}$; if $\Xi_n = \Sigma_n$, then (2) yields the well-known GLS estimator, $\hat{\beta}_{\mathrm{GLS}}$. If the structure of Σ_n is known,

but its elements have to be estimated, i.e., $\Xi_n = \hat{\Sigma}_n$, then we will refer to (2) as feasible GLS and denote it as $\hat{\beta}_{\text{FGLS}}$. We will refer to (2) as the GLS-AR, $\hat{\beta}_{\text{GLS-AR}}$, if Ξ_n is based on high-order AR processes.

We also consider the Newey-West's HAC estimators, which as the GLS-AR estimators, do not require the identification of the residual autocovariance structures. We note, however, that the HAC procedure is designed to provide consistent estimates of covariance matrices of residuals in regression models with both autocorrelated and heteroscedastic residuals. Nevertheless, we include HAC in our simulation comparisons because it is often used by practitioners analyzing linear models with autocorrelated residuals regardless of the form of the covariance structure.

Finally, we note that the idea of using a sequence of autoregressions to whiten a stationary time series has a long history in statistics, and has played an important role in developing new statistical techniques in many research areas. For example, we may reference the works of Koreisha and Pukkila (1985) and Pukkila et al. (1990) on estimation and identification of ARMA models; Saikkonen (1991) on developing asymptotically efficient estimation of cointegration regressions, and more recently, Harvey et al. (2010) on detecting multiple level breaks in autocorrelated time series; Choi and Kurozumi (2012) on model selection criteria for the leads-and-lags cointegrating regression; Chambers et al. (2014) on testing for seasonal unit roots; and Chambers (2013) on Jackknife estimation of stationary autoregressive models. We are optimistic that our theoretical developments in the time domain will lead to similar advances in time series analysis.

The article will be organized as follows: in Section 2 we present the large-sample distribution properties of GLS-AR estimators; in Section 3 we provide the results from an exhaustive Monte Carlo study contrasting the empirical sizes and power of test statistics based on several GLS, GLS-AR, and HAC estimators; in Section 4 we present an empirical application using global warming data; and finally, in Section 5 we offer some concluding remarks. All technical proofs and additional simulation results are available at Online Supplement.

2 Large-Sample Distributional Properties of GLS-AR Estimators

In this section we will first derive the asymptotic properties of the feasible GLS-AR estimator ($\hat{\beta}_{\text{FGLS-AR}}$). In Theorem 1, we will show that under some classic regression regularity conditions and for sufficiently large p, $\hat{\beta}_{\text{FGLS-AR}}$ is asymptotically efficient, normally distributed with variance approaching that of the GLS estimator $\hat{\beta}_{\text{GLS}}$. Then, in Section 2.2, we will show how to conduct parameter inferences about the regression coefficients.

2.1 Asymptotic Properties of Feasible GLS-AR Estimators

To construct $\hat{\beta}_{\text{FGLS-AR}}$ we first run a least squares regression to obtain an estimator for u and its covariance matrix Σ_n assuming that u can be approximated by a finite-order autoregressive process.

Let $\hat{u} = Y - X\hat{\beta}_{\text{OLS}}$, then

$$\hat{u} = Y - X(\beta + (X'X)^{-1}X'u) = (I_n - X(X'X)^{-1}X')u. \tag{3}$$

We define

$$\hat{\beta}_{\text{FGLS-AR}} \equiv (X'\hat{\Sigma}_n^{-1}X)^{-1}X'\hat{\Sigma}_n^{-1}Y, \tag{4}$$

where $\hat{\Sigma}_n = (\hat{\sigma}_{i,j}, 1 \leqslant i, j \leqslant n)$ and the entries in $\hat{\Sigma}_n$, $\hat{\sigma}_{i,j} = \hat{r}_{|j-i|}$ are estimated using

$$\hat{r}_j \equiv \hat{r}_j(p) = \begin{cases} \frac{1}{n-p}\sum_{i=1}^{n-p} \hat{u}_i \hat{u}_{i+j}, & \text{for } 0 \leqslant j \leqslant p \leqslant n \\ 0 & \text{for } j > p. \end{cases} \tag{5}$$

To show that $\hat{\beta}_{\text{FGLS-AR}}$ is efficient and normally distributed it is necessary to make the following assumptions on X and u:

Assumptions.

(H1) There exists $\{a_i\}$ such that for $i \geqslant 1$,

$$\sum_{l=0}^{\infty} a_l u_{i-l} = \epsilon_i, \tag{6}$$

where $\{\epsilon_i\}$ are i.i.d. with $E(\epsilon_i) = 0$, $E(\epsilon_i^2) = \sigma^2$, $E(|\epsilon_i|^4) < \infty$.

(H2) $\sum_{i=0}^{\infty} |a_i| < \infty$.

(H3) $A(z) = 1 + \sum_{l=1}^{\infty} a_i z^i$ is nonzero for z such as $|z| \leqslant 1$.

(H4) $\lim_{n\to\infty} n^{-\alpha} X'\Sigma_n^{-1} X = A_0$ for some $\alpha > 0$, where A_0 is finite and non-singular.

(H5) Let h_i be the i-th diagonal element of the matrix $H \equiv X(X'X)^{-1}X'$. As $n \to \infty$, $p\, h_n^* = O(1)$, where $p = o(n^{1/4})$ and $h_n^* = \max\{h_i\}$.

Assumptions (H1)–(H3) are the standard assumptions (Kunitomo and Yamamoto, 1985; Bhansali, 1986; Bhansali and Papangeloup, 1991) imposed for stationarity of $\{u_i\}$ so that the covariance function is absolutely summable

$$c_r \equiv 2\sum_{j=0}^{\infty} |r_j| < \infty. \tag{7}$$

These assumptions also guarantee that $\{u_i\}$ has an infinite order moving average representation (Wold's decomposition theorem (Amemiya, 1973)),

$$u_i = \sum_{l=0}^{\infty} b_l \epsilon_{i-l} \tag{8}$$

for some $\{b_i\}$, where

$$\sum_{i=0}^{\infty} |b_i| < \infty. \tag{9}$$

Assumption (H4) is commonly required in constructing GLS estimators (see Judge et al., 1985). It is a sufficient condition used to show the consistency and the asymptotic normality of $\hat{\beta}_{\text{FGLS-AR}}$. Note that we make this assumption in a very unrestrictive way such that α does not have to be 1 to include a wide class of dynamics of regressors. For example, it includes the time trend regressor as a special case (i.e., $\alpha = 3$). It rules out, however, regressions with slowly varying regressors, in which case the regressors are asymptotically collinear (Phillips, 2007). We further note that the value of α could be different for each regressor. However, we present the results only for the same normalization in terms of n^α to avoid more complications in notation.

Assumption (H5) focuses on the behavior of H, which is sometimes referred to as the Hat Matrix. It ensures that $\hat{u}$ converges to u in distribution as $p \to \infty$ with the sample size increasing

at an appropriate rate (Davidson and Mackinnon, 1993). Note that $\sum_{i=1}^{n} h_i = k$. The average size of the diagonal element of the hat matrix, then, is k/n. Assumption (H5) is not restrictive and should be satisfied by design matrices associated with most real world applications unless the sample contains some extreme outliers.

In Theorem 1, we show that GLS-AR estimators are asymptotically efficient as GLS estimators, as both the number of AR lag, p, and the number of observations, n, increase together so that $p = o(n^{1/4})$ in the limit. In the next Section of Finite-sample Properties, we provide practical guidelines of how to select p with given n in finite samples.

Theorem 1. *Assume that (H1)–(H5) are satisfied. Let p be chosen as a function of n so that $p \to \infty$ and $p = o(n^{1/4})$ as $n \to \infty$. Then*

$$n^{\alpha/2}(\hat{\beta}_{FGLS\text{-}AR} - \beta) \xrightarrow{d} N(0, A_0^{-1}), \tag{10}$$

where A_0^{-1} is the same as the variance of the GLS estimator.

We note that if $\sigma_{i,j} = r_{|j-i|} = 0$ for $|j - i| > p_0$ for some finite p_0, then Theorem 1 remains valid for any $p \geqslant p_0$.

To prove Theorem 1 we must show that for any given vector $\tau \in R^k$,

$$n^{\alpha/2}\tau'(\hat{\beta}_{\text{FGLS-AR}} - \beta) \xrightarrow{d} N(0, \tau' A_0^{-1}\tau). \tag{11}$$

Although we leave the details of the proof to the Appendix, there we show that by decomposing $(\hat{\beta}_{\text{FGLS-AR}} - \beta)$ as the sum of three terms,

$$\begin{aligned}\hat{\beta}_{\text{FGLS-AR}} - \beta &= (X'\hat{\Sigma}_n^{-1}X)^{-1}X'\hat{\Sigma}_n^{-1}u \\ &= (X'\Sigma_n^{-1}X)^{-1}X'\Sigma_n^{-1}u \\ &\quad + \left((X'\hat{\Sigma}_n^{-1}X)^{-1} - (X'\Sigma_n^{-1}X)^{-1}\right)X'\Sigma_n^{-1}u \\ &\quad + (X'\hat{\Sigma}_n^{-1}X)^{-1}X'(\hat{\Sigma}_n^{-1} - \Sigma_n^{-1})u,\end{aligned} \tag{12}$$

we can evaluate each term separately as the residual autoregressive order p increases at an appropriate rate dependent on sample size.

Thus, by denoting the three terms in (12) as $W_{n,1}$, $W_{n,2}$, and $W_{n,3}$, respectively, it will suffice to show that:

Lemma 1. *Under (H1)–(H5), we have*

$$n^{\alpha/2}\tau' W_{n,1} \xrightarrow{d} N(0, \tau' A_0^{-1}\tau). \tag{13}$$

Lemma 2. *Assume (H1)–(H4) are satisfied. If $p \to \infty$ and $p = o(n^{1/2})$, then*

$$n^{\alpha/2}\tau' W_{n,2} \xrightarrow{p} 0. \tag{14}$$

Lemma 3. *Assume (H1)–(H4) are satisfied. If $p \to \infty$ and $p = o(n^{1/4})$, then*

$$n^{\alpha/2}\tau' W_{n,3} \xrightarrow{p} 0. \tag{15}$$

Lemma 1 corresponds to the case for which the covariance matrix Σ_n is known. Lemmas 2 and 3 ensure that $\hat{\Sigma}_n$ can be replaced by Σ_n in $(\hat{\beta}_{\text{FGLS-AR}} - \beta)$.

Although the basic idea of the method is simple, implementing it in the time domain is a challenging task. The difficulty revolves in finding an efficient estimate for $\hat{r}_j - E(\hat{r}_j)$, which depends both on the order of autoregression approximation and the dynamics of regressors. In the Appendix we show how to overcome this problem by extending the classic results on estimates of autocorrelation functions in pure time series settings (for example, results in Fuller, 1996) to regression settings.

2.2 Testing Hypotheses of Regression Coefficients

The results from Theorem 1 can be used to construct an F-statistic

$$F_{j,n-k} = \frac{(R\hat{\beta}_{\text{FGLS}} - \mu)'[R(X'\hat{\Sigma}_n^{-1}X)^{-1}R']^{-1}(R\hat{\beta}_{\text{FGLS-AR}} - \mu)/j}{\hat{u}'\hat{\Sigma}_n^{-1}\hat{u}/(n-k)}, \tag{16}$$

for testing j linear restrictions of the form

$$H_0 : R\beta = \mu, \tag{17}$$

against the alternative

$$H_1 : R\beta \neq \mu \tag{18}$$

where R is a $(j \times k)$ matrix with each row representing a linear restriction on the coefficient vector and μ is a $(j \times 1)$ vector.

In particular, when $j = 1$, the t-statistic

$$t_{n-k} = \frac{n^{\alpha/2}(\hat{\beta}_{\text{FGLS-AR}})_i}{\sqrt{\hat{u}'\hat{\Sigma}_n^{-1}\hat{u}/(n-k)(X'\hat{\Sigma}_n^{-1}X/n^{\alpha})_{ii}^{-1}}} \tag{19}$$

can be used to test the null hypothesis $\beta_i = 0$.

We note that the $F_{j,n-k}$ test in (16) and the t_{n-k} statistic in (19) are based on the asymptotic results in Theorem 1. Similarly to the usual inference procedures based on feasible GLS or HAC estimators, the exact distributions of these test statistics depend on the data and the parameters, and do not, in general, follow F and t distributions in finite samples. They are viewed as large-sample approximations whose quality improves as the sample size increases. In the finite-sample simulations considered in the next section we will use the critical values from the F and the standard normal distributions to approximate the true distributions of the test statistics $F_{j,n-k}$ and t_{n-k}.

3 Finite-Sample Properties

In this section we provide the results from an extensive Monte Carlo study contrasting the empirical sizes and power of test statistics based on OLS, feasible GLS, feasible GLS-AR, and HAC estimators.

3.1 The Simulation Designs

In the previous section we have shown that if p is chosen to be sufficiently large, the errors introduced by using the FGLS-AR(p) and the estimated covariance matrix have no impact on the limiting distribution of the regression estimates nor their corresponding test statistics. In this section we will present results from an exhaustive Monte Carlo study contrasting the finite sample performance of several estimators. We will compare the resulting empirical sizes and the power of tests based on the t-statistics of regression coefficients obtained from various estimation procedures to their nominal sizes of 1% and 5%. This latter approach involves obtaining consistent standard errors for the OLS estimates of the regression (Newey and West, 1987). This is accomplished through the weighted estimator of $X'\Sigma_n X$, i.e., the Bartlett kernel, $w(\cdot)$,

$$\hat{\Gamma} = n^{-1} \sum_{t=1,s=1,|t-s|\leqslant L}^{n} w(t-s)\hat{u}_t\hat{u}_s X[t,.]'X[s,.], \tag{20}$$

where $\hat{u}_t = y_t - X[t,.]\hat{\beta}_{\text{OLS}}$, and L is an arbitrarily selected truncation value for which the magnitude of the $cov(u_t, u_s)$ is assumed to be inconsequential (Mittelhammer et al., 2000). While appealing in terms of their asymptotic properties, the finite sample performances of the inference procedures based on HAC are known to be unreliable, especially when there is strong autocorrelation in the data (e.g., Kiefer and Vogelsang, 2005). Although there have been some more recent developments that attempt to address the size distortion problem in finite samples (e.g., Kiefer and Vogelsang, 2002; PC Sun P Y and Jin, 2011; Politis, 2011), those new procedures are complicated and not readily available in statistical packages. Due to its popularity in the empirical research, we only contrast the original version of HAC (by Newey and West, 1987) with our approach.

We generated, for sample sizes ranging from 50 to 1,000 observations, 1,000 realizations for each of a variety of stationary and invertible Gaussian ARMA structures with varying parameter values as the residuals of a regression model with one exogenous variable generated by an AR(1) process. The parameter values for the residual ARMA structures were chosen to not only conform with other previously published studies such as Engle (1974), Pukkila et al. (1990), and Zinde-Walsh and Galberaith (1991), but also to provide a representative set of examples of possible autocorrelated error structures in regression models. Then, we created the regression model

$$y_t = 2.0 + 0.5x_t + u_t, \tag{21}$$

where u_t follows various ARMA(p, q) series and the generating process for the exogenous variable x_t followed an AR(1) process, $(1 - \zeta B)x_t = w_t$, with $w_t \sim IN(0, \sigma^2)$, and $\text{E}(u_t, w_s) = 0$, for all t and s, and $\zeta = \{-0.5, 0.0, 0.5\}$. Only one set of random numbers was generated for each of the AR(1) model structures of the exogenous variable used in (21). Breusch (1980) has shown that for a fixed regressor the distribution of $(\hat{\beta}_{\text{FGLS}} - \beta)/\sigma$ does not depend on β and σ^2. In addition, the result also holds if the covariance matrix is misspecified (Koreisha and Fang, 2001). This implies that in simulation studies, only one point in the parameter space (β, σ^2) needs to be considered for feasible GLS and GLS-AR estimators. We let $\sigma = 1$ in our simulations. For the HAC approach, Newey and West (1987) recommended setting the truncation value at $L_{NW} = 4(n/100)^{2/9}$ while others such Stock and Watson (2003) and Kiefer and Vogelsang (2005) have suggested using $L_{SW} = 0.75n^{1/3}$. For sample sizes considered in this study the two truncation rules differ only for sample sizes of 200 and 1,000 observations. Since there are no substantial differences in empirical sizes for $n = 1{,}000$, we will only report empirical size values based on both L_{NW} and L_{SW} for the sample size of 200.

For the approach based on AR(p) corrections, we found that $\hat{\beta}_{\text{FGLS-AR}}$ has the best performance when p is set near $4n^{1/6}$ or $2n^{1/4}$. For the sake of brevity, since for sample sizes considered here the integer difference between $4n^{1/6}$ and $2n^{1/4}$ is typically just 1, we will only report empirical sizes for just one value for p, denoted as $\tilde{p}$.

3.2 Empirical Sizes of Tests (Nominal Size 5%)

Figure 1 contains the empirical sizes of the t-statistics for the regression coefficient β_1 obtained from feasible GLS-AR procedures for selected ARMA(p, q) residual autocorrelation structures for wide range of sample of sizes. Table 1 provides the comparisons of the empirical sizes of OLS, HAC, feasible GLS, and feasible GLS-AR procedures. Again, for the sake of brevity and to avoid a great deal of repetitiveness, the figure and table do not include all permutations of residual serial correlation and nominal sizes. We have also omitted the results of the simulations for β_0

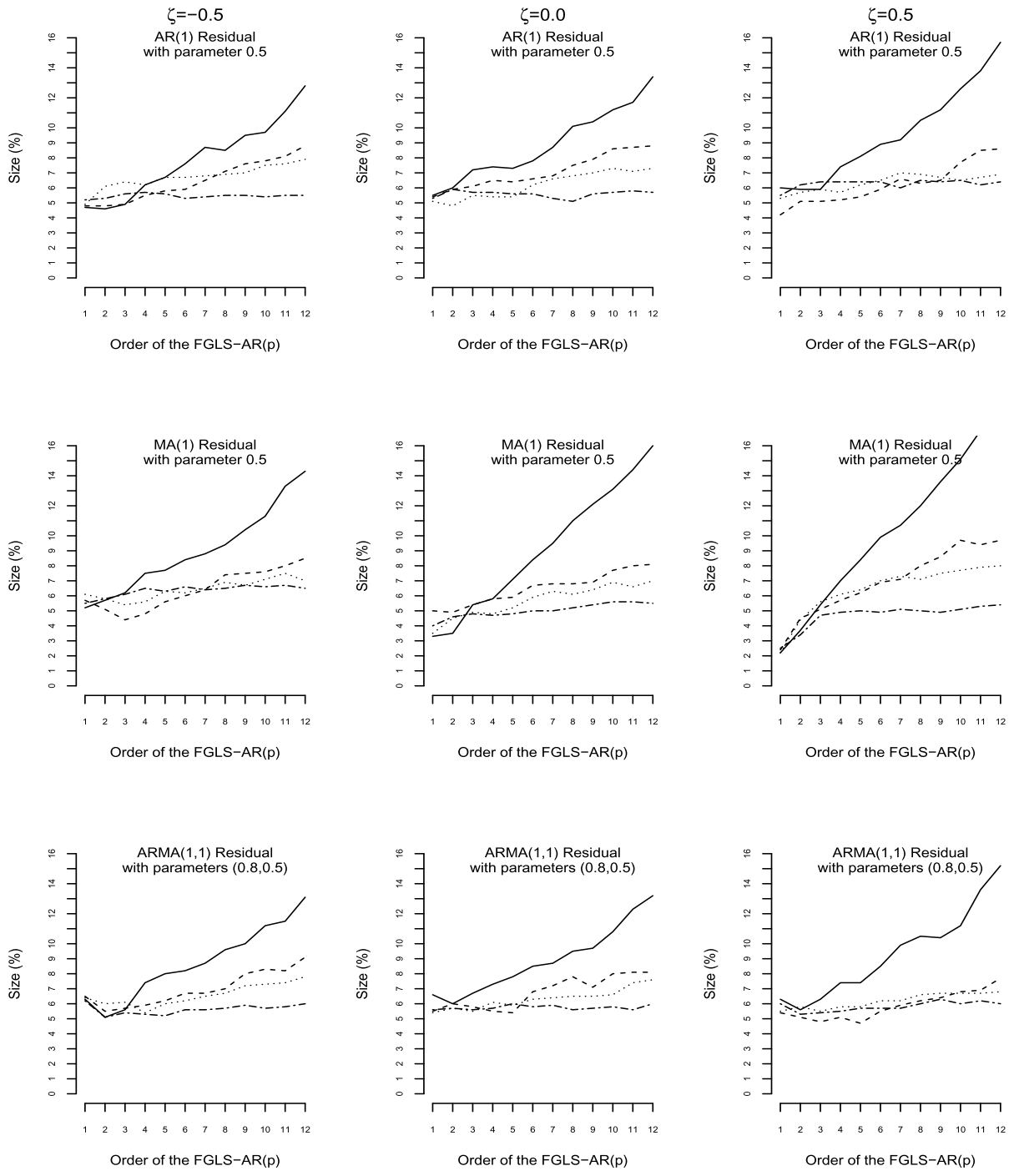

Figure 1: Empirical size of the t-statistics for β_1 based on GLS-AR(p) estimator for a variety of serial correlations and sample sizes. Nominal size is 5%. ζ is the AR coefficient associated with the generating process for the exogenous variable x_t. The plots in the left, the middle, and the right columns are results based on $\zeta = -0.5$, 0.0, and 0.5, respectively. Legend: solid line for $n = 50$; dashed line for $n = 100$; dotted line for $n = 200$; and dashed/dotted line for $n = 500$.

Table 1: Empirical size of the t-statistic for β_1 based on different estimators for a variety of serial correlations (nominal size 5%).

	$\zeta = -0.5^a$					$\zeta = 0.0^a$					$\zeta = 0.5^a$				
	Estimators^b					Estimators^b					Estimators^b				
n	OLS	FGLS	HAC	FGLS-AR(1)	FGLS-AR($\tilde{p}$)	OLS	FGLS	HAC	FGLS-AR(1)	FGLS-AR($\tilde{p}$)	OLS	FGLS	HAC	FGLS-AR(1)	FGLS-AR($\tilde{p}$)
Residual Serial Correlation: AR(1) $\phi_1 = 0.5$															
50	1.0	4.7	9.7	4.7	5.6	4.8	5.5	11.6	5.5	6.8	11.3	6.0	14.2	6.0	6.5
100	0.8	4.8	9.1	4.8	5.8	4.3	5.4	8.9	5.4	6.4	11.7	4.2	12.3	4.2	5.4
200	1.1	4.9	8.4	4.9	6.2	5.2	5.1	8.7	5.1	6.2	13.1	5.3	10.9	5.3	6.4
500	1.4	5.2	7.0	5.2	5.5	6.4	5.3	8.2	5.3	5.8	13.4	5.5	8.5	5.5	6.2
1000	1.5	5.3	6.6	5.2	5.1	6.5	5.2	7.2	5.3	5.3	13.3	5.1	6.8	5.6	5.4
Residual Serial Correlation: MA(1) $\theta_1 = 0.5$															
50	9.0	8.9	12.2	5.2	6.9	4.2	9.9	10.1	3.3	5.8	1.2	14.3	8.0	2.2	7.0
100	10.7	6.0	12.1	5.7	5.6	5.7	6.7	10.1	5.0	5.9	1.5	7.1	8.9	2.4	6.2
200	11.7	5.7	10.9	6.1	6.5	5.5	5.6	8.2	3.5	5.7	1.2	6.9	6.7	2.2	6.7
500	9.6	5.3	7.6	5.5	6.7	5.5	5.3	7.7	4.0	5.6	1.1	5.6	7.1	2.5	5.3
1000	10.2	5.2	6.7	5.5	5.4	5.3	5.1	6.2	4.7	5.2	1.5	5.5	6.1	3.1	5.0
Residual Serial Correlation: ARMA(1,1) $\phi_1 = 0.8$; $\theta_1 = 0.5$															
50	2.6	6.8	11.2	6.3	6.3	5.3	6.5	12.1	6.6	7.3	10.3	6.5	15.6	6.3	6.8
100	1.9	6.3	9.8	6.5	6.2	4.2	5.9	9.5	5.5	5.4	10.5	4.5	11.7	5.4	4.7
200	2.1	5.8	9.3	6.4	6.5	5.8	5.7	9.8	5.4	6.4	13.1	5.2	11.8	5.5	6.2
500	2.7	5.5	7.3	6.2	5.8	5.7	5.6	8.0	5.6	5.6	12.9	5.4	8.9	6.0	6.2
1000	3.0	5.2	6.2	5.8	5.3	5.5	5.2	7.2	5.1	4.9	12.2	4.8	6.7	5.3	5.4
Residual Serial Correlation: ARMA(1,1) $\phi_1 = 0.8$; $\theta_1 = -0.7$															
50	0.4	14.6	5.5	0.2	3.3	4.3	13.1	10.8	1.1	3.6	19.5	13.4	16.3	3.4	4.3
100	0.2	9.4	7.9	0.1	3.7	4.3	8.6	10.8	0.7	4.6	19.8	8.6	13.5	3.0	4.4
200	0.1	6.1	7.2	0.2	4.6	5.4	7.7	9.8	1.0	5.8	21.7	7.2	11.5	3.6	5.9
500	0.3	5.3	7.3	0.6	5.0	5.5	4.8	8.3	1.7	5.7	20.7	6.1	8.5	4.2	5.6
1000	0.8	5.3	7.1	1.1	4.8	5.4	4.7	7.0	2.1	5.1	19.0	5.8	7.2	4.5	5.5
Residual Serial Correlation: ARMA(1,2) $\phi_1 = -0.8$; $\theta_1 = 1.4$, $\theta_2 = -0.6$															
50	21.4	14.3	15.3	3.0	3.4	4.9	13.8	9.4	0.8	3.5	0.2	15.9	5.4	0.0	3.4
100	22.2	10.9	14.6	3.7	3.6	6.2	12.6	11.2	1.4	4.2	0.3	13.8	7.9	0.1	3.0
200	21.4	9.1	11.6	3.4	5.0	6.8	9.3	9.5	1.1	4.4	0.1	9.4	7.6	0.2	3.9
500	20.2	6.1	7.8	3.3	5.1	5.5	5.7	6.9	1.0	5.0	0.2	5.9	5.6	0.7	5.9
1000	19.8	5.4	7.9	3.5	4.6	5.8	5.5	6.1	1.7	4.9	0.8	5.4	5.5	1.8	5.6
Residual Serial Correlation: ARMA(2,1) $\phi_1 = 1.4$, $\phi_2 = -0.6$; $\theta_1 = -0.8$															
50	0.4	14.6	5.4	0.0	2.7	4.3	13.2	10.7	0.9	3.6	19.6	12.5	16.5	3.1	3.7
100	0.1	11.6	7.8	0.1	3.2	4.5	12.5	10.6	0.5	3.9	20.2	12.0	13.7	2.8	3.9
200	0.0	7.0	7.4	0.0	4.1	5.4	7.8	9.7	1.1	4.7	21.6	8.6	11.6	3.7	5.1
500	0.2	5.4	7.4	0.5	4.9	5.7	5.5	8.1	1.4	4.9	21.0	5.5	8.6	4.1	5.3
1000	0.8	5.0	5.8	1.1	4.9	5.5	5.4	7.0	2.7	5.1	20.8	5.4	8.5	4.3	5.2

a. ζ=AR coefficient associated the generating process for the exogenous variable x_t.

b. Estimators: OLS=Ordinary Least Squares; AR(1)=Estimated Generalized Least Squares Estimator with AR(1) Correction for the Residual Serial Correlation; HAC=Heteroscedasticity and Autocorrelation Consistent Estimator (Truncation Rule $L_{NW}(L_{SW})$; for $n = 200$, $L_{NW} = 5$ ($L_{SW} = 4$)); FGLS=Feasible Generalized Least Squares Estimator with the Corrected Residual Autocorrelation Structure; AR($\tilde{p}$)=Feasible Generalized Least Squares Estimator with AR($\tilde{p}$) Correction for the Residual Serial Correlation ($\tilde{p} = [2n^{1/4}]$).

because they are similar to what we observed for β_1. These supplementary simulation results are available from the authors upon request.

Figure 1 shows the empirical size of the tests based on t-ratios derived from estimates of the regression coefficient β_1 and its corresponding standard error when the nominal size was set at 5% using the feasible GLS-AR(p) procedure. Each graph in this figure summarizes the results associated with 48,000 trials. As can be seen when $p = \tilde{p}$ the empirical sizes were generally close to the specified significance levels (results for 1% can be obtained from the authors) regardless of sample size or stochastic nature of X. Even for serial correlation structures that would clearly benefit from applying a lower AR order such as the AR(1) parametrization or the nearly white noise ARMA(1,1) parametrization with $\phi_1 = 0.8$ and $\theta_1 = 0.5$, the differences in empirical sizes obtained from lower AR orders were not very large. However, for other parametrizations setting the order of the autoregressive correction either too low or too high (relative to $\tilde{p}$) can lead to substantial under or over rejection of the null hypothesis that $\beta_1 = 0$. Using a feasible GLS-AR(1) correction, for instance, as is so frequently done in practice, can severely under estimate the size of the test. For example, for the ARMA (1,1) residual structure with $\phi_1 = 0.8$ and $\theta_1 = -0.7$, and for the ARMA(1,2) parametrization, $\phi_1 = -0.8; \theta_1 = 1.4$, $\theta_2 = -0.6$, such a correction would reject the null hypothesis well below the nominal 5% regardless of the stochastic nature of the exogenous variable or the sample size. Too large an order for the AR correction, on the other hand, can result in over-rejecting the null hypothesis. For the MA(1) parametrization with $\theta_1 = 0.5$ with sample size of 100 observations, for example, when p is set at 10, the null hypothesis is rejected between 7.6% and 9.7% of the time depending on the stochastic nature of the exogenous variable. As one would expect, however, as the sample size increases the variation in the empirical size of the tests is not as pronounced as p increases, i. e., the slopes of the lines tend to flatten out approaching the nominal size.

Table 1 contrasts the empirical size of the t statistic associated with the regression coefficient β_1 among the different procedures. As noted earlier for sample sizes typically available to model builders the feasible GLS-AR procedure yields empirical sizes that are close to their nominal sizes. The empirical sizes based on feasible GLS are generally higher (often considerably higher) than the nominal size of the test (5% reported here), and dependent on the structure of the serial correlation. The percentage of the time the null is over-rejected, however, decreases as the sample size increases. For the ARMA(2,1) parametrization with $\phi_1 = 1.4$, $\phi_2 = -0.6; \theta_1 = -0.8$, for instance, the ranges for the empirical sizes associated with samples from 50 to 1,000 observations were between 14.6% and 5.0%, 13.2% and 5.4%, and 12.5% to 5.4%, for $\zeta = \{-0.5, 0.0, 0.5\}$, respectively, whereas for the AR(1) parametrization with $\phi_1 = 0.5$ and for the processes relative close to being white noise such as the ARMA(1,1) parametrization $\phi_1 = 0.8$ and $\theta_1 = 0.5$ the variations in the empirical sizes were much smaller and closer to their nominal sizes. These observations regarding the tendency of GLS estimators to over-reject the null hypothesis are in line with previous findings that show that although the parameter estimates obtained from feasible GLS for such sample sizes are efficient in terms of mean square error (Koreisha and Fang, 2001), their corresponding standard errors can be seriously underestimated (Kennedy, 1988). As sample sizes increase the empirical sizes approach their nominal sizes. For the ARMA(2,1) parametrization just discussed, for instance, the empirical sizes (averaged across ζ) go down to 5.5% for $n = 500$ and 5.3% for $n = 1000$. The results show that although GLS is asymptotically efficient, it requires relatively large samples (e.g., $n \geqslant 500$) for the t statistic based on it to converge to a Gaussian distribution.

As expected, since HAC estimates are known to be downward biased in finite samples (Kiefer et al., 2000; and Kiefer and Vogelsang, 2005), the HAC procedure tended to over-reject

the null hypothesis. As can be seen the resulting empirical sizes were substantially higher than their nominal sizes (particularly for samples sizes less than or equal to 100 observations), and very dependent on the structure of the serial correlation as well as the stochastic nature of the exogenous variable. With few exceptions the degree of over-rejection decreased as the sample size increased. For the AR(1) parametrization with $\phi_1 = 0.5$, for example, the ranges for the empirical sizes associated with samples from 50 to 1,000 observations were between 9.7% and 6.6%, 11.6% and 7.2%, and 14.2% and 6.8%, for $\zeta = \{-0.5, 0.0, 0.5\}$, respectively, whereas for the ARMA(1,2) parametrization with $\phi_1 = -0.8; \theta_1 = 1.4, \theta_2 = -0.6$, the corresponding ranges were from 15.3% to 7.8%, 9.4% to 6.1%, and 5.4% to 5.5%, respectively. For sample sizes of 200 observations the L_{SW} truncation rule appeared to be marginally better than the L_{MW} rule.

The variation in the size of the test based on OLS was much greater than the other procedures, and very dependent on the stochastic nature of X. When $\zeta = 0.0$, i.e., when X was generated as an independently, identically distributed (i.i.d.) random variable, the empirical size of OLS t-tests were near the nominal size, but when X was stochastic, i.e., when $\zeta = \{-0.5, 0.5\}$ the empirical size of the tests varied considerably from their nominal sizes with no evident discernable pattern leading to over or under rejection of the null hypothesis. For example, for the ARMA(1,1) parametrization with $\phi_1 = 0.8$ and $\theta_1 = -0.7$, the empirical size of the OLS test oscillated in the range from less than 1% to 21% while for the ARMA(1,2) parametrization with $\phi_1 = -0.8; \theta_1 = 1.4, \theta_2 = -0.6$ the oscillation was in the reverse direction from approximately 22% to less than 1% for $\zeta = \{-0.5, 0.5\}$, respectively. Further increases in the sample size did not appear to impact the variation of the empirical sizes of OLS tests. This very interesting finding regarding the impact that the stochastic nature of the exogenous variable has on the size of the test may shed some light on why authors such as Judge et al. (1985) and Choudhury et al. (1999) may have admonished practitioners that "OLS may often be better than assuming another incorrect truncation of the actual process." Increasing the lag length of the autoregressive correction, as we have already observed results in more efficient parameter estimates as well as standard errors.

A more detailed explanation why the variation in the size of the test for β_1 based on OLS is dependent on the stochastic nature of X can be obtained from the authors upon request. For a single regressors case, if the residual follows an AR(1) process with autocorrelation ρ, for example, it is easy to show that the variance of the OLS estimate for β_1 is under-estimated (over-estimated) when ρ and ζ have the same (opposite) sign, and that the order of the magnitude of the bias can be considerable.

3.3 Empirical Power of Tests (Nominal Size 1%)

Table 2 contains results associated with the power of the tests with nominal size of 1% for the regression coefficient β_1, obtained from OLS, HAC, feasible GLS, and feasible GLS-AR procedures. From Table 2 we see that although the power of the tests for β_1 based on all procedures improved as sample size increased, for some serial correlation structures, power of the OLS and HAC tests (unlike those of feasible GLS and feasible GLS-AR) did not reach 100% even for relatively large sample sizes (e.g., $n \geqslant 500$). The power of the tests based on feasible GLS and feasible GLS-AR(p) with $p = \tilde{p}$ was typically substantially higher than that of the OLS or HAC tests regardless of sample size or stochastic nature of the exogenous variable. For sample sizes greater than or equal to 100 observations the power of feasible GLS and feasible GLS-AR (even though not shown here, also for other order p) nearly always approached 100% (ranging typically from 96% to 100%), whereas the ranges for the power of the OLS and HAC tests were

Table 2: Empirical power of the t-statistic for β_1 based on different estimators for a variety of serial correlations (nominal size 1%).

	$\zeta = -0.5^a$					$\zeta = 0.0^a$					$\zeta = 0.5^a$				
	Estimators^b					Estimators^b					Estimators^b				
n	OLS	FGLS	HAC	FGLS-AR(1)	FGLS-AR($\tilde{p}$)	OLS	FGLS	HAC	FGLS-AR(1)	FGLS-AR($\tilde{p}$)	OLS	FGLS	HAC	FGLS-AR(1)	FGLS-AR($\tilde{p}$)
Residual Serial Correlation: AR(1) $\phi_1 = 0.5$															
50	81.5	98.2	93.2	97.9	96.3	63.9	86.6	73.8	84.1	82.9	72.0	75.7	69.4	75.7	74.3
100	99.4	99.9	100.0	99.9	100.0	95.1	99.5	94.8	99.5	99.1	96.5	97.9	91.2	97.9	97.8
200	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0	99.9	99.9
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Residual Serial Correlation: MA(1) $\theta_1 = 0.5$															
50	75.1	81.6	73.6	72.8	75.9	64.4	87.4	72.1	78.4	82.6	79.1	98.4	91.6	94.8	97.2
100	97.0	99.2	95.2	97.5	98.5	95.3	99.8	96.1	99.2	99.7	99.7	100.0	99.7	100.0	100.0
200	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Residual Serial Correlation: ARMA(1,1) $\phi_1 = 0.8$; $\theta_1 = 0.5$															
50	84.6	93.2	91.7	94.7	92.2	70.3	80.9	77.0	80.2	78.3	75.3	78.9	75.3	75.8	76.5
100	99.4	98.9	99.7	99.9	99.8	95.7	98.4	96.4	98.9	98.1	96.8	98.7	94.1	98.3	98.3
200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.6	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.6	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.6	100.0	100.0
Residual Serial Correlation: ARMA(1,1) $\phi_1 = 0.8$; $\theta_1 = -0.7$															
50	11.6	100.0	60.6	99.2	100.0	14.0	99.7	30.0	88.1	99.5	29.3	98.2	26.2	65.6	95.7
100	34.1	100.0	84.6	100.0	100.0	29.3	100.0	44.2	99.8	100.0	43.2	99.9	31.3	96.2	99.9
200	78.9	100.0	97.8	100.0	100.0	56.0	100.0	65.6	100.0	100.0	65.1	100.0	45.0	100.0	100.0
500	100.0	100.0	100.0	100.0	100.0	93.4	100.0	93.3	100.0	100.0	92.3	100.0	76.5	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	91.7	100.0	100.0
Residual Serial Correlation: ARMA(1,2) $\phi_1 = -0.8$; $\theta_1 = 1.4$, $\theta_2 = -0.6$															
50	24.2	96.0	16.4	61.7	94.7	9.1	99.9	19.9	83.4	99.3	4.9	100.0	44.9	99.1	100.0
100	37.8	99.7	24.5	95.5	100.0	21.3	100.0	33.0	99.4	100.0	21.8	100.0	75.5	100.0	100.0
200	57.9	100.0	35.2	100.0	100.0	48.0	100.0	56.0	100.0	100.0	67.3	100.0	94.9	100.0	100.0
500	88.9	100.0	68.3	100.0	100.0	90.4	100.0	90.2	100.0	100.0	99.6	100.0	100.0	100.0	100.0
1000	97.9	100.0	87.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Residual Serial Correlation: ARMA(2,1) $\phi_1 = 1.4$, $\phi_2 = -0.6$; $\theta_1 = -0.8$															
50	8.2	99.7	56.8	98.7	100.0	12.6	99.5	26.0	85.2	100.0	27.8	98.6	24.9	61.0	96.0
100	27.7	99.9	81.4	100.0	100.0	25.7	100.0	40.0	99.6	100.0	40.7	100.0	28.6	95.3	100.0
200	71.3	100.0	97.1	100.0	100.0	51.2	100.0	61.5	100.0	100.0	61.4	100.0	40.1	100.0	100.0
500	99.8	100.0	100.0	100.0	100.0	91.3	100.0	91.1	100.0	100.0	90.2	100.0	70.6	100.0	100.0
1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.6	100.0	100.0

a. ζ=AR coefficient associated the generating process for the exogenous variable x_t.

b. Estimators: OLS=Ordinary Least Squares; AR(1)=Estimated Generalized Least Squares Estimator with AR(1) Correction for the Residual Serial Correlation; HAC=Heteroscedasticity and Autocorrelation Consistent Estimator (Truncation Rule $L_{NW}(L_{SW})$; for $n = 200$, $L_{NW} = 5$ ($L_{SW} = 4$)); FGLS=Feasible Generalized Least Squares Estimator with the Corrected Residual Autocorrelation Structure; AR($\tilde{p}$)=Feasible Generalized Least Squares Estimator with AR($\tilde{p}$) Correction for the Residual Serial Correlation ($\tilde{p} = [2n^{1/4}]$).

much broader. For instance, across serial correlation structures, for $n = 200$ observations power based on OLS ranged from 34% to 100% (when $\zeta = -0.5$), 54% to 100% (when $\zeta = 0.0$), and 61% to 100% (when $\zeta = 0.5$). The corresponding ranges for HAC were 58% to 100% (when $\zeta = -0.5$), 48% to 100% (when $\zeta = 0.0$), and 39% to 100% (when $\zeta = 0.5$). (The ranges for OLS and HAC were much wider when $n = 100$.)

As would be expected, for sample sizes of 50 observations power levels associated with β_1 were lower for all procedures. For feasible GLS and feasible GLS-AR (with $\tilde{p} = 4$) power levels across all serial correlation structures and ζ were not less than 74%. (For serial correlation structures that were significantly different from white noise the range is much closer to 95% to 100%, but somewhat dependent on the stochastic nature of X.)

For this sample size, variation in power levels based on the feasible GLS-AR(1) correction was more pronounced than with feasible GLS and feasible GLS-AR(4). For some parametrizations power levels based on AR(1) corrections were as low as 61%.

It should also be noted that even for the AR(1) and the nearly white noise ARMA(1,1) ($\phi_1 = 0.8$ and $\theta_1 = 0.5$) serial correlation structures, for which setting a low order autoregressive correction would be more appropriate, the difference in power levels between a feasible GLS-AR(1) and a feasible GLS-AR(4) correction was inconsequential (2%).

For OLS and HAC, power levels manifested a greater dependency on the stochastic nature of X and on the form of the serial correlation than observed for larger sample sizes. Moreover, their magnitudes were generally much lower than those of feasible GLS and feasible GLS-AR(p) with $p = \tilde{p}$. Across serial correlation structures, for $n = 50$ observations, power based on OLS ranged from 8% to 84% (when $\zeta = -0.5$), 78% to 84% (when $\zeta = 0.0$), and 61% to 95% (when $\zeta = 0.5$). The corresponding ranges for HAC were 16% to 93% (when $\zeta = -0.5$), 20% to 77% (when $\zeta = 0.0$), and 25% to 92% (when $\zeta = 0.5$).

Note that caution is needed as we look at power since some tests do not maintain their sizes. Nevertheless, these power results combined with our observations regarding the size of the tests indicate that for finite samples, inferences based on tests generated from the feasible GLS-AR correction are more robust than from feasible GLS, HAC, or OLS. Furthermore, they cast serious doubts on the reliability of inferences based on generic, multipurpose procedures such as feasible GLS-AR(1) and HAC. We also conduct a joint test of significance. The results are available at Online Supplement.

4 Empirical Application: An Analysis of Global Temperature Trends

The issue of global warming has received considerable attention over the past two decades. In this section we apply our tests to examine whether the observed changes in temperature are due to natural variability or if we are witnessing an upward trend which is statistically significant.

The data analyzed in this paper (displayed in Figure 2) are the annual global temperature anomaly time series, which incorporate land and marine data, from 1850 to 2012. The original series were constructed by Jones et al. (1986), and have been continually updated and expanded by P. Jones of the Climatic Research Unit at the University of East Anglia, UK. The series are relative to the 1961–1990 means, measured in degrees Celsius. They have been analyzed over different spans of time by several researchers including Bloomfield and Nychka (1992), Wu et al. (2001), and Fomby and Vogelsang (2002).

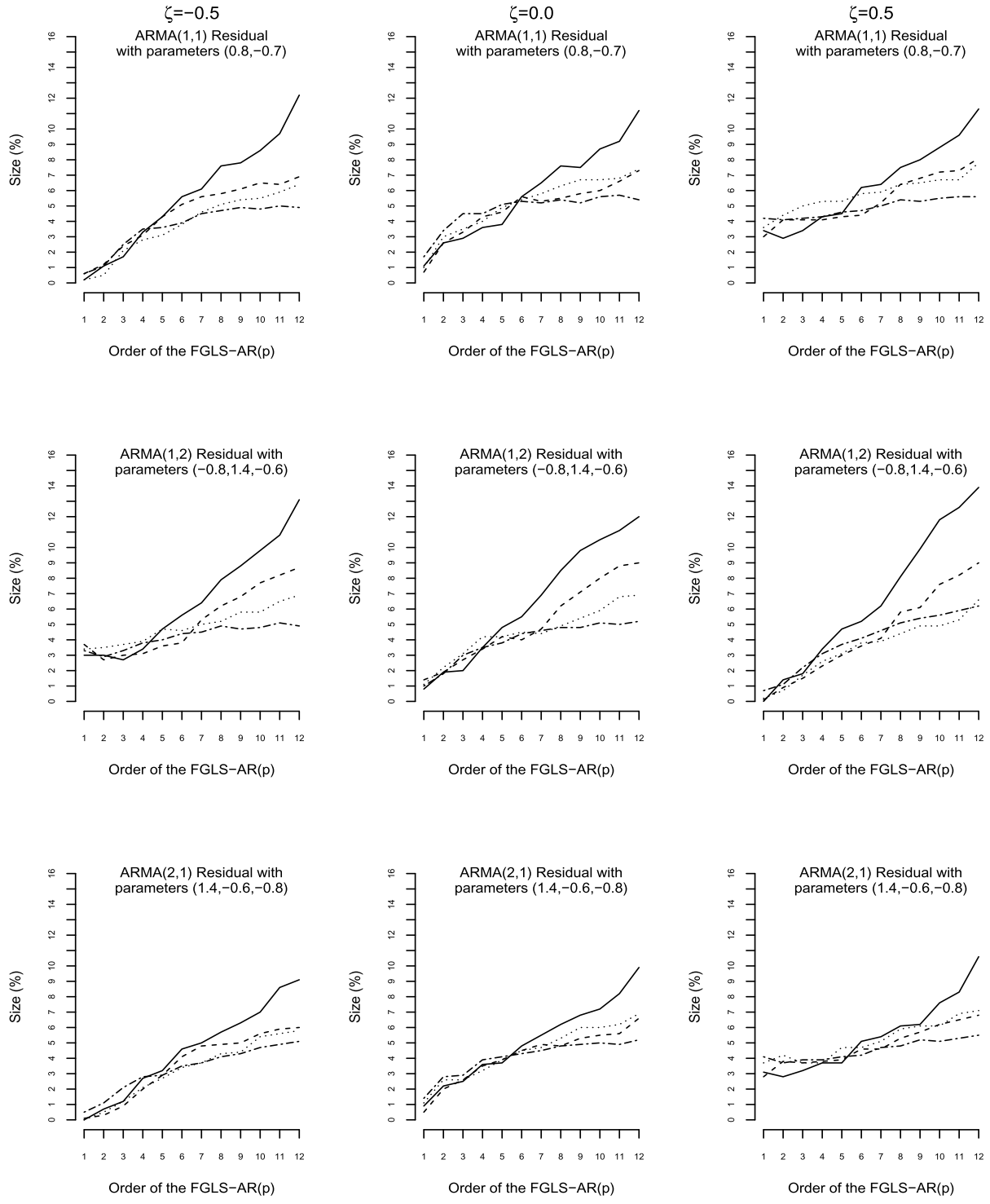

Figure 2: Annual global temperature anomalies from 1850 to 2012 relative to 1961–1990 means and measured in degrees Celsius.

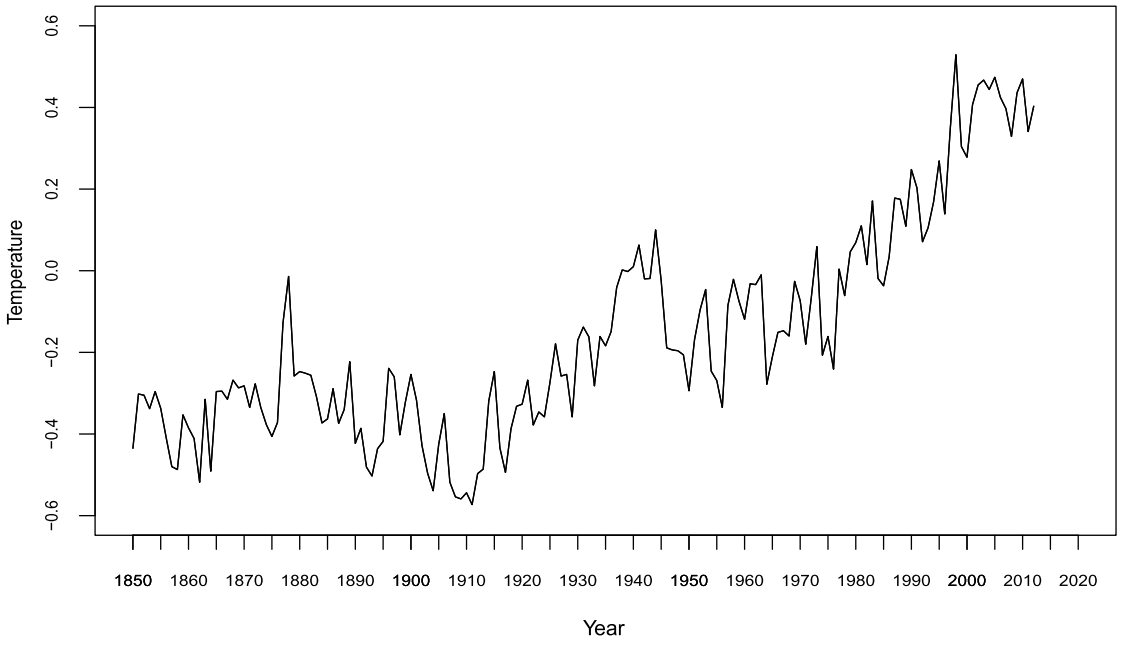

Figure 2: (Continued.)

Table 3: Analysis results of the t-statistic for β_1 (10^{-3}) and standard error (10^{-3}). Based on different estimators for the global temperature series (1850–1945).

	Estimate	Standard Error	t-value	p-value
OLS	2.32	0.46	5.05	< 0.0001
HAC	2.32	0.81	2.86	0.0051
FGLS-AR(1)	2.32	1.10	2.11	0.0352
FGLS-AR(5)	2.21	1.37	1.61	0.1081
FGLS-ARMA(1,2)	2.25	1.35	1.66	0.0967

We will begin by first analyzing the first one-hundred years of the annual global temperature anomalies to illustrate how important it is to account for serial correlation in regression models. Following Bloomfield and Nychka (1992) we consider the deterministic trend model with a stationary error:

$$y_t = \beta_0 + \beta_1 t + u_t,$$

where y_t is the temperature and t is the year. The error u_t was identified as an ARMA(1,2) process based on the Minimum Information Criterion (MINIC) method (Hannan and Rissanen, 1982). Table 3 contains the estimates for β_1 based on OLS, HAC, feasible GLS-AR(1), feasible GLS-AR(p) (where $p = \sqrt{100}/2 = 5$), and feasible GLS-ARMA(1,2) estimators.

As can be seen, although the coefficient estimates are very similar across all methods, the $|t|$ values for significance of β_1 are drastically different across estimation approaches. Tests based on OLS and HAC estimators are statistically significant at 99% level; the test based on FGLS-AR(1) estimator is statistically significant at 95% but not at the 99% level, and tests based on FGLS-AR(5) and FGLS-ARMA(1,2) estimators are statistically significant only at round 90% level, i.e., suggesting that β_1 is insignificant.

Although the OLS estimator is supposedly efficient for this deterministic trend model, tests of significance depend entirely on how the estimate for the error variance is calculated which in this case was severely underestimated. To see this more clearly, let us write the variance of $\hat{\beta}_1$ based on the OLS estimator in terms of autocorrelations of the error,

$$var(\hat{\beta}_1^{\mathrm{OLS}}) = \frac{\gamma_0 + 2\sum_{h=1}^{n-1} U(h)\gamma_h}{n}, \tag{22}$$

where $U(h) = \sum_{t=1}^{n-h}(t+h-\bar{t})(t-\bar{t}) / \sum_{t=1}^{n}(t-\bar{t})^2$ (Lee and Lund, 2004). Since the variance of β_1 based on OLS estimator ignores the term $2\sum_{h=1}^{n-1} U(h)\gamma_h$, which, in our case, is significant and positive, $var(\hat{\beta}_1)$ is underestimated; hence, the test based on OLS over-rejects the null. It is interesting to see that although the test statistic based on HAC or FGLS-AR(1) estimator reduces the frequency of over-rejection, they still underestimate the variance of u_t by quite a bit; thus, one should interpret the evidence based on HAC or FGLS-AR(1) estimators with caution. The results based on either FGLS-AR(5) (or FGLS-ARMA(1,2)) estimator are more robust because the tests take into account more fully the residual autocorrelation. It is also worth noting that estimation of the model using our proposed approach did not require identification of the residual serial correlation structure.

We will now analyze the full data set (1850–2012). A quick glance at Figure 2 suggests that a structural break in the deterministic trend occurred around the first decade of the twentieth century, and the possibility of a shift in the trend function. We will, thus, fit the following model to the data,

$$y_t = \beta_0 + \beta_1 t + \beta_2 D_t + \beta_3 L_t + u_t, \tag{23}$$

where T_B is the structural breaking points and the level dummy variable $D_t = 1$, if $t > T_B$, zero otherwise; and L_t indicates potential shift in the trend function defined as: $L_t = (t - T_B)D_t$, and test the hypotheses that $\beta_1 = 0$ and/or $\beta_3 = 0$. Using the FGLS-AR(6) estimator both Akaike information criterion (AIC) and Bayesian information criterion (BIC) identified 1911 as the year when the structural break in the trend occurred.

From Table 4 which contains the tests of significance for both β_1 and β_3 based on different estimators we do not find any evidence of a warming trend prior to 1911, but that there is a significant upward trend subsequently. The average increase in global temperature after 1911 appears to be about 0.9 degrees Celsius. It is also interesting to note that for the period before 1911 both OLS and HAC erroneously identified a statistically significant downward trend in temperatures at 99% and 95% level, respectively.

This global warming example clearly highlights the importance of accounting for the serial correlation in the error structure of regression models, and demonstrates the simplicity and power of our proposed approach.

5 Conclusions

In this article we have shown that the generalized least squares estimator based on autoregressive corrections (GLS-AR) has the same limiting distribution as GLS, and that parameter inferences on the regression coefficients can be conducted using the usual t and F distributions. Our novel derivations of asymptotic normality in the time domain extend the works of many others in pure time series settings in the frequency domain to time series regression models under very general structures of design matrices. Moreover, based on a large Monte Carlo study we have

Table 4: Analysis results of the t-statistic for β_1 (10^{-3}), β_3 (10^{-3}) and standard error (10^{-3}). Based on different estimators for the global temperature series (1850–2012).

Case for β_1	Estimate	Standard Error	t-value	p-value
OLS	−1.68	0.63	−2.67	0.0084
HAC	−1.68	0.86	−1.96	0.0523
FGLS-AR(1)	−1.51	1.29	−1.17	0.2409
FGLS-AR(6)	−1.58	1.73	−0.91	0.3610

Case for β_3	Estimate	Standard Error	t-value	p-value
OLS	9.16	0.88	10.45	< 0.0001
HAC	9.16	1.27	7.22	< 0.0001
FGLS-AR(1)	9.01	1.80	5.01	< 0.0001
FGLS-AR(6)	9.26	2.44	3.80	0.0001

also demonstrated that for finite samples the GLS-AR(p) procedure with p set at or near $[n^{1/2}]$ or $[2n^{1/4}]$ yields more robust estimators in terms of size and power than OLS, HAC or EGLS. Thus, model builders upon detecting serial correlation in the residuals generated from an OLS regression should simply re-estimate the regression equation using GLS with an AR correction since there is not much to be gained by identifying the form of the residual serial correlation. Automatic procedures for performing this estimation without even providing initial estimates for the autoregressive coefficients can be found in most widely used statistical packages such as SAS (Autoreg), Splus (ARIMA), and SPSS (ARIMA).

One possible extension is to use the framework employed here to the nonlinear regression model with correlated errors. The methods in technical proofs might prove valuable.

Supplementary Material

Online Supplements.

References

Amemiya T (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica*, 41: 723–732. https://doi.org/10.2307/1914092

Berk KN (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, 2: 489–502.

Bhansali RJ (1986). The criterion autoregressive transfer function of Parzen. *Journal of Time Series Analysis*, 7: 79–103. https://doi.org/10.1111/j.1467-9892.1986.tb00487.x

Bhansali RJ, Papangeloup F (1991). Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order. *The Annals of Statistics*, 19: 1155–1162.

Bloomfield P, Nychka D (1992). Climate spectra and detecting climate change. *Climate Change*, 21: 275–287. https://doi.org/10.1007/BF00139727

Breusch T (1980). Useful invariance results for generalized regression models. *Journal of Econometrics*, 13: 327–340. https://doi.org/10.1016/0304-4076(80)90083-4

Chambers M (2013). Jackknife estimation of stationary autoregressive models. *Journal of Econometrics*, 172: 142–157. https://doi.org/10.1016/j.jeconom.2012.09.003

Chambers M, Ercolane JS, Taylor AMR (2014). Testing for seasonal unit roots by frequency domain regression. *Journal of Econometrics*, 178: 243–258. https://doi.org/10.1016/j.jeconom.2013.08.025

Choi I, Kurozumi E (2012). Model selection criteria for the leads-and-lags cointegrating regression. *Journal of Econometrics*, 169: 224–238. https://doi.org/10.1016/j.jeconom.2012.01.021

Choudhury A, Hubata R, Louis R (1999). Understanding time-series regression estimators. *American Statistician*, 53: 342–348.

Cochrane D, Orcutt GH (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44: 32–61. https://doi.org/10.1080/01621459.1949.10483312

Davidson R, Mackinnon JG (1993). *Estimation and Inference in Econometrics*. Oxford University Press, Oxford, UK.

Engle R (1974). Specification of the disturbance for efficient estimation. *Econometrica*, 42: 135–146. https://doi.org/10.2307/1913690

Fomby TB, Vogelsang TJ (2002). The application of size-robust trend statistics to global-warming temperature series. *Journal of Climate*, 15: 117–123.

Fuller WA (1996). *Introduction to Statistical Time Series*. John Wiley & Sons, Inc., New York, NY.

Goldberger AS (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57: 369–375. https://doi.org/10.1080/01621459.1962.10480665

Hannan EJ, Rissanen J (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69: 81–94. https://doi.org/10.1093/biomet/69.1.81

Harvey DI, Leybourne SJ, Taylor AMR (2010). Robust methods for detecting multiple level breaks in autocorrelated time series. *Journal of Econometrics*, 157: 342–358. https://doi.org/10.1016/j.jeconom.2010.02.003

Jones WTML, D P, Wright PB (1986). Global temperature variations between 1861 and 1984. *Nature*, 332: 430–434. https://doi.org/10.1038/322430a0

Judge GG, Griffiths WE, Hill RC, Lütkepohl H, Lee T (1985). *The Theory and Practice of Econometrics*. John Wiley & Sons, New York, NY.

Kadiyala K (1970). Testing for the independence of regression disturbances. *Econometrica*, 38: 97–117. https://doi.org/10.2307/1909244

Kennedy P (1988). *A Guide to Econometrics*. The MIT Press, Cambridge, MA.

Kiefer N, Vogelsang T, Bunzel H (2000). Simple robust testing of regression hypotheses. *Econometrica*, 68: 695–714. https://doi.org/10.1111/1468-0262.00128

Kiefer N, Vogelsang TJ (2005). A new asymptotic theory for heteroscedasticity-autocorrelation robust tests. *Econometric Theory*, 21: 1130–1164. https://doi.org/10.1017/S0266466605050565

Kiefer NM, Vogelsang TJ (2002). Heteroscedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70: 2093–2095. https://doi.org/10.1111/1468-0262.00366

King M (1983). Testing for autoregressive against moving average errors in the linear regression

model. *Journal of Econometrics*, 21: 35–51. https://doi.org/10.1016/0304-4076(83)90118-5

Koreisha S, Fang Y (2001). Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society, Series B*, 63: 515–531. https://doi.org/10.1111/1467-9868.00296

Koreisha S, Pukkila T (1985). Properties of predictors in misspecified autoregressive time series model. *Journal of the American Statistical Association*, 80: 941–950. https://doi.org/10.1080/01621459.1985.10478208

Kunitomo N, Yamamoto T (1985). Properties of predictors in misspecified autoregressive time series model. *Journal of the American Statistical Association*, 80: 941–950. https://doi.org/10.1080/01621459.1985.10478208

Lee J, Lund R (2004). Revisiting simple linear regression with autocorrelated errors. *Biometrika*, 91: 240–245. https://doi.org/10.1093/biomet/91.1.240

Maddala GS (2001). *Introduction to Econometrics*. John Wiley & Sons, New York, NY.

Mittelhammer R, Judge G, Miller D (2000). *Econometric Foundations*. Cambridge University Press, Cambridge, MA.

Newey W, West K (1987). A simple positive semi-definite, heteroskedastic and autocorrelation consistent covariance matrix. *Econometrica*, 55: 703–708. https://doi.org/10.2307/1913610

Phillips PCB (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, 23: 557–614. https://doi.org/10.1017/S0266466607070491

Politis DN (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, 27: 703–744. https://doi.org/10.1017/S0266466610000484

Prais SJ, Winsten CB (1954). Trend estimators and serial correlation. Chicago. Cowles Commission Discussion Paper, No. 383.

Pukkila T, Koreisha S, Kallinen A (1990). The identification of ARMA models. *Biometrika*, 77: 537–549. https://doi.org/10.1093/biomet/77.3.537

Saikkonen P (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7: 21.

Stock J, Watson M (2003). *Introduction to Econometrics*. Addison Wesley, Boston, MA.

Sun Y, Philips PCB, Jin S (2011). Power maximization and size control in heteroscedasticity and autocorrelation robust tests with exponentiated kernels. *Econometric Theory*, 27: 1320–1368. https://doi.org/10.1017/S0266466611000077

Thursby J (1987). OLS or GLS in the presence of specification error? *Journal of Econometrics*, 35: 359–374. https://doi.org/10.1016/0304-4076(87)90033-9

Walker A (1967). Some tests of separate families of hypotheses in time series analysis. *Biometrika*, 54: 39–68.

Wu WB, Woodroofe M, Mentz G (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika*, 88: 793–804. https://doi.org/10.1093/biomet/88.3.793

Zinde-Walsh V, Galberaith J (1991). Estimation of a linear regression model with stationary ARMA(p, q) errors. *Journal of Econometrics*, 47: 333–357. https://doi.org/10.1016/0304-4076(91)90106-N