# Feasible GLS for Time Series Regression*

Pierre Perron† Emilio González-Coya‡

Boston University Morgan Stanley

September 8, 2022; revised August 17, 2024

**Abstract**

We consider a linear regression model with serially correlated errors. It is well known that with fixed regressors Generalized Least-Squares is more efficient than Ordinary Least-Squares (OLS). However, there are usually three main reasons advanced for adopting OLS instead of GLS. The first is that it is generally believed that OLS is valid whether the regressors are exogenous (uncorrelated with past innovations) or not, while GLS is only consistent when dealing with pre-determined regressors (uncorrelated with future innovations). Second, OLS is more robust than GLS. Third, the gains in accuracy can be minor and the inference can be misleading (e.g., bad coverage rates of the confidence intervals). We argue that all three claims are wrong, in general and under some weak conditions. The first contribution is to dispel the fact that OLS is valid with non-exogenous regressors, while GLS is valid only with exogenous regressors. Under some regularity conditions, we show the opposite to be true. The second contribution is to show that GLS is much more robust that OLS even when the regressors are exogenous. By that we mean that even a blatantly incorrect GLS correction can achieve a lower MSE than OLS. The third contribution is to devise a feasible GLS procedure valid whether or not the regressors are exogenous, which achieves a MSE close to that of the correctly specified infeasible GLS. We also briefly address issues related to correcting for heteroskedastic errors.

**Keywords**: Feasible Generalized Least-Squares, Mean-Squared Error, Confidence Intervals, sieve approximation, Non-parametric Methods, Linear Model.

**JEL Classification**: C22

*We acknowledge the computing support from the Boston University Shared Computing Cluster. We are grateful to Zhongjun Qu for useful conversations, as well as the editor in charge of the review process, Michael Jansson, and two referees who went beyond the call of duty to provide many useful comments. We also thank Guiseppe Cavaliere, Ivan Fernández-Val, Jean-Jacques Forneron, Hiroaki Kaido, Guido Kuersteiner, Serena Ng, Marco Olivari and Robert Taylor for useful comments.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

‡Morgan Stanley, 1585 Broadway Avenue, New York, NY 10036 (emilio.gonzalez.coya.sandoval @morganstanley.com).

## 1 Introduction

We consider a linear regression model with serially correlated errors. If the regressors are fixed or strictly exogenous (i.e., uncorrelated with the innovations at all leads and lags), Generalized Least-Squares (GLS) is BLUE, hence more efficient than Ordinary Least-Squares (OLS). If the regressors are pre-determined (i.e., uncorrelated with future innovations), GLS is no longer unbiased but is consistent and asymptotically efficient. With exogenous regressors OLS is consistent, though not efficient. Early work concentrated on fixed regressors or equivalently strictly exogenous regressors. This remained the case well into the 80s; e.g., Amemiya (1986). Contributions to construct GLS estimates include Cochrane and Orcutt (1949), Prais and Winsten (1954), Durbin (1970), Amemiya (1973), among others.

The limit distributions of both the OLS and GLS estimators were well known but it was not well established how to consistently estimate the limit variance of the OLS estimate. Spurred by the development of the Generalized Method of Moments (GMM) by Hansen (1982) econometricians started to tackle this problem. Early contributions (in a more general non-linear context) include White and Domowitz (1984a), White and Domowitz (1984b), Newey and West (1987) and a comprehensive treatment was provided by Andrews (1991) who used results from the theory of spectral density estimation developed much earlier. Since then all the theoretical and empirical work has concentrated on OLS and a flood of papers have been devoted to deliver improved estimates of the limit variance of OLS so that the confidence intervals have accurate finite sample coverage rates. This continues to this day. There is barely any mention or work about GLS in the theoretical and empirical literature when dealing with the linear model with serially correlated errors, at least in econometrics. One is satisfied using OLS with a disregard for ways to improve the properties of the estimate *per se*; e.g., bias, variance and MSE (mean-squared errors). The goal is only to provide good estimates of the confidence interval of the OLS estimate.

There are generally three main reasons for adopting OLS instead of GLS. 1) There seems to be a misconception, though not shared by all, about whether OLS is valid with the regressors being exogenous or not (i.e., uncorrelated with past innovations or not), while GLS is inconsistent with non-exogenous regressors. This view is now taught early on in undergraduate textbooks; e.g., Stock and Watson (2019), ch. 16. 2) When applying GLS one needs to choose a specification to model the nature of the serial correlation in the errors. It is then argued that an incorrect specification can lead to worse results than using OLS; i.e., it is believed that while OLS is sub-optimal relative to GLS, it is more robust than

GLS, which can deliver worse outcomes (e.g., higher MSE) when not choosing a proper specification for the serial correlation in the errors; see, e.g., Engle (1974), Judge et al. (1985), p. 281, and Choudhury et al. (1999). 3) Even with a decent specification, the gains in accuracy can be minor and the inference can be misleading; e.g., bad coverage rates using standard estimates of the asymptotic variance to construct the confidence intervals. Our goal is to show that all three claims are, in general, wrong under weak conditions.

Our focus is on the linear model. Of course, our results relies on some crucial assumptions. The first is that the regressors are pre-determined, which is often viewed as less controversial for applications than the requirement of exogenous regressors. The second is that the errors are a stationary process with a Wold linear invertible representation. This is usually satisfied but may fail in some models, especially those involving rational expectations arguments.

Under the stated conditions, the first contribution is to dispel the belief that OLS is valid with non-exogenous regressors, while GLS is valid only with exogenous regressors. We show the opposite to be true, in general. The misconception likely arose from a misconceived notion of exogenous versus pre-determined regressors when the errors are correlated. We consider the linear model $y = X\beta + u$ and $u_t$ stationary so that it has a linear representation in terms of a (possibly) infinite linear model of the form $C(L) = \sum_{j=0}^{\infty} c_t e_{t-j}$ with $e_t$ being an $i.i.d.$ sequence. The usual argument for the consistency of GLS relies of whether $x_t$ is exogenous with respect to $u_t$. We argue that this leads to an incorrect result. One should analyze the issue of the consistency of GLS by assessing whether $x_t$ is exogenous with respect to the innovations $e_t$. For OLS, it does not matter since the condition remains $E(x_t u_t) = 0$. But this implies $E(x_t \sum_{j=0}^{\infty} c_t e_{t-j}) = 0$, which requires regressors exogenous with respect to $e_t$. Theoretical and simulation evidence substantiate these statements. Non-exogenous regressors can imply inconsistent OLS estimates, while the GLS estimates are consistent. Also, unlike OLS, GLS is consistent with lagged dependent variables as regressors.

The second contribution is to show that GLS is more robust that OLS, in that even a blatantly incorrect GLS correction can achieve a lower MSE than OLS when both are consistent. For illustration, we take a simple $AR(1)$ correction with parameter $\rho$ applied to a model with exogenous regressors so that both OLS and GLS are consistent. We show that, in most cases, GLS will have lower MSE than OLS for a wide range of processes and values of $\rho$, as long as $\rho$ is of the same sign as the first-order covariance of the residuals, say $cor_u(1)$. A simple procedure that pre-tests for serial correlation and applies a GLS correction with a randomly drawn value of $\rho$ with the same sign as the estimate of $cor_u(1)$ based on the estimated residuals will not do worse than OLS. This shows that GLS can be applied with a misspecified structure and still yield improvements over OLS. Also, it shows that issues of bias in the estimates of the parameters used to apply GLS will not make GLS less efficient than OLS. However, we can do better by choosing a good specification for the error process in order to achieve the lowest possible MSE and good finite-samples coverage rates for the confidence intervals. This calls for a good feasible GLS (FGLS) procedure.

The third contribution is to devise a FGLS procedure valid with pre-determined regressors whether or not they are exogenous, which achieves a MSE close to that of the infeasible GLS procedure that uses the true structure (and parameters) of the serial correlation in the errors. Care must be applied. For instance, for an $AR(1)$ process the usual procedure of Cochrane and Orcutt (1949) will not work. It is based on estimating the autocorrelation parameter using the OLS residuals. Since OLS is inconsistent when the regressors are not exogenous, this approach fails. Instead, we propose a procedure based on a generalization of the so-called Durbin (1970) regression, whose coefficients are consistent with or without exogenous regressors. Using the resulting quasi-differenced series, we apply an autoregressive approximation of order, say $k_T$, with $k_T$ chosen using the Bayesian Information Criterion (BIC); see Schwarz (1978). The simulations show that the resulting FGLS estimate performs surprisingly well in finite samples. It delivers estimates having lower MSE than OLS, often by a wide margin. The finite sample coverage rates of the confidence intervals constructed using the standard asymptotic distribution are very close to the nominal level with lengths much shorter than using OLS with heteroskedasticity and autocorrelation consistent standard errors. We provide extensive evidence for both exogenous and non-exogenous regressors. In most cases, the MSE of the FGLS is close to that of the infeasible GLS estimate.

A non-trivial exception for which OLS remains valid with serially correlated errors and non-exogenous regressors pertains to $h$ steps ahead predictive regressions as in, e.g., Hansen and Hodrick (1980). Under rational expectations, the errors are $MA(h-1)$ and the regressors are uncorrelated with the errors. Still, we show that GLS is valid and leads to much more efficient estimates, contrary to what is asserted in Hansen and Hodrick (1980), provided the MA process is invertible. In the Supplement, we also consider the case with both serial correlation and heteroskedasticity. We propose a two-step GLS procedure suggested by González-Coya and Perron (2024a) to fit the heteroskedasticity and further reduce the MSE.

The consistency of the GLS and FGLS procedure requires pre-determined regressors (uncorrelated with future innovations). This condition is certainly less contentious than the exogeneity assumption that requires the regressors to be uncorrelated with past innovations, at least in well specified models, otherwise one could forecast future innovations. Nevertheless, it is still possible to have a misspecified model or a model with some lagged endogeneity, which implies that OLS is consistent while GLS is not because the regressors are not pre-determined. However, correlation between past regressors and future innovations implies that the innovations are correlated with some variables. This is a problem of omitted variables being available or not as observations. If the omitted variable is observed (e.g., a lagged value of some covariate), then one includes the relevant lag as regressor. This purges all correlation between past regressors and current innovations so we are back with pre-determined regressors and GLS is efficient. When the omitted variable is unobserved, things are more complex. OLS can be consistent while GLS is not. However, these are knife-edge cases in the sense that minor changes in the specification renders OLS inconsistent; e.g., adding lagged regressors or having the omitted unobserved variable being serially correlated.

The rest of the paper is structured as follows. Section 2 provides the general setup and motivation. It also provides results about the conditions under which OLS and GLS are consistent. Section 3 discusses the robustness of GLS. Section 4 presents preliminary issues related to the feasible GLS estimate proposed. Section 5 presents the main Feasible GLS procedures for the general case with an invertible short-memory stationary process for the errors. Issues related to the inclusion of lagged dependent variables and the importance of the assumption of pre-determined regressors are also included. Section 6 presents extensive simulations about the finite sample properties of the OLS and FGLS estimates and how close they are to achieving the precision of the infeasible GLS estimate, for a wide variety of processes for the serial correlation in the errors. Both cases with exogenous and non-exogenous regressors are covered. Section 7 provides brief concluding remarks. A Supplement contains some technical derivations, additional material and simulation results.

## 2 General setup and motivation[1]

Consider a scalar time series of random variable $y_t$ generated by:

$$y_t = x_t'\beta + u_t, \quad t = 1, \ldots, T, \tag{1}$$

where $x_t' = (x_{1t}, \ldots, x_{kt})$ is a vector of regressors (or covariates), $\beta' = (\beta_1, \ldots, \beta_k)$ a vector of unknown coefficients, $T$ is the sample size. In matrix notation: $y = X\beta + u$, with $y = (y_1, ..., y_T)'$, $u = (u_1, ..., u_T)'$ and $X = (x_1', ..., x_T')'$. The ordinary least-squares (OLS) estimate of $\beta$ is $\hat{\beta} = (X'X)^{-1} X'y$. We assume that the error sequence $u_t$ is a stationary

[1] The material in this section was first discussed in Perron (2021). This paper now supersedes it.

process so that it admits a Wold representation of the form

$$u_t = C(L)e_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \tag{2}$$

where $c_0 = 1$. The roots of $C(L)$ are assumed to be outside the unit circle, so that $u_t$ is invertible and has an infinite autoregressive representation. Also, $\sum_{j=0}^{\infty} j|c_j| < \infty$, so that $u_t$ is a short-memory processes. For now, we assume that $e_t \sim i.i.d.\ (0, \sigma_e^2)$ (independent and identically distributed innovations). We consider heteroskedastic innovations in the Supplement. We also consider later what happens when the process is non-invertible. As a matter of terminology, we label $u_t$ as the errors and $e_t$ as the innovations.

We assume that $E[e_t x_t] = 0$, otherwise some instrumental variable procedure would be needed. We say that the regressors are "pre-determined" if:

$$E\left[x_t(e_{t+1}, ..., e_T)\right] = 0, \tag{3}$$

i.e., regressors uncorrelated with future innovations. Throughout, we shall maintain that this is the case with some comments about what happens when it does not hold in Section 5.2. We label the regressors as exogenous if

$$E\left[x_t(e_{t-1}, ..., e_1)\right] = 0, \tag{4}$$

i.e., regressors uncorrelated with past innovations. This last condition is often seen as problematic, e.g., , Stock and Watson (2019), pp. 588-597. The assumption of pre-determined regressors is usually seen as much less contentious, at least in well specified models, otherwise one could forecast future innovations. The terminology used differ in the literature. What we label as pre-determined is sometimes referred to as exogenous, and what we refer to as exogenous is labeled as strictly exogenous; e.g., Stock and Watson (2019), p. 573. We shall continue with our terminology. Also, the conditions are usually stated in terms of condition expectations, i.e., $E\left[x_t|e_{t+1}, ..., e_T\right] = 0$ or $E\left[x_t|e_{t-1}, ..., e_1\right] = 0$. Since these imply (3) and (4), respectively, and we make use of the latter only, this is without loss of generality. More importantly, we define the relation between the regressors and the innovations $e_t$, not the errors $u_t$ as is commonly done in the literature. The benefits of doing this will become clear.

### 2.1 Conditions for the Consistency of OLS

It is well known that the main condition (again apart from technical issues) for the consistency of the OLS estimate is that $E(x_t u_t) = 0$. This condition is usually seen as unproblematic apart from obvious cases of omitted variables in $u_t$ correlated with some regressor, or the presence of lagged dependent variables. The only problem is then that the limit variance is different from that obtained assuming *i.i.d.* errors and calls for the use of heteroskedasticity and autocorrelation consistent covariance matrix estimates, HAC estimates for short.

However, this condition requires, in general, exogenous regressors, since $E(x_t \sum_{j=0}^{t} c_j e_{t-j}) = 0$ is required. In general, this implies the requirement $E(x_t e_{t-j}) = 0$ or $E(e_t x_{t+j}) = 0$, which is unlikely to be satisfied when the regressors are not exogenous. We state that this is the case "in general" since there are many ways in which the regressors could be non-exogenous and $E(x_t u_t) = 0$. We view these as knife-edge cases. For example, $x_t$ is correlated with $e_{t-2}$ but $u_t = e_t + c_1 e_{t-1} + c_3 e_{t-3}$. Also, the correlation between $x_t$ and various lags of $e_t$ is such that the stated condition holds. For instance, suppose that $u_t$ is an $MA(2)$. Then, if $c_1 E(xe_{t-1}) = -c_2 E(x_t e_{t-2})$ and $E(x_t e_t) = 0$, we have $E(x_t u_t) = 0$. Such cases are, however, unlikely to hold in practice. See also Section 2.1.1.

Another way of assessing this result is to argue that a regression with serially correlated errors is dynamically misspecified. Consider an $AR(1)$ model of the form $u_t = \rho u_{t-1} + e_t$. Then, $E(u_t x_t) = 0$ implies that $x_t$ is exogenous with respect to $e_t$ since $E(u_t x_t) = \rho E(u_{t-1} x_t) + E(e_t x_t) = 0$ if $E(u_{t-1} x_t) = 0$ or equivalently $E(e_{t-j} x_t) = 0$, in general. In other words, $E(y_t | x_t) = x_t' \beta$ if $x_t$ is exogenous, except for some knife-edge cases.

**Example 1.** *Consider a simple regression of consumption on income, say $c_t = \mu + \beta inc_t + u_t$. Suppose you have a unexpected event that affect your potential consumption beyond your regular income, which is reflected here in $e_t$. Then most likely one would consume part of it this period, and smaller parts in future periods. This would lead, as a rough approximation, to an error process $u_t$ of an $AR(1)$ type given by $u_t = \rho u_{t-1} + e_t$. It is plausible to argue it is uncorrelated with past income (since it was unforecastable) and also with current income (because of time to adjust). Then, income is pre-determined with respect to the shocks $e_t$. In future periods, agents will most likely adjust to re-establish the desired balance. For instance, future income is to some extent forecastable. If you are young you may expect it to be higher. Then, you are likely to consume almost all your windfall (maybe more via borrowing). This affects the error in the regression. Hence, income is not exogenous with respect to past innovations. If $\rho \neq 0$, the shocks will persist for some time. Hence, current income will be correlated with past shocks. Therefore, current income will also be correlated with the current value of $u_t$ causing inconsistency of the OLS estimate. Note that the example given pertains to past shocks affecting future levels of consumption, not the past levels of income, which may affects current consumption; e.g., because of habit formation. This suggests including lagged income as a regressors, which restores regressors to be pre-determined, see Section 5.2.*

### 2.1.1 The Rational Expectations (RE) case

There is one non-trivial and empirically relevant exception for which OLS remains valid when the errors are serially correlated and the regressors are not exogenous. This pertains to multi-steps ahead predictive regressions as examined, for instance, in the influential work of Hansen and Hodrick (1980). In their framework, it is supposed that $E(y_{t+h}|\Phi_t) = x_t'\beta$, where $\Phi_t$ is the information set available at time $t$. Then,

$$y_{t+h} = x_t'\beta + u_{t+h}, \tag{5}$$

with $u_{t+h} = y_{t+h} - E(y_{t+h}|\Phi_t)$ so that the errors terms are forecast errors from using the best predictor based on $x_t$. It can be shown that $u_{t+h}$ is an $MA(h-1)$ process. Since $x_t \subset \Phi_t$, $E(x_t u_{t+h}) = 0$ and OLS is consistent. Following our notation, we can write (5) as $y_t = x_{t-h}'\beta + u_t$, where $u_t = \sum_{j=0}^{h-1} c_j e_{t-j}$. OLS is then consistent only requiring pre-determined regressors so that $E[x_{t-h}\sum_{j=0}^{h-1} c_j e_{t-j}] = 0$. Hence, such cases involve no issue related to exogenous regressors and the fact that the regressors are pre-determined is an implication of the rational expectations hypothesis. Still, as discussed in Remark 4 below, GLS remains consistent with non-exogenous regressors.

### 2.1.2 Summary

Our purpose is to clarify the conditions under which OLS is consistent. Nothing new is offered. The main condition still remains $E(x_t u_t) = 0$. One often reads that GLS should not be applied because it requires exogenous regressors. Since OLS is routinely applied, some researchers may think that issues of exogeneity are irrelevant for the consistency of OLS and only argue that it is enough to ensure that the regressors and the innovations (the $e_t$) are contemporaneously uncorrelated. Stating the condition as $E(x_t \sum_{j=0}^{t} c_j e_{t-j}) = 0$ (for the linear processes considered) makes it clear that exogeneity of the regressors with respect to all past innovations is needed except for the "RE case" and some knife-edge occurrences. Of course, this requires working with the Wold representation for $u_t$. It may well be the case that one has some structural model not in this form and is able to deduce that $E(x_t u_t) = 0$ directly. Then issues of exogeneity with respect to $u_t$ (or $e_t$) become irrelevant.

## 2.2 Conditions for the Consistency of GLS

Since $u_t$ is assumed stationary, let $V(u) = \sigma_e^2\Omega$, a symmetric, non-singular, and positive definite matrix. Then, there exists a non-singular matrix $D$ such that $D'D = \Omega^{-1}$. The GLS estimate is given by $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$ and, using (1),

$$\hat{\beta}_{GLS} - \beta = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}u = (X'D'DX)^{-1} X'D'Du.$$

The main condition for consistency is that

$$p\lim_{T\to\infty} T^{-1}X'\Omega^{-1}u = p\lim_{T\to\infty} T^{-1}X'D'Du = 0. \tag{6}$$

In other words, $DX$ and $Du$ must be uncorrelated, at least in large samples. Consistency can be achieved as follows. Note first that we can choose $D$ to be lower triangular. For instance, the Cholesky decomposition gives $\Omega = LL'$ with $L$ lower triangular. We can set $D = L^{-1}$, which will be lower triangular. The elements of $DX$ are of the form $\sum_{j=1}^{t} d_{tj}x'_j$, which for row $t$ involves only current and past $x$'s. The next condition is to ensure that $Du$ recovers the vector of innovations $(e_1, e_2, ..., e_t, ...)$ at least in large samples. This is where the assumption of the invertibility of the MA representation is important, i.e., that the roots of $C(L)$ be all outside the unit circle. Then, $u_t$ has an autoregressive representation of the form $A(L)u_t = e_t$. A common practice is to approximate this possibly infinite AR process by a finite order one, with the order increasing with $T$, i.e., use the process

$$u_t = \sum_{j=1}^{k_T} \rho_j u_{t-j} + e_{t,k_T},$$

with $k_T$ increasing at some appropriate rate as $T$ increases. This is a standard approach in the time series literature with a long history of useful applications. Note that as $T$ increases, $e_{t,k_T}$ approaches $e_t$. The details for the implementation are in Section 5. Then, we have,

$$\lim_{T\to\infty} E[X'D'Du] = E[\textstyle\sum_{t=1}^{\infty}(\sum_{j=1}^{t} d_{tj}x'_j)'e_t] = 0, \tag{7}$$

requiring only pre-determined regressors. Therefore, GLS is consistent without the need for exogenous regressors.

**Remark 1.** *Since $D'D = \Omega^{-1}$, GLS is invariant to the choice of $D$. Hence, only pre-determined regressors are needed whatever the choice of $D$, provided the invertibility condition holds. Consider the $AR(1)$ model with a forward filter, i.e., $D$ chosen to be upper triangular, call it $F$. Ignoring the first and last observations $F = D'$, the condition for consistency is*

$$E[(x_t - \rho x_{t+1})(u_t - \rho u_{t+1})] = E[(x_t - \rho x_{t+1})((1-\rho^2)u_t - \rho e_{t+1})] = 0,$$

*which requires a) $E[x_{t+1}e_{t+1}] = 0$, holding by assumption; b) $E[x_te_{t+1}] = 0$, satisfied with predetermined regressors; and c) $E[(x_t - \rho x_{t+1})u_t] = 0$, also holding with non-exogenous* *regressors since*

$$\begin{aligned} E[(x_t - \rho x_{t+1})u_t] &= E[(x_t - \rho x_{t+1}) \textstyle\sum_{j=0}^{t} \rho^j e_{t-j}] = E[x_t \rho^j e_{t-j}] - \rho E[x_{t+1} \rho^{j-1} e_{t-j+1}] \\ &= \rho^j E[x_t e_{t-j}] - \rho^j E[x_{t+1} e_{t-j+1}] = 0, \end{aligned}$$

*since the last two terms are equivalent. What is needed is solely that there exist one decomposition of* $\Omega^{-1}$ *with* $D$ *lower triangular and* $Du = e$*, at least in large samples.*

Consider $AR(1)$ errors, $u_t = \rho u_{t-1} + e_t$. Ignoring the first observation for simplicity,

$$D = \begin{bmatrix} 1 & 0 & & 0 \\ -\rho & 1 & & \\ & & \ddots & \\ 0 & & -\rho & 1 \end{bmatrix} \tag{8}$$

and

$$p\lim_{T\to\infty} T^{-1} X'D'Du = p\lim_{T\to\infty} T^{-1} \textstyle\sum_{t=2}^{T} (x_t - \rho x_{t-1})(u_t - \rho u_{t-1}).$$

For this quantity to converge to zero, the conditions often advanced for (6) to hold are $E(x_t u_t) = E(x_t u_{t-1}) = E(x_{t-1} u_t) = 0$. It is often argued that the condition $E(x_t u_{t-1}) = 0$ is problematic following (4); see Stock and Watson (2019), pp. 584-585, who use this reasoning to argue that GLS requires exogenous regressors and, hence, have limited appeal in practice. But this overlooks the fact that $u_t$ is a composite of the fundamental sources of variations, namely $e_t$, and ignores the structure of the model. Also, assessing exogeneity conditions based on the relation between $x_t$ and $u_t$ is not appropriate. Since the GLS regression is OLS applied to the regression $y^* = X^*\beta + e$, where $y^* = Dy$ and $X^* = DX$, issues related to the exogeneity of the regressors need to be analyzed via the relation of $X^*$ to $e$ and not of $X$ to $u$. There are no more $u$'s in the model. Indeed, we can write (6) as

$$T^{-1} (DX)' (Du) = T^{-1} \textstyle\sum_{t=2}^{T} (x_t - \rho x_{t-1}) e_t. \tag{9}$$

Thus, for consistency, we need $E(x_t - \rho x_{t-1}) e_t = 0$, or $E(x_t e_t) = E(x_{t-1} e_t) = 0$, for all $t$, which is satisfied as long as the regressors are predetermined. There is no need to assume exogenous regressors. Then, assuming $\rho$ known, one can consistently estimate $\beta$ using the quasi-difference regression

$$(y_t - \rho y_{t-1}) = (x_t - \rho x_{t-1})'\beta + e_t, \ (t = 2, ..., T). \tag{10}$$

When $u_t$ is general linear process, GLS simply amounts to OLS applied to the regression

$$(y_t - \sum_{j=1}^{k_T} \rho_j y_{t-j}) = (x_t - \sum_{j=1}^{k_T} \rho_j x_{t-j})'\beta + e_{kt},\ (t = k_T + 1, ..., T).$$

**Remark 2.** *It is useful to expand on the condition (9). Suppose we apply GLS with some arbitrary value $|\rho^*| < 1$. Then, with $D^*$ as defined by (8) with $\rho^*$ instead of $\rho$,*

$$\begin{aligned} T^{-1}(D^*X)'(D^*u) &= T^{-1}\sum_{t=2}^{T}(x_t - \rho^* x_{t-1})(u_t - \rho^* u_{t-1}) \\ &= T^{-1}\sum_{t=2}^{T}(x_t - \rho^* x_{t-1})(e_t - (\rho - \rho^*)u_{t-1}) \\ &= T^{-1}\sum_{t=2}^{T}(x_t - \rho^* x_{t-1})(e_t - (\rho - \rho^*)(e_{t-1} + \rho u_{t-2})). \end{aligned}$$

*Therefore, assuming pre-determined regressors, i.e., $E(x_t e_t) = E(x_{t-1} e_t) = 0$, for all $t$, what is needed for consistency is either a) exogenous regressors so that $E(x_t e_{t-1}) = E(x_t e_{t-2}) = E(x_{t-1} e_{t-2}) = 0$, irrespective of the value of $\rho$ and $\rho^*$; or b) non-exogenous regressors and $\rho = \rho^*$. Accordingly, if the regressors are exogenous, GLS is consistent using any value of $\rho^*$, including 0, so that OLS is consistent, a well-known result, see above. On the other hand, with non-exogenous regressors, we need $\rho = \rho^*$ for consistency, i.e., the correct value of the parameter of the serial correlation in $u_t$. Of importance is the fact that when $\rho \neq 0$, the value $\rho^* = 0$ is not permitted, showing that OLS is indeed inconsistent as claimed above using other arguments. This result can be extended to more general cases. The fact that the correct GLS transformation is needed is exemplified by the arguments advanced by Flood and Garber (1980) who argued, correctly, that applying an $AR(1)$ correction to a model with $ARMA(1,1)$ errors leads to GLS being inconsistent. However, unlike what they stated, this does not mean that GLS is not applicable. It simply needs to be applied correctly.*

**Remark 3.** *An important implication of our result is the fact that unlike OLS, GLS is consistent with lagged dependent variables as regressors. This follows given that (7) remains 0 when $x_t$ includes lagged dependent variables given $E(y_{t-j} e_t) = 0$ $(j \geq 1)$. Since in the original model estimated by OLS, a lagged dependent variable implies $E(x_t u_t) \neq 0$, OLS is inconsistent. The GLS transformation can be viewed as a way to obtain a regression with pre-determined regressors with respect to the relevant innovations $e_t$.*

**Remark 4.** *Contrary to the claim made by Hansen and Hodrick (1980), and re-iterated in Hansen and West (2002), GLS is, in general, consistent with predictive regressions of the type discussed in Section 2.1.1, provided the MA process is invertible. This follows trivially since (7) is satisfied if the regressors only include lagged values at delay $h$, i.e., the GLS regression still only involves predetermined regressors with respect to the innovations $e_t$. We show in the Supplement, Section S.4, that even for this case GLS performs much better.*

## 3 The Robustness of GLS

It is often argued that GLS may be less robust than OLS because a wrong specification of the process for $u_t$ may lead GLS to have higher MSE than OLS. We show that this is incorrect. To have meaningful comparisons, we assume exogenous regressors so that both OLS and GLS are consistent. Note first that GLS is consistent even when using a misspecified model when the regressors are exogenous and pre-determined. Suppose you assume that $V(u) = \sigma_e^2 \Omega_*$ while the correct specification is $V(u) = \sigma_e^2 \Omega$. Let $\Omega_*^{-1} = D_*' D_*$ and $\Omega^{-1} = D'D$. Then,

$$T^{-1} X' \Omega_*^{-1} u = T^{-1} X' \Omega_*^{-1} D^{-1} e = T^{-1} (HX)' e \xrightarrow{p} 0,$$

since $HX$ with $H = X' \Omega_*^{-1} D^{-1}$ is simply a linear combination of all the regressors, which are uncorrelated with the innovations at all leads and lags (and current value). We shall show that when adopting a simple $AR(1)$ specification, it is possible to obtain GLS estimates that performs no worse than OLS, and most often much better, irrespective of the true data-generating process for the errors, as long as it is stationary. For reasons that will become clear, we apply an $AR(1)$ GLS with some known value $\rho$, i.e., OLS applied to the regression (10). We ignore the initial condition for simplicity. We have the following results about the relative MSE of OLS and GLS.

**Theorem 1.** *Let $u_t$ be a zero mean stationary process and $\hat{\beta}_{GLS}$ the estimate applying OLS to (10) for a given value $\rho$. The scalar exogenous variable $x_t$ satisfies $p\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T-j} x_t x_{t+j} = R_x(j)$, $cor_x(j) = R_x(j)/R_x(1)$, with similar definitions for $cor_u(j)$. Also, $h_{xu}(0)$ is the spectral density function at frequency zero of $x_t u_t$, $\widetilde{R}_{xu}(1) = \int_{-\pi}^{\pi} \cos(\lambda) h_x(\lambda) h_u(\lambda) d\lambda$, and $\widetilde{R}_{xu}(2) = \int_{-\pi}^{\pi} \cos(2\lambda) h_x(\lambda) h_u(\lambda) d\lambda$ with $h_x(\lambda)$ and $h_u(\lambda)$, the spectral density function of $x_t$ and $u_t$, respectively. Then,* $\lim_{T\to\infty} (\mathrm{MSE}(\hat{\beta}_{GLS}) / \mathrm{MSE}(\hat{\beta}_{OLS})) < 1$*if*

$$\rho^2 - 2\rho(1+\rho^2)\widetilde{R}_{xu}(1)/h_{xu}(0) + \rho^2 \widetilde{R}_{xu}(2)/h_{xu}(0) < 2\rho^2 \operatorname{cor}_x(1)^2 - 2\rho(1+\rho^2)\operatorname{cor}_x(1).$$

The result in the previous Theorem, proved in the Supplement, is useful but opaque as far as obtaining useful insights given the level of generality. The following corollary considers the case with *i.i.d.* regressors. While restrictive, the results allow important insights that still apply with a serially correlated regressor; see Section S.2 of the Supplement.

**Corollary 1.** *Under the same conditions as in Theorem 1, except that $x_t \sim i.i.d.(0, \sigma_x^2)$,* $\lim_{T\to\infty} (\mathrm{MSE}(\hat{\beta}_{GLS}) / \mathrm{MSE}(\hat{\beta}_{OLS})) < 1$ *if*

$$\begin{aligned} \rho/(2(1+\rho^2))(1+\operatorname{cor}_u(2)) &< \operatorname{cor}_u(1) \quad \textit{when } \rho > 0, \\ \rho/(2(1+\rho^2))(1+\operatorname{cor}_u(2)) &> \operatorname{cor}_u(1) \quad \textit{when } \rho < 0. \end{aligned}$$

A necessary condition for such inequalities to hold is that $\rho\, \mathrm{cor}_u(1) > 0$. To explore the intuitive content, suppose that $u_t$ is an $AR(1)$ process with parameter $\rho_u$ and $\rho > 0$. Then,

$$\lim_{T\to\infty}(\mathrm{MSE}(\hat{\beta}_{\mathrm{GLS}})/\,\mathrm{MSE}(\hat{\beta}_{\mathrm{OLS}})) < 1 \iff \rho(1+\rho_u^2) - 2\rho_u(1+\rho^2) < 0.$$

If $\rho = \rho_u$, the condition is trivially satisfied, as expected. Moreover, it is satisfied unless $\rho_u < 0.27$, in which case we need $0 < \rho < 2\rho_u$. As will transpire from the simulations results, $\rho\, \mathrm{cor}_u(1) > 0$ is nearly also a sufficient condition unless $\mathrm{cor}_u(1)$ is small. This is quite a strong result. It says that applying GLS with an $AR(1)$ specification will lead to an estimate with lower MSE than OLS for a wide range of data-generating processes for $u_t$ by simply quasi-differencing the data with a parameter $\rho$ that has the same sign as $cor_u(1)$, the first-order correlation coefficient of $u_t$. If $cor_u(1) = 0$, OLS performs better. This can occur with serial correlation implying $cor_u(1) = 0$ and $cor_u(j) \neq 0$ for some $j > 1$. An example is an $MA(2)$ process of the form $u_t = e_t + \theta_2 e_{t-2}$. We view such cases as knife-edge ones. When $\mathrm{cor}_u(1)$ is small, the same results hold for a range given by $0 < \rho < 2\rho_u$.

GLS with a simple $AR(1)$ specification will beat OLS for a wide range of quasi-difference parameters whatever the true DGP for $u_t$. So not only can we misspecify the nature of the serial correlation but also allow a wide range of values for the quasi-difference parameter, and still have GLS perform better than OLS. Of course, we are not saying that adopting a simple $AR(1)$ with a value of $\rho$ having the same sign as $cor_u(1)$ is the best. For that, we need a FGLS procedure that yields an estimate asymptotically equivalent to GLS with the correct specification for $u_t$. We will cover in Sections 4-5, a method to achieve this goal. We could extend the results to have alternative GLS procedures, e.g., some $AR(k)$. The results would be much more complex, though qualitatively similar. Hence, such extensions would add little to the main message, namely the robustness of GLS.

We illustrate these issues using simulations in the Supplement. The results are in accordance with the theory. When $cor_u(1)$ is "large", GLS has smaller MSE than OLS when the sign of the quasi-difference parameter is the same as the sign of $cor_u(1)$. If $cor_u(1)$ is "small" GLS is better when $\rho$ is in the vicinity of $cor_u(1)$. We also consider a very simple procedure to obtain a GLS estimate that is (almost) never worse than OLS, subject to very minor random deviations. First use a test for serial correlation at delay one; we use the LM test of Godfrey (1978). If the test does not reject the null hypothesis of no serial correlation, then use OLS. This will occur when $cor_u(1)$ is "small". If the test rejects, estimate $cor_u(1)$ via the sample first-order serial correlation of the OLS residuals. If it is positive (negative), use any positive (negative) value of the quasi-differencing parameter $\rho$. To make clear that any value of $\rho$ will do, in the simulations we simply draw $\rho$ from a Uniform distribution with support $(0.1, 0.9)$ when positive value are required and with support $(-0.9, -0.1)$ when negative values are in order. The hybrid-GLS procedure yields more precise estimates for almost all cases, with minor exceptions when $cor_u(1)$ is "small".

Of course, using the incorrect quasi-differences does not lead to the best outcome as GLS is optimal only when the correct specification is used. Hence, in order to have estimates as good as possible (lowest MSE), we need to obtain a parameterization of the DGP for the errors that is a good approximation to the true one without any prior knowledge about the true structure. This leads to consider Feasible GLS (FGLS), which we tackle in the next section. Still, the results of this section are important in that they suggest that some departures from the true specification due to misspecification or biased parameter estimates will not make FGLS being less precise than OLS.

## 4 Issues Related to Constructing a Feasible GLS Estimate

We consider first the case with $AR(1)$ residuals to present the main issues of interest. The model with non-exogenous regressors is

$$y_t = \beta x_t + u_t, \quad u_t = \rho u_{t-1} + e_t, \tag{11}$$

with $x_t = (1, w_t)'$, $w_t = v_t + e_{t-1}$, $v_t, e_t \sim i.i.d.N(0,1)$ independent of each other. In practice, one needs a feasible version of the GLS estimate. Here, the Cochrane and Orcutt (1949) procedure will not work since it estimates $\rho$ using the OLS residuals, i.e., $\hat{\rho}^{CO} = \sum_{t=2}^{T} \hat{u}_{t-1}\hat{u}_t / \sum_{t=2}^{T} \hat{u}_{t-1}^2$, where $\hat{u}_t = y_t - x_t'\hat{\beta}_{OLS}$. Without exogenous regressors, $\hat{\beta}_{OLS}$ is inconsistent and so will $\hat{\rho}^{CO}$. A method valid without exogenous regressors is to first estimate $\rho$ using Durbin's regression (Durbin (1970)), which simply re-writes (10) as

$$y_t = \rho y_{t-1} + x_t'\beta - \rho x_{t-1}'\beta + e_t. \tag{12}$$

Then, a consistent estimate of $\rho$, say $\hat{\rho}^D$, can be obtained estimating (12) by OLS. Using the estimate on the lagged dependent variable, one can then construct a feasible version of the quasi-difference regression (10) using

$$(y_t - \hat{\rho}^D y_{t-1}) = (x_t - \hat{\rho}^D x_{t-1})'\beta + e_t, \ (t = 2, ..., T), \tag{13}$$

to estimate $\beta$. The estimates of $\beta$ and $\rho$ will be consistent with regressors exogenous or not as long as they are pre-determined. Alternatively, one can simply estimate $\beta$ using OLS applied directly to the Durbin regression (12), though this is less efficient since it does not amount to a GLS procedure. Of course, one can iterate though we do not pursue this avenue.

It is useful to illustrate the issues via simple simulation experiments. The specifications are the same as (11) for the $AR(1)$ case and is $y_t = x_t'\beta + u_t$, where $x_t = (1, w_t)'$ with $w_t = v_t + e_{t-1}$, and $u_t = \rho u_{t-1} + e_t$ is an $AR(1)$ process; $v_t, e_t \sim i.i.d.N(0,1)$ independent of each other. We set $u_0 = 0$, without loss of generality, $\beta = (1,1)'$, $\rho = 0.8$ and $T = 500$. The simulations are based on 10,000 replications. Note that $E\left(e_t x_{t+1}\right) \neq 0$, so that the regressors are not exogenous. Accordingly, $E(x_t u_t) \neq 0$ and OLS is inconsistent. Note that $E\left(e_t x_t\right) = 0$ so that no "classical" endogeneity problem is present. Also $E(x_t e_{t-j}) = 0$ $(j > 0)$ so that GLS is consistent. We consider the following regressions, where $\delta = \rho\beta$:

a) $y_t = x_t'\beta + u_t \quad (OLS)$; b) $y_t = x_t'\beta + \rho y_{t-1} + x_{t-1}'\delta + \widetilde{u}_t \quad$ (Durbin)

c) $y_t - \rho y_{t-1} = \left(x_t - \rho x_{t-1}\right)'\beta + e_t \quad (GLS)$; d) $y_t - \hat{\rho} y_{t-1} = \left(x_t - \hat{\rho} x_{t-1}\right)'\beta + e_t \quad (FGLS)$

The first is simply OLS; the second is the Durbin regression from which consistent estimates of $\rho$ and $\beta$ can be obtained. The third is the infeasible GLS based on the known value of $\rho$ (to be used as a benchmark). The fourth is a feasible GLS regression for which we shall use two estimates of $\rho$: a) that used in the Cochrane and Orcutt procedure based on

$$\hat{\rho} = \sum_{t=2}^{T} \hat{u}_{t-1}\hat{u}_t / \sum_{t=2}^{T} \hat{u}_t^2, \tag{14}$$

where $\hat{u}_t = y_t - x_t'\hat{\beta}_{OLS}$. As argued above, this should lead to an inconsistent estimate of $\beta$. This method is labelled CO-FGLS. b) The estimate of $\rho$ obtained from the Durbin regression, with the method labelled as FGLS. The results are presented in Table 1.

Obviously, the bias and MSE of OLS is very large, in accordance with the fact that it is inconsistent. The Durbin and FGLS methods lead to very small biases, in accordance with the fact that they yield consistent estimates. The FGLS has better finite sample properties and performs nearly as well as the infeasible GLS method. The CO-FGLS method has surprisingly small bias (and MSE) despite being inconsistent. This can be explained as follows. The estimate of $\rho$ given by (14) has a substantial bias so that the mean of the estimate of $\rho$ is 0.63 instead of 0.8. As argued in Section 3, it is better to do any kind of GLS method instead of OLS. Here, the quasi-differencing operation is biased but still effective in substantially reducing the bias in the estimate of $\beta$, though not as well as when using a less biased and consistent estimate as provided by the Durbin regression and used in the FGLS method. Using simulations with $T = 10,000$, we verified that the bias and MSE of OLS and CO-FGLS remains the same, while those for Durbin, GLS and FGLS are nearly zero.

The FGLS estimate of $\beta$ is, however, more efficient than that obtained from the Durbin regression with a MSE 31% smaller in the simulations. FGLS also remains more efficient in large samples since the Durbin regression does not impose relevant restrictions; see Remark 6 for more details. Hence, we shall only consider the FGLS method.

## 5 FGLS for the general case

We now present the recommended feasible method, applicable for all cases except with lagged dependent variables as regressors, discussed later. Assuming invertibility, we can approximate the linear processes (2) by some autoregression whose order increases with $T$, i.e., use $u_t = \sum_{j=1}^{k_T} \rho_j u_{t-j} + e_{kt}$, with $k_T \to \infty$ at some appropriate rate so that $e_{kt}$ is nearly $i.i.d.$. Then (12) and (13) are replaced by

$$y_t = \sum_{j=1}^{k_T} \rho_j y_{t-j} + x_t'\beta - \sum_{j=1}^{k_T} x_{t-j}'\delta_j + e_{kt}, \tag{15}$$

$$(y_t - \sum_{j=1}^{k_T} \hat{\rho}_j^D y_{t-j}) = (x_t - \sum_{j=1}^{k_T} \hat{\rho}_j^D x_{t-j})'\beta + e_{kt},\ (t = k_T + 1, ..., T), \tag{16}$$

where $\hat{\rho}_j^D$ $(j = 1, ..., k_T)$ are the OLS estimates of the coefficients associated with the lagged dependent variables from regression (15). Of course, one can iterate starting with any consistent estimate. However, as our simulations will show the estimates have very good properties so that iterations are not warranted. The FGLS estimate can then be computed in two steps: 1) For any given $k_T$, estimate (15) by OLS and use BIC to select the lag length $k_T^*$. The search is made for $k_T \in [0, k_T^{\max}]$ and the method suggested by Ng and Perron (2005) is used to ensure a proper comparison across models with different values of $k_T$, i.e., using the same effective number of observations, namely $T - k_T^{\max}$. The maximal order $k_T^{\max}$ increases with $T$, but in practice the method is robust to reasonable values. We use $k_T^{\max} = 12$ when $T = 200, 500$. Hence, BIC selects $k_T^* = \arg\min_{k_T}[\ln(\hat{\sigma}_{ek^*}^2) + (\ln(T - k_T^{\max})/(T - k_T^{\max}))k_T]$, where $\hat{\sigma}_{ek^*}^2 = (T - k_T^{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \hat{e}_{kt}^2$ and $\hat{e}_{kt}$ are the residuals from applying OLS to (15) using observations $t = k_T^{\max} + 1, ..., T$ for each value of $k_T$. 2) From step 1, use the estimates $\hat{\rho}_j^D$ $(j = 1, ..., k_T^*)$ to construct the quasi-differenced variables $(y_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D y_{t-j})$ and $(x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})$. The FGLS estimate of $\beta$ is then obtained applying OLS to the regression (16) with $k_T = k_T^*$ using the observations $t = k_T^* + 1, ..., T$.

The FGLS estimate will have the same asymptotic properties as the infeasible GLS estimate. The arguments are as follows. If the process is an $AR(p)$, BIC will select a value $k_T^*$ that converges in probability to $p$. The estimates $\hat{\rho}_j^D$ are consistent for $\rho_j$ $(j = 1, ..., k_T^*)$. For general linear short-memory processes $k_T^* = O_p(\ln(T))$, which increases to infinity. Hence,

$||\hat{\rho}_j^D - \rho_j|| = O_p(T^{-1/2})$, where $||\cdot||$ is the Euclidean norm of the vector. This holds following Berk (1974) under the same conditions, basically that $k_T \to \infty$ and $k_T^3/T \to 0$. Since these rates allow a log rate of increase for $k_T$, the same result holds when selecting $k_T$ using BIC, which implies a log rate of increase as shown in Hannan and Deistler (2012). Given the consistency and rate of convergence of $\hat{\rho}_j^D$, it is then easy to show the equivalence between FGLS and the infeasible GLS. The estimation of the parameters $\hat{\rho}_j^D$ has no first-order effect. Since the technical arguments involve only modifications of already established results, we omit the details. Hence, the asymptotic distribution is given by

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, p\lim_{T\to\infty} \sigma_e^2 (X'\Omega^{-1}X)^{-1}),$$

and the limit variance is consistently estimated by $\hat{\sigma}_{ek}^2[(x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})'(x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})]^{-1}$, where $\hat{\sigma}_{ek}^2 = (T - k_T^*)^{-1} \sum_{t=k_T+1}^{T} \hat{e}_{kt}^2$, with $\hat{e}_{kt}$ the estimated residuals from applying OLS to the FGLS regression (16) with $k_T = k_T^*$. The main idea is to have some transformations to make the regression residuals as close as possible to the contemporaneous true errors and then have this regression involve only past regressors so that only pre-determined regressors are required. Asymptotically, it works. It is a standard approach in the time series literature. Of course, in finite samples, some leftover correlation might be present. Then, it is an issue about whether the asymptotic approximation and the choice of the tuning parameters $k_T^*$ provide good approximations in finite sample. In Section 6, we provide extensive simulations to show that a) the mean, variance and MSE are close to that which could be obtained using the infeasible GLS procedure; b) the coverage rates of the confidence intervals are near the nominal level, i.e., the asymptotic distribution is a good approximation; c) the length of the confidence intervals are shorter (higher precision) compared to other methods.

**Remark 5.** *Amemiya (1973) analyzed feasible GLS when the errors $u_t$ are an $ARMA(p,q)$ process approximated by an $AR(k_T)$ with $k_T$ increasing with $T$. He uses the OLS residuals and assumes "non-stochastic" regressors. Our results show that his proposed method is valid only under the assumption of exogenous regressors. Still, our approach is closely related. For a similar more recent treatment, see Fang et al. (2023). For more advanced treatments, see Hannan and Kavalieris (1986) and Hannan and Deistler (2012), among many others.*

**Remark 6.** *In order to improve upon OLS, Baillie et al. (2024) proposed using the Durbin regression (15). They claim correctly that the estimate of $\beta$ is consistent whether the regressors are exogenous or not. However, this leads to a less efficient estimates compared to FGLS, which can be substantial even though it remains more efficient than OLS. Additional* *simulation experiments showed our FGLS procedure to be more efficient mostly due to the fact that with serially correlated regressors issues of multicolinearity reduces efficiency; see also González-Coya and Perron (2024b) who present evidence of very poor power of tests when using the Durbin regression for cases calibrated to real data. Hence, we shall not further consider methods to estimate $\beta$ based on (15). As discussed below, it offers no additional advantage in extended contexts such as regressors with lagged dependent variables and non-predetermined regressors. Their method need not impose the so-called common factor restriction. But our goal is to establish the most efficient method for the linear model adopted, which is GLS. If one views the common factor restriction as unreasonable, the problem becomes different and we have nothing to say about this* [2]*.*

**Example 2.** *González-Coya and Perron (2024b) consider issues related to testing the Uncovered Interest Parity (UIP) condition. Let $s_t = \ln(S_t)$ and $f_{t,h} = \ln(F_{t,h})$, where $S_t$ and $F_{t,h}$ are the levels of the spot exchange rate and the h-period forward exchange rate at time t. The UIP condition is also frequently expressed as $E(s_{t+h} - s_t|\Phi_t) = (f_{t,h} - s_t)$. Since $s_{t+h} - f_{t,h}$ is an approximate measure of the rate of return to speculation, we can express the efficient-markets hypothesis as $f_{t,h} = E(s_{t+h}|\Phi_t)$. This implies forecast errors $s_{t+h} - f_{t,h}$ uncorrelated with information available at time t, $\Phi_t$. The Fama (1984) regression is*

$$s_{t+h} - s_t = \alpha + \beta(f_{t,h} - s_t) + u_{t+h}, \tag{17}$$

*with $h = 4$ for 1-month forward rates and $h = 12$ for 3-month forward rates, when using weekly data. Under the efficient-market hypothesis, $H_0 : \alpha = 0, \beta = 1$ so that the log of the forward rate provides an unbiased forecast of the log of the future spot exchange rate. Derivations from $\beta = 1$ are sometimes interpreted as a measure of the variation of the premium in the forward rate. The component $u_{t+h}$ is a forecast error having an $MA(h-1)$ structure. It is derived from an assumption of rational expectations with no frictions. It seems reasonable to argue that frictions are present such that $E(s_{t+h} - s_t|\Phi_t) = (f_{t,h} - s_t)$ holds on average and that temporary serially correlated deviations could be present. Then the model is still (17) but with $v_{t+h}$ instead of $u_{t+h}$, where $v_{t+h}$ is a more general linear process that depends on past innovations; e.g. an $ARMA(1,h)$. Then, OLS is consistent,*

[2] For the record, which is incorrectly stated in their paper, their prior versions (e.g., arXiv:2203.04080v1) before presenting our work at the NBER-NSF time series conference in September 2022 labelled the method as DynReg and argued that it was a device to improve the finite sample coverage rate over OLS+HAC. It continued to claim that OLS was consistent and GLS not when the errors are serially correlated. Their newer versions changed the label of the method as the Durbin regression and they now claim that OLS is inconsistent while GLS is. These changes were in no doubt fostered by our work, but improperly acknowledged.

*though not efficient under (17) but inconsistent with $v_{t+h}$ instead of $u_{t+h}$ if $(f_{t,h} - s_t)$ is not exogenous with respect to past innovations. On the other hand, FGLS is consistent and efficient under both cases. A negative estimate of $\beta$ in regression (17) is a robust finding in the literature; see Engel (1996). This is known as the "forward discount anomaly"; it is a widespread empirical finding that the returns on nominal exchange rates appear to be negatively correlated with the lagged forward premium. Using data from a variety of country pairs, González-Coya and Perron (2024b) replicate this finding using the OLS estimates. However, the FGLS estimates are drastically different. They are mostly above 0. They interpret the large differences between the OLS and the FGLS estimates as evidence of OLS being inconsistent, which could account for the forward discount anomaly. To substantiate this claim, they present evidence of serial correlation in the errors at lags greater than $h-1$.*

### 5.1 The case with lagged dependent variables as regressors

As stated in Remark 3, GLS is consistent with lagged dependent variables as regressors. However, alternative methods to get consistent estimate of the parameters $\rho_j$ $(j = 1, ..., k_T^*)$ are needed to construct the FGLS estimate. Consider the model

$$y_t = \sum_{j=1}^{p_y} \alpha_j y_{t-j} + x_t'\beta + u_t,$$

where $u_t = C(L)e_t$ is again a linear invertible stationary short-memory process and $x_{jt}$ $(j = 1, ..., k)$ are pre-determined regressors. When constructing the Durbin regression, one pre-multiply both sides by $(1 - \sum_{i=1}^{k_T^*} \rho_j L^i)$ for some $k_T^*$ selected via the BIC information criterion. Assuming $k_T^* = p_y$ for simplicity, this leads to the regression

$$y_t = \sum_{j=1}^{k_T^*} \alpha_j^* y_{t-j} + (x_t'\beta_j - \sum_{j=1}^{k_T^*} x_{t-j}'\delta_j) + e_{kt}, \tag{18}$$

where $\alpha_j^* = \alpha_j + \rho_j$ for $j = p_y$ and $\delta_j = (\delta_{j1}, ..., \delta_{jk})$ with $\delta_{ji} = \beta_{ji}\rho_j$. Accordingly, the parameters $\rho_j$ cannot be identified using the coefficient on the lagged dependent variable $\alpha_j^*$ since $\alpha_j$ is unknown. However, as suggested by Wallis (1967), one can obtain consistent estimates using the fact that $\rho_j = \delta_{ji}/\beta_{ji}$. Hence, one simply estimate the regression model (18) by OLS, get estimates $\hat{\beta}_j$ and $\hat{\delta}_{ji}$ and construct the estimates $\hat{\rho}_j^D = \hat{\delta}_{ji}/\hat{\beta}_j$. One can then proceed to construct the FGLS estimates as described in Step 2 above. The only drawback is that if the number of regressors $x_{jt}$ is greater than one, there are multiple choices for each value of $i$. In principle, choosing anyone will lead to a consistent estimate in well specified models. Simulations and applications reported in González-Coya and Perron (2024b) show that the results are not sensitive to the choice of the variable used. This is because GLS is quite robust to small variations in the quasi-differencing parameters $\rho_j$.

In the case of predictive regressions assuming rational expectations, only lagged dependent variables may be included as regressors, in which case the procedure described above is not applicable. These take the form $y_{t+h} = \beta_0 + \sum_{j=1}^{m} \beta_j y_{t-j} + u_{t+h}$, where $m < h$. For instance, in Hansen and Hodrick (1980), $h = 13$ and $m = 2$ with $y_{t+h} = s_{t+h} - f_t$, where $s_{t+h}$ is the (log) spot exchange rate at time $t+h$ and $f_t$ the (log) $h$-period forward exchange rate at time $t$. Under rational expectations, all coefficients should be 0. Both OLS and GLS are consistent since past regressors are uncorrelated with $u_{t+k}$, even if the latter have an $MA(12)$ structure given the assumption of rational expectations. One can construct consistent estimates of $\rho_j$ using the OLS residuals, say $\widetilde{u}_t$, given that OLS is consistent. Let the fitted value obtained for an OLS regression of $\widetilde{u}_t$ on $k_T$ lags be $\widetilde{u}_t = \sum_{j=1}^{k_T^*} \hat{\rho}_j^O \widetilde{u}_{t-j} + \widetilde{e}_{tk}$. The FGLS estimates are obtained using $\hat{\rho}_j^O$ instead of $\hat{\rho}_j^D$ in (16). If rational expectations does not hold so that the errors are, say, an $AR(p)$ process (e.g., adaptive expectations), then both OLS and FGLS are inconsistent, though infeasible GLS remains consistent. Another example is the estimation of local projections to estimate impulse response functions as suggested by Jordà (2005). For a particular equation of the VAR model, these take the form

$$y_{t+s} = c^s + B_1^{s+1} y_{t-1} + B_2^{s+1} y_{t-2} + ... + B_p^{s+1} y_{t-p} + u_{t+s}, \tag{19}$$

for $s = 0, 1, ..., h$, where $h$ is the maximal horizon for the impulse response functions. Then, $u_{t+s}$ is a forecast error having an $MA(s-1)$ structure, uncorrelated with all past values of the variables. OLS is consistent and we can estimate the parameters for the autoregressive representation using the OLS estimates $\hat{u}_{t+s}$ obtained from (19). Lusompa (2023) presents an alternative more complex procedure.

### 5.2 Issues related to pre-determined regressors

The crucial condition for GLS to be consistent is that the regressors be pre-determined. of course it is possible to concoct a model which implies that OLS is consistent while GLS is not because the regressors are not pre-determined. Take the following example:

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, \ldots, T, \tag{20}$$

$$x_t = v_t + \eta_t, \quad u_t = \rho_u u_{t-1} + \varepsilon_t^u + \lambda \eta_{t-1} = \rho_u u_{t-1} + e_t, \tag{21}$$

where $e_t = \varepsilon_t^u + \lambda \eta_{t-1}$, $\eta_t, \varepsilon_t^u \sim i.i.d.N(0,1)$ are independent of each other. We allow $v_t$ to be serially correlated, with $v_t = \rho_v v_{t-1} + \varepsilon_t^v$, where $\varepsilon_t^v \sim i.i.d.N(0,1)$ independent of $\eta_t$ and

$\varepsilon_t^u$. It is then the case that $E(x_t u_t) = 0$ so that OLS is consistent and when using (20) as the regression, $E(x_{t-1} e_t) \neq 0$, so that GLS is inconsistent. This is indeed the case. Note, however, that allowing $\eta_t$ to be serially correlated renders OLS inconsistent. This case is one with an unobserved variable in the innovations correlated with only the past regressors. If we simply change $\eta_{t-1}$ in (21) to $\eta_t$ or allow $\eta_t$ to be serially correlated, OLS, GLS, Durbin and so one are no longer consistent. What is common is the case with $\eta_{t-1}$ being an observed variable; e.g., the lagged value of some covariate $x_{t-1}$. Here, one can simply introduce $x_{t-1}$ as a regressor and use the regression

$$y_t = \alpha + \beta x_t + \delta x_{t-1} + u_t^*, \quad t = 1, \ldots, T. \tag{22}$$

The error term $u_t^* = \rho_u u_{t-1}^* + \varepsilon_t^u$ is then purged of the component $\eta_{t-1}$ and one can apply GLS provided the lagged values $\{x_{t-2}, x_{t-3}, ...\}$ are not subject to any other source of correlation with $e_t$ independent of $\eta_{t-1}$. In other words, all lagged vales of $x_{t-1}$ can be a function of $\eta_{t-1}$ but not correlated with $u_t$ via some other independent component. If that would be the case then, one could simply add a further lagged value $x_{t-2}$ as a regressor in (22). And so on, if needed. Hence, with innovations affected by omitted observable variables, the problem is easy to fix. Simply include enough lags of the covariates as regressors. This is in fact the reason why Baillie et al. (2024) advocate using the Durbin regression as a means to have estimates robust to non-predetermined regressors. They include all lags of both the dependent and original regressors as covariates. Doing so, they lose considerable efficiency. Our aim is geared to provide an efficient method.

One can test whether the regressors are pre-determined or not. What causes the correlation between the innovations and the regressors is of no consequence. It could be some omitted lagged variable, some errors in variables correlated with lagged regressors, or whatever. The fact is that non-determinedness implies correlation between some variables and the errors means that tests can be performed for its potential presence. What is needed are estimates of the residuals based on a consistent estimate of $\beta$ in (20) whether or not exogeneity or pre-determinedness hold. When the omitted variable is observed, this can be achieved via the Durbin regression (12). The main idea is very simple and involves using a standard variable addition test (e.g., Engle (1982)). The steps are the following: a) Estimate the Durbin regression (15) and get the estimate $\hat{\beta}^D$; b) construct the residuals $\hat{u}_t^D = y_t - \hat{\beta}^D x_t$; c) De-mean the residuals to obtain $\widetilde{u}_t^D = \hat{u}_t^D - T^{-1} \sum_{t=1}^T \hat{u}_t^D$; d) Perform an LM test for variable addition using lagged values of $x_t$. This can be done sequentially using the first, then second, and so on lags. Upon a rejection, include the relevant lagged variables as regressors in the main equation (20); e) Apply FGLS as outlined above to this regression. This will lead to a consistent of estimate of $\beta$ with regressors exogenous or not. One can also select the lagged regressors to be included via information criteria, such as the BIC.

When the omitted variable is unobserved, things are more complex. In general, none of the procedures discussed here will be consistent except in some special cases such as the model described in (20). If this type of one-period lag endogeneity is deemed relevant, or some variations that imply the same qualitative results, then one can use the OLS estimate to construct the residuals since it is consistent. Upon a rejection using the variable addition test, GLS or FGLS should not be applied if such lagged endogeneity issues are a concern. If the researcher is confident that the regressors are exogenous and contemporaneously uncorrelated with the innovations, then OLS is preferred as it is consistent, while GLS is not. Baillie et al. (2024) also cannot handle innovations correlated with past regressors via some unobserved variable. Cases with OLS consistent while GLS is not can be viewed as knife-edge cases in the sense that minor changes in the specification renders OLS inconsistent; e.g., adding lagged regressors or having the omitted unobserved variable being serially correlated. Surely, many cases with exogenous regressors and non-pre-determined variables for which OLS is consistent and GLS is not exist. Practitioners must be judicious in applying any method.

## 6 Simulation results

The issues addressed are the following: the bias, variance and MSE of the FGLS estimates as well as the exact coverage rate and lengths of the confidence intervals. We also report similar results for the infeasible GLS procedure that uses the true value of $\Omega$ to construct the estimate $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$, with $Var(\hat{\beta}_{GLS}|X) = \sigma_e^2(X'\Omega^{-1}X)^{-1}$, which is specific to the data-generating process and uses the true values of the parameters. This is done to assess the extent to which the FGLS procedure is able to provide as precise an estimate as possible. For $AR(1)$ processes, we also report results for the Cochrane and Orcutt (1949), labelled CO. For the FGLS procedure, we considered three methods to select the lag length of the autoregressive approximation: AIC (Akaike (1973)), BIC (Schwarz (1978)) and the MAIC suggested by Ng and Perron (2001). The best results were obtained using the BIC, as suggested in Section 5, so we omit results with AIC or MAIC.

It is often the case, with rational expectations models, that the theory predicts $MA(h-1)$ errors whenever forecasts at horizons $h$ are involved. In the simulations, we shall consider errors generated from $MA(1)$ processes. It is useful to also consider an approximate GLS procedure for $MA(1)$ errors for the following reasons: a) an autoregressive approximation selected using the BIC may yield a rather parsimonious model that fails to capture the extent of the serial correlation in the errors; b) we may have prior knowledge that the errors are an $MA(1)$ process. Hence, we also consider the following approximate GLS procedure, labelled, GMA. It is based on the OLS regression $y_t^* = x_t^*\beta + e_t$, where $y_t^* = \sum_{j=0}^{t-1}(-\hat{\theta})^j y_{t-j}$, $x_t^* = \sum_{j=0}^{t-1}(-\hat{\theta})^j x_{t-j}$ with $\hat{\theta}$ the MLE (exact or approximate) of $\theta$ for $\widetilde{u}_t = e_t + \hat{\theta}e_{t-1}$, where $\widetilde{u}_t = y_t - x_t\widetilde{\beta}$ with $\widetilde{\beta}$ the OLS estimate from the regression (15) with $k_T = int[4(T/100)^{2/9}]$.

We consider the DGP $y_t = \alpha + \beta x_{1t} + u_t$. We set $(\alpha, \beta) = (0, 1)$, without loss of generality. The sample size is $T = 200$. For the errors $u_t$, we consider the following specifications: 1) $AR(1)$: $u_t = \rho_u u_{t-1} + e_t$; $\rho_u = \{-0.5, 0.0, 0.2, 0.5, 0.8\}$; 2) $AR(2)$: $u_t = \rho_{u1}u_{t-1} + \rho_{u2}u_{t-2} + e_t$; $(\rho_{u1}, \rho_{u2}) = \{(1.34, -0.42), (0.5, -0.3), (-0.5, 0.3), (0.0, 0.3), (0.5, 0.3)\}$; 3) $MA(1)$: $u_t = e_t + \theta e_{t-1}$; $\theta = \{-0.7, -0.4, 0.5\}$; 4) $ARMA(1,1)$: $u_t = \rho_u u_{t-1} + e_t + \theta e_{t-1}$; $(\rho_u, \theta) = \{(-0.5, -0.4), (0.2, -0.4), (0.2, 0.5), (0.5, -0.4), (0.5, 0.5), (0.8, -0.4), (0.8, 0.5)\}$. Throughout, $e_t \sim i.i.d.\ N(0, 1)$ and $x_{1t} = \rho_x x_{1t-1} + v_t + \gamma e_{t-1}$ with $v_t \sim i.i.d. N(0, 1)$ independent of $e_t$. When $\gamma = 0$, the regressors are exogenous, while $\gamma \neq 0$ imply non-exogenous regressors. We report results for $\rho_x = 0.8$, while the Supplement reports results for $\rho_x = 0$; see Tables S.4-S.7. We use 10,000 replications and $T = 200$, $500$. The results are presented in Tables 2-5. We focus our discussion on the MSE and the confidence intervals.

To construct the confidence intervals, we simply use the fact that, for some given lag length $k_T^*$, the FGLS estimate is simply OLS obtained from the regression (16), so that an estimate of ($T$ times) the asymptotic covariance matrix is $Var(\hat{\beta}_{\text{FGLS}}) = \hat{\sigma}_e^2(X_{k_T^*}'X_{k_T^*})^{-1}$, where $X_{k_T^*} = (x_{k_T^*+1}', ..., x_T')'$, $x_t = (1, x_{1t}^*)$ for $t = k_T^* + 1, ..., T$, with $x_t^* = x_t - \sum_{j=1}^{k_T^*}\hat{\rho}_j^D x_{t-j}$ and $\hat{\sigma}_e^2 = (T - k_T^*)^{-1}\sum_{t=k_T^*+1}^{T}\hat{e}_{tk_T}^2$, with $\hat{e}_{tk_T}$ the OLS residuals from estimating regression (16) by OLS. For the GMA procedure the variance is estimated similarly, except that $Var(\hat{\beta}_{GMA}) = \hat{\sigma}_e^2(X^{*\prime}X^*)^{-1}$, where $X^* = (x_1^{*\prime}, ..., x_T^{*\prime})'$, $x_t^* = (1, x_{1t}^*)$ for $t = 1, ..., T$, with $x_t^* = \sum_{j=0}^{t-1}(-\hat{\theta})^j x_{t-j}$. To construct the confidence interval of the OLS estimate, we use the so-called HAC standard errors based on the weighting scheme suggested by Andrews (1991) with automatic bandwidth selection. This leads to the following estimate of the asymptotic covariance matrix: $Var(\hat{\beta}_{\text{OLS}}) = (T^{-1}X'X)^{-1}\hat{\Sigma}(T^{-1}X'X)^{-1}$, where $\hat{\Sigma} = T^{-1}\sum_{j=-T+1}^{T-1}w(j/m)\hat{\Gamma}_v(j)$ with $\hat{\Gamma}_v(j) = \hat{v}_t\hat{v}_{t-j}'$ for $j \geq 0$ and $\hat{\Gamma}_v(j) = T^{-1}\sum_{t=-j+1}^{T}\hat{v}_{t+j}\hat{v}_t'$ for $j < 0$, and $\hat{v}_t = x_t(y_t - x_t\hat{\beta}_{\text{OLS}})$. We use the quadratic spectral kernel recommended by Andrews (1991) for which $w(z) = (3/z^2)(\sin(z)/z - \cos(z))$, where $z = 6\pi z/5$, and $m$ is the bandwidth parameter constructed using the automatic bandwidth selection using an $AR(1)$ approximation. The confidence intervals are constructed in the usual way, via $\hat{\beta}_{A,i} \pm z_{1-\alpha/2} \cdot Var(\hat{\beta}_A)_{ii}^{1/2}$, where $A$ refers to the estimator (OLS, GLS, FGLS, etc...), $i$ is the index for the coefficient, $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the $N(0,1)$ distribution. We use $\alpha = 0.05$, i.e., two-sided 95% confidence sets. We first present results with exogenous regressors that allows a proper comparison since both OLS and FGLS are consistent.

### 6.1 Simulations with exogenous regressors

Here $\gamma = 0$ and both OLS and FGLS are consistent. The following features are noteworthy: 1) The MSE of the FGLS estimate is never higher than when using OLS. It can be dramatically lower; e.g., the empirically relevant case of the $AR(2)$ with parameters 1.34 and -0.42 for which the reduction is 96% when $T = 200$. Overall, the reductions can be very substantial. 2) In most cases, the MSE of FGLS are near those obtained using the infeasible GLS, so the suggested procedure nearly achieves the best possible outcome. This is even the case for processes having an MA component, which are notoriously difficult to approximate using low order autoregressions. 3) When the error process is strongly correlated the reduction in MSE comes from both a reduction in bias and variance. When the extent of the correlation is small, most of the reduction is due to a decrease in variance. 4) As discussed in Section 3, an $AR(2)$ with parameters (0.0,0.3) causes problems when applying a first-order correction. This is no longer the case selecting $k_T$ using the BIC. 5) For the $AR(1)$ case, using the Cochrane and Orcutt (1949) procedure (valid here because of exogenous regressors) yields results that are nearly identical to using the more general method advocated. This shows that FGLS adapts well to the generating process in that a method tailored to work for an $AR(1)$ does not perform better. 6) For the $MA(1)$ case, the GMA performs as well as FGLS and the infeasible GLS. In all cases, the gains are mostly due to a decrease in variance.

The results for the coverage rates of the confidence intervals with nominal level 95% are presented in the last two column-panels of Tables 2-5. The following features are noteworthy. 1) In most cases, the exact coverage rates for the FGLS method are within 1% of the nominal level, hence not statistically different. This holds even with strong correlation in the errors unlike the method based on OLS, which is subject to high size distortions as extensively documented previously in the literature. The main reason for why the coverage rates of the FGLS estimates are near the nominal 95% level is because it involves residuals that are nearly *i.i.d.*, in which case the Central Limit Theorem (CLT) is a good approximation even for small samples. The OLS method involves the product $x_t u_t$ which can be strongly correlated, in which case a much large sample is needed for the CLT to provide a good approximation. 2) The length of the confidence set using FGLS is always shorter than that obtained with OLS. The differences are larger as the process is more strongly correlated. For instance, in the case of the $AR(2)$ with parameters 1.34 and -0.42, the length of the confidence interval with FGLS is 77% smaller. With $i.i.d.$ regressors ($\rho_x = 0$), see the Supplement, the same qualitative results hold, though the coverage rates of the confidence intervals for OLS are close to the nominal level 95% in all cases (similar to FGLS) given that $x_t u_t$ is less correlated.

Overall, the simulations show that the suggested FGLS procedure with BIC to select the lag length can do no worse than OLS even with near zero correlation. It yields estimates with much lower MSE, especially as the strength of the serial correlation increases. This is achieved with no cost and some benefits to the coverage rates of the confidence intervals and a substantial reduction in their lengths.

### 6.2 Simulations with non-exogenous regressors

The specifications are the same except that now $\gamma \neq 0$. Accordingly, $x_t$ is not an exogenous regressor, it is simply pre-determined. We consider two values of $\gamma$, namely $\gamma = 0.25$ (weak correlation) and $\gamma = 0.50$ (strong correlation). The results are presented in the second and third horizontal panels of Tables 2-5. Note that the condition $E(x_t u_{t-1}) = 0$ usually used to justify the consistency of GLS is not satisfied. Still, the results will show its irrelevance as FGLS performs very well while OLS very poorly. This accords with the theoretical discussion.

The following features are noteworthy. 1) For the MSE (and bias and variance), much of the same results hold as with exogenous regressors. Again, FGLS performs almost as well as the infeasible GLS. 2) For $MA(1)$ processes the approximate GLS, labelled GMA, performs slightly better than FGLS, when $T = 200$; the differences are substantially reduced when $T = 500$, in which case both performs nearly as well as the infeasible GLS. 3) Across all cases, the main difference is the very large bias and MSE of OLS. For instance, for an $AR(1)$ with parameter $\rho_u = 0.8$, the MSE is about 23 times larger than FGLS when $T = 200$ and $\gamma = 0.5$ (and 55 times larger when $T = 500$). There are even more pronounced examples like the $AR(2)$ with parameters $(1.34, -0.42)$ for which the differences are 149 times larger when $T = 200$ and 363 times when $T = 500$. Both the bias and variance of OLS are much larger than those with FGLS for both $T = 200, 500$, given that OLS is inconsistent.

The results for the coverage rates of the confidence intervals are presented in the last two column segments of Tables 2-5. The following features are noteworthy. 1) The results for OLS are meaningless. The coverage rates are all over the map and can be near 0 with strong correlation in the errors. Also, they get noticeably worse as $T$ increases. 2) For FGLS, the coverage rates are near 95% for $AR(1)$ errors. For $AR(2)$ errors, we see some less accurate coverage rates for $\gamma = 0.5$. 3) For $MA(1)$ errors, the coverage rates of GMA and FGLS are good when $\gamma = 0.25$, but more precise with GMA when $\gamma = 0.5$. 4) For $ARMA(1,1)$ errors, the coverage rates of FGLS are good for $\gamma = 0.25$ but less so for $\gamma = 0.5$.

The results for the case with *i.i.d.* regressors ($\rho_x = 0$) are presented in the Supplement. The same qualitative conclusions hold. Overall, the simulations show that the suggested FGLS procedure with BIC to select the lag length is by far superior compared to OLS.

**Remark 7.** *As discussed in Section 2.1.1 and Remark 4, in the rational expectations case, both OLS and GLS are consistent. Simulation experiments in the Supplement show that, with exogenous or non-exogenous regressors, FGLS is by far superior to OLS in terms of MSE and length of the coverage rates, with results similar to the case with exogenous regressors.*

**Remark 8.** *González-Coya and Perron (2024b) present simulation results about the power of tests on $\beta$ for cases calibrated to real data. With exogenous regressors, the tests based on all methods have nearly the correct size while FGLS has the highest power by a wide margin over the Durbin regression and OLS, which have very little power. When the regressors are non-exogenous, OLS has distorted size, as expected, but otherwise the relative power functions remain the same. The poor performance of the tests based on the Durbin regression steams from the fact that with regressors that are serially correlated, as is usually the case, the introduction of all the lagged regressors creates a collinearity problem that inflates the MSE of the estimates and thereby reduces power. This is avoided when using FGLS since the final regression is a simple transformation of the original regressors. Also, the relative power can be deduced by looking at the length of the confidence intervals, which we report.*

**Remark 9.** *If heteroskedasticity in the innovations is a concern, two avenues are possible. The first is to correct the standard errors using a heteroskedasticity-robust covariance matrix as suggested by, e.g., White (1980) or subsequent variations. Our recommendation is to apply a further FGLS correction as suggested by González-Coya and Perron (2024a). It is based on an Adaptive Lasso procedure to fit the skedastic function. The method and some simulation results are presented in the Supplement, Section S.5. Overall, further reduction in the MSE are possible even using incorrect covariates to estimate the skedastic function as long as there is some correlation between the covariates used in the Lasso specification and those in the true skedastic function. The coverage rate of the confidence intervals have an exact size close to the nominal level and the lengths are smaller compared to applying OLS or correcting only for serial correlation. With homoskedastic innovations, the results are equivalent to those obtained correcting only for serial correlation. Hence, correcting for heteroskedasticity when it is not present has no detrimental effect.*

### 6.3 The case with a non-invertible process

We now consider the case with non-invertible errors with the roots of $C(L)$ inside the unit circle. For motivation, let us revisit the example discussed in Section 2.1.1. The predictive model states that $E(y_{t+k}|\Phi_t) = x_t'\beta$, where $\Phi_t$ is the information set available at time $t$. Then, $y_{t+k} = x_t'\beta + u_{t+k}$, with $u_{t+k} = y_{t+k} - E(y_{t+k}|\Phi_t)$ so that the error terms are forecast errors from using the best predictor based on $x_t$. It can be shown that $u_{t+k}$ is an $MA(k-1)$ process of the form $u_{t+k} = e_{t+k} + c_1 e_{t+k-1} + .... + c_{k-1}e_{t+1}$, with $e_t \sim i.i.d.\ (0, \sigma_e^2)$. Since $x_t \subset \Phi_t$, $E(x_t u_{t+k}) = 0$, OLS is consistent and can be applied with the relevant HAC correction. For simplicity and without loss of generality we shall restrict ourselves to the case of $MA(1)$ errors. Suppose that $y_t$ is an $AR(2)$ process with parameters $(1.34, -0.42)$. Suppose that $k = 2$, then $u_{t+k}$ is an $MA(1)$ with parameter 1.34. Hence, the root is inside the unit circle and the process is non-invertible. In this case, OLS is consistent since it only requires $E(x_t u_{t+2}) = 0$ which is guaranteed by the rational expectations hypothesis.

Things are more complex with GLS. First, there does not exist a matrix $D$ such that $D'D = \Omega^{-1}$ and $Du = e$, with the vector of innovations having elements $e_t$ for $t = k, ..., T$, even in large samples. Continuing with the $MA(1)$ example with $u_{t+2} = e_t + ce_{t-1}$, we have that the covariance matrix of $u$ when the MA parameter is $c$ is simply a scaled version of the covariance matrix of $u$ when the MA parameter is $c^{-1}$. Hence, the GLS estimates are the same using either values since the scale factor cancels. In other words, let $\beta_{GLS}(c)$ and $\beta_{GLS}(c^{-1})$ be the GLS estimates with MA parameter $c$ or $c^{-1}$, then $\beta_{GLS}(c) = \beta_{GLS}(c^{-1})$. Does that mean GLS is inconsistent? No. It is simply a consequence of the well known observational equivalence. If two processes are observationally equivalent, then estimators based on them will be identical. We can gain some insights by looking at what the transformation of the model by pre-multiplying by $D$ does. It applies a filter such that $\alpha(L)u_{t+k}$ has the same autocovariance function as $C(L)e_{t+k}$. If $C(L)$ is invertible, then $\alpha(L) = C(L)^{-1}$ and $\alpha(L)u_{t+k} = e_{t+k}$. This is the case discussed above with consistent and efficient GLS estimates.

When the process is non-invertible, the transformation will involve the observationally equivalent representation with $\alpha(L) = (1 + c^{-1}L)$. A researcher using the invertible model would not recover the true structural shocks, but rather

$$\begin{aligned}(1-c^{-1}L)^{-1}u_{t+k} &= (1-c^{-1}L)^{-1}(1-cL)e_{t+k} = (1-c^{-1}L)^{-1}(1-c^{-1}L+c^{-1}L-cL)e_{t+k} \\ &= e_t + (c^{-1}-c)(1-c^{-1}L)^{-1}e_{t+k-1} = e_t + (c^{-1}-c)\textstyle\sum_{i=0}^{\infty}(c^{-1})^i e_{t+k-1-i}.\end{aligned}$$

A discussion of these issues is contained in Hannan (1971) and Rozanov (1967). The problem is with the second term, which involves all past values of the innovations. Since the $DX$

involves past values of $x_t$, GLS will be consistent with exogenous regressors but will be inconsistent otherwise. If we consider a model of the form $y = X\beta + u$, with $u_t$ a general non-invertible process that is correlated beyond period $t$, e.g., some non-invertible ARMA process, then both OLS and GLS fail to be consistent. The problem is that it is very difficult, given the observational equivalence between the non-invertible and invertible representations, to ascertain whether the process is invertible or not.

## 7 Conclusions

We showed that, contrary to the widely held view, a) OLS is, in general, inconsistent with non-exogenous regressors, while GLS is consistent; 2) GLS is very robust in that an incorrect specification still allows a lower MSE than OLS; 3) a simple FGLS procedure based on estimating an approximating $AR(k_T^*)$ process with $k_T^*$ chosen using the BIC works very well and delivers estimates that a) are by far superior to OLS (lower MSE); b) robust to a wide variety of data-generating process; c) have confidence intervals with exact coverage rates close to the nominal level with length much shorter than with OLS. With heteroskedastic innovations, a method is suggested to further improve the precision of the estimate.

We used the simple linear model as it is the leading case of interest. Our results should extend to more complex non-linear models estimated by non-linear least-squares or the generalized method of moments approach. A similar treatment for models with endogenous regressors contemporarily correlated with the innovations and estimated via some instrumental variable procedure is covered in Olivari and Perron (2024). Our results provides a strong case for abandoning the often-used OLS+HAC approach so common nowadays. In most cases, it is outright inconsistent in the case of non-exogenous regressors, while GLS is consistent. Even if the regressors are exogenous, GLS yields estimates with substantially lower MSE and confidence intervals with adequate coverage rates and shorter lengths. This holds whether the regressors are exogenous or not, provided a) the regressors are predetermined, i.e., past regressors are not correlated with some unobserved component in the contemporaneous innovations, and b) the stationary linear error process is invertible.

### References


Akaike, H. (1973). "Information theory and an extension of maximum likelihood principle." *Proc. 2nd Int. Symp. on Information Theory.* 267–281.

Amemiya, T. (1973). "Generalized least squares with an estimated autocovariance matrix." *Econometrica*, 41, 723–732.

Amemiya, T. (1986). *Advanced Econometrics.* Harvard University Press.

Andrews, D. W. K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59, 817–858.

Baillie, R. T., Diebold, F. X., Kapetanios, G., Kim, K. H., and Mora, A. (2024). "On robust inference in time series regression." *National Bureau of Economic Research, Woking paper No. 32554.*

Berk, K. N. (1974). "Consistent autoregressive spectral estimates." *The Annals of Statistics*, 2, 489–502.

Choudhury, A. H., Hubata, R., and Louis, R. D. S. (1999). "Understanding time-series regression estimators." *The American Statistician*, 53, 342–348.

Cochrane, D. and Orcutt, G. H. (1949). "Application of least squares regression to relationships containing auto-correlated error terms." *Journal of the American Statistical Association*, 44, 32–61.

Durbin, J. (1970). "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables." *Econometrica*, 38, 410–421.

Engel, C. (1996). "The forward discount anomaly and the risk premium: A survey of recent evidence." *Journal of Empirical Finance*, 3, 123–192.

Engle, R. F. (1974). "Specification of the disturbance for efficient estimation." *Econometrica*, 42, 135–146.

Engle, R. F. (1982). "A general approach to Lagrange multiplier diagnostics." *Journal of Econometrics*, 20, 83–104.

Fama, E. F. (1984). "Forward and spot exchange rates." *Journal of Monetary Economics*, 14, 319–338.

Fang, Y., Koreisha, S. G., and Shao, Q. (2023). "Revisiting the use of generalized least squares in time series regression models." *Journal of Data Science*, 1–19.

Flood, R. P. and Garber, P. M. (1980). "A pitfall in estimation of models with rational expectations." *Journal of Monetary Economics*, 6, 433–435.

Godfrey, L. G. (1978). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables." *Econometrica*, 46, 1293–1302.

González-Coya, E. and Perron, P. (2024a). "Estimation in the presence of heteroskedasticty of unknown form: a Lasso-based approach." *Journal of Econometric Methods*, 13, 29–48.

González-Coya, E. and Perron, P. (2024b). "Uncovering bias in uncovered interest parity tests." *Unpublished Manuscript, Depertment of Economics, Boston University.*

Hannan, E. J. (1971). "Non-linear time series regression." *Journal of Applied Probability*, 8, 767–780.

Hannan, E. J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems.* Wiley, New York.

Hannan, E. J. and Kavalieris, L. (1986). "Regression, autoregressive models." *Journal of Time Series Analysis*, 7, 27–49.

Hansen, B. E. and West, K. D. (2002). "Generalized method of moments and macroeconomics." *Journal of Business and Economic Statistics*, 20, 460–469.

Hansen, L. P. (1982). "Large sample properties of Generalized Method of Moments estimators." *Econometrica*, 50, 1029–1054.

Hansen, L. P. and Hodrick, R. J. (1980). "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis." *Journal of Political Economy*, 88, 829–853.

Jordà, Ò. (2005). "Estimation and inference of impulse responses by local projections." *American Economic Review*, 95, 161–182.

Judge, G. C., Griffiths, W. E., Hill, R. C., Lutkepohl, H., and Lee, T.-C. (1985). *The Theory and Practice of Econometrics, 2nd ed.* New York: Wiley.

Lusompa, A. (2023). "Local projections, autocorrelation, and efficiency." *Quantitative Economics*, 14, 1199–1220.

Newey, W. and West, K. (1987). "A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, 55, 703–708.

Ng, S. and Perron, P. (2001). "Lag length selection and the construction of unit root tests with good size and power." *Econometrica*, 69, 1519–1554.

Ng, S. and Perron, P. (2005). "A note on the selection of time series models." *Oxford Bulletin of Economics and Statistics*, 67, 115–134.

Olivari, M. and Perron, P. (2024). "GLS-IV for time series regressions." *Manuscript, Department of Economics, Boston University.*

Perron, P. (2021). "Conditions for OLS and GLS to be consistent in models with serially correlated errors." *Unpulished Manuscript, Department of Eonomics, Boston University.*

Prais, S. and Winsten, C. (1954). "Trend estimators and serial correlation." *Cowles Foundation, Discussion Paper 383.*

Rozanov, Y. A. (1967). *Stationary Random Processes.* Holden-Day, San Francisco.

Schwarz, G. (1978). "Estimating the dimension of a model." *The Annals of Statistics*, 6, 461–464.

Stock, J. H. and Watson, M. W. (2019). *Introduction to Econometrics, 4th Ed.* Pearson, New York, NY.

Wallis, K. F. (1967). "Lagged dependent variables and serially correlated errors: a reappraisal of three-pass least squares." *The Review of Economics and Statistics*, 49, 555–567.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica*, 48, 817–838.

White, H. and Domowitz, I. (1984a). "Nonlinear regression with dependent observations." *Econometrica*, 52, 143–161.

White, H. and Domowitz, I. (1984b). "Nonlinear regression with dependent observations." *Econometrica*, 52, 143–161.

Table 1: Root mean squared errors, bias and variance of estimators of $\beta$ and $\rho$; AR(1) model.

| | $\beta$ | | | | | $\rho$ | |
|---|---|---|---|---|---|---|---|
| | OLS | Durbin | GLS | FGLS | CO-FGLS | FGLS | CO-FGLS |
| RMSE | 0.400 | 0.036 | 0.025 | 0.025 | 0.041 | 0.034 | 0.175 |
| Bias | 0.400 | 0.029 | 0.012 | 0.020 | 0.035 | 0.027 | 0.171 |
| Variance | 0.0031 | 0.0013 | 0.0006 | 0.0006 | 0.0008 | 0.0010 | 0.0013 |

Table 2: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, AR(1) case with $\rho_x = 0.8$. (First 3 columns are multiplied by 100).

| | | | MSE | | | | Bias | | | | Variance | | | | Coverage | | | Lenght | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AR(1) | OLS | GLS | CO | FGLS | OLS | GLS | CO | FGLS | OLS | GLS | CO | FGLS | OLS | CO | FGLS | OLS | CO | FGLS |
| $T = 200$ | $\gamma = 0$ | -0.5 | 0.11 | 0.10 | 0.10 | 0.10 | 2.64 | 2.45 | 2.45 | 2.44 | 0.13 | 0.10 | 0.10 | 0.10 | 0.95 | 0.95 | 0.95 | 0.14 | 0.12 | 0.12 |
| | | 0 | 0.19 | 0.19 | 0.19 | 0.19 | 3.46 | 3.46 | 3.46 | 3.45 | 0.20 | 0.20 | 0.20 | 0.19 | 0.95 | 0.95 | 0.95 | 0.17 | 0.17 | 0.17 |
| | | 0.2 | 0.27 | 0.26 | 0.26 | 0.26 | 4.14 | 4.04 | 4.05 | 4.06 | 0.26 | 0.27 | 0.27 | 0.26 | 0.93 | 0.95 | 0.94 | 0.20 | 0.20 | 0.20 |
| | | 0.5 | 0.59 | 0.40 | 0.41 | 0.41 | 6.06 | 5.04 | 5.09 | 5.10 | 0.50 | 0.42 | 0.41 | 0.40 | 0.91 | 0.94 | 0.94 | 0.27 | 0.25 | 0.25 |
| | | 0.8 | 2.27 | 0.50 | 0.51 | 0.51 | 11.91 | 5.66 | 5.72 | 5.72 | 1.68 | 0.51 | 0.50 | 0.50 | 0.88 | 0.95 | 0.95 | 0.50 | 0.28 | 0.28 |
| | $\gamma = 0.25$ | -0.5 | 0.24 | 0.09 | 0.09 | 0.09 | 4.01 | 2.37 | 2.38 | 2.39 | 0.12 | 0.09 | 0.09 | 0.09 | 0.85 | 0.96 | 0.95 | 0.14 | 0.12 | 0.12 |
| | | 0 | 0.18 | 0.18 | 0.18 | 0.18 | 3.36 | 3.36 | 3.40 | 3.37 | 0.18 | 0.19 | 0.18 | 0.18 | 0.95 | 0.96 | 0.95 | 0.17 | 0.17 | 0.17 |
| | | 0.2 | 0.27 | 0.24 | 0.25 | 0.27 | 4.21 | 3.93 | 4.00 | 4.11 | 0.24 | 0.26 | 0.25 | 0.25 | 0.92 | 0.95 | 0.94 | 0.19 | 0.20 | 0.19 |
| | | 0.5 | 0.98 | 0.38 | 0.41 | 0.41 | 8.16 | 4.93 | 5.12 | 5.14 | 0.45 | 0.39 | 0.38 | 0.38 | 0.78 | 0.93 | 0.94 | 0.26 | 0.24 | 0.24 |
| | | 0.8 | 5.34 | 0.48 | 0.55 | 0.51 | 19.60 | 5.50 | 5.89 | 5.72 | 1.49 | 0.48 | 0.47 | 0.47 | 0.62 | 0.93 | 0.94 | 0.47 | 0.27 | 0.27 |
| | $\gamma = 0.5$ | -0.5 | 0.50 | 0.08 | 0.08 | 0.08 | 6.13 | 2.22 | 2.27 | 2.27 | 0.11 | 0.08 | 0.08 | 0.08 | 0.58 | 0.95 | 0.95 | 0.13 | 0.11 | 0.11 |
| | | 0 | 0.16 | 0.16 | 0.17 | 0.16 | 3.14 | 3.14 | 3.23 | 3.17 | 0.16 | 0.16 | 0.16 | 0.16 | 0.95 | 0.94 | 0.94 | 0.15 | 0.16 | 0.15 |
| | | 0.2 | 0.28 | 0.21 | 0.22 | 0.27 | 4.29 | 3.65 | 3.77 | 4.12 | 0.19 | 0.22 | 0.21 | 0.21 | 0.88 | 0.93 | 0.93 | 0.17 | 0.18 | 0.18 |
| | | 0.5 | 1.74 | 0.32 | 0.40 | 0.40 | 11.82 | 4.53 | 5.11 | 5.05 | 0.34 | 0.34 | 0.31 | 0.32 | 0.47 | 0.88 | 0.92 | 0.23 | 0.22 | 0.22 |
| | | 0.8 | 11.17 | 0.40 | 0.79 | 0.48 | 31.21 | 5.05 | 7.00 | 5.53 | 1.09 | 0.41 | 0.40 | 0.40 | 0.19 | 0.82 | 0.93 | 0.40 | 0.25 | 0.25 |
| $T = 500$ | $\gamma = 0$ | -0.5 | 0.04 | 0.04 | 0.04 | 0.04 | 1.67 | 1.56 | 1.56 | 1.56 | 0.05 | 0.04 | 0.04 | 0.04 | 0.95 | 0.94 | 0.94 | 0.08 | 0.07 | 0.07 |
| | | 0 | 0.08 | 0.08 | 0.08 | 0.08 | 2.21 | 2.22 | 2.22 | 2.21 | 0.07 | 0.07 | 0.07 | 0.07 | 0.94 | 0.94 | 0.94 | 0.11 | 0.11 | 0.11 |
| | | 0.2 | 0.11 | 0.11 | 0.11 | 0.11 | 2.64 | 2.60 | 2.60 | 2.60 | 0.10 | 0.10 | 0.10 | 0.10 | 0.93 | 0.94 | 0.94 | 0.12 | 0.13 | 0.13 |
| | | 0.5 | 0.23 | 0.17 | 0.17 | 0.17 | 3.85 | 3.28 | 3.27 | 3.27 | 0.20 | 0.16 | 0.16 | 0.16 | 0.92 | 0.94 | 0.94 | 0.18 | 0.16 | 0.16 |
| | | 0.8 | 0.90 | 0.21 | 0.21 | 0.21 | 7.59 | 3.68 | 3.67 | 3.67 | 0.77 | 0.20 | 0.20 | 0.20 | 0.92 | 0.94 | 0.94 | 0.34 | 0.18 | 0.18 |
| | $\gamma = 0.25$ | -0.5 | 0.15 | 0.04 | 0.04 | 0.04 | 3.40 | 1.51 | 1.51 | 1.51 | 0.04 | 0.03 | 0.03 | 0.03 | 0.66 | 0.95 | 0.95 | 0.08 | 0.07 | 0.07 |
| | | 0 | 0.07 | 0.07 | 0.07 | 0.07 | 2.14 | 2.14 | 2.16 | 2.15 | 0.07 | 0.07 | 0.07 | 0.07 | 0.95 | 0.94 | 0.95 | 0.10 | 0.10 | 0.10 |
| | | 0.2 | 0.13 | 0.10 | 0.10 | 0.10 | 2.89 | 2.51 | 2.54 | 2.56 | 0.09 | 0.10 | 0.10 | 0.10 | 0.89 | 0.94 | 0.94 | 0.12 | 0.12 | 0.12 |
| | | 0.5 | 0.67 | 0.16 | 0.17 | 0.17 | 7.05 | 3.16 | 3.24 | 3.24 | 0.18 | 0.15 | 0.15 | 0.15 | 0.64 | 0.94 | 0.94 | 0.17 | 0.15 | 0.15 |
| | | 0.8 | 4.15 | 0.20 | 0.21 | 0.20 | 18.38 | 3.55 | 3.67 | 3.60 | 0.68 | 0.19 | 0.19 | 0.19 | 0.41 | 0.93 | 0.94 | 0.32 | 0.17 | 0.17 |
| | $\gamma = 0.5$ | -0.5 | 0.36 | 0.03 | 0.03 | 0.03 | 5.57 | 1.40 | 1.42 | 1.41 | 0.04 | 0.03 | 0.03 | 0.03 | 0.24 | 0.95 | 0.95 | 0.08 | 0.07 | 0.07 |
| | | 0 | 0.06 | 0.06 | 0.07 | 0.06 | 1.99 | 1.99 | 2.04 | 2.00 | 0.06 | 0.06 | 0.06 | 0.06 | 0.95 | 0.94 | 0.94 | 0.10 | 0.10 | 0.10 |
| | | 0.2 | 0.17 | 0.09 | 0.09 | 0.10 | 3.47 | 2.33 | 2.39 | 2.51 | 0.08 | 0.08 | 0.08 | 0.08 | 0.77 | 0.93 | 0.93 | 0.11 | 0.11 | 0.11 |
| | | 0.5 | 1.52 | 0.13 | 0.18 | 0.16 | 11.71 | 2.92 | 3.39 | 3.21 | 0.14 | 0.13 | 0.12 | 0.13 | 0.14 | 0.88 | 0.92 | 0.15 | 0.14 | 0.14 |
| | | 0.8 | 10.43 | 0.17 | 0.35 | 0.19 | 31.35 | 3.26 | 4.68 | 3.45 | 0.50 | 0.16 | 0.16 | 0.16 | 0.01 | 0.82 | 0.93 | 0.27 | 0.16 | 0.16 |

Table 3: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, AR(2) case with $\rho_x = 0.8$. (First 3 columns are multiplied by 100).

| | | | MSE | | | Bias | | | Variance | | | Coverage | | Lenght | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AR(2) | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | FGLS | OLS | FGLS |
| $T = 200$ | $\gamma = 0$ | 0.5,-0.3 | 0.32 | 0.28 | 0.29 | 4.54 | 4.23 | 4.28 | 0.34 | 0.29 | 0.28 | 0.95 | 0.95 | 0.23 | 0.21 |
| | | -0.5,0.3 | 0.17 | 0.13 | 0.13 | 3.22 | 2.83 | 2.86 | 0.16 | 0.13 | 0.13 | 0.94 | 0.95 | 0.16 | 0.14 |
| | | 1.34,-0.42 | 11.45 | 0.42 | 0.42 | 26.96 | 5.15 | 5.19 | 8.09 | 0.40 | 0.39 | 0.87 | 0.94 | 1.08 | 0.25 |
| | | 0,0.3 | 0.31 | 0.26 | 0.28 | 4.39 | 4.09 | 4.20 | 0.22 | 0.27 | 0.26 | 0.90 | 0.93 | 0.18 | 0.20 |
| | | 0.5,0.3 | 1.86 | 0.47 | 0.48 | 10.79 | 5.50 | 5.60 | 1.27 | 0.47 | 0.45 | 0.86 | 0.94 | 0.43 | 0.27 |
| | $\gamma = 0.25$ | 0.5,-0.3 | 0.36 | 0.26 | 0.28 | 4.83 | 4.04 | 4.18 | 0.31 | 0.27 | 0.26 | 0.92 | 0.94 | 0.22 | 0.20 |
| | | -0.5,0.3 | 0.21 | 0.12 | 0.13 | 3.61 | 2.74 | 2.84 | 0.16 | 0.12 | 0.12 | 0.91 | 0.94 | 0.15 | 0.13 |
| | | 1.34,-0.42 | 27.51 | 0.39 | 0.41 | 44.90 | 5.00 | 5.16 | 7.29 | 0.37 | 0.37 | 0.58 | 0.94 | 1.02 | 0.24 |
| | | 0,0.3 | 0.32 | 0.25 | 0.28 | 4.58 | 3.94 | 4.20 | 0.20 | 0.26 | 0.24 | 0.86 | 0.92 | 0.18 | 0.19 |
| | | 0.5,0.3 | 3.79 | 0.44 | 0.49 | 16.41 | 5.32 | 5.61 | 1.14 | 0.44 | 0.43 | 0.64 | 0.93 | 0.41 | 0.26 |
| | $\gamma = 0.5$ | 0.5,-0.3 | 0.41 | 0.22 | 0.25 | 5.35 | 3.72 | 4.00 | 0.25 | 0.23 | 0.22 | 0.85 | 0.94 | 0.19 | 0.18 |
| | | -0.5,0.3 | 0.30 | 0.10 | 0.12 | 4.41 | 2.54 | 2.79 | 0.14 | 0.10 | 0.10 | 0.84 | 0.92 | 0.15 | 0.12 |
| | | 1.34,-0.42 | 58.25 | 0.33 | 0.39 | 71.48 | 4.61 | 5.00 | 5.38 | 0.32 | 0.32 | 0.16 | 0.92 | 0.88 | 0.22 |
| | | 0,0.3 | 0.36 | 0.21 | 0.29 | 4.91 | 3.65 | 4.25 | 0.17 | 0.22 | 0.20 | 0.79 | 0.89 | 0.16 | 0.18 |
| | | 0.5,0.3 | 7.50 | 0.38 | 0.49 | 25.20 | 4.90 | 5.60 | 0.83 | 0.37 | 0.36 | 0.25 | 0.91 | 0.35 | 0.24 |
| $T = 500$ | $\gamma = 0$ | 0.5,-0.3 | 0.13 | 0.11 | 0.11 | 2.86 | 2.66 | 2.66 | 0.13 | 0.11 | 0.11 | 0.94 | 0.94 | 0.14 | 0.13 |
| | | -0.5,0.3 | 0.06 | 0.05 | 0.05 | 2.02 | 1.81 | 1.81 | 0.06 | 0.05 | 0.05 | 0.94 | 0.94 | 0.10 | 0.09 |
| | | 1.34,-0.42 | 4.61 | 0.16 | 0.16 | 17.11 | 3.23 | 3.23 | 3.95 | 0.16 | 0.16 | 0.91 | 0.94 | 0.77 | 0.15 |
| | | 0,0.3 | 0.12 | 0.11 | 0.11 | 2.79 | 2.62 | 2.62 | 0.08 | 0.10 | 0.10 | 0.89 | 0.94 | 0.11 | 0.13 |
| | | 0.5,0.3 | 0.75 | 0.19 | 0.19 | 6.91 | 3.51 | 3.51 | 0.60 | 0.18 | 0.18 | 0.91 | 0.94 | 0.30 | 0.17 |
| | $\gamma = 0.25$ | 0.5,-0.3 | 0.18 | 0.11 | 0.11 | 3.40 | 2.57 | 2.61 | 0.12 | 0.10 | 0.10 | 0.88 | 0.94 | 0.13 | 0.12 |
| | | -0.5,0.3 | 0.11 | 0.05 | 0.05 | 2.67 | 1.76 | 1.78 | 0.06 | 0.05 | 0.05 | 0.86 | 0.94 | 0.09 | 0.08 |
| | | 1.34,-0.42 | 21.68 | 0.15 | 0.16 | 41.99 | 3.12 | 3.19 | 3.51 | 0.15 | 0.15 | 0.40 | 0.94 | 0.72 | 0.15 |
| | | 0,0.3 | 0.16 | 0.10 | 0.11 | 3.21 | 2.54 | 2.61 | 0.08 | 0.10 | 0.10 | 0.82 | 0.94 | 0.11 | 0.12 |
| | | 0.5,0.3 | 2.84 | 0.18 | 0.19 | 14.90 | 3.40 | 3.50 | 0.53 | 0.17 | 0.17 | 0.48 | 0.94 | 0.28 | 0.16 |
| | $\gamma = 0.5$ | 0.5,-0.3 | 0.28 | 0.09 | 0.10 | 4.59 | 2.39 | 2.51 | 0.09 | 0.09 | 0.09 | 0.69 | 0.93 | 0.12 | 0.12 |
| | | -0.5,0.3 | 0.18 | 0.04 | 0.05 | 3.67 | 1.63 | 1.72 | 0.05 | 0.04 | 0.04 | 0.68 | 0.93 | 0.09 | 0.08 |
| | | 1.34,-0.42 | 54.49 | 0.13 | 0.15 | 71.60 | 2.85 | 3.08 | 2.57 | 0.12 | 0.12 | 0.01 | 0.92 | 0.62 | 0.14 |
| | | 0,0.3 | 0.24 | 0.09 | 0.11 | 4.15 | 2.35 | 2.59 | 0.07 | 0.08 | 0.08 | 0.64 | 0.92 | 0.10 | 0.11 |
| | | 0.5,0.3 | 6.90 | 0.15 | 0.19 | 25.24 | 3.10 | 3.45 | 0.39 | 0.15 | 0.15 | 0.04 | 0.92 | 0.24 | 0.15 |

Table 4: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, MA(1) case with $\rho_x = 0.8$. (First 3 columns are multiplied by 100).

| | | | MSE | | | | Bias | | | | Variance | | | | Coverage | | | Lenght | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MA(1) | OLS | GLS | GMA | FGLS | OLS | GLS | GMA | FGLS | OLS | GLS | GMA | FGLS | OLS | GMA | FGLS | OLS | GMA | FGLS |
| $T = 200$ | $\gamma = 0$ | -0.7 | 0.08 | 0.03 | 0.03 | 0.04 | 2.22 | 1.44 | 1.47 | 1.51 | 0.12 | 0.03 | 0.04 | 0.05 | 0.98 | 0.96 | 0.96 | 0.14 | 0.08 | 0.08 |
| | | -0.4 | 0.11 | 0.09 | 0.09 | 0.10 | 2.61 | 2.41 | 2.42 | 2.44 | 0.13 | 0.09 | 0.09 | 0.10 | 0.97 | 0.95 | 0.96 | 0.14 | 0.12 | 0.12 |
| | | 0.5 | 0.42 | 0.35 | 0.36 | 0.38 | 5.11 | 4.67 | 4.70 | 4.84 | 0.36 | 0.34 | 0.33 | 0.32 | 0.92 | 0.95 | 0.93 | 0.23 | 0.23 | 0.22 |
| | $\gamma = 0.25$ | -0.7 | 0.53 | 0.03 | 0.04 | 0.05 | 6.58 | 1.40 | 1.56 | 1.73 | 0.12 | 0.03 | 0.04 | 0.04 | 0.55 | 0.96 | 0.95 | 0.14 | 0.08 | 0.08 |
| | | -0.4 | 0.27 | 0.09 | 0.09 | 0.10 | 4.21 | 2.35 | 2.40 | 2.49 | 0.13 | 0.08 | 0.09 | 0.10 | 0.84 | 0.95 | 0.95 | 0.14 | 0.12 | 0.12 |
| | | 0.5 | 0.55 | 0.33 | 0.34 | 0.37 | 6.01 | 4.54 | 4.64 | 4.79 | 0.33 | 0.32 | 0.31 | 0.30 | 0.85 | 0.94 | 0.93 | 0.22 | 0.22 | 0.22 |
| | $\gamma = 0.5$ | -0.7 | 1.36 | 0.03 | 0.06 | 0.09 | 11.06 | 1.30 | 1.83 | 2.24 | 0.12 | 0.02 | 0.04 | 0.04 | 0.04 | 0.92 | 0.87 | 0.13 | 0.07 | 0.08 |
| | | -0.4 | 0.57 | 0.08 | 0.10 | 0.12 | 6.62 | 2.17 | 2.40 | 2.68 | 0.12 | 0.07 | 0.08 | 0.08 | 0.54 | 0.93 | 0.92 | 0.13 | 0.11 | 0.11 |
| | | 0.5 | 0.74 | 0.27 | 0.30 | 0.33 | 7.38 | 4.14 | 4.33 | 4.60 | 0.26 | 0.27 | 0.26 | 0.26 | 0.70 | 0.93 | 0.91 | 0.20 | 0.20 | 0.20 |
| $T = 500$ | $\gamma = 0$ | -0.7 | 0.03 | 0.01 | 0.01 | 0.01 | 1.31 | 0.82 | 0.82 | 0.83 | 0.04 | 0.01 | 0.01 | 0.01 | 0.96 | 0.96 | 0.96 | 0.08 | 0.04 | 0.05 |
| | | -0.4 | 0.04 | 0.03 | 0.03 | 0.03 | 1.53 | 1.39 | 1.40 | 1.40 | 0.05 | 0.03 | 0.03 | 0.04 | 0.96 | 0.96 | 0.96 | 0.08 | 0.07 | 0.07 |
| | | 0.5 | 0.15 | 0.13 | 0.13 | 0.13 | 3.05 | 2.80 | 2.81 | 2.85 | 0.14 | 0.13 | 0.13 | 0.13 | 0.94 | 0.95 | 0.95 | 0.15 | 0.14 | 0.14 |
| | $\gamma = 0.25$ | -0.7 | 0.41 | 0.01 | 0.01 | 0.02 | 6.14 | 0.79 | 0.87 | 0.99 | 0.04 | 0.01 | 0.01 | 0.01 | 0.08 | 0.95 | 0.94 | 0.08 | 0.04 | 0.05 |
| | | -0.4 | 0.17 | 0.03 | 0.03 | 0.03 | 3.61 | 1.36 | 1.39 | 1.46 | 0.05 | 0.03 | 0.03 | 0.03 | 0.63 | 0.95 | 0.95 | 0.08 | 0.07 | 0.07 |
| | | 0.5 | 0.31 | 0.12 | 0.12 | 0.13 | 4.64 | 2.74 | 2.78 | 2.85 | 0.13 | 0.12 | 0.12 | 0.12 | 0.77 | 0.95 | 0.94 | 0.14 | 0.14 | 0.13 |
| | $\gamma = 0.5$ | -0.7 | 1.14 | 0.01 | 0.02 | 0.03 | 10.46 | 0.74 | 0.98 | 1.31 | 0.04 | 0.01 | 0.01 | 0.01 | 0.00 | 0.92 | 0.85 | 0.08 | 0.04 | 0.04 |
| | | -0.4 | 0.41 | 0.03 | 0.03 | 0.04 | 6.07 | 1.26 | 1.38 | 1.60 | 0.04 | 0.03 | 0.03 | 0.03 | 0.14 | 0.94 | 0.92 | 0.08 | 0.06 | 0.07 |
| | | 0.5 | 0.60 | 0.10 | 0.11 | 0.13 | 7.14 | 2.55 | 2.67 | 2.84 | 0.10 | 0.10 | 0.10 | 0.10 | 0.40 | 0.93 | 0.92 | 0.13 | 0.13 | 0.12 |

Table 5: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, ARMA(1,1) case with $\rho_x = 0.8$. (First 3 columns are multiplied by 100).

| | | | MSE | | | Bias | | | Variance | | | Coverage | | Lenght | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ARMA(1,1) | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | FGLS | OLS | FGLS |
| | | -0.5,-0.4 | 0.09 | 0.04 | 0.04 | 2.38 | 1.60 | 1.61 | 0.11 | 0.04 | 0.05 | 0.97 | 0.96 | 0.13 | 0.09 |
| | | 0.2,-0.4 | 0.13 | 0.13 | 0.13 | 2.90 | 2.82 | 2.84 | 0.16 | 0.13 | 0.15 | 0.96 | 0.96 | 0.16 | 0.15 |
| | | 0.2,0.5 | 0.59 | 0.39 | 0.41 | 6.07 | 4.93 | 5.10 | 0.51 | 0.39 | 0.38 | 0.92 | 0.94 | 0.28 | 0.24 |
| | $\gamma = 0$ | 0.5,-0.4 | 0.25 | 0.24 | 0.25 | 3.96 | 3.91 | 3.95 | 0.22 | 0.25 | 0.22 | 0.93 | 0.93 | 0.18 | 0.18 |
| | | 0.5,0.5 | 1.30 | 0.43 | 0.46 | 9.00 | 5.19 | 5.40 | 1.05 | 0.43 | 0.43 | 0.90 | 0.94 | 0.40 | 0.26 |
| | | 0.8,-0.4 | 0.88 | 0.43 | 0.46 | 7.41 | 5.21 | 5.41 | 0.59 | 0.43 | 0.41 | 0.86 | 0.93 | 0.30 | 0.25 |
| | | 0.8,0.5 | 5.12 | 0.39 | 0.41 | 17.83 | 4.93 | 5.09 | 3.65 | 0.38 | 0.40 | 0.87 | 0.94 | 0.73 | 0.25 |
| | | -0.5,-0.4 | 0.48 | 0.04 | 0.04 | 6.08 | 1.52 | 1.63 | 0.12 | 0.04 | 0.05 | 0.60 | 0.96 | 0.13 | 0.08 |
| | | 0.2,-0.4 | 0.19 | 0.12 | 0.14 | 3.48 | 2.69 | 2.96 | 0.15 | 0.12 | 0.14 | 0.93 | 0.95 | 0.15 | 0.14 |
| | | 0.2,0.5 | 1.00 | 0.35 | 0.40 | 8.30 | 4.73 | 5.04 | 0.47 | 0.37 | 0.36 | 0.79 | 0.94 | 0.27 | 0.24 |
| $T = 200$ | $\gamma = 0.25$ | 0.5,-0.4 | 0.23 | 0.22 | 0.23 | 3.84 | 3.73 | 3.82 | 0.21 | 0.23 | 0.21 | 0.92 | 0.93 | 0.18 | 0.18 |
| | | 0.5,0.5 | 3.15 | 0.39 | 0.46 | 15.28 | 4.96 | 5.44 | 0.96 | 0.41 | 0.41 | 0.64 | 0.93 | 0.38 | 0.25 |
| | | 0.8,-0.4 | 1.59 | 0.39 | 0.46 | 10.52 | 4.99 | 5.39 | 0.54 | 0.41 | 0.39 | 0.69 | 0.93 | 0.28 | 0.25 |
| | | 0.8,0.5 | 13.75 | 0.35 | 0.43 | 32.12 | 4.67 | 5.21 | 3.30 | 0.36 | 0.38 | 0.56 | 0.93 | 0.70 | 0.24 |
| | | -0.5,-0.4 | 1.23 | 0.04 | 0.05 | 10.38 | 1.48 | 1.82 | 0.12 | 0.03 | 0.04 | 0.08 | 0.93 | 0.13 | 0.08 |
| | | 0.2,-0.4 | 0.32 | 0.11 | 0.19 | 4.64 | 2.61 | 3.36 | 0.13 | 0.10 | 0.12 | 0.81 | 0.90 | 0.14 | 0.13 |
| | | 0.2,0.5 | 1.77 | 0.31 | 0.42 | 11.94 | 4.42 | 5.19 | 0.36 | 0.32 | 0.30 | 0.49 | 0.89 | 0.23 | 0.22 |
| | $\gamma = 0.5$ | 0.5,-0.4 | 0.22 | 0.21 | 0.24 | 3.81 | 3.62 | 3.94 | 0.17 | 0.20 | 0.18 | 0.90 | 0.90 | 0.16 | 0.16 |
| | | 0.5,0.5 | 6.58 | 0.34 | 0.53 | 23.97 | 4.60 | 5.73 | 0.71 | 0.35 | 0.35 | 0.21 | 0.88 | 0.32 | 0.23 |
| | | 0.8,-0.4 | 2.90 | 0.36 | 0.52 | 15.33 | 4.77 | 5.73 | 0.40 | 0.35 | 0.33 | 0.35 | 0.88 | 0.24 | 0.23 |
| | | 0.8,0.5 | 29.47 | 0.30 | 0.53 | 51.13 | 4.36 | 5.73 | 2.39 | 0.31 | 0.32 | 0.14 | 0.88 | 0.59 | 0.22 |
| | | -0.5,-0.4 | 0.03 | 0.02 | 0.02 | 1.45 | 0.98 | 0.98 | 0.04 | 0.02 | 0.02 | 0.96 | 0.96 | 0.08 | 0.05 |
| | | 0.2,-0.4 | 0.05 | 0.05 | 0.05 | 1.83 | 1.76 | 1.77 | 0.06 | 0.05 | 0.05 | 0.95 | 0.96 | 0.09 | 0.09 |
| | | 0.2,0.5 | 0.23 | 0.16 | 0.17 | 3.85 | 3.26 | 3.30 | 0.21 | 0.15 | 0.15 | 0.94 | 0.93 | 0.18 | 0.15 |
| | $\gamma = 0$ | 0.5,-0.4 | 0.10 | 0.10 | 0.10 | 2.51 | 2.50 | 2.50 | 0.09 | 0.09 | 0.09 | 0.93 | 0.93 | 0.11 | 0.12 |
| | | 0.5,0.5 | 0.51 | 0.18 | 0.19 | 5.70 | 3.45 | 3.51 | 0.45 | 0.17 | 0.17 | 0.92 | 0.93 | 0.26 | 0.16 |
| | | 0.8,-0.4 | 0.35 | 0.18 | 0.19 | 4.69 | 3.40 | 3.46 | 0.27 | 0.17 | 0.17 | 0.90 | 0.94 | 0.20 | 0.16 |
| | | 0.8,0.5 | 2.05 | 0.16 | 0.17 | 11.31 | 3.22 | 3.27 | 1.73 | 0.15 | 0.16 | 0.91 | 0.94 | 0.51 | 0.15 |
| | | -0.5,-0.4 | 0.38 | 0.01 | 0.02 | 5.83 | 0.97 | 1.05 | 0.04 | 0.01 | 0.02 | 0.15 | 0.95 | 0.08 | 0.05 |
| | | 0.2,-0.4 | 0.11 | 0.05 | 0.05 | 2.68 | 1.72 | 1.85 | 0.06 | 0.05 | 0.05 | 0.85 | 0.95 | 0.09 | 0.09 |
| | | 0.2,0.5 | 0.68 | 0.15 | 0.17 | 7.15 | 3.06 | 3.25 | 0.19 | 0.14 | 0.14 | 0.64 | 0.93 | 0.17 | 0.15 |
| $T = 500$ | $\gamma = 0.25$ | 0.5,-0.4 | 0.11 | 0.09 | 0.10 | 2.57 | 2.40 | 2.52 | 0.08 | 0.09 | 0.08 | 0.91 | 0.92 | 0.11 | 0.11 |
| | | 0.5,0.5 | 2.41 | 0.16 | 0.19 | 14.06 | 3.20 | 3.46 | 0.40 | 0.16 | 0.16 | 0.41 | 0.93 | 0.25 | 0.16 |
| | | 0.8,-0.4 | 1.14 | 0.17 | 0.19 | 9.28 | 3.27 | 3.49 | 0.24 | 0.16 | 0.16 | 0.55 | 0.94 | 0.19 | 0.16 |
| | | 0.8,0.5 | 10.88 | 0.14 | 0.17 | 30.18 | 2.98 | 3.26 | 1.54 | 0.14 | 0.15 | 0.34 | 0.94 | 0.48 | 0.15 |
| | | -0.5,-0.4 | 1.00 | 0.01 | 0.02 | 9.71 | 0.89 | 1.11 | 0.04 | 0.01 | 0.01 | 0.00 | 0.92 | 0.08 | 0.05 |
| | | 0.2,-0.4 | 0.19 | 0.04 | 0.06 | 3.84 | 1.60 | 1.99 | 0.05 | 0.04 | 0.04 | 0.62 | 0.91 | 0.09 | 0.08 |
| | | 0.2,0.5 | 1.56 | 0.13 | 0.17 | 11.87 | 2.89 | 3.28 | 0.14 | 0.12 | 0.12 | 0.13 | 0.90 | 0.15 | 0.14 |
| | $\gamma = 0.5$ | 0.5,-0.4 | 0.12 | 0.08 | 0.11 | 2.81 | 2.26 | 2.64 | 0.07 | 0.08 | 0.07 | 0.84 | 0.88 | 0.10 | 0.10 |
| | | 0.5,0.5 | 6.04 | 0.15 | 0.22 | 23.86 | 3.07 | 3.71 | 0.30 | 0.14 | 0.14 | 0.02 | 0.89 | 0.21 | 0.15 |
| | | 0.8,-0.4 | 2.66 | 0.15 | 0.19 | 15.54 | 3.06 | 3.47 | 0.18 | 0.14 | 0.13 | 0.08 | 0.90 | 0.16 | 0.14 |
| | | 0.8,0.5 | 27.70 | 0.13 | 0.21 | 51.30 | 2.89 | 3.58 | 1.11 | 0.12 | 0.12 | 0.01 | 0.88 | 0.41 | 0.14 |

**"Feasible GLS for Time Series Regression"**
**by Pierre Perron and Emilio González-Coya**
**Supplementary material for online publication**

In Section S-1, we present the proofs of Theorem 1 and Corollary 2. In Section S-2, we report the simulation discussed in Section 3. Additional simulations results are reported in Section S-3 that complement those in Section 6, while Section S-4 presents simulation results for predictive regressions. Section S-5 discusses our suggested method to correct for possible heteroskedasticity in the errors. The method is presented as well as simulations showing that further reductions in MSE can be achieved.

## S-1 Proof of some results

**Proof of Theorem 1**. The GLS estimator is the OLS estimator of the quasi-differenced equation

$$(y_t - \rho y_{t-1}) = (x_t - \rho x_{t-1})'\beta + e_t, \ (t = 2, ..., T).$$

Let $w_t = u_t - \rho u_{t-1}$ and note that $w_t$ is a filter: $w_t = \psi(L)u_t$ with $\psi(L) = (1 - \rho L)$. Let $\Lambda = E[ww']$ so that

$$\Lambda^{-1} = \begin{bmatrix} 1 & -\rho & & & & \\ -\rho & 1+\rho^2 & -\rho & & 0 & \\ & -\rho & 1+\rho^2 & -\rho & & \\ & & \ddots & & & \\ & 0 & & -\rho & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{bmatrix}.$$

Hence, the GLS estimator can be written as

$$\hat{\beta}_{\text{GLS}} = (X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}y, \hat{\beta}_{\text{GLS}} - \beta = (X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}u.$$

The variance of the GLS estimator is

$$\text{Var}(\hat{\beta}_{\text{GLS}}) = (X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}\Omega\Lambda^{-1}X(X'\Lambda^{-1}X)^{-1}.$$

The OLS estimator can be written as

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'y, \hat{\beta}_{\text{OLS}} - \beta = (X'X)^{-1}X'u.$$

with $\text{Var}(\hat{\beta}_{\text{OLS}}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$. Since both estimators are consistent the limit of their MSE is equivalent to the limit of their variance. We have,

$$\begin{aligned}
\lim_{T\to\infty} T\,\text{Var}(\hat{\beta}_{\text{OLS}}) &= p\lim_{T\to\infty}(T^{-1}X'X)^{-1}T^{-1}X'\Omega X(T^{-1}X'X)^{-1} \\
&= R_x(0)^{-2}2\pi h_{xu}(0).
\end{aligned}$$

Note that $h_{xu}(0)$ is ($2\pi$ times) the spectral density function of the process $z_t = x_t u_t$. By the Convolution Theorem, we have,

$$h_{xu}(\omega) = \int_{-\pi}^{\pi} h_x(\lambda)h_u(\omega-\lambda)d\lambda,$$

and thus

$$h_{xu}(0) = \int_{-\pi}^{\pi} h_x(\lambda)h_u(-\lambda)d\lambda = \int_{-\pi}^{\pi} h_x(\lambda)h_u(\lambda)d\lambda,$$

since $h_u(-\lambda) = h_u(\lambda)$. The asymptotic variance of the GLS estimator is

$$\begin{aligned}
\lim_{T\to\infty} T\,\text{Var}(\hat{\beta}_{\text{GLS}}) &= p\lim_{T\to\infty}(T^{-1}X'\Lambda^{-1}X)^{-1}T^{-1}X'\Lambda^{-1}\Omega\Lambda^{-1}X(T^{-1}X'\Lambda^{-1}X)^{-1} \\
&= ((1+\rho^2)R_x(0) - 2\rho R_x(1))^{-2}2\pi h_{x^*u^*}(0), \qquad \text{(A.1)}
\end{aligned}$$

where $x_t^* = x_t - \rho x_{t-1}$ and $u_t^* = u_t - \rho u_{t-1}$. The spectral density function of $x_t^*$ is thus given by

$$\begin{aligned}
h_{x^*}(\omega) &= |\psi(e^{-i\omega})|^2 h_x(\omega) \\
&= (1-\rho e^{-i\omega})(1-\rho e^{i\omega})h_x(\omega) \\
&= (1+\rho^2 - 2\rho\cos(\omega))h_x(\omega).
\end{aligned}$$

Analogously, the spectral density function of $u_t^*$, is given by

$$h_{u^*}(\omega) = (1+\rho^2 - 2\rho\cos(\omega))h_u(\omega).$$

Hence, the spectral density function at frequency zero of the process $z_t^* = x_t^* u_t^*$ is

$$\begin{aligned}
h_{x^*u^*}(0) &= \int_{-\pi}^{\pi} h_x^*(\lambda)h_u^*(-\lambda)d\lambda \\
&= \int_{-\pi}^{\pi} (1+\rho^2 - 2\rho\cos(\lambda))^2 h_x(\lambda)h_u(\lambda)d\lambda \\
&= (1+\rho^2)^2 h_{xu}(0) - 4\rho(1+\rho^2)\int_{-\pi}^{\pi}\cos(\lambda)h_x(\lambda)h_u(\lambda)d\lambda \\
&\qquad +4\rho^2\int_{-\pi}^{\pi}\cos(\lambda)^2 h_x(\lambda)h_u(\lambda)d\lambda \\
&= (1+\rho^2)^2 h_{xu}(0) - 4\rho(1+\rho^2)\int_{-\pi}^{\pi}\cos(\lambda)h_x(\lambda)h_u(\lambda)d\lambda \\
&\qquad +2\rho^2\int_{-\pi}^{\pi}(1+\cos(2\lambda))h_x(\lambda)h_u(\lambda)d\lambda \\
&= (2\rho^2 + (1+\rho^2)^2)h_{xu}(0) - 4\rho(1+\rho^2)\widetilde{R}_{xu}(1) + 2\rho^2\widetilde{R}_{xu}(2).
\end{aligned}$$

Now, we can write equation (A.1) as

$$\begin{aligned}\lim_{T\to\infty} T\,\mathrm{Var}(\hat\beta_{\mathrm{GLS}}) &= ((1+\rho^2)R_x(0)-2\rho R_x(1))^{-2}2\pi((2\rho^2+(1+\rho^2)^2)h_{xu}(0)\\ &\quad -4\rho(1+\rho^2)\widetilde{R}_{xu}(1)+2\rho^2\widetilde{R}_{xu}(2))\end{aligned}$$

and the ratio of interest is

$$\lim_{T\to\infty}\left(\frac{\mathrm{MSE}(\hat\beta_{\mathrm{GLS}})}{\mathrm{MSE}(\hat\beta_{\mathrm{OLS}})}\right)=\frac{\lim_{T\to\infty} T\,\mathrm{Var}(\hat\beta_{\mathrm{GLS}})}{\lim_{T\to\infty} T\,\mathrm{Var}(\hat\beta_{\mathrm{OLS}})}$$

$$=\frac{R_x(0)^2}{((1+\rho^2)R_x(0)-2\rho R_x(1))^2}\frac{(2\rho^2+(1+\rho^2)^2)h_{xu}(0)-4\rho(1+\rho^2)\widetilde{R}_{xu}(1)+2\rho^2\widetilde{R}_{xu}(2)}{h_{xu}(0)},$$

and thus,

$$\begin{aligned}\lim_{T\to\infty}\left(\frac{\mathrm{MSE}(\hat\beta_{\mathrm{GLS}})}{\mathrm{MSE}(\hat\beta_{\mathrm{OLS}})}\right) &< 1\\ \text{iff } (2\rho^2+(1+\rho^2)^2)-4\rho(1+\rho^2)\frac{\widetilde{R}_{xu}(1)}{h_{xu}(0)}+2\rho^2\frac{\widetilde{R}_{xu}(2)}{h_{xu}(0)} &< ((1+\rho^2)-2\rho\,\mathrm{cor}_x(1))^2\\ \text{iff } \rho^2-2\rho(1+\rho^2)\frac{\widetilde{R}_{xu}(1)}{h_{xu}(0)}+\rho^2\frac{\widetilde{R}_{xu}(2)}{h_{xu}(0)} &< 2\rho^2\,\mathrm{cor}_x(1)^2-2\rho(1+\rho^2)\,\mathrm{cor}_x(1).\square\end{aligned}$$

**Proof of Corollary 1**: Note that if $x_t$ is *i.i.d.*, its spectral density function is $h_x(\omega)=(2\pi)^{-1}R_x(0)$ for all $\omega$. Thus, using the results in Theorem 1:

$$\begin{aligned}h_{xu}(\omega) &= \int_{-\pi}^{\pi} h_x(\lambda)h_u(\lambda)d\lambda = h_x(0)\int_{-\pi}^{\pi} h_u(\lambda)d\lambda\\ &= \frac{1}{2\pi}R_x(0)R_u(0)\end{aligned}$$

and

$$\widetilde{R}_{xu}(1)=\int_{-\pi}^{\pi}\cos(\lambda)h_x(\lambda)h_u(\lambda)d\lambda=h_x(0)\int_{-\pi}^{\pi}\cos(\lambda)h_u(\lambda)d\lambda=\frac{1}{2\pi}R_x(0)R_u(1),$$

$$\widetilde{R}_{xu}(2)=\int_{-\pi}^{\pi}\cos(2\lambda)h_x(\lambda)h_u(\lambda)d\lambda=h_x(0)\int_{-\pi}^{\pi}\cos(2\lambda)h_u(\lambda)d\lambda=\frac{1}{2\pi}R_x(0)R_u(2).$$

Hence,

$$\begin{gathered}\lim_{T\to\infty}\left(\mathrm{MSE}(\hat\beta_{\mathrm{GLS}})/\,\mathrm{MSE}(\hat\beta_{\mathrm{OLS}})\right)<1\\ \text{iff } \rho^2-2\rho(1+\rho^2)\,\mathrm{cor}_u(1)+\rho^2\,\mathrm{cor}_u(2)<0\\ \text{iff } \frac{\rho}{2(1+\rho^2)}(1+\mathrm{cor}_u(2))<\mathrm{cor}_u(1)\quad\text{when } \rho>0,\\ \text{iff } \frac{\rho}{2(1+\rho^2)}(1+\mathrm{cor}_u(2))>\mathrm{cor}_u(1)\quad\text{when } \rho<0.\square\end{gathered}$$

## S-2 Simulations related to Section 3.

We illustrate the issues discussed in Section 3 using simulations. We consider the following DGP:

$$y_t = \alpha + \beta x_t + u_t,$$

where $x_t \sim i.i.d.\ (0, 1)$. We set $(\alpha, \beta) = (0, 1)$, without loss of generality. The sample size is $T = 200$. For the errors $u_t$, we consider the following specifications: 1) $AR(1)$: $u_t = \rho_u u_{t-1} + e_t$; $\rho_u = \{-0.5, 0.0, 0.2, 0.5, 0.8\}$; 2) $AR(2)$: $u_t = \rho_{u1} u_{t-1} + \rho_{u2} u_{t-2} + e_t$; $(\rho_{u1}, \rho_{u2}) = \{(1.34, -0.42), (0.5, -0.3), (-0.5, 0.3), (0.0, 0.3), (0.5, 0.3)\}$; 3) $MA(1)$: $u_t = e_t + \theta e_{t-1}$; $\theta = \{-0.7, -0.4, 0.5\}$; 4) $ARMA(1, 1)$: $u_t = \rho_u u_{t-1} + e_t + \theta e_{t-1}$; $(\rho_u, \theta) = \{(-0.5, -0.4), (0.2, -0.4), (0.2, 0.5), (0.5, -0.4), (0.5, 0.5), (0.8, -0.4), (0.8, 0.5)\}$. Throughout, $e_t \sim i.i.d.\ N(0, \sigma_e^2)$ independent of $x_j$ for all $t$ and $j$ so that the regressors are exogenous, otherwise OLS would be inconsistent and the comparisons meaningless. We set $\sigma_x^2 = \sigma_e^2 = 1$. For all cases, we consider a range of values for the parameters. These are chosen mostly arbitrarily, except for the first pair of the $AR(2)$ case, which are typical estimates for detrended U.S. real GDP; e.g., Blanchard (1981). In all cases, we adopt an AR(1) specification with different values of the quasi-differencing parameter $\rho$. The results are presented in Table S.1. The first column reports the value of $cor_u(1)$ and the main entries are the MSE of GLS relative to the MSE of OLS for various value of $\rho$ in the range $(-0.9, 0.9)$. We shall discuss the purpose of the values reported in the last column later.

It is most instructive to start with the $AR(1)$ case. When $\rho_u = 0$, as expected OLS is best and GLS has higher MSE. When $\rho_u = -0.5$, GLS has lower MSE for all negative values of $\rho$ and, vice versa, when $\rho_u = 0.5, 0.8$, GLS has lower MSE for all positive values of $\rho$. When $\rho_u = 0.2$, a small value, things are more complex. Here, GLS is best when $\rho \in (0.1, 0.4)$ but marginally worse than OLS when $\rho \in (0.5, 0.9)$ (and, of course also worse when $\rho$ is negative). These results are what one would expect from Corollary 1, in particular the fact that when $\rho_u < 0.5$ GLS is better when $0 < \rho < 2\rho_u$. The results for the other cases are qualitatively similar and in accordance with the theory. When $cor_u(1)$ is "large", GLS has smaller MSE than OLS when the sign of the quasi-difference parameter is the same as the sign of $cor_u(1)$. If $cor_u(1)$ is "small" GLS is better when $\rho$ is in the vicinity of $cor_u(1)$. Of special interest is the $AR(2)$ case with $(\rho_{u1}, \rho_{u2}) = (1.34, -0.42)$, which is roughly typical of many macroeconomic time series given the strong serial correlation. In this case, the gains in MSE reduction over OLS are of the order of 95% when $\rho \in (0.6, 0.9)$. These are substantial gains, which can be obtained by merely using an incorrect $AR(1)$ process with a wide range of values of $\rho$. This illustrates strong robustness to using GLS.

The theoretical and simulation results suggest a very simple procedure to obtain a GLS estimate that is (almost) never worse than OLS, subject to very minor random deviations. First use a test for serial correlation at delay one; we use the LM test of Godfrey (1978). If the test does not reject the null hypothesis of no serial correlation, then use OLS. This will occur when $cor_u(1)$ is "small". If the test rejects, estimate $cor_u(1)$ via the sample first-order serial correlation of the OLS residuals. If it is positive (negative), use any positive (negative) value of the quasi-differencing parameter $\rho$. To make clear that any value of $\rho$ will do, in the simulations we simply draw $\rho$ from a Uniform distribution with support $(0.1, 0.9)$ when positive value are required and with support $(-0.9, -0.1)$ when negative values are in order. The results for the relative MSE of GLS over that of OLS are reported in the last column of Table S.1 under the heading "hybrid". They show that this hybrid-GLS procedure yields more precise estimates for all cases, except for few minor cases due to random variations when $cor_u(1)$ is "small". An exception is when $cor_u(1) = 0$ and there is correlation at higher lags; see the $AR(2)$ case with $(\rho_{u1}, \rho_{u2}) = (0.0, 0.3)$. We view this as a knife-edge case.

Tables S.2-S.3 report corresponding results when $x_t$ is an AR(1) process given by $x_t = \rho_x x_{t-1} + v_t$ with $v_t \sim i.i.d.\ N(0,1)$, with $\rho_x = 0.5$ and $\rho_x = 0.8$. The results are qualitatively similar.

**Remark 1.** *In the hybrid procedure discussed above, we use the OLS residuals to construct an estimate of $cor_u(1)$. From the results in Section 2.1, the OLS estimates of the parameters are inconsistent when the regressors are not exogenous. Here, however, the regressors are exogenous. When constructing a FGLS estimate, we do not need this hybrid procedure.*

**Remark 2.** *After the first draft of this paper was completed, we became aware of the work by Koreisha and Fang (2001). They present exact bounds for the relative variance of OLS, GLS and Feasible GLS allowing for misspecification of the process generating the errors when constructing the FGLS estimate. The results depend on the covariance matrix of the errors, the exact nature of the GLS structure used and the method to construct the FGLS estimate, the regressors and the sample size. The bounds are, however, not informative and quite complex. Accordingly they resort to simulation experiments using approximate autoregressive processes of order 1, 7 and 14 when $T = 200$ to construct the FGLS estimate. In the paper, they report results for few selected cases, which do not allow addressing several of the issues discussed above, e.g., the effect of the sign of the quasi-difference parameter, the strength of the correlation in the errors. They wrongly conclude that GLS (constructed using an AR misspecification) is always better than OLS. As shown above this is not the case.*

## S-3 Additional simulations related to Section 6

Tables S.4-S.7 present simulations results related to those presented in Section 6. The setup is exactly the same, except that we set $\rho_x = 0$, instead of $\rho_x = 0.8$. The goal is simply to show robustness of the results. The are indeed qualitatively similar.

## S-4 Simulations with predictive regressions

As discussed in Section 2.1.1 and Remark 4, in the case of predictive regressions assuming rational expectations, both OLS and GLS are consistent. We present the results of a small simulation experiment to show that, with exogenous or non-exogenous regressors, FGLS is by far superior to OLS in terms of MSE and length of the coverage rates, when the MA process is invertible. The setup adopted corresponds to regression

$$y_{t+k} = x_t'\beta + u_{t+k}$$

with $k = 2$ so that the errors are $MA(1)$. The data-generating process is similar to that used above except that the regressors are lagged two periods so that $y_t = \alpha + \beta x_{t-2} + u_t$, $u_t = e_t + \theta e_{t-1}$ and $x_t = \rho_x x_{t-1} + v_t + \gamma e_{t-1}$ with $v_t$ and $e_t$ independent $i.i.d.N(0,1)$ variables. We set $(\alpha, \beta) = (0, 1)$, $\rho_x = 0$ and again $\gamma = 0$ (exogenous regressors), $\gamma = 0.25$ (weak correlation) and $\gamma = 0.50$ (strong correlation). We also consider $\theta = -0.7$, $-0.4$ and 0.5.

The results are presented in Table S.8. With $\gamma = 0$, the results are similar to those in Table 4. FGLS and GMA have much lower MSE than OLS and are nearly as efficient as the infeasible GLS, especially when $T = 500$. The coverage rates for all methods are near the nominal 95% level, except when the MA parameter is strongly negative. Again, the length of the confidence intervals are shorter with FGLS and GMA compared to OLS. With non-exogenous regressors, the results are broadly similar. The only exception is that the coverage rates for GMA are substantially lower than the nominal level. Those for FGLS are adequate except when $\theta = -0.7$. This is in line with our theoretical results.

## S-5 Correcting for heteroskedasticity

In this section, we now consider a FGLS procedure for heteroskedasticity in the errors $e_t$. We describe the method suggested by González-Coya and Perron (2024) based on an Adaptive Lasso procedure to fit the skedastic function. Lasso is a non-parametric estimation method first proposed by Tibshirani (1996). It selects regressors amongst a potentially large set $w_{tj}$ $(j = 1, ..., d)$, where $d$ can be very large, by imposing a $\ell_1$ penalty on their size. Lasso forces the coefficients to be equally penalized. We can, however, assign different weights to different coefficients. If the weights are data-dependent and properly chosen, this can enhance the properties of Lasso, in particular when the irrelevant covariates are highly correlated with the relevant ones. To that effect, Zou (2006) considered the adaptive Lasso given by

$$\hat{\phi}^{\mathrm{A}} = \arg\min_{\phi}\{(1/2)\textstyle\sum_{t=1}^{T}(\log(v_t^2) - \phi_0 - \sum_{j=1}^{d} w_{tj}\phi_j)^2 + \lambda \sum_{j=1}^{d} \hat{\vartheta}_j \left|\phi_j\right|\}, \qquad (A.2)$$

where $\hat{\vartheta}_j = |\hat{\phi}_j|^{-\psi}$, $\psi > 0$ and $\hat{\phi}_j$ is a root-$T$-consistent estimator of $\phi_j$. Here, $v_t$ is some process exhibiting heteroskedasticity, though no serial correlation, to be specified below. The implementation of Adaptive Lasso to obtain a fit to the skedastic function is as follows. 1) Compute the first-step estimate of $\phi$ as the solution to the Ridge regression problem:

$$\hat{\phi}^{\text{ridge}} = \arg\min_{\phi}\{(1/2)\sum_{t=1}^{T}(\log(v_t^2) - \phi_0 - \sum_{j=1}^{d} w_{tj}\phi_j)^2 + \lambda^r \sum_{j=1}^{d} \phi_j^2\},$$

where $\lambda^r$ is selected via cross-validation. 2) Compute the weights as $\hat{\vartheta}_j = |\hat{\phi}_j^{ridge}|^{-\psi}$. The Adaptive Lasso estimates are then

$$\hat{\phi}^{\text{A}} = \arg\min_{\phi}\{(1/2)\sum_{t=1}^{T}(\log(v_t^2) - \phi_0 - \sum_{j=1}^{d} w_{tj}\phi_j)^2 + \lambda^A \sum_{j=1}^{d} |\hat{\phi}_j^{ridge}|^{-\psi}\,|\phi_j|\},$$

where the two tuning parameters, $\lambda^A$ and $\psi$ are selected via the following $K$-cross-validation method: a) Fix $L$ possible values for $\psi$; we use $L = 6$ and $\psi^c = (0, 0.25, 0.5, 0.75, 1, 2)$. b) Fix a partition for the $K$-fold cross-validation, i.e., split the data into $K$ roughly equal-sized parts. We use $K = 10$. Let $\kappa : \{1, \ldots, T\} \mapsto \{1, \ldots, K\}$ be an indexing function that indicates the partition to which observation $t$ is allocated to by the randomization. c) For every $\psi_i^c$, compute the optimal cross-validated $\lambda_i^A$ and the mean cross-validated error. For the $k$th part, we fit the model to the other $K-1$ parts of the data, and calculate the prediction error of the fitted model when predicting the $k$th part of the data. We do this for $k = 1, \ldots, K$ and combine the $K$ estimates of the prediction error. Denote by $\hat{f}_i^{-k}(w)$ the fitted function, computed with the $k$th part of the data removed and using $\psi_i^c$. Then the cross-validation estimate of the prediction error is

$$\text{CV}(\hat{f}_i) = T^{-1}\sum_{t=1}^{T} L\left(\log(v_t^2), \hat{f}_i^{-\kappa(t)}(w)\right),$$

where $L(\cdot)$ is a loss function; we use the MSE. Let $\lambda_i^A$ be the value that minimizes $\text{CV}(\hat{f}_i)$. d) The cross-validated pair $(\lambda^{A*}, \psi^{c*})$ used is the one that minimizes $\text{CV}(\lambda_i^A, \psi_i^c)$ for $i = 1, \ldots, L$. Note that we do not have in mind any oracle model. The aim is to be agnostic about such knowledge and to try to devise a method as robust as possible that allows a reduction in the MSE. Since the skedastic function is, in general, not consistently estimated, there is a need to further correct the variance estimate of the FGLS estimator using a Heteroskedasticity Robust version. We denote the resulting fitted value of the skedastic function by $\widetilde{v}_t^2$.

Here, $v_t \equiv \hat{e}_{tk}$, the residuals from applying the GLS regression

$$(y_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D y_{t-j}) = (x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})'\beta + e_{kt},\ (t = k_T^* + 1, ..., T), \tag{A.3}$$

Let $\hat{\beta}_{F-C}$ denote the GLS estimate that corrects only for serial correlation and $\hat{\beta}_{F-CH}$, the one that corrects for both serial correlation and heteroskedasticity. To be more precise, we apply the following steps: a) Estimate by OLS the quasi-differenced regression (A.3) to obtain the residuals $\hat{e}_{tk}$; b) Estimate the model $\log(\max\{\hat{e}_{tk}^2, \delta^2\}) = \phi_0 - \sum_{j=1}^{d} z_{tj}\phi_j$, via Adaptive Lasso, where $\delta = 0.1$ is some small positive number to avoid dealing with residuals that are nearly zero. Note that $z_t$ may include some or all elements of $x_t$ or transformations of them. Denote the predicted values from this model by $\widetilde{v}_t \equiv \widetilde{e}_{tk}^2$; c) $\hat{\beta}_{F-CH}$ is the weighted least squares (WLS) estimator of the quasi-differenced regression (A.3), with weights given by $\widetilde{e}_{tk}^{-2}$.

In order to construct confidence intervals for the parameter $\beta$ of interest, introducing some finite sample refinements can be beneficial. Here, we describe the particular form adopted, following Miller and Startz (2019) and Rothenberg (1988). We focus on the estimate of the asymptotic variance of the FGLS estimator:

$$Var(\hat{\beta}_{F-CH}) = (T^{-1}X'\widetilde{W}^{-1}X)^{-1}\hat{\Omega}(T^{-1}X'\widetilde{W}^{-1}X)^{-1}, \tag{A.4}$$

where $\widetilde{W}$ is a diagonal matrix with entries $\widetilde{w}_{tt} = \widetilde{v}_t(w)^2 \equiv \widetilde{e}_{tk}^2$, the predicted values obtained from the procedure to fit the skedastic function $v_t(w)$, $X$ is the matrix of regressors, $\hat{\Omega} = T^{-1}X'\hat{\Sigma}^{F-CH}X$ with $\hat{\Sigma}^{F-CH}$ a diagonal matrix with $t^{th}$ entry given by:

$$\hat{\Sigma}_{tt}^{F-CH} = \frac{\hat{e}_{tk-F-CH}^2}{(\widetilde{e}_{tk}^2)^2}\left(\frac{1}{(1-h_{t,F-CH})^2} + 4\frac{h_{t,F-C}}{k}\hat{df}\right), \tag{A.5}$$

where $\hat{e}_{F-CH} = [\hat{e}_{1,F-CH}, ..., \hat{e}_{T,F-CH}]'$ are the estimated residuals from the FGLS regression correcting for serial correlation and heteroskedasticity, i.e., $\hat{e}_{tF-CH} = y_t^* - \hat{\beta}_{F-CH}x_t^*$, with

$$y_t^* = (y_t - \textstyle\sum_{j=1}^{k_T^*} \hat{\rho}_j^D y_{t-j})/(\widetilde{e}_{tk}^2)^{1/2}, \tag{A.6}$$

$$x_t^* = (x_t - \textstyle\sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})/(\widetilde{e}_{tk}^2)^{1/2}. \tag{A.7}$$

$\hat{df}$ is an estimate of the degrees of freedom used in the estimation of the weights. For Lasso, the number of nonzero coefficients is an unbiased estimate for the degrees of freedom (Zou et al. (2007)). The confidence intervals for the $k$th coefficient is then obtained using $\hat{\beta}_{F-CH,k}$ $\pm$ $z_{1-\alpha/2}SE(\hat{\beta}_{F-CH_k})$,where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the normal distribution and $SE(\hat{\beta}_{FGLS,k}) := (Var(\hat{\beta}_{F-CH}))_{kk}^{1/2}$, with $Var(\hat{\beta}_{F-CH})$ defined in (A.4).

### S-5.1 Simulation results with heteroskedasticity

We consider the linear model (1) with serially correlated and heteroskedastic errors. The specifications are the same as in the text except that $e_t \sim N(0, v_t(z))$ or, equivalently, $e_t = \sqrt{v_t(z)}\varepsilon_t$, where $\varepsilon_t \sim i.i.d.\ N(0,1)$. We apply a FGLS accounting for heteroskedasticity in the FGLS regression used to correct for serial correlation,

$$y_t - \textstyle\sum_{j=1}^{k_T^*} \hat{\rho}_j^D y_{t-j} = (x_t - \sum_{j=1}^{k_T^*} \hat{\rho}_j^D x_{t-j})'\beta + e_{tk},\ (t = k_T^* + 1, ..., T),$$

This is then equivalent to applying OLS to the regression $y_t^* = x_t^*\beta + e_{tk-F-CH}$, where $y_t^*$ and $x_t^*$ are defined by (A.6) and (A.7) and the estimate of $\widetilde{e}_{tk}^2$ is constructed as outlined in the previous section. We only consider a subset of the cases used earlier with $T = 200$. These are: 1) $AR(1)$: $u_t = 0.5u_{t-1} + v_t(z)^{1/2}\varepsilon_t$; 2) $AR(2)$: $u_t = 1.34u_{t-1} - 0.42u_{t-2} + v_t(z)^{1/2}\varepsilon_t$; 3) $MA(1)$: $u_t = v_t(z)^{1/2}\varepsilon_t + 0.5v_{t-1}(z)^{1/2}\varepsilon_{t-1}$; 4) $ARMA(1,1)$: $u_t = 0.8u_{t-1} + v_t(z)^{1/2}\varepsilon_t - 0.4v_{t-1}(z)^{1/2}\varepsilon_{t-1}$, where $\varepsilon_t \sim i.i.d.\ N(0,1)$. We consider three specifications for the skedastic function $\nu_t(\cdot)$ as in Romano and Wolf (2017). These are, from weak to strong heteroskedasticity: a) Power function: $\nu_t(x)_1 = x_t^2$; b) Squared log function: $\nu_t(x)_2 = [\log(x_t)]^2$; c) Exponential of a second-degree polynomial: $\nu_t(x)_3 = \exp(0.2x_t + 0.2x_t^2)$. The input matrix is $W = (1, w, w^2, cos(w), cos(2w), cos(3w))$. We consider two cases: a) $w_t = x_t$, which assumes that we select the correct variable influencing the skedastic function; b) $w_t = \phi x_t + (1-\phi)q_t$ with $q_t \sim U(1,4)$ and $\phi \sim$Bernouli$(\rho)$ with $\rho = 0.75$. In this case, the covariate used to model the skedastic function is not the same as the true one but is correlated with it, the correlation being $\rho$. Note that in practice, one can include a vast set of potential covariates. Hence, with the parsimonious set considered, the improvements obtained in terms of MSE and length of the confidence intervals should be viewed as conservative.

The results are reported in Table S.9; the first panel for $w_t = x_t$ and the second for $w_t = \phi x_t + (1-\phi)q_t$. We present the MSE, bias and variance of the FGLS estimate as well as the coverage rates and lengths of the confidence intervals obtained using the method discussed in the previous section. We also present results for the OLS estimate, the FGLS estimate that accounts only for serial correlation (F-C) and the FGLS estimate that accounts for both serial correlation and heteroskedasticity (F-CH). This is done to gauge the extent of the improvement provided by the correction for heteroskedasticity. Note that when using F-C, we construct the confidence intervals that correct for serial correlation the same way as we do for F-CH, i.e., applying the same correction for potential remaining heteroskedasticity.

When the covariate used is the correct one, we see important reduction in the MSE of the F-CH estimate relative to F-C, more so as the heteroskedasticity is stronger. Both the variance and the bias contribute to the reductions in the MSE. Since correcting for serial correlation via a FGLS procedure provides substantially more precise estimates relative to OLS, needless to say that the same applies when further correcting for heteroskedasticity. The coverage rates of the confidence intervals have an exact size close to the nominal level. The OLS estimates also have good coverage rates in most cases but can be sensitive to the strength of the serial correlation; e.g., the $AR(2)$ case. However, the lengths are substantially smaller using F-CH compared to OLS and to a lesser extent compared to F-C.

The results in the bottom panel pertains to the case with an incorrect covariate, though correlated with the correct one. The results are similar with the exception that the incremental reductions in MSE, bias and variance provided by the correction for heteroskedasticity are smaller, as expected. Nevertheless, they are still important enough in magnitude. Hence, using incorrect covariates to estimate the skedastic function can still lead to more precise estimates, as long as there is some correlation between the two sets of covariates. The coverage rate of the confidence intervals have an exact size close to the nominal level and the lengths are much smaller than those with OLS and, to some extent, than with F-C.

We also performed simulation experiments with homoskedastic errors. The results were then essentially equivalent to those obtained with F-C. This means that correcting for heteroskedasticity when it is not present has no detrimental effect on the precision of the estimate, a result emphasized by González-Coya and Perron (2024). Overall, the results show that a further correction for heteroskedasticity can lead to more precise estimates and smaller lengths of the confidence intervals compared to only correcting for serial correlation.

## References


Blanchard, O. J. (1981). "What is left of the multiplier accelerator?." *The American Economic Review, Papers and Proceedings*, 71, 132–143.

Godfrey, L. G. (1978). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables." *Econometrica*, 46, 1293–1302.

González-Coya, E. and Perron, P. (2024). "Estimation in the presence of heteroskedasticty of unknown form: a Lasso-based approach." *Journal of Econometric Methods*, 13, 29–48.

Koreisha, S. G. and Fang, Y. (2001). "Generalized Least Squares with misspecified serial correlation structures." *Journal of the Royal Statistical Society: Series B*, 63, 515–531.

Miller, S. and Startz, R. (2019). "Feasible Generalized Least Squares using machine learning." *Economics Letters*, 175, 28–31.

Romano, J. P. and Wolf, M. (2017). "Resurrecting Weighted Least Squares." *Journal of Econometrics*, 197, 1–19.

Rothenberg, T. J. (1988). "Approximate power functions for some robust tests of regression coefficients." *Econometrica*, 56, 997–1019.

Tibshirani, R. (1996). "Regression shrinkage and selection via the Lasso." *Journal of the Royal Statistical Society: Series B*, 58, 267–288.

Zou, H. (2006). "The Adaptive Lasso and its oracle properties." *Journal of the American Statistical Association*, 101, 1418–1429.

Zou, H., Hastie, T., and Tibshirani, R. (2007). "On the degrees of freedom of the Lasso." *The Annals of Statistics*, 35, 2173–2192.

Table S.1: AR(1)-GLS with parameter $\rho$. Empirical Mean Squared Error of GLS relative to OLS, $T = 200$, $\rho_x = 0$.

| | | | $\rho$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $Cor_u(1)$ | -0.9 | -0.8 | -0.7 | -0.6 | -0.5 | -0.4 | -0.3 | -0.2 | -0.1 | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Hybrid |
| | -0.5 | -0.5 | 0.64 | 0.63 | 0.63 | 0.61 | 0.61 | 0.59 | 0.65 | 0.71 | 0.83 | 1.00 | 1.22 | 1.46 | 1.71 | 1.97 | 2.15 | 2.31 | 2.48 | 2.52 | 2.57 | 0.60 |
| | 0 | 0 | 1.54 | 1.49 | 1.41 | 1.43 | 1.32 | 1.26 | 1.15 | 1.06 | 1.02 | 1.00 | 1.01 | 1.07 | 1.14 | 1.22 | 1.31 | 1.38 | 1.45 | 1.48 | 1.51 | 1.01 |
| AR(1) | 0.2 | 0.2 | 1.94 | 1.92 | 1.84 | 1.76 | 1.64 | 1.53 | 1.35 | 1.24 | 1.10 | 1.00 | 0.95 | 0.92 | 0.95 | 0.96 | 1.01 | 1.04 | 1.09 | 1.10 | 1.11 | 1.01 |
| | 0.5 | 0.5 | 2.61 | 2.55 | 2.51 | 2.35 | 2.19 | 1.98 | 1.73 | 1.46 | 1.22 | 1.00 | 0.83 | 0.71 | 0.65 | 0.62 | 0.60 | 0.62 | 0.59 | 0.63 | 0.61 | 0.89 |
| | 0.8 | 0.8 | 3.42 | 3.35 | 3.22 | 3.06 | 2.77 | 2.50 | 2.10 | 1.72 | 1.34 | 1.00 | 0.72 | 0.52 | 0.38 | 0.29 | 0.25 | 0.25 | 0.22 | 0.23 | 0.24 | 0.93 |
| | 0.5,-0.3 | 0.38 | 2.29 | 2.25 | 2.19 | 2.09 | 1.95 | 1.78 | 1.59 | 1.38 | 1.18 | 1.00 | 0.86 | 0.76 | 0.67 | 0.66 | 0.64 | 0.64 | 0.64 | 0.64 | 0.64 | 0.74 |
| | -0.5,0.3 | -0.71 | 0.46 | 0.45 | 0.44 | 0.43 | 0.43 | 0.44 | 0.49 | 0.59 | 0.76 | 1.00 | 1.31 | 1.66 | 2.02 | 2.36 | 2.65 | 2.89 | 3.06 | 3.17 | 3.23 | 0.42 |
| AR(2) | 1.34,-0.42 | 0.94 | 3.80 | 3.73 | 3.59 | 3.38 | 3.09 | 2.73 | 2.31 | 1.85 | 1.41 | 1.00 | 0.67 | 0.42 | 0.25 | 0.14 | 0.09 | 0.06 | 0.05 | 0.04 | 0.04 | 0.17 |
| | 0,0.3 | 0 | 1.03 | 1.77 | 1.72 | 1.64 | 1.53 | 1.41 | 1.27 | 1.15 | 1.05 | 1.00 | 1.00 | 1.05 | 1.14 | 1.24 | 1.33 | 1.42 | 1.48 | 1.52 | 1.55 | 1.05 |
| | 0.5,0.3 | 0.68 | 3.26 | 3.20 | 3.08 | 2.91 | 2.67 | 2.38 | 2.03 | 1.67 | 1.31 | 1.00 | 0.75 | 0.58 | 0.47 | 0.41 | 0.39 | 0.38 | 0.39 | 0.39 | 0.40 | 0.71 |
| | -0.7 | -0.47 | 0.56 | 0.57 | 0.58 | 0.56 | 0.57 | 0.61 | 0.64 | 0.72 | 0.83 | 1.00 | 1.21 | 1.42 | 1.64 | 1.88 | 2.09 | 2.22 | 2.32 | 2.39 | 2.45 | 0.60 |
| MA(1) | -0.4 | -0.34 | 0.84 | 0.80 | 0.79 | 0.79 | 0.76 | 0.75 | 0.78 | 0.81 | 0.89 | 1.00 | 1.15 | 1.33 | 1.51 | 1.69 | 1.91 | 1.96 | 2.09 | 2.09 | 2.21 | 0.76 |
| | 0.5 | 0.40 | 2.28 | 2.29 | 2.21 | 2.08 | 1.97 | 1.80 | 1.58 | 1.38 | 1.18 | 1.00 | 0.86 | 0.77 | 0.72 | 0.69 | 0.68 | 0.67 | 0.69 | 0.69 | 0.72 | 0.69 |
| | -0.5,-0.4 | -0.57 | 0.30 | 0.29 | 0.30 | 0.30 | 0.34 | 0.37 | 0.45 | 0.57 | 0.75 | 1.00 | 1.29 | 1.63 | 1.94 | 2.27 | 2.50 | 2.69 | 2.83 | 2.97 | 3.02 | 0.40 |
| | 0.2,-0.4 | -0.14 | 1.11 | 1.08 | 1.08 | 1.05 | 1.00 | 0.98 | 0.94 | 0.94 | 0.94 | 1.00 | 1.08 | 1.20 | 1.33 | 1.48 | 1.58 | 1.71 | 1.75 | 1.80 | 1.82 | 0.95 |
| | 0.2,0.5 | 0.57 | 2.60 | 2.59 | 2.46 | 2.38 | 2.21 | 1.98 | 1.74 | 1.48 | 1.23 | 1.00 | 0.81 | 0.67 | 0.59 | 0.53 | 0.51 | 0.49 | 0.49 | 0.49 | 0.51 | 0.54 |
| ARMA(1,1) | 0.5,-0.4 | 0.07 | 1.72 | 1.68 | 1.62 | 1.59 | 1.51 | 1.40 | 1.27 | 1.14 | 1.06 | 1.00 | 0.98 | 0.99 | 1.06 | 1.13 | 1.15 | 1.20 | 1.28 | 1.32 | 1.35 | 1.01 |
| | 0.5,0.5 | 0.75 | 3.08 | 3.05 | 2.93 | 2.77 | 2.56 | 2.31 | 1.97 | 1.65 | 1.30 | 1.00 | 0.75 | 0.55 | 0.42 | 0.34 | 0.28 | 0.28 | 0.27 | 0.26 | 0.26 | 0.34 |
| | 0.8,-0.4 | 0.42 | 2.72 | 2.72 | 2.51 | 2.46 | 2.27 | 2.05 | 1.77 | 1.48 | 1.22 | 1.00 | 0.83 | 0.70 | 0.65 | 0.62 | 0.64 | 0.63 | 0.64 | 0.67 | 0.67 | 0.52 |
| | 0.8,0.5 | 0.99 | 3.60 | 3.53 | 3.45 | 3.23 | 2.94 | 2.61 | 2.23 | 1.80 | 1.38 | 1.00 | 0.69 | 0.45 | 0.29 | 0.19 | 0.13 | 0.11 | 0.09 | 0.08 | 0.08 | 0.18 |

Table S.2: Empirical Mean Squared Error of GLS relative to OLS, $T = 200$, $\rho_x = 0.5$.

| | | $Cor_u(1)$ | $\rho$ | | | | | | | | | | | | | | | | | | Hybrid |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | -0.9 | -0.8 | -0.7 | -0.6 | -0.5 | -0.4 | -0.3 | -0.2 | -0.1 | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | |
| | -0.5 | -0.5 | 0.73 | 0.73 | 0.73 | 0.73 | 0.72 | 0.73 | 0.74 | 0.77 | 0.85 | 1.00 | 1.25 | 1.65 | 2.21 | 2.92 | 3.73 | 4.54 | 5.25 | 5.77 | 6.08 | 0.72 |
| | 0 | 0 | 1.16 | 1.15 | 1.15 | 1.13 | 1.12 | 1.09 | 1.06 | 1.03 | 1.01 | 1.00 | 1.02 | 1.09 | 1.21 | 1.40 | 1.63 | 1.88 | 2.11 | 2.28 | 2.38 | 1.01 |
| AR(1) | 0.2 | 0.2 | 1.30 | 1.29 | 1.28 | 1.27 | 1.24 | 1.21 | 1.16 | 1.11 | 1.05 | 1.00 | 0.96 | 0.94 | 0.97 | 1.04 | 1.15 | 1.28 | 1.41 | 1.50 | 1.56 | 1.09 |
| | 0.5 | 0.5 | 1.51 | 1.50 | 1.48 | 1.46 | 1.42 | 1.37 | 1.30 | 1.22 | 1.11 | 1.00 | 0.88 | 0.77 | 0.69 | 0.64 | 0.62 | 0.64 | 0.67 | 0.70 | 0.72 | 0.66 |
| | 0.8 | 0.8 | 1.70 | 1.69 | 1.67 | 1.63 | 1.58 | 1.52 | 1.42 | 1.31 | 1.16 | 1.00 | 0.82 | 0.64 | 0.48 | 0.35 | 0.26 | 0.22 | 0.20 | 0.19 | 0.20 | 0.35 |
| | 0.5,-0.3 | 0.38 | 1.33 | 1.32 | 1.31 | 1.29 | 1.27 | 1.23 | 1.19 | 1.13 | 1.07 | 1.00 | 0.93 | 0.88 | 0.85 | 0.86 | 0.89 | 0.94 | 1.00 | 1.04 | 1.07 | 0.93 |
| | -0.5,0.3 | -0.71 | 0.62 | 0.62 | 0.61 | 0.61 | 0.60 | 0.60 | 0.62 | 0.67 | 0.78 | 1.00 | 1.37 | 1.95 | 2.76 | 3.80 | 4.97 | 6.16 | 7.19 | 7.95 | 8.39 | 0.62 |
| AR(2) | 1.34,-0.42 | 0.94 | 1.75 | 1.74 | 1.72 | 1.68 | 1.63 | 1.56 | 1.46 | 1.33 | 1.18 | 1.00 | 0.80 | 0.60 | 0.42 | 0.27 | 0.16 | 0.09 | 0.06 | 0.04 | 0.04 | 0.26 |
| | 0,0.3 | 0 | 1.33 | 1.32 | 1.31 | 1.29 | 1.26 | 1.22 | 1.17 | 1.11 | 1.05 | 1.00 | 0.98 | 1.00 | 1.09 | 1.24 | 1.46 | 1.69 | 1.91 | 2.08 | 2.18 | 1.01 |
| | 0.5,0.3 | 0.68 | 1.69 | 1.68 | 1.66 | 1.63 | 1.58 | 1.51 | 1.42 | 1.30 | 1.16 | 1.00 | 0.83 | 0.66 | 0.51 | 0.39 | 0.32 | 0.29 | 0.29 | 0.29 | 0.30 | 0.40 |
| | -0.7 | -0.47 | 0.51 | 0.51 | 0.51 | 0.52 | 0.53 | 0.55 | 0.59 | 0.67 | 0.79 | 1.00 | 1.33 | 1.80 | 2.45 | 3.25 | 4.14 | 5.02 | 5.79 | 6.35 | 6.68 | 0.56 |
| MA(1) | -0.4 | -0.34 | 0.84 | 0.77 | 0.77 | 0.77 | 0.76 | 0.77 | 0.78 | 0.81 | 0.88 | 1.00 | 1.21 | 1.53 | 1.98 | 2.55 | 3.20 | 3.84 | 4.41 | 4.83 | 5.07 | 0.77 |
| | 0.5 | 0.40 | 1.40 | 1.39 | 1.38 | 1.36 | 1.33 | 1.29 | 1.23 | 1.17 | 1.09 | 1.00 | 0.91 | 0.84 | 0.79 | 0.78 | 0.80 | 0.85 | 0.90 | 0.95 | 0.97 | 0.86 |
| | -0.5,-0.4 | -0.57 | 0.33 | 0.33 | 0.34 | 0.34 | 0.35 | 0.38 | 0.43 | 0.53 | 0.71 | 1.00 | 1.46 | 2.15 | 3.09 | 4.25 | 5.55 | 6.84 | 7.96 | 8.78 | 9.25 | 0.41 |
| | 0.2,-0.4 | -0.14 | 0.97 | 0.97 | 0.96 | 0.96 | 0.95 | 0.93 | 0.93 | 0.93 | 0.95 | 1.00 | 1.11 | 1.30 | 1.58 | 1.95 | 2.38 | 2.83 | 3.21 | 3.50 | 3.67 | 0.95 |
| | 0.2,0.5 | 0.57 | 1.48 | 1.47 | 1.45 | 1.43 | 1.40 | 1.35 | 1.29 | 1.21 | 1.11 | 1.00 | 0.89 | 0.78 | 0.69 | 0.63 | 0.60 | 0.60 | 0.62 | 0.64 | 0.65 | 0.66 |
| ARMA(1,1) | 0.5,-0.4 | 0.07 | 1.28 | 1.28 | 1.27 | 1.25 | 1.22 | 1.19 | 1.15 | 1.10 | 1.04 | 1.00 | 0.98 | 0.99 | 1.05 | 1.17 | 1.34 | 1.53 | 1.70 | 1.84 | 1.92 | 1.06 |
| | 0.5,0.5 | 0.75 | 1.59 | 1.58 | 1.56 | 1.54 | 1.49 | 1.44 | 1.36 | 1.26 | 1.14 | 1.00 | 0.85 | 0.70 | 0.56 | 0.45 | 0.37 | 0.33 | 0.31 | 0.31 | 0.31 | 0.44 |
| | 0.8,-0.4 | 0.42 | 1.70 | 1.69 | 1.67 | 1.63 | 1.58 | 1.52 | 1.42 | 1.31 | 1.16 | 1.00 | 0.82 | 0.65 | 0.50 | 0.38 | 0.31 | 0.28 | 0.28 | 0.28 | 0.29 | 0.41 |
| | 0.8,0.5 | 0.99 | 1.76 | 1.75 | 1.73 | 1.69 | 1.64 | 1.56 | 1.46 | 1.34 | 1.18 | 1.00 | 0.80 | 0.60 | 0.42 | 0.27 | 0.16 | 0.09 | 0.06 | 0.04 | 0.04 | 0.25 |

Table S.3: Empirical Mean Squared Error of GLS relative to OLS, $T = 200$, $\rho_x = 0.8$.

| | | $Cor_u(1)$ | $\rho$ -0.9 | -0.8 | -0.7 | -0.6 | -0.5 | -0.4 | -0.3 | -0.2 | -0.1 | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Hybrid |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AR(1) | -0.5 | -0.5 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.85 | 0.87 | 0.91 | 1.00 | 1.18 | 1.51 | 2.12 | 3.17 | 4.83 | 7.18 | 10.06 | 12.87 | 14.87 | 0.85 |
| | 0 | 0 | 1.06 | 1.05 | 1.05 | 1.05 | 1.04 | 1.03 | 1.02 | 1.01 | 1.00 | 1.00 | 1.01 | 1.06 | 1.18 | 1.41 | 1.83 | 2.45 | 3.24 | 4.04 | 4.61 | 1.00 |
| | 0.2 | 0.2 | 1.11 | 1.11 | 1.11 | 1.10 | 1.09 | 1.08 | 1.07 | 1.05 | 1.03 | 1.00 | 0.98 | 0.97 | 0.99 | 1.07 | 1.25 | 1.56 | 1.98 | 2.42 | 2.74 | 1.28 |
| | 0.5 | 0.5 | 1.18 | 1.18 | 1.18 | 1.17 | 1.16 | 1.14 | 1.12 | 1.09 | 1.05 | 1.00 | 0.94 | 0.87 | 0.79 | 0.73 | 0.70 | 0.73 | 0.81 | 0.94 | 1.04 | 0.82 |
| | 0.8 | 0.8 | 1.24 | 1.24 | 1.23 | 1.22 | 1.21 | 1.19 | 1.16 | 1.12 | 1.07 | 1.00 | 0.91 | 0.80 | 0.67 | 0.53 | 0.40 | 0.29 | 0.23 | 0.22 | 0.22 | 0.46 |
| AR(2) | 0.5,-0.3 | 0.38 | 1.09 | 1.09 | 1.09 | 1.08 | 1.08 | 1.07 | 1.06 | 1.04 | 1.02 | 1.00 | 0.98 | 0.97 | 0.98 | 1.04 | 1.19 | 1.43 | 1.75 | 2.10 | 2.35 | 1.35 |
| | -0.5,0.3 | -0.71 | 0.81 | 0.81 | 0.80 | 0.80 | 0.80 | 0.80 | 0.81 | 0.83 | 0.88 | 1.00 | 1.24 | 1.71 | 2.58 | 4.09 | 6.51 | 9.97 | 14.22 | 18.40 | 21.38 | 0.81 |
| | 1.34,-0.42 | 0.94 | 1.75 | 1.74 | 1.72 | 1.68 | 1.63 | 1.56 | 1.46 | 1.33 | 1.18 | 1.00 | 0.80 | 0.60 | 0.42 | 0.27 | 0.16 | 0.09 | 0.06 | 0.04 | 0.04 | 0.38 |
| | 0,0.3 | 0 | 1.14 | 1.14 | 1.13 | 1.13 | 1.12 | 1.10 | 1.08 | 1.06 | 1.03 | 1.00 | 0.97 | 0.96 | 1.00 | 1.12 | 1.37 | 1.79 | 2.36 | 2.96 | 3.39 | 1.03 |
| | 0.5,0.3 | 0.68 | 1.24 | 1.24 | 1.23 | 1.22 | 1.21 | 1.19 | 1.16 | 1.12 | 1.07 | 1.00 | 0.91 | 0.80 | 0.68 | 0.54 | 0.42 | 0.33 | 0.29 | 0.30 | 0.32 | 0.52 |
| MA(1) | -0.7 | -0.47 | 0.55 | 0.55 | 0.55 | 0.56 | 0.57 | 0.59 | 0.62 | 0.69 | 0.80 | 1.00 | 1.36 | 2.01 | 3.13 | 5.00 | 7.91 | 12.00 | 16.95 | 21.80 | 25.24 | 0.60 |
| | -0.4 | -0.34 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.84 | 0.85 | 0.87 | 0.91 | 1.00 | 1.17 | 1.48 | 2.04 | 3.00 | 4.52 | 6.68 | 9.33 | 11.93 | 13.79 | 0.85 |
| | 0.5 | 0.40 | 1.14 | 1.14 | 1.13 | 1.13 | 1.12 | 1.10 | 1.09 | 1.06 | 1.04 | 1.00 | 0.96 | 0.92 | 0.90 | 0.91 | 0.98 | 1.14 | 1.38 | 1.64 | 1.84 | 1.13 |
| ARMA(1,1) | -0.5,-0.4 | -0.57 | 0.47 | 0.47 | 0.47 | 0.48 | 0.49 | 0.50 | 0.54 | 0.61 | 0.74 | 1.00 | 1.48 | 2.35 | 3.89 | 6.50 | 10.60 | 16.41 | 23.48 | 30.43 | 35.36 | 0.52 |
| | 0.2,-0.4 | -0.14 | 0.96 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.96 | 1.00 | 1.08 | 1.25 | 1.56 | 2.11 | 3.01 | 4.30 | 5.90 | 7.48 | 8.62 | 0.97 |
| | 0.2,0.5 | 0.57 | 1.16 | 1.16 | 1.16 | 1.15 | 1.14 | 1.13 | 1.11 | 1.08 | 1.05 | 1.00 | 0.95 | 0.88 | 0.82 | 0.78 | 0.77 | 0.82 | 0.93 | 1.07 | 1.18 | 0.87 |
| | 0.5,-0.4 | 0.07 | 1.11 | 1.11 | 1.10 | 1.10 | 1.09 | 1.08 | 1.06 | 1.04 | 1.02 | 1.00 | 0.98 | 0.99 | 1.04 | 1.18 | 1.45 | 1.88 | 2.45 | 3.04 | 3.47 | 1.13 |
| | 0.5,0.5 | 0.75 | 1.21 | 1.20 | 1.20 | 1.19 | 1.18 | 1.16 | 1.14 | 1.10 | 1.06 | 1.00 | 0.93 | 0.83 | 0.73 | 0.62 | 0.53 | 0.46 | 0.45 | 0.47 | 0.49 | 0.60 |
| | 0.8,-0.4 | 0.42 | 1.24 | 1.24 | 1.24 | 1.23 | 1.21 | 1.19 | 1.16 | 1.12 | 1.07 | 1.00 | 0.91 | 0.80 | 0.67 | 0.53 | 0.40 | 0.31 | 0.27 | 0.27 | 0.30 | 0.48 |
| | 0.8,0.5 | 0.99 | 1.25 | 1.25 | 1.24 | 1.23 | 1.22 | 1.20 | 1.17 | 1.13 | 1.07 | 1.00 | 0.91 | 0.79 | 0.64 | 0.49 | 0.33 | 0.19 | 0.10 | 0.06 | 0.05 | 0.39 |

Table S.4: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, AR(1) case with $\rho_x = 0$. (First 3 columns are multiplied by 100).

| | | | MSE | | | | Bias | | | | Variance | | | | Coverage | | | Lenght | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AR(1) | OLS | GLS | CO | FGLS | OLS | GLS | CO | FGLS | OLS | GLS | CO | FGLS | OLS | CO | FGLS | OLS | CO | FGLS |
| $T = 200$ | $\gamma = 0$ | -0.5 | 0.64 | 0.38 | 0.38 | 0.38 | 6.38 | 4.95 | 4.95 | 4.98 | 0.68 | 0.41 | 0.41 | 0.41 | 0.95 | 0.96 | 0.96 | 0.32 | 0.25 | 0.25 |
| | | 0 | 0.49 | 0.49 | 0.49 | 0.49 | 5.59 | 5.62 | 5.63 | 5.62 | 0.50 | 0.51 | 0.50 | 0.50 | 0.95 | 0.95 | 0.95 | 0.28 | 0.28 | 0.28 |
| | | 0.2 | 0.51 | 0.48 | 0.48 | 0.49 | 5.72 | 5.55 | 5.56 | 5.61 | 0.52 | 0.49 | 0.48 | 0.48 | 0.95 | 0.95 | 0.94 | 0.28 | 0.27 | 0.27 |
| | | 0.5 | 0.66 | 0.41 | 0.41 | 0.42 | 6.51 | 5.13 | 5.13 | 5.14 | 0.66 | 0.41 | 0.41 | 0.40 | 0.95 | 0.94 | 0.94 | 0.32 | 0.25 | 0.25 |
| | | 0.8 | 1.37 | 0.32 | 0.32 | 0.32 | 9.33 | 4.51 | 4.51 | 4.53 | 1.33 | 0.31 | 0.31 | 0.31 | 0.95 | 0.95 | 0.95 | 0.45 | 0.22 | 0.22 |
| | $\gamma = 0.25$ | -0.5 | 1.98 | 0.37 | 0.39 | 0.39 | 12.04 | 4.87 | 4.95 | 4.99 | 0.64 | 0.39 | 0.39 | 0.38 | 0.70 | 0.95 | 0.95 | 0.31 | 0.25 | 0.24 |
| | | 0 | 0.47 | 0.48 | 0.50 | 0.49 | 5.49 | 5.50 | 5.67 | 5.57 | 0.47 | 0.48 | 0.47 | 0.47 | 0.95 | 0.95 | 0.95 | 0.27 | 0.27 | 0.27 |
| | | 0.2 | 0.72 | 0.46 | 0.49 | 0.55 | 6.87 | 5.43 | 5.62 | 5.94 | 0.49 | 0.46 | 0.46 | 0.46 | 0.89 | 0.94 | 0.93 | 0.27 | 0.27 | 0.27 |
| | | 0.5 | 2.01 | 0.39 | 0.42 | 0.41 | 12.17 | 5.00 | 5.16 | 5.13 | 0.62 | 0.38 | 0.39 | 0.38 | 0.67 | 0.94 | 0.94 | 0.31 | 0.25 | 0.24 |
| | | 0.8 | 4.67 | 0.30 | 0.31 | 0.31 | 18.81 | 4.38 | 4.43 | 4.44 | 1.26 | 0.29 | 0.30 | 0.29 | 0.62 | 0.95 | 0.94 | 0.44 | 0.22 | 0.21 |
| | $\gamma = 0.5$ | -0.5 | 4.52 | 0.32 | 0.43 | 0.36 | 19.95 | 4.48 | 5.18 | 4.80 | 0.55 | 0.33 | 0.35 | 0.32 | 0.22 | 0.89 | 0.94 | 0.29 | 0.23 | 0.22 |
| | | 0 | 0.40 | 0.40 | 0.46 | 0.43 | 5.05 | 5.05 | 5.45 | 5.20 | 0.40 | 0.41 | 0.40 | 0.40 | 0.95 | 0.92 | 0.94 | 0.25 | 0.25 | 0.25 |
| | | 0.2 | 1.05 | 0.39 | 0.49 | 0.67 | 8.62 | 4.99 | 5.61 | 6.59 | 0.42 | 0.39 | 0.39 | 0.39 | 0.76 | 0.90 | 0.87 | 0.25 | 0.25 | 0.25 |
| | | 0.5 | 4.47 | 0.33 | 0.47 | 0.38 | 19.80 | 4.61 | 5.51 | 4.97 | 0.54 | 0.33 | 0.35 | 0.32 | 0.23 | 0.89 | 0.93 | 0.29 | 0.23 | 0.22 |
| | | 0.8 | 10.84 | 0.25 | 0.29 | 0.27 | 31.19 | 4.05 | 4.28 | 4.19 | 1.09 | 0.25 | 0.27 | 0.25 | 0.13 | 0.94 | 0.94 | 0.41 | 0.20 | 0.20 |
| $T = 500$ | $\gamma = 0$ | -0.5 | 0.27 | 0.17 | 0.17 | 0.17 | 4.15 | 3.27 | 3.27 | 3.28 | 0.27 | 0.16 | 0.16 | 0.16 | 0.95 | 0.94 | 0.94 | 0.20 | 0.16 | 0.16 |
| | | 0 | 0.20 | 0.21 | 0.21 | 0.21 | 3.62 | 3.63 | 3.63 | 3.63 | 0.20 | 0.20 | 0.20 | 0.20 | 0.95 | 0.95 | 0.95 | 0.18 | 0.18 | 0.18 |
| | | 0.2 | 0.22 | 0.20 | 0.20 | 0.20 | 3.72 | 3.55 | 3.55 | 3.56 | 0.21 | 0.19 | 0.19 | 0.19 | 0.95 | 0.95 | 0.95 | 0.18 | 0.17 | 0.17 |
| | | 0.5 | 0.28 | 0.16 | 0.16 | 0.16 | 4.21 | 3.22 | 3.22 | 3.23 | 0.27 | 0.16 | 0.16 | 0.16 | 0.94 | 0.95 | 0.95 | 0.20 | 0.16 | 0.16 |
| | | 0.8 | 0.57 | 0.12 | 0.12 | 0.12 | 6.02 | 2.80 | 2.80 | 2.80 | 0.55 | 0.12 | 0.12 | 0.12 | 0.94 | 0.95 | 0.95 | 0.29 | 0.14 | 0.14 |
| | $\gamma = 0.25$ | -0.5 | 1.98 | 0.37 | 0.39 | 0.39 | 12.04 | 4.87 | 4.95 | 4.99 | 0.64 | 0.39 | 0.39 | 0.38 | 0.70 | 0.95 | 0.95 | 0.31 | 0.25 | 0.24 |
| | | 0 | 0.47 | 0.48 | 0.50 | 0.49 | 5.49 | 5.50 | 5.67 | 5.57 | 0.47 | 0.48 | 0.47 | 0.47 | 0.95 | 0.95 | 0.95 | 0.27 | 0.27 | 0.27 |
| | | 0.2 | 0.72 | 0.46 | 0.49 | 0.55 | 6.87 | 5.43 | 5.62 | 5.94 | 0.49 | 0.46 | 0.46 | 0.46 | 0.89 | 0.94 | 0.93 | 0.27 | 0.27 | 0.27 |
| | | 0.5 | 2.01 | 0.39 | 0.42 | 0.41 | 12.17 | 5.00 | 5.16 | 5.13 | 0.62 | 0.38 | 0.39 | 0.38 | 0.67 | 0.94 | 0.94 | 0.31 | 0.25 | 0.24 |
| | | 0.8 | 4.67 | 0.30 | 0.31 | 0.31 | 18.81 | 4.38 | 4.43 | 4.44 | 1.26 | 0.29 | 0.30 | 0.29 | 0.62 | 0.95 | 0.94 | 0.44 | 0.22 | 0.21 |
| | $\gamma = 0.5$ | -0.5 | 4.21 | 0.13 | 0.21 | 0.15 | 19.97 | 2.89 | 3.64 | 3.07 | 0.22 | 0.13 | 0.14 | 0.13 | 0.01 | 0.89 | 0.93 | 0.18 | 0.15 | 0.14 |
| | | 0 | 0.17 | 0.17 | 0.19 | 0.17 | 3.25 | 3.26 | 3.51 | 3.30 | 0.16 | 0.16 | 0.16 | 0.16 | 0.94 | 0.92 | 0.94 | 0.16 | 0.16 | 0.16 |
| | | 0.2 | 0.81 | 0.16 | 0.22 | 0.23 | 8.03 | 3.19 | 3.68 | 3.74 | 0.17 | 0.15 | 0.16 | 0.15 | 0.51 | 0.90 | 0.90 | 0.16 | 0.16 | 0.15 |
| | | 0.5 | 4.17 | 0.13 | 0.21 | 0.15 | 19.85 | 2.91 | 3.67 | 3.09 | 0.22 | 0.13 | 0.14 | 0.13 | 0.01 | 0.89 | 0.93 | 0.18 | 0.15 | 0.14 |
| | | 0.8 | 10.42 | 0.10 | 0.11 | 0.11 | 31.55 | 2.53 | 2.68 | 2.58 | 0.45 | 0.10 | 0.11 | 0.10 | 0.00 | 0.94 | 0.94 | 0.26 | 0.13 | 0.12 |

Table S.5: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, AR(2) case with $\rho_x = 0$. (First 3 columns are multiplied by 100).

| | | | MSE | | | Bias | | | Variance | | | Coverage | | Lenght | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | AR(2) | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | FGLS | OLS | FGLS |
| $T = 200$ | $\gamma = 0$ | 0.5,-0.3 | 0.63 | 0.38 | 0.39 | 6.26 | 4.96 | 5.03 | 0.64 | 0.38 | 0.38 | 0.95 | 0.94 | 0.31 | 0.24 |
| | | -0.5,0.3 | 1.11 | 0.38 | 0.38 | 8.36 | 4.93 | 4.97 | 1.11 | 0.38 | 0.38 | 0.95 | 0.95 | 0.41 | 0.24 |
| | | 1.34,-0.42 | 5.24 | 0.17 | 0.17 | 18.02 | 3.32 | 3.33 | 5.10 | 0.17 | 0.17 | 0.94 | 0.95 | 0.88 | 0.16 |
| | | 0,0.3 | 0.54 | 0.45 | 0.46 | 5.89 | 5.39 | 5.44 | 0.55 | 0.47 | 0.46 | 0.95 | 0.95 | 0.29 | 0.27 |
| | | 0.5,0.3 | 1.08 | 0.37 | 0.38 | 8.29 | 4.86 | 4.92 | 1.06 | 0.38 | 0.38 | 0.94 | 0.95 | 0.40 | 0.24 |
| | $\gamma = 0.25$ | 0.5,-0.3 | 1.88 | 0.36 | 0.38 | 11.86 | 4.82 | 4.97 | 0.60 | 0.36 | 0.35 | 0.68 | 0.94 | 0.30 | 0.23 |
| | | -0.5,0.3 | 2.51 | 0.35 | 0.38 | 13.08 | 4.73 | 4.92 | 1.07 | 0.36 | 0.35 | 0.79 | 0.94 | 0.40 | 0.23 |
| | | 1.34,-0.42 | 13.88 | 0.17 | 0.17 | 31.60 | 3.28 | 3.31 | 4.75 | 0.16 | 0.16 | 0.73 | 0.95 | 0.85 | 0.16 |
| | | 0,0.3 | 0.52 | 0.42 | 0.45 | 5.80 | 5.21 | 5.26 | 0.53 | 0.44 | 0.44 | 0.94 | 0.94 | 0.28 | 0.26 |
| | | 0.5,0.3 | 2.25 | 0.35 | 0.38 | 12.42 | 4.73 | 4.95 | 1.00 | 0.36 | 0.35 | 0.82 | 0.94 | 0.39 | 0.23 |
| | $\gamma = 0.5$ | 0.5,-0.3 | 4.46 | 0.31 | 0.38 | 19.89 | 4.37 | 4.90 | 0.49 | 0.31 | 0.30 | 0.18 | 0.92 | 0.27 | 0.22 |
| | | -0.5,0.3 | 5.06 | 0.31 | 0.40 | 20.39 | 4.40 | 5.02 | 0.94 | 0.31 | 0.30 | 0.43 | 0.90 | 0.38 | 0.22 |
| | | 1.34,-0.42 | 30.83 | 0.14 | 0.15 | 51.46 | 2.96 | 3.12 | 4.06 | 0.14 | 0.14 | 0.26 | 0.94 | 0.78 | 0.15 |
| | | 0,0.3 | 0.50 | 0.38 | 0.47 | 5.61 | 4.89 | 5.47 | 0.47 | 0.37 | 0.37 | 0.94 | 0.92 | 0.27 | 0.24 |
| | | 0.5,0.3 | 4.62 | 0.31 | 0.42 | 19.25 | 4.39 | 5.16 | 0.88 | 0.30 | 0.30 | 0.46 | 0.89 | 0.36 | 0.22 |
| $T = 500$ | $\gamma = 0$ | 0.5,-0.3 | 0.27 | 0.15 | 0.15 | 4.14 | 3.07 | 3.08 | 0.26 | 0.15 | 0.15 | 0.95 | 0.95 | 0.20 | 0.15 |
| | | -0.5,0.3 | 0.45 | 0.16 | 0.16 | 5.34 | 3.17 | 3.17 | 0.45 | 0.15 | 0.15 | 0.95 | 0.95 | 0.26 | 0.15 |
| | | 1.34,-0.42 | 2.17 | 0.07 | 0.07 | 11.62 | 2.04 | 2.04 | 2.17 | 0.07 | 0.07 | 0.95 | 0.96 | 0.57 | 0.10 |
| | | 0,0.3 | 0.23 | 0.19 | 0.19 | 3.80 | 3.48 | 3.48 | 0.22 | 0.18 | 0.18 | 0.95 | 0.95 | 0.18 | 0.17 |
| | | 0.5,0.3 | 0.45 | 0.15 | 0.15 | 5.33 | 3.09 | 3.09 | 0.44 | 0.15 | 0.15 | 0.95 | 0.95 | 0.26 | 0.15 |
| | $\gamma = 0.25$ | 0.5,-0.3 | 1.62 | 0.13 | 0.14 | 11.79 | 2.91 | 3.02 | 0.24 | 0.14 | 0.14 | 0.33 | 0.95 | 0.19 | 0.15 |
| | | -0.5,0.3 | 1.77 | 0.14 | 0.15 | 11.81 | 3.00 | 3.08 | 0.43 | 0.14 | 0.14 | 0.58 | 0.94 | 0.26 | 0.15 |
| | | 1.34,-0.42 | 11.79 | 0.06 | 0.06 | 31.31 | 1.98 | 2.00 | 2.03 | 0.06 | 0.06 | 0.40 | 0.93 | 0.56 | 0.10 |
| | | 0,0.3 | 0.21 | 0.17 | 0.18 | 3.62 | 3.26 | 3.37 | 0.21 | 0.17 | 0.17 | 0.95 | 0.95 | 0.18 | 0.16 |
| | | 0.5,0.3 | 1.79 | 0.14 | 0.15 | 11.86 | 2.94 | 3.06 | 0.42 | 0.14 | 0.14 | 0.56 | 0.94 | 0.25 | 0.15 |
| | $\gamma = 0.5$ | 0.5,-0.3 | 4.17 | 0.12 | 0.14 | 19.94 | 2.74 | 2.99 | 0.19 | 0.12 | 0.12 | 0.01 | 0.92 | 0.17 | 0.13 |
| | | -0.5,0.3 | 4.34 | 0.12 | 0.15 | 19.91 | 2.82 | 3.14 | 0.37 | 0.12 | 0.12 | 0.08 | 0.91 | 0.24 | 0.13 |
| | | 1.34,-0.42 | 29.61 | 0.05 | 0.06 | 52.75 | 1.86 | 1.91 | 1.72 | 0.05 | 0.05 | 0.01 | 0.95 | 0.51 | 0.09 |
| | | 0,0.3 | 0.20 | 0.15 | 0.19 | 3.56 | 3.10 | 3.45 | 0.19 | 0.15 | 0.15 | 0.94 | 0.91 | 0.17 | 0.15 |
| | | 0.5,0.3 | 4.26 | 0.12 | 0.15 | 19.68 | 2.77 | 3.10 | 0.37 | 0.12 | 0.12 | 0.09 | 0.92 | 0.24 | 0.13 |

Table S.6: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, MA(1) case with $\rho_x = 0$. (First 3 columns are multiplied by 100).

| | | | MSE | | | | Bias | | | | Variance | | | | Coverage | | | Lenght | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MA(1) | OLS | GLS | GMA | FGLS | OLS | GLS | GMA | FGLS | OLS | GLS | GMA | FGLS | OLS | GMA | FGLS | OLS | GMA | FGLS |
| $T = 200$ | $\gamma = 0$ | -0.7 | 0.76 | 0.28 | 0.29 | 0.31 | 6.93 | 4.21 | 4.26 | 4.45 | 0.75 | 0.27 | 0.27 | 0.30 | 0.94 | 0.94 | 0.94 | 0.34 | 0.20 | 0.22 |
| | | -0.4 | 0.59 | 0.43 | 0.44 | 0.46 | 6.12 | 5.24 | 5.27 | 5.38 | 0.58 | 0.43 | 0.42 | 0.44 | 0.94 | 0.94 | 0.94 | 0.30 | 0.26 | 0.26 |
| | | 0.5 | 0.62 | 0.38 | 0.39 | 0.41 | 6.31 | 4.95 | 4.98 | 5.12 | 0.62 | 0.38 | 0.37 | 0.40 | 0.95 | 0.95 | 0.95 | 0.31 | 0.24 | 0.25 |
| | $\gamma = 0.25$ | -0.7 | 3.29 | 0.26 | 0.29 | 0.37 | 16.34 | 4.04 | 4.30 | 4.79 | 0.69 | 0.38 | 0.25 | 0.29 | 0.51 | 0.94 | 0.93 | 0.33 | 0.20 | 0.21 |
| | | -0.4 | 1.39 | 0.40 | 0.44 | 0.48 | 9.94 | 5.11 | 5.31 | 5.50 | 0.54 | 0.46 | 0.39 | 0.41 | 0.77 | 0.94 | 0.93 | 0.29 | 0.25 | 0.25 |
| | | 0.5 | 1.95 | 0.36 | 0.39 | 0.44 | 12.07 | 4.80 | 4.95 | 5.25 | 0.58 | 0.44 | 0.35 | 0.38 | 0.66 | 0.94 | 0.93 | 0.30 | 0.23 | 0.24 |
| | $\gamma = 0.5$ | -0.7 | 8.53 | 0.22 | 0.32 | 0.51 | 28.23 | 3.72 | 4.41 | 5.61 | 0.56 | 0.32 | 0.22 | 0.25 | 0.03 | 0.91 | 0.85 | 0.29 | 0.18 | 0.20 |
| | | -0.4 | 3.11 | 0.34 | 0.44 | 0.53 | 16.35 | 4.69 | 5.31 | 5.80 | 0.45 | 0.40 | 0.34 | 0.35 | 0.32 | 0.91 | 0.90 | 0.26 | 0.23 | 0.23 |
| | | 0.5 | 4.26 | 0.30 | 0.39 | 0.50 | 19.46 | 4.41 | 5.02 | 5.68 | 0.47 | 0.38 | 0.30 | 0.32 | 0.19 | 0.91 | 0.89 | 0.27 | 0.22 | 0.22 |
| $T = 500$ | $\gamma = 0$ | -0.7 | 0.32 | 0.11 | 0.11 | 0.12 | 4.52 | 2.57 | 2.60 | 2.66 | 0.30 | 0.10 | 0.10 | 0.11 | 0.94 | 0.94 | 0.94 | 0.21 | 0.13 | 0.13 |
| | | -0.4 | 0.25 | 0.18 | 0.18 | 0.19 | 4.01 | 3.36 | 3.39 | 3.41 | 0.23 | 0.17 | 0.17 | 0.17 | 0.94 | 0.94 | 0.94 | 0.19 | 0.16 | 0.16 |
| | | 0.5 | 0.27 | 0.16 | 0.16 | 0.16 | 4.04 | 3.18 | 3.19 | 3.24 | 0.25 | 0.15 | 0.15 | 0.16 | 0.94 | 0.94 | 0.94 | 0.20 | 0.15 | 0.15 |
| | $\gamma = 0.25$ | -0.7 | 2.88 | 0.10 | 0.10 | 0.12 | 16.18 | 2.47 | 2.58 | 2.81 | 0.28 | 0.10 | 0.10 | 0.11 | 0.13 | 0.94 | 0.93 | 0.21 | 0.12 | 0.13 |
| | | -0.4 | 1.05 | 0.16 | 0.17 | 0.18 | 9.27 | 3.17 | 3.28 | 3.39 | 0.22 | 0.16 | 0.16 | 0.16 | 0.50 | 0.94 | 0.94 | 0.18 | 0.16 | 0.16 |
| | | 0.5 | 1.61 | 0.14 | 0.15 | 0.16 | 11.79 | 2.99 | 3.08 | 3.22 | 0.23 | 0.14 | 0.14 | 0.15 | 0.31 | 0.95 | 0.94 | 0.19 | 0.15 | 0.15 |
| | $\gamma = 0.5$ | -0.7 | 8.11 | 0.08 | 0.11 | 0.17 | 28.10 | 2.27 | 2.64 | 3.31 | 0.22 | 0.08 | 0.08 | 0.08 | 0.00 | 0.92 | 0.87 | 0.19 | 0.11 | 0.12 |
| | | -0.4 | 2.78 | 0.13 | 0.17 | 0.20 | 16.13 | 2.89 | 3.26 | 3.55 | 0.18 | 0.14 | 0.14 | 0.14 | 0.03 | 0.92 | 0.91 | 0.17 | 0.14 | 0.15 |
| | | 0.5 | 4.11 | 0.12 | 0.16 | 0.19 | 19.79 | 2.77 | 3.16 | 3.41 | 0.19 | 0.12 | 0.12 | 0.12 | 0.00 | 0.92 | 0.90 | 0.17 | 0.14 | 0.14 |

Table S.7: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, ARMA(1,1) case with $\rho_x = 0$. (First 3 columns are multiplied by 100).

| | | | MSE | | | Bias | | | Variance | | | Coverage | | Lenght | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ARMA(1,1) | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | GLS | FGLS | OLS | FGLS | OLS | FGLS |
| $T=200$ | $\gamma=0$ | -0.5,-0.4 | 1.13 | 0.27 | 0.29 | 8.52 | 4.56 | 4.28 | 1.06 | 0.38 | 0.28 | 0.96 | 0.95 | 0.40 | 0.21 |
| | | 0.2,-0.4 | 0.55 | 0.51 | 0.52 | 5.91 | 5.67 | 5.74 | 0.52 | 0.49 | 0.49 | 0.94 | 0.95 | 0.28 | 0.27 |
| | | 0.2,0.5 | 0.79 | 0.31 | 0.34 | 7.04 | 4.49 | 4.67 | 0.74 | 0.31 | 0.33 | 0.94 | 0.95 | 0.34 | 0.23 |
| | | 0.5,-0.4 | 0.54 | 0.52 | 0.53 | 5.85 | 5.77 | 5.80 | 0.50 | 0.50 | 0.50 | 0.94 | 0.94 | 0.28 | 0.28 |
| | | 0.5,0.5 | 1.22 | 0.22 | 0.24 | 8.77 | 3.78 | 3.94 | 1.14 | 0.22 | 0.24 | 0.94 | 0.95 | 0.42 | 0.19 |
| | | 0.8,-0.4 | 0.75 | 0.43 | 0.45 | 6.95 | 5.29 | 5.42 | 0.69 | 0.43 | 0.42 | 0.94 | 0.95 | 0.33 | 0.25 |
| | | 0.8,0.5 | 2.83 | 0.16 | 0.17 | 13.47 | 3.20 | 3.30 | 2.68 | 0.16 | 0.17 | 0.95 | 0.95 | 0.64 | 0.16 |
| | $\gamma=0.25$ | -0.5,-0.4 | 5.55 | 0.26 | 0.31 | 21.44 | 4.10 | 4.51 | 0.99 | 0.25 | 0.26 | 0.43 | 0.93 | 0.39 | 0.20 |
| | | 0.2,-0.4 | 0.70 | 0.44 | 0.52 | 6.68 | 5.34 | 5.75 | 0.49 | 0.46 | 0.46 | 0.90 | 0.94 | 0.27 | 0.26 |
| | | 0.2,0.5 | 3.34 | 0.27 | 0.33 | 16.59 | 4.19 | 4.62 | 0.69 | 0.29 | 0.31 | 0.47 | 0.96 | 0.33 | 0.21 |
| | | 0.5,-0.4 | 0.50 | 0.44 | 0.48 | 5.65 | 5.28 | 5.55 | 0.48 | 0.47 | 0.47 | 0.94 | 0.94 | 0.27 | 0.26 |
| | | 0.5,0.5 | 6.56 | 0.20 | 0.25 | 23.67 | 3.58 | 3.98 | 1.07 | 0.21 | 0.22 | 0.36 | 0.95 | 0.40 | 0.18 |
| | | 0.8,-0.4 | 1.52 | 0.37 | 0.41 | 10.23 | 4.87 | 5.15 | 0.67 | 0.40 | 0.39 | 0.80 | 0.94 | 0.32 | 0.24 |
| | | 0.8,0.5 | 11.95 | 0.14 | 0.18 | 30.92 | 3.06 | 3.42 | 2.54 | 0.15 | 0.16 | 0.53 | 0.94 | 0.62 | 0.15 |
| | $\gamma=0.5$ | -0.5,-0.4 | 13.91 | 0.21 | 0.36 | 36.08 | 3.61 | 4.65 | 0.82 | 0.21 | 0.22 | 0.01 | 0.89 | 0.35 | 0.18 |
| | | 0.2,-0.4 | 1.07 | 0.39 | 0.71 | 8.71 | 4.95 | 6.78 | 0.41 | 0.39 | 0.39 | 0.75 | 0.86 | 0.25 | 0.24 |
| | | 0.2,0.5 | 8.29 | 0.25 | 0.48 | 27.75 | 4.01 | 5.45 | 0.57 | 0.25 | 0.27 | 0.04 | 0.87 | 0.30 | 0.19 |
| | | 0.5,-0.4 | 0.56 | 0.40 | 0.57 | 6.01 | 5.04 | 6.09 | 0.40 | 0.40 | 0.40 | 0.91 | 0.90 | 0.25 | 0.24 |
| | | 0.5,0.5 | 16.59 | 0.18 | 0.36 | 39.56 | 3.40 | 4.73 | 0.89 | 0.18 | 0.19 | 0.01 | 0.87 | 0.37 | 0.16 |
| | | 0.8,-0.4 | 3.01 | 0.34 | 0.50 | 15.63 | 4.61 | 5.66 | 0.58 | 0.34 | 0.33 | 0.47 | 0.89 | 0.82 | 0.23 |
| | | 0.8,0.5 | 27.92 | 0.13 | 0.27 | 50.66 | 2.89 | 4.18 | 2.16 | 0.13 | 0.14 | 0.06 | 0.84 | 0.57 | 0.14 |
| $T=500$ | $\gamma=0$ | -0.5,-0.4 | 0.43 | 0.10 | 0.11 | 5.22 | 2.59 | 2.63 | 0.42 | 0.10 | 0.11 | 0.95 | 0.95 | 0.25 | 0.13 |
| | | 0.2,-0.4 | 0.21 | 0.19 | 0.20 | 3.66 | 3.49 | 3.54 | 0.21 | 0.19 | 0.19 | 0.95 | 0.94 | 0.18 | 0.17 |
| | | 0.2,0.5 | 0.29 | 0.12 | 0.12 | 4.34 | 2.69 | 2.74 | 0.30 | 0.12 | 0.13 | 0.96 | 0.96 | 0.21 | 0.14 |
| | | 0.5,-0.4 | 0.20 | 0.20 | 0.20 | 3.56 | 3.56 | 3.56 | 0.20 | 0.20 | 0.20 | 0.95 | 0.95 | 0.18 | 0.17 |
| | | 0.5,0.5 | 0.43 | 0.08 | 0.09 | 5.32 | 2.25 | 2.30 | 0.46 | 0.09 | 0.09 | 0.97 | 0.95 | 0.27 | 0.12 |
| | | 0.8,-0.4 | 0.27 | 0.17 | 0.17 | 4.16 | 3.28 | 3.32 | 0.28 | 0.17 | 0.17 | 0.96 | 0.95 | 0.21 | 0.16 |
| | | 0.8,0.5 | 1.03 | 0.06 | 0.06 | 8.17 | 1.91 | 1.94 | 1.11 | 0.06 | 0.07 | 0.96 | 0.95 | 0.41 | 0.10 |
| | $\gamma=0.25$ | -0.5,-0.4 | 4.85 | 0.10 | 0.11 | 21.12 | 2.48 | 2.61 | 0.39 | 0.10 | 0.10 | 0.07 | 0.93 | 0.24 | 0.12 |
| | | 0.2,-0.4 | 0.43 | 0.19 | 0.21 | 5.39 | 3.47 | 3.63 | 0.19 | 0.18 | 0.18 | 0.79 | 0.93 | 0.17 | 0.16 |
| | | 0.2,0.5 | 2.97 | 0.12 | 0.14 | 16.39 | 2.79 | 3.02 | 0.28 | 0.12 | 0.11 | 0.13 | 0.93 | 0.21 | 0.13 |
| | | 0.5,-0.4 | 0.26 | 0.20 | 0.23 | 4.11 | 3.55 | 3.85 | 0.19 | 0.19 | 0.18 | 0.90 | 0.93 | 0.17 | 0.17 |
| | | 0.5,0.5 | 5.90 | 0.09 | 0.11 | 23.35 | 2.33 | 2.56 | 0.43 | 0.08 | 0.09 | 0.06 | 0.92 | 0.26 | 0.11 |
| | | 0.8,-0.4 | 1.16 | 0.17 | 0.18 | 9.46 | 3.26 | 3.41 | 0.27 | 0.16 | 0.16 | 0.57 | 0.93 | 0.20 | 0.15 |
| | | 0.8,0.5 | 10.22 | 0.06 | 0.07 | 30.21 | 1.97 | 2.15 | 1.05 | 0.06 | 0.06 | 0.15 | 0.93 | 0.40 | 0.10 |
| | $\gamma=0.5$ | -0.5,-0.4 | 13.54 | 0.09 | 0.15 | 36.35 | 2.34 | 3.05 | 0.33 | 0.08 | 0.09 | 0.00 | 0.88 | 0.23 | 0.12 |
| | | 0.2,-0.4 | 0.84 | 0.16 | 0.24 | 8.28 | 3.20 | 3.90 | 0.16 | 0.15 | 0.15 | 0.47 | 0.89 | 0.16 | 0.15 |
| | | 0.2,0.5 | 8.00 | 0.10 | 0.17 | 27.85 | 2.49 | 3.31 | 0.23 | 0.10 | 0.10 | 0.00 | 0.87 | 0.19 | 0.12 |
| | | 0.5,-0.4 | 0.32 | 0.17 | 0.28 | 4.62 | 3.26 | 4.31 | 0.16 | 0.16 | 0.16 | 0.84 | 0.86 | 0.16 | 0.15 |
| | | 0.5,0.5 | 16.23 | 0.07 | 0.13 | 39.83 | 2.10 | 2.92 | 0.36 | 0.07 | 0.07 | 0.00 | 0.87 | 0.23 | 0.10 |
| | | 0.8,-0.4 | 2.74 | 0.14 | 0.19 | 15.78 | 3.00 | 3.45 | 0.24 | 0.13 | 0.14 | 0.09 | 0.90 | 0.19 | 0.14 |
| | | 0.8,0.5 | 27.60 | 0.05 | 0.09 | 51.74 | 1.78 | 2.43 | 0.89 | 0.05 | 0.05 | 0.00 | 0.87 | 0.37 | 0.09 |

Table S.8: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals. Predictive regression with $\rho_x = 0$. (First 3 columns are multiplied by 100).

| | | | MSE | | | | Bias | | | | Variance | | | | Coverage | | | Lenght | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MA(1) | OLS | GLS | GMA | FGLS | OLS | GLS | GMA | FGLS | OLS | GLS | GMA | FGLS | OLS | GMA | FGLS | OLS | GMA | FGLS |
| $T = 200$ | $\gamma = 0$ | -0.7 | 0.75 | 0.27 | 0.30 | 0.33 | 6.87 | 4.10 | 4.27 | 4.52 | 0.76 | 0.27 | 0.28 | 0.32 | 0.94 | 0.93 | 0.94 | 0.34 | 0.21 | 0.22 |
| | | -0.4 | 0.57 | 0.41 | 0.44 | 0.44 | 5.99 | 5.07 | 5.23 | 5.26 | 0.59 | 0.43 | 0.43 | 0.46 | 0.95 | 0.94 | 0.95 | 0.30 | 0.26 | 0.27 |
| | | 0.5 | 0.63 | 0.36 | 0.39 | 0.41 | 6.25 | 4.78 | 4.98 | 5.09 | 0.63 | 0.39 | 0.39 | 0.43 | 0.94 | 0.94 | 0.96 | 0.31 | 0.24 | 0.26 |
| | $\gamma = 0.25$ | -0.7 | 0.73 | 0.27 | 0.30 | 0.39 | 6.77 | 4.08 | 4.37 | 4.88 | 0.71 | 0.26 | 0.21 | 0.31 | 0.94 | 0.88 | 0.91 | 0.33 | 0.18 | 0.22 |
| | | -0.4 | 0.57 | 0.43 | 0.49 | 0.47 | 6.05 | 5.24 | 5.56 | 5.51 | 0.55 | 0.41 | 0.35 | 0.43 | 0.94 | 0.91 | 0.94 | 0.29 | 0.23 | 0.26 |
| | | 0.5 | 0.63 | 0.38 | 0.47 | 0.43 | 6.38 | 4.88 | 5.43 | 5.23 | 0.59 | 0.36 | 0.48 | 0.41 | 0.95 | 0.95 | 0.94 | 0.30 | 0.27 | 0.25 |
| | $\gamma = 0.5$ | -0.7 | 0.61 | 0.23 | 0.31 | 0.45 | 6.15 | 3.77 | 4.41 | 5.35 | 0.61 | 0.22 | 0.16 | 0.27 | 0.95 | 0.83 | 0.87 | 0.30 | 0.15 | 0.20 |
| | | -0.4 | 0.48 | 0.36 | 0.55 | 0.41 | 5.51 | 4.82 | 5.87 | 5.10 | 0.47 | 0.35 | 0.27 | 0.37 | 0.95 | 0.83 | 0.94 | 0.27 | 0.20 | 0.24 |
| | | 0.5 | 0.54 | 0.31 | 0.54 | 0.38 | 5.90 | 4.45 | 5.89 | 4.91 | 0.50 | 0.31 | 0.50 | 0.35 | 0.95 | 0.95 | 0.94 | 0.28 | 0.28 | 0.23 |
| $T = 500$ | $\gamma = 0$ | -0.7 | 0.30 | 0.10 | 0.11 | 0.11 | 4.31 | 2.52 | 2.56 | 2.63 | 0.30 | 0.10 | 0.11 | 0.11 | 0.95 | 0.95 | 0.96 | 0.22 | 0.13 | 0.13 |
| | | -0.4 | 0.23 | 0.17 | 0.17 | 0.18 | 3.80 | 3.26 | 3.29 | 3.32 | 0.23 | 0.17 | 0.17 | 0.18 | 0.95 | 0.95 | 0.95 | 0.19 | 0.16 | 0.17 |
| | | 0.5 | 0.25 | 0.15 | 0.16 | 0.16 | 4.01 | 3.13 | 3.20 | 3.23 | 0.25 | 0.15 | 0.15 | 0.16 | 0.96 | 0.95 | 0.95 | 0.20 | 0.15 | 0.16 |
| | $\gamma = 0.25$ | -0.7 | 0.28 | 0.10 | 0.11 | 0.13 | 4.19 | 2.45 | 2.65 | 2.93 | 0.28 | 0.10 | 0.08 | 0.11 | 0.95 | 0.89 | 0.92 | 0.21 | 0.11 | 0.13 |
| | | -0.4 | 0.22 | 0.16 | 0.18 | 0.17 | 3.69 | 3.14 | 3.38 | 3.24 | 0.22 | 0.16 | 0.14 | 0.17 | 0.95 | 0.91 | 0.96 | 0.18 | 0.15 | 0.16 |
| | | 0.5 | 0.23 | 0.14 | 0.17 | 0.16 | 3.88 | 3.04 | 3.30 | 3.25 | 0.24 | 0.14 | 0.19 | 0.15 | 0.95 | 0.96 | 0.95 | 0.19 | 0.17 | 0.15 |
| | $\gamma = 0.5$ | -0.7 | 0.24 | 0.08 | 0.16 | 0.18 | 3.88 | 2.26 | 3.30 | 3.48 | 0.24 | 0.08 | 0.06 | 0.09 | 0.95 | 0.73 | 0.85 | 0.19 | 0.09 | 0.12 |
| | | -0.4 | 0.18 | 0.13 | 0.28 | 0.15 | 3.40 | 2.89 | 4.36 | 3.02 | 0.19 | 0.14 | 0.11 | 0.14 | 0.95 | 0.75 | 0.94 | 0.17 | 0.13 | 0.15 |
| | | 0.5 | 0.20 | 0.12 | 0.20 | 0.16 | 3.57 | 2.80 | 3.58 | 3.25 | 0.20 | 0.12 | 0.20 | 0.13 | 0.95 | 0.95 | 0.93 | 0.18 | 0.18 | 0.14 |

Table S.9: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals. Serially correlated and heteroskedastic errors with $\rho_x = 0$. (First 3 columns are multiplied by 100).

| | | | MSE | | | Bias | | | Variance | | | Coverage | | | Lenght | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | OLS | F-C | F-CH | OLS | F-C | F-CH | OLS | F-C | F-CH | OLS | F-C | F-CH | OLS | F-C | F-CH |
| $w = x$ | $\nu = x^2$ | AR(1) | 6.36 | 3.76 | 2.65 | 19.85 | 15.42 | 13.05 | 6.32 | 3.70 | 2.96 | 0.95 | 0.94 | 0.96 | 0.98 | 0.76 | 0.68 |
| | | AR(2) | 48.61 | 1.51 | 1.02 | 54.53 | 9.74 | 8.14 | 39.56 | 1.59 | 1.16 | 0.91 | 0.95 | 0.96 | 2.38 | 0.50 | 0.42 |
| | | MA(1) | 2.37 | 0.42 | 0.40 | 12.17 | 5.12 | 4.96 | 2.36 | 0.43 | 0.38 | 0.94 | 0.94 | 0.94 | 0.60 | 0.25 | 0.24 |
| | | ARMA(1,1) | 6.86 | 4.13 | 3.11 | 20.62 | 16.18 | 14.15 | 6.62 | 3.80 | 3.15 | 0.94 | 0.94 | 0.96 | 1.00 | 0.77 | 0.70 |
| | $\nu = \log(x)^2$ | AR(1) | 0.77 | 0.46 | 0.18 | 6.90 | 5.40 | 3.39 | 0.75 | 0.46 | 0.20 | 0.94 | 0.95 | 0.97 | 0.34 | 0.27 | 0.17 |
| | | AR(2) | 6.03 | 0.19 | 0.06 | 19.22 | 3.43 | 1.96 | 4.88 | 0.20 | 0.07 | 0.91 | 0.95 | 0.96 | 0.84 | 0.17 | 0.10 |
| | | MA(1) | 0.29 | 0.05 | 0.05 | 4.28 | 1.80 | 1.68 | 0.28 | 0.05 | 0.04 | 0.94 | 0.95 | 0.92 | 0.21 | 0.09 | 0.08 |
| | | ARMA(1,1) | 0.83 | 0.50 | 0.27 | 7.22 | 5.63 | 4.12 | 0.79 | 0.47 | 0.25 | 0.94 | 0.95 | 0.95 | 0.35 | 0.27 | 0.20 |
| | $\nu = \exp(0.2(x + x^2))$ | AR(1) | 13.30 | 8.10 | 3.61 | 28.88 | 22.67 | 15.26 | 13.25 | 6.10 | 4.00 | 0.94 | 0.91 | 0.96 | 1.41 | 0.97 | 0.78 |
| | | AR(2) | 82.05 | 2.81 | 1.19 | 71.07 | 13.16 | 8.72 | 68.01 | 2.62 | 1.34 | 0.91 | 0.94 | 0.96 | 3.11 | 0.64 | 0.45 |
| | | MA(1) | 4.63 | 0.72 | 0.60 | 17.01 | 6.63 | 6.08 | 4.61 | 0.70 | 0.52 | 0.94 | 0.94 | 0.93 | 0.83 | 0.32 | 0.28 |
| | | ARMA(1,1) | 14.09 | 8.94 | 4.48 | 29.73 | 23.70 | 16.96 | 13.75 | 6.25 | 4.37 | 0.93 | 0.90 | 0.95 | 1.44 | 0.98 | 0.82 |
| $w = \phi x + (1 - \phi)q$ | $\nu = x^2$ | AR(1) | 6.90 | 4.15 | 3.43 | 20.93 | 16.38 | 14.66 | 6.34 | 3.72 | 3.50 | 0.93 | 0.94 | 0.95 | 0.98 | 0.76 | 0.73 |
| | | AR(2) | 49.07 | 1.74 | 1.36 | 55.64 | 10.51 | 9.27 | 39.78 | 1.59 | 1.41 | 0.89 | 0.94 | 0.95 | 2.38 | 0.50 | 0.46 |
| | | MA(1) | 2.59 | 0.41 | 0.39 | 12.82 | 5.05 | 4.96 | 2.36 | 0.42 | 0.42 | 0.93 | 0.95 | 0.95 | 0.60 | 0.25 | 0.25 |
| | | ARMA(1,1) | 7.19 | 4.38 | 3.73 | 21.37 | 16.85 | 15.25 | 6.69 | 3.82 | 3.65 | 0.93 | 0.94 | 0.94 | 1.01 | 0.77 | 0.75 |
| | $\nu = \log(x)^2$ | AR(1) | 0.82 | 0.49 | 0.37 | 7.22 | 5.65 | 4.78 | 0.76 | 0.46 | 0.40 | 0.93 | 0.95 | 0.96 | 0.34 | 0.27 | 0.24 |
| | | AR(2) | 5.96 | 0.21 | 0.15 | 19.40 | 3.66 | 3.01 | 4.90 | 0.20 | 0.17 | 0.89 | 0.95 | 0.96 | 0.84 | 0.17 | 0.15 |
| | | MA(1) | 0.31 | 0.05 | 0.05 | 4.44 | 1.76 | 1.70 | 0.29 | 0.05 | 0.05 | 0.93 | 0.96 | 0.95 | 0.21 | 0.09 | 0.08 |
| | | ARMA(1,1) | 0.85 | 0.52 | 0.40 | 7.36 | 5.84 | 4.95 | 0.81 | 0.47 | 0.41 | 0.93 | 0.94 | 0.95 | 0.35 | 0.27 | 0.25 |
| | $\nu = \exp(0.2(x + x^2))$ | AR(1) | 13.80 | 8.42 | 5.22 | 29.76 | 23.39 | 18.05 | 12.79 | 6.14 | 5.40 | 0.93 | 0.91 | 0.95 | 1.39 | 0.97 | 0.90 |
| | | AR(2) | 84.12 | 3.13 | 1.90 | 72.78 | 14.18 | 10.92 | 67.91 | 2.63 | 1.96 | 0.89 | 0.93 | 0.95 | 3.10 | 0.64 | 0.54 |
| | | MA(1) | 4.91 | 0.71 | 0.63 | 17.76 | 6.66 | 6.25 | 4.48 | 0.70 | 0.63 | 0.93 | 0.96 | 0.94 | 0.82 | 0.32 | 0.30 |
| | | ARMA(1,1) | 14.21 | 8.96 | 5.93 | 30.15 | 24.20 | 19.25 | 13.39 | 6.31 | 5.74 | 0.93 | 0.91 | 0.95 | 1.42 | 0.99 | 0.93 |