Feasible GLS for Time Series Regression*

Pierre Perron† Boston University

Emilio González-Coya‡ Boston University

September 8, 2022; revised September 15, 2026

Abstract

We consider GLS and OLS estimation in a linear regression model with serially correlated errors, and we provide the following contributions. First, we clarify when OLS is consistent or not. Second, we give sufficient conditions such that GLS is valid without the assumption of exogenous regressors (uncorrelated with past innovations). Third, we devise a feasible GLS procedure valid whether or not the regressors are exogenous, and which achieves a MSE close to that of the correctly specified infeasible GLS. Fourth, we derive the asymptotic distribution of this feasible GLS estimate: with exogenous regressors, the limit distribution is the same as that of the infeasible GLS estimator. However, with non-exogenous regressors, an additional term is present in the asymptotic variance. We provide a variance estimator, free of tuning parameters, which delivers confidence intervals with asymptotically correct coverage in both cases. We also illustrate how GLS can be more robust than OLS when the regressors are exogenous, even when GLS is based on an incorrect correction. The main assumptions are: a) the regressors are pre-determined (uncorrelated with future innovations); b) the errors are stationary thereby admitting a Wold representation in terms of some unpredictable innovations; c) the moving-average representation of the errors is invertible and implies a short-memory process. We also briefly address issues related to 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, Marco Olivari and Robert Taylor for useful comments.

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

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

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 the Best Linear Unbiased Estimate (BLUE). 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.

Spurred by the development of the Generalized Method of Moments (GMM) by Hansen (1982) econometricians started to tackle the problem of estimating the limit variance of the OLS estimate. Early contributions include White and Domowitz (1984), Newey and West (1987) and a comprehensive treatment was provided by Andrews (1991). 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 little 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).

Three main reasons are generally advanced for adopting OLS instead of GLS. 1) It is often argued that GLS is inconsistent when the regressors are not exogenous (i.e., correlated with past innovations), a view reflected in leading textbooks; e.g., Stock and Watson (2019), ch. 16. 2) When applying GLS one needs to choose a specification for the serial correlation in the errors, and it is argued that an incorrect specification can lead to worse results than using OLS; i.e., while OLS is sub-optimal relative to GLS, it is viewed as more robust; see, e.g., Engle (1974), Judge et al. (1985), p. 281, and Choudhury et al. (1999). 3) Even with an adequate specification, the gains in accuracy may be minor and the inference misleading; e.g., inaccurate coverage rates when using standard estimates of the asymptotic distribution. Our goal is to re-examine these arguments. We show that, under conditions we view as weak, each of them can fail, and we delineate the conditions under which each conclusion does or does not hold, so that readers can assess the relative merits of the methods in their own

applications.

Our focus is on the linear model $y = X\beta + u$. Of course, our results rely on some crucial assumptions. The first is that the regressors are pre-determined, which is often viewed as less controversial than the requirement of exogenous regressors. The second is that the errors are a stationary process and the associated Wold linear representation is invertible. This is usually satisfied but may fail in some models, especially those involving rational expectations arguments. Under the stated conditions, the first and second contributions are 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. We assume that u_t is a stationary process so that it has a linear representation in terms of a (possibly) infinite linear model of the form $u_t = \sum_{j=0}^{\infty} c_j e_{t-j}$ with e_t being an $i.i.d.$ sequence satisfying $E(e_t|\Phi_{t-1}) = 0$ for some information set Φ_t, thereby making e_t the sequence of innovations of interest. The usual argument for the consistency of GLS relies on whether x_t is exogenous with respect to u_t. We argue that this leads to an incorrect result. One should analyze this issue 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 imply, in general, inconsistent OLS estimates, while the GLS estimates are consistent. Also, unlike OLS, GLS is consistent with lagged dependent variables as regressors.

We 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, as one cannot base such estimates on the OLS residuals. 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 asymptotic distribution and the variance estimator derived in Section 4.1, are, in general, very close to the nominal level whether or not the regressors are exogenous, with lengths much shorter than using OLS with heteroskedasticity and autocorrelation consistent standard errors. We provide extensive evidence for both ex-

ogenous and non-exogenous regressors. In most cases, the MSE of the FGLS is close to that of the infeasible GLS estimate.

We provide a formal analysis of the asymptotic distribution of the proposed FGLS estimate. We show that it is $\sqrt{T}$-consistent and asymptotically normal provided only that the regressors are pre-determined. Whether the first-step estimation of the autoregressive approximation affects the limiting distribution is governed by an orthogonality condition between the GLS-transformed regressors and the lagged errors. Under exogenous regressors, the condition holds, the estimation of the serial correlation parameters has no effect, and FGLS attains the same asymptotic distribution as the infeasible GLS. With regressors that are non-exogenous, the condition generally fails and the first step inflates the asymptotic variance of FGLS relative to that of the infeasible GLS, though it affects neither consistency nor the rate of convergence. An important practical implication is that the usual variance estimate yields confidence intervals that under-cover. Accordingly, we introduce a corrected variance estimator constructed from by-products of the two estimation steps, involving no kernel, bandwidth or other tuning parameters, which is consistent whether or not the regressors are exogenous and restores asymptotically correct coverage.

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 FGLS is valid and leads to more efficient estimates 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 (2024) to fit the heteroskedasticity and further reduce the MSE. We also show that GLS is more robust than OLS, in that even a blatantly incorrect GLS correction can achieve a lower MSE than OLS when both are consistent; see Remark 6.

The rest of the paper is structured as follows. Section 2 provides the general setup, motivation and conditions under which OLS and GLS are consistent. Section 3 presents preliminary issues related to the feasible GLS estimate proposed. Section 4 presents the procedures for the general case with an invertible short-memory stationary process for the errors. Section 4.1 derives the asymptotic distribution of the proposed FGLS estimate and introduces the associated variance estimator. Issues related to the inclusion of lagged dependent variables are discussed in Section 4.2. Section 5 presents extensive simulations about the finite sample properties of the OLS and FGLS for a wide variety of processes for the

serial correlation in the errors for both exogenous and non-exogenous regressors. Section 6 provides brief concluding remarks including some practical guidance. The Supplement contains the preliminary lemmas and the proofs of the results stated in Section 4.1. It also contains some technical derivations, additional material and simulation results. An empirical illustration of the importance of our results is presented in González-Coya and Perron (2025) who consider issues related to testing the Uncovered Interest Parity (UIP) condition.

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 $x_t' = (x_{1t}, \ldots, x_{qt})$ is a vector of regressors assumed to be stationary, $\beta' = (\beta_1, \ldots, \beta_q)$ 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 β is $\hat{\beta} = (X'X)^{-1} X'y$. We assume that the error sequence u_t is a stationary 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. Note that $e_t \sim i.i.d.\ (0, \sigma_e^2)$ (independent and identically distributed) with $E(e_t|\Phi_{t-1}) = 0$ for some information set Φ_t, thereby making e_t the sequence of innovations of interest. 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 say that the regressors are "pre-determined" if:

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

i.e., regressors uncorrelated with current and future innovations. In particular, the condition $E[e_t x_t] = 0$ is maintained throughout; if it fails, some instrumental variable procedure would be needed. Throughout, we shall maintain that this is the case with some comments about what happens when it does not hold in Remark 9. We label the regressors as exogenous if

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

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

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 differs in the literature. What we label as pre-determined is sometimes referred to as exogenous (or past and present exogenous), and what we refer to as exogenous is labeled as strictly exogenous (or present and future exogenous); e.g., Stock and Watson (2019), p. 573. We shall continue with our terminology. Also, the conditions are usually stated in terms of conditional expectations, i.e., $E\left[x_t|e_t,e_{t+1},...,e_T\right]=0$ 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 it is equivalent to $E(x_t\sum_{j=0}^{t}c_j e_{t-j})=0$. In general, this implies the requirement $E(x_t e_{t-j})=0$, which is unlikely to be satisfied with non-exogenous regressors. 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 $u_t=e_t+c_1e_{t-1}+c_3e_{t-3}$. Also, suppose that u_t is an $MA(2)$. Then, if $c_1E(x_te_{t-1})=-c_2E(x_te_{t-2})$ and $E(x_te_t)=0$, we have $E(x_tu_t)=0$. Such cases are, however, unlikely to hold in practice.

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, $E(u_tx_t)=0$ implies that x_t is exogenous with respect to e_t since $E(u_tx_t)=\rho E(u_{t-1}x_t)+E(e_tx_t)=0$ 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.

Remark 1. *The Rational Expectations (RE) case. There is one non-trivial and empir-*

ically relevant exception for which OLS remains valid with serially correlated errors and non-exogenous regressors. This pertains to multi-steps ahead predictive regressions as examined, for instance, in Hansen and Hodrick (1980). In their framework, it is supposed that $E(y_{t+h}|\Phi_t) = x_t'\beta$, *where* Φ_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 error 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* $u_t = {}_{j=0}^{h-1}\, c_j e_{t-j}$. *OLS is then consistent only requiring pre-determined regressors so that* $E[x_{t-h}{}_{j=0}^{h-1} c_j e_{t-j}] = 0$. *Hence, this type of models 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 5 below, GLS remains consistent with non-exogenous regressors.*

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. Stating the condition as $E(x_t {}_{j=0}^{t-1} c_j e_{t-j}) = 0$ (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

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

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 [2]. This is a standard approach in the time series literature. Note that as T increases, e_{t,k_T} approaches e_t; see Section 4. Then, with pre-determined regressors,

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

Therefore, GLS is consistent without the need for exogenous regressors.

Remark 2. *Since $D'D = \Omega^{-1}$, GLS is invariant to the choice of D. Hence, only predetermined 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) $E[x_t e_{t+1}] = 0$, satisfied with predetermined regressors; and c) $E[(x_t - \rho x_{t+1})u_t] = 0$, also holding given

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

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

[2]Note that strictly speaking one should append a subscript k_T to ρ_j. We omit this explicit dependence in order to alleviate issues of cumbersome notation. The same applies to further estimates discussed below.

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

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

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; 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 variation, 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. The transformation eliminates u_t. Indeed, we can write (6) as

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

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 with predetermined regressors. There is no need to assume exogenous regressors. Then, with ρ known, one can consistently estimate β using the regression

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

Remark 3. *Keeping the AR(1) example, suppose we apply GLS with some arbitrary value $|\rho^*| < 1$. Then, with D^* as defined by (8) with ρ^* instead of ρ,*

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

Therefore, assuming pre-determined regressors what is needed for consistency is either a) exogenous regressors irrespective of the value of ρ and ρ^*; or b) non-exogenous regressors

and $\rho = \rho^*$. Accordingly, if the regressors are exogenous, GLS is consistent using any value of ρ^*, including 0, so that OLS is consistent, a well-known result. 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.

Remark 4. *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 5. *GLS is, in general, consistent with predictive regressions of the type discussed in Remark 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.3, that even for this case GLS performs much better.*

Remark 6. *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. Section S.1 in the supplement considers a very simple AR(1)-based procedure to obtain a GLS estimate that is (almost) never worse than OLS, subject to very minor random deviations. 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. Still, the results 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. The general message is that, within the class of designs considered, GLS-type corrections, even when based on misspecified structures, deliver an MSE no larger, and usually smaller, than OLS; the full benefits, of course, require an appropriate specification, an issue addressed by the FGLS procedure proposed below.*

3 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 ρ 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 $\hat{\rho}^{CO}$ be. A method valid without exogenous regressors is to first estimate ρ 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 ρ, say $\hat{\rho}^D$, can be obtained estimating (12) by OLS. 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 β. The estimates of β and ρ will be consistent with regressors exogenous or not as long as they are pre-determined. Alternatively, one can simply estimate β 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 $E(x_{t+1}e_t) \neq 0$, so that the regressors are not exogenous. Accordingly, $E(x_t u_t) \neq 0$ and OLS is inconsistent. Note that $E(e_t x_t) = 0$ 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) $y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + e_t \quad (GLS)$; d) $y_t - \hat{\rho} y_{t-1} = (x_t - \hat{\rho} x_{t-1})'\beta + e_t \quad (FGLS)$

The first is OLS; the second is the Durbin regression from which consistent estimates of ρ and β can be obtained. The third is the infeasible GLS based on the known value of ρ (to be used as a benchmark). The fourth is a feasible GLS regression for which we shall use two estimates of ρ: a) that used in the Cochrane and Orcutt procedure, labelled CO-FGLS. b) The estimate of ρ obtained from the Durbin regression, labelled as FGLS.

The results are presented in Table 1. The bias and MSE of OLS is very large, in accordance with its inconsistency. The Durbin and FGLS methods lead to very small biases, since

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. The implied estimate of ρ has a substantial bias with mean 0.63 instead of 0.8. Here, the quasi-differencing operation is biased but still effective in reducing the bias in the estimate of β, though not as well as with the estimate from the Durbin regression. Using simulations with $T = 10,000$, we verified that the the FGLS estimate of β is more efficient than that from the Durbin regression with a MSE 31% smaller. Hence, we shall only consider the FGLS method.

4 FGLS for the general case

We now present the recommended feasible method, applicable to 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

$$y_t = \sum_{j=1}^{k_T} \rho_j y_{t-j} + x_t'\beta - \sum_{j=1}^{k_T} x_{t-j}'\delta_j + e_{kt}, \tag{14}$$

$$(y_t - \sum_{j=1}^{k_T} \hat{\rho}_j^D y_{t-j}) = (x_t - \sum_{j=1}^{k_T} \hat{\rho}_j^D x_{t-j})'\beta + e_{kt}, \ (t = k_T + 1, ..., T), \tag{15}$$

where $\hat{\rho}_j^D$ $(j = 1, ..., k_T)$ are the OLS estimates of the coefficients associated with the lagged dependent variables from regression (14). 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 (14) 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 $\hat{\sigma}_{ek^*}^2 = (T - k_T^{\max})_{t=k_{\max}+1}^{-1\ T} \hat{e}_{kt}^2$ and $\hat{e}_{kt}$ are the residuals from applying OLS to (14) 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 β is then obtained applying OLS to the regression (15) with $k_T = k_T^*$ using the observations $t = k_T^* + 1, ..., T$.

The asymptotic properties of the resulting estimate, together with the construction of the confidence intervals, are analyzed in Section 4.1.

Remark 7. *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 8. *To improve upon OLS, Baillie et al. (2025) proposed using the Durbin regression (14). They claim correctly that the estimate of β is consistent whether the regressors are exogenous or not. However, this leads to a less efficient estimate compared to FGLS. 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 (2025) 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 this method. As discussed below, it offers no additional advantage in extended contexts such as models with lagged dependent variables and non-predetermined regressors.*

Remark 9. *The crucial condition for GLS to be consistent is that the regressors be predetermined. With innovations correlated with some 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. (2025) advocate using the Durbin regression 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. 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 β whether or not exogeneity or pre-determinedness hold. When the omitted variable is observed, this can be achieved via the Durbin regression (12) using a variable addition test (e.g., Engle (1982)). The steps are the following: a) Estimate the Durbin regression (14) and get the estimate $\hat{\beta}^D$; b) construct the residuals $\hat{u}_t^D = y_t - x_t'\hat{\beta}^D$; 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 (1); e) Apply FGLS as outlined above to this regression. This will lead to a consistent of estimate of β with regressors pre-determined 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 none of the procedures discussed here will be consistent except in some special cases.

4.1 The asymptotic distribution of the FGLS estimate

We now provide a formal analysis of the asymptotic distribution of the two-step FGLS estimate defined by (14) and (15), under the data-generating process of Section 2. The analysis delivers the following results: the FGLS estimate is $\sqrt{T}$-consistent and asymptotically normal provided the regressors are pre-determined; with exogenous regressors, the first-step estimation of the coefficients ρ_j has no impact, so that FGLS attains the asymptotic distribution of the infeasible GLS; with regressors that non-exogenou, the first step inflates the asymptotic variance; and a variance estimator constructed from by-products of the two steps delivers asymptotically valid confidence intervals in both cases. The preliminary lemmas and all proofs are gathered in the Supplement.

We first define the notation used. Recall from Section 2 that u_t has the autoregressive representation $A(L)u_t = e_t$ with $A(L) = C(L)^{-1} = 1 - \sum_{j=1}^{\infty} \rho_j L^j$. Applying this whitening filter to the regressors defines

$$\widetilde{x}_t = A(L)x_t = x_t - \sum_{j=1}^{\infty} \rho_j x_{t-j}, \tag{16}$$

so that the GLS transformation maps (1) into the regression of $A(L)y_t$ on $\widetilde{x}_t$ with errors e_t, and

$$Q = E[\widetilde{x}_t \widetilde{x}_t'] = p\lim_{T\to\infty} T^{-1} X'\Omega^{-1} X. \tag{17}$$

Throughout this section and the Supplement, we write k for k_T and let $\hat{\rho}^D = (\hat{\rho}_1^D, ..., \hat{\rho}_k^D)'$ and $\rho^{(k)} = (\rho_1, ..., \rho_k)'$. Define the feasible and (truncated) infeasible filtered regressors and errors

$$\hat{\widetilde{x}}_t = x_t - \sum_{j=1}^{k} \hat{\rho}_j^D x_{t-j}, \qquad \check{\widetilde{x}}_t = x_t - \sum_{j=1}^{k} \rho_j x_{t-j}, \tag{18}$$

$$\hat{e}_{kt} = u_t - \sum_{j=1}^{k} \hat{\rho}_j^D u_{t-j}, \qquad \check{e}_{kt} = u_t - \sum_{j=1}^{k} \rho_j u_{t-j} = e_t + r_{kt}, \tag{19}$$

where $r_{kt} = \sum_{j>k} \rho_j u_{t-j}$ is the truncation remainder. Since $y_t - x_t'\beta = u_t$, the FGLS estimate obtained from (15) satisfies the exact identity

$$\hat{\beta}_{\text{FGLS}} - \beta = (\sum_t \hat{\widetilde{x}}_t \hat{\widetilde{x}}_t')^{-1} \sum_t \hat{\widetilde{x}}_t \hat{e}_{kt}, \tag{20}$$

where here and below $_t$ denotes summation over the effective sample $t = k+1, ..., T$. Since $k/T \to 0$, replacing $(T-k)^{-1}$ by T^{-1} (and $t = k+1, ..., T$ by $t = 1, ..., T$ in population-limit statements) affects no asymptotic statement, we do so without loss. For the first step, write $x_t = (1, w_t')'$ with w_t the stochastic regressors, define the vector of lagged errors $U_{t-1,k} = (u_{t-1}, ..., u_{t-k})'$ and the regressor block $W_{t,k} = (1, w_t', w_{t-1}', ..., w_{t-k}')'$ (a single intercept is understood in all regressions and, without loss of generality, we take $E(w_t) = 0$ in population second-moment calculations since the intercept absorbs means). Let $\psi_{t,k} = U_{t-1,k} - \Pi_k W_{t,k}$ denote the population residual of the linear projection of $U_{t-1,k}$ on $W_{t,k}$, with $\Sigma_k = E[\psi_{t,k}\psi_{t,k}']$. Finally, define the cross-moments between the filtered regressors and the lagged errors:

$$g_j = E[\widetilde{x}_t u_{t-j}] \ (j = 1, 2, ...), \qquad G_k = [g_1, ..., g_k]. \tag{21}$$

Throughout, $\|\cdot\|$ denotes the Euclidean norm for vectors and the induced (spectral) norm for matrices, $\|\cdot\|_F$ the Frobenius norm, and $\xrightarrow{p}$, $\xrightarrow{d}$ convergence in probability and in distribution. The analysis relies on the following assumptions, formalizing the conditions of Section 2.

Assumption 1. *$\{e_t\}$ is a stationary and ergodic martingale difference sequence with respect to the filtration Φ_t: $E(e_t|\Phi_{t-1}) = 0$, $E(e_t^2|\Phi_{t-1}) = \sigma_e^2 \in (0, \infty)$, and $E(e_t^4) < \infty$.*

Assumption 2. *$u_t = C(L)e_t$ as in (2) with $c_0 = 1$, $\sum_{j=0}^{\infty} j|c_j| < \infty$ and $C(z) \neq 0$ for $|z| \leq 1$, so that $A(L)u_t = e_t$ with $\sum_{j=1}^{\infty} j|\rho_j| < \infty$.[3] The spectral density of u_t, $f_u(\lambda) = (\sigma_e^2/2\pi)|C(e^{-i\lambda})|^2$, is bounded and bounded away from zero on $[-\pi, \pi]$.*

Assumption 3. *The process $\{(w_t', e_t)'\}$ is jointly stationary and ergodic, with $E\|x_t\|^4 < \infty$. Define the enlarged information set $\mathcal{G}_{t-1} = \sigma(x_t, x_{t-1}, ...; e_{t-1}, e_{t-2}, ...)$, which contains the contemporaneous regressors. The regressors are pre-determined in the conditional-mean sense*

$$E(e_t|\mathcal{G}_{t-1}) = 0, \tag{22}$$

which implies the orthogonality conditions $E(x_t e_t) = 0$ and $E(x_{t-j}e_t) = 0$ for all $j \geq 1$, i.e., condition (3). In addition, conditionally on the enlarged information set, the innovations are homoskedastic with bounded conditional kurtosis: $E(e_t^2|\mathcal{G}_{t-1}) = \sigma_e^2$ and $E(e_t^4|\mathcal{G}_{t-1}) \leq \kappa_4 < \infty$ a.s. The matrix Q in (17) is finite and positive definite.

[3]The weighted summability of $\{\rho_j\}$ follows from that of $\{c_j\}$ because the weighted Wiener algebra $\{\phi :_j (1+|j|)|\phi_j| < \infty\}$ is inverse-closed; see, e.g., Hannan and Deistler (2012), ch. 6.6. Readers preferring a primitive statement may simply impose $_j j|\rho_j| < \infty$ directly.

Assumption 4. *The autoregressive order $k = k_T$ satisfies*

$$k_T^3/T \to 0 \qquad \text{and} \qquad k_T\sqrt{T}\,{}_{j>k_T}|\rho_j| \to 0. \tag{23}$$

Assumption 5. *Let $z_t = (w_t', u_t)'$ and $\Gamma_z(h) = E[z_t z_{t+h}']$. (i) The autocovariances and the joint fourth-order cumulants of z_t are absolutely summable, i.e., ${}_h\|\Gamma_z(h)\| < \infty$ and ${}_{h_1,h_2,h_3}|cum(z_{at}, z_{b,t+h_1}, z_{c,t+h_2}, z_{d,t+h_3})| < \infty$ for all components a, b, c, d; see, e.g., Brillinger (2001), ch. 2. (ii) The joint spectral density matrix $f_z(\lambda)$ satisfies $c_z I \preceq f_z(\lambda) \preceq C_z I$ for all $\lambda \in [-\pi, \pi]$ and some $0 < c_z \leq C_z < \infty$. (iii) $\Gamma^* = \lim_{k\to\infty} G_k \Sigma_k^{-1} G_k'$ exists.*

These conditions are the analogues of those in Berk (1974) and Amemiya (1973), and they are satisfied in our setup. Assumption 1 holds, in particular, for the *i.i.d.* $(0, \sigma_e^2)$ innovations of Section 2 provided the fourth moment is finite, and Assumption 2 simply collects the conditions already imposed on (2). Some comments on the remaining conditions are in order. a) Condition (22) is the conditional-mean version of the pre-determinedness hypothesis (3); it is weaker than exogeneity, which would in addition require $E(x_t e_s) = 0$ for all $s < t$. The conditional moment restrictions on e_t^2 and e_t^4 hold, in particular, in the designs of Sections 3 and 5, where e_t is *i.i.d.* and independent of $\{v_s\}$ and of past innovations, so that conditioning on $\mathcal{G}_{t-1}$ does not affect the moments of e_t. b) Assumption 4 does not separately impose $k_T \to \infty$. For a true $AR(p)$ process the tail ${}_{j>k}|\rho_j|$ vanishes for any $k \geq p$, so bounded orders are permitted, while for a genuine $AR(\infty)$ process the tail condition forces $k_T \to \infty$. The condition holds for exponentially decaying ρ_j (e.g., invertible ARMA processes) with any $k_T \asymp (\ln T)^{1+\epsilon}$, $\epsilon > 0$, and for $|\rho_j| = O(j^{-m})$ provided $m > 7/2$. Regarding data-selected orders, Lemma A.2 in the Supplement shows that for $AR(p)$ errors the BIC-selected order satisfies Assumption 4 with probability approaching one, since $P(k_T^* = p) \to 1$. For processes with an infinite autoregressive representation, the BIC order grows at a logarithmic rate, which satisfies $k_T^3/T \to 0$ but need not satisfy the tail condition in (23): the BIC balances the squared truncation bias against the penalty, a phenomenon akin to that documented by Ng and Perron (1995) for information criteria in unit root autoregressions. The adjusted order $\hat{k}_T = \max\{k_T^*, \lceil (\ln T)^{3/2} \rceil\}$ restores the tail condition (Lemma A.2), and the results below should be understood as applying to such an order when the errors are not a finite-order autoregression; in practice the difference is immaterial for the sample sizes considered and the simulations use the BIC order throughout. c) Assumption 5(i) quantifies the short memory of the joint process, which is needed to obtain convergence rates for sample cross-moments whose dimension grows with k; a simple sufficient condition is that z_t be a linear process in *i.i.d.* innovations with finite fourth moments, absolutely summable coefficient matrices and a

non-degenerate transfer function, as in the designs of Sections 3 and 5. d) Assumption 5(ii) is an identification (no perfect collinearity) condition for the Durbin regression (14) and is not innocuous: it fails if x_t contains a lagged dependent variable, since then $u_{t-1} = y_{t-1} - x'_{t-1}\beta$ lies in the span of the regressors and (14) is perfectly collinear. This is the reason why the case with lagged dependent variables requires the separate treatment of Section 4.2. e) Assumption 5(iii) ensures that the asymptotic variance below is well defined when $k_T \to \infty$; it is not needed for $AR(p)$ errors, since $P(k_T^* = p) \to 1$ under the BIC and the relevant matrix is then $\Gamma_p = G_p \Sigma_p^{-1} G_p'$.

The following representation isolates the exact channel through which the first-step estimation of the autoregressive coefficients affects the second step.

Proposition 1. *Under Assumptions 1-5,*

$$\sqrt{T}(\hat{\beta}_{FGLS} - \beta) = Q^{-1}[T_t^{-1/2}\widetilde{x}_t e_t - \sqrt{T} G_k(\hat{\rho}^D - \rho^{(k)})] + o_p(1). \tag{24}$$

The first term inside the brackets is the score of the infeasible GLS estimate. By a martingale central limit theorem, $T_t^{-1/2}\widetilde{x}_t e_t \xrightarrow{d} N(0, \sigma_e^2 Q)$, so that $\sqrt{T}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, \sigma_e^2 Q^{-1})$ under pre-determined regressors only (Lemma A.1 in the Supplement). The second term captures the effect of the first-step estimation. Lemma A.3 in the Supplement establishes that $\|\hat{\rho}^D - \rho^{(k)}\| = O_p(\sqrt{k/T})$ whether or not the regressors are exogenous together with the linear representation $\sqrt{T}(\hat{\rho}^D - \rho^{(k)}) = \Sigma_k^{-1} T_t^{-1/2} \psi_{t,k} e_t + o_p(1)$. Two features of (24) deserve emphasis. First, the $\sqrt{T}$-consistency of $\hat{\rho}^D$ is not, by itself, sufficient for the first step to be asymptotically irrelevant: the term $\sqrt{T} G_k(\hat{\rho}^D - \rho^{(k)})$ is $O_p(1)$ in general. It vanishes precisely when $G_k = 0$, i.e., when the cross-moments $g_j = E[\widetilde{x}_t u_{t-j}]$ between the GLS-filtered regressors and the lagged errors are all zero. Second, each g_j is a linear combination of moments of the form $E(x_{t-i} e_{t-j-l})$ ($i \geq 0$, $l \geq 0$, $j \geq 1$), all of which vanish under exogenous regressors, whereas non-exogenous regressors leave the moments involving past innovations unrestricted, so that $g_j \neq 0$ in general. The main result follows.

Theorem 1. *Suppose Assumptions 1-5 hold. (i) If the orthogonality condition*

$$g_j = E[\widetilde{x}_t u_{t-j}] = 0 \text{ for all } j \geq 1 \tag{25}$$

holds—in particular, if the regressors are exogenous—then the first-step estimation has no effect and

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

(ii) In general,

$$\sqrt{T}(\hat{\beta}_{FGLS} - \beta) = Q^{-1}T_t^{-1/2}s_{t,k} + o_p(1) \xrightarrow{d} N(0, Q^{-1}\Lambda Q^{-1}),$$

where $s_{t,k} = (\widetilde{x}_t - G_k\Sigma_k^{-1}\psi_{t,k})e_t$ *and*

$$\Lambda = \sigma_e^2(Q + \Gamma^*), \qquad \Gamma^* = \lim_{k\to\infty} G_k\Sigma_k^{-1}G_k' \succeq 0, \tag{26}$$

so that $Q^{-1}\Lambda Q^{-1} = \sigma_e^2 Q^{-1} + \sigma_e^2 Q^{-1}\Gamma^* Q^{-1} \succeq \sigma_e^2 Q^{-1}$.

Part (i) states the precise sense in which the estimation of the parameters $\hat{\rho}_j^D$ has no asymptotic effect. Under the orthogonality condition (25), FGLS is asymptotically equivalent to the infeasible GLS, in line with the classical equivalence result of Amemiya (1973), who assumed non-stochastic (hence exogenous) regressors. Part (ii) shows that pre-determinedness alone preserves $\sqrt{T}$-consistency and asymptotic normality—the limit law remains centred at β—but the plug-in step inflates the asymptotic variance by the positive semi-definite matrix $\sigma_e^2 Q^{-1}\Gamma^* Q^{-1}$. En route to (26), the finite-k variance of the effective score $s_{t,k}$ also involves the cross-moments $E[\widetilde{x}_t\psi_{t,k}']$, which vanish as $k \to \infty$ (and exactly, for $AR(p)$ errors, once $k \geq p$), leaving the clean limit Λ. When the errors are a finite-order $AR(p)$ process and k is selected by the BIC, $P(k_T^* = p) \to 1$ and the operative correction is Γ_p rather than the limit Γ^*; the variance estimator introduced below automatically tracks whichever applies.

The result has an immediate implication for inference. The standard variance estimate associated with the quasi-differenced regression (15), namely

$$\widehat{AVar} = \hat{\sigma}_{e,k}^2[(T-k)_t^{-1}\hat{\widetilde{x}}_t\hat{\widetilde{x}}_t']^{-1}, \qquad \hat{\sigma}_{e,k}^2 = (T-k)_t^{-1}\widetilde{e}_{kt}^2, \tag{27}$$

with $\widetilde{e}_{kt}$ the OLS residuals from (15), converges in probability to $\sigma_e^2 Q^{-1}$ (Lemma A.4 in the Supplement). It is therefore consistent for the asymptotic variance of FGLS under the orthogonality condition (25), but it omits the correction term $\sigma_e^2 Q^{-1}\Gamma^* Q^{-1}$ otherwise, so that the associated Wald confidence intervals under-cover with non-exogenous regressors. It is worth noting that this deficiency cannot be repaired by applying a HAC correction to the quasi-differenced regression: the effective score $s_{t,k}$ is a martingale difference sequence and the autocovariances of $\widetilde{x}_t e_t$ vanish at all lags, so a HAC estimator applied to $\hat{\widetilde{x}}_t\widetilde{e}_{kt}$ still converges to $\sigma_e^2 Q$. The missing term arises from the estimated-filter channel in (24) and must be targeted directly. This motivates the following estimator, which is valid in the general case.

All the ingredients are by-products of the two estimation steps. Let $\hat{u}_t = y_t - x_t'\hat{\beta}_{\text{FGLS}}$ $(t = 1, ..., T)$ denote the unfiltered FGLS residuals and define

$$\hat{g}_j = (T-k)_t^{-1}\hat{\widetilde{x}}_t\hat{u}_{t-j}, \qquad \hat{G}_k = [\hat{g}_1, ..., \hat{g}_k]. \tag{28}$$

Let $\hat{\psi}_{t,k}$ denote the sample residuals from the (multivariate) OLS regression of $(y_{t-1}, ..., y_{t-k})'$ on $W_{t,k}$, i.e., the first-stage regression underlying (14). Since $y_{t-j} = x_{t-j}'\beta + u_{t-j}$ and x_{t-j} lies in the span of $W_{t,k}$ for $0 \leq j \leq k$, these equal the sample projection residuals of the latent $U_{t-1,k}$ exactly. Set $\hat{\Sigma}_k = (T-k)_t^{-1}\hat{\psi}_{t,k}\hat{\psi}_{t,k}'$ and $\hat{Q}_k = (T-k)_t^{-1}\hat{\widetilde{x}}_t\hat{\widetilde{x}}_t'$. The proposed estimator is

$$\widehat{AVar}_R = \hat{\sigma}_{e,k}^2[\hat{Q}_k^{-1} + \hat{Q}_k^{-1}\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}_k'\hat{Q}_k^{-1}], \tag{29}$$

with the convention that the correction term (i.e. the second term in the bracket in equation 29) is zero when $k = 0$. No kernel, bandwidth or truncation choice beyond the already-selected order k is involved. Note that $\hat{g}_j$ pairs $\hat{\widetilde{x}}_t$ with the raw lagged residual $\hat{u}_{t-j}$, not with the projected residual $\hat{\psi}_{j,t,k}$. The target is $G_k = E[\widetilde{x}_t U_{t-1,k}']$, whereas $E[\widetilde{x}_t\psi_{t,k}'] \to 0$, so pairing with the projected residuals would estimate a vanishing matrix.

Theorem 2. *Suppose Assumptions 1-5 hold, part (iii) of Assumption 5 being used only where indicated. Then: (i)* $\|\hat{G}_k - G_k\| = O_p(\sqrt{k/T})$ *and* $\|\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}_k' - G_k\Sigma_k^{-1}G_k'\| = O_p(k/\sqrt{T}) = o_p(1)$. *(ii) Consequently,* $\widehat{AVar}_R - \sigma_e^2(Q^{-1} + Q^{-1}\Gamma_k Q^{-1}) \xrightarrow{p} 0$, *where* $\Gamma_k = G_k\Sigma_k^{-1}G_k'$ *is evaluated along the order sequence actually used. (iii) In particular: a) under the orthogonality condition (25), e.g., with exogenous regressors, the estimated correction satisfies* $\|\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}_k'\| = O_p(k/T)$ *and* $\widehat{AVar}_R \xrightarrow{p} \sigma_e^2 Q^{-1}$, *the same limit as (27); b) if* Γ_k *converges along the operative order sequence—as holds with limit* Γ^* *under Assumption 5(iii) when* $k \to \infty$, *and with limit* Γ_p *for AR(p) errors under the BIC—then, with* Γ° *denoting the corresponding limit,*

$$\widehat{AVar}_R \xrightarrow{p} \sigma_e^2(Q^{-1} + Q^{-1}\Gamma^\circ Q^{-1}) = Q^{-1}\Lambda Q^{-1},$$

the asymptotic variance of $\sqrt{T}(\hat{\beta}_{FGLS} - \beta)$ *in Theorem 1(ii). Hence* $\widehat{AVar}_R$ *consistently estimates the asymptotic variance of the FGLS estimate whether or not the regressors are exogenous.*

Several features are worth noting. First, every ingredient of (29) is available as a by-product of the two estimation steps: $\hat{\widetilde{x}}_t$ and $\hat{\sigma}_{e,k}^2$ from the quasi-differenced regression (15), $\hat{u}_t$ from $\hat{\beta}_{\text{FGLS}}$, and $\hat{\psi}_{t,k}$ from the first stage of the Durbin regression; there are no tuning

parameters. Second, the correction is positive semi-definite by construction. Third, the magnitude of the estimated correction $\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}_k'$ provides a direct diagnostic for the strength of non-exogeneity: values near zero indicate that (25) is approximately satisfied and that the two variance estimates will essentially coincide. The implications for confidence intervals are as follows, where $\Phi(\cdot)$ denotes the standard normal distribution function and $z_{1-\alpha/2}$ its $1-\alpha/2$ quantile.

Corollary 1. *Let c be a non-zero vector of the same dimension as β. Under the conditions of Theorem 2(iii), the Wald interval based on the corrected estimator, $c'\hat{\beta}_{FGLS} \pm z_{1-\alpha/2}(c'\widehat{AVar}_R\, c/T)^{1/2}$, has asymptotic coverage $1-\alpha$ whether or not the regressors are exogenous.*

Accordingly, all the FGLS confidence intervals reported in this paper are based on $\widehat{AVar}_R$. In Section 5, 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.

4.2 The case with lagged dependent variables as regressors

As stated in Remark 4, GLS is consistent with lagged dependent variables as regressors. However, alternative methods to get consistent estimate of the parameters ρ_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} $(j = 1, ..., q)$ are pre-determined regressors. When constructing the Durbin regression, one pre-multiplies both sides by $(1 - \sum_{j=1}^{k_T^*} \rho_j L^j)$ for some k_T^* selected via the BIC information criterion. Since the product of the two lag polynomials has order $k_T^* + p_y$, this leads to the regression

$$y_t = \sum_{j=1}^{k_T^*+p_y} \alpha_j^* y_{t-j} + x_t'\beta - \sum_{j=1}^{k_T^*} x_{t-j}'\delta_j + e_{kt}, \tag{30}$$

where the coefficients on the lagged dependent variables are obtained from the convolution of the two lag polynomials, i.e., $\alpha_j^* = \alpha_j + \rho_j - \sum_{i=1}^{j-1} \rho_i \alpha_{j-i}$ (with the conventions $\alpha_j = 0$ for $j > p_y$ and $\rho_j = 0$ for $j > k_T^*$), and $\delta_j = (\delta_{j1}, ..., \delta_{jq})$ with $\delta_{ji} = \beta_i \rho_j$. Accordingly, the parameters

ρ_j cannot be identified using the coefficient on the lagged dependent variable α_j^* since α_j is unknown. However, as suggested by Wallis (1967), one can obtain consistent estimates using the fact that $\rho_j = \delta_{ji}/\beta_i$, given by $\hat{\rho}_j^D = \hat{\delta}_{ji}/\hat{\beta}_i$. One can then proceed to construct the FGLS estimates as described in Step 2 above. A drawback is that with multiple regressors x_{jt}, there are many ways to construct an estimate of $\hat{\rho}_j^D$, one for each i. Simulations and applications reported in González-Coya and Perron (2025) show that the results are not sensitive to the choice of the variable used because GLS is robust to small variations in ρ_j.

5 Simulation results

The issues addressed are the following: the bias, variance and MSE of the FGLS estimates, 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 Ω to construct the estimate $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$, with $Var(\hat{\beta}_{GLS}|X) = \sigma_e^2(X'\Omega^{-1}X)^{-1}$, 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. 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 θ 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 (14) 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 regres-

sors. 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. For the FGLS estimate, the autoregressive order k_T^* is selected using the BIC as described in Section 4, with $k_T^{\max} = 12$; the median selected order across replications is reported in the last column of each table (labelled Med. k). The results are presented in Tables 2-5. We focus our discussion on the MSE and the confidence intervals.

To construct the confidence intervals for the FGLS estimate, we use the variance estimator of Theorem 2, which is valid whether or not the regressors are exogenous, i.e., $Var(\hat{\beta}_{\text{FGLS}}) = (T - k_T^*)^{-1}\widehat{AVar}_R$ with $\widehat{AVar}_R$ defined by (29) using $k = k_T^*$. Under exogenous regressors, it is asymptotically equivalent to the standard estimator $\hat{\sigma}_e^2(X_{k_T^*}'X_{k_T^*})^{-1}$ associated with the quasi-differenced regression (15), where $X_{k_T^*} = (x_{k_T^*+1}^{*\prime}, ..., x_T^{*\prime})'$, $x_t^* = x_t - \sum_{j=1}^{k_T^*}\hat{\rho}_j^D x_{t-j}$ 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 (15); under non-exogenous regressors it incorporates the correction required by Theorem 1 (see Corollary 1). 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 $\hat{\Sigma} = \sum_{j=-T+1}^{T-1} w(j/m)\hat{\Gamma}_v(j)$ with $\hat{\Gamma}_v(j) = T^{-1}\sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}'$ 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.

5.1 Simulations with exogenous regressors

We first present results with exogenous regressors, i.e., $\gamma = 0$, that allows a proper comparison since 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 case of the $AR(2)$ with parameters 1.34 and -0.42 for which the reduction is 96% when $T = 200$. Overall, the reductions are 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) All estimators are essentially unbiased, so that the reduction in MSE comes entirely from a reduction in variance. 4) 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. 5) 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 within random errors due to simulations. 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 in the literature. The main reason for why the coverage rates of FGLS 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. Note also that with exogenous regressors the correction embedded in $\widehat{AVar}_R$ is asymptotically negligible (Theorem 2), and indeed using the standard variance estimator yields essentially identical results in this case, so using the robust version is costless. 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 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.

5.2 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 γ, 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 56 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 159 times larger when $T = 200$ and 404 times when $T = 500$. The bias of OLS is much larger than that of FGLS for both $T = 200, 500$, given that OLS is inconsistent, and so is its variance when the errors are strongly correlated.

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, with the confidence intervals based on the variance estimator $\widehat{AVar}_R$, the coverage rates are close to the nominal 95% level throughout: for $AR(1)$ and $AR(2)$ errors they range between 0.91 and 0.95 when $T = 200$ and between 0.94 and 0.95 when $T = 500$, for all values of γ. This contrasts sharply with the under-coverage that obtains, with non-exogenous regressors, when the standard variance estimator (27) is used instead, in accordance with Corollary 1. 3) For $MA(1)$ errors, the coverage rates of both GMA and FGLS are good when $\gamma = 0.25$. When $\gamma = 0.5$, some under-coverage remains for FGLS when $\theta = -0.7$ (0.87 when $T = 200$; 0.85 when $T = 500$), while GMA is more accurate (0.91). The source of the problem here is the quality of the autoregressive approximation rather than the estimate of the variance. With $\theta = -0.7$ the implied autoregressive coefficients, $\rho_j = -(-\theta)^j$, decay slowly, while the BIC selects a median order of only 3 to 4 (see the Med. k column). The resulting truncation

induces a bias, apparent from the larger bias of FGLS relative to the infeasible GLS in Table 4, which the confidence interval is not designed to account for. Being tailored to the MA structure, the GMA transformation avoids the truncation and performs better. 4) For $ARMA(1,1)$ errors, the coverage rates of FGLS are good for $\gamma = 0.25$ (0.92 to 0.95) and adequate, though slightly less accurate, for $\gamma = 0.5$ (0.90 to 0.93 when $T = 200$; 0.88 to 0.94 when $T = 500$). The least accurate case is the process with parameters $(0.5, -0.4)$, for which the autoregressive and moving-average roots nearly cancel: the errors are then close to white noise, the BIC selects a median order $k_T^* = 0$, and FGLS essentially reduces to OLS, whose coverage exhibits similar distortions in that design. The results for the case with *i.i.d.* regressors are presented in the Supplement, with similar conclusions.

Remark 10. *As discussed in Remarks 1 and 5, 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 delivers substantially lower MSE and shorter confidence intervals than OLS, with results similar to the case with exogenous regressors.*

Remark 11. *González-Coya and Perron (2025) present simulation results about the power of tests on β 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 arises from the fact that with regressors that are serially correlated, as is usually the case, the introduction of many 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.*

Remark 12. *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 (2024). 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.4. 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.

5.3 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 Remark 1. The predictive model states that $E(y_{t+h}|\Phi_t) = x_t'\beta$, where Φ_t is the information set available at time t. Then, $y_{t+h} = x_t'\beta + u_{t+h}$, with $u_{t+h} = y_{t+h} - E(y_{t+h}|\Phi_t)$ so that the error terms are forecast errors from using the best predictor based on x_t. It can be shown that u_{t+h} is an $MA(h-1)$ process of the form $u_{t+h} = e_{t+h} + c_1 e_{t+h-1} + ... + c_{h-1}e_{t+1}$, with $e_t \sim i.i.d.\ (0, \sigma_e^2)$. Since $x_t \subset \Phi_t$, $E(x_t u_{t+h}) = 0$, OLS is consistent and can be applied with the relevant HAC correction. For simplicity, 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 $h = 2$, then u_{t+h} 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 $t = h, ..., T$, even in large samples. Continuing with the $MA(1)$ example with $u_{t+2} = e_{t+2} + ce_{t+1}$, 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. It does not imply that GLS is inconsistent since 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. Note that the observationally equivalent representation involves both the MA coefficient and the variance of the innovations: with $u_t = e_t + ce_{t-1}$, $e_t \sim i.i.d.(0, \sigma_e^2)$ and $|c| > 1$, the invertible representation is $u_t = \widetilde{e}_t + c^{-1}\widetilde{e}_{t-1}$ with $V(\widetilde{e}_t) = c^2\sigma_e^2$. Accordingly, writing $V(u) = \sigma_e^2\Omega$ as in Section 2, the matrix Ω is identified only up to this normalization, and the transformation that standardizes the errors is $\sigma_e^{-1}D$ rather than D, where $D'D = \Omega^{-1}$. The scale factor cancels from the GLS estimate of β, as just noted.

For FGLS, we can gain some insights by looking at the transformation of the model applying an autoregressive filter $\alpha(L)$. Then, $\alpha(L)u_{t+h}$ has the same autocovariance function

as $C(L)e_{t+h}$. If $C(L)$ is invertible, $\alpha(L) = C(L)^{-1}$ and $\alpha(L)u_{t+h} = e_{t+h}$. 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 MA polynomial $(1+c^{-1}L)$, i.e., the whitening filter $\alpha(L) = (1+c^{-1}L)^{-1}$. A researcher using the invertible model would not recover the true structural shocks, but rather

$$\begin{aligned}\widetilde{e}_{t+h} := (1+c^{-1}L)^{-1}u_{t+h} &= (1+c^{-1}L)^{-1}(1+cL)e_{t+h} \\ &= (1+c^{-1}L)^{-1}(1+c^{-1}L-c^{-1}L+cL)e_{t+h} \\ &= e_{t+h} + (c-c^{-1})(1+c^{-1}L)^{-1}e_{t+h-1} \\ &= e_{t+h} + (c-c^{-1})\textstyle\sum_{i=0}^{\infty}(-c^{-1})^i e_{t+h-1-i}.\end{aligned}$$

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 DX involves past values of x_t, FGLS will be consistent with exogenous regressors but will be inconsistent otherwise. The transformed regression has regressors $\widetilde{x}_{t+h} = (1+c^{-1}L)^{-1}x_t = \sum_{m=0}^{\infty}(-c^{-1})^m x_{t-m}$ and errors $\widetilde{e}_{t+h}$ as in the last display, and the probability limit of the resulting estimate differs from β by $(E[\widetilde{x}_{t+h}\widetilde{x}'_{t+h}])^{-1}E[\widetilde{x}_{t+h}\widetilde{e}_{t+h}]$. Since $E(x_{t-m}e_s) = 0$ for all $s \geq t-m$ when the regressors are pre-determined, the only surviving terms in the cross-moment are those involving innovations dated strictly prior to the regressors, i.e.,

$$E[\widetilde{x}_{t+h}\widetilde{e}_{t+h}] = (c-c^{-1})\textstyle\sum_{m=0}^{\infty}\sum_{i=h+m}^{\infty}(-c^{-1})^{m+i}E[x_{t-m}e_{t+h-1-i}].$$

These expectations all vanish if the regressors are exogenous, in which case FGLS remains consistent; with regressors that are non-exogenous they are generally non-zero, and FGLS is inconsistent. 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.

6 Conclusions

We showed that 1) OLS is, in general, inconsistent with non-exogenous regressors, while GLS is consistent; 2) 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) have MSE lower than OLS, often by a wide margin; b) perform well across the wide variety of

data-generating processes considered; c) have confidence intervals with exact coverage rates close to the nominal level with length much shorter than with OLS; 3) the FGLS estimate is $\sqrt{T}$-consistent and asymptotically normal, attaining the asymptotic distribution of the infeasible GLS under exogenous regressors, while with non-exogenous regressors the first-step estimation inflates its asymptotic variance. Inference should be based on the proposed variance estimator $\widehat{AVar}_R$, which is valid in both cases and involves no tuning parameters. Our results hold whether the regressors are exogenous or not, provided a) the regressors are pre-determined, and b) the stationary linear error process is invertible. We used the simple linear model as it is the leading case of interest. A similar treatment for models with endogenous regressors contemporaneously correlated with the innovations and estimated via some instrumental variable procedure is covered in Olivari and Perron (2024).

Our results suggest the following practical guidance. First, the FGLS procedure with k_T^* selected by the BIC, together with confidence intervals based on $\widehat{AVar}_R$, can be used as a default in linear time series regressions with stationary errors: under exogenous regressors the additional correction is asymptotically negligible, so nothing is lost, while under non-exogenous regressors it is required for reliable inference. Second, whether the regressors are pre-determined can be assessed using the variable addition test described in Remark 9; if pre-determinedness itself fails, lags of the relevant omitted variables should be included when observable, otherwise an instrumental variable procedure should be used. Third, when the errors may have a moving-average component with a large negative coefficient (e.g., multi-step forecast errors), the autoregressive approximation may require more lags than selected by the BIC; prior knowledge of the MA structure can be exploited via a tailored transformation such as the GMA procedure of Section 5. Fourth, with lagged dependent variables, OLS is inconsistent and the FGLS estimate should be constructed as described in Section 4.2. Fifth, in multi-step predictive regressions under rational expectations, both OLS and FGLS are consistent and FGLS is more efficient provided the implied MA process is invertible; if it is non-invertible, FGLS retains consistency only under exogenous regressors, and caution is warranted. Finally, whether GLS-type methods are preferable to OLS with HAC standard errors ultimately depends on the maintained assumptions, notably stationarity, invertibility and pre-determined regressors. When these conditions are plausible, as we believe is often the case, our results indicate that substantial gains in precision and in the reliability of inference are available.

References

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

Amemiya, T., 1986. Advanced Econometrics. Harvard University Press, Cambridge, MA.

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., Mora, A., 2025. On robust inference in time series regression. Econom. J. 28, 131–173.

Berk, K.N., 1974. Consistent autoregressive spectral estimates. Ann. Statist. 2, 489–502.

Brillinger, D.R., 2001. Time Series: Data Analysis and Theory. SIAM, Philadelphia.

Choudhury, A.H., Hubata, R., St. Louis, R.D., 1999. Understanding time-series regression estimators. Amer. Statist. 53, 342–348.

Cochrane, D., Orcutt, G.H., 1949. Application of least squares regression to relationships containing auto-correlated error terms. J. Amer. Statist. Assoc. 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.

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. J. Econometrics 20, 83–104.

Fang, Y., Koreisha, S.G., Shao, Q.m., 2023. Revisiting the use of generalized least squares in time series regression models. J. Data Sci. 22, 486–504.

González-Coya, E., Perron, P., 2024. Estimation in the presence of heteroskedasticity of unknown form: a lasso-based approach. J. Econom. Methods 13, 29–48.

González-Coya, E., Perron, P., 2025. Uncovering bias in uncovered interest parity tests. Forthcoming in Applied Economic Analysis.

Hannan, E.J., 1971. Non-linear time series regression. J. Appl. Probab. 8, 767–780.

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

Hannan, E.J., Kavalieris, L., 1986. Regression, autoregressive models. J. Time Ser. Anal. 7, 27–49.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L.P., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. J. Polit. Econ. 88, 829–853.

Judge, G.C., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.C., 1985. The Theory and Practice of Econometrics, 2nd Ed. 2nd ed., Wiley, New York.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Ng, S., Perron, P., 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. J. Amer. Statist. Assoc. 90, 268–281.

Ng, S., Perron, P., 2005. A note on the selection of time series models. Oxf. Bull. Econ. Stat. 67, 115–134.

Olivari, M., Perron, P., 2024. GLS-IV for time series regressions. Unpublished manuscript, Department of Economics, Boston University.

Perron, P., 2021. Conditions for OLS and GLS to be consistent in models with serially correlated errors. Unpublished manuscript, Department of Economics, Boston University.

Prais, S.J., Winsten, C.B., 1954. Trend estimators and serial correlation. Discussion Paper 383. Cowles Commission.

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

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Stock, J.H., Watson, M.W., 2019. Introduction to Econometrics, 4th Ed. 4th ed., Pearson, New York, NY.

Wallis, K.F., 1967. Lagged dependent variables and serially correlated errors: a reappraisal of three-pass least squares. Rev. Econ. Stat. 49, 555–567.

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

White, H., Domowitz, I., 1984. Nonlinear regression with dependent observations. Econometrica 52, 143–161.

Table 1: Root mean squared errors, bias and variance of estimators of β and ρ; $AR(1)$ model. Bias and Variance are the mean and the variance of the estimation error across replications, so that $\text{RMSE}^2 = \text{Bias}^2 + \text{Variance}$.

	β					ρ	
	OLS	Durbin	GLS	FGLS	CO-FGLS	FGLS	CO-FGLS
RMSE	0.399	0.036	0.025	0.025	0.041	0.034	0.174
Bias	0.395	0.002	0.000	0.001	0.031	-0.008	-0.171
Variance	0.0031	0.0013	0.0006	0.0006	0.0008	0.0011	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). Bias and Variance are the mean and the variance of $\hat{\beta} - \beta$ across replications, so that MSE = Bias^2+ Variance before the scaling by 100. FGLS confidence intervals and lengths are based on the exogeneity-robust variance estimator $\widehat{\text{AVar}}_R$. Med. k is the median autoregressive lag order selected by BIC across replications.

			MSE				Bias				Variance				Coverage			Length			
		AR(1)	OLS	GLS	CO	FGLS	OLS	GLS	CO	FGLS	OLS	GLS	CO	FGLS	OLS	CO	FGLS	OLS	CO	FGLS	Med. k
$T = 200$	$\gamma = 0$	-0.5	0.11	0.10	0.10	0.10	0.01	0.00	0.00	0.00	0.11	0.10	0.10	0.10	0.95	0.95	0.94	0.14	0.12	0.12	1
		0	0.20	0.20	0.20	0.20	0.00	0.00	0.00	0.01	0.20	0.20	0.20	0.20	0.94	0.94	0.95	0.17	0.17	0.17	0
		0.2	0.28	0.27	0.27	0.27	0.08	0.08	0.08	0.09	0.28	0.27	0.27	0.27	0.93	0.94	0.94	0.19	0.20	0.20	1
		0.5	0.58	0.42	0.43	0.43	-0.05	-0.04	-0.04	-0.03	0.58	0.42	0.43	0.43	0.91	0.94	0.94	0.27	0.25	0.25	1
		0.8	2.21	0.52	0.53	0.53	0.02	-0.02	-0.02	-0.03	2.21	0.52	0.53	0.53	0.89	0.94	0.94	0.49	0.28	0.28	1
	$\gamma = 0.25$	-0.5	0.25	0.09	0.09	0.09	-3.54	-0.32	-0.35	-0.32	0.12	0.09	0.09	0.09	0.83	0.95	0.95	0.13	0.12	0.12	1
		0	0.19	0.19	0.19	0.19	-0.40	-0.40	-0.37	-0.39	0.19	0.19	0.19	0.19	0.94	0.94	0.95	0.16	0.17	0.17	0
		0.2	0.28	0.25	0.26	0.27	1.67	-0.43	-0.22	0.00	0.25	0.25	0.26	0.27	0.91	0.94	0.93	0.19	0.19	0.19	1
		0.5	0.99	0.40	0.42	0.43	6.79	-0.34	0.43	-0.12	0.53	0.40	0.42	0.43	0.77	0.93	0.94	0.25	0.24	0.25	1
		0.8	5.48	0.48	0.55	0.51	18.52	-0.05	1.34	0.30	2.05	0.48	0.53	0.51	0.60	0.93	0.94	0.46	0.27	0.27	1
	$\gamma = 0.5$	-0.5	0.50	0.08	0.08	0.09	-6.03	-0.50	-0.61	-0.49	0.13	0.08	0.08	0.08	0.57	0.94	0.94	0.13	0.11	0.11	1
		0	0.17	0.17	0.18	0.17	-0.68	-0.68	-0.66	-0.67	0.16	0.16	0.17	0.17	0.93	0.94	0.94	0.15	0.15	0.15	0
		0.2	0.29	0.23	0.24	0.29	2.82	-0.74	-0.21	-0.06	0.21	0.22	0.24	0.29	0.86	0.93	0.91	0.17	0.18	0.19	1
		0.5	1.74	0.34	0.42	0.41	11.65	-0.50	1.95	-0.12	0.39	0.33	0.38	0.41	0.45	0.90	0.94	0.22	0.22	0.25	1
		0.8	11.26	0.41	0.81	0.49	31.28	0.07	4.76	0.72	1.47	0.41	0.58	0.48	0.18	0.83	0.94	0.39	0.25	0.26	1
$T = 500$	$\gamma = 0$	-0.5	0.04	0.04	0.04	0.04	0.02	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.95	0.95	0.95	0.08	0.07	0.07	1
		0	0.07	0.07	0.07	0.07	0.01	0.01	0.01	0.01	0.07	0.07	0.07	0.07	0.95	0.95	0.95	0.11	0.11	0.11	0
		0.2	0.10	0.10	0.10	0.10	-0.01	-0.01	-0.01	-0.01	0.10	0.10	0.10	0.10	0.94	0.95	0.95	0.12	0.12	0.12	1
		0.5	0.23	0.16	0.17	0.17	-0.04	-0.07	-0.08	-0.08	0.23	0.16	0.17	0.17	0.92	0.95	0.95	0.18	0.16	0.16	1
		0.8	0.91	0.20	0.20	0.20	-0.05	-0.06	-0.06	-0.06	0.91	0.20	0.20	0.20	0.91	0.95	0.95	0.34	0.18	0.18	1
	$\gamma = 0.25$	-0.5	0.15	0.03	0.04	0.04	-3.27	-0.15	-0.17	-0.15	0.05	0.03	0.03	0.03	0.67	0.95	0.95	0.08	0.07	0.07	1
		0	0.07	0.07	0.07	0.07	-0.16	-0.16	-0.15	-0.16	0.07	0.07	0.07	0.07	0.95	0.95	0.95	0.10	0.10	0.10	0
		0.2	0.13	0.10	0.10	0.10	1.89	-0.15	-0.04	-0.11	0.10	0.10	0.10	0.10	0.89	0.94	0.94	0.12	0.12	0.12	1
		0.5	0.69	0.15	0.17	0.17	6.98	-0.11	0.38	-0.03	0.21	0.15	0.16	0.16	0.61	0.93	0.95	0.17	0.15	0.16	1
		0.8	4.32	0.19	0.21	0.20	18.73	0.00	0.81	0.15	0.81	0.19	0.20	0.20	0.38	0.94	0.95	0.32	0.17	0.17	1
	$\gamma = 0.5$	-0.5	0.35	0.03	0.03	0.03	-5.50	-0.20	-0.30	-0.21	0.05	0.03	0.03	0.03	0.24	0.95	0.95	0.08	0.07	0.07	1
		0	0.06	0.06	0.06	0.06	-0.28	-0.28	-0.27	-0.28	0.06	0.06	0.06	0.06	0.95	0.94	0.95	0.09	0.10	0.10	0
		0.2	0.18	0.08	0.09	0.10	3.22	-0.27	0.15	-0.22	0.08	0.08	0.09	0.10	0.76	0.93	0.94	0.11	0.11	0.12	1
		0.5	1.56	0.13	0.19	0.16	11.87	-0.19	1.85	-0.07	0.15	0.13	0.15	0.16	0.13	0.87	0.94	0.15	0.14	0.16	1
		0.8	10.67	0.16	0.35	0.19	31.75	-0.05	3.68	0.23	0.59	0.16	0.22	0.19	0.01	0.81	0.94	0.27	0.16	0.16	1

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). Bias and Variance are the mean and the variance of $\hat{\beta} - \beta$ across replications, so that MSE = Bias^2+ Variance before the scaling by 100. FGLS confidence intervals and lengths are based on the exogeneity-robust variance estimator $\widehat{\text{AVar}}_R$. Med. k is the median autoregressive lag order selected by BIC across replications.

			MSE			Bias			Variance			Coverage		Length		
		AR(2)	OLS	GLS	FGLS	OLS	GLS	FGLS	OLS	GLS	FGLS	OLS	FGLS	OLS	FGLS	Med. k
$T = 200$	$\gamma = 0$	0.5,-0.3	0.33	0.28	0.29	0.00	-0.02	-0.02	0.33	0.28	0.29	0.94	0.94	0.22	0.20	2
		-0.5,0.3	0.17	0.13	0.13	0.04	0.04	0.03	0.17	0.13	0.13	0.94	0.94	0.15	0.14	2
		1.34,-0.42	11.50	0.40	0.41	0.40	0.01	0.00	11.50	0.40	0.41	0.87	0.95	1.08	0.25	2
		0,0.3	0.31	0.28	0.29	-0.04	-0.05	-0.06	0.31	0.28	0.29	0.88	0.93	0.18	0.20	2
		0.5,0.3	1.84	0.46	0.48	-0.13	-0.07	-0.06	1.84	0.46	0.48	0.86	0.94	0.43	0.26	2
	$\gamma = 0.25$	0.5,-0.3	0.36	0.27	0.29	2.52	-0.47	-0.40	0.29	0.27	0.29	0.91	0.94	0.21	0.20	2
		-0.5,0.3	0.22	0.12	0.13	-2.30	-0.29	-0.19	0.17	0.12	0.13	0.90	0.93	0.15	0.14	2
		1.34,-0.42	27.76	0.37	0.39	41.97	0.05	0.18	10.15	0.37	0.39	0.59	0.94	1.02	0.24	2
		0,0.3	0.33	0.25	0.28	2.16	-0.29	0.06	0.28	0.25	0.28	0.85	0.93	0.17	0.20	2
		0.5,0.3	3.91	0.44	0.49	14.87	-0.15	0.04	1.70	0.44	0.49	0.63	0.94	0.40	0.26	2
	$\gamma = 0.5$	0.5,-0.3	0.41	0.24	0.27	4.27	-0.74	-0.65	0.23	0.23	0.27	0.85	0.94	0.19	0.19	2
		-0.5,0.3	0.32	0.11	0.13	-3.89	-0.47	-0.32	0.17	0.10	0.13	0.83	0.93	0.14	0.13	2
		1.34,-0.42	58.86	0.32	0.37	71.73	0.16	0.35	7.42	0.32	0.36	0.16	0.94	0.87	0.23	2
		0,0.3	0.36	0.22	0.29	3.63	-0.56	0.02	0.23	0.21	0.29	0.78	0.92	0.16	0.19	2
		0.5,0.3	7.55	0.37	0.50	25.16	-0.14	0.16	1.22	0.37	0.50	0.24	0.93	0.34	0.26	2
$T = 500$	$\gamma = 0$	0.5,-0.3	0.12	0.11	0.11	-0.06	-0.06	-0.06	0.12	0.11	0.11	0.95	0.95	0.14	0.13	2
		-0.5,0.3	0.06	0.05	0.05	0.03	0.02	0.02	0.06	0.05	0.05	0.94	0.95	0.10	0.09	2
		1.34,-0.42	4.71	0.16	0.16	0.38	0.07	0.08	4.71	0.16	0.16	0.91	0.95	0.76	0.15	2
		0,0.3	0.12	0.10	0.10	-0.02	-0.02	-0.02	0.12	0.10	0.10	0.90	0.95	0.11	0.13	2
		0.5,0.3	0.75	0.19	0.19	0.09	0.04	0.04	0.75	0.19	0.19	0.91	0.95	0.30	0.17	2
	$\gamma = 0.25$	0.5,-0.3	0.18	0.10	0.10	2.66	-0.20	-0.17	0.11	0.10	0.10	0.87	0.95	0.13	0.13	2
		-0.5,0.3	0.10	0.05	0.05	-2.05	-0.14	-0.11	0.06	0.05	0.05	0.86	0.95	0.09	0.09	2
		1.34,-0.42	22.58	0.15	0.15	42.77	0.14	0.16	4.29	0.15	0.15	0.37	0.94	0.72	0.15	2
		0,0.3	0.17	0.10	0.11	2.38	-0.10	-0.03	0.11	0.10	0.11	0.80	0.94	0.11	0.13	2
		0.5,0.3	2.96	0.17	0.18	15.01	-0.06	0.03	0.70	0.17	0.18	0.46	0.95	0.28	0.17	2
	$\gamma = 0.5$	0.5,-0.3	0.29	0.09	0.10	4.55	-0.30	-0.27	0.08	0.09	0.10	0.67	0.94	0.12	0.12	2
		-0.5,0.3	0.18	0.04	0.05	-3.44	-0.20	-0.17	0.06	0.04	0.05	0.69	0.95	0.09	0.08	2
		1.34,-0.42	56.58	0.12	0.14	73.15	0.14	0.20	3.08	0.12	0.14	0.01	0.95	0.61	0.15	2
		0,0.3	0.25	0.08	0.10	4.02	-0.20	-0.09	0.09	0.08	0.10	0.61	0.94	0.10	0.12	2
		0.5,0.3	7.10	0.14	0.17	25.69	0.03	0.22	0.49	0.14	0.17	0.03	0.95	0.24	0.16	2

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). Bias and Variance are the mean and the variance of $\hat{\beta} - \beta$ across replications, so that MSE = Bias^2+ Variance before the scaling by 100. FGLS confidence intervals and lengths are based on the exogeneity-robust variance estimator $\widehat{\text{AVar}}_R$. Med. k is the median autoregressive lag order selected by BIC across replications.

			MSE				Bias				Variance				Coverage			Length			
		MA(1)	OLS	GLS	GMA	FGLS	OLS	GLS	GMA	FGLS	OLS	GLS	GMA	FGLS	OLS	GMA	FGLS	OLS	GMA	FGLS	Med. k
$T = 200$	$\gamma = 0$	-0.7	0.08	0.03	0.03	0.03	0.03	0.01	0.01	0.00	0.08	0.03	0.03	0.03	0.98	0.96	0.96	0.13	0.08	0.08	3
		-0.4	0.11	0.09	0.09	0.09	-0.03	-0.04	-0.04	-0.04	0.11	0.09	0.09	0.09	0.96	0.95	0.95	0.14	0.12	0.12	2
		0.5	0.40	0.34	0.34	0.35	-0.02	-0.02	-0.03	-0.03	0.40	0.34	0.34	0.35	0.92	0.95	0.93	0.23	0.22	0.22	2
	$\gamma = 0.25$	-0.7	0.54	0.03	0.04	0.05	-6.62	-0.28	-0.70	-1.06	0.10	0.03	0.03	0.04	0.53	0.95	0.94	0.13	0.08	0.08	3
		-0.4	0.27	0.08	0.09	0.10	-3.95	-0.34	-0.50	-0.90	0.11	0.08	0.09	0.09	0.83	0.94	0.95	0.14	0.11	0.12	2
		0.5	0.52	0.31	0.33	0.35	4.03	-0.35	-0.35	0.21	0.36	0.31	0.32	0.34	0.86	0.94	0.94	0.22	0.22	0.22	2
	$\gamma = 0.5$	-0.7	1.36	0.03	0.06	0.08	-11.10	-0.41	-1.28	-1.86	0.13	0.03	0.04	0.05	0.04	0.91	0.87	0.13	0.07	0.08	3
		-0.4	0.58	0.07	0.09	0.12	-6.72	-0.57	-0.90	-1.61	0.12	0.07	0.08	0.09	0.50	0.94	0.93	0.13	0.11	0.11	1
		0.5	0.73	0.28	0.30	0.33	6.78	-0.67	-0.66	0.26	0.27	0.27	0.30	0.33	0.70	0.93	0.93	0.20	0.20	0.22	2
$T = 500$	$\gamma = 0$	-0.7	0.03	0.01	0.01	0.01	0.01	-0.01	-0.01	-0.01	0.03	0.01	0.01	0.01	0.98	0.96	0.97	0.08	0.04	0.05	4
		-0.4	0.04	0.03	0.03	0.03	0.01	0.01	0.01	0.01	0.04	0.03	0.03	0.03	0.96	0.94	0.96	0.08	0.07	0.07	2
		0.5	0.15	0.13	0.13	0.13	0.01	0.01	0.01	0.01	0.15	0.13	0.13	0.13	0.94	0.95	0.94	0.15	0.14	0.14	2
	$\gamma = 0.25$	-0.7	0.42	0.01	0.01	0.02	-6.17	-0.07	-0.27	-0.57	0.04	0.01	0.01	0.01	0.09	0.95	0.94	0.08	0.04	0.05	4
		-0.4	0.17	0.03	0.03	0.04	-3.59	-0.12	-0.19	-0.51	0.04	0.03	0.03	0.03	0.62	0.95	0.95	0.08	0.07	0.07	2
		0.5	0.31	0.12	0.12	0.13	4.11	-0.20	-0.19	0.18	0.14	0.12	0.12	0.13	0.78	0.94	0.94	0.14	0.14	0.14	2
	$\gamma = 0.5$	-0.7	1.15	0.01	0.02	0.03	-10.49	-0.14	-0.56	-1.10	0.05	0.01	0.01	0.01	0.00	0.91	0.85	0.08	0.04	0.04	4
		-0.4	0.41	0.03	0.03	0.04	-6.06	-0.19	-0.32	-0.90	0.05	0.03	0.03	0.03	0.15	0.93	0.92	0.08	0.06	0.07	2
		0.5	0.60	0.10	0.11	0.13	7.06	-0.24	-0.25	0.38	0.10	0.10	0.11	0.13	0.40	0.94	0.94	0.12	0.13	0.14	2

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). Bias and Variance are the mean and the variance of $\hat{\beta} - \beta$ across replications, so that MSE = Bias^2+ Variance before the scaling by 100. FGLS confidence intervals and lengths are based on the exogeneity-robust variance estimator $\widehat{\text{AVar}}_R$. Med. k is the median autoregressive lag order selected by BIC across replications.

			MSE			Bias			Variance			Coverage		Length		
		ARMA(1,1)	OLS	GLS	FGLS	OLS	GLS	FGLS	OLS	GLS	FGLS	OLS	FGLS	OLS	FGLS	Med. k
$T = 200$	$\gamma = 0$	-0.5,-0.4	0.10	0.04	0.04	0.05	0.02	0.02	0.10	0.04	0.04	0.97	0.96	0.13	0.08	2
		0.2,-0.4	0.14	0.13	0.13	0.01	0.00	0.00	0.14	0.13	0.13	0.96	0.96	0.15	0.15	1
		0.2,0.5	0.57	0.39	0.42	-0.03	-0.03	-0.04	0.57	0.39	0.42	0.92	0.94	0.28	0.24	2
		0.5,-0.4	0.25	0.25	0.25	0.06	0.06	0.07	0.25	0.25	0.25	0.92	0.92	0.18	0.18	0
		0.5,0.5	1.27	0.43	0.47	0.13	0.02	0.01	1.27	0.43	0.47	0.90	0.94	0.39	0.26	3
		0.8,-0.4	0.88	0.43	0.47	0.20	0.13	0.11	0.88	0.43	0.47	0.86	0.93	0.30	0.25	2
		0.8,0.5	4.99	0.39	0.42	-0.04	-0.05	-0.05	4.99	0.39	0.42	0.88	0.94	0.73	0.25	3
	$\gamma = 0.25$	-0.5,-0.4	0.48	0.04	0.05	-6.06	-0.21	-0.55	0.11	0.04	0.04	0.58	0.95	0.13	0.08	2
		0.2,-0.4	0.20	0.12	0.15	-2.49	-0.37	-1.11	0.14	0.12	0.13	0.91	0.95	0.15	0.15	1
		0.2,0.5	0.99	0.37	0.42	6.92	-0.22	0.65	0.52	0.37	0.41	0.79	0.93	0.26	0.24	2
		0.5,-0.4	0.24	0.23	0.25	1.07	-0.38	0.52	0.23	0.23	0.24	0.91	0.92	0.17	0.17	0
		0.5,0.5	3.22	0.41	0.49	14.40	0.07	1.35	1.15	0.41	0.47	0.64	0.93	0.37	0.26	3
		0.8,-0.4	1.60	0.40	0.48	9.19	-0.15	-0.06	0.76	0.40	0.48	0.68	0.93	0.28	0.25	2
		0.8,0.5	13.57	0.37	0.46	30.14	0.07	1.47	4.49	0.37	0.44	0.57	0.93	0.69	0.25	3
	$\gamma = 0.5$	-0.5,-0.4	1.21	0.04	0.05	-10.33	-0.37	-1.03	0.15	0.03	0.04	0.08	0.92	0.13	0.08	2
		0.2,-0.4	0.32	0.11	0.18	-4.28	-0.67	-2.02	0.13	0.10	0.14	0.80	0.91	0.14	0.14	1
		0.2,0.5	1.76	0.31	0.40	11.75	-0.44	1.08	0.38	0.31	0.39	0.47	0.93	0.23	0.23	2
		0.5,-0.4	0.22	0.20	0.24	1.81	-0.64	0.86	0.19	0.20	0.23	0.89	0.90	0.16	0.16	0
		0.5,0.5	6.54	0.35	0.53	23.87	-0.02	2.28	0.84	0.35	0.48	0.21	0.91	0.32	0.25	2
		0.8,-0.4	2.93	0.35	0.50	15.31	-0.35	-0.37	0.59	0.35	0.50	0.34	0.92	0.24	0.25	2
		0.8,0.5	29.72	0.31	0.53	51.49	0.29	2.98	3.21	0.31	0.45	0.12	0.91	0.59	0.24	3
$T = 500$	$\gamma = 0$	-0.5,-0.4	0.03	0.01	0.01	0.00	0.00	0.00	0.03	0.01	0.01	0.97	0.96	0.08	0.05	3
		0.2,-0.4	0.05	0.05	0.05	-0.06	-0.06	-0.06	0.05	0.05	0.05	0.96	0.96	0.09	0.09	1
		0.2,0.5	0.22	0.15	0.16	0.03	0.02	0.01	0.22	0.15	0.16	0.93	0.94	0.18	0.15	3
		0.5,-0.4	0.10	0.09	0.10	0.08	0.07	0.08	0.10	0.09	0.10	0.93	0.93	0.11	0.11	1
		0.5,0.5	0.49	0.17	0.18	0.02	0.03	0.03	0.49	0.17	0.18	0.93	0.94	0.26	0.16	3
		0.8,-0.4	0.35	0.17	0.18	-0.01	0.03	0.03	0.35	0.17	0.18	0.90	0.94	0.20	0.16	2
		0.8,0.5	2.02	0.15	0.16	0.04	0.02	0.02	2.02	0.15	0.16	0.91	0.95	0.51	0.15	3
	$\gamma = 0.25$	-0.5,-0.4	0.37	0.01	0.02	-5.72	-0.09	-0.36	0.04	0.01	0.02	0.15	0.95	0.08	0.05	3
		0.2,-0.4	0.10	0.05	0.05	-2.22	-0.15	-0.61	0.05	0.05	0.05	0.86	0.94	0.09	0.09	1
		0.2,0.5	0.69	0.14	0.16	6.99	-0.11	0.38	0.20	0.14	0.16	0.62	0.94	0.17	0.15	3
		0.5,-0.4	0.11	0.09	0.10	1.28	-0.14	0.53	0.09	0.09	0.10	0.90	0.92	0.11	0.11	0
		0.5,0.5	2.45	0.16	0.19	14.20	0.02	0.76	0.44	0.16	0.18	0.38	0.94	0.25	0.16	3
		0.8,-0.4	1.17	0.16	0.18	9.29	-0.05	-0.20	0.31	0.16	0.18	0.51	0.94	0.19	0.16	2
		0.8,0.5	11.14	0.14	0.16	30.64	0.07	0.78	1.76	0.14	0.15	0.31	0.94	0.48	0.15	3
	$\gamma = 0.5$	-0.5,-0.4	1.00	0.01	0.02	-9.74	-0.15	-0.67	0.05	0.01	0.02	0.00	0.91	0.08	0.05	2
		0.2,-0.4	0.19	0.04	0.06	-3.73	-0.22	-1.07	0.05	0.04	0.05	0.63	0.92	0.09	0.08	1
		0.2,0.5	1.55	0.12	0.16	11.84	-0.20	0.71	0.15	0.12	0.16	0.13	0.93	0.15	0.15	3
		0.5,-0.4	0.12	0.08	0.10	2.14	-0.28	0.92	0.07	0.08	0.10	0.85	0.88	0.10	0.10	0
		0.5,0.5	6.09	0.14	0.20	24.00	-0.01	1.32	0.33	0.14	0.19	0.01	0.92	0.21	0.16	3
		0.8,-0.4	2.71	0.14	0.18	15.72	-0.17	-0.49	0.24	0.14	0.18	0.07	0.94	0.16	0.16	2
		0.8,0.5	28.14	0.12	0.20	51.78	0.18	1.64	1.33	0.12	0.17	0.01	0.92	0.41	0.15	3