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

In Section S-1, we present detailed theoretical and simulation evidence about the claims made in Remark 6. Additional simulation results are reported in Section S-2 that complement those in Section 5, while Section S-3 presents simulation results for predictive regressions. Section S-4 discusses our suggested method to further correct for possible heteroskedasticity in the errors. The method is presented as well as simulations showing that further reductions in MSE can be achieved. Section S-5 contains the preliminary lemmas and the proofs of the results stated in Section 4.1 of the paper.

S-1 The Robustness of GLS

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

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

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

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

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

The result in the previous Theorem, proved in Section S-1.1, is useful but opaque as far as obtaining useful insights given the level of generality. The following corollary considers the case with *i.i.d.* regressors. While restrictive, the results allow important insights that still apply with a serially correlated regressor.

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

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

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

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

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

S-1.1 Proof of some results

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

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

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

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

Hence, the GLS estimator can be written as

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

The variance of the GLS estimator is

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

The OLS estimator can be written as

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

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

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

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

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

and thus

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

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

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

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

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

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

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

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

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

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

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

and the ratio of interest is

$$\lim_{T\to\infty} \left(\frac{MSE(\hat{\beta}_{\text{GLS}})}{MSE(\hat{\beta}_{\text{OLS}})} \right) = \frac{\lim_{T\to\infty} TVar(\hat{\beta}_{\text{GLS}})}{\lim_{T\to\infty} TVar(\hat{\beta}_{\text{OLS}})}$$

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

and thus,

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

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

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

and

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

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

Hence,

$$\begin{aligned} &\lim_{T\to\infty} \left(MSE(\hat{\beta}_{\text{GLS}}) / MSE(\hat{\beta}_{\text{OLS}}) \right) < 1 \\ &\text{iff } \rho^2 - 2\rho(1+\rho^2) cor_u(1) + \rho^2 cor_u(2) < 0 \\ &\text{iff } \frac{\rho}{2(1+\rho^2)} (1 + cor_u(2)) < cor_u(1) \quad \text{when } \rho > 0, \\ &\text{iff } \frac{\rho}{2(1+\rho^2)} (1 + cor_u(2)) > cor_u(1) \quad \text{when } \rho < 0. \square \end{aligned}$$

S-1.2 Simulations

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

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

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

It is most instructive to start with the $AR(1)$ case. When $\rho_u = 0$, as expected OLS is best and GLS has higher MSE. When $\rho_u = -0.5$, GLS has lower MSE for all negative values of ρ and, vice versa, when $\rho_u = 0.5, 0.8$, GLS has lower MSE for all positive values of ρ. When $\rho_u = 0.2$, a small value, things are more complex. Here, GLS is best when $\rho \in (0.1, 0.4)$

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

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

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

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

Remark 2. *After the first draft of this paper was completed, we became aware of the work by Koreisha and Fang (2001). They present exact bounds for the relative variance of OLS, GLS and Feasible GLS allowing for misspecification of the process generating the errors when constructing the FGLS estimate. The results depend on the covariance matrix of the errors, the exact nature of the GLS structure used and the method to construct the FGLS estimate,*

the regressors and the sample size. The bounds are, however, not informative and quite complex. Accordingly they resort to simulation experiments using approximate autoregressive processes of order 1, 7 and 14 when $T = 200$ to construct the FGLS estimate. In the paper, they report results for few selected cases, which do not allow addressing several of the issues discussed above, e.g., the effect of the sign of the quasi-difference parameter, the strength of the correlation in the errors. They wrongly conclude that GLS (constructed using an AR misspecification) is always better than OLS. As shown above this is not the case.

S-2 Additional simulations related to Section 5

Tables S.4-S.7 present simulations results related to those presented in Section 5. The setup is exactly the same, except that we set $\rho_x = 0$, instead of $\rho_x = 0.8$. The goal is simply to show robustness of the results. They are indeed qualitatively similar. As in the paper, the FGLS confidence intervals are based on the exogeneity-robust variance estimator $\widehat{AVar}_R$ of Theorem 2 in the paper, and the median autoregressive order selected by the BIC is reported in the last column of each table. The coverage rates are close to the nominal level, with the same qualifications as discussed in Section 5 of the paper.

S-3 Simulations with predictive regressions

As discussed in Remarks 1 and 5, in the case of predictive regressions assuming rational expectations, both OLS and GLS are consistent. We present the results of a small simulation experiment to show that, with exogenous or non-exogenous regressors, FGLS delivers substantially lower MSE and shorter confidence intervals than OLS, when the MA process is invertible. The setup adopted corresponds to regression

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

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

The results are presented in Table S.8. With $\gamma = 0$, the results are similar to those in Table 4. FGLS and GMA have much lower MSE than OLS and are nearly as efficient as the infeasible GLS, especially when $T = 500$. The coverage rates for all methods are near the nominal 95% level. Again, the length of the confidence intervals are shorter with FGLS and GMA compared to OLS. With non-exogenous regressors, the results are broadly similar, with one qualification: the coverage rates for FGLS show some under-coverage when

$\gamma = 0.5$, most notably when $\theta = -0.7$ (0.87 when $T = 200$ and 0.84 when $T = 500$), for the reason discussed in Section 5 of the paper, namely that the implied autoregressive coefficients decay slowly while the BIC selects a small order, so that the truncation of the autoregressive approximation induces a bias. The coverage rates for GMA, which is tailored to the MA structure, remain close to the nominal level. This is in line with our theoretical results.

S-4 Correcting for heteroskedasticity

In this section, we now consider a FGLS procedure for heteroskedasticity in the errors e_t. We describe the method suggested by González-Coya and Perron (2024) based on an Adaptive Lasso procedure to fit the skedastic function. Lasso is a non-parametric estimation method first proposed by Tibshirani (1996). It selects regressors amongst a potentially large set w_{tj} $(j = 1, ..., d)$, where d can be very large, by imposing a ℓ_1 penalty on their size. Lasso forces the coefficients to be equally penalized. We can, however, assign different weights to different coefficients. If the weights are data-dependent and properly chosen, this can enhance the properties of Lasso, in particular when the irrelevant covariates are highly correlated with the relevant ones. To that effect, Zou (2006) considered the adaptive Lasso given by

$$\hat{\phi}^{\mathrm{A}} = \arg\min_{\phi}\{(1/2)_{t=1}^{T}(\log(v_t^2) - \phi_0 - {}_{j=1}^{d}\, w_{tj}\phi_j)^2 + \lambda_{j=1}^{d}\hat{\vartheta}_j\,|\phi_j|\}, \tag{A.2}$$

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

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

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

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

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

$k = 1, \ldots, K$ and combine the K estimates of the prediction error. Denote by $\hat{f}_i^{-k}(w)$ the fitted function, computed with the kth part of the data removed and using ψ_i^c. Then the cross-validation estimate of the prediction error is

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

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

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

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

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

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

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

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

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

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

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

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

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

S-4.1 Simulation results with heteroskedasticity

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

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

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

The results are reported in Table S.9; the first panel for $w_t = x_t$ and the second for $w_t = \phi x_t + (1-\phi)q_t$. We present the MSE, bias and variance of the FGLS estimate as well as the coverage rates and lengths of the confidence intervals obtained using the method discussed

in the previous section. We present results for the OLS estimate, the FGLS estimate that accounts only for serial correlation (F-C) and the FGLS estimate that accounts for both serial correlation and heteroskedasticity (F-CH). This is done to gauge the extent of the improvement provided by the correction for heteroskedasticity. Note that when using F-C, we construct the confidence intervals that correct for serial correlation the same way as we do for F-CH, i.e., applying the same correction for potential remaining heteroskedasticity.

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

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

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

S-5 Preliminary lemmas and proofs related to Section 4.1[1]

This section states and proves the preliminary lemmas underlying the results of Section 4.1 of the paper, and then proves Proposition 1, Theorems 1 and 2, and Corollary 1. Throughout this section, references to Assumptions 1–5, Proposition 1, Theorems 1 and 2, Corollary 1 and to equations numbered without a prefix are to the paper; lemmas and equations numbered A.x are those of this Supplement. The notation is that of Section 4.1; in addition, c and C denote generic positive constants, not necessarily the same at each occurrence, $\theta_k = _{j>k} |\rho_j|$ denotes the tail of the autoregressive coefficients, and $\Delta = \hat{\rho}^D - \rho^{(k)}$. All stochastic-order bounds are derived from moment bounds that are uniform over $k \leq k_T^{\max}$; the paragraph following Lemma A.2 explains how data-driven orders are accommodated.

Lemma A.1. *(CLT for infeasible GLS.) Under Assumptions 1-3,* $\sqrt{T}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, \sigma_e^2 Q^{-1})$.

Proof. Write $\hat{\beta}_{GLS} - \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u = [(DX)'(DX)]^{-1}(DX)'(Du)$.

Step 1: the whitening rows. The t-th row of D is $(\sigma_e/\sigma_{(t)})(-\phi_{t-1,t-1}, ..., -\phi_{t-1,1}, 1)$, where $\{\phi_{t-1,i}\}_{i=1}^{t-1}$ are the coefficients of the best linear predictor of u_t given $(u_{t-1}, ..., u_1)$ and $\sigma_{(t)}^2$ is the corresponding prediction variance; this is what makes $Var(Du) = \sigma_e^2 D\Omega D' = \sigma_e^2 I$. Two approximations connect these rows to the infinite-order weights $a_i = -\rho_i$ (with $a_0 = 1$): (i) truncation of the infinite past, with t-th error bounded in L^1 by a multiple of $\tau_t = _{i \geq t} |a_i|$; and (ii) replacement of the finite-predictor coefficients by $a_1, ..., a_{t-1}$, with L^1 error bounded, for t beyond some fixed t_0, by a multiple of $\sum_{i=1}^{t-1} |\phi_{t-1,i} - (-a_i)| \leq c\,\tau_t$ by Baxter's inequality (Baxter (1962)); the same inequality yields $\sigma_{(t)}^2 - \sigma_e^2 = O(\tau_t^2)$, hence $|\sigma_e/\sigma_{(t)} - 1| = O(\tau_t^2)$. (For a true $AR(p)$ process, $\phi_{t-1,i} = -a_i$ exactly once $t > p$, and $\tau_t = 0$ for $t > p$.) Consequently,

$$(Du)_t = e_t + d_t^u, \qquad (DX)_t' = \widetilde{x}_t + d_t^x, \qquad E|d_t^u| + E\|d_t^x\| \leq c\,\tau_t,$$

and $_t\tau_t = _i\, i|a_i| < \infty$ by Assumption 2. Note that the approximation errors d_t are not confined to finitely many t; it is their cumulated size that is bounded. By the Cauchy-Schwarz and Markov inequalities, the discrepancies satisfy

$$\|T^{-1/2}(DX)'(Du) - T_t^{-1/2}\widetilde{x}_t e_t\| = O_p(T_t^{-1/2}\tau_t) = O_p(T^{-1/2}),$$

and similarly $T^{-1}(DX)'(DX) = T_t^{-1}\widetilde{x}_t\widetilde{x}_t' + O_p(T^{-1})$. Since $\{(w_t', e_t)'\}$ is jointly stationary and ergodic (Assumption 3) and $\widetilde{x}_t$ is a measurable function of $(x_t, x_{t-1}, ...)$ with $E\|\widetilde{x}_t\|^2 < \infty$, the ergodic theorem gives $T_t^{-1}\widetilde{x}_t\widetilde{x}_t' \xrightarrow{p} Q$.

[1]The authors gratefully acknowledge the use of a generative AI assistant (Anthropic Claude Fable 5) in a supporting role during the preparation of this work. The tool was used to help replicate and independently re-corroborate the numerical simulation results, and to assist with the formalization and checking of the proofs in this section. It did not write any part of the main text, and no theoretical or empirical results were derived directly from it; all proofs, derivations, and simulation outputs were formulated and verified independently by the authors, who reviewed all AI-assisted output and take full responsibility for the content.

Step 2: martingale CLT. The filtered score $\widetilde{x}_t = _{i\geq 0}\ a_i x_{t-i}$ is $\mathcal{G}_{t-1}$-measurable (this is where the enlarged information set is essential, as $\widetilde{x}_t$ depends on the contemporaneous x_t and is not Φ_{t-1}-measurable), and (22) gives $E(e_t|\mathcal{G}_{t-1}) = 0$. Hence $\{\widetilde{x}_t e_t, \mathcal{G}_t\}$ is a stationary, ergodic, square-integrable martingale difference sequence with $E[\widetilde{x}_t e_t] = _{i\geq 0}\ a_i E[x_{t-i} e_t] = 0$ (the interchange justified by dominated convergence and the orthogonality conditions of Assumption 3) and conditional second moment $E[(\widetilde{x}_t e_t)(\widetilde{x}_t e_t)'|\mathcal{G}_{t-1}] = \widetilde{x}_t \widetilde{x}_t' E(e_t^2|\mathcal{G}_{t-1}) = \sigma_e^2 \widetilde{x}_t \widetilde{x}_t'$, whose mean is $\sigma_e^2 Q$. The martingale CLT for stationary ergodic martingale difference sequences (Billingsley (1968), Theorem 23.1) gives $T_t^{-1/2} \widetilde{x}_t e_t \xrightarrow{d} N(0, \sigma_e^2 Q)$, and the result follows by Slutsky's theorem. Note that only pre-determinedness is used, not exogeneity. □

Lemma A.2. *(Order selection.) Let k_T^* be the BIC-selected order over $k \in [0, k_T^{\max}]$ with $k_T^{\max} = \lfloor (T/\ln T)^{1/3} \rfloor$, applied to the Durbin regression (14) with a common effective sample (Ng and Perron (2005)). Under Assumptions 1-3 and 5: (i) if $\rho_j = 0$ for $j > p$, then $P(k_T^* = p) \to 1$, and Assumption 4 holds along k_T^* with probability approaching one; (ii) if $|\rho_j| \leq c_\rho \lambda^j$ for some $\lambda \in (0,1)$ (e.g., invertible ARMA processes), then $k_T^* = O_p(\ln T)$. However, the BIC alone does not guarantee the tail condition in (23): since the BIC balances the squared bias against the penalty, generically $\sqrt{T}_{j>k_T^*}|\rho_j| \asymp_p \sqrt{k_T^* \ln T} \to \infty$. The adjusted order $\hat{k}_T = \max\{k_T^*, \lceil (\ln T)^{3/2} \rceil\}$ satisfies Assumption 4 with probability approaching one.*

Proof. (i) Write the BIC objective as $\ln \hat{\sigma}_{e,k}^2 + k \ln(\bar{T})/\bar{T}$, $\bar{T} = T - k_T^{\max}$, where $\hat{\sigma}_{e,k}^2$ is the residual variance of the Durbin regression with k lags. Let σ_k^2 denote the population residual variance of that regression; for $k \geq p$, $\sigma_k^2 = \sigma_e^2$, while $\sigma_k^2 > \sigma_e^2$ for $k < p$ (under-fitting inflates $\ln \hat{\sigma}_{e,k}^2$ by a positive constant in the limit). Over-fitting is ruled out because for each $k > p$ the difference $\ln \hat{\sigma}_{e,p}^2 - \ln \hat{\sigma}_{e,k}^2$ is $O_p((k-p)/T)$—a χ^2-type improvement per extra (asymptotically irrelevant) regressor—while the penalty increases by $(k-p)\ln(\bar{T})/\bar{T}$; standard exponential tail bounds give $P(\exists\, k \in (p, k_T^{\max}] : BIC(k) < BIC(p)) \to 0$; see Hannan and Deistler (2012), ch. 5, for the classical argument. Hence $P(k_T^* = p) \to 1$; along $k_T^* = p$ the tail in (23) is exactly zero and $p^3/T \to 0$ trivially.

(ii) With $|\rho_j| \leq c_\rho \lambda^j$, the bias term satisfies $\sigma_k^2 - \sigma_e^2 = O(\lambda^{2k})$, using the Baxter-type upper bound $O((_{j>k}|\rho_j|)^2)$, and the population trade-off between this term and the penalty $k \ln T/T$ is minimised at $k \asymp \ln T$, whence $k_T^* = O_p(\ln T)$; cf. Hannan and Deistler (2012), ch. 6.6. For the negative claim, note that at the balancing order the squared tail is of the same magnitude as the penalty, $(_{j>k^*}|\rho_j|)^2 \asymp k^* \ln T/T$, so $\sqrt{T}_{j>k^*}|\rho_j| \asymp \sqrt{k^* \ln T}$ diverges; this is the same phenomenon documented by Ng and Perron (1995) for information-criterion lag selection in unit-root autoregressions. For the adjusted order, eventually $\hat{k}_T = \lceil (\ln T)^{3/2} \rceil \geq k_T^*$ with probability approaching one, and then $\hat{k}_T^3/T = (\ln T)^{9/2}/T \to 0$ while

$$\hat{k}_T \sqrt{T}_{j>\hat{k}_T}|\rho_j| \leq c\,(\ln T)^{3/2} \sqrt{T} \lambda^{(\ln T)^{3/2}} = c\,(\ln T)^{3/2} T^{1/2 - |\ln \lambda|(\ln T)^{1/2}} \to 0.$$

□

Data-driven orders. All stochastic-order bounds in Lemmas A.3 and A.4 and in the proofs below are derived from moment bounds that are uniform over $k \leq k_T^{\max}$. Consequently, for any (possibly data-driven) order $\hat{k}_T$ for which there exist deterministic sequences $k_T^- \leq k_T^+ \leq k_T^{\max}$ with $P(k_T^- \leq \hat{k}_T \leq k_T^+) \to 1$ and such that Assumption 4 holds uniformly over $[k_T^-, k_T^+]$, all conclusions hold with $k = \hat{k}_T$. On an event of probability approaching one, every bound applies simultaneously for all k in the bracket. Lemma A.2 verifies this bracketing for the BIC order under $AR(p)$ errors, and for the adjusted BIC order under exponential decay. We henceforth write k for either a deterministic sequence satisfying Assumption 4 or such a data-driven order.

Lemma A.3. *(Berk-type rate and linear representation, robust to non-exogeneity.) Under Assumptions 1-5: (i)* $\|\hat{\rho}^D - \rho^{(k)}\| = O_p(\sqrt{k/T}) + O_p(\sqrt{k}_{j>k}|\rho_j|) = O_p(\sqrt{k/T})$*, and in particular* $\max_{1\leq j\leq k} |\hat{\rho}_j^D - \rho_j| = O_p(\sqrt{k/T})$*; (ii)* $\sqrt{T}(\hat{\rho}^D - \rho^{(k)}) = \Sigma_k^{-1} T_t^{-1/2} \psi_{t,k} e_t + \zeta_T$*, with* $\|\zeta_T\| = O_p(k^{3/2}/\sqrt{T}) + O_p(\sqrt{kT}_{j>k}|\rho_j|) = o_p(1)$*. The estimator is consistent whether or not the regressors are exogenous, provided they are pre-determined.*

Proof. By the Frisch-Waugh-Lovell theorem, $\hat{\rho}^D$ equals the OLS coefficient vector from regressing y_t on the sample residuals $\hat{\psi}_{t,k}$ of $(y_{t-1}, ..., y_{t-k})'$ on $W_{t,k}$. Since $y_{t-j} = x_{t-j}'\beta + u_{t-j}$ and $x_{t-j}'\beta$ lies in the span of $W_{t,k}$ for $0 \leq j \leq k$, residualising y_{t-j} on $W_{t,k}$ is the same as residualising u_{t-j}; thus $\hat{\psi}_{t,k} = U_{t-1,k} - \hat{\Pi}_k W_{t,k}$ with $\hat{\Pi}_k$ the sample projection coefficient, an object computable from the data even though $U_{t-1,k}$ is latent. Using $u_t = \rho^{(k)\prime} U_{t-1,k} + e_t + r_{kt}$ and the sample orthogonality ${}_t\hat{\psi}_{t,k} W_{t,k}' = 0$,

$$\hat{\rho}^D - \rho^{(k)} = (T_t^{-1} \hat{\psi}_{t,k} \hat{\psi}_{t,k}')^{-1} T_t^{-1} \hat{\psi}_{t,k} (e_t + r_{kt}). \tag{A.8}$$

(a) The design. The Toeplitz covariance matrix of any finite stack of z-lags has eigenvalues in $[2\pi c_z, 2\pi C_z]$ by Assumption 5(ii). These eigenvalue bounds carry over to Σ_k as follows. For the upper bound, $\Gamma_{u,k} = E[U_{t-1,k} U_{t-1,k}']$ is a principal submatrix of such a covariance matrix, so $\lambda_{\max}(\Gamma_{u,k}) \leq 2\pi C_z$; moreover, since $\psi_{t,k}$ is a projection residual, $a'\Sigma_k a = \min_b Var(a'U_{t-1,k} - b'W_{t,k}) \leq Var(a'U_{t-1,k}) = a'\Gamma_{u,k} a$ for every $a \in \mathbb{R}^k$. Hence $\Sigma_k \preceq \Gamma_{u,k}$ and $\lambda_{\max}(\Sigma_k) \leq 2\pi C_z$. This comparison cannot deliver the lower bound, because projection can only reduce the quadratic form: a lower bound on the eigenvalues of $\Gamma_{u,k}$ implies nothing for the smaller matrix Σ_k. Instead, note that $a'\Sigma_k a = Var(a'U_{t-1,k} - a'\Pi_k W_{t,k})$ is the variance of a linear combination of $z_t, ..., z_{t-k}$ whose coefficients on $(u_{t-1}, ..., u_{t-k})$ are exactly the entries of a (the intercept does not affect variances). If c denotes its full coefficient vector, then $\|c\| \geq \|a\|$, so that $a'\Sigma_k a \geq 2\pi c_z \|c\|^2 \geq 2\pi c_z \|a\|^2$ by Assumption 5(ii). Thus $2\pi c_z I \preceq \Sigma_k \preceq 2\pi C_z I$ uniformly in k. For the sample design, each entry of the joint second-moment matrix of $(U_{t-1,k}', W_{t,k}')'$ has sampling error of order $O(T^{-1/2})$ in L^2, uniformly over entries, by the covariance and cumulant summability of Assumption 5(i) (Berk (1974), Lemma 3). Summing the squared sampling errors over the $O(k^2)$ entries bounds the

Frobenius norm. Hence, writing $\hat{\Sigma}_k = T_t^{-1}\hat{\psi}_{t,k}\hat{\psi}'_{t,k}$ for the sample counterpart of Σ_k defined in Section 4.1, and since the spectral norm of a matrix is bounded by its Frobenius norm,

$$\|\hat{\Sigma}_k - \Sigma_k\| \leq \|\hat{\Sigma}_k - \Sigma_k\|_F = O_p(k/\sqrt{T}) = o_p(1) \quad \text{under } k^2/T \to 0, \tag{A.9}$$

where the passage from the raw second moments to the projected ones uses $\|\hat{\Pi}_k - \Pi_k\| = O_p(k/\sqrt{T})$ (same entrywise bounds, plus the uniform eigenvalue bounds on the W-block) and $\|\Pi_k\| = O(1)$. Hence the inverse in (A.8) is $O_p(1)$ in spectral norm. The rate in (A.9) is $k/\sqrt{T}$, not $\sqrt{k/T}$; only $o_p(1)$ is needed here.

(b) The score. Decompose $T_t^{-1}\hat{\psi}_{t,k}e_t = T_t^{-1}\psi_{t,k}e_t - (\hat{\Pi}_k - \Pi_k)T_t^{-1}W_{t,k}e_t$. Both $\psi_{t,k}$ and $W_{t,k}$ are $\mathcal{G}_{t-1}$-measurable (they involve $u_{t-1}, ..., u_{t-k}$ and $x_t, ..., x_{t-k}$ only), so by (22) and the conditional moment restrictions of Assumption 3, both scores are martingale differences with exact second moments: $E\|T_t^{-1}\psi_{t,k}e_t\|^2 = T^{-1}\sigma_e^2\text{tr}(\Sigma_k) = O(k/T)$ and likewise $\|T_t^{-1}W_{t,k}e_t\| = O_p(\sqrt{k/T})$. Hence,

$$T_t^{-1}\hat{\psi}_{t,k}e_t = T_t^{-1}\psi_{t,k}e_t + O_p(k/\sqrt{T})O_p(\sqrt{k/T}) = O_p(\sqrt{k/T}) + O_p(k^{3/2}/T).$$

(c) The truncation term. Each component satisfies

$$|T_t^{-1}\hat{\psi}_{j,t,k}r_{kt}| \leq (T_t^{-1}\hat{\psi}_{j,t,k}^2)^{1/2}(T_t^{-1}r_{kt}^2)^{1/2},$$

and $E[r_{kt}^2]^{1/2} \leq \theta_k E[u_t^2]^{1/2}$, so $\|T_t^{-1}\hat{\psi}_{t,k}r_{kt}\| = O_p(\sqrt{k}\theta_k)$—note the factor $\sqrt{k}$ from the growing dimension.

Combining (a)-(c) in (A.8) gives (i); the truncation contribution is $o_p(\sqrt{k/T})$ because $\sqrt{T}\theta_k \to 0$ under Assumption 4. For (ii), multiply (A.8) by $\sqrt{T}$ and collect the discrepancies from the leading term $\Sigma_k^{-1}T_t^{-1/2}\psi_{t,k}e_t$: (1) $[(T_t^{-1}\hat{\psi}_{t,k}\hat{\psi}'_{t,k})^{-1} - \Sigma_k^{-1}]T_t^{-1/2}\hat{\psi}_{t,k}e_t$ has norm $O_p(k/\sqrt{T})O_p(\sqrt{k}) = O_p(k^{3/2}/\sqrt{T})$, using (A.9) and $\|T_t^{-1/2}\hat{\psi}_{t,k}e_t\| = O_p(\sqrt{k})$ from (b); (2) $\Sigma_k^{-1}(\hat{\Pi}_k - \Pi_k)T_t^{-1/2}W_{t,k}e_t = O_p(k^{3/2}/\sqrt{T})$; (3) the truncation part, $\sqrt{T}O_p(\sqrt{k}\theta_k)$. All are $o_p(1)$ under Assumption 4; this is the step for which the rate $k^3/T \to 0$ (rather than $k^2/T \to 0$) and the strengthened tail condition are used. Crucially, step (b) uses only pre-determinedness; exogeneity is not required. □

Lemma A.4. *(Consistency of the design and scale estimates.) Under Assumptions 1-5 (with k as above), (i)* $(T-k)_t^{-1}\hat{\tilde{x}}_t\hat{\tilde{x}}'_t \xrightarrow{p} Q$*; (ii)* $\hat{\sigma}_{e,k}^2 \xrightarrow{p} \sigma_e^2$*; hence* $\widehat{AVar}$ *in (27) satisfies* $\widehat{AVar} \xrightarrow{p} \sigma_e^2 Q^{-1}$*. Consequently,* $\widehat{AVar}$ *consistently estimates the asymptotic variance of* $\sqrt{T}(\hat{\beta}_{FGLS} - \beta)$ *whenever the orthogonality condition (25) holds, while under non-exogenous regressors it estimates the leading term* $\sigma_e^2 Q^{-1}$ *and omits the correction* $\sigma_e^2 Q^{-1}\Gamma^* Q^{-1} \succeq 0$ *of Theorem 1(ii).*

Proof. (i) Write $\hat{\tilde{x}}_t = \check{\tilde{x}}_t - \sum_{j\leq k}\Delta_j x_{t-j}$. Then

$$(T-k)_t^{-1}\hat{\tilde{x}}_t\hat{\tilde{x}}'_t = (T-k)_t^{-1}\check{\tilde{x}}_t\check{\tilde{x}}'_t + R_T,$$

with

$$\|R_T\| \leq 2\|\Delta\| \, \|(T_t^{-1}\check{\widetilde{x}}_t x'_{t-j})_{j\leq k}\|_F + \|\Delta\|^2 \lambda_{\max}(T_t^{-1}\bar{X}_{t,k}\bar{X}'_{t,k}),$$

where $\bar{X}_{t,k}$ stacks $(x'_{t-1}, ..., x'_{t-k})'$. The first factor is $O_p(\sqrt{k/T})O_p(\sqrt{k}) = O_p(k/\sqrt{T})$ and the second $O_p(k/T)O_p(1) = O_p(k/T)$ (the stacked Gram matrix has $O_p(1)$ spectral norm under Assumption 5), so $\|R_T\| = O_p(k/\sqrt{T}) = o_p(1)$ under $k^2/T \to 0$. Also $(T-k)_t^{-1}\check{\widetilde{x}}_t\check{\widetilde{x}}'_t \xrightarrow{p} E[\widetilde{x}_t\widetilde{x}'_t] = Q$: the gap between $\check{\widetilde{x}}_t$ and $\widetilde{x}_t$ is $s_{kt} =_{j>k} \rho_j x_{t-j}$ with $E\|s_{kt}\|^2 = O(\theta_k^2) \to 0$, and $T_t^{-1}\widetilde{x}_t\widetilde{x}'_t \xrightarrow{p} Q$ by the ergodic theorem. Note that this part uses only Lemma A.3(i).

(ii) Expand $\widetilde{e}_{kt} = e_t + r_{kt} -_{j\leq k} \Delta_j u_{t-j} - \hat{\check{\widetilde{x}}}'_t(\hat{\beta}_{\text{FGLS}} - \beta)$; the squares and cross products of the last three terms are $o_p(1)$ in average by, respectively, Assumption 4 ($E[r_{kt}^2] = O(\theta_k^2)$), Lemma A.3(i) ($\|\Delta\|^2\lambda_{\max}(T_t^{-1}U_{t-1,k}U'_{t-1,k}) = O_p(k/T)$), and the $\sqrt{T}$-consistency of $\hat{\beta}_{\text{FGLS}}$ (Theorem 1(ii), whose proof relies only on part (i) of this lemma, so the argument is not circular), while $(T-k)_t^{-1}e_t^2 \xrightarrow{p} \sigma_e^2$ by the ergodic theorem. Slutsky's theorem then gives $\widehat{AVar} \xrightarrow{p} \sigma_e^2 Q^{-1}$. □

Proof of Proposition 1

From (18)-(19), write $\hat{\rho}_j^D = \rho_j + \Delta_j$, so that

$$\hat{\check{\widetilde{x}}}_t = \check{\widetilde{x}}_t -_{j=1}^{k} \Delta_j x_{t-j}, \qquad \hat{e}_{kt} = \check{e}_{kt} -_{j=1}^{k} \Delta_j u_{t-j}.$$

Substituting into $T_t^{-1/2}\hat{\check{\widetilde{x}}}_t\hat{e}_{kt}$ and using $\check{e}_{kt} = e_t + r_{kt}$, $\check{\widetilde{x}}_t = \widetilde{x}_t + s_{kt}$ with $s_{kt} =_{j>k} \rho_j x_{t-j}$, gives five groups of terms:

$$\begin{aligned} T_t^{-1/2}\hat{\check{\widetilde{x}}}_t\hat{e}_{kt} \;=\; & \underbrace{T_t^{-1/2}\widetilde{x}_t e_t}_{\text{(A)}} + \underbrace{T_t^{-1/2}(\widetilde{x}_t r_{kt} + s_{kt}e_t + s_{kt}r_{kt})}_{\text{(Trunc)}} \\ & - \underbrace{{}_{j=1}^{k}\Delta_j T_t^{-1/2} x_{t-j}\check{e}_{kt}}_{\text{(C)}} - \underbrace{{}_{j=1}^{k}\Delta_j T_t^{-1/2}\check{\widetilde{x}}_t u_{t-j}}_{\text{(B)}} \\ & + \underbrace{T_{i=1}^{-1/2k}{}_{j=1}^{k}\Delta_i\Delta_{jt}x_{t-i}u_{t-j}}_{\text{(D)}}. \end{aligned}$$

We treat each block in turn.

(Trunc). Since $r_{kt} =_{j>k} \rho_j u_{t-j}$ has $E[r_{kt}^2]^{1/2} \leq \theta_k E[u_t^2]^{1/2}$, the Cauchy-Schwarz inequality gives $E\|T_t^{-1/2}\widetilde{x}_t r_{kt}\| \leq \sqrt{T}E[\|\widetilde{x}_t\|^2]^{1/2}E[r_{kt}^2]^{1/2} = O(\sqrt{T}\theta_k) = o(1)$ by Assumption 4, and the same bound applies to $s_{kt}e_t$ and $s_{kt}r_{kt}$. Hence (Trunc)$= o_p(1)$. The bound for ${}_t s_{kt}e_t$ is crude—being a martingale difference score it is in fact $O_p(\theta_k)$—but nothing sharper is needed.

(C). Since $\check{e}_{kt} = e_t + r_{kt}$, and x_{t-j} is $\mathcal{G}_{t-1}$-measurable for every $j \geq 0$, each $T_t^{-1/2}x_{t-j}e_t$ is a martingale difference score with exact second moment $\sigma_e^2 E\|x_t\|^2$, uniform in j. By the

Cauchy-Schwarz inequality,

$$\begin{aligned}\|_{j\le k}\Delta_j T_t^{-1/2}x_{t-j}e_t\| &\le \|\Delta\|(\textstyle\sum_{j=1}^{k}\|T_t^{-1/2}x_{t-j}e_t\|^2)^{1/2}\\ &= O_p(\sqrt{k/T})O_p(\sqrt{k}) = O_p(k/\sqrt{T}) = o_p(1),\end{aligned}$$

because $k^2/T \to 0$. The truncation part of (C) is bounded by

$$\|\Delta\|\sqrt{k}\max_{j\le k}\|T_t^{-1/2}x_{t-j}r_{kt}\| = O_p(\sqrt{k/T})O_p(\sqrt{k}\sqrt{T}\theta_k) = O_p(k\theta_k) = o_p(1)$$

by Assumption 4.

(D). For each regressor component m, let $\bar{P}_k^{(m)}$ be the $k\times k$ matrix with entries $(\bar{P}_k^{(m)})_{ij} = T_t^{-1}x_{m,t-i}u_{t-j}$. Its spectral norm is $O_p(1)$: the mean $E[\bar{P}_k^{(m)}]$ is a section of the cross-covariance Toeplitz operator of (x_m, u), bounded by $2\pi C_z$ under Assumption 5(ii), and the sampling fluctuation is $O_p(k/\sqrt{T}) = o_p(1)$ by the entrywise bounds of Assumption 5(i), as in (A.9). Hence, componentwise, $|(\mathrm{D})_m| = |\sqrt{T}\Delta'\bar{P}_k^{(m)}\Delta| \le \sqrt{T}\|\bar{P}_k^{(m)}\|\|\Delta\|^2 = O_p(\sqrt{T}\cdot 1\cdot k/T) = O_p(k/\sqrt{T}) = o_p(1)$, which holds already under $k^2/T \to 0$.

(B). Write $T_t^{-1}\tilde{\tilde{x}}_t u_{t-j} = g_j + \eta_j$, where η_j collects two terms: (1) the bias $E[s_{kt}u_{t-j}] =_{i>k} \rho_i E[x_{t-i}u_{t-j}]$, which is $O(\theta_k)$ uniformly in j; and (2) the mean-zero sampling fluctuation, with $E\|\eta_j - E\eta_j\|^2 = O(T^{-1})$ uniformly in $j \le k$ by Assumption 5(i). Then

$$(\mathrm{B}) = \sqrt{T}G_k\Delta + \sqrt{T}_{j\le k}\Delta_j\eta_j, \qquad |\sqrt{T}_{j\le k}\Delta_j\eta_j| \le \sqrt{T}\|\Delta\|(_{j\le k}\|\eta_j\|^2)^{1/2},$$

and $(_{j\le k}\|\eta_j\|^2)^{1/2} = O_p(\sqrt{k/T}) + O(\sqrt{k}\theta_k)$, so the remainder is $O_p(k/\sqrt{T}) + O_p(k\theta_k) = o_p(1)$. The leading term is $O_p(1)$: by Lemma A.3(ii), $\sqrt{T}G_k\Delta = G_k\Sigma_k^{-1}T_t^{-1/2}\psi_{t,k}e_t + G_k\zeta_T$ with $\|G_k\zeta_T\| \le \|G_k\|\|\zeta_T\| = o_p(1)$, and the main part has exact second moment $\sigma_e^2\mathrm{tr}(G_k\Sigma_k^{-1}G_k') \le \sigma_e^2(2\pi c_z)^{-1}\|G_k\|_F^2 = O(1)$ by the martingale difference property, the conditional moment restrictions of Assumption 3, and the uniform eigenvalue bounds. Here we used the fact that, under Assumptions 2, 3 and 5(i), the g_j are absolutely summable: writing $g_j =_{i\ge 0} a_i E[x_{t-i}u_{t-j}]$ and noting that $E[x_{t-i}u_{t-j}]$ is a subvector of $\Gamma_z(i-j)$, $_j\|g_j\| \le (_i|a_i|)(_h\|\Gamma_z(h)\|) < \infty$; in particular $\sup_k\|G_k\| \le \sup_k\|G_k\|_F < \infty$.

Finally, $(T_t^{-1}\hat{\tilde{x}}_t\hat{\tilde{x}}_t')^{-1} \xrightarrow{p} Q^{-1}$ by Lemma A.4(i), which uses only Lemma A.3(i) and Assumptions 4-5, so no circularity is involved. Collecting (A), (B) and the negligible blocks, the identity (20) yields (24). ■

Proof of Theorem 1

(i) If $g_j = 0$ for all j then $G_k = 0$, so (24) reduces to $\sqrt{T}(\hat{\beta}_{\mathrm{FGLS}} - \beta) = Q^{-1}T_t^{-1/2}\tilde{x}_t e_t + o_p(1)$, exactly the expansion underlying Lemma A.1; the displayed form follows from (17). Exogeneity implies (25) because $g_j =_{i\ge 0\ l\ge 0} a_i c_l E[x_{t-i}e_{t-j-l}] = 0$ when every $E[x_{t-i}e_{t-j-l}] = 0$ (the interchange of summation and expectation being justified by absolute summability).

(ii) Substituting Lemma A.3(ii) into (24) and using $\|G_k\zeta_T\| \le \|G_k\|\|\zeta_T\| = o_p(1)$,

$$\sqrt{T}(\hat{\beta}_{\text{FGLS}} - \beta) = Q^{-1}T_t^{-1/2}(\widetilde{x}_t - G_k\Sigma_k^{-1}\psi_{t,k})e_t + o_p(1) = Q^{-1}T_t^{-1/2}s_{t,k} + o_p(1).$$

Both $\widetilde{x}_t$ and $\psi_{t,k}$ are $\mathcal{G}_{t-1}$-measurable, so $\{s_{t,k}, \mathcal{G}_t\}$ is, for each T, a square-integrable martingale difference array (the dimension of $\psi_{t,k}$ grows with T). We verify the two conditions of the martingale CLT for arrays (Hall and Heyde (1980), Corollary 3.1), for fixed c of the same dimension as β, writing $b_k' = c'G_k\Sigma_k^{-1}$ and noting $\|b_k\|^2 \le (2\pi c_z)^{-1}c'G_k\Sigma_k^{-1}G_k'c = O(1)$.

Conditional variances. By the conditional moment restrictions of Assumption 3,

$$T_t^{-1}E[(c's_{t,k})^2|\mathcal{G}_{t-1}] = \sigma_e^2[c'\hat{Q}c - 2b_k'\hat{m} + b_k'\hat{\Sigma}_k b_k], \tag{A.10}$$

with $\hat{Q} = T_t^{-1}\widetilde{x}_t\widetilde{x}_t'$, $\hat{m} = T_t^{-1}\psi_{t,k}\widetilde{x}_t'c$, $\hat{\Sigma}_k = T_t^{-1}\psi_{t,k}\psi_{t,k}'$. Now (1) $\hat{Q} \xrightarrow{p} Q$ (ergodic theorem); (2) $\|\hat{\Sigma}_k - \Sigma_k\| = O_p(k/\sqrt{T}) \to 0$ as in (A.9), so $b_k'\hat{\Sigma}_k b_k = b_k'\Sigma_k b_k + o_p(1) = c'G_k\Sigma_k^{-1}G_k'c + o_p(1)$; (3) $\|E[\psi_{t,k}\widetilde{x}_t']\| = O(\sqrt{k}\theta_k) \to 0$, because $\widetilde{x}_t = \sum_{i=0}^{k} a_i x_{t-i} - s_{kt}$ and the first piece lies in the span of $W_{t,k}$, which is orthogonal to $\psi_{t,k}$ in population. So $E[\psi_{t,k}\widetilde{x}_t'] = -E[\psi_{t,k}s_{kt}']$, each of whose k rows is $O(\theta_k)$ (and which vanishes exactly for a true $AR(p)$ process once $k \ge p$); the sampling fluctuation of $\hat{m}$ is $O_p(\sqrt{k/T})$ by Assumption 5(i). Hence, using Assumption 5(iii),

$$T_t^{-1}E[(c's_{t,k})^2|\mathcal{G}_{t-1}] \xrightarrow{p} \sigma_e^2 c'(Q + \Gamma^*)c = c'\Lambda c.$$

Lyapunov condition. By the conditional moment restrictions of Assumption 3 and the c_r-inequality, $E[(c's_{t,k})^4] \le 8\kappa_4(E[(c'\widetilde{x}_t)^4] + E[(b_k'\psi_{t,k})^4])$. The first term is finite since the L^4 norm of $\widetilde{x}_t$ is bounded by $(\sum_i |a_i|)$ times that of x_t, which is finite; for the second, $E[(b_k'\psi_{t,k})^4] = 3(b_k'\Sigma_k b_k)^2 + cum_4(b_k'\psi_{t,k})$, and the summable-cumulant condition of Assumption 5(i) bounds $|cum_4(b_k'\psi_{t,k})| \le C\|b_k\|^4$ uniformly in k (the projection coefficients defining $\psi_{t,k}$ have uniformly bounded ℓ_1-norm per row and the quadruple sums converge absolutely). Hence, $\sup_{t,k} E[(c's_{t,k})^4] < \infty$ and $T_t^{-2}E[(c's_{t,k})^4] = O(T^{-1}) \to 0$, which implies the conditional Lindeberg condition. The array CLT and the Cramér-Wold device then give $T_t^{-1/2}s_{t,k} \xrightarrow{d} N(0, \Lambda)$, and Slutsky's theorem completes the proof of (ii). If $G_k = 0$ (equivalently, under (25)) then $s_{t,k} = \widetilde{x}_t e_t$ and $\Lambda = \sigma_e^2 Q$, recovering part (i); the comparison $Q^{-1}\Lambda Q^{-1} \succeq \sigma_e^2 Q^{-1}$ is immediate from $\Gamma^* \succeq 0$. Finally, note from (A.10) that the finite-k variance of the effective score is

$$\Lambda_k = \sigma_e^2(Q - G_k\Sigma_k^{-1}E[\psi_{t,k}\widetilde{x}_t'] - E[\widetilde{x}_t\psi_{t,k}']\Sigma_k^{-1}G_k' + G_k\Sigma_k^{-1}G_k'), \tag{A.11}$$

whose cross terms are $O(\sqrt{k}\theta_k) \to 0$ (exactly zero for $AR(p)$ errors with $k \ge p$), yielding the clean limit $\Lambda = \sigma_e^2(Q + \Gamma^*)$ used in the statement; for bounded operative orders, e.g., $AR(p)$ errors with the BIC, the argument applies verbatim with Γ_p in place of Γ^*, since the conditional-variance step requires only that Γ_k converge along the realized order sequence, which is trivial when $k = p$ eventually. ■

Proof of Theorem 2

Recall from (A.9) that $\|\hat{\Sigma}_k - \Sigma_k\| = O_p(k/\sqrt{T})$; since $\Sigma_k \succeq 2\pi c_z I$ (part (a) of the proof of Lemma A.3), on an event of probability approaching one $\lambda_{\min}(\hat{\Sigma}_k) \geq \pi c_z$, whence $\|\hat{\Sigma}_k^{-1}\| = O_p(1)$ and

$$\|\hat{\Sigma}_k^{-1} - \Sigma_k^{-1}\| \leq \|\hat{\Sigma}_k^{-1}\|\|\Sigma_k - \hat{\Sigma}_k\|\|\Sigma_k^{-1}\| = O_p(k/\sqrt{T}). \tag{A.12}$$

Step 1: convergence of $\hat{G}_k$. Using $\hat{u}_{t-j} = u_{t-j} - x'_{t-j}(\hat{\beta}_{\text{FGLS}} - \beta)$ and $\hat{\tilde{x}}_t = \tilde{x}_t - \sum_{i\leq k} \Delta_i x_{t-i}$,

$$\hat{g}_j - g_j = \underbrace{(T_t^{-1}\tilde{x}_t u_{t-j} - g_j)}_{\eta_j} - \underbrace{\sum_{i\leq k}\Delta_i T_t^{-1} x_{t-i}u_{t-j}}_{\xi_j} - \underbrace{(T_t^{-1}\hat{\tilde{x}}_t x'_{t-j})(\hat{\beta}_{\text{FGLS}} - \beta)}_{\nu_j}.$$

For η_j: exactly as in block (B) of the proof of Proposition 1, $(\sum_{j\leq k}\|\eta_j\|^2)^{1/2} = O_p(\sqrt{k/T}) + O(\sqrt{k}\theta_k)$. For ξ_j: for each regressor component m, the k-vector $((\xi_j)_m)_{j\leq k}$ equals $\bar{P}_k^{(m)\prime}\Delta$ with $\bar{P}_k^{(m)}$ as in block (D) of that proof, so $(\sum_{j\leq k}(\xi_j)_m^2)^{1/2} \leq \|\bar{P}_k^{(m)}\|\|\Delta\| = O_p(1)O_p(\sqrt{k/T})$ by Lemma A.3(i); summing over the components gives $O_p(\sqrt{k/T})$. For ν_j: by the Cauchy-Schwarz inequality, simultaneously for all $j \leq k$,

$$\|T_t^{-1}\hat{\tilde{x}}_t x'_{t-j}\|_F \leq (T_t^{-1}\|\hat{\tilde{x}}_t\|^2)^{1/2}(T_s^{-1}\|x_s\|^2)^{1/2} = O_p(1),$$

because $T_t^{-1}\|\hat{\tilde{x}}_t\|^2 = \text{tr}(\hat{Q}_k) \xrightarrow{p} \text{tr}(Q)$ by Lemma A.4(i) and $E\|x_t\|^2 < \infty$; since $\sqrt{T}(\hat{\beta}_{\text{FGLS}} - \beta) = O_p(1)$ by Theorem 1(ii)—whose proof involves no variance estimation, so no circularity arises—we get $(\sum_{j\leq k}\|\nu_j\|^2)^{1/2} = O_p(\sqrt{k})O_p(T^{-1/2}) = O_p(\sqrt{k/T})$. Collecting,

$$\|\hat{G}_k - G_k\| \leq \|\hat{G}_k - G_k\|_F = O_p(\sqrt{k/T}) + O(\sqrt{k}\theta_k) = O_p(\sqrt{k/T}), \tag{A.13}$$

the tail term being dominated because $\sqrt{T}\theta_k \to 0$ under Assumption 4; in particular $\|\hat{G}_k\| \leq \sup_k \|G_k\|_F + o_p(1) = O_p(1)$.

Step 2: the correction. Decomposing

$$\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}'_k - G_k\Sigma_k^{-1}G'_k = (\hat{G}_k - G_k)\hat{\Sigma}_k^{-1}\hat{G}'_k + G_k(\hat{\Sigma}_k^{-1} - \Sigma_k^{-1})\hat{G}'_k + G_k\Sigma_k^{-1}(\hat{G}_k - G_k)'$$

and using (A.12), (A.13), $\sup_k \|G_k\| < \infty$ and $\|\Sigma_k^{-1}\| \leq (2\pi c_z)^{-1}$,

$$\|\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}'_k - \Gamma_k\| = O_p(\sqrt{k/T}) + O_p(k/\sqrt{T}) = O_p(k/\sqrt{T}) = o_p(1)$$

under Assumption 4, proving (i).

Step 3: assembly. By Lemma A.4, $\hat{Q}_k^{-1} \xrightarrow{p} Q^{-1}$ and $\hat{\sigma}^2_{e,k} \xrightarrow{p} \sigma_e^2$; moreover, $\|\Gamma_k\| \leq (2\pi c_z)^{-1}\sup_k \|G_k\|_F^2$, which is $O(1)$. Hence all products of the convergent factors converge, giving (ii). For (iii)a): under (25), $G_k = 0$, so (A.13) gives $\|\hat{G}_k\| = O_p(\sqrt{k/T})$, hence $\|\hat{G}_k\hat{\Sigma}_k^{-1}\hat{G}'_k\| \leq \|\hat{\Sigma}_k^{-1}\|\|\hat{G}_k\|^2 = O_p(k/T)$, and (ii) reduces to $\widehat{AVar}_R \xrightarrow{p} \sigma_e^2 Q^{-1}$. Part (iii)b) is immediate from (ii): for $AR(p)$ errors, $P(k_T^* = p) \to 1$ by Lemma A.2(i) and $\Gamma_k = \Gamma_p$ eventually; when $k \to \infty$, Assumption 5(iii) gives $\Gamma_k \to \Gamma^*$. ■

Proof of Corollary 1

The result follows from Theorem 1, Lemma A.4, Theorem 2 and Slutsky's theorem. ■

References

Baxter, G., 1962. An asymptotic result for the finite predictor. Math. Scand. 10, 137–144.

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

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Blanchard, O.J., 1981. What is left of the multiplier accelerator? Amer. Econ. Rev. Pap. Proc. 71, 132–143.

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

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

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic Press, New York.

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

Koreisha, S.G., Fang, Y., 2001. Generalized least squares with misspecified serial correlation structures. J. R. Stat. Soc. Ser. B 63, 515–531.

Miller, S., Startz, R., 2019. Feasible generalized least squares using machine learning. Econom. Lett. 175, 28–31.

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.

Romano, J.P., Wolf, M., 2017. Resurrecting weighted least squares. J. Econometrics 197, 1–19.

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

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288.

Zou, H., 2006. The adaptive Lasso and its oracle properties. J. Amer. Statist. Assoc. 101, 1418–1429.

Zou, H., Hastie, T., Tibshirani, R., 2007. On the degrees of freedom of the Lasso. Ann. Statist. 35, 2173–2192.

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

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

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

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

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

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

Table S.4: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, AR(1) case with $\rho_x = 0$. (First 3 columns are multiplied by 100). 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	0.68	0.42	0.42	0.42	0.09	0.05	0.06	0.06	0.68	0.42	0.42	0.42	0.94	0.95	0.94	0.32	0.25	0.25	1
		0	0.52	0.52	0.53	0.52	-0.07	-0.07	-0.06	-0.06	0.52	0.52	0.53	0.52	0.94	0.94	0.94	0.27	0.28	0.28	0
		0.2	0.53	0.49	0.50	0.51	0.02	-0.01	-0.01	-0.01	0.53	0.49	0.50	0.51	0.94	0.94	0.94	0.28	0.27	0.27	1
		0.5	0.67	0.41	0.41	0.42	0.21	0.19	0.19	0.20	0.67	0.41	0.41	0.42	0.94	0.95	0.95	0.32	0.25	0.25	1
		0.8	1.37	0.31	0.32	0.32	-0.11	-0.02	-0.02	-0.03	1.37	0.31	0.32	0.32	0.94	0.95	0.94	0.45	0.22	0.22	1
	$\gamma = 0.25$	-0.5	2.08	0.39	0.41	0.41	-11.95	-0.20	-0.72	-0.22	0.65	0.39	0.41	0.41	0.67	0.94	0.94	0.31	0.24	0.24	1
		0	0.48	0.48	0.52	0.50	-0.06	-0.06	0.05	-0.02	0.48	0.48	0.52	0.50	0.94	0.94	0.94	0.27	0.27	0.27	0
		0.2	0.69	0.46	0.48	0.54	4.47	-0.09	0.43	0.91	0.49	0.46	0.48	0.53	0.89	0.94	0.93	0.27	0.26	0.27	1
		0.5	1.98	0.38	0.41	0.41	11.57	0.06	0.71	0.22	0.64	0.38	0.40	0.41	0.68	0.94	0.94	0.31	0.24	0.24	1
		0.8	4.72	0.29	0.29	0.29	18.46	0.09	0.38	0.16	1.31	0.29	0.29	0.29	0.62	0.95	0.95	0.43	0.21	0.21	1
	$\gamma = 0.5$	-0.5	4.71	0.34	0.48	0.39	-20.31	-0.44	-2.83	-0.43	0.58	0.34	0.40	0.39	0.20	0.91	0.94	0.29	0.23	0.23	1
		0	0.41	0.41	0.47	0.44	-0.32	-0.32	-0.12	-0.25	0.41	0.41	0.47	0.44	0.94	0.93	0.94	0.24	0.25	0.25	0
		0.2	1.02	0.38	0.48	0.65	7.75	-0.08	1.66	1.87	0.42	0.38	0.45	0.62	0.77	0.92	0.88	0.25	0.24	0.26	1
		0.5	4.45	0.32	0.45	0.37	19.72	0.12	2.74	0.40	0.56	0.32	0.38	0.37	0.22	0.92	0.94	0.28	0.23	0.23	1
		0.8	10.80	0.25	0.27	0.27	31.09	0.16	1.18	0.28	1.13	0.25	0.26	0.27	0.13	0.95	0.94	0.40	0.20	0.20	1
$T = 500$	$\gamma = 0$	-0.5	0.28	0.16	0.17	0.17	0.05	0.01	0.01	0.01	0.28	0.16	0.17	0.17	0.95	0.95	0.95	0.20	0.16	0.16	1
		0	0.19	0.19	0.20	0.20	0.00	0.00	0.00	0.00	0.19	0.19	0.20	0.20	0.95	0.95	0.95	0.17	0.18	0.18	0
		0.2	0.20	0.19	0.19	0.19	0.00	-0.01	-0.01	0.00	0.20	0.19	0.19	0.19	0.95	0.95	0.95	0.18	0.17	0.17	1
		0.5	0.27	0.16	0.16	0.16	-0.01	0.00	0.00	0.01	0.27	0.16	0.16	0.16	0.95	0.95	0.95	0.20	0.16	0.16	1
		0.8	0.55	0.12	0.12	0.13	0.14	0.05	0.05	0.05	0.55	0.12	0.12	0.13	0.95	0.95	0.94	0.29	0.14	0.14	1
	$\gamma = 0.25$	-0.5	1.65	0.15	0.16	0.16	-11.82	-0.09	-0.51	-0.11	0.25	0.15	0.16	0.16	0.34	0.95	0.95	0.20	0.15	0.15	1
		0	0.19	0.19	0.20	0.19	-0.05	-0.05	-0.01	-0.03	0.19	0.19	0.20	0.19	0.95	0.94	0.95	0.17	0.17	0.17	0
		0.2	0.41	0.18	0.19	0.20	4.66	-0.03	0.32	0.10	0.20	0.18	0.19	0.20	0.81	0.94	0.94	0.17	0.17	0.17	1
		0.5	1.62	0.15	0.16	0.16	11.65	-0.01	0.44	0.05	0.26	0.15	0.16	0.16	0.35	0.95	0.95	0.20	0.15	0.15	1
		0.8	3.98	0.12	0.12	0.12	18.60	-0.01	0.16	0.02	0.52	0.12	0.12	0.12	0.26	0.95	0.95	0.28	0.13	0.13	1
	$\gamma = 0.5$	-0.5	4.29	0.13	0.21	0.14	-20.18	-0.18	-2.40	-0.21	0.22	0.13	0.15	0.14	0.01	0.90	0.95	0.18	0.15	0.15	1
		0	0.17	0.17	0.19	0.17	-0.15	-0.15	-0.10	-0.14	0.17	0.17	0.19	0.17	0.94	0.92	0.94	0.16	0.16	0.16	0
		0.2	0.79	0.15	0.20	0.21	7.87	-0.06	1.55	0.31	0.17	0.15	0.18	0.21	0.51	0.92	0.93	0.16	0.15	0.17	1
		0.5	4.18	0.13	0.20	0.14	19.88	0.02	2.33	0.15	0.22	0.13	0.15	0.14	0.01	0.90	0.95	0.18	0.15	0.15	1
		0.8	10.45	0.10	0.11	0.10	31.62	0.04	0.82	0.08	0.45	0.10	0.10	0.10	0.00	0.95	0.95	0.26	0.13	0.12	1

Table S.5: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, AR(2) case with $\rho_x = 0$. (First 3 columns are multiplied by 100). 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	0.65	0.38	0.39	0.01	0.03	0.02	0.65	0.38	0.39	0.94	0.94	0.31	0.24	2
		-0.5,0.3	1.14	0.38	0.39	0.12	0.06	0.05	1.14	0.38	0.39	0.94	0.95	0.41	0.24	2
		1.34,-0.42	5.33	0.17	0.17	0.20	-0.01	0.00	5.33	0.17	0.17	0.94	0.95	0.88	0.16	2
		0,0.3	0.56	0.47	0.49	-0.05	-0.03	-0.04	0.56	0.47	0.49	0.94	0.95	0.29	0.27	2
		0.5,0.3	1.12	0.38	0.39	-0.04	0.01	0.00	1.12	0.38	0.39	0.94	0.94	0.40	0.24	2
	$\gamma = 0.25$	0.5,-0.3	1.96	0.36	0.39	11.60	0.09	0.26	0.61	0.36	0.39	0.67	0.95	0.30	0.24	2
		-0.5,0.3	2.55	0.35	0.39	-12.06	-0.23	0.04	1.09	0.35	0.39	0.78	0.94	0.40	0.24	2
		1.34,-0.42	14.20	0.16	0.17	30.23	0.08	0.08	5.06	0.16	0.17	0.72	0.95	0.85	0.16	2
		0,0.3	0.54	0.44	0.48	-0.21	-0.10	-0.03	0.54	0.44	0.48	0.94	0.95	0.28	0.26	2
		0.5,0.3	2.31	0.36	0.39	11.27	0.03	-0.09	1.04	0.36	0.39	0.79	0.94	0.39	0.24	2
	$\gamma = 0.5$	0.5,-0.3	4.43	0.31	0.39	19.84	0.12	0.45	0.50	0.31	0.38	0.18	0.94	0.27	0.23	2
		-0.5,0.3	5.08	0.30	0.42	-20.29	-0.18	0.35	0.96	0.30	0.42	0.44	0.92	0.38	0.24	2
		1.34,-0.42	30.60	0.14	0.15	51.36	0.12	0.17	4.22	0.14	0.15	0.25	0.94	0.78	0.15	2
		0,0.3	0.49	0.37	0.47	-0.25	-0.09	0.10	0.49	0.37	0.47	0.94	0.94	0.27	0.26	2
		0.5,0.3	4.59	0.30	0.42	19.11	-0.02	-0.41	0.94	0.30	0.42	0.47	0.92	0.37	0.24	2
$T = 500$	$\gamma = 0$	0.5,-0.3	0.26	0.15	0.15	0.00	0.00	0.00	0.26	0.15	0.15	0.95	0.95	0.20	0.15	2
		-0.5,0.3	0.45	0.15	0.15	0.03	0.05	0.05	0.45	0.15	0.15	0.95	0.95	0.26	0.15	2
		1.34,-0.42	2.18	0.07	0.07	0.06	0.00	0.00	2.18	0.07	0.07	0.95	0.95	0.57	0.10	2
		0,0.3	0.22	0.18	0.19	-0.04	-0.01	-0.01	0.22	0.18	0.19	0.95	0.95	0.18	0.17	2
		0.5,0.3	0.45	0.15	0.15	0.11	0.03	0.03	0.45	0.15	0.15	0.94	0.95	0.26	0.15	2
	$\gamma = 0.25$	0.5,-0.3	1.62	0.14	0.15	11.76	0.03	0.08	0.24	0.14	0.15	0.32	0.95	0.19	0.15	2
		-0.5,0.3	1.85	0.14	0.15	-11.91	-0.04	0.03	0.44	0.14	0.15	0.56	0.95	0.26	0.15	2
		1.34,-0.42	11.63	0.06	0.07	30.97	0.04	0.04	2.05	0.06	0.07	0.41	0.95	0.56	0.10	2
		0,0.3	0.22	0.17	0.18	-0.11	-0.08	-0.03	0.22	0.17	0.18	0.94	0.95	0.18	0.17	2
		0.5,0.3	1.74	0.14	0.15	11.51	-0.03	-0.05	0.42	0.14	0.15	0.58	0.95	0.25	0.15	2
	$\gamma = 0.5$	0.5,-0.3	4.18	0.12	0.14	19.97	0.09	0.16	0.20	0.12	0.14	0.00	0.95	0.17	0.15	2
		-0.5,0.3	4.40	0.12	0.15	-20.02	-0.05	0.07	0.39	0.12	0.15	0.08	0.95	0.24	0.15	2
		1.34,-0.42	29.66	0.05	0.06	52.80	0.08	0.08	1.78	0.05	0.06	0.01	0.94	0.51	0.09	2
		0,0.3	0.19	0.15	0.18	-0.04	0.00	0.10	0.19	0.15	0.18	0.94	0.94	0.17	0.16	2
		0.5,0.3	4.23	0.12	0.15	19.63	0.02	0.00	0.38	0.12	0.15	0.08	0.94	0.24	0.15	2

Table S.6: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, MA(1) case with $\rho_x = 0$. (First 3 columns are multiplied by 100). 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	0.76	0.28	0.29	0.32	0.08	-0.05	-0.05	-0.03	0.76	0.28	0.29	0.32	0.94	0.93	0.94	0.34	0.20	0.22	3
		-0.4	0.60	0.43	0.43	0.45	0.05	0.00	0.01	0.03	0.60	0.43	0.43	0.45	0.94	0.95	0.95	0.30	0.25	0.26	2
		0.5	0.64	0.39	0.40	0.42	-0.02	-0.02	-0.03	0.00	0.64	0.39	0.40	0.42	0.94	0.94	0.94	0.31	0.24	0.25	2
	$\gamma = 0.25$	-0.7	3.44	0.26	0.28	0.36	-16.57	-0.44	-0.56	-1.96	0.69	0.25	0.28	0.32	0.48	0.93	0.93	0.32	0.19	0.21	3
		-0.4	1.45	0.41	0.44	0.47	-9.55	-0.36	-0.26	-1.25	0.54	0.41	0.43	0.45	0.74	0.93	0.94	0.29	0.24	0.26	1
		0.5	1.98	0.36	0.39	0.44	11.80	0.28	0.45	1.53	0.58	0.36	0.39	0.42	0.65	0.94	0.94	0.30	0.23	0.24	2
	$\gamma = 0.5$	-0.7	8.50	0.22	0.29	0.49	-28.18	-0.73	-1.03	-3.70	0.56	0.21	0.28	0.35	0.03	0.91	0.89	0.29	0.18	0.21	3
		-0.4	3.08	0.35	0.44	0.52	-16.20	-0.51	-0.38	-2.24	0.46	0.34	0.44	0.47	0.32	0.91	0.92	0.26	0.23	0.25	1
		0.5	4.41	0.30	0.40	0.51	19.83	0.19	0.46	2.60	0.48	0.30	0.39	0.45	0.17	0.91	0.91	0.27	0.21	0.24	2
$T = 500$	$\gamma = 0$	-0.7	0.29	0.11	0.11	0.11	-0.01	-0.02	-0.02	-0.04	0.29	0.11	0.11	0.11	0.95	0.94	0.95	0.21	0.13	0.13	4
		-0.4	0.23	0.17	0.17	0.17	0.08	0.05	0.05	0.05	0.23	0.17	0.17	0.17	0.95	0.95	0.95	0.19	0.16	0.16	2
		0.5	0.25	0.15	0.15	0.16	0.01	0.03	0.03	0.03	0.25	0.15	0.15	0.16	0.95	0.95	0.95	0.19	0.15	0.15	2
	$\gamma = 0.25$	-0.7	3.00	0.10	0.10	0.13	-16.50	-0.17	-0.20	-1.08	0.28	0.10	0.10	0.11	0.12	0.94	0.94	0.20	0.12	0.13	4
		-0.4	1.13	0.17	0.18	0.19	-9.51	-0.18	-0.14	-0.75	0.23	0.16	0.18	0.18	0.46	0.94	0.94	0.18	0.16	0.16	2
		0.5	1.62	0.14	0.15	0.16	11.77	0.05	0.11	0.76	0.23	0.14	0.15	0.16	0.31	0.95	0.95	0.19	0.15	0.15	2
	$\gamma = 0.5$	-0.7	8.09	0.08	0.11	0.17	-28.06	-0.34	-0.47	-2.17	0.22	0.08	0.11	0.13	0.00	0.91	0.90	0.18	0.11	0.13	4
		-0.4	2.77	0.14	0.17	0.21	-16.08	-0.20	-0.16	-1.39	0.18	0.14	0.17	0.19	0.03	0.92	0.92	0.17	0.14	0.16	2
		0.5	4.18	0.12	0.15	0.19	19.97	0.11	0.26	1.52	0.19	0.12	0.15	0.17	0.00	0.92	0.92	0.17	0.14	0.15	2

Table S.7: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals, ARMA(1,1) case with $\rho_x = 0$. (First 3 columns are multiplied by 100). 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	1.05	0.26	0.28	-0.11	0.01	0.01	1.05	0.26	0.28	0.94	0.95	0.40	0.20	2
		0.2,-0.4	0.54	0.49	0.51	-0.05	-0.06	-0.06	0.54	0.49	0.51	0.94	0.95	0.28	0.27	1
		0.2,0.5	0.78	0.31	0.34	0.02	-0.03	-0.04	0.78	0.31	0.34	0.94	0.95	0.34	0.22	2
		0.5,-0.4	0.52	0.51	0.52	-0.14	-0.12	-0.13	0.52	0.51	0.52	0.94	0.94	0.28	0.28	0
		0.5,0.5	1.17	0.22	0.23	-0.08	0.05	0.04	1.17	0.22	0.23	0.94	0.95	0.41	0.19	3
		0.8,-0.4	0.73	0.43	0.45	-0.05	-0.07	-0.08	0.73	0.43	0.45	0.94	0.94	0.33	0.25	2
		0.8,0.5	2.82	0.16	0.17	-0.08	-0.05	-0.05	2.82	0.16	0.17	0.94	0.95	0.64	0.16	3
	$\gamma = 0.25$	-0.5,-0.4	5.61	0.25	0.31	-21.44	-0.33	-1.27	1.01	0.25	0.29	0.41	0.93	0.38	0.20	2
		0.2,-0.4	0.73	0.46	0.56	-4.80	-0.19	-1.33	0.49	0.46	0.54	0.89	0.93	0.27	0.27	1
		0.2,0.5	3.38	0.30	0.37	16.29	0.20	1.45	0.72	0.30	0.35	0.49	0.93	0.32	0.22	2
		0.5,-0.4	0.52	0.47	0.52	2.12	-0.17	0.96	0.48	0.47	0.51	0.94	0.94	0.27	0.27	0
		0.5,0.5	6.49	0.21	0.26	23.21	0.10	1.29	1.11	0.21	0.24	0.37	0.94	0.40	0.19	3
		0.8,-0.4	1.49	0.40	0.46	8.98	-0.13	-0.71	0.68	0.40	0.46	0.80	0.93	0.32	0.25	2
		0.8,0.5	11.66	0.15	0.19	30.00	0.16	1.26	2.67	0.15	0.18	0.52	0.93	0.62	0.16	3
	$\gamma = 0.5$	-0.5,-0.4	14.12	0.21	0.37	-36.43	-0.57	-2.40	0.85	0.21	0.31	0.01	0.90	0.35	0.20	2
		0.2,-0.4	1.08	0.39	0.69	-8.15	-0.34	-2.65	0.41	0.39	0.62	0.74	0.87	0.25	0.26	1
		0.2,0.5	8.36	0.25	0.44	27.87	0.29	2.67	0.59	0.25	0.37	0.04	0.90	0.29	0.22	2
		0.5,-0.4	0.56	0.40	0.57	3.79	-0.09	1.98	0.41	0.40	0.53	0.90	0.91	0.25	0.25	0
		0.5,0.5	16.61	0.18	0.33	39.62	0.30	2.64	0.92	0.17	0.26	0.01	0.90	0.37	0.18	2
		0.8,-0.4	3.02	0.34	0.50	15.56	0.04	-1.18	0.60	0.34	0.49	0.46	0.91	0.30	0.24	2
		0.8,0.5	27.95	0.13	0.26	50.74	0.25	2.36	2.20	0.13	0.20	0.04	0.88	0.57	0.15	3
$T = 500$	$\gamma = 0$	-0.5,-0.4	0.43	0.10	0.10	0.01	0.03	0.01	0.43	0.10	0.10	0.94	0.95	0.25	0.13	3
		0.2,-0.4	0.21	0.19	0.20	0.01	-0.01	-0.01	0.21	0.19	0.20	0.95	0.95	0.18	0.17	1
		0.2,0.5	0.31	0.12	0.13	0.01	0.01	0.01	0.31	0.12	0.13	0.94	0.95	0.21	0.14	3
		0.5,-0.4	0.20	0.19	0.19	-0.04	-0.03	-0.04	0.20	0.19	0.19	0.95	0.95	0.18	0.17	1
		0.5,0.5	0.47	0.09	0.09	-0.08	0.05	0.05	0.47	0.09	0.09	0.95	0.95	0.27	0.12	3
		0.8,-0.4	0.29	0.17	0.17	-0.05	0.00	0.00	0.29	0.17	0.17	0.95	0.95	0.21	0.16	2
		0.8,0.5	1.12	0.06	0.06	0.33	0.03	0.03	1.12	0.06	0.06	0.95	0.95	0.41	0.10	3
	$\gamma = 0.25$	-0.5,-0.4	4.89	0.10	0.11	-21.21	-0.05	-0.72	0.39	0.10	0.11	0.07	0.94	0.24	0.13	2
		0.2,-0.4	0.42	0.18	0.20	-4.72	-0.06	-0.54	0.19	0.18	0.20	0.81	0.94	0.17	0.17	1
		0.2,0.5	2.96	0.11	0.14	16.37	0.00	0.77	0.28	0.11	0.13	0.12	0.94	0.21	0.14	3
		0.5,-0.4	0.25	0.19	0.22	2.36	0.04	0.74	0.19	0.19	0.22	0.91	0.93	0.17	0.17	1
		0.5,0.5	5.92	0.08	0.10	23.42	0.07	0.78	0.43	0.08	0.09	0.05	0.94	0.26	0.12	3
		0.8,-0.4	1.15	0.16	0.17	9.33	0.03	-0.05	0.28	0.16	0.17	0.56	0.95	0.20	0.16	2
		0.8,0.5	10.23	0.06	0.07	30.24	0.05	0.69	1.09	0.06	0.07	0.15	0.94	0.40	0.10	3
	$\gamma = 0.5$	-0.5,-0.4	13.36	0.08	0.14	-36.08	-0.23	-1.48	0.34	0.08	0.12	0.00	0.91	0.22	0.12	2
		0.2,-0.4	0.81	0.15	0.23	-8.06	-0.14	-1.11	0.16	0.15	0.22	0.49	0.92	0.16	0.17	1
		0.2,0.5	8.07	0.10	0.16	27.98	0.11	1.60	0.24	0.09	0.13	0.00	0.92	0.19	0.14	3
		0.5,-0.4	0.32	0.16	0.27	3.95	0.00	1.33	0.17	0.16	0.25	0.83	0.88	0.16	0.16	0
		0.5,0.5	16.23	0.07	0.13	39.82	0.13	1.51	0.38	0.07	0.10	0.00	0.89	0.24	0.11	3
		0.8,-0.4	2.73	0.14	0.18	15.75	-0.03	-0.18	0.25	0.14	0.18	0.09	0.94	0.19	0.16	2
		0.8,0.5	27.38	0.05	0.09	51.45	0.10	1.30	0.91	0.05	0.08	0.00	0.89	0.37	0.10	3

Table S.8: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals. Predictive regression with $\rho_x = 0$. (First 3 columns are multiplied by 100). 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	0.78	0.27	0.28	0.31	0.06	0.01	0.00	0.02	0.78	0.27	0.28	0.31	0.94	0.94	0.94	0.34	0.20	0.21	3
		-0.4	0.60	0.43	0.44	0.45	-0.03	-0.03	-0.02	-0.01	0.60	0.43	0.44	0.45	0.94	0.94	0.94	0.30	0.25	0.26	2
		0.5	0.64	0.39	0.39	0.41	0.03	-0.02	-0.03	-0.03	0.64	0.39	0.39	0.41	0.94	0.94	0.94	0.31	0.24	0.25	2
	$\gamma = 0.25$	-0.7	0.70	0.26	0.27	0.36	-0.09	-0.34	-0.38	-2.08	0.70	0.26	0.27	0.32	0.95	0.94	0.92	0.33	0.19	0.21	3
		-0.4	0.57	0.41	0.42	0.45	-0.08	-0.15	-0.13	-0.65	0.57	0.41	0.42	0.44	0.94	0.94	0.94	0.29	0.25	0.25	2
		0.5	0.58	0.35	0.36	0.40	-0.11	-0.01	0.04	1.15	0.58	0.35	0.36	0.39	0.94	0.95	0.94	0.30	0.23	0.24	2
	$\gamma = 0.5$	-0.7	0.61	0.22	0.23	0.45	-0.06	-0.57	-0.66	-3.73	0.61	0.22	0.23	0.31	0.94	0.93	0.87	0.30	0.18	0.20	3
		-0.4	0.47	0.34	0.34	0.40	-0.18	-0.33	-0.30	-1.21	0.47	0.33	0.34	0.39	0.94	0.95	0.93	0.26	0.23	0.23	2
		0.5	0.51	0.31	0.32	0.42	-0.17	0.07	0.12	2.09	0.51	0.31	0.32	0.37	0.94	0.94	0.92	0.27	0.21	0.23	2
$T = 500$	$\gamma = 0$	-0.7	0.30	0.11	0.11	0.11	0.02	0.06	0.06	0.07	0.30	0.11	0.11	0.11	0.95	0.94	0.95	0.21	0.13	0.13	4
		-0.4	0.24	0.17	0.17	0.18	-0.06	-0.06	-0.06	-0.05	0.23	0.17	0.17	0.18	0.95	0.95	0.95	0.19	0.16	0.16	2
		0.5	0.25	0.15	0.16	0.16	-0.01	-0.01	-0.01	-0.01	0.25	0.15	0.16	0.16	0.95	0.94	0.94	0.20	0.15	0.15	2
	$\gamma = 0.25$	-0.7	0.29	0.10	0.10	0.13	0.04	-0.09	-0.11	-1.34	0.29	0.10	0.10	0.12	0.95	0.94	0.92	0.21	0.12	0.13	4
		-0.4	0.22	0.16	0.16	0.17	0.01	-0.06	-0.05	-0.78	0.22	0.16	0.16	0.17	0.95	0.95	0.94	0.18	0.16	0.16	2
		0.5	0.24	0.14	0.14	0.17	-0.13	-0.06	-0.05	1.04	0.24	0.14	0.14	0.16	0.95	0.95	0.94	0.19	0.15	0.15	2
	$\gamma = 0.5$	-0.7	0.24	0.08	0.09	0.19	-0.05	-0.24	-0.29	-2.66	0.24	0.08	0.09	0.12	0.95	0.94	0.84	0.19	0.11	0.12	4
		-0.4	0.18	0.13	0.13	0.17	-0.05	-0.14	-0.12	-1.41	0.18	0.13	0.13	0.15	0.95	0.95	0.92	0.17	0.14	0.15	2
		0.5	0.20	0.12	0.12	0.19	-0.04	0.00	0.03	2.06	0.20	0.12	0.12	0.15	0.95	0.95	0.89	0.17	0.14	0.14	2

Table S.9: Bias, empirical Mean Squared Error, Variance, Coverage Rate and Length of Confidence Intervals. Serially correlated and heteroskedastic errors with $\rho_x = 0$. (First 3 columns are multiplied by 100). 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. Med. k is the median autoregressive lag order selected by BIC across replications.

			MSE			Bias			Variance			Coverage			Length			
			OLS	F-C	F-CH	OLS	F-C	F-CH	OLS	F-C	F-CH	OLS	F-C	F-CH	OLS	F-C	F-CH	Med. k
$w = x$	$\nu = x^2$	AR(1)	6.74	4.17	2.95	-0.41	-0.37	-0.30	6.74	4.16	2.95	0.94	0.94	0.94	0.97	0.79	0.65	1
		AR(2)	50.10	1.65	1.12	1.16	0.21	0.22	50.09	1.65	1.12	0.90	0.94	0.94	2.40	0.50	0.40	2
		MA(1)	6.31	4.09	2.98	-0.28	-0.16	0.04	6.31	4.09	2.98	0.94	0.94	0.94	0.95	0.78	0.65	2
		ARMA(1,1)	7.04	4.39	3.25	0.19	-0.20	-0.18	7.03	4.39	3.25	0.94	0.94	0.93	1.00	0.80	0.67	2
	$\nu = \log(x)^2$	AR(1)	0.80	0.48	0.23	-0.05	0.05	0.04	0.80	0.48	0.23	0.93	0.95	0.94	0.33	0.27	0.18	1
		AR(2)	6.01	0.20	0.08	0.19	-0.06	-0.01	6.01	0.20	0.08	0.90	0.94	0.94	0.84	0.18	0.10	2
		MA(1)	0.72	0.47	0.25	-0.03	-0.01	-0.04	0.72	0.47	0.25	0.94	0.95	0.94	0.33	0.27	0.19	2
		ARMA(1,1)	0.84	0.52	0.29	0.04	-0.02	-0.03	0.84	0.52	0.29	0.93	0.94	0.93	0.34	0.28	0.20	2
	$\nu = \exp(0.2(x + x^2))$	AR(1)	14.12	8.76	3.89	-0.26	-0.04	-0.03	14.12	8.76	3.89	0.93	0.95	0.94	1.40	1.15	0.75	1
		AR(2)	84.69	3.11	1.26	1.93	0.40	0.26	84.66	3.10	1.26	0.90	0.94	0.94	3.14	0.68	0.43	2
		MA(1)	13.29	8.44	3.94	-0.47	-0.51	-0.34	13.29	8.44	3.94	0.94	0.94	0.94	1.38	1.13	0.76	2
		ARMA(1,1)	14.45	9.52	4.47	-0.48	-0.24	-0.24	14.45	9.51	4.47	0.93	0.94	0.93	1.43	1.16	0.78	2
$w = \phi x + (1 - \phi)q$	$\nu = x^2$	AR(1)	6.54	4.12	3.97	0.33	0.18	0.18	6.54	4.12	3.97	0.94	0.95	0.94	0.97	0.79	0.76	1
		AR(2)	49.86	1.66	1.46	-0.92	-0.15	-0.11	49.85	1.66	1.46	0.90	0.94	0.94	2.40	0.50	0.46	2
		MA(1)	6.21	4.02	3.88	-0.12	-0.16	0.02	6.21	4.02	3.88	0.94	0.94	0.94	0.95	0.78	0.75	2
		ARMA(1,1)	7.05	4.37	4.28	0.09	0.03	-0.02	7.05	4.37	4.28	0.94	0.94	0.93	1.00	0.80	0.77	2
	$\nu = \log(x)^2$	AR(1)	0.77	0.48	0.51	0.04	-0.02	-0.02	0.77	0.48	0.51	0.94	0.95	0.93	0.34	0.27	0.26	1
		AR(2)	5.96	0.20	0.17	0.06	0.01	0.02	5.96	0.20	0.17	0.90	0.95	0.95	0.84	0.18	0.15	2
		MA(1)	0.74	0.47	0.47	0.01	-0.06	-0.09	0.74	0.47	0.47	0.94	0.94	0.94	0.33	0.27	0.26	2
		ARMA(1,1)	0.84	0.53	0.52	0.13	0.09	0.11	0.84	0.53	0.52	0.94	0.94	0.93	0.34	0.28	0.26	2
	$\nu = \exp(0.2(x + x^2))$	AR(1)	13.96	8.73	8.34	0.16	-0.02	-0.03	13.96	8.73	8.34	0.93	0.95	0.94	1.40	1.15	1.09	1
		AR(2)	84.42	3.05	2.39	-1.47	-0.14	-0.18	84.40	3.05	2.39	0.90	0.95	0.94	3.12	0.68	0.58	2
		MA(1)	13.38	8.61	7.84	-0.48	-0.23	0.03	13.38	8.61	7.84	0.94	0.94	0.94	1.38	1.13	1.06	2
		ARMA(1,1)	14.29	9.24	8.82	-0.38	-0.22	0.10	14.29	9.24	8.82	0.93	0.94	0.94	1.43	1.16	1.10	2