The Econometrics Journal

Econometrics Journal (2024), volume **00**, pp. 1–43.
https://doi.org/10.1093/ectj/utae019

On robust inference in time-series regression

RICHARD T. BAILLIE[†], FRANCIS X. DIEBOLD[‡], GEORGE KAPETANIOS[§], KUN HO KIM[‖] AND AARON MORA[¶]

[†]*Department of Economics, Michigan State University, East Lansing, MI 48824, USA; and Kings College Business School, University of London, London WC2B, UK.*
Email: baillie@msu.edu

[‡]*Department of Economics, University of Pennsylvania, Philadelphia, PA 19104; and NBER, USA.*
Email: fdiebold@sas.upenn.edu

[§]*Kings College Business School, University of London, London WC2B, UK.*
Email: george.kapetanios@kcl.ac.uk

[‖]*John Molson School of Business, Concordia University, Montreal, QC H3H 0A1, Canada.*
Email: kunho.kim@concordia.ca

[¶]*Darla Moore School of Business, University of South Carolina, Columbia, SC 29208, USA.*
Email: aaron.mora@moore.sc.edu

First version received: 18 October 2023; final version accepted: 10 September 2024.

Summary: Least squares regression with heteroskedasticity consistent standard errors (‘OLS-HC regression’) has proved very useful in cross-section environments. However, several major difficulties, which are generally overlooked, must be confronted when transferring the HC technology to time-series environments via heteroskedasticity and autocorrelation consistent standard errors (‘OLS-HAC regression’). First, in plausible time-series environments, OLS parameter estimates can be inconsistent, so that OLS-HAC inference fails even asymptotically. Second, most economic time series have autocorrelation, which renders OLS parameter estimates inefficient. Third, autocorrelation similarly renders conditional predictions based on OLS parameter estimates inefficient. Finally, the structure of popular HAC covariance matrix estimators is ill-suited for capturing the autoregressive autocorrelation typically present in economic time series, which produces large size distortions and reduced power in HAC-based hypothesis testing, in all but the largest samples. We show that all four problems are largely avoided by the use of a simple and easily implemented dynamic regression procedure, which we call DURBIN. We demonstrate the advantages of DURBIN with detailed simulations covering a range of practical issues.

Keywords: *DURBIN regression*, *dynamic regression*, *heteroskedasticity and autocorrelation consistent (HAC) regression*, *serial correlation*.

JEL codes: *C13*, *C22*, *C31*.

1. INTRODUCTION

For nearly a century, regression with heteroskedastic and/or autocorrelated disturbances has featured prominently in empirical economics research. For many decades, attention centred on modelling the heteroskedasticity or autocorrelation in the context of feasible generalised least squares (FGLS) estimation.

The dominant estimation approach in recent decades, however, is ordinary least squares (OLS) with standard errors adjusted to achieve valid asymptotic inference without taking a stand on the form of heteroskedasticity or autocorrelation. The idea traces to the classic contribution of White (1980), who considered OLS regression with heteroskedasticity consistent (HC) standard errors ('OLS-HC regression') in cross-sectional environments, where sample sizes are typically very large, little or no information is available regarding the form of any possible heteroskedasticity, and serial correlation is irrelevant. In such environments HC standard errors are appropriate and justly emphasised (e.g., Angrist and Pischke, 2008).

In an elegant extension, Newey and West (1987) generalise White's estimator from cross sections to time series, with possible heteroskedasticity *and* serial correlation, by replacing White's covariance matrix estimator with an appropriate time-series analogue based on an estimator of a spectral density at frequency zero.[1] Such OLS regression with heteroskedasticity and autocorrelation consistent (HAC) standard errors ('OLS-HAC regression') has become extremely popular in time-series environments.

In this paper we argue, however, that, in contrast to cross-section OLS-HC regression, time-series OLS-HAC regression as typically implemented is likely to be problematic, for a variety of reasons:

(1) In plausible time-series environments, OLS parameter estimates can be inconsistent, so that OLS-HAC inference fails even asymptotically.
And moreover, even when OLS parameter estimates are consistent:
(2) OLS parameter estimates can be highly inefficient in the presence of serial correlation, compared to estimators that account for the serial correlation.
(3) OLS-HAC regression discards valuable predictive information in serially correlated disturbances and hence produces suboptimal (inefficient) forecasts, whereas accurate out-of-sample prediction is often a central concern in time-series econometrics.
(4) Newey–West-style HAC covariance matrix estimators are ill-suited for capturing the autoregressive autocorrelation typically present in economic time series, which can produce large size distortions, and large power reductions even when the size is not distorted.

Claim 1 is not widely appreciated, with the exception of Perron and González-Coya (2022), whose results and approach complement ours.[2] Claim 2 is well known, but its importance in finite samples is ignored when using OLS-HAC regression. Claim 3 is obvious, but again ignored when using OLS-HAC regression. Claim 4 is appreciated and has motivated several important

[1] The Newey–West estimator collapses to the White (1980) estimator if serial correlation is absent, but appropriately incorporates serial correlation in the calculation of robust standard errors when serial correlation is present.

[2] By now parts of our paper and theirs are entangled. A preliminary version of our paper was presented at the 2016 NBER-NSF Time Series Conference at Columbia University. Our first-draft working paper was released in March 2022, with no knowledge of their work-in-progress. Their first-draft working paper was released in September 2022, with knowledge of ours. Our second-draft working paper was released in June 2022, with knowledge of theirs. This third draft of our paper was released on 2024/10/18 15:03:00.

refinements of the Newey–West HAC covariance matrix estimator (e.g., Andrews, 1991; Kiefer and Vogelsang, 2002; Lazarus et al., 2018), as well as use of spectral density estimators that differ from the Newey–West lag-window estimator (e.g., Müller, 2014). However, those refinements have been only partially successful.

Against the background of the above claims 1–4, which we will substantiate in detail, we proceed to make a constructive contribution. We propose an alternative to OLS-HAC regression based on so-called DURBIN regressions (Durbin, 1970). Working in a very general environment that includes most dynamic specifications of interest as special cases, we show that the new procedure simultaneously addresses claims 1–4 above. Indeed, the DURBIN regression procedure performs well in all situations, dominating the traditional OLS-HAC and FGLS procedures.

Our paper proceeds as follows. In Section 2, we introduce the basic data-generating process and estimators, including, not only traditional OLS-HAC regression and our DURBIN regression, but also traditional FGLS and a recently proposed modified FGLS procedure. In Section 3, we present a generalised modelling framework. In Section 4, we present extensive simulation evidence. We conclude in Section 5, and we present supplementary results in three appendices.

2. DATA-GENERATING PROCESS AND ESTIMATORS

Traditional OLS-HAC regression focuses exclusively on OLS parameter estimation, assuming consistency and surrendering on efficiency. But, as we emphasise in this section, even OLS consistency cannot be assumed without significant loss of generality. Moreover, aspects of the consistency and efficiency of OLS and various competitors, under various conditions, are nuanced and not widely appreciated. Hence in this section we begin by reviewing aspects of OLS consistency and efficiency in comparison to competitors—in particular, a new procedure that we propose based on Durbin (1970) regressions, a new modified FGLS procedure, and traditional FGLS—in a sequence of progressively richer dynamic environments.

2.1. Data-Generating Process

We start with the standard data-generating process (DGP) in the OLS-HAC regression literature,

$$y_t = x_t'\beta + u_t, \tag{2.1}$$

where $t = 1, 2, ..., T$, β is a k-vector of parameters, x_t is a k-vector of covariance-stationary covariates and u_t is a scalar covariance-stationary disturbance with $E(u_t u_t') = \sigma^2\Omega$.[3] DGP (2.1) is usually augmented with conditions such that OLS is consistent. Then the econometrician generally aims to provide standard error corrections that enable asymptotically valid inference. Note that such OLS-HAC regression involves just a static regression of y_t on x_t, basically imported directly from cross-sectional micro-econometrics, with dynamics allowed only through u_t. We will later argue that such a framework is uncompelling in time-series environments, but it is the industry standard in OLS-HAC regression, so we maintain it for now.

[3] Because u_t is covariance-stationary, it can be serially correlated and/or conditionally heteroskedastic. In this paper we emphasise serial correlation exclusively, because serial correlation is the unique feature of time-series data relative to cross-section data. Cross sections do of course sometimes have a spatial dimension and therefore a natural ordering in space if not in time, and spatial correlation has recently begun to receive attention from a HAC estimation perspective, as in Müller and Watson (2022). Spatial HAC estimation is, however, beyond the scope of this paper.

Crucial insights will flow from adopting a starting point that allows for significant generality regarding possible relationships between x_t and u_t. In particular, consider the Wold representation of the Gaussian vector process $z_t = (x_t', u_t)'$,

$$z_t = \sum_{i=0}^{\infty} \Xi_i \varepsilon_{t-i}. \tag{2.2}$$

The coefficient matrices are $\Xi_0 = I$ and

$$\Xi_i = \begin{pmatrix} \xi_{x,i} & \xi_{xu,i} \\ \xi_{ux,i}' & \xi_{u,i} \end{pmatrix},$$

and $\varepsilon_t = (\varepsilon_{x,t}', \varepsilon_{u,t})'$ is a vector white noise innovation process with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s') = 0$ for $s \neq t$, and contemporaneous covariance matrix $E(\varepsilon_t \varepsilon_t') = \Sigma$, where

$$\Sigma = \begin{pmatrix} \Sigma_x & \Sigma_{xu} \\ \Sigma_{xu}' & \sigma_u \end{pmatrix}.$$

Under mild regularity conditions, the infinite vector moving average representation (2.2) is equivalent to the infinite vector-autoregressive (VAR) representation[4]

$$z_t = \sum_{i=1}^{\infty} \Psi_i z_{t-i} + \varepsilon_t, \tag{2.3}$$

where

$$\Psi_i = \begin{pmatrix} \Psi_{x,i} & \Psi_{xu,i} \\ \Psi_{ux,i} & \psi_{u,i} \end{pmatrix}.$$

This setting encompasses a variety of DGPs, and we will consider the consistency and efficiency properties of different estimators under various restrictions imposed on (2.3).

We now proceed to consider various estimation strategies that may be appropriate in the environment given by (2.1) and (2.3).

2.2. *OLS Parameter Estimation and HAC Covariance Matrix Estimation*

In the standard notation, the OLS estimator of the regression parameter is of course

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

If $\Omega = I$, the limiting distribution of the OLS estimator is

$$T^{1/2}\left(\hat{\beta}_{OLS} - \beta\right) \to N\left(0, \sigma^2 Q^{-1}\right),$$

where $Q = p\lim_{T\to\infty}\left(T^{-1}X'X\right)$.

Based on the VAR representation (2.3), we define 'block diagonality' (*BD*) as holding when $\Psi_{ux,i} = \Psi_{xu,i} = 0$, for all i, and $\Sigma_{xu} = 0$. The *BD* condition implies strong exogeneity, namely that $E(u_s|x_t) = 0$ for all s and t.[5] In the *BD* environment OLS is consistent, but asymptotically

[4] Such regularity conditions include assumptions on the rate of decline of $\|\Xi_i\|$ towards zero as $i \to \infty$, for suitable norm $\|.\|$, to control the persistence of the process and to avoid phenomena such as long memory that complicate the analysis. For details see, e.g., Davidson (2002) and references cited therein.

[5] Strong exogeneity is sometimes called strict exogeneity.

inefficient, with limiting distribution

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

where $V = Q^{-1}\Omega Q^{-1}$. The key object in V is Ω, which is the spectrum of $x_t u_t$ at frequency zero. HAC inference estimates V using

$$\widehat{V} = Q^{-1}\widehat{\Omega}Q^{-1},$$

where $\widehat{\Omega}$ is a consistent estimator of Ω, so that $\widehat{V}$ is consistent for V. Different choices for $\widehat{\Omega}$ therefore define different HAC covariance matrix estimators and are the main issue in implementing OLS-HAC regression, as we discuss subsequently in Section 4.2.1.

2.3. FGLS Estimation

If condition *BD* holds, and if the matrix Ω is known, then generalised least square (GLS) is a consistent and asymptotically efficient estimator of β. However, Ω is almost always unknown, in which case attention turns to FGLS as defined by Amemiya (1973), which is again both consistent and asymptotically efficient provided that condition *BD* holds.[6]

The OLS-HAC regression literature was historically motivated by environments where OLS is consistent for β, but where condition *BD* simultaneously fails in such a way that FGLS is inconsistent. Such situations are possible, and we will discuss such a classic situation (Hansen and Hodrick, 1980) at some length in Section 3.4, but they are by no means the only or the most important possibility. Indeed, there is much more to investigate when *BD* fails, as emphasised in the insightful work of Perron and González-Coya (2022).

We now consider an alternative estimation procedure that avoids the above discussed OLS-HAC and FGLS complications and *always* delivers consistent (and sometimes fully efficient) estimates of β, together with reliable asymptotic inference.

2.4. DURBIN Estimation and Its Relatives

A natural third approach to estimation and inference, which we will argue is generally preferable to both OLS-HAC and FGLS, is based on the 'Durbin (1970) regression', given by

$$y_t = x_t'\beta + \sum_{j=1}^{\infty}\phi_j y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\gamma_j + \varepsilon_{y,t}, \tag{2.4}$$

where $\varepsilon_{y,t}$ is serially uncorrelated, and uncorrelated with y_{t-j} and x_{t-j} for all j. The DURBIN regression 'cleans out' disturbance dynamics by its direct inclusion of y_{t-j} and x_{t-j}, so that standard OLS estimation and inference are trustworthy. We refer to the DURBIN regression, and the associated estimator of β, as DURBIN. Crucially, note that the DGP remains (2.1) and (2.3); DURBIN is simply a certain procedure (regression) that can be implemented on data from that DGP, just as OLS and FGLS are certain procedures that can be implemented on data from that DGP.

[6] Recent contributions to the FGLS literature include Romano and Wolf (2017) for heteroskedastic environments, and Kapetanios and Psaradakis (2016) for dynamic environments.

Operationally, it is of course necessary to use a finite order approximation to the infinite order DURBIN regression (2.4),

$$y_t = x_t'\beta + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{p} x_{t-j}'\gamma_j + \varepsilon_{y,t},$$

with finite lag order p selected using a data based procedure, typically an information criterion, and increasing at a suitable rate. The theoretical validity of such a procedure for producing valid asymptotic estimation and inference is well known (see, e.g., Lewis and Reinsel, 1985; Hannan and Deistler, 1988), and we shall have more to say about it when we later implement DURBIN in the simulations of Section 4.

We can also write the finite order DURBIN approximation as

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p}\sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + \varepsilon_{y,t}, \tag{2.5}$$

which emphasises the extent of the parameterisation and lag structure. We will later explore in greater detail the relationship between the DGP given by (2.1) and (2.3) and the DURBIN regression (2.5), which is effectively one equation of a VAR and appears to be the originator of the autoregressive distributed lag (ADL) model, which is widely used in empirical econometric work.

An estimator closely related to DURBIN, recently proposed by Perron and González-Coya (2022), is a variation on FGLS. We refer to it as FGLS-D (short for 'FGLS-DURBIN'). While FGLS uses a first-stage OLS regression, FGLS-D uses a first-stage DURBIN regression (2.5). Under *BD*, it follows that FGLS-D is also efficient. However, when *BD* does not hold, FGLS-D may not be efficient or even consistent, while DURBIN remains consistent.

Given that condition *BD* may not hold, it is important to consider the implications of its violation for the various methods of estimation and inference. To see the effects of the various sub-conditions embedded in condition *BD*, we will relax it in sequential stages. First, we impose only that $\Psi_{ux,i} = 0$ for all i and that $\Sigma_{xu} = 0$, so that x is weakly exogenous (that is, $E(u_s|x_t) = 0$, for all s with $t<s$), but not strongly exogenous.[7] x_t now depends on lags of u_t, but not vice versa. We refer to this restriction as *GEXOG* ('GLS exogeneity'). Clearly, OLS is now inconsistent, as is FGLS, which uses OLS residuals, while FGLS-D remains consistent and efficient. Importantly, DURBIN remains consistent, even if not fully efficient, throughout.

Second, we impose only $\Sigma_{xu} = 0$, so that x is neither strongly nor weakly exogenous. We denote this condition by *EBD* ('error variance block diagonal'). u_t now depends on lags of x_t, and the finite-ordered FGLS autoregression for u_t is no longer valid. Therefore, neither FGLS nor FGLS-D is consistent. DURBIN, however, remains consistent under *EBD*, and moreover it is also efficient.

[7] Weak exogeneity is sometimes called predeterminedness, and sometimes defined as $E(u_s|x_t) = 0$, for all s with $t \leq s$ rather than $t < s$. See Mikusheva and Sølvsten (2023).

To see the consistency and efficiency of DURBIN under EBD, note that, using (2.3) and $u_t = y_t - x_t'\beta$, we can write (2.1) as

$$\begin{aligned} y_t &= x_t'\beta + u_t \\ &= x_t'\beta + \sum_{j=1}^{\infty} \psi_{u,j} u_{t-j} + \sum_{j=1}^{\infty} \Psi_{xu,j} x_{t-j} + \varepsilon_{u,t} \\ &= x_t'\beta + \sum_{j=1}^{\infty} \psi_{u,j} \left(y_{t-j} - x_{t-j}'\beta \right) + \sum_{j=1}^{\infty} \Psi_{xu,j} x_{t-j} + \varepsilon_{u,t} \\ &= x_t'\beta + \sum_{j=1}^{\infty} \psi_{u,j} y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}' \left(\Psi_{xu,j}' - \psi_{u,j}\beta \right) + \varepsilon_{u,t}. \end{aligned} \tag{2.6}$$

Noting that the relationship $\gamma_j = \Psi_{xu,j} - \psi_{u,j}\beta$ gives a one-to-one mapping between γ_j and $\Psi_{xu,j}$, given values for $\psi_{u,j}$ and β, we immediately obtain efficiency for DURBIN.[8]

Finally, we impose no restrictions at all, in which case all methods become inconsistent and the use of instrumentation appears to be the only way forward.

In summary, OLS requires stronger conditions for consistency than the FGLS variants. The FGLS variants, in turn, require stronger conditions for consistency than DURBIN. Hence, overall, DURBIN has attractive consistency features in comparison with OLS and the FGLS variants. However, when the FGLS variants are consistent, they are also fully efficient. We shall see how such trade-offs resolve themselves in the simulations in Section 4.

3. A GENERALISED DATA-GENERATING PROCESS

We now move from the basic DGP (2.1) to a generalised version that subsumes all cases of interest.

3.1. Data-Generating Process

Henceforth, we work with the data-generating process given by

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + u_t. \tag{3.1}$$

We emphasise that u_t may also be a dynamic process, to allow, for example, for missing covariates. In particular, we continue to allow $(x_t', u_t)'$ to follow the vector moving average (2.2), or equivalently, the vector autoregression (2.3). This generalised DGP covers most linear dynamic relationships of conceivable interest. We use *NDY* ('no dynamics in y') to refer to the restriction imposed on the generalised DGP (3.1) to get the basic DGP (2.1), namely $\phi_j = \gamma_{i,j} = 0 \, \forall i, j$.

We also emphasise that (3.1) is now the data-generating process, and various regressions could be fit to its data realisations in various attempts at estimation and inference for β.

[8] Of course, if $\Psi_{xu,j} = 0$, then DURBIN, which estimates γ_j, is over-parameterised, providing a simple argument showing that DURBIN is inefficient under *GEXOG*.

One such regression, for example, is FGLS. Clearly, the use of FGLS in environments characterised by the generalised DGP (3.1) accounts only for $x_t'\beta$ and therefore ignores all terms involving lags, resulting in misspecification of the conditional mean part of the fitted regression. That is, the only way lagged information is used in FGLS is through estimation of the error covariance matrix, which neglects the problem of misspecification of the conditional mean.

DURBIN is another such regression that can be fit to the generalised DGP (3.1). Indeed, DURBIN can perfectly accommodate the generalised DGP, because, in precise parallel to (2.6), we have

$$\begin{aligned}
y_t &= x_t'\beta + \sum_{j=1}^{\infty}\lambda_j y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\theta_j + u_t \\
&= x_t'\beta + \sum_{j=1}^{\infty}\lambda_j y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\theta_j + \sum_{j=1}^{\infty}\psi_{u,j}u_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\psi_{xu,j} + \varepsilon_{u,t} \\
&= x_t'\beta + \sum_{j=1}^{\infty}\lambda_j y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\theta_j \\
&\quad + \sum_{j=1}^{\infty}\psi_{u,j}\left(y_{t-j} - x_{t-j}'\beta - \sum_{s=1}^{\infty}\lambda_s y_{t-j-s} - \sum_{s=1}^{\infty} x_{t-j-s}'\theta_s\right) + \sum_{j=1}^{\infty} x_{t-j}'\Psi_{xu,j} + \varepsilon_{u,t} \\
&= x_t'\beta + \sum_{j=0}^{\infty}\left(\lambda_j + \psi_{u,j} - \sum_{s=1}^{j-1}\psi_{u,s}\lambda_{j-s}\right) y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\left(\Psi_{xu,j} - \psi_{u,j}\beta - \sum_{s=1}^{j-1}\psi_{u,s}\theta_{j-s}\right) + \varepsilon_{u,t},
\end{aligned}$$

which is a DURBIN regression with

$$\phi_j = \lambda_j + \psi_{u,j} - \sum_{s=1}^{j-1}\psi_{u,s}\lambda_{j-s}$$

$$\gamma_j = \psi_{xu,j} - \psi_{u,j}\beta - \sum_{s=1}^{j-1}\psi_{u,s}\theta_{j-s}.$$

The above relationships between the parameters of the generalised DGP (3.1) and the DURBIN regression (2.5) also show that the generalised DGP is so richly parameterised that not all parameters are identified through estimation of (2.5) alone. β is always identified and consistently estimable via DURBIN, however, even in cases where OLS, FGLS, and FGLS-D are inconsistent.

3.2. Estimator Comparisons

In Table 1 we summarise the consistency and efficiency properties of all estimators in the leading environments that we have considered, all of which are specialisations of the generalised DGP given by (3.1) and (2.3). Table 1 makes clear the important trade-off between the occasional efficiency of FGLS/FGLS-D and the robust consistency of DURBIN. That is, although FGLS is sometimes efficient when DURBIN is not (under $NDY + BD$ and $NDY + GEXOG$), DURBIN is *always* at least consistent, and FGLS is not.

Table 1. Estimator consistency and efficiency under various conditions.

Restriction	Estimator			
	OLS	DURBIN	FGLS	FGLS-D
NDY + BD	✓×	✓×	✓✓	✓✓
NDY + GEXOG	××	✓×	××	✓✓
NDY + EBD	××	✓✓	××	××
EBD	××	✓✓	××	××
None	××	××	××	××

Notes: We show the consistency and efficiency properties of various estimators under various restrictions on the generalised DGP (3.1) with $(x_t, u_t)'$ governed by (2.3). In each cell of the table, the first checkmark, or lack thereof, relates to consistency and the second to efficiency.

Indeed, the *EBD* row of Table 1 is starkly revealing, as, for example, it includes simple and natural DGPs like

$$y_t = x_t\beta + \phi y_{t-1} + x_{t-1}\gamma + u_t.$$

The conventional FGLS procedure would be to regress y_t on x_t, and then to regress the residuals on lagged residuals, thereby obtaining the Cochrane–Orcutt filter to apply to the y_t and x_t series. One strongly suspects, and our subsequent simulations in Section 4 show clearly, that FGLS will perform poorly in this environment unless $\gamma \approx \beta\phi$, in which case the DGP reduces (approximately) to just a static regression of y_t on x_t with AR(1) disturbances.[9]

3.3. Hausman Tests

Table 1 also highlights the potential usefulness of tests for validity of the various restrictions. If for example, one 'knew' that *NDY + GEXOG* held, then FGLS or FGLS-D would be fully appealing estimators (consistent and efficient) whereas DURBIN would be less appealing (consistent, but not efficient). Alternatively, if one knew that instead *NDY + EBD* held, then FGLS or FGLS-D would be highly *un*appealing (inconsistent) whereas DURBIN would be fully appealing (consistent and efficient).

Hausman tests are available, as follows. Clearly, restrictions on the parameters of (2.3) determine the comparative desirability of alternative methods of estimation and inference for β. The key restriction is *BD*. Under the null hypothesis that *BD* holds with u serially correlated, OLS is consistent, but not efficient, while FGLS is both consistent and efficient. Under the alternative hypothesis that *BD* fails, OLS and FGLS are generally both inconsistent, but have different limits, which depend on the parameters of (2.3). As a result, Hausman tests can be used.

In particular, one may wish to query whether $\beta_1 = \beta_2$, where

$$E(y_t|x_t) = x_t'\beta_1$$

and

$$E(y_t|x_t, x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}, ...) = x_t'\beta_2 + \sum_{j=1}^{\infty}\phi_j y_{t-j} + \sum_{j=1}^{\infty}x_{t-j}'\gamma_j.$$

[9] The restriction $\gamma \approx \beta\phi$ is known as the common factor restriction.

Under the null hypothesis, FGLS should be used. Otherwise, one should consider using FGLS-D or DURBIN if one is interested in β_2 as would typically be the case, or consider using OLS if for some reason β_1 is of interest.

Overall, however, we find it preferable simply to use DURBIN under all circumstances, unless there is some compelling reason to do otherwise. There are three reasons:

(1) An acceptable HAC estimator of the variance of the OLS estimator may not be available when implementing a Hausman test. Indeed, the poor performance of OLS-HAC is the theme of this paper.
(2) As regards consistent/efficient estimation, it will be clear from the simulation results in Section 4 that the Mean Squared Error (MSE) cost of using DURBIN when a more efficient estimator is available (i.e., when *BD* or at least *GEXOG* holds) is generally small, whereas the MSE cost of *not* using DURBIN can be very large when neither *BD* nor *GEXOG* holds.
(3) As regards consistent inference, it will also be clear from the simulation results in Section 4 that DURBIN-based inference performs well in all circumstances that we investigate, both in terms of test size and power, in contrast to all other methods that we consider, where inference often fails.

We will shortly turn to the extensive simulation results alluded to in points 2 and 3 above, but first we briefly consider DURBIN versus other estimation approaches in the important context of predictive inference.

3.4. *Predictive Inference*

As is clear from Table 1, OLS is rarely consistent in time-series situations of interest. One case where OLS *is* consistent and simultaneously FGLS is inconsistent involves multi-step forecast evaluation, where one tests whether a forecast x_t is unbiased for y_{t+k}. That is, one tests whether

$$E(y_{t+k}|x_t) = x_t,$$

for $k \geq 1$.

One of the earliest analyses of this problem was by Hansen and Hodrick (1980), where y_{t+k} represented the k-period-ahead spot exchange rate and x_t represented the current k-period forward rate. The null hypothesis of $\beta = 1$ implies moving average disturbances, producing a violation of strong exogeneity while nevertheless satisfying weak exogeneity. Hansen and Hodrick (1980) recognised that FGLS can be inconsistent in such a situation, whereas OLS remains consistent, albeit inefficient. They recognised, moreover, that the OLS standard error was inconsistent and therefore required a 'correction'—and OLS-HAC was born.

Note, however, that DURBIN is also perfectly applicable in the Hansen–Hodrick environment, delivering, not only consistent standard errors, but also efficient as opposed to merely consistent parameter estimates.[10] In particular, under the null of unbiasedness, the error term,

$$u_{t+k} = y_{t+k} - x_t,$$

[10] For a full empirical analysis, see Baillie et al. (2023).

satisfies $Cov(u_{t+j}u_t) = 0$ for $j > k$, which implies that u_{t+k} can be represented by an $MA(k-1)$ process. Hence we can write

$$y_{t+k} = x_t\beta + \theta(L)\varepsilon_{t+k}, \tag{3.2}$$

where ε_t is a white noise process and $\theta(L)$ is a polynomial in the lag operator of order $k-1$.

Conceptually, (3.2) is merely a restricted DURBIN model, because on using the filter $\theta(L)^{-1}$ we obtain

$$\left\{\theta(L)^{-1}y_{t+k}\right\} = \beta\left\{\theta(L)^{-1}x_t\right\} + \varepsilon_{t+k}. \tag{3.3}$$

The filtered explanatory variable is uncorrelated with current and future innovations, ε_{t+k}, so that estimation of (3.3) by OLS will produce consistent and asymptotically efficient estimates of the regression parameters. In practice, it is convenient to use the approximation $\theta(L)^{-1} \approx \pi(L)$, where $\pi(L) = \left(1 - \pi_1 L - \ldots - \pi_p L^p\right)$ is a pth-order lag-operator polynomial with all roots outside the unit circle. DURBIN will then be

$$\pi(L)y_{t+k} = \beta\pi(L)x_t + \varepsilon_{t+k},$$

which is a restricted version of the generalised DGP (3.1) and can also be estimated by restricted OLS.

4. SIMULATION EVIDENCE ON ESTIMATION AND TESTING

In this section we examine, via simulation, the sampling properties of the various estimators, the properties of forecasts that use those estimated parameters, and crucially, the size and power of associated hypothesis tests. Supplementary results are contained in Appendices A (AR disturbances; see Table A1), B (MA disturbances; see Tables B1–B3, and Figures B1–B2), and C (ARMA disturbances; see Tables C1–C3 and Figures C1–C2).

4.1. Simulation Design

The main simulation results will comprise four data-generation processes that impose different assumptions on the generalised DGP given by (3.1) and (2.3):

(1) Autoregressive disturbances, AR(1) (*NDY* + *BD*)

$$\begin{aligned} y_t &= \beta x_t + u_t \\ \begin{pmatrix} x_t \\ u_t \end{pmatrix} &= \begin{pmatrix} 0.7 & 0 \\ 0 & \rho \end{pmatrix}\begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix}. \end{aligned} \tag{4.1}$$

(2) Triangular vector autoregression (VAR) on (2.3) (*NDY* + *GEXOG*)

$$\begin{aligned} y_t &= \beta x_t + u_t \\ \begin{pmatrix} x_t \\ u_t \end{pmatrix} &= \begin{pmatrix} \psi_{11} & \psi_{12} \\ 0 & \psi_{22} \end{pmatrix}\begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix}. \end{aligned} \tag{4.2}$$

(3) Unrestricted VAR on (2.3) (*NDY* + *EBD*)

$$y_t = \beta x_t + u_t$$
$$\begin{pmatrix} x_t \\ u_t \end{pmatrix} = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix}. \quad (4.3)$$

(4) Dynamic regression (*EBD*)

$$y_t = \beta x_t + \rho y_{t-1} - 0.5 x_{t-1} + u_t$$
$$\begin{pmatrix} x_t \\ u_t \end{pmatrix} = \begin{pmatrix} 0.7 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix}. \quad (4.4)$$

In all cases, $(\varepsilon_{x,t}, \varepsilon_{u,t})' \sim iidN(0, I)$ with $t = 1, ..., T$. We explore $T \in \{50, 200, 600, 2500\}$, which also spans the relevant range for macroeconomics, where structural change and other considerations tend to keep sample spans to roughly 'the most recent fifty years'; that is, sample sizes of 50 years, 200 quarters, 600 months, or approximately 2,500 weeks. Including $T = 2500$ also lets us check our Monte Carlo results against known large-sample results.

The autoregressive DGP in (4.1) matches the design in Lazarus et al. (2018). We explore $\rho \in \{0, .3, .5, .7, .9, .95, .99\}$, which spans the relevant range for economics. All ρ values are positive, as economic time series are generally positively serially correlated, and they range from white noise to the very strong serial correlation often of relevance in macroeconomic series. Including the white noise case ($\rho = 0$) allows us to check our Monte Carlo results against known results for the independent and identically distributed (i.i.d.) case.

In the simulations for the triangular VAR DGP in (4.2), we consider the following values for the matrix Ψ:

$$\Psi_1 = \begin{pmatrix} 0.4 & 0.7 \\ 0 & 0.5 \end{pmatrix} \qquad \Psi_1^* = \begin{pmatrix} 0.4 & 0.7 \\ 0 & 0.6 \end{pmatrix}.$$

Ψ_1^* has a larger leading eigenvalue than Ψ_1 (0.6 versus 0.5) and hence exhibits stronger autoregressive features.

For the unrestricted VAR DGP in (4.3), we consider the following values:

$$\Psi_2 = \begin{pmatrix} 0.4 & 0.7 \\ 0.3 & 0.5 \end{pmatrix} \qquad \Psi_2^* = \begin{pmatrix} 0.4 & 0.7 \\ 0.3 & 0.6 \end{pmatrix}.$$

As before, Ψ_2^* was selected to be similar to Ψ_2, but with a larger leading eigenvalue (0.97 for Ψ_2^* and 0.91 for Ψ_2).

For the dynamic regression DGP in (4.4), we consider various parameter values for the coefficient on y_{t-1}, namely $\rho \in \{0, 0.5, 0.7, 0.95\}$. When $\rho = 0.5$, the common factor restriction introduced in Note 10 holds, in which case we expect FGLS and FGLS-D to perform well.

For all DGPs in our simulations we perform 10,000 Monte Carlo replications. We simulate exact realisations of x and u by drawing x_0 and u_0 from their stationary distribution at each Monte Carlo replication, and we use common random numbers whenever appropriate.

4.2. *Operational Considerations*

Next, we detail operational matters relating to our implementation of the various estimators we use in our simulations.

4.2.1. OLS-HAC. OLS-HAC estimation proceeds from the approach previously outlined in Section 2.2; namely

$$T^{1/2}(\hat{\beta}_{OLS} - \beta) \rightarrow N(0, V),$$

where $V = Q^{-1}\Omega Q^{-1}$ and

$$\Omega = \sum_{\tau=-\infty}^{\infty} \Gamma(\tau),$$

where $\Gamma(\tau) = cov(x_t u_t, x_{t-\tau} u_{t-\tau})$, and $\tau = 0, \pm 1, \ldots$

The key object in V is Ω, the spectrum of xu at frequency zero. The OLS-HAC approach uses

$$\widehat{V} = Q^{-1}\widehat{\Omega} Q^{-1},$$

where $\widehat{\Omega}$ is a consistent estimator of Ω and hence $\widehat{V}$ delivers a consistent estimator of V.

A large literature on consistent estimation of Ω can be traced back to at least Hansen and Hodrick (1980). The most popular approach is due to Newey and West (1987), who propose lag-window estimation with linearly decreasing (Bartlett) lag window:

$$\widehat{\Omega} = \left(\frac{1}{T} \sum_{t=1}^{T} (x_t x_t') \hat{u}_t^2 + \sum_{\tau=1}^{h} \left(1 - \frac{\tau}{h+1} \right) (\widehat{\boldsymbol{\Gamma}}_\tau + \widehat{\boldsymbol{\Gamma}}_{-\tau}) \right), \tag{4.5}$$

where

$$\widehat{\boldsymbol{\Gamma}}_\tau = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t x_t x_{t-\tau}' \hat{u}_{t-\tau},$$

the $\hat{u}_t$ are OLS regression residuals, and T is sample size. Indeed, many leading HAC estimators are of the form (4.5), distinguished only by their choice of truncation lag h.

We will explore several leading truncation lag choices, including:

(1) NW: Newey–West (4.5) with $h = \lceil (T/100)^{2/9} \rceil$. This h choice is a standard textbook recommendation (e.g., Wooldridge, 2015).
(2) NW-A: Newey–West (4.5) with $h = \lceil 0.75 T^{1/3} \rceil$. This h choice is also standard, arising when a formula in Andrews (1991) is specialised to the case of a first-order autoregression with coefficient 0.25.
(3) NW-LLSW: Newey–West (4.5) with $h = \lceil 1.3 T^{1/2} \rceil$, as proposed by Lazarus et al. (2018). Its use of $T^{1/2}$ rather than $T^{2/9}$ or $T^{1/3}$ as in NW or NW-A, respectively, produces higher truncation lags. For example, if $T = 200$, then NW selects $h = 5$, but NW-LLSW selects $h = 19$.
(4) NW-KV: Newey–West (4.5) with $h = T$, as proposed by Kiefer and Vogelsang (2002), which builds on Kiefer et al. (2000). Setting $h = T$ is of course the maximum possible truncation lag.

We will also explore the Müller (2014) HAC estimator (we denote it by M), which is not in the Newey–West family. Instead, it is an orthogonal series estimator, that uses a type-II discrete cosine transform to produce an equally weighted average of projections on cosines. The M estimator is:

$$\widehat{\Omega} = \frac{1}{\nu}\sum_{j=1}^{\nu}\widehat{\Lambda}_j\widehat{\Lambda}_j',$$

where

$$\widehat{\Lambda}_j = \sqrt{\frac{2}{T}}\sum_{t=1}^{T}(x_t\hat{u}_t)\cos\left(\pi j\left(\frac{t-1/2}{T}\right)\right).$$

The M truncation parameter, ν, is the total number of cosines included in the average projection. Lazarus et al. (2018) suggest setting $\nu = \lfloor 0.4T^{2/3}\rfloor$, producing the M-LLSW estimator.

4.2.2. FGLS and FGLS-D. If the data follow the DGP in (2.1), namely $y_t = x_t\beta + u_t$, and there exists a known lag operator polynomial (filter) $\Phi(L)$ that reduces u_t to white noise ε_t (i.e., $\Phi(L)u_t = \varepsilon_t$), then GLS estimation of β is appropriate, and it amounts to running an OLS regression on transformed data. Specifically, one regresses $\tilde{y}_t$ on $\tilde{x}_t$, where $\tilde{y}_t = \Phi(L)y_t$ and $\tilde{x}_t = \Phi(L)x_t$.

In practice, however, $\Phi(L)$ is unknown and needs to be approximated. The FGLS estimator uses $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ and proceeds as follows:

(1) Run an OLS regression of y_t on x_t, and obtain the residuals $\hat{u}_t$.
(2) Fit an AR(p) model to $\hat{u}_t$ (in particular, run an OLS regression of $\hat{u}_t$ on $\hat{u}_{t-1}, \ldots, \hat{u}_{t-p}$, with p selected by Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC)), and obtain the coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_p$.
(3) Construct the transformed data,

$$\tilde{x}_t = x_t - \hat{\phi}_1 x_{t-1} - \ldots - \hat{\phi}_p x_{t-p}$$
$$\tilde{y}_t = y_t - \hat{\phi}_1 y_{t-1} - \ldots - \hat{\phi}_p y_{t-p}.$$

(4) Run an OLS regression of $\tilde{y}_t$ on $\tilde{x}_t$ to obtain the FGLS estimator of β.

The FGLS-D estimator relies on a different first-stage procedure, replacing the regressions in steps 1 and 2 above with a single DURBIN regression, proceeding as follows:

(1) Run the OLS DURBIN regression (with p selected by AIC or BIC),

$$y_t = \sum_{j=1}^{p}\varphi_j y_{t-j} + \sum_{i=1}^{k}\beta_i x_{i,t} + \sum_{j=1}^{p}\sum_{i=1}^{k}\gamma_{i,j}x_{i,t-j} + \varepsilon_t.$$

(2) Use the estimated coefficients on the lags of y_t, $\hat{\varphi}_1, \ldots, \hat{\varphi}_p$, to construct the transformed data,

$$\tilde{x}_t = x_t - \hat{\varphi}_1 x_{t-1} - \ldots - \hat{\varphi}_p x_{t-p}$$
$$\tilde{y}_t = y_t - \hat{\varphi}_1 y_{t-1} - \ldots - \hat{\varphi}_p y_{t-p}.$$

(3) Run the OLS regression of $\tilde{y}_t$ on $\tilde{x}_t$ to obtain the FGLS-D estimator of β.

4.2.3. DURBIN. As previously noted, the DURBIN regression augments regression (2.1) with lags of y and x to capture dynamics, very much in the spirit of an arbitrary equation in a vector autoregression, as suggested by Durbin (1970).[11] The p^{th}-order DURBIN regression is

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + \varepsilon_t, \tag{4.6}$$

which has $p + k + kp$ parameters.

If u_t in (2.1) is a finite-ordered AR(p) process with p known, then DURBIN holds exactly. In particular, we have[12]

$$\begin{aligned} y_t &= \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \beta_i \phi_j x_{i,t-j} + \varepsilon_t \\ &= \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + \varepsilon_t . \end{aligned} \tag{4.7}$$

Hence the usual asymptotic inference is immediately available:

$$T^{1/2}(\widehat{\vartheta}_{OLS} - \vartheta) \rightarrow N(0, Q^{-1}), \tag{4.8}$$

where ϑ_{OLS} is the vector of DURBIN parameters,

$$Q = plim\left(T^{-1} \sum_{t=1}^{T} z_t z_t'\right),$$

and $z_t' = \left(y_{t-1}, \ldots, y_{t-p},\ x_{1,t}, \ldots, x_{k,t},\ x_{1,t-1}, \ldots, x_{k,t-1},\ \ldots,\ x_{1,t-p}, \ldots, x_{k,t-p}\right)$.

In the more compelling case, where p is *un*known and must be selected (implemented in our Monte Carlo below), the DURBIN regression (4.6) is approximate rather than exact. However, the limiting distribution (4.8) remains valid if p is selected suitably (Grenander, 1981; Hannan and Deistler, 1988), as achieved by standard criteria with well-known optimality properties.[13] In particular, if a p_{max} is known such that $p \leq p_{max}$, then a consistent selection criterion (in the model selection sense) like BIC is a natural choice. Alternatively, in the absence of a p_{max}, an efficient selection criterion (in the model selection sense) like AIC is a natural choice.[14]

4.3. Estimation Accuracy

We first examine the accuracy of our four estimators (OLS, FGLS, FGLS-D, and DURBIN) under our four DGPs (*NDY* + *BD*, *NDY* + *GEXOG*, *NDY* + *EBD*, *BD*). The key object of interest is

[11] Also related is the important recent work of Montiel Olea and Plagborg-Møller (2021), who study lag-augmented local projection estimators of impulse-response functions in vector autoregressions.

[12] Note that DURBIN does *not* impose the common factor restriction embedded in (4.7), namely that $\gamma_{i,j} = \beta_i \phi_j\ \forall i, j$, in which case DURBIN coincides with FGLS. See Sargan (1964) and Hendry and Mizon (1978).

[13] OLS-HAC regression, in contrast, typically relies on one or another of various 'rules of thumb' for bandwidth (truncation, h or ν) selection. 'Automatic' bandwidth selection has, however, been considered in Andrews–Newey–West environments by Andrews (1991), Andrews and Monahan (1992), and Newey and West (1994), among others.

[14] In the Gaussian case, we have BIC $= T\log(\text{SSE}) + \log(T)(p + k + kp)$ and AIC $= T\log(\text{SSE}) + 2(p + k + kp)$, where SSE is the DURBIN regression sum of squared errors.

RE_{est}, the efficiency of DURBIN relative to OLS, FGLS or FGLS-D. For example:

$$\text{RE}_{\text{est}}(\text{OLS}) = \frac{\text{MSE(OLS)}}{\text{MSE(DURBIN)}}.$$

We also show MSE and bias.[15]

4.3.1. Autoregressive disturbances DGP (NDY + BD). Results appear in Table 2. Let us begin directly with the RE_{est} results for DURBIN relative to OLS. For any fixed sample size T, RE_{est} is increasing in serial correlation strength ρ. Consider, for example, a leading case like $T = 200$ corresponding, to fifty years of quarterly data. For $\rho = 0$, RE_{est} is close to 1, as it should be since there is no serial correlation. RE_{est} grows quickly as ρ increases, however, reaching 2.9 when $\rho = 0.7$ and 36.3 when $\rho = 0.95$.

In contrast, for any fixed serial correlation strength ρ, RE_{est} stabilises quickly in sample size T and remains approximately constant. Consider, for example, a realistic case like $\rho = 0.9$. RE_{est} remains at approximately $\text{RE}_{\text{est}} = 12$ for all sample sizes $T \in \{50, 200, 600, 2500\}$. Hence RE_{est} is clearly driven by serial correlation strength and not by sample size.

In Figure 1 we provide a visual representation of the RE_{est} of DURBIN relative to OLS presented in Table 2. It reveals clearly that RE_{est} is driven entirely by the degree of serial correlation and not by sample size.

Now consider separately the MSEs for OLS and DURBIN that underlie RE_{est}. For any fixed sample size T, the MSE of OLS is strongly increasing in serial correlation strength ρ (because the OLS estimator ignores serial correlation), whereas the MSE from DURBIN is invariant to serial correlation strength (because the DURBIN estimator controls for serial correlation). That is why the RE_{est} ratio is also strongly increasing in ρ, as documented earlier. In contrast, for any fixed serial correlation strength ρ, the MSEs for *both* OLS and DURBIN decrease with sample size T (as they must, since both OLS and DURBIN are consistent), but they decrease proportionately, so that the RE_{est} ratio is invariant to T, as documented earlier.

Next, let us examine the bias and variance components that underlie the MSEs. First consider bias. Both the OLS and DURBIN estimators are theoretically unbiased for any serial correlation strength and sample size, and the Monte Carlo confirms the theory: the estimated biases are always negligible and invariant to ρ.[16] Moreover, given the scale of the bias, the patterns mentioned above for MSE will correspond to patterns in variance: OLS variance increases sharply with serial correlation strength (because OLS ignores serial correlation), whereas DURBIN variance does not (because DURBIN controls for serial correlation), and both variances decrease with sample size (by consistency), but they do so proportionately. That is, the MSE patterns between OLS and DURBIN, and hence the corresponding RE_{est} patterns, are driven entirely by variance.

4.3.2. Triangular and unrestricted VAR DGPs (NDY + GEXOG, NDY + EBD). Results appear in Table 3. Ψ_1 and Ψ_1^* correspond to different parameterisations of the *NDY + GEXOG* DGP, and Ψ_2 and Ψ_2^* correspond to different parameterisations of the *NDY + EBD* DGP.[17] For all sample sizes, OLS and FGLS exhibit large bias and MSE, which is expected since they are indeed inconsistent under both *NDY + GEXOG* and *NDY + EBD*. As a result, the RE_{est}'s for DURBIN relative to OLS and FGLS in Table 3 are very large: DURBIN dominates both.

[15] Note that all OLS-HAC estimators simply use the OLS estimator of β. Particular HAC estimators will have particular effects on the *standard errors* of $\hat{\beta}$, but not on $\hat{\beta}$ itself, which always remains just $\hat{\beta}_{OLS}$.

[16] Moreover, the estimated biases decrease with T, as expected, by consistency.

[17] Recall that Ψ_1^* has a larger leading eigenvalue than does Ψ_1, and Ψ_2^* has a larger leading eigenvalue than Ψ_2.

Table 2. Bias, MSE, and relative efficiency estimators: OLS, FGLS, FGLS-D, DURBIN, DGP: autoregressive disturbances, $NDY + BD$.

		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
T = 50								
Bias	OLS	0.0012	0.0011	0.0011	0.0013	0.0006	– 0.0038	– 0.0270
	FGLS	0.0010	0.0005	0.0000	– 0.0002	– 0.0012	– 0.0008	– 0.0034
	FGLS-D	0.0013	0.0005	– 0.0001	– 0.0001	0.0000	0.0000	– 0.0001
	DURBIN	0.0005	– 0.0004	– 0.0005	0.0001	0.0002	– 0.0001	0.0001
MSE	OLS	0.0114	0.0185	0.0295	0.0589	0.3019	1.3097	78.5014
	FGLS	0.0124	0.0181	0.0220	0.0241	0.0246	0.0308	0.2431
	FGLS-D	0.0116	0.0186	0.0225	0.0232	0.0207	0.0197	0.0187
	DURBIN	0.0131	0.0207	0.0238	0.0232	0.0229	0.0230	0.0230
RE_{est}	OLS	0.8754	0.8922	1.2407	2.5407	13.1694	56.8642	3414.0760
	FGLS	0.9449	0.8755	0.9260	1.0405	1.0743	1.3379	10.5731
	FGLS-D	0.8889	0.8975	0.9481	1.0004	0.9009	0.8534	0.8136
T = 200								
		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
Bias	OLS	0.0004	0.0007	0.0010	0.0017	0.0034	0.0043	0.0023
	FGLS	0.0003	0.0005	0.0006	0.0007	0.0006	0.0006	0.0004
	FGLS-D	0.0004	0.0006	0.0007	0.0007	0.0006	0.0005	0.0005
	DURBIN	0.0005	0.0007	0.0006	0.0006	0.0005	0.0005	0.0005
MSE	OLS	0.0026	0.0044	0.0071	0.0147	0.0641	0.1859	9.2794
	FGLS	0.0027	0.0040	0.0048	0.0052	0.0047	0.0046	0.0048
	FGLS-D	0.0026	0.0040	0.0048	0.0051	0.0047	0.0045	0.0043
	DURBIN	0.0027	0.0051	0.0051	0.0051	0.0051	0.0051	0.0051
RE_{est}	OLS	0.9564	0.8625	1.3966	2.8777	12.5220	36.3137	1815.8133
	FGLS	0.9759	0.7859	0.9416	1.0073	0.9242	0.8920	0.9410
	FGLS-D	0.9579	0.7895	0.9378	1.0010	0.9163	0.8777	0.8454
T = 600								
		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
Bias	OLS	0.0001	0.0001	0.0000	– 0.0004	– 0.0011	– 0.0014	0.0002
	FGLS	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004
	FGLS-D	0.0001	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004
	DURBIN	0.0001	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
MSE	OLS	0.0009	0.0014	0.0024	0.0048	0.0201	0.0494	1.1912
	FGLS	0.0009	0.0013	0.0016	0.0017	0.0016	0.0015	0.0015
	FGLS-D	0.0009	0.0013	0.0016	0.0017	0.0016	0.0015	0.0014
	DURBIN	0.0009	0.0017	0.0017	0.0017	0.0017	0.0017	0.0017
RE_{est}	OLS	0.9855	0.8593	1.4071	2.8956	12.0262	29.4751	710.3845
	FGLS	0.9891	0.7636	0.9275	0.9990	0.9285	0.8933	0.8884
	FGLS-D	0.9857	0.7631	0.9261	0.9977	0.9269	0.8911	0.8590
T = 2,500								
		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
Bias	OLS	– 0.0002	– 0.0003	– 0.0004	– 0.0007	– 0.0011	– 0.0011	– 0.0014
	FGLS	– 0.0002	– 0.0003	– 0.0004	– 0.0005	– 0.0005	– 0.0005	– 0.0005
	FGLS-D	– 0.0002	– 0.0003	– 0.0004	– 0.0005	– 0.0005	– 0.0005	– 0.0005
	DURBIN	– 0.0002	– 0.0005	– 0.0005	– 0.0005	– 0.0005	– 0.0005	– 0.0005

Table 2. Continued

		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
T = 2,500								
MSE	OLS	0.0002	0.0003	0.0006	0.0012	0.0049	0.0109	0.1118
	FGLS	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0003
	FGLS-D	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0003
	DURBIN	0.0002	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	0.9958	0.8683	1.4285	2.9573	12.1117	27.2062	279.2032
	FGLS	0.9974	0.7675	0.9319	1.0002	0.9232	0.8858	0.8562
	FGLS-D	0.9957	0.7673	0.9316	1.0001	0.9231	0.8859	0.8533

Notes: All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

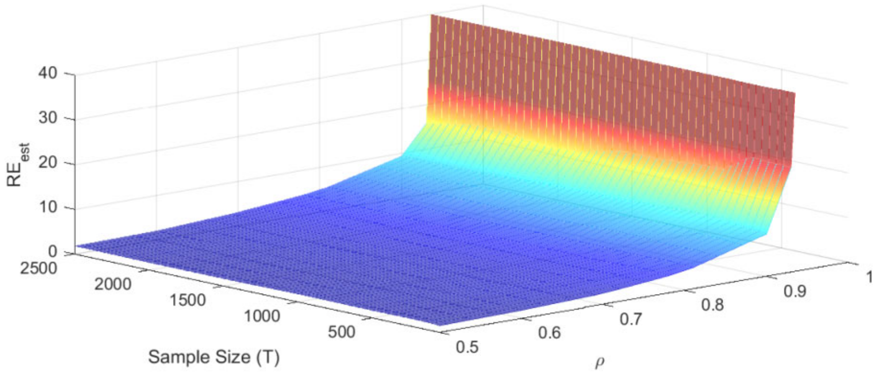

Figure 1. Efficiency of DURBIN relative to OLS, DGP: autoregressive disturbances, $NDY + BD$. All shocks are $N(0, 1)$ white noise. We select DURBIN lag order using BIC. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. We do not plot values for $\rho = 0.99$, due to their extreme magnitude as shown in Table 2. See text for details.

4.3.3. Dynamic regression DGP (EBD). Results appear in Table 4. In the *EBD* case, OLS, FGLS, and FGLS-D are in general inconsistent, whereas DURBIN remains consistent. This is reflected in the large biases and MSEs of the other estimators compared to DURBIN, and hence the high efficiency of DURBIN relative to OLS and FGLS.

A notable exception is when $\rho = 0.5$, in which case the common factor restriction holds, so that it is possible to write the dynamic regression as a single-regressor equation (with just x_t) and a disturbance with AR(1) serial correlation. Put differently, in this case the DGP in (4.4) can be rewritten in the form of (4.1), so that FGLS and FGLS-D are consistent and efficient and should have lower MSE than DURBIN. Table 4 shows that this is the case for all sample sizes. This result highlights the role that the common factor restriction plays; if it holds, it guarantees that all dynamics enter through the disturbance term, so that FGLS and FGLS-D dominate DURBIN, but if it does not hold (and there is no reason why it should hold), DURBIN dominates.

Table 3. Bias, MSE, and relative efficiency estimators: OLS, FGLS, FGLS-D, DURBIN, DGPs: (1) triangular VAR, $NDY + GEXOG$, (2) unrestricted VAR, $NDY + EBD$.

		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
T = 50					
Bias	OLS	0.2342	0.3058	0.5257	0.6807
	FGLS	0.1064	0.1493	0.4592	0.6343
	FGLS-D	0.0692	0.0542	0.1886	0.2883
	DURBIN	0.0619	0.0405	0.0276	0.0070
MSE	OLS	0.0691	0.1087	0.2959	0.4797
	FGLS	0.0286	0.0409	0.2473	0.4343
	FGLS-D	0.0391	0.0373	0.0979	0.1654
	DURBIN	0.0413	0.0397	0.0414	0.0298
RE_{est}	OLS	1.6710	2.7379	7.1500	16.1016
	FGLS	0.6917	1.0301	5.9738	14.5764
	FGLS-D	0.9446	0.9401	2.3660	5.5518
T = 200					
		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Bias	OLS	0.2452	0.3194	0.5630	0.7253
	FGLS	0.1012	0.1436	0.5047	0.6994
	FGLS-D	0.0036	0.0047	0.1959	0.3438
	DURBIN	0.0010	0.0004	0.0005	0.0005
MSE	OLS	0.0636	0.1058	0.3216	0.5294
	FGLS	0.0145	0.0253	0.2651	0.4957
	FGLS-D	0.0053	0.0052	0.0558	0.1469
	DURBIN	0.0053	0.0051	0.0051	0.0051
RE_{est}	OLS	11.9309	20.6053	62.6078	102.9589
	FGLS	2.7240	4.9382	51.6035	96.4078
	FGLS-D	0.9983	1.0037	10.8684	28.5674
T = 600					
		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Bias	OLS	0.2462	0.3209	0.5713	0.7383
	FGLS	0.0988	0.1408	0.5036	0.7180
	FGLS-D	0.0014	0.0019	0.2013	0.3671
	DURBIN	0.0005	0.0005	0.0004	0.0004
MSE	OLS	0.0618	0.1042	0.3279	0.5461
	FGLS	0.0112	0.0214	0.2584	0.5176
	FGLS-D	0.0017	0.0017	0.0468	0.1461
	DURBIN	0.0017	0.0017	0.0017	0.0017

Table 3. Continued

		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
T = 600					
RE_{est}	OLS	36.8447	62.1620	195.6287	325.6477
	FGLS	6.6678	12.7621	154.1354	308.6328
	FGLS-D	1.0178	1.0135	27.9311	87.1036
T = 2,500					
Bias	OLS	0.2468	0.3217	0.5748	0.7435
	FGLS	0.0977	0.1396	0.4643	0.7209
	FGLS-D	– 0.0003	– 0.0001	0.2020	0.3756
	DURBIN	– 0.0005	– 0.0005	– 0.0005	– 0.0005
MSE	OLS	0.0612	0.1038	0.3307	0.5531
	FGLS	0.0099	0.0199	0.2168	0.5204
	FGLS-D	0.0004	0.0004	0.0424	0.1439
	DURBIN	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	152.7636	259.1858	826.1758	1,380.9176
	FGLS	24.6839	49.6294	541.6495	1,299.4068
	FGLS-D	1.0260	1.0150	105.8071	359.3683

Notes: All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

4.4. *Prediction Accuracy*

One of the primary uses of regression and dynamic regression is for ex ante prediction. There is substantial previous literature related to the task of prediction. In particular, Baillie (1979) has considered the situation of predictions from the regression model with AR(p) errors and the properties of prediction from static regressions and also with optimal multi-step predictions in the sense of minimum MSE predictions. Baillie (1979) also derived results on the efficiency of these predictors with and without estimated parameters. One conclusion concerns the importance of including the full effects of dynamics from the AR(p) regression model in the predictor. In this case, the complete structural dynamic predictor generally has substantial asymptotic and small sample efficiency gains over predictors from static regressions. Similar effects and properties are found in more complicated dynamic models such as the DGP considered in Section 3 of this paper.

We now consider one-step-ahead predictions relying on the OLS and DURBIN estimation strategies. The results reflect that an explicit modelling of autocorrelation can be used for improved prediction. OLS estimators neglect this and therefore produce suboptimal predictions. To see this, first consider the case of a DGP with autoregressive disturbances and known parameter $\beta = 1$.[18]

[18] We start with the case of known parameter β, as it can easily be solved analytically.

Table 4. Bias, MSE, and relative efficiency estimators: OLS, FGLS, FGLS-D, DURBIN, DGP: dynamic regression, *EBD*.

		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
T = 50						
Bias	OLS	– 0.3355	0.0010	0.2569	0.6915	0.8533
	FGLS	– 0.3362	0.0001	0.0343	– 0.0970	– 0.1550
	FGLS-D	– 0.3334	0.0000	0.0182	– 0.1312	– 0.1895
	DURBIN	– 0.0479	– 0.0004	0.0000	0.0004	0.0002
MSE	OLS	0.1274	0.0299	0.1319	0.8955	1.8657
	FGLS	0.1300	0.0222	0.0301	0.0423	0.0572
	FGLS-D	0.1281	0.0226	0.0271	0.0399	0.0563
	DURBIN	0.0444	0.0240	0.0232	0.0230	0.0231
RE_{est}	OLS	2.8674	1.2501	5.6956	38.9434	80.9010
	FGLS	2.9255	0.9249	1.2996	1.8379	2.4803
	FGLS-D	2.8815	0.9438	1.1702	1.7344	2.4418
T = 200						
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	– 0.3461	0.0009	0.2706	0.7405	0.9198
	FGLS	– 0.3459	0.0006	0.0087	– 0.1362	– 0.1897
	FGLS-D	– 0.3453	0.0005	0.0052	– 0.1419	– 0.1946
	DURBIN	0.0004	0.0005	0.0005	0.0006	0.0006
MSE	OLS	0.1230	0.0072	0.0901	0.6718	1.1902
	FGLS	0.1233	0.0048	0.0062	0.0239	0.0408
	FGLS-D	0.1228	0.0048	0.0060	0.0251	0.0425
	DURBIN	0.0051	0.0051	0.0051	0.0051	0.0051
RE_{est}	OLS	23.9427	1.3956	17.5892	131.0656	232.2652
	FGLS	24.0002	0.9424	1.2036	4.6640	7.9642
	FGLS-D	23.9027	0.9384	1.1677	4.9018	8.2866
T = 600						
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	– 0.3488	0.0000	0.2726	0.7500	0.9323
	FGLS	– 0.3489	0.0004	0.0033	– 0.1424	– 0.1942
	FGLS-D	– 0.3486	0.0004	0.0023	– 0.1440	– 0.1956
	DURBIN	0.0004	0.0005	0.0005	0.0005	0.0004
MSE	OLS	0.1227	0.0024	0.0799	0.6047	0.9892
	FGLS	0.1229	0.0016	0.0020	0.0220	0.0393
	FGLS-D	0.1227	0.0016	0.0019	0.0224	0.0398
	DURBIN	0.0017	0.0017	0.0017	0.0017	0.0017

Table 4. Continued

T = 600						
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
RE_{est}	OLS	73.2154	1.4047	47.6893	360.9901	590.2263
	FGLS	73.3214	0.9270	1.1709	13.1098	23.4335
	FGLS-D	73.2098	0.9260	1.1527	13.3502	23.7366
T = 2,500						
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	– 0.3499	– 0.0004	0.2733	0.7541	0.9377
	FGLS	– 0.3499	– 0.0004	0.0001	– 0.1454	– 0.1966
	FGLS-D	– 0.3498	– 0.0004	– 0.0001	– 0.1457	– 0.1969
	DURBIN	– 0.0005	– 0.0005	– 0.0005	– 0.0005	– 0.0005
MSE	OLS	0.1227	0.0006	0.0761	0.5790	0.9085
	FGLS	0.1227	0.0004	0.0005	0.0215	0.0390
	FGLS-D	0.1226	0.0004	0.0005	0.0216	0.0391
	DURBIN	0.0004	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	306.1939	1.4289	189.9424	1,445.9862	2,267.9542
	FGLS	306.2177	0.9319	1.1363	53.7495	97.4095
	FGLS-D	306.0911	0.9316	1.1296	53.9678	97.6422

Notes: All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Specifically consider the DGP given by

$$\begin{aligned} y_t &= x_t + u_t \\ x_t &= \rho x_{t-1} + \epsilon_{x,t} \\ u_t &= \rho u_{t-1} + \epsilon_{u,t}, \end{aligned}$$

with all shocks $N(0, 1)$ and orthogonal at all leads and lags. For this DGP, the optimal prediction accounting for serial correlation in u is

$$\begin{aligned} y^{opt}_{t+1,t} &= x_{t+1,t} + u_{t+1,t} \\ &= \rho x_t + \rho u_t, \end{aligned} \tag{4.9}$$

and the corresponding prediction error is $e^{opt}_{t+1} = \varepsilon_{x,t+1} + \varepsilon_{u,t+1}$, with variance $\sigma^2_{opt} = 2$.

The suboptimal prediction, failing to account for serial correlation in u, is just the first term in (4.9),

$$y^{subopt}_{t+1,t} = \rho x_t,$$

with corresponding prediction error $e^{subopt}_{t+1} = \varepsilon_{x,t+1} + u_{t+1}$, and variance $\sigma^2_{subopt} = 1 + \frac{1}{1-\rho^2}$.

Table 5. Prediction efficiency of DURBIN relative to OLS, DGP: autoregressive disturbances, $NDY + EBD$.

	Relative Prediction Efficiency (RE_{pred})						
T	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
50	0.995	1.037	1.142	1.428	2.932	5.745	159.490
200	0.998	1.047	1.164	1.473	3.077	5.537	49.223
600	0.999	1.044	1.157	1.466	3.067	5.452	25.423
2,500	1.000	1.047	1.161	1.477	3.148	5.627	25.318

Notes: All shocks are $N(0, 1)$ white noise. RE_{pred} is the relative predictive efficiency of DURBIN, $RE_{pred} = MSPE(OLS)/MSPE(DURBIN)$, where MSPE is 1-step-ahead mean squared prediction error. We select the DURBIN lag order using BIC. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Both predictions are unbiased, so the prediction efficiency of DURBIN relative to OLS (RE_{pred}) is just the relative variance, which is

$$RE_{pred} = \frac{\sigma^2_{subopt}}{\sigma^2_{opt}} = \frac{1}{2} + \frac{1}{2(1-\rho^2)}. \tag{4.10}$$

RE_{pred} is bounded below by 1, which occurs when $\rho = 0$, and $RE_{pred} \to \infty$ monotonically as $\rho \to 1$.

Now we consider the case of estimated parameters, which is more complicated. In Table 5 we show RE_{pred} estimated by Monte Carlo, accounting for parameter estimation uncertainty. For all but the most extreme cases (e.g., $T = 50$ with $\rho = 0.99$) the Monte Carlo results are almost identical to the analytic result (4.10) that ignores parameter estimation uncertainty.[19] Hence RE_{pred} depends strongly on ρ, but not on T. More precisely, for any T we of course obtain $RE_{pred} = 1$ in the white noise case ($\rho = 0$), but then RE_{pred} grows quickly in ρ, and for any ρ, RE_{pred} stabilises extremely quickly in T and is basically constant.

4.5. Inference

Now we consider the finite-sample properties of hypothesis tests associated with the various estimation procedures. We first consider test sizes, after which we consider rejection frequencies. In all tables in this section we consider the following estimators: OLS with unadjusted standard errors, five OLS-HAC estimators (NW, NW-A, NW-LLSW, NW-KV, and M-LLSW), FGLS, FGLS-D, and two implementations of DURBIN, one using BIC for lag order selection and the other using AIC. Additionally, we have included two Hausman tests; the first null hypothesis is that FGLS is efficient relative to OLS, and the second is that FGLS-D is efficient relative to DURBIN.

4.5.1. Size. Table 6 contains results for the autoregressive disturbances DGP, $NDY + BD$. First, tests based on OLS are incorrectly sized for all (ρ, T) combinations, except when $\rho = 0$, and the size distortions become huge as ρ grows. Second, the various NW HAC corrections reduce, but

[19] This is because the effects of parameter estimation uncertainty on MSPE vanish quickly (like $1/T$ rather than $1/\sqrt{T}$), as is well known. Hence the earlier-documented poor estimation efficiency of OLS relative to DURBIN, although a large problem for some purposes, is not an important problem for prediction.

Table 6. Empirical size of nominal 5% t-test of $H_0 : \beta = 1$, DGP: autoregressive disturbances, $NDY + BD$.

T = 50

	Truncation	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
OLS	–	0.054	0.115	0.167	0.242	0.346	0.386	0.409
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.067	0.091	0.113	0.141	0.197	0.223	0.267
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.067	0.095	0.124	0.159	0.228	0.261	0.298
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.068	0.085	0.095	0.108	0.122	0.140	0.202
NW-KV	$h = T$	0.063	0.079	0.089	0.094	0.093	0.087	0.159
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.069	0.077	0.081	0.084	0.088	0.105	0.165
FGLS	BIC	0.074	0.078	0.079	0.077	0.074	0.077	0.054
FGLS-D	BIC	0.057	0.103	0.089	0.069	0.056	0.055	0.053
DURBIN	BIC	0.061	0.104	0.080	0.056	0.054	0.055	0.056
DURBIN	AIC	0.087	0.099	0.081	0.077	0.079	0.081	0.079
Hausman 1	OLS vs FGLS		0.740	0.639	0.452	0.252	0.240	0.276
Hausman 2	DURBIN vs FGLS-D		0.046	0.087	0.116	0.122	0.113	0.099

T = 200

	Truncation	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
OLS	–	0.051	0.114	0.174	0.251	0.352	0.382	0.403
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.056	0.071	0.082	0.108	0.142	0.154	0.172
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.056	0.071	0.082	0.108	0.142	0.154	0.172
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.058	0.064	0.065	0.074	0.078	0.073	0.084
NW-KV	$h = T$	0.055	0.059	0.062	0.065	0.058	0.043	0.028
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.060	0.062	0.062	0.067	0.065	0.060	0.078
FGLS	BIC	0.055	0.058	0.057	0.054	0.050	0.051	0.047
FGLS-D	BIC	0.052	0.063	0.056	0.052	0.049	0.049	0.049
DURBIN	BIC	0.054	0.061	0.050	0.050	0.049	0.049	0.049
DURBIN	AIC	0.066	0.054	0.052	0.053	0.052	0.051	0.053
Hausman 1	OLS vs FGLS		0.623	0.443	0.216	0.123	0.112	0.122
Hausman 2	DURBIN vs FGLS-D		0.047	0.060	0.095	0.067	0.063	0.061

T=600

	Truncation	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
OLS	–	0.050	0.112	0.172	0.248	0.347	0.376	0.399
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.051	0.062	0.072	0.087	0.113	0.125	0.136
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.051	0.062	0.070	0.080	0.104	0.113	0.123
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.054	0.058	0.057	0.060	0.066	0.065	0.051
NW-KV	$h = T$	0.054	0.053	0.054	0.053	0.050	0.047	0.020
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.054	0.057	0.056	0.058	0.058	0.057	0.052
FGLS	BIC	0.050	0.051	0.052	0.048	0.048	0.048	0.049
FGLS-D	BIC	0.050	0.051	0.051	0.048	0.047	0.049	0.048
DURBIN	BIC	0.050	0.047	0.047	0.048	0.048	0.048	0.047
DURBIN	AIC	0.061	0.048	0.048	0.049	0.048	0.049	0.049
Hausman 1	OLS vs FGLS		0.533	0.299	0.116	0.088	0.083	0.069
Hausman 2	DURBIN vs FGLS-D		0.051	0.054	0.091	0.056	0.055	0.053

T = 2,500

	Truncation	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
OLS	–	0.050	0.118	0.181	0.255	0.357	0.385	0.405
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.051	0.058	0.065	0.074	0.089	0.098	0.100
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.051	0.057	0.062	0.070	0.081	0.088	0.090
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.052	0.053	0.053	0.056	0.057	0.060	0.043
NW-KV	$h = T$	0.051	0.051	0.052	0.053	0.050	0.049	0.028
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.051	0.052	0.052	0.055	0.054	0.055	0.046
FGLS	BIC	0.052	0.050	0.050	0.050	0.049	0.049	0.047
FGLS-D	BIC	0.050	0.050	0.050	0.050	0.049	0.049	0.048
DURBIN	BIC	0.051	0.051	0.050	0.050	0.050	0.050	0.050
DURBIN	AIC	0.062	0.050	0.050	0.050	0.050	0.050	0.050
Hausman 1	OLS vs FGLS		0.423	0.156	0.072	0.068	0.067	0.052
Hausman 2	DURBIN vs FGLS-D		0.050	0.050	0.087	0.054	0.053	0.053

Notes: All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Table 7. Empirical size of nominal 5% t-test of $H_0 : \beta = 1$, DGPs: (1) triangular VAR, *NDY* + *GEXOG*, (2) unrestricted VAR, *NDY* + *EBD*.

	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
T = 50					
OLS	–	0.605	0.793	0.967	0.993
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.548	0.732	0.953	0.989
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.561	0.744	0.957	0.990
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.493	0.669	0.925	0.980
NW-KV	$h = T$	0.429	0.584	0.870	0.959
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.437	0.606	0.896	0.969
FGLS	BIC	0.219	0.325	0.878	0.960
FGLS-D	BIC	0.319	0.261	0.506	0.632
DURBIN	BIC	0.280	0.211	0.127	0.067
DURBIN	AIC	0.139	0.106	0.082	0.076
Hausman 1	OLS vs FGLS	0.872	0.894	0.698	0.738
Hausman 2	DURBIN vs FGLS-D	0.006	0.008	0.426	0.642
T = 200					
	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
OLS	–	0.990	1.000	1.000	1.000
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.984	0.999	1.000	1.000
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.984	0.999	1.000	1.000
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.973	0.998	1.000	1.000
NW-KV	$h = T$	0.872	0.959	0.999	1.000
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.969	0.997	1.000	1.000
FGLS	BIC	0.459	0.699	1.000	1.000
FGLS-D	BIC	0.132	0.134	0.736	0.906
DURBIN	BIC	0.053	0.050	0.049	0.050
DURBIN	AIC	0.053	0.052	0.053	0.053
Hausman 1	OLS vs FGLS	0.994	0.998	0.517	0.464
Hausman 2	DURBIN vs FGLS-D	0.000	0.000	0.916	0.979
T = 600					
	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
OLS	–	1.000	1.000	1.000	1.000
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	1.000	1.000	1.000
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	1.000	1.000	1.000
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	1.000	1.000	1.000	1.000
NW-KV	$h = T$	0.997	1.000	1.000	1.000
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	1.000	1.000	1.000	1.000
FGLS	BIC	0.836	0.980	1.000	1.000
FGLS-D	BIC	0.130	0.133	0.955	0.997
DURBIN	BIC	0.047	0.047	0.046	0.046
DURBIN	AIC	0.049	0.048	0.048	0.048

Table 7. Continued

T = 600					
	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Hausman 1	OLS vs FGLS	1.000	1.000	0.671	0.278
Hausman 2	DURBIN vs FGLS-D	0.000	0.000	1.000	1.000

T = 2,500					
	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
OLS	–	1.000	1.000	1.000	1.000
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	1.000	1.000	1.000
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	1.000	1.000	1.000
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	1.000	1.000	1.000	1.000
NW-KV	$h = T$	1.000	1.000	1.000	1.000
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	1.000	1.000	1.000	1.000
FGLS	BIC	1.000	1.000	1.000	1.000
FGLS-D	BIC	0.127	0.128	1.000	1.000
DURBIN	BIC	0.050	0.050	0.049	0.049
DURBIN	AIC	0.050	0.050	0.049	0.049
Hausman 1	OLS vs FGLS	1.000	1.000	0.999	0.474
Hausman 2	DURBIN vs FGLS-D	0.000	0.000	1.000	1.000

Notes: All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

do not eliminate the size distortion. In particular, distortion generally remains in the economically crucial region of $\rho \in [0.5, 0.99]$, depending on the sample size and the precise NW version used. NW and NW-A are worst, NW-LLSW are better, and NW-KV is the best. The M-LLSW HAC correction is different in that it exhibits an approximately correct size across (ρ, T) combinations. Finally, tests based on FGLS, FGLS-D and DURBIN, in contrast, are correctly sized for all (ρ, T) combinations, even with extremely strong autocorrelation. This holds regardless of whether DURBIN lag order selection is done with BIC or AIC.

Table 7 contains results for the two VAR DGPs, *NDY* + *GEXOG* and *NDY* + *EBD*. In the *NDY* + *GEXOG* environment, OLS and FGLS are inconsistent, which produces large size distortions. In contrast, DURBIN and FGLS-D are consistent; they should outperform OLS and FGLS, and they do. DURBIN and FLGS-D should perform similarly, and they do. In the *NDY* + *EBD* environment, OLS, FGLS, *and* FGLS-D are inconsistent, and all have large size distortions. DURBIN, however, remains consistent and performs admirably.

Finally, Table 8 contains results for the dynamic regression DGP, *EBD*. In this environment DURBIN should perform well, and it does, whereas all other test sizes are distorted, except at or near the very special common factor case of $\rho = 0.5$.

4.5.2. Power. Only tests that are correctly sized are of real interest, because only correctly sized tests produce trustworthy and interpretable rejections. As we have shown, DURBIN satisfies that requirement, whereas OLS-HAC regression does not. One could simply stop there, but it is of interest to compare rejection frequencies in a few laboratory environments where the DGP is

Table 8. Empirical size of nominal 5% t-test of $H_0 : \beta = 1$, DGP: dynamic regression, *EBD*.

T = 50						
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
OLS	–	0.778	0.168	0.447	0.503	0.400
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.774	0.112	0.312	0.336	0.242
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.778	0.123	0.341	0.376	0.280
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.737	0.093	0.244	0.231	0.155
NW-KV	$h = T$	0.664	0.084	0.205	0.188	0.116
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.689	0.081	0.201	0.179	0.120
FGLS	BIC	0.769	0.078	0.095	0.115	0.174
FGLS-D	BIC	0.763	0.087	0.080	0.122	0.198
DURBIN	BIC	0.207	0.079	0.055	0.053	0.054
DURBIN	AIC	0.101	0.079	0.074	0.076	0.076
Hausman 1	OLS vs FGLS	0.786	0.637	0.599	0.496	0.334
Hausman 2	DURBIN vs FGLS-D	0.772	0.087	0.576	0.930	0.956
T = 200						
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
OLS	–	1.000	0.175	0.822	0.901	0.769
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	0.082	0.657	0.734	0.512
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	0.082	0.657	0.734	0.512
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.999	0.065	0.562	0.627	0.353
NW-KV	$h = T$	0.978	0.062	0.424	0.447	0.243
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.999	0.061	0.532	0.582	0.318
FGLS	BIC	0.999	0.056	0.065	0.410	0.691
FGLS-D	BIC	0.999	0.055	0.060	0.437	0.722
DURBIN	BIC	0.050	0.049	0.049	0.049	0.049
DURBIN	AIC	0.053	0.054	0.053	0.053	0.053
Hausman 1	OLS vs FGLS	0.714	0.441	0.828	0.894	0.667
Hausman 2	DURBIN vs FGLS-D	1.000	0.061	0.975	1.000	1.000
T = 600						
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
OLS	–	1.000	0.172	0.994	0.999	0.987
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	0.072	0.969	0.991	0.921
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	0.070	0.967	0.991	0.915
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	1.000	0.057	0.950	0.985	0.877
NW-KV	$h = T$	1.000	0.053	0.804	0.841	0.639
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	1.000	0.055	0.948	0.982	0.855
FGLS	BIC	1.000	0.052	0.058	0.903	0.996
FGLS-D	BIC	1.000	0.052	0.056	0.912	0.997
DURBIN	BIC	0.046	0.046	0.046	0.047	0.047
DURBIN	AIC	0.047	0.047	0.047	0.048	0.048

Table 8. Continued

T = 600						
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Hausman 1	OLS vs FGLS	0.664	0.297	0.996	1.000	0.989
Hausman 2	DURBIN vs FGLS-D	1.000	0.054	1.000	1.000	1.000
T = 2,500						
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
OLS	–	1.000	0.182	1.000	1.000	1.000
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	0.064	1.000	1.000	1.000
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	0.062	1.000	1.000	1.000
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	1.000	0.053	1.000	1.000	1.000
NW-KV	$h = T$	1.000	0.052	0.997	0.999	0.981
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	1.000	0.052	1.000	1.000	1.000
FGLS	BIC	1.000	0.049	0.058	1.000	1.000
FGLS-D	BIC	1.000	0.049	0.056	1.000	1.000
DURBIN	BIC	0.050	0.050	0.050	0.050	0.050
DURBIN	AIC	0.050	0.050	0.049	0.050	0.050
Hausman 1	OLS vs FGLS	0.643	0.156	1.000	1.000	1.000
Hausman 2	DURBIN vs FGLS-D	1.000	0.049	1.000	1.000	1.000

Notes: All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

known. We do so in Figure 2 for three of our DGPs with $T = 200$ and various persistence parameters, comparing OLS-HAC (Kiefer-Vogelsang, LLSW), FGLS, and FGLS-D.

In the top row of Figure 2 we show rejection frequencies for the autoregressive disturbances environment, *NDY + BD*. All estimators are consistent, and all tests have correct size when $\beta = 1$, i.e., when the true parameter equals its value under the null hypothesis. Moving away from the null, however, it is clear that OLS-HAC power is inferior to that of DURBIN, because OLS is inefficient relative to DURBIN. Moreover, the inferior power performance of OLS-HAC increases with disturbance persistence (ρ), precisely because the relative inefficiency of OLS increases with persistence. Finally, DURBIN, FGLS and FGLS-D have virtually identical power curves.

In the middle row of Figure 2 we show rejection frequencies for the triangular VAR case, *NDY + GEXOG*. OLS-HAC and FGLS are so badly mis-sized that it is not worth discussing them, whereas FGLS-D is asymptotically correctly sized, but is still over-sized for $T = 200$. Only DURBIN is trustworthy. Moving from the middle-left to middle-right panel (higher persistence), the superiority of DURBIN is amplified.

In the bottom row of Figure 2 we show rejection frequencies for the unrestricted VAR case, *NDY + EBD*. FGLS-D fails even asymptotically, so it is not surprising that its finite-sample performance is much worse than in the middle-row triangular VAR *NDY + GEXOG* case. DURBIN, however, remains trustworthy. Moving from the bottom-left to bottom-right panel (higher persistence), the superiority of DURBIN is amplified, just as in the triangular case.

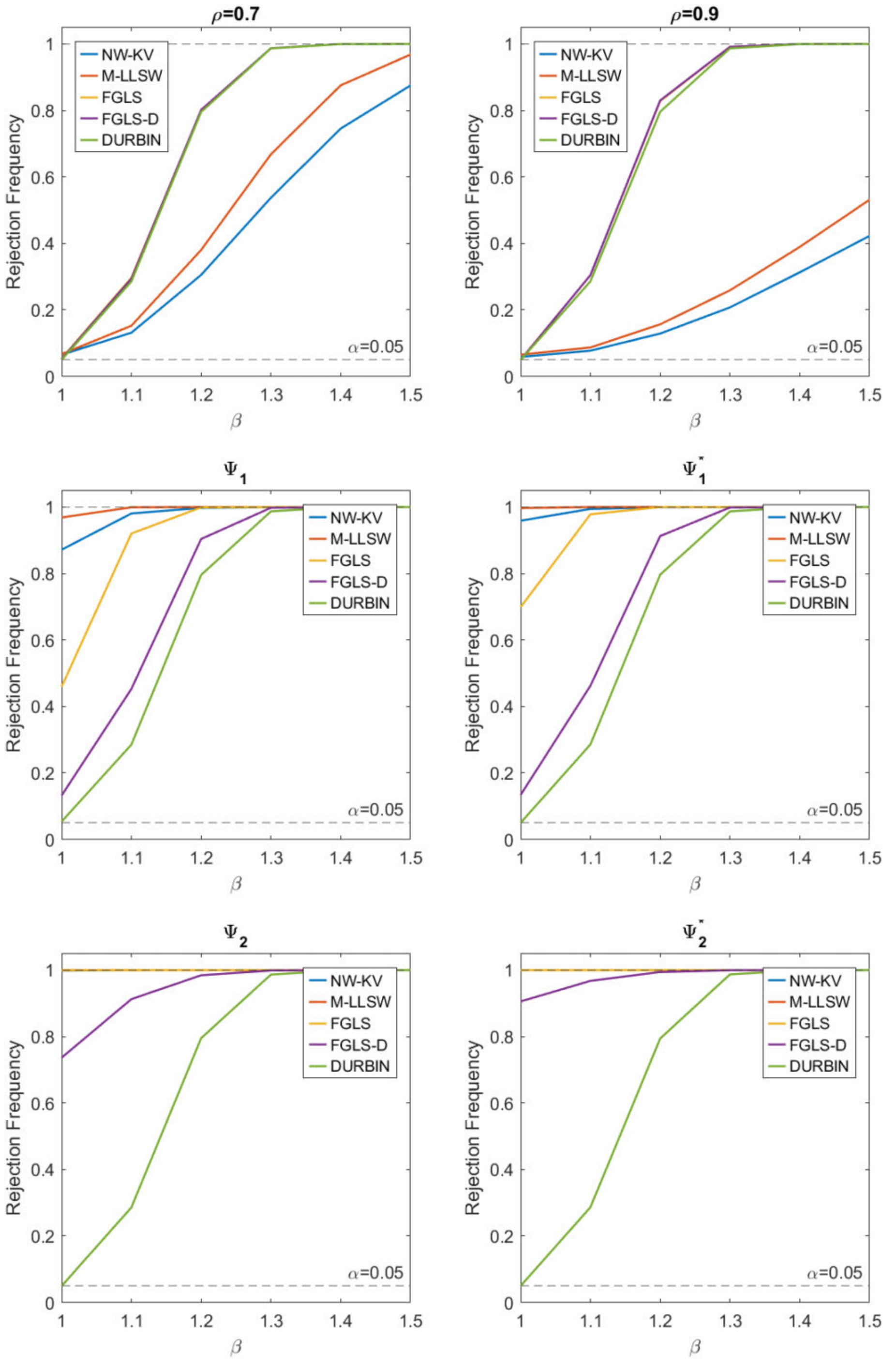

Figure 2. Empirical rejection frequencies of nominal 5% t-tests of H_0: $\beta = 1$, $T = 200$. DGPs: $NDY + BD$ (top row), $NDY + GEXOG$ (middle row), and $NDY + EBD$ (bottom row). See text for details.

5. CONCLUDING REMARKS AND DIRECTIONS FOR FUTURE RESEARCH

We have considered issues surrounding the time-series application of OLS regression with HAC standard errors. Although the OLS-HAC methodology is often sensible in cross-section regression situations, we argued that it is not generally an effective procedure in time-series regressions. Such regressions usually possess persistent autocorrelation, which causes OLS-HAC regressions to be highly suboptimal for parameter estimation (in terms of efficiency), inference (in terms of both test size and power), and prediction.

We showed that the OLS-HAC problems are largely avoided by the use of a simple dynamic regression procedure, DURBIN. We demonstrated the significant advantages of DURBIN with detailed simulations covering a range of practical environments and issues. Effectively, DURBIN is a powerful tool for pre-whitened HAC estimation, in the tradition of Andrews and Monahan (1992)—indeed *such* a good pre-whitening tool that there's rarely any need for subsequent HAC estimation.

However, DURBIN is of course not a panacea. For example, DURBIN may struggle in small samples when dynamics have a strong moving average component. Our Monte Carlo makes clear, however, that for all but the most extreme environments, DURBIN with lag order selected using standard information criteria performs consistently well. Indeed, that is the key message of our paper.

In future work, one could generalise the DURBIN regression in various ways. For example, one could allow different lag lengths for y and the x_i's. One could also allow for heteroskedasticity, which we suppressed in this paper so as to focus exclusively on autocorrelation, for example, by allowing for *GARCH* disturbances in the DURBIN regression.

ACKNOWLEDGEMENTS

For detailed comments we are greatly indebted to the editor, co-editor, and two referees. In addition, we gratefully acknowledge useful discussions and/or comments from Rob Engle, Domenico Giannone, Jim Hamilton, Daniel Lewis, Nour Meddahi, Ulrich Müller, Serena Ng, Lasse Pedersen, Pierre Perron, Peter Phillips, Mikkel Plagborg-Møller, Peter Schmidt, George Tauchen, Tim Volgelsang, Mark Watson, Ken West, and Jeff Wooldridge. We are also grateful to seminar participants at Michigan State University and the University of Pennsylvania, and conference participants at the 2022 NBER Summer Institute, the 2023 joint meetings of the Royal Economic Society and Scottish Economic Society, and the 2023 Copenhagen Conference on Advances in Financial Econometrics. All remaining errors or misunderstandings are ours alone.

REFERENCES

Amemiya, T. (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica 41*(4), 723–32.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–58.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–66.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Baillie, R. T. (1979). The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors. *Journal of the American Statistical Association 74*, 175–84.

Baillie, R. T., F. X. Diebold, G. Kapetanios and K. H. Kim (2023). A new test for market efficiency and uncovered interest parity. *Journal of International Money and Finance 130*, 102765.

Davidson, J. (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics 106*, 243–69.

Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica 38*, 410–21.

Grenander, U. (1981). *Abstract Inference*. New York: Wiley.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory Of Linear Systems*. New York: Wiley.

Hansen, L. P. and R. J. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy 88*(5), 829–53.

Hendry, D. F. and G. E. Mizon (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal 88*(351), 549–63.

Kapetanios, G. and Z. Psaradakis (2016). Semiparametric sieve-type generalized least squares inference. *Econometric Reviews 35*(6), 951–85.

Kiefer, N. M. and T. J. Vogelsang (2002). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica 70*(5), 2093–95.

Kiefer, N. M., T. J. Vogelsang and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Lazarus, E., D. J. Lewis, J. H. Stock and M. W. Watson (2018). Har inference: Recommendations for practice. *Journal of Business and Economic Statistics 36*(4), 541–59.

Lewis, R. and G. C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis 16*(3), 393–411.

Mikusheva, A. and M. Sølvsten (2023). Linear regression with weak exogeneity. *arXiv: Econometrics* 2308.08958.

Montiel Olea, J. L. and M. Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think. *Econometrica 89*(4), 1789–823.

Müller, U. K. (2014). Hac corrections for strongly autocorrelated time series. *Journal of Business and Economic Statistics 32*(3), 311–22.

Müller, U. K. and M. W. Watson (2022). Spatial correlation robust inference. *Econometrica 90*(6), 2901–35.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–8.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*(4), 631–53.

Perron, P. and E. González-Coya (2022). Feasible GLS for time series regression, Ms., Boston University.

Romano, J. P. and M. Wolf (2017). Resurrecting weighted least squares. *Journal of Econometrics 197*, 1–19.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology. In P. E. Hart, G. Mills and J. K. Whitaker (Eds.), *Econometric Analysis for National Economic Planning*, 25–54. New York: Butterworths.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 817–38.

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Boston, MA: Cengage Learning.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Replication Package

Co-editor Jaap Abbring handled this manuscript.

APPENDIX A: ADDITIONAL MONTE CARLO: AR DISTURBANCES

Table A1. Selected lags by test estimators: FGLS, DURBIN AIC, DURBIN BIC, DGP: autoregressive disturbances, $NDY + BD$.

T			$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
50	Median	FGLS	1	1	1	1	1	1	1
		DURBIN BIC	0	0	1	1	1	1	1
		DURBIN AIC	0	1	1	1	1	1	1
	Mean	FGLS BIC	1.2	1.2	1.2	1.2	1.3	1.3	1.5
		DURBIN BIC	0.1	0.4	0.9	1.1	1.1	1.1	1.1
		DURBIN AIC	2.0	2.6	3.0	3.0	3.1	3.1	3.1
200	Median	FGLS	1	1	1	1	1	1	1
		DURBIN BIC	0	1	1	1	1	1	1
		DURBIN AIC	0	1	1	1	1	1	1
	Mean	FGLS BIC	1.1	1.1	1.1	1.1	1.1	1.1	1.5
		DURBIN BIC	0.0	0.9	1.0	1.0	1.0	1.0	1.0
		DURBIN AIC	1.1	2.1	2.1	2.1	2.2	2.2	2.2
600	Median	FGLS	1	1	1	1	1	1	1
		DURBIN BIC	0	1	1	1	1	1	1
		DURBIN AIC	0	1	1	1	1	1	1
	Mean	FGLS BIC	1.0	1.0	1.0	1.0	1.0	1.0	1.5
		DURBIN BIC	0.0	1.0	1.0	1.0	1.0	1.0	1.0
		DURBIN AIC	0.7	1.7	1.7	1.7	1.8	1.8	1.8
2,500	Median	FGLS	1	1	1	1	1	1	1
		DURBIN BIC	0	1	1	1	1	1	1
		DURBIN AIC	0	1	1	1	1	1	1
	Mean	FGLS BIC	1.0	1.0	1.0	1.0	1.0	1.0	1.1
		DURBIN BIC	0.0	1.0	1.0	1.0	1.0	1.0	1.0
		DURBIN AIC	0.7	1.7	1.7	1.7	1.7	1.7	1.7

Notes: All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

APPENDIX B: ADDITIONAL MONTE CARLO: MA DISTURBANCES

Table B1. Selected lags by test estimators: FGLS, DURBIN AIC, DURBIN BIC, DGP: moving average disturbances.

T			$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
50	Median	FGLS BIC	1	1	1	2	2	2	3
		DURBIN BIC	0	0	1	1	2	2	2
		DURBIN AIC	0	1	2	3	5	5	5
	Mean	FGLS BIC	1.3	1.3	1.6	2.2	2.9	3.0	3.0
		DURBIN BIC	0.1	0.3	0.8	1.4	1.9	2.0	2.0
		DURBIN AIC	1.9	2.6	3.3	4.2	5.4	5.5	5.6
200	Median	FGLS BIC	1	1	2	3	5	5	5
		DURBIN BIC	0	1	1	2	3	3	3
		DURBIN AIC	0	1	2	4	8	9	10
	Mean	FGLS BIC	1.1	1.2	1.9	3.1	4.9	5.3	5.4
		DURBIN BIC	0.0	0.8	1.4	2.3	3.3	3.5	3.5
		DURBIN AIC	1.1	2.5	3.5	5.4	9.3	10.7	11.6
600	Median	FGLS BIC	1	1	2	4	7	8	9
		DURBIN BIC	0	1	2	3	5	6	6
		DURBIN AIC	0	2	3	6	12	14	16
	Mean	FGLS BIC	1.0	1.4	2.5	4.3	7.6	8.6	9.0
		DURBIN BIC	0.0	1.1	2.1	3.4	5.4	5.9	6.1
		DURBIN AIC	0.7	2.4	3.8	6.2	12.3	15.2	17.1
2,500	Median	FGLS BIC	1	2	3	6	12	15	17
		DURBIN BIC	0	2	3	5	9	11	12
		DURBIN AIC	0	2	4	8	17	24	28
	Mean	FGLS BIC	1.0	2.0	3.4	5.9	12.2	15.3	17.3
		DURBIN BIC	0.0	1.6	2.9	4.9	9.5	11.2	12.1
		DURBIN AIC	0.7	3.0	4.7	8.0	18.0	24.2	27.2

Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Table B2. Bias, MSE, and relative efficiency estimators: OLS, FGLS, FGLS-D, DURBIN, DGP: moving average disturbances.

		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
T = 50								
Bias	OLS	0.0005	0.0003	0.0002	0.0000	– 0.0001	– 0.0002	– 0.0002
	FGLS	0.0006	0.0003	0.0002	0.0003	– 0.0002	– 0.0003	– 0.0003
	FGLS-D	0.0004	0.0000	0.0002	0.0002	– 0.0002	– 0.0001	– 0.0002
	DURBIN	0.0000	0.0000	0.0002	0.0001	– 0.0005	– 0.0005	– 0.0005
MSE	OLS	0.0112	0.0165	0.0212	0.0267	0.0332	0.0349	0.0364
	FGLS	0.0119	0.0167	0.0195	0.0216	0.0231	0.0238	0.0246
	FGLS-D	0.0113	0.0170	0.0206	0.0227	0.0248	0.0257	0.0266
	DURBIN	0.0130	0.0195	0.0239	0.0268	0.0313	0.0328	0.0341
RE_{est}	OLS	0.8604	0.8465	0.8846	0.9954	1.0603	1.0650	1.0656
	FGLS	0.9168	0.8584	0.8169	0.8055	0.7391	0.7250	0.7195
	FGLS-D	0.8723	0.8745	0.8605	0.8468	0.7914	0.7839	0.7804
T = 200								
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	– 0.0005	– 0.0006	– 0.0006	– 0.0007	– 0.0008	– 0.0008	– 0.0008
	FGLS	– 0.0005	– 0.0008	– 0.0007	– 0.0006	– 0.0003	– 0.0002	– 0.0003
	FGLS-D	– 0.0005	– 0.0007	– 0.0008	– 0.0009	– 0.0002	– 0.0002	– 0.0001
	DURBIN	– 0.0005	– 0.0009	– 0.0012	– 0.0012	– 0.0009	– 0.0010	– 0.0009
MSE	OLS	0.0026	0.0039	0.0050	0.0063	0.0079	0.0083	0.0086
	FGLS	0.0026	0.0037	0.0043	0.0044	0.0040	0.0039	0.0040
	FGLS-D	0.0026	0.0037	0.0043	0.0046	0.0045	0.0045	0.0046
	DURBIN	0.0027	0.0049	0.0053	0.0056	0.0061	0.0063	0.0066
RE_{est}	OLS	0.9585	0.7885	0.9385	1.1317	1.2929	1.3093	1.3145
	FGLS	0.9704	0.7530	0.8030	0.7887	0.6591	0.6232	0.6107
	FGLS-D	0.9593	0.7547	0.8102	0.8206	0.7353	0.7143	0.7072
T = 600								
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002
	FGLS	0.0003	0.0003	0.0002	0.0000	0.0000	– 0.0001	– 0.0002
	FGLS-D	0.0003	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000
	DURBIN	0.0004	0.0001	0.0000	0.0001	0.0002	0.0003	0.0002
MSE	OLS	0.0008	0.0013	0.0016	0.0021	0.0026	0.0027	0.0028
	FGLS	0.0009	0.0012	0.0013	0.0013	0.0010	0.0009	0.0009
	FGLS-D	0.0008	0.0012	0.0013	0.0014	0.0011	0.0011	0.0011
	DURBIN	0.0009	0.0017	0.0017	0.0017	0.0018	0.0019	0.0019
RE_{est}	OLS	0.9852	0.7694	0.9826	1.2120	1.4225	1.4576	1.4644
	FGLS	0.9881	0.7158	0.8009	0.7692	0.5566	0.4912	0.4651
	FGLS-D	0.9854	0.7182	0.8026	0.7892	0.6250	0.5831	0.5667
T = 2,500								
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	– 0.0001	– 0.0002	– 0.0003	– 0.0003	– 0.0004	– 0.0004	– 0.0004
	FGLS	– 0.0001	– 0.0002	– 0.0003	– 0.0003	– 0.0003	– 0.0002	– 0.0002
	FGLS-D	– 0.0001	– 0.0002	– 0.0003	– 0.0003	– 0.0003	– 0.0002	– 0.0002
	DURBIN	– 0.0001	– 0.0002	– 0.0001	– 0.0002	– 0.0002	– 0.0001	– 0.0001

Table B2. Continued

		T = 2,500						
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
MSE	OLS	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0007
	FGLS	0.0002	0.0003	0.0003	0.0003	0.0002	0.0002	0.0001
	FGLS-D	0.0002	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002
	DURBIN	0.0002	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	0.9906	0.7571	0.9752	1.2276	1.4949	1.5520	1.5774
	FGLS	0.9916	0.7056	0.7909	0.7644	0.4806	0.3639	0.3003
	FGLS-D	0.9905	0.7078	0.7936	0.7710	0.5209	0.4293	0.3847

Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Table B3. Empirical size of nominal 5% t-test of $H_0 : \beta = 1$, DGP: moving average disturbances.

	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
T = 50								
OLS	–	0.052	0.095	0.116	0.127	0.131	0.132	0.132
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.070	0.083	0.088	0.090	0.092	0.093	0.092
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.069	0.085	0.090	0.094	0.096	0.096	0.096
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.065	0.074	0.079	0.080	0.080	0.081	0.081
NW-KV	$h = T$	0.063	0.070	0.075	0.078	0.079	0.079	0.080
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.066	0.073	0.073	0.074	0.074	0.074	0.074
FGLS	BIC	0.067	0.072	0.081	0.092	0.098	0.098	0.098
FGLS-D	BIC	0.054	0.091	0.092	0.090	0.092	0.093	0.092
DURBIN	BIC	0.061	0.095	0.090	0.076	0.077	0.078	0.078
DURBIN	AIC	0.085	0.093	0.086	0.080	0.081	0.083	0.084
Hausman 1	OLS vs FGLS		0.743	0.681	0.576	0.473	0.459	0.455
Hausman 2	DURBIN vs FGLS-D		0.047	0.071	0.087	0.098	0.097	0.099
T = 200								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.050	0.094	0.114	0.127	0.129	0.129	0.129
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.055	0.064	0.067	0.069	0.069	0.069	0.069
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.055	0.064	0.067	0.069	0.069	0.069	0.069
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.058	0.059	0.061	0.061	0.061	0.061	0.062
NW-KV	$h = T$	0.053	0.055	0.057	0.057	0.058	0.058	0.057
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.059	0.058	0.058	0.058	0.059	0.059	0.059
FGLS	BIC	0.052	0.054	0.060	0.063	0.061	0.060	0.060
FGLS-D	BIC	0.050	0.057	0.057	0.060	0.057	0.056	0.055
DURBIN	BIC	0.052	0.061	0.050	0.053	0.051	0.051	0.052
DURBIN	AIC	0.063	0.054	0.054	0.054	0.060	0.060	0.060
Hausman 1	OLS vs FGLS		0.668	0.542	0.385	0.224	0.201	0.195
Hausman 2	DURBIN vs FGLS-D		0.047	0.055	0.063	0.060	0.060	0.062
T = 600								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.048	0.095	0.114	0.124	0.129	0.130	0.130
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.049	0.054	0.057	0.058	0.059	0.059	0.059
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.049	0.054	0.057	0.058	0.057	0.057	0.057
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.048	0.052	0.052	0.051	0.052	0.052	0.051
NW-KV	$h = T$	0.051	0.049	0.049	0.048	0.048	0.048	0.048
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.051	0.050	0.051	0.051	0.051	0.052	0.052
FGLS	BIC	0.048	0.049	0.052	0.049	0.046	0.044	0.042
FGLS-D	BIC	0.048	0.046	0.052	0.050	0.043	0.041	0.041
DURBIN	BIC	0.049	0.046	0.046	0.047	0.048	0.048	0.048
DURBIN	AIC	0.058	0.046	0.047	0.048	0.049	0.051	0.053
Hausman 1	OLS vs FGLS		0.585	0.413	0.239	0.105	0.090	0.086
Hausman 2	DURBIN vs FGLS-D		0.051	0.053	0.054	0.052	0.051	0.051
T = 2,500								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.051	0.098	0.114	0.128	0.134	0.134	0.134
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.053	0.055	0.055	0.056	0.056	0.056	0.056
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.053	0.054	0.054	0.055	0.055	0.055	0.055
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.054	0.054	0.055	0.054	0.053	0.053	0.052
NW-KV	$h = T$	0.047	0.049	0.050	0.050	0.051	0.051	0.051
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.054	0.055	0.054	0.053	0.052	0.052	0.052
FGLS	BIC	0.051	0.052	0.050	0.050	0.045	0.038	0.035
FGLS-D	BIC	0.051	0.051	0.050	0.050	0.041	0.037	0.033
DURBIN	BIC	0.052	0.051	0.050	0.050	0.052	0.051	0.052
DURBIN	AIC	0.066	0.051	0.050	0.051	0.052	0.051	0.051
Hausman 1	OLS vs FGLS		0.480	0.305	0.124	0.065	0.061	0.058
Hausman 2	DURBIN vs FGLS-D		0.050	0.048	0.052	0.053	0.056	0.054

Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

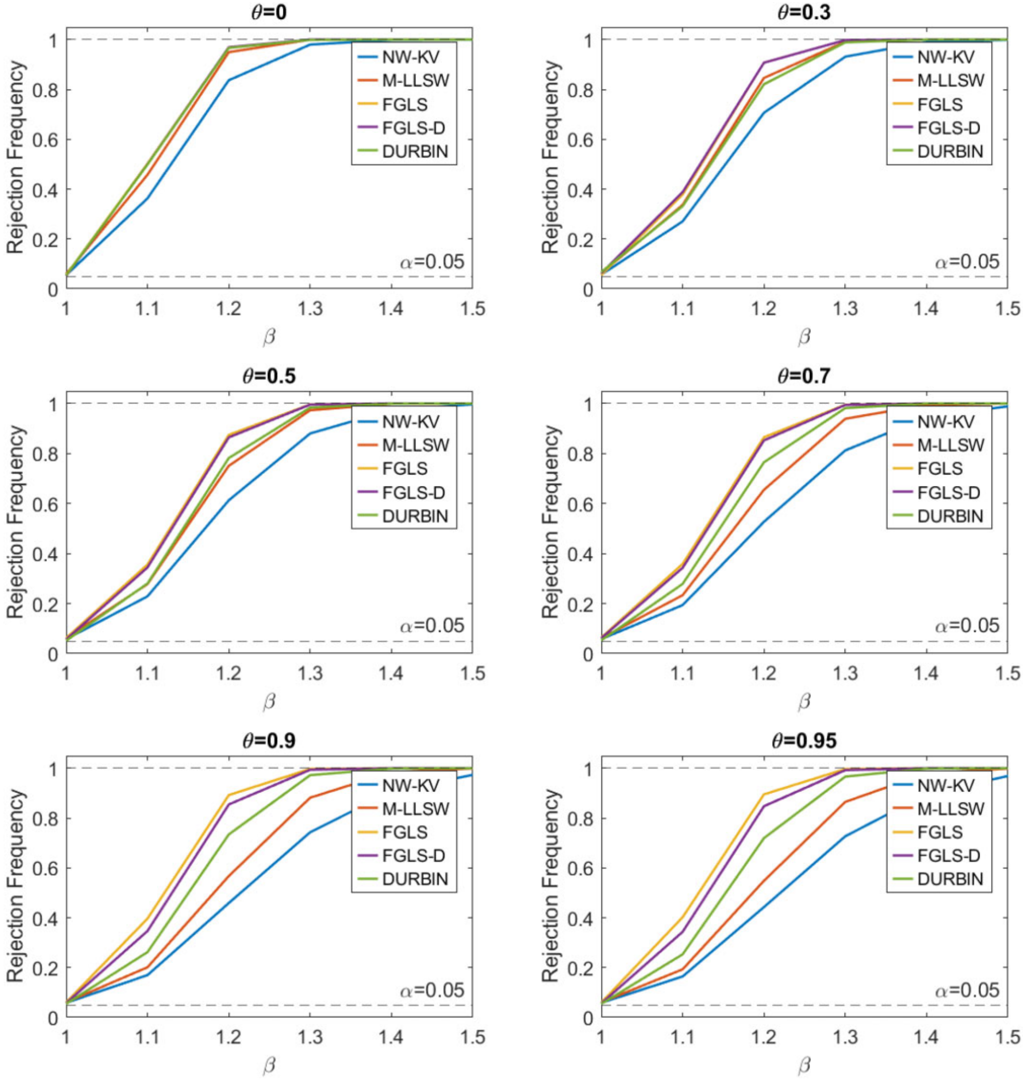

Figure B1. Empirical rejection frequencies of nominal 5% t-test of H_0: $\beta = 1$, DGP: moving average disturbances, $T = 200$. The data-generating process is $y_t = \beta x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = \rho\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., 200$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

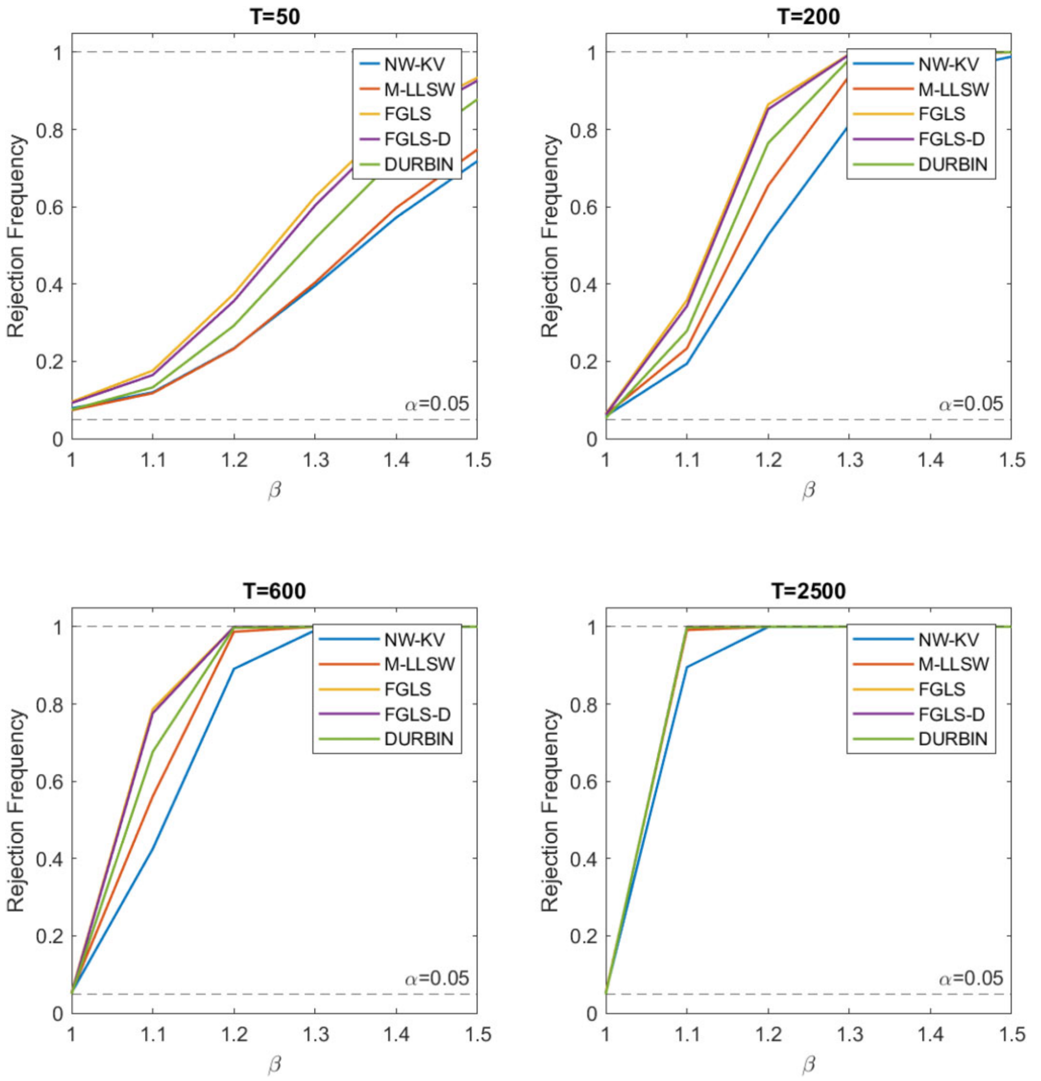

Figure B2. Empirical rejection frequencies of nominal 5% t-test of H_0: $\beta = 1$, DGP: moving average disturbances, $\theta = 0.7$. The data-generating process is $y_t = \beta x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

APPENDIX C: ADDITIONAL MONTE CARLO: ARMA DISTURBANCES

Table C1. Selected lags by test estimators: FGLS, DURBIN AIC, DURBIN BIC, DGP: ARMA disturbances.

T			$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
50	Median	FGLS BIC	1	1	2	2	2	2	1
		DURBIN BIC	1	1	2	2	2	3	3
		DURBIN AIC	1	2	3	4	6	6	6
	Mean	FGLS BIC	1.3	1.6	2.1	2.6	2.9	2.7	2.0
		DURBIN BIC	1.1	1.3	1.7	2.3	2.9	3.0	3.1
		DURBIN AIC	3.0	3.5	4.1	5.0	6.0	6.2	6.3
200	Median	FGLS BIC	1	2	3	4	5	5	4
		DURBIN BIC	1	2	2	3	4	4	4
		DURBIN AIC	1	2	3	5	9	10	11
	Mean	FGLS BIC	1.1	2.0	2.7	3.7	4.9	4.9	3.9
		DURBIN BIC	1.0	1.6	2.3	3.1	4.2	4.4	4.5
		DURBIN AIC	2.2	3.5	4.5	6.3	10.2	11.7	12.6
600	Median	FGLS BIC	1	2	3	5	7	8	7
		DURBIN BIC	1	2	3	4	6	7	7
		DURBIN AIC	1	3	4	6	12	15	17
	Mean	FGLS BIC	1.0	2.2	3.3	4.9	7.5	8.0	7.1
		DURBIN BIC	1.0	2.0	2.9	4.2	6.3	6.7	7.0
		DURBIN AIC	1.8	3.2	4.7	7.0	13.0	16.0	18.0
2,500	Median	FGLS BIC	1	3	4	7	12	14	15
		DURBIN BIC	1	2	4	6	10	12	13
		DURBIN AIC	1	3	5	8	18	25	28
	Mean	FGLS BIC	1.0	2.8	4.2	6.6	12.3	14.6	15.0
		DURBIN BIC	1.0	2.3	3.7	5.8	10.3	12.0	13.0
		DURBIN AIC	1.7	3.8	5.6	8.9	18.9	24.7	27.5

Notes: The data-generating process is $y_t = x_t + u_t, x_t = 0.7x_{t-1} + \epsilon_{x,t}, u_t = 0.7u_{t-1} + \theta\epsilon_{u,t-1} + \epsilon_{u,t}, t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Table C2. Bias, MSE, and relative efficiency estimators: OLS, FGLS, FGLS-D, DURBIN, DGP: ARMA disturbances.

		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
T = 50								
Bias	OLS	0.0024	0.0029	0.0032	0.0036	0.0045	0.0054	0.0126
	FGLS	0.0014	0.0020	0.0021	0.0013	0.0015	0.0010	− 0.0017
	FGLS-D	0.0019	0.0021	0.0024	0.0025	0.0009	0.0008	0.0008
	DURBIN	0.0022	0.0019	0.0023	0.0023	0.0013	0.0021	0.0027
MSE	OLS	0.0566	0.0928	0.1243	0.1678	0.3406	0.8135	15.7532
	FGLS	0.0239	0.0245	0.0240	0.0234	0.0265	0.0333	0.1499
	FGLS-D	0.0231	0.0229	0.0218	0.0197	0.0187	0.0190	0.0197
	DURBIN	0.0231	0.0245	0.0266	0.0292	0.0335	0.0351	0.0366
RE_{est}	OLS	2.4507	3.7877	4.6689	5.7560	10.1796	23.1566	430.2889
	FGLS	1.0346	0.9990	0.9028	0.8022	0.7926	0.9492	4.0948
	FGLS-D	0.9996	0.9342	0.8202	0.6755	0.5599	0.5415	0.5386
T = 200								
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0003	0.0003	0.0003	0.0004	0.0010	0.0020	0.0094
	FGLS	− 0.0001	0.0001	0.0000	0.0000	− 0.0001	− 0.0003	0.0000
	FGLS-D	− 0.0001	0.0000	− 0.0002	− 0.0001	− 0.0001	− 0.0002	− 0.0003
	DURBIN	− 0.0001	0.0000	0.0000	0.0000	− 0.0001	0.0000	− 0.0002
MSE	OLS	0.0145	0.0240	0.0318	0.0414	0.0600	0.0919	1.0251
	FGLS	0.0051	0.0049	0.0042	0.0034	0.0027	0.0028	0.0045
	FGLS-D	0.0051	0.0050	0.0042	0.0034	0.0026	0.0026	0.0026
	DURBIN	0.0051	0.0054	0.0054	0.0057	0.0062	0.0064	0.0067
RE_{est}	OLS	2.8280	4.4747	5.8510	7.2895	9.7269	14.2970	153.9034
	FGLS	0.9997	0.9173	0.7771	0.5953	0.4373	0.4379	0.6732
	FGLS-D	0.9951	0.9244	0.7782	0.5905	0.4216	0.3988	0.3921
T = 600								
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	− 0.0006	− 0.0008	− 0.0010	− 0.0012	− 0.0014	− 0.0015	− 0.0022
	FGLS	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002
	FGLS-D	0.0003	0.0002	0.0001	0.0001	0.0000	0.0001	0.0000
	DURBIN	0.0003	0.0002	0.0002	0.0003	0.0002	0.0003	0.0003
MSE	OLS	0.0048	0.0080	0.0105	0.0136	0.0179	0.0221	0.1314
	FGLS	0.0017	0.0015	0.0013	0.0009	0.0006	0.0005	0.0007
	FGLS-D	0.0017	0.0015	0.0013	0.0010	0.0006	0.0006	0.0006
	DURBIN	0.0017	0.0017	0.0017	0.0017	0.0018	0.0019	0.0020
RE_{est}	OLS	2.8927	4.7354	6.1781	7.8037	9.7536	11.6809	66.9742
	FGLS	0.9999	0.9061	0.7480	0.5354	0.3093	0.2781	0.3499
	FGLS-D	0.9991	0.9059	0.7558	0.5478	0.3281	0.2928	0.2810
T = 2,500								
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0000	− 0.0001	− 0.0001	− 0.0001	− 0.0002	− 0.0002	− 0.0005
	FGLS	0.0000	− 0.0001	− 0.0001	− 0.0001	− 0.0001	− 0.0001	− 0.0001
	FGLS-D	0.0000	− 0.0001	− 0.0001	− 0.0001	− 0.0001	− 0.0001	− 0.0001
	DURBIN	0.0000	0.0000	− 0.0001	0.0000	0.0000	0.0000	0.0000
MSE	OLS	0.0012	0.0020	0.0026	0.0033	0.0042	0.0046	0.0109
	FGLS	0.0004	0.0004	0.0003	0.0002	0.0001	0.0001	0.0001
	FGLS-D	0.0004	0.0004	0.0003	0.0002	0.0001	0.0001	0.0001
	DURBIN	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	2.9101	4.7804	6.3157	8.0661	10.0045	10.7792	24.7226
	FGLS	0.9992	0.9055	0.7473	0.5178	0.2293	0.1695	0.1602
	FGLS-D	0.9991	0.9086	0.7538	0.5239	0.2444	0.1875	0.1630

Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible.

Table C3. Empirical size of nominal 5% t-test of $H_0 : \beta = 1$, DGP: ARMA disturbances.

	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
T = 50								
OLS	–	0.241	0.264	0.271	0.274	0.286	0.299	0.325
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.142	0.150	0.151	0.152	0.140	0.115	0.044
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.159	0.170	0.172	0.173	0.163	0.139	0.068
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.107	0.108	0.108	0.106	0.090	0.066	0.020
NW-KV	$h = T$	0.092	0.094	0.095	0.091	0.071	0.050	0.014
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.082	0.081	0.080	0.075	0.059	0.041	0.013
FGLS	BIC	0.075	0.076	0.082	0.081	0.073	0.075	0.102
FGLS-D	BIC	0.068	0.063	0.068	0.065	0.060	0.058	0.059
DURBIN	BIC	0.056	0.057	0.061	0.063	0.065	0.067	0.068
DURBIN	AIC	0.077	0.078	0.082	0.084	0.083	0.084	0.083
Hausman 1	OLS vs FGLS		0.311	0.243	0.205	0.162	0.120	0.038
Hausman 2	DURBIN vs FGLS-D		0.125	0.117	0.108	0.105	0.104	0.106
T = 200								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.244	0.268	0.275	0.279	0.282	0.291	0.308
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.104	0.105	0.108	0.109	0.108	0.099	0.047
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.104	0.105	0.108	0.109	0.108	0.099	0.047
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.071	0.074	0.074	0.075	0.073	0.063	0.026
NW-KV	$h = T$	0.065	0.065	0.064	0.064	0.057	0.047	0.017
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.064	0.066	0.066	0.064	0.058	0.046	0.019
FGLS	BIC	0.052	0.056	0.050	0.050	0.038	0.036	0.038
FGLS-D	BIC	0.052	0.053	0.050	0.045	0.039	0.038	0.038
DURBIN	BIC	0.049	0.050	0.048	0.050	0.054	0.054	0.053
DURBIN	AIC	0.051	0.052	0.054	0.056	0.057	0.058	0.061
Hausman 1	OLS vs FGLS		0.136	0.118	0.106	0.094	0.085	0.034
Hausman 2	DURBIN vs FGLS-D		0.072	0.053	0.055	0.059	0.060	0.059
T = 600								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.250	0.273	0.278	0.281	0.284	0.287	0.308
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.084	0.087	0.086	0.086	0.088	0.085	0.051
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.079	0.080	0.080	0.080	0.081	0.079	0.047
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.060	0.061	0.063	0.062	0.061	0.056	0.030
NW-KV	$h = T$	0.053	0.052	0.053	0.052	0.048	0.041	0.019
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.057	0.057	0.057	0.057	0.053	0.045	0.021
FGLS	BIC	0.050	0.051	0.049	0.045	0.034	0.031	0.027
FGLS-D	BIC	0.049	0.051	0.048	0.043	0.035	0.034	0.033
DURBIN	BIC	0.049	0.049	0.051	0.052	0.052	0.053	0.052
DURBIN	AIC	0.050	0.051	0.051	0.053	0.054	0.053	0.055
Hausman 1	OLS vs FGLS		0.087	0.080	0.075	0.072	0.068	0.039
Hausman 2	DURBIN vs FGLS-D		0.049	0.053	0.056	0.056	0.057	0.058
T = 2,500								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.251	0.274	0.280	0.284	0.285	0.288	0.302
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.076	0.079	0.078	0.078	0.077	0.075	0.061
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.072	0.073	0.074	0.073	0.073	0.071	0.056
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.057	0.058	0.059	0.060	0.058	0.057	0.044
NW-KV	$h = T$	0.053	0.052	0.053	0.052	0.050	0.046	0.031
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.056	0.057	0.057	0.056	0.054	0.052	0.033
FGLS	BIC	0.053	0.051	0.051	0.049	0.039	0.030	0.025
FGLS-D	BIC	0.052	0.051	0.050	0.049	0.039	0.033	0.031
DURBIN	BIC	0.052	0.054	0.053	0.054	0.053	0.053	0.053
DURBIN	AIC	0.053	0.053	0.053	0.053	0.054	0.053	0.053
Hausman 1	OLS vs FGLS		0.070	0.067	0.066	0.064	0.062	0.048
Hausman 2	DURBIN vs FGLS-D		0.053	0.053	0.052	0.054	0.055	0.056

Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

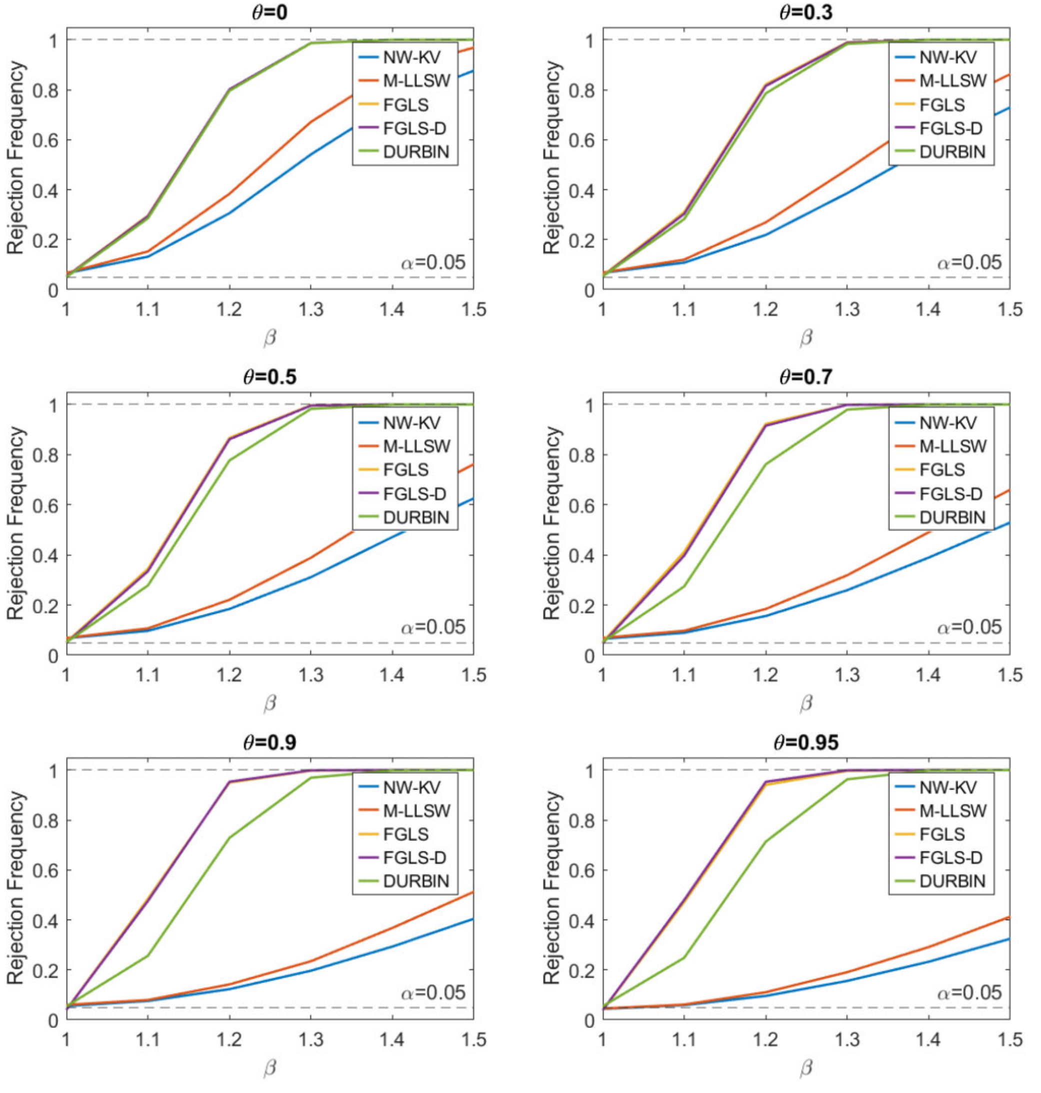

Figure C1. Empirical rejection frequencies of nominal 5% t-test of H_0: $\beta = 1$, DGP: ARMA disturbances, $T = 200$. The data-generating process is $y_t = \beta x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., 200$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

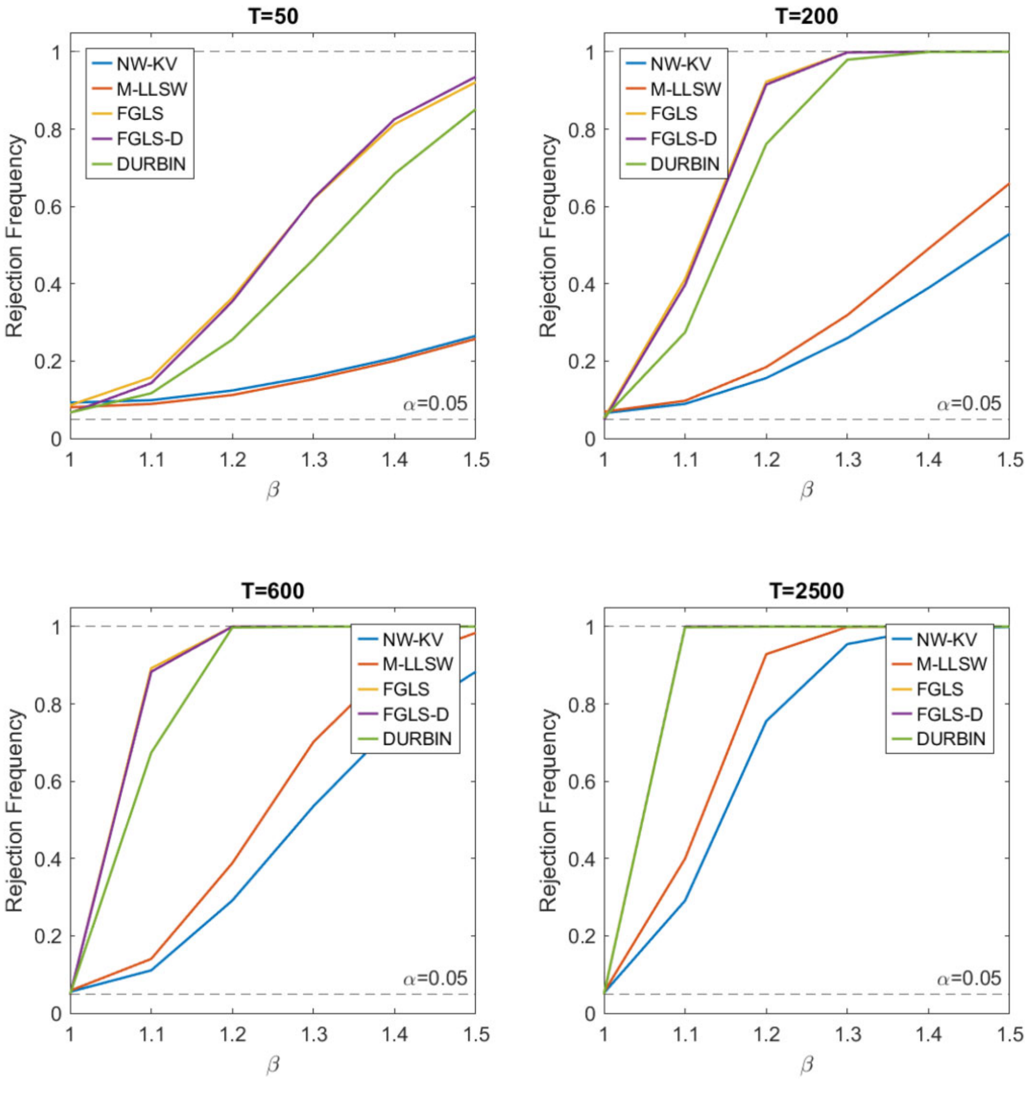

Figure C2. Empirical rejection frequencies of nominal 5% t-test of H_0: $\beta = 1$, DGP: ARMA disturbances, $\theta = 0.5$. The data-generating process is $y_t = \beta x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + 0.5\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, \ldots, T$. All shocks are $N(0, 1)$ white noise. We perform 10,000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.