NBER WORKING PAPER SERIES

ON ROBUST INFERENCE IN TIME SERIES REGRESSION

Richard T. Baillie
Francis X. Diebold
George Kapetanios
Kun Ho Kim
Aaron Mora

Working Paper 32554
http://www.nber.org/papers/w32554

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2024

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-Moller, 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

On Robust Inference in Time Series Regression
Richard T. Baillie, Francis X. Diebold, George Kapetanios, Kun Ho Kim, and Aaron Mora
NBER Working Paper No. 32554
June 2024
JEL No. C13,C22,C31

ABSTRACT

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.

Richard T. Baillie
Department of Economics
Michigan State University
East Lansing, MI 48824
baillie@msu.edu

Francis X. Diebold
Department of Economics
University of Pennsylvania
133 South 36th Street
Philadelphia, PA 19104
and NBER
fdiebold@sas.upenn.edu

George Kapetanios
King's College London
Banking & Finance Department
Bush House
30 Aldwych
London, WC2B 4BG
United Kingdom
george.kapetanios@kcl.ac.uk

Kun Ho Kim
Department of Finance
John Molson School of Business
Concordia University
1450 Guy Street
Montreal, Quebec H3H 0A1
Canada
kunho.kim@concordia.ca

Aaron Mora
Darla Moore School of Business
University of South Carolina
1014 Green Street
Columbia, SC 29208
moramela@mailbox.sc.edu

A data appendix is available at http://www.nber.org/data-appendix/w32554

1 Introduction

For nearly a century, regression with heteroskedastic and/or autocorrelated disturbances has featured prominently in empirical economics research. For many decades, attention centered on modeling the heteroskedasticity or autocorrelation in the context of feasible generalized 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 emphasized (e.g. Angrist and Pischke (2008)).

In an elegant extension, Newey and West (1987) generalize 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 analog 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 sub-optimal (inefficient) forecasts, whereas accurate out-of-sample prediction is often a central concern in time series econometrics.

[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.

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 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 generalized modeling 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 emphasize in this section, even OLS

[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 June 1, 2024.

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,

$$y_t = x_t'\beta + u_t, \tag{1}$$

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 (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.

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 $z_t = (x_t', u_t)'$,

$$z_t = \sum_{i=0}^{\infty} \Xi_i \varepsilon_{t-i}. \tag{2}$$

[3]Because u_t is covariance-stationary, it can be serially correlated and/or conditionally heteroskedastic. In this paper we emphasize 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.

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) is equivalent to the infinite vector-autoregressive (VAR) representation[4]

$$z_t = \sum_{i=1}^{\infty} \Psi_i z_{t-i} + \varepsilon_t, \tag{3}$$

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 (3).

We now proceed to consider various estimation strategies that may be appropriate in the environment given by (1) and (3).

2.2 OLS Parameter Estimation and HAC Covariance Matrix Estimation

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

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

[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.

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

Based on the VAR representation (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 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 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 a classic such situation (Hansen and Hodrick, 1980) at some length in section 3.4 below, but they are by no means the only or the most important possibility. Indeed there is much more to investigate when BD fails, as emphasized 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.

[5] Strong exogeneity is sometimes called strict exogeneity.

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

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

$$y_t = x_t'\beta + \sum_{j=1}^{\infty} \phi_j y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}'\gamma_j + \varepsilon_{y,t}, \tag{4}$$

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 well that the DGP remains (1) and (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.

Operationally, it is of course necessary to use a finite order approximation to the infinite order DURBIN regression (4),

$$y_t = x_t'\beta + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{p} x_{t-j}'\gamma_j + \varepsilon_{y,t}, \tag{5}$$

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) or 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

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + \varepsilon_{y,t}, \tag{6}$$

which emphasizes the extent of the parameterization and lag structure. We will later explore in greater detail the relationship between the DGP given by (1) and (3) and the DURBIN regression (6), which is effectively one equation of a VAR and appears to be the originator of the autoregressive distributed lag (ADL) model 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 (6). 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, $E(u_s|x_t) = 0, \forall s, t \leq s$) 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.

To see the consistency and efficiency of DURBIN under EBD, note that, using (3) and $u_t = y_t - x_t'\beta$, we can write (1) as

$$\begin{aligned} y_t &= x_t'\beta + u_t \\ &= x_t'\beta + \sum_{j=1}^{\infty} \psi_{u,j} u_{t-j} + \sum_{j=1}^{\infty} \Psi_{xu,j} x_{t-j} + \varepsilon_{u,t} \\ &= x_t'\beta + \sum_{j=1}^{\infty} \psi_{u,j} \left(y_{t-j} - x_{t-j}'\beta\right) + \sum_{j=1}^{\infty} \Psi_{xu,j} x_{t-j} + \varepsilon_{u,t} \\ &= x_t'\beta + \sum_{j=1}^{\infty} \psi_{u,j} y_{t-j} + \sum_{j=1}^{\infty} x_{t-j}' \left(\Psi_{xu,j}' - \psi_{u,j}\beta\right) + \varepsilon_{u,t}. \end{aligned} \tag{7}$$

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.

[7] Weak exogeneity is sometimes called predeterminedness. See Mikusheva and Sølvsten (2023).

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

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. On the other hand, when the FGLS variants are consistent, they are also fully efficient. We shall see how such trade-offs resolve themselves in the simulations of section 4 below.

3 A Generalized Data-Generating Process

We now move from the basic DGP (1) to a generalized version that subsumes all cases of interest.

3.1 Data-Generating Process

Henceforth we work with the data-generating process given by

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + u_t. \tag{8}$$

We emphasize 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 the vector moving average (2), or equivalently, the vector autoregression (3). This generalized DGP covers most linear dynamic relationships of conceivable interest. We use NDY ("no dynamics in y") to refer to the restriction imposed on the generalized DGP (8) to get the basic DGP (1), namely $\phi_j = \gamma_{i,j} = 0 \ \forall i, j$.

We also emphasize that (8) is now the data generating process, and various regressions could be fit to its data realizations in various attempts at estimation and inference for β. One such regression, for example, is FGLS. Clearly the use of FGLS in environments characterized by the generalized DGP (8) 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 generalized DGP (8). Indeed DURBIN can perfectly accommodate the generalized DGP, because, in precise parallel to

(7), we have

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

which is a DURBIN regression with

$$
\begin{aligned}
\phi_j &= \lambda_j + \psi_{u,j} - \sum_{s=1}^{j-1}\psi_{u,s}\lambda_{j-s} \\
\gamma_j &= \psi_{xu,j} - \psi_{u,j}\beta - \sum_{s=1}^{j-1}\psi_{u,s}\theta_{j-s}.
\end{aligned}
$$

The above relationships between the parameters of the generalized DGP (8) and the DURBIN regression (6) also show that the generalized DGP is so richly parameterized that not all parameters are identified through estimation of (6) 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 summarize the consistency and efficiency properties of all estimators in the leading environments that we have considered, all of which are specializations of the generalized DGP given by (8) and (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 generalized DGP (8) with $(x_t, u_t)'$ governed by (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 (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 inconsis-

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

tent but have different limits, which depend on the parameters of (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

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

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 below that the 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 below 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 vs 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) recognized that FGLS can be inconsistent in such a situation, whereas OLS remains consistent, albeit inefficient. They recognized, 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,

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

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

$$y_{t+k} = x_t\beta + \theta(L)\varepsilon_t, \tag{9}$$

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

Conceptually, equation (9) is merely a restricted DURBIN model, because on using the filter $\theta(L)^{-1}$ we obtain

$$\left\{\theta(L)^{-1}y_{t+k}\right\} = \beta\left\{\theta(L)^{-1}x_t\right\} + \varepsilon_{t+k}. \tag{10}$$

The filtered explanatory variable is uncorrelated with current and future innovations, ε_{t+k}, so that estimation of equation (10) 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 $\pi(L) = (1 - \pi_1 L - ... - \pi_p L^p)$ is a pth-order lag-operator polynomial

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

with all roots outside the unit circle. DURBIN will then be

$$\pi(L)y_{t+k} = \beta\pi(L)x_t + \varepsilon_{t+k}, \tag{11}$$

which is a restricted version of the generalized DGP (8) 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.

4.1 Simulation Design

The main simulation results will comprise four data generation processes that impose different assumptions on the generalized DGP given by (8) and (3):

1. Autoregressive Disturbances, AR(1) ($NDY + BD$)

$$\begin{aligned} y_t &= \beta x_t + u_t \\ \begin{pmatrix} x_t \\ u_t \end{pmatrix} &= \begin{pmatrix} 0.7 & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix} \end{aligned} \tag{12}$$

2. Triangular vector autoregression (VAR) on (3) ($NDY + GEXOG$)

$$\begin{aligned} y_t &= \beta x_t + u_t \\ \begin{pmatrix} x_t \\ u_t \end{pmatrix} &= \begin{pmatrix} \psi_{11} & \psi_{12} \\ 0 & \psi_{22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix} \end{aligned} \tag{13}$$

3. Unrestricted VAR on (3) ($NDY + EBD$)

$$\begin{aligned} y_t &= \beta x_t + u_t \\ \begin{pmatrix} x_t \\ u_t \end{pmatrix} &= \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix} \end{aligned} \tag{14}$$

4. Dynamic Regression (EBD)

$$y_t = \beta x_t + \rho y_{t-1} - 0.5 x_{t-1} + u_t$$
$$\begin{pmatrix} x_t \\ u_t \end{pmatrix} = \begin{pmatrix} 0.7 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{u,t} \end{pmatrix}. \tag{15}$$

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 2500 weeks. Including $T = 2500$ also lets us check our Monte Carlo results against known large-sample results.

The autoregressive DGP in (12) 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 iid case.

In the simulations for the triangular VAR DGP in (13), 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 (14), 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 (15), we consider various parameter values for the 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 footnote 9 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 realizations 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

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

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 $\tau = 0, \pm 1, ...$

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:

$$\widehat{\Omega} = \left(\frac{1}{T} \sum_{t=1}^{T} (x_t x_t') \hat{u}_t^2 + \sum_{\tau=1}^{h} \left(1 - \frac{\tau}{h+1} \right) (\widehat{\mathbf{\Gamma}}_\tau + \widehat{\mathbf{\Gamma}}_{-\tau}) \right), \tag{16}$$

where

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

the $\hat{u}_t$ are OLS regression residuals, and T is sample size. Indeed, many leading HAC estimators are of the form (16), distinguished only by their choice of truncation lag h.

We will explore several leading truncation lag choices, including:

1. NW: Newey-West (16) 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 (16) with $h = \lceil 0.75T^{1/3} \rceil$. This h choice is also standard, arising when a formula in Andrews (1991) is specialized to the case of a first-order autoregression with coefficient 0.25.

3. NW-LLSW: Newey-West (16) with $h = \lceil 1.3T^{1/2} \rceil$, 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 (16) 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 (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 $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$ 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 $\hat{u}_{t-1}, ..., \hat{u}_{t-p}$, with p selected by AIC or BIC), and obtain the coefficients $\hat{\phi}_1, ..., \hat{\phi}_p$.

3. Construct the transformed data,

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

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, $\hat{\varphi}_1, ..., \hat{\varphi}_p$, to construct the transformed data,

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

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 (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

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + \varepsilon_t, \tag{17}$$

[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.

which has $p + k + kp$ parameters.

If u_t in equation (1) is a finite-ordered AR(p) process with p known, then DURBIN holds exactly. In particular, we have[12]

$$\begin{aligned} y_t &= \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \beta_i \phi_j x_{i,t-j} + \varepsilon_t \\ &= \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{p} \sum_{i=1}^{k} \gamma_{i,j} x_{i,t-j} + \varepsilon_t. \end{aligned} \tag{18}$$

Hence the usual asymptotic inference is immediately available:

$$T^{1/2}(\widehat{\vartheta}_{OLS} - \vartheta) \to N(0, Q^{-1}), \tag{19}$$

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

$$Q = plim \left(T^{-1} \sum_{t=1}^{T} z_t z_t' \right), \tag{20}$$

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

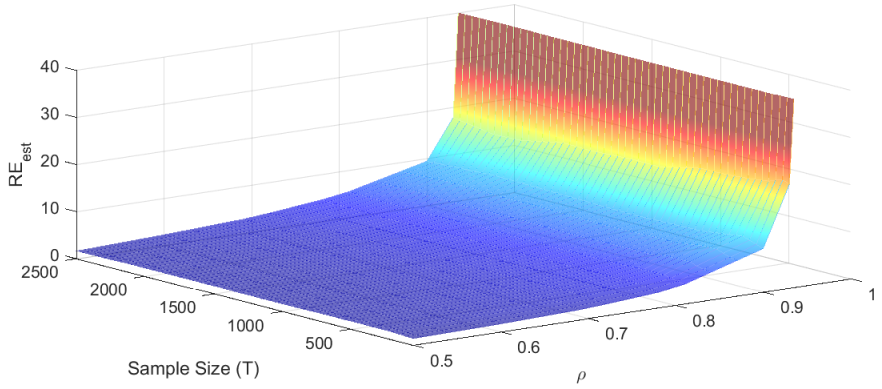
In the more compelling case where p is *un*known and must be selected (implemented in our Monte Carlo below), the DURBIN regression (17) is approximate rather than exact. However, the limiting distribution (19) remains valid if p is selected suitably (Grenander, 1981; Hannan and Deistler, 1988), as achieved 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]

[12]Note that DURBIN does *not* impose the common factor restriction embedded in (18), 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.

Figure 1: Efficiency of DURBIN Relative to OLS
DGP: Autoregressive Disturbances, $NDY + BD$



Notes: All shocks are $N(0,1)$ white noise. We select DURBIN lag order using BIC. We perform 10000 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 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 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]

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} stabilizes quickly in sample size T and remains approximately constant. Consider, for example, a realistic case like $\rho =$

[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 $\hat{\beta}_{OLS}$.

Table 2: Bias, MSE, and Relative Efficiency
Estimators: OLS, FGLS, FGLS-D, DURBIN
DGP: Autoregressive Disturbances, $NDY + BD$

				T=50				
		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
Bias	OLS	0.0006	-0.0021	-0.0008	0.0025	0.0093	0.0098	-0.0270
	FGLS	0.0006	-0.0015	-0.0002	0.0000	0.0030	-0.0002	-0.0068
	FGLS-D	0.0007	-0.0016	-0.0002	0.0006	0.0021	-0.0002	0.0014
	DURBIN	0.0006	-0.0010	0.0006	0.0001	0.0034	-0.0006	0.0006
MSE	OLS	0.0112	0.0180	0.0289	0.0599	0.3013	1.3368	80.0070
	FGLS	0.0121	0.0174	0.0216	0.0237	0.0251	0.0304	0.2953
	FGLS-D	0.0114	0.0179	0.0222	0.0231	0.0207	0.0198	0.0183
	DURBIN	0.0131	0.0201	0.0237	0.0229	0.0226	0.0234	0.0227
RE_{est}	OLS	0.8597	0.8952	1.2189	2.6140	13.3211	57.1114	3530.5932
	FGLS	0.9261	0.8666	0.9116	1.0356	1.1115	1.2986	13.0298
	FGLS-D	0.8685	0.8869	0.9370	1.0063	0.9142	0.8476	0.8092

				T=200				
		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
Bias	OLS	-0.0003	-0.0006	-0.0007	-0.0012	0.0029	-0.0082	-0.0441
	FGLS	-0.0003	-0.0008	-0.0004	-0.0009	-0.0005	-0.0008	0.0001
	FGLS-D	-0.0003	-0.0008	-0.0004	-0.0008	-0.0005	-0.0005	0.0000
	DURBIN	-0.0002	-0.0009	-0.0003	-0.0009	-0.0002	-0.0004	-0.0003
MSE	OLS	0.0026	0.0043	0.0072	0.0147	0.0635	0.1884	8.9415
	FGLS	0.0027	0.0039	0.0048	0.0051	0.0048	0.0046	0.0048
	FGLS-D	0.0026	0.0040	0.0048	0.0051	0.0048	0.0045	0.0044
	DURBIN	0.0027	0.0051	0.0052	0.0051	0.0052	0.0052	0.0051
RE_{est}	OLS	0.9611	0.8526	1.3750	2.8763	12.3138	36.3636	1753.3931
	FGLS	0.9761	0.7730	0.9319	1.0083	0.9296	0.8909	0.9385
	FGLS-D	0.9614	0.7801	0.9306	1.0024	0.9221	0.8767	0.8567

				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	-0.0008	0.0009	0.0005	0.0007	0.0119
	FGLS	0.0001	0.0000	-0.0006	0.0001	0.0002	0.0002	-0.0006
	FGLS-D	0.0001	0.0000	-0.0006	0.0001	0.0002	0.0002	-0.0006
	DURBIN	0.0001	0.0001	-0.0005	0.0001	0.0003	0.0004	-0.0006
MSE	OLS	0.0009	0.0015	0.0024	0.0048	0.0203	0.0493	1.1932
	FGLS	0.0009	0.0013	0.0016	0.0017	0.0015	0.0015	0.0015
	FGLS-D	0.0009	0.0013	0.0016	0.0017	0.0015	0.0015	0.0014
	DURBIN	0.0009	0.0017	0.0017	0.0017	0.0016	0.0017	0.0017
RE_{est}	OLS	0.9900	0.8551	1.3794	2.8902	12.4793	29.5992	705.7406
	FGLS	0.9959	0.7625	0.9287	1.0027	0.9257	0.8918	0.8731
	FGLS-D	0.9899	0.7622	0.9278	1.0026	0.9250	0.8912	0.8464

				T=2500				
		$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
Bias	OLS	0.0000	0.0002	-0.0001	0.0001	-0.0008	0.0008	-0.0010
	FGLS	0.0000	0.0002	0.0000	0.0001	0.0002	0.0002	-0.0001
	FGLS-D	0.0000	0.0002	0.0000	0.0001	0.0002	0.0002	-0.0001
	DURBIN	0.0000	0.0004	0.0001	0.0001	0.0001	0.0003	0.0000
MSE	OLS	0.0002	0.0003	0.0006	0.0011	0.0048	0.0106	0.1116
	FGLS	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004
	FGLS-D	0.0002	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004
	DURBIN	0.0002	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	0.9946	0.8436	1.4542	2.8534	12.0802	26.5763	272.6119
	FGLS	0.9954	0.7551	0.9355	0.9998	0.9212	0.8891	0.8598
	FGLS-D	0.9946	0.7549	0.9353	1.0000	0.9213	0.8887	0.8569

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 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

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.

Triangular and Unrestricted VAR DGPs ($NDY+GEXOG$, $NDY+EBD$). Results appear in Table 3. Ψ_1 and Ψ_1^* correspond to different parameterizations of the $NDY + GEXOG$ DGP, and Ψ_2 and Ψ_2^* correspond to different parameterizations 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.

[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.

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 (15) can be rewritten in the form of (12), 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.

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 modeling 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

Table 3: Bias, MSE, and Relative Efficiency
Estimators: OLS, FGLS, FGLS-D, DURBIN
DGPs: (1) Triangular VAR, $NDY + GEXOG$, (2) Unrestricted VAR, $NDY + EBD$

		T=50			
		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Bias	OLS	0.2376	0.3054	0.5286	0.6828
	FGLS	0.1083	0.1484	0.4622	0.6367
	FGLS-D	0.0708	0.0520	0.1889	0.2902
	DURBIN	0.0637	0.0386	0.0268	0.0074
MSE	OLS	0.0705	0.1082	0.2989	0.4834
	FGLS	0.0286	0.0400	0.2494	0.4385
	FGLS-D	0.0384	0.0352	0.0980	0.1672
	DURBIN	0.0408	0.0374	0.0409	0.0300
RE_{est}	OLS	1.7284	2.8912	7.3025	16.1337
	FGLS	0.7021	1.0693	6.0917	14.6353
	FGLS-D	0.9430	0.9393	2.3938	5.5796

		T=200			
		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Bias	OLS	0.2445	0.3174	0.5621	0.7252
	FGLS	0.1001	0.1413	0.5037	0.6986
	FGLS-D	0.0033	0.0034	0.1947	0.3410
	DURBIN	0.0009	-0.0007	-0.0003	0.0002
MSE	OLS	0.0632	0.1045	0.3206	0.5291
	FGLS	0.0141	0.0246	0.2644	0.4944
	FGLS-D	0.0052	0.0051	0.0557	0.1451
	DURBIN	0.0052	0.0051	0.0052	0.0052
RE_{est}	OLS	12.0344	20.6612	62.1667	101.5652
	FGLS	2.6880	4.8570	51.2578	94.9083
	FGLS-D	0.9968	1.0006	10.8053	27.8557

		T=600			
		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Bias	OLS	0.2460	0.3218	0.5721	0.7378
	FGLS	0.0980	0.1409	0.5045	0.7176
	FGLS-D	0.0000	0.0012	0.2007	0.3667
	DURBIN	-0.0008	-0.0002	0.0001	0.0008
MSE	OLS	0.0617	0.1048	0.3288	0.5454
	FGLS	0.0110	0.0214	0.2592	0.5169
	FGLS-D	0.0017	0.0017	0.0464	0.1454
	DURBIN	0.0017	0.0017	0.0017	0.0017
RE_{est}	OLS	37.0807	61.9455	199.1306	329.6430
	FGLS	6.6214	12.6578	156.9929	312.4485
	FGLS-D	1.0163	1.0083	28.1260	87.8873

		T=2500			
		Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
Bias	OLS	0.2472	0.3220	0.5756	0.7438
	FGLS	0.0984	0.1400	0.4654	0.7213
	FGLS-D	0.0004	0.0001	0.2034	0.3763
	DURBIN	0.0002	-0.0003	0.0001	-0.0001
MSE	OLS	0.0614	0.1040	0.3316	0.5534
	FGLS	0.0100	0.0200	0.2179	0.5209
	FGLS-D	0.0004	0.0004	0.0430	0.1444
	DURBIN	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	153.9376	258.1795	830.3128	1370.2607
	FGLS	25.1163	49.6206	545.5996	1289.7121
	FGLS-D	1.0261	1.0147	107.5807	357.5905

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 10000 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 4: Bias, MSE, and Relative Efficiency
Estimators: OLS, FGLS, FGLS-D, DURBIN
DGP: Dynamic Regression, EBD

				T=50		
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	-0.3366	-0.0003	0.2531	0.6886	0.8700
	FGLS	-0.3369	-0.0002	0.0329	-0.1000	-0.1536
	FGLS-D	-0.3336	-0.0004	0.0170	-0.1317	-0.1891
	DURBIN	-0.0484	-0.0001	-0.0009	0.0000	-0.0004
MSE	OLS	0.1275	0.0292	0.1305	0.9036	1.9497
	FGLS	0.1297	0.0215	0.0303	0.0423	0.0597
	FGLS-D	0.1275	0.0220	0.0267	0.0400	0.0570
	DURBIN	0.0435	0.0234	0.0224	0.0231	0.0234
RE_{est}	OLS	2.9300	1.2443	5.8337	39.1776	83.4520
	FGLS	2.9827	0.9161	1.3561	1.8348	2.5563
	FGLS-D	2.9310	0.9401	1.1944	1.7358	2.4410
				T=200		
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	-0.3462	-0.0004	0.2678	0.7371	0.9235
	FGLS	-0.3464	-0.0004	0.0082	-0.1369	-0.1895
	FGLS-D	-0.3457	-0.0004	0.0049	-0.1422	-0.1946
	DURBIN	0.0000	-0.0003	0.0004	0.0005	0.0005
MSE	OLS	0.1231	0.0071	0.0879	0.6648	1.1970
	FGLS	0.1236	0.0047	0.0061	0.0242	0.0408
	FGLS-D	0.1231	0.0047	0.0060	0.0254	0.0425
	DURBIN	0.0051	0.0050	0.0051	0.0053	0.0051
RE_{est}	OLS	24.1127	1.4081	17.2231	124.8315	232.6470
	FGLS	24.1985	0.9384	1.2010	4.5375	7.9253
	FGLS-D	24.0998	0.9342	1.1725	4.7644	8.2576
				T=600		
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	-0.3489	0.0005	0.2734	0.7502	0.9393
	FGLS	-0.3488	0.0002	0.0036	-0.1431	-0.1938
	FGLS-D	-0.3486	0.0002	0.0026	-0.1447	-0.1952
	DURBIN	-0.0002	0.0002	0.0009	-0.0001	0.0010
MSE	OLS	0.1228	0.0024	0.0803	0.6050	0.9995
	FGLS	0.1229	0.0015	0.0020	0.0221	0.0391
	FGLS-D	0.1227	0.0015	0.0020	0.0226	0.0396
	DURBIN	0.0017	0.0017	0.0017	0.0017	0.0017
RE_{est}	OLS	73.7851	1.4430	46.0534	364.0187	596.3733
	FGLS	73.8221	0.9363	1.1720	13.3167	23.3308
	FGLS-D	73.7053	0.9356	1.1602	13.5727	23.6306
				T=2500		
		$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
Bias	OLS	-0.3495	0.0001	0.2740	0.7544	0.9374
	FGLS	-0.3494	0.0002	0.0007	-0.1453	-0.1959
	FGLS-D	-0.3494	0.0002	0.0005	-0.1456	-0.1962
	DURBIN	0.0003	0.0002	0.0000	-0.0002	0.0002
MSE	OLS	0.1224	0.0006	0.0764	0.5791	0.9073
	FGLS	0.1224	0.0004	0.0005	0.0215	0.0387
	FGLS-D	0.1223	0.0004	0.0004	0.0216	0.0388
	DURBIN	0.0004	0.0004	0.0004	0.0004	0.0004
RE_{est}	OLS	311.2664	1.4398	193.3029	1445.5630	2249.2027
	FGLS	311.1810	0.9367	1.1432	53.6830	96.0556
	FGLS-D	311.1251	0.9365	1.1370	53.8808	96.2890

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 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

and known parameter $\beta = 1$.[18] 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

$$\begin{aligned} y^{opt}_{t+1,t} &= x_{t+1,t} + u_{t+1,t} \\ &= \rho x_t + \rho u_t, \end{aligned} \tag{21}$$

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 (21),

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

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

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

$$\text{RE}_{\text{pred}} = \frac{\sigma^2_{\text{subopt}}}{\sigma^2_{\text{opt}}} = \frac{1}{2} + \frac{1}{2(1-\rho^2)}. \tag{22}$$

RE_{pred} is bounded below by 1, which occurs when $\rho=0$, and $\text{RE}_{\text{pred}} \to \infty$ 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 (22) 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 $\text{RE}_{\text{pred}} = 1$ in the white noise case ($\rho = 0$), but then RE_{pred} grows quickly in ρ, and for any ρ, RE_{pred} stabilizes extremely quickly in T and is basically constant.

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

[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 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	0.989	1.042	1.160	1.452	3.033	5.865	391.908
200	0.997	1.051	1.168	1.476	3.121	5.698	47.361
600	1.000	1.047	1.152	1.505	3.214	5.605	25.569
2500	1.000	1.049	1.163	1.469	3.101	5.656	25.648

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 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

4.5 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 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 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	–	0.051	0.106	0.167	0.245	0.346	0.380	0.407
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.066	0.088	0.113	0.141	0.200	0.227	0.263
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.064	0.093	0.121	0.160	0.230	0.264	0.294
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.064	0.078	0.091	0.110	0.123	0.137	0.196
NW-KV	$h = T$	0.061	0.075	0.081	0.097	0.091	0.090	0.155
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.065	0.069	0.077	0.087	0.086	0.102	0.167
FGLS	BIC	0.066	0.076	0.082	0.076	0.076	0.084	0.054
FGLS-D	BIC	0.054	0.095	0.090	0.069	0.057	0.059	0.052
DURBIN	BIC	0.060	0.099	0.082	0.058	0.053	0.058	0.051
DURBIN	AIC	0.086	0.093	0.088	0.078	0.076	0.080	0.076
Hausman 1	OLS vs FGLS		0.738	0.632	0.446	0.253	0.239	0.274
Hausman 2	DURBIN vs FGLS-D		0.052	0.091	0.121	0.119	0.118	0.100
		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	0.110	0.174	0.252	0.352	0.386	0.413
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.059	0.067	0.085	0.107	0.142	0.157	0.174
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.059	0.067	0.085	0.107	0.142	0.157	0.174
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.059	0.060	0.068	0.072	0.080	0.076	0.084
NW-KV	$h = T$	0.053	0.056	0.061	0.060	0.061	0.049	0.032
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.060	0.059	0.063	0.063	0.066	0.063	0.076
FGLS	BIC	0.055	0.054	0.056	0.055	0.052	0.053	0.048
FGLS-D	BIC	0.051	0.061	0.054	0.054	0.053	0.051	0.050
DURBIN	BIC	0.053	0.064	0.049	0.052	0.051	0.053	0.049
DURBIN	AIC	0.066	0.055	0.053	0.056	0.056	0.057	0.052
Hausman 1	OLS vs FGLS		0.622	0.455	0.212	0.126	0.115	0.123
Hausman 2	DURBIN vs FGLS-D		0.048	0.067	0.093	0.074	0.066	0.062
		T=600						
	Truncation	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
OLS	–	0.049	0.117	0.175	0.249	0.351	0.379	0.408
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.049	0.065	0.070	0.086	0.114	0.120	0.134
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.049	0.064	0.068	0.082	0.104	0.109	0.121
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.051	0.057	0.057	0.057	0.065	0.058	0.050
NW-KV	$h = T$	0.049	0.050	0.053	0.053	0.050	0.042	0.017
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.051	0.054	0.054	0.055	0.057	0.051	0.052
FGLS	BIC	0.050	0.055	0.051	0.049	0.047	0.051	0.051
FGLS-D	BIC	0.049	0.055	0.050	0.049	0.047	0.051	0.051
DURBIN	BIC	0.049	0.055	0.049	0.049	0.045	0.048	0.049
DURBIN	AIC	0.063	0.056	0.050	0.050	0.046	0.048	0.051
Hausman 1	OLS vs FGLS		0.533	0.296	0.115	0.088	0.077	0.069
Hausman 2	DURBIN vs FGLS-D		0.052	0.052	0.089	0.059	0.051	0.054
		T=2500						
	Truncation	$\rho = 0$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
OLS	–	0.050	0.110	0.179	0.246	0.352	0.376	0.405
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.050	0.053	0.063	0.068	0.088	0.093	0.093
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.050	0.052	0.060	0.065	0.081	0.085	0.082
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.051	0.048	0.052	0.052	0.057	0.056	0.042
NW-KV	$h = T$	0.050	0.047	0.050	0.049	0.050	0.046	0.027
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.051	0.047	0.052	0.051	0.053	0.051	0.042
FGLS	BIC	0.050	0.048	0.051	0.050	0.048	0.050	0.052
FGLS-D	BIC	0.049	0.048	0.050	0.050	0.048	0.050	0.053
DURBIN	BIC	0.050	0.050	0.049	0.050	0.048	0.053	0.049
DURBIN	AIC	0.064	0.050	0.049	0.050	0.047	0.052	0.050
Hausman 1	OLS vs FGLS		0.429	0.155	0.072	0.068	0.066	0.049
Hausman 2	DURBIN vs FGLS-D		0.051	0.050	0.086	0.052	0.052	0.052

Notes: All shocks are $N(0, 1)$ white noise. We perform 10000 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, $NDY + EBD$

T=50					
	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
OLS	–	0.613	0.787	0.969	0.992
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.558	0.730	0.954	0.989
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.568	0.743	0.959	0.990
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.507	0.676	0.928	0.980
NW-KV	$h = T$	0.438	0.586	0.874	0.958
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.447	0.607	0.899	0.970
FGLS	BIC	0.221	0.322	0.883	0.959
FGLS-D	BIC	0.309	0.248	0.505	0.627
DURBIN	BIC	0.276	0.200	0.124	0.070
DURBIN	AIC	0.133	0.099	0.087	0.078
Hausman 1	OLS vs FGLS	0.880	0.899	0.693	0.740
Hausman 2	DURBIN vs FGLS-D	0.006	0.009	0.427	0.640
T=200					
	Truncation	Ψ_1	Ψ_1^*	Ψ_2	Ψ_2^*
OLS	–	0.990	0.999	1.000	1.000
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.983	0.999	1.000	1.000
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.983	0.999	1.000	1.000
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.974	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	0.996	1.000	1.000
FGLS	BIC	0.453	0.688	0.999	1.000
FGLS-D	BIC	0.125	0.123	0.735	0.903
DURBIN	BIC	0.047	0.047	0.048	0.052
DURBIN	AIC	0.047	0.051	0.054	0.054
Hausman 1	OLS vs FGLS	0.994	0.998	0.516	0.457
Hausman 2	DURBIN vs FGLS-D	0.000	0.000	0.910	0.978
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	0.831	0.978	1.000	1.000
FGLS-D	BIC	0.122	0.128	0.956	0.998
DURBIN	BIC	0.048	0.051	0.051	0.048
DURBIN	AIC	0.049	0.053	0.052	0.047
Hausman 1	OLS vs FGLS	1.000	1.000	0.669	0.273
Hausman 2	DURBIN vs FGLS-D	0.000	0.000	1.000	1.000
T=2500					
	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	0.123	0.129	1.000	1.000
DURBIN	BIC	0.048	0.050	0.050	0.050
DURBIN	AIC	0.048	0.049	0.050	0.050
Hausman 1	OLS vs FGLS	1.000	1.000	0.999	0.465
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 10000 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 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	–	0.785	0.162	0.441	0.504	0.406
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.779	0.105	0.302	0.330	0.247
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.783	0.117	0.331	0.372	0.282
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.742	0.087	0.237	0.228	0.159
NW-KV	$h = T$	0.662	0.082	0.200	0.184	0.122
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.687	0.075	0.194	0.170	0.124
FGLS	BIC	0.775	0.077	0.097	0.113	0.183
FGLS-D	BIC	0.768	0.085	0.078	0.124	0.198
DURBIN	BIC	0.205	0.078	0.052	0.055	0.059
DURBIN	AIC	0.098	0.082	0.071	0.078	0.078
Hausman 1	OLS vs FGLS	0.782	0.624	0.583	0.491	0.332
Hausman 2	DURBIN vs FGLS-D	0.770	0.085	0.565	0.930	0.955
		T=200				
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
OLS	–	1.000	0.169	0.820	0.903	0.779
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	0.081	0.656	0.738	0.516
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	0.081	0.656	0.738	0.516
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.999	0.063	0.560	0.620	0.352
NW-KV	$h = T$	0.979	0.058	0.421	0.447	0.248
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.999	0.060	0.536	0.583	0.319
FGLS	BIC	1.000	0.053	0.066	0.415	0.683
FGLS-D	BIC	1.000	0.051	0.062	0.441	0.716
DURBIN	BIC	0.049	0.047	0.050	0.055	0.052
DURBIN	AIC	0.051	0.051	0.054	0.058	0.058
Hausman 1	OLS vs FGLS	0.714	0.443	0.827	0.899	0.665
Hausman 2	DURBIN vs FGLS-D	1.000	0.061	0.974	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	0.174	0.995	0.999	0.989
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	0.075	0.971	0.991	0.928
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	0.072	0.970	0.991	0.920
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	1.000	0.058	0.950	0.985	0.882
NW-KV	$h = T$	1.000	0.050	0.801	0.843	0.646
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	1.000	0.057	0.946	0.984	0.864
FGLS	BIC	1.000	0.049	0.066	0.912	0.995
FGLS-D	BIC	1.000	0.049	0.065	0.922	0.997
DURBIN	BIC	0.050	0.047	0.053	0.050	0.050
DURBIN	AIC	0.051	0.048	0.055	0.050	0.050
Hausman 1	OLS vs FGLS	0.675	0.293	0.997	1.000	0.990
Hausman 2	DURBIN vs FGLS-D	1.000	0.053	1.000	1.000	1.000
		T=2500				
Test	Truncation	$\rho = 0$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.95$
OLS	–	1.000	0.173	1.000	1.000	1.000
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	1.000	0.062	1.000	1.000	1.000
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	1.000	0.060	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	0.050	0.997	0.998	0.981
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	1.000	0.053	1.000	1.000	1.000
FGLS	BIC	1.000	0.051	0.056	1.000	1.000
FGLS-D	BIC	1.000	0.051	0.053	1.000	1.000
DURBIN	BIC	0.049	0.048	0.049	0.052	0.050
DURBIN	AIC	0.049	0.049	0.050	0.052	0.050
Hausman 1	OLS vs FGLS	0.636	0.154	1.000	1.000	1.000
Hausman 2	DURBIN vs FGLS-D	1.000	0.051	1.000	1.000	1.000

Notes: All shocks are $N(0, 1)$ white noise. We perform 10000 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 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

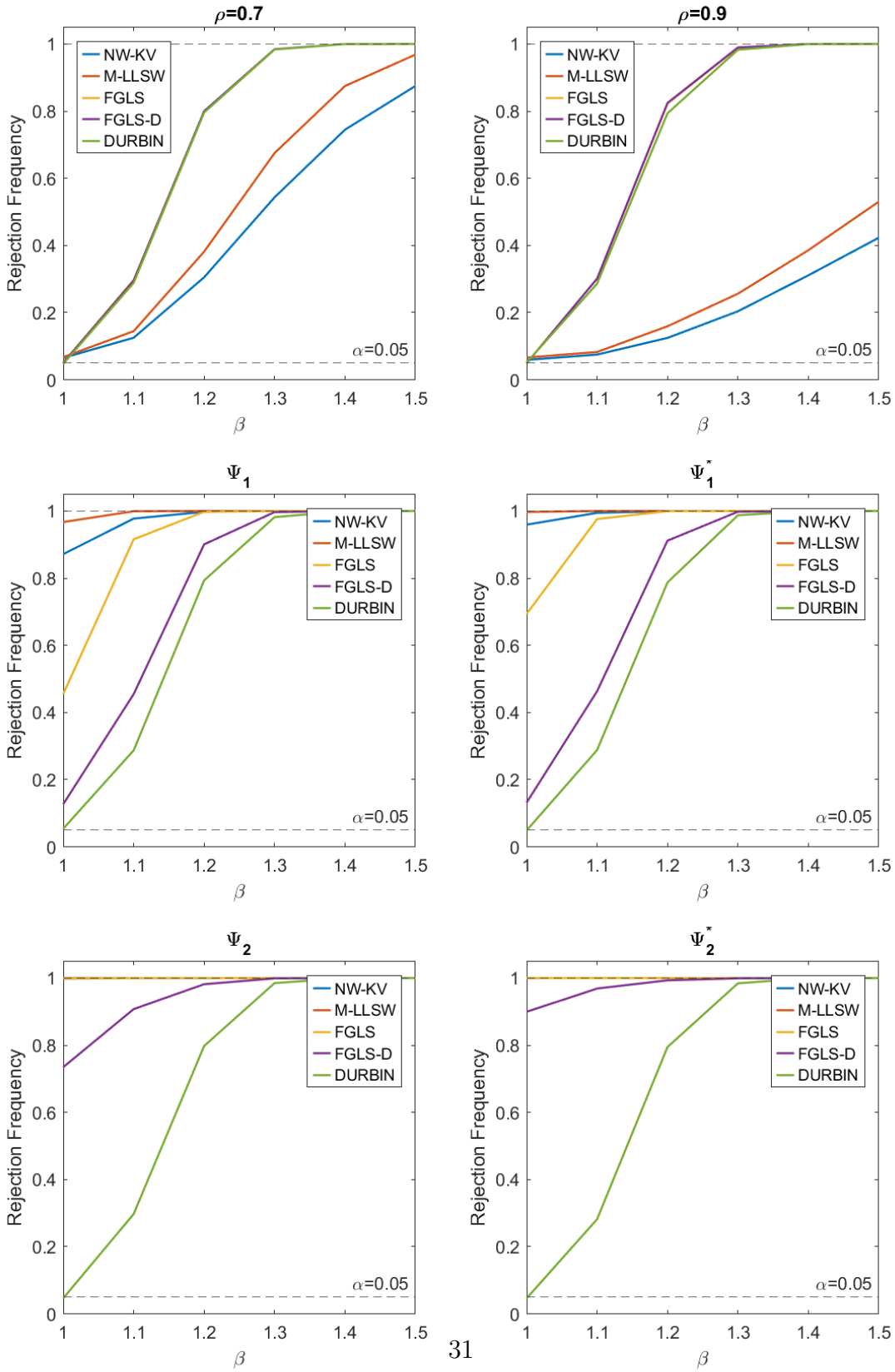
Only tests that are correctly sized are of real interest, because only correctly-sized tests produce trustworthy and interpretable rejections. As we have shown, DURBIN satisfies that requirement, whereas OLS-HAC regression does not. One could simply stop there, but it is of interest to compare rejection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

Figure 2: Empirical Rejection Frequencies of Nominal 5% t-Tests of H_0: β=1, $T = 200$



Notes: DGPs: $NDY + BD$ (top row), $NDY + GEXOG$ (middle row), and $NDY + EBD$ (bottom row). See text for details.

(higher persistence), the superiority of DURBIN is amplified, just as in the triangular case.

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 sub-optimal 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.

On the other hand, 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 generalize 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.

Appendices

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 BIC	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	1.3	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	2.5	2.9	3.0	3.0	3.1	3.1
200	Median	FGLS BIC	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	2.2	2.2	2.2	2.2	2.2
600	Median	FGLS BIC	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	1.8	1.7	1.7	1.8	1.8	1.7
2500	Median	FGLS BIC	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 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

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	1	1	1	2	2	3	3
		DURBIN BIC	0	0	1	1	2	2	2
		DURBIN AIC	0	1	2	3	5	5	5
	Mean	FGLS	1.3	1.3	1.6	2.2	2.9	3.0	3.1
		DURBIN BIC	0.1	0.3	0.8	1.4	1.9	2.0	2.0
		DURBIN AIC	2.0	2.6	3.4	4.2	5.2	5.6	5.6
200	Median	FGLS	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	1.1	1.2	1.9	3.1	4.9	5.2	5.4
		DURBIN BIC	0.0	0.8	1.4	2.3	3.3	3.5	3.5
		DURBIN AIC	1.1	2.4	3.6	5.4	9.1	10.7	11.5
600	Median	FGLS	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	1.0	1.4	2.5	4.3	7.6	8.5	8.9
		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	17.0
2500	Median	FGLS	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	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	12.0
		DURBIN AIC	0.7	3.0	4.8	8.1	18.1	24.1	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 10000 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

				T=50				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0004	-0.0013	0.0008	0.0018	0.0024	-0.0002	-0.0034
	FGLS	0.0004	-0.0010	0.0004	0.0018	0.0027	-0.0008	-0.0014
	FGLS-D	0.0004	-0.0010	0.0013	0.0009	0.0028	-0.0022	-0.0022
	DURBIN	0.0004	-0.0002	0.0028	0.0011	0.0032	-0.0020	-0.0028
MSE	OLS	0.0112	0.0165	0.0208	0.0279	0.0329	0.0344	0.0360
	FGLS	0.0121	0.0167	0.0195	0.0224	0.0231	0.0237	0.0246
	FGLS-D	0.0114	0.0168	0.0203	0.0235	0.0248	0.0256	0.0267
	DURBIN	0.0131	0.0192	0.0237	0.0275	0.0314	0.0334	0.0350
RE_{est}	OLS	0.8570	0.8612	0.8795	1.0140	1.0485	1.0300	1.0288
	FGLS	0.9250	0.8730	0.8225	0.8167	0.7355	0.7084	0.7032
	FGLS-D	0.8668	0.8755	0.8573	0.8542	0.7900	0.7654	0.7638

				T=200				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	-0.0008	-0.0005	-0.0005	0.0001	-0.0006	-0.0015	-0.0004
	FGLS	-0.0008	-0.0003	-0.0004	0.0004	-0.0005	0.0001	-0.0004
	FGLS-D	-0.0008	-0.0004	-0.0003	0.0003	-0.0004	-0.0001	-0.0002
	DURBIN	-0.0007	0.0000	-0.0001	0.0000	-0.0005	0.0000	-0.0005
MSE	OLS	0.0026	0.0039	0.0049	0.0064	0.0076	0.0083	0.0088
	FGLS	0.0026	0.0037	0.0043	0.0044	0.0040	0.0039	0.0040
	FGLS-D	0.0026	0.0037	0.0043	0.0045	0.0044	0.0045	0.0047
	DURBIN	0.0027	0.0050	0.0053	0.0056	0.0061	0.0063	0.0066
RE_{est}	OLS	0.9666	0.7802	0.9311	1.1610	1.2500	1.3162	1.3237
	FGLS	0.9824	0.7383	0.8057	0.7966	0.6489	0.6166	0.6115
	FGLS-D	0.9673	0.7484	0.8185	0.8160	0.7254	0.7078	0.7102

				T=600				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0004	0.0001	0.0003	-0.0004	0.0005	0.0001	-0.0011
	FGLS	0.0004	0.0000	0.0006	-0.0001	0.0002	0.0005	-0.0006
	FGLS-D	0.0004	0.0000	0.0005	-0.0002	0.0001	0.0005	-0.0007
	DURBIN	0.0003	0.0001	0.0007	0.0000	0.0005	0.0005	-0.0015
MSE	OLS	0.0009	0.0013	0.0017	0.0021	0.0026	0.0027	0.0029
	FGLS	0.0009	0.0012	0.0014	0.0014	0.0010	0.0009	0.0009
	FGLS-D	0.0009	0.0012	0.0014	0.0014	0.0012	0.0011	0.0011
	DURBIN	0.0009	0.0017	0.0017	0.0017	0.0018	0.0019	0.0020
RE_{est}	OLS	0.9762	0.7668	0.9680	1.2116	1.4440	1.4420	1.4588
	FGLS	0.9805	0.7113	0.7987	0.7756	0.5660	0.4812	0.4708
	FGLS-D	0.9761	0.7112	0.8060	0.7905	0.6386	0.5750	0.5733

				T=2500				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0001	0.0003	0.0000	0.0001	-0.0001	0.0002	0.0004
	FGLS	0.0001	0.0002	0.0000	0.0002	-0.0001	0.0001	0.0001
	FGLS-D	0.0001	0.0002	0.0000	0.0002	-0.0001	0.0002	0.0001
	DURBIN	0.0001	0.0002	0.0000	0.0000	0.0000	0.0003	0.0001
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	0.9948	0.7708	0.9920	1.2350	1.5383	1.5773	1.5744
	FGLS	0.9952	0.7127	0.8024	0.7610	0.4866	0.3710	0.3012
	FGLS-D	0.9948	0.7146	0.8034	0.7692	0.5248	0.4369	0.3862

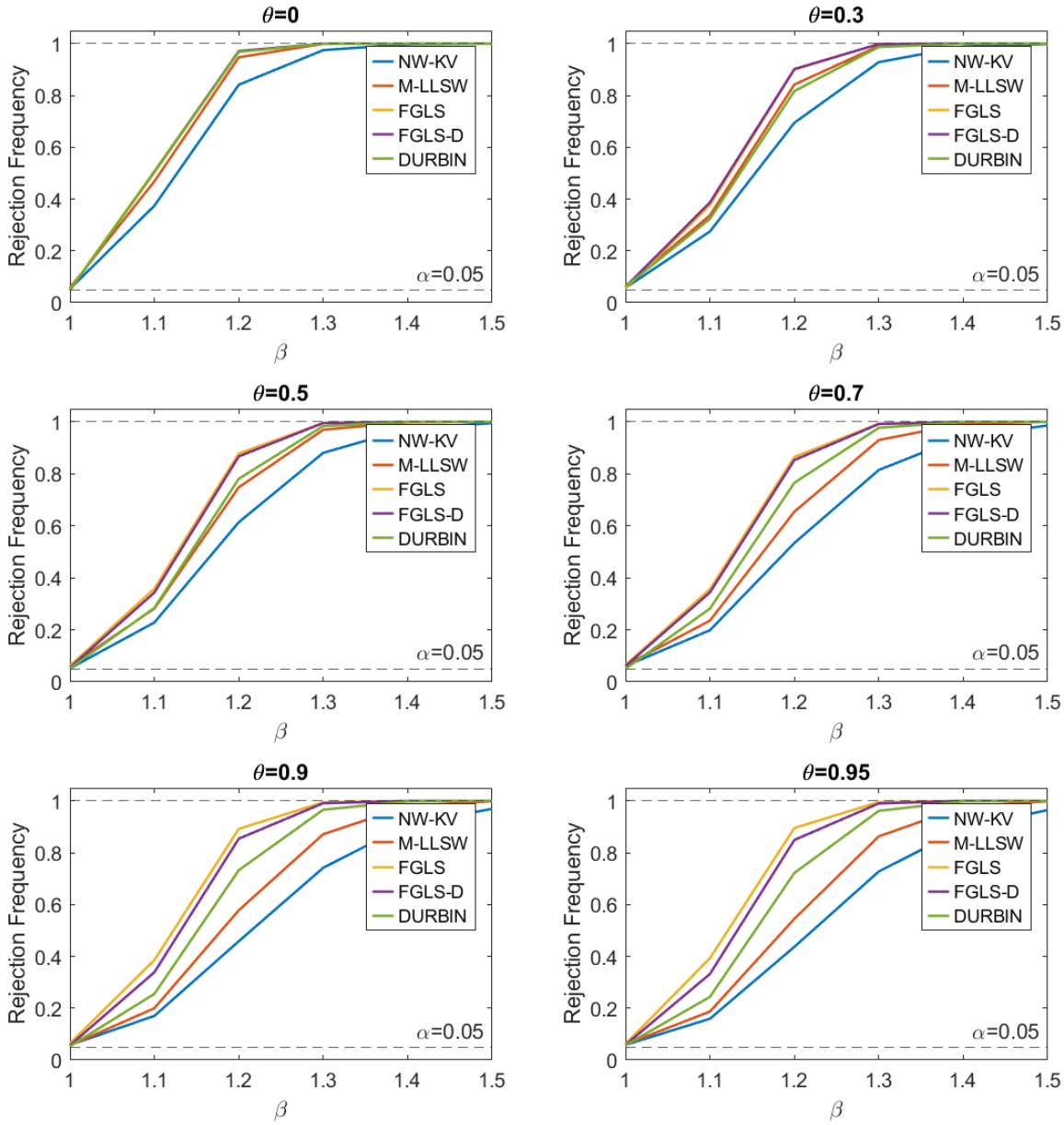
Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10000 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

T=50								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.051	0.097	0.115	0.132	0.129	0.129	0.130
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.066	0.082	0.085	0.094	0.092	0.088	0.093
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.064	0.084	0.088	0.097	0.096	0.093	0.095
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.064	0.074	0.080	0.083	0.082	0.082	0.086
NW-KV	$h = T$	0.061	0.073	0.077	0.079	0.075	0.075	0.076
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.065	0.072	0.072	0.073	0.076	0.072	0.077
FGLS	BIC	0.067	0.076	0.082	0.097	0.101	0.099	0.104
FGLS-D	BIC	0.054	0.092	0.095	0.093	0.092	0.092	0.094
DURBIN	BIC	0.060	0.096	0.094	0.075	0.075	0.074	0.078
DURBIN	AIC	0.086	0.096	0.087	0.080	0.085	0.085	0.088
Hausman 1	OLS vs FGLS		0.753	0.680	0.574	0.477	0.466	0.449
Hausman 2	DURBIN vs FGLS-D		0.047	0.073	0.088	0.097	0.099	0.095
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	0.090	0.110	0.132	0.123	0.133	0.134
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.054	0.060	0.065	0.070	0.064	0.070	0.066
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.054	0.060	0.065	0.070	0.064	0.070	0.066
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.053	0.057	0.055	0.059	0.056	0.066	0.062
NW-KV	$h = T$	0.049	0.055	0.056	0.059	0.053	0.057	0.060
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.056	0.057	0.056	0.059	0.053	0.062	0.061
FGLS	BIC	0.052	0.053	0.059	0.063	0.056	0.057	0.059
FGLS-D	BIC	0.050	0.059	0.057	0.061	0.053	0.055	0.056
DURBIN	BIC	0.052	0.062	0.050	0.051	0.055	0.055	0.055
DURBIN	AIC	0.067	0.053	0.054	0.056	0.059	0.063	0.058
Hausman 1	OLS vs FGLS		0.666	0.545	0.385	0.224	0.208	0.199
Hausman 2	DURBIN vs FGLS-D		0.049	0.055	0.059	0.059	0.064	0.060
T=600								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.052	0.094	0.117	0.125	0.132	0.130	0.134
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.054	0.057	0.063	0.066	0.060	0.060	0.066
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.054	0.056	0.062	0.065	0.060	0.058	0.065
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.055	0.055	0.056	0.059	0.055	0.054	0.057
NW-KV	$h = T$	0.052	0.050	0.047	0.054	0.052	0.051	0.049
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.055	0.053	0.055	0.057	0.054	0.052	0.055
FGLS	BIC	0.053	0.047	0.058	0.053	0.047	0.040	0.043
FGLS-D	BIC	0.052	0.046	0.058	0.050	0.047	0.039	0.041
DURBIN	BIC	0.053	0.051	0.055	0.052	0.049	0.050	0.052
DURBIN	AIC	0.067	0.051	0.054	0.056	0.048	0.050	0.052
Hausman 1	OLS vs FGLS		0.588	0.421	0.242	0.101	0.095	0.088
Hausman 2	DURBIN vs FGLS-D		0.050	0.051	0.049	0.048	0.054	0.054
T=2500								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.046	0.098	0.115	0.132	0.133	0.131	0.127
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.047	0.057	0.057	0.057	0.058	0.059	0.056
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.046	0.056	0.057	0.055	0.058	0.058	0.056
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.047	0.053	0.054	0.052	0.055	0.052	0.050
NW-KV	$h = T$	0.050	0.049	0.046	0.050	0.050	0.047	0.048
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.046	0.055	0.053	0.051	0.054	0.051	0.050
FGLS	BIC	0.046	0.055	0.051	0.053	0.047	0.045	0.034
FGLS-D	BIC	0.046	0.053	0.051	0.053	0.044	0.039	0.035
DURBIN	BIC	0.046	0.049	0.050	0.050	0.048	0.051	0.050
DURBIN	AIC	0.058	0.049	0.050	0.049	0.049	0.051	0.048
Hausman 1	OLS vs FGLS		0.473	0.291	0.119	0.067	0.057	0.054
Hausman 2	DURBIN vs FGLS-D		0.052	0.050	0.052	0.053	0.050	0.053

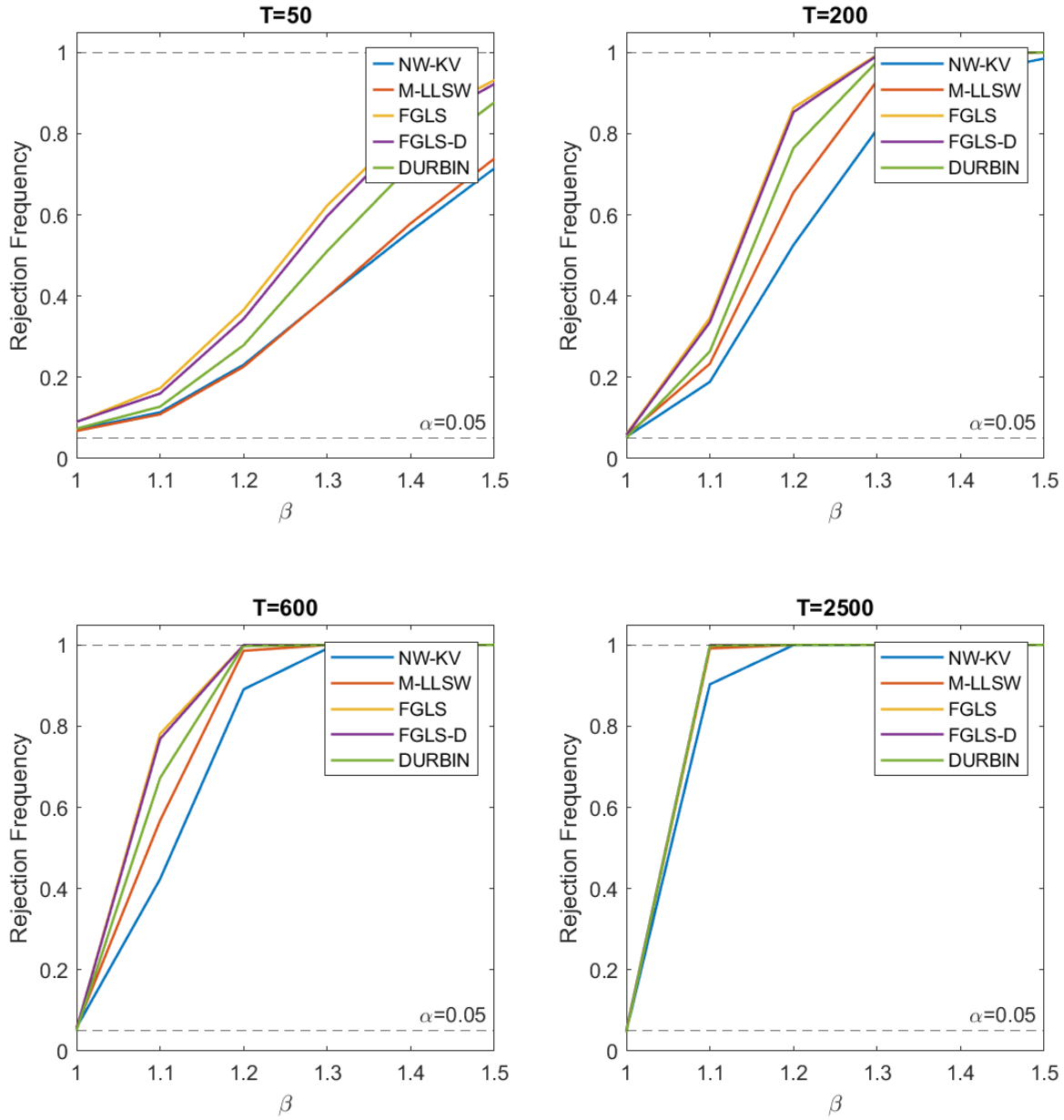
Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Figure B1: Empirical Rejection frequencies of Nominal 5% t-Test of H_0: β=1
DGP: Moving Average Disturbances, $T = 200$



Notes: 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 10000 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: β=1
DGP: Moving Average Disturbances, $\theta = 0.7$



Notes: 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 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

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	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	1.2	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	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	1.1	2.0	2.6	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	3.4	4.4	6.3	10.2	11.6	12.5
600	Median	FGLS	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	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	6.8	7.0
		DURBIN AIC	1.7	3.3	4.7	7.0	13.1	16.1	18.0
2500	Median	FGLS	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	1.0	2.8	4.2	6.7	12.3	14.6	14.9
		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 10000 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

				T=50				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	-0.0026	-0.0017	-0.0022	0.0006	0.0023	-0.0039	-0.0234
	FGLS	-0.0002	0.0009	0.0009	0.0018	0.0005	-0.0046	0.0023
	FGLS-D	0.0001	0.0006	0.0008	0.0030	-0.0006	-0.0024	0.0021
	DURBIN	0.0003	0.0004	0.0002	0.0031	0.0001	-0.0053	0.0014
MSE	OLS	0.0559	0.0945	0.1271	0.1698	0.3421	0.8138	15.4933
	FGLS	0.0242	0.0250	0.0244	0.0232	0.0276	0.0358	0.1380
	FGLS-D	0.0235	0.0231	0.0219	0.0198	0.0191	0.0199	0.0200
	DURBIN	0.0234	0.0245	0.0268	0.0292	0.0349	0.0369	0.0376
RE_{est}	OLS	2.3893	3.8547	4.7469	5.8198	9.8056	22.0441	411.8199
	FGLS	1.0369	1.0207	0.9108	0.7964	0.7900	0.9687	3.6672
	FGLS-D	1.0059	0.9441	0.8198	0.6778	0.5486	0.5383	0.5310
				T=200				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0002	0.0016	-0.0029	-0.0022	-0.0004	-0.0032	-0.0024
	FGLS	-0.0006	0.0009	-0.0010	-0.0007	-0.0001	-0.0003	-0.0005
	FGLS-D	-0.0006	0.0008	-0.0009	-0.0003	0.0000	0.0001	-0.0004
	DURBIN	-0.0006	0.0007	-0.0008	-0.0004	-0.0006	0.0010	-0.0006
MSE	OLS	0.0146	0.0238	0.0318	0.0420	0.0591	0.0932	1.0482
	FGLS	0.0051	0.0048	0.0042	0.0034	0.0026	0.0029	0.0046
	FGLS-D	0.0051	0.0048	0.0042	0.0034	0.0026	0.0027	0.0027
	DURBIN	0.0051	0.0052	0.0053	0.0057	0.0060	0.0066	0.0068
RE_{est}	OLS	2.8789	4.5810	5.9923	7.3801	9.7832	14.1853	155.2254
	FGLS	1.0071	0.9277	0.7980	0.5970	0.4372	0.4432	0.6771
	FGLS-D	1.0026	0.9303	0.7937	0.5921	0.4311	0.4053	0.3961
				T=600				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	-0.0007	0.0007	-0.0012	-0.0007	-0.0005	0.0004	-0.0019
	FGLS	0.0007	0.0000	-0.0001	0.0001	0.0000	0.0003	-0.0003
	FGLS-D	0.0007	0.0000	0.0000	0.0002	-0.0001	0.0004	-0.0003
	DURBIN	0.0007	-0.0001	0.0002	-0.0001	0.0002	0.0007	-0.0006
MSE	OLS	0.0049	0.0082	0.0106	0.0135	0.0178	0.0227	0.1273
	FGLS	0.0016	0.0015	0.0013	0.0009	0.0006	0.0005	0.0007
	FGLS-D	0.0016	0.0015	0.0013	0.0010	0.0006	0.0006	0.0005
	DURBIN	0.0016	0.0017	0.0017	0.0017	0.0019	0.0019	0.0020
RE_{est}	OLS	2.9909	4.8906	6.1570	7.8409	9.6024	11.9589	63.9303
	FGLS	0.9998	0.9163	0.7628	0.5497	0.3109	0.2779	0.3544
	FGLS-D	0.9988	0.9152	0.7705	0.5623	0.3314	0.2911	0.2746
				T=2500				
		$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
Bias	OLS	0.0002	0.0006	0.0004	0.0002	-0.0005	-0.0003	0.0020
	FGLS	-0.0003	0.0003	0.0000	0.0002	0.0001	0.0000	0.0000
	FGLS-D	-0.0003	0.0003	0.0000	0.0002	0.0001	0.0000	0.0000
	DURBIN	-0.0003	0.0004	0.0000	0.0001	0.0002	0.0002	0.0002
MSE	OLS	0.0011	0.0019	0.0026	0.0033	0.0041	0.0044	0.0111
	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	2.8997	4.6975	6.4698	7.9824	9.8683	10.4866	25.2690
	FGLS	0.9991	0.9153	0.7524	0.5116	0.2320	0.1695	0.1637
	FGLS-D	0.9989	0.9183	0.7556	0.5188	0.2460	0.1876	0.1657

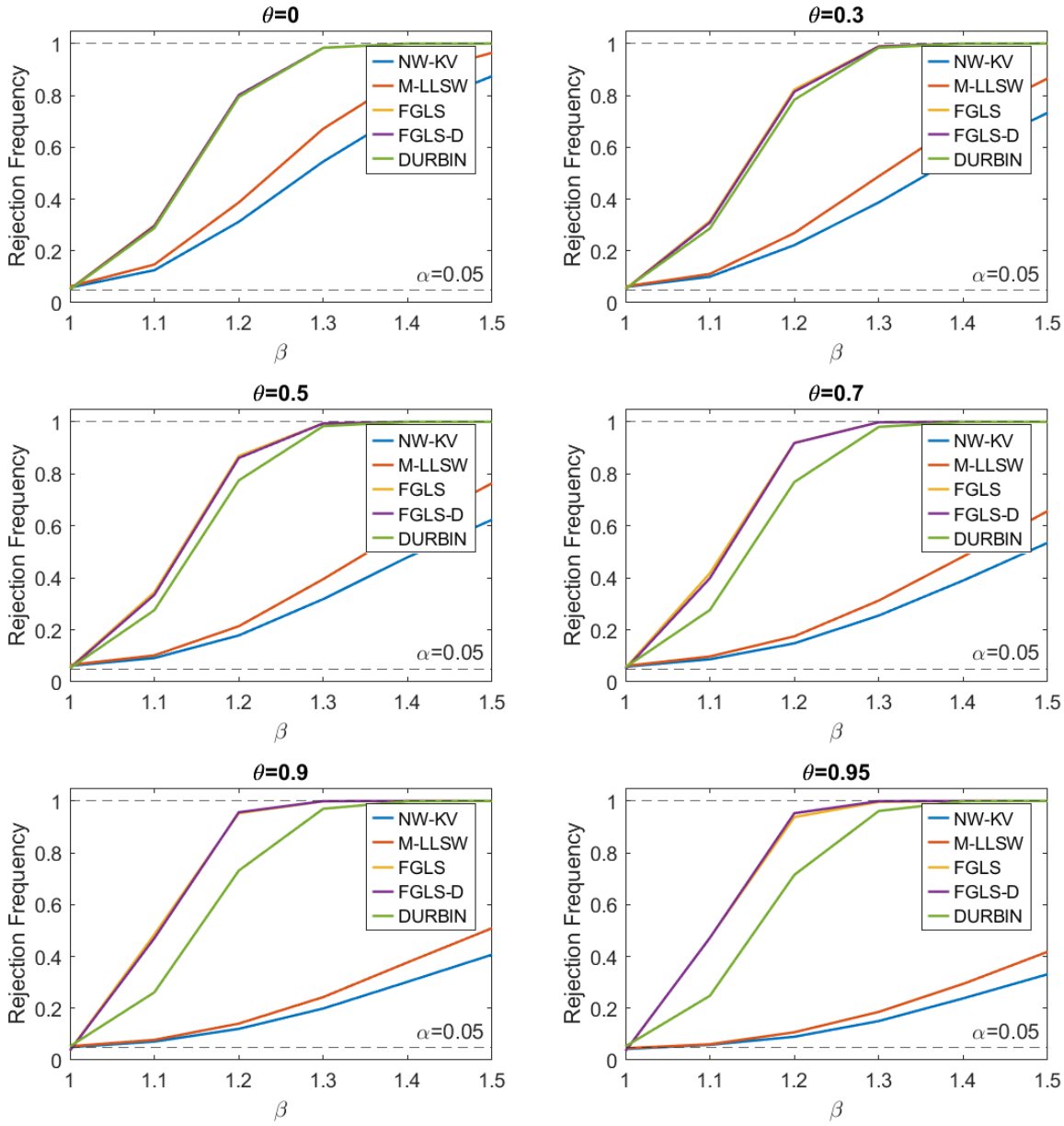
Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We select FGLS, FGLS-D, and DURBIN lag orders using BIC. RE_{est} denotes the relative estimation efficiency of DURBIN. We perform 10000 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

T=50								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.237	0.263	0.277	0.279	0.291	0.299	0.325
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.136	0.143	0.151	0.150	0.136	0.119	0.045
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.154	0.165	0.172	0.172	0.164	0.148	0.065
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.103	0.107	0.109	0.111	0.090	0.071	0.021
NW-KV	$h = T$	0.088	0.096	0.097	0.093	0.073	0.052	0.013
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.081	0.084	0.088	0.084	0.061	0.045	0.014
FGLS	BIC	0.076	0.084	0.086	0.081	0.081	0.081	0.104
FGLS-D	BIC	0.068	0.068	0.068	0.065	0.062	0.060	0.061
DURBIN	BIC	0.055	0.057	0.063	0.062	0.074	0.073	0.073
DURBIN	AIC	0.074	0.078	0.082	0.075	0.088	0.085	0.081
Hausman 1	OLS vs FGLS		0.312	0.250	0.213	0.164	0.120	0.038
Hausman 2	DURBIN vs FGLS-D		0.129	0.124	0.105	0.104	0.104	0.107
T=200								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.248	0.269	0.274	0.285	0.283	0.296	0.304
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.106	0.105	0.112	0.114	0.102	0.101	0.042
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.106	0.105	0.112	0.114	0.102	0.101	0.042
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.074	0.074	0.077	0.078	0.068	0.063	0.022
NW-KV	$h = T$	0.065	0.062	0.063	0.061	0.056	0.045	0.016
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.066	0.066	0.068	0.067	0.053	0.045	0.017
FGLS	BIC	0.051	0.053	0.051	0.049	0.038	0.040	0.039
FGLS-D	BIC	0.051	0.051	0.048	0.045	0.037	0.040	0.039
DURBIN	BIC	0.049	0.048	0.050	0.056	0.052	0.056	0.055
DURBIN	AIC	0.053	0.053	0.055	0.057	0.058	0.060	0.061
Hausman 1	OLS vs FGLS		0.140	0.123	0.108	0.095	0.087	0.034
Hausman 2	DURBIN vs FGLS-D		0.076	0.052	0.058	0.063	0.060	0.060
T=600								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.257	0.278	0.276	0.276	0.281	0.298	0.298
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.086	0.090	0.088	0.087	0.089	0.091	0.052
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.081	0.084	0.083	0.082	0.082	0.086	0.047
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.058	0.064	0.061	0.061	0.058	0.060	0.030
NW-KV	$h = T$	0.054	0.054	0.051	0.053	0.048	0.045	0.019
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.054	0.058	0.056	0.058	0.051	0.047	0.020
FGLS	BIC	0.046	0.050	0.049	0.045	0.038	0.033	0.024
FGLS-D	BIC	0.046	0.049	0.047	0.044	0.038	0.034	0.031
DURBIN	BIC	0.046	0.050	0.050	0.050	0.050	0.050	0.053
DURBIN	AIC	0.047	0.051	0.049	0.051	0.052	0.049	0.054
Hausman 1	OLS vs FGLS		0.089	0.082	0.074	0.071	0.073	0.037
Hausman 2	DURBIN vs FGLS-D		0.052	0.050	0.052	0.054	0.056	0.058
T=2500								
	Truncation	$\theta = 0$	$\theta = 0.3$	$\theta = 0.5$	$\theta = 0.7$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 0.99$
OLS	–	0.247	0.273	0.284	0.284	0.282	0.278	0.300
NW	$h = \lceil 4(T/100)^{2/9} \rceil$	0.067	0.067	0.076	0.072	0.074	0.068	0.057
NW-A	$h = \lceil 0.75T^{1/3} \rceil$	0.063	0.064	0.071	0.068	0.069	0.064	0.053
NW-LLSW	$h = \lceil 1.3T^{1/2} \rceil$	0.051	0.051	0.056	0.056	0.053	0.050	0.043
NW-KV	$h = T$	0.045	0.050	0.053	0.052	0.050	0.045	0.033
M-LLSW	$\nu = \lfloor 4(T/100)^{2/9} \rfloor$	0.048	0.049	0.054	0.053	0.052	0.045	0.029
FGLS	BIC	0.047	0.050	0.050	0.047	0.038	0.029	0.026
FGLS-D	BIC	0.047	0.050	0.049	0.047	0.039	0.033	0.032
DURBIN	BIC	0.047	0.050	0.051	0.052	0.048	0.051	0.051
DURBIN	AIC	0.047	0.051	0.052	0.053	0.049	0.050	0.050
Hausman 1	OLS vs FGLS		0.063	0.067	0.065	0.060	0.056	0.047
Hausman 2	DURBIN vs FGLS-D		0.051	0.050	0.053	0.051	0.054	0.055

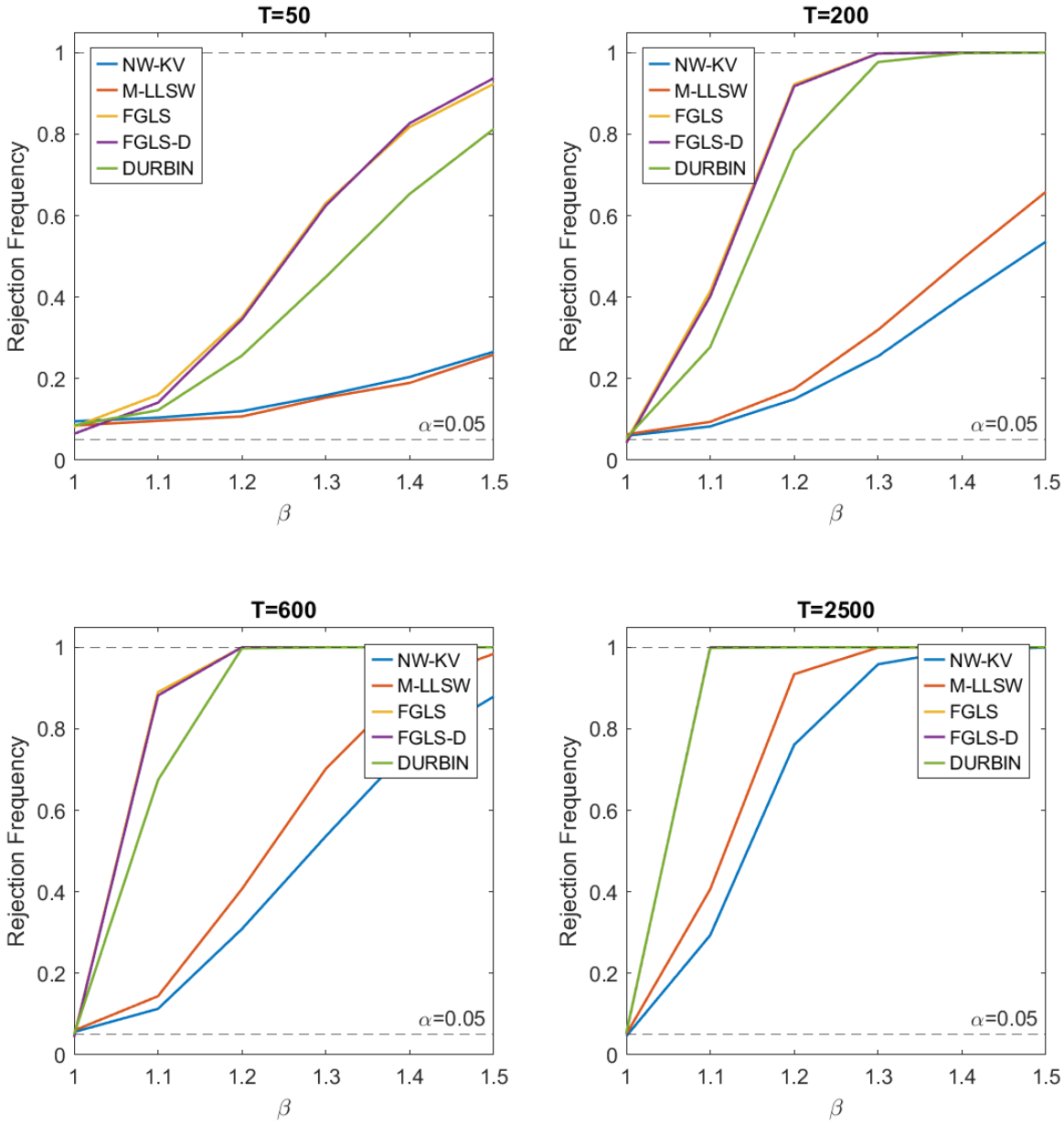
Notes: The data-generating process is $y_t = x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + \theta\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

Figure C1: Empirical Rejection Frequencies of Nominal 5% t-Test of H_0: β=1, DGP: ARMA Disturbances, $T = 200$



Notes: 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 10000 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: β=1, DGP: ARMA Disturbances, $\theta = 0.5$



Notes: The data-generating process is $y_t = \beta x_t + u_t$, $x_t = 0.7x_{t-1} + \epsilon_{x,t}$, $u_t = 0.7u_{t-1} + 0.5\epsilon_{u,t-1} + \epsilon_{u,t}$, $t = 1, ..., T$. All shocks are $N(0, 1)$ white noise. We perform 10000 Monte Carlo replications, drawing x_0 and u_0 from their stationary distributions and using common random numbers whenever possible. See text for details.

References

Amemiya, T. (1973), "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica*, 41, 723–732.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 817–858.

Andrews, D.W.K. and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

Angrist, J.D. and J.-S. Pischke (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, 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, 175–184.

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, 243–269.

Durbin, J. (1970), "Testing for Serial Correlation in Least-Squares Regression when some of the Regressors are Lagged Dependent Variables," *Econometrica*, 410–421.

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

Hannan, E.J. and M. Deistler (1988), *The Statistical Theory Of Linear Systems*, 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, 829–853.

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," *The Economic Journal*, 88, 549–563.

Kapetanios, G. and Z. Psaradakis (2016), "Semiparametric Sieve-Type Generalized Least Squares Inference," *Econometric Reviews*, 35, 951–985.

Kiefer, N.M. and T.J. Vogelsang (2002), "Heteroskedasticity-Autocorrelation Robust Standard Errors using the Bartlett Kernel without Truncation," *Econometrica*, 70, 2093–2095.

Kiefer, N.M., T.J. Vogelsang, and H. Bunzel (2000), "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 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, 541–559.

Lewis, R. and G.C. Reinsel (1985), "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393–411.

Mikusheva, A. and M. Sølvsten (2023), "Linear Regression with Weak Exogeneity," Working Paper at arXiv:2308.08958.

Montiel Olea, J.L. and M. Plagborg-Møller (2021), "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 89, 1789–1823.

Müller, U.K. (2014), "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32, 311–322.

Müller, U.K. and M.W. Watson (2022), "Spatial Correlation Robust Inference," *Econometrica*, 90, 2901–2935.

Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newey, W.K. and K.D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Perron, P. and E. González-Coya (2022), "Feasible GLS for Time Series Regression," Manuscript, Department of Economics, 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*, Butterworths, New York, 25-54.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

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