Econometric Theory, **31**, 2015, 1331–1358.
doi:10.1017/S0266466614000723

DIFFERENCING TRANSFORMATIONS AND INFERENCE IN PREDICTIVE REGRESSION MODELS

LORENZO CAMPONOVO
University of St. Gallen

The limit distribution of conventional test statistics for predictability may depend on the degree of persistence of the predictors. Therefore, diverging results and conclusions may arise because of the different asymptotic theories adopted. Using differencing transformations, we introduce a new class of estimators and test statistics for predictive regression models with Gaussian limit distribution that is instead insensitive to the degree of persistence of the predictors. This desirable feature allows to construct Gaussian confidence intervals for the parameter of interest in stationary, nonstationary, and even locally explosive settings. Besides the limit distribution, we also study the efficiency and the rate of convergence of our new class of estimators. We show that the rate of convergence is $\sqrt{n}$ in stationary cases, while it can be arbitrarily close to n in nonstationary settings, still preserving the Gaussian limit distribution. Monte Carlo simulations confirm the high reliability and accuracy of our test statistics.

1. INTRODUCTION

Consider the predictive regression model

$$y_t = \alpha + \beta x_{t-1} + u_t, \qquad t = 1, \ldots, n, \tag{1}$$

where x_{t-1} is the lagged variable assumed to predict the response variable y_t, α is the unknown intercept, β is the unknown parameter of interest, and u_t is the disturbance of the regression. Predictive regression models play an important role in many empirical economic and financial studies. Very popular examples can be found in the stock return predictability analysis, where the predictive regression model (1) is used for studying the predictive power of some economic variables, such as dividend yields, earnings price ratio, interest rates, to predict future returns; see, e.g., Fama and French (1988), Campbell and Shiller (1988),

I thank the co-editor and two anonymous referees for useful comments. The author would also like to thank Peter Phillips, Olivier Scaillet, and Fabio Trojani for many useful suggestions. Finally, we thank participants at the Econometrics Research Seminar at Yale University 2011, the 2012 Conference on Robust Econometric Methods for Modeling Economic and Financial Variables in Lisbon, the 2012 NBER-NSF Time Series Conference at Texas A&M University in College Station, the 2013 North-American Summer Meeting of the Econometric Society at University of Southern California, Los Angeles, and the 2013 European Meeting of the Econometric Society in Gothenburg for helpful comments. Part of this research was done when the author was visiting Yale University, financed by a fellowship of the Swiss National Science Foundation (SNSF). Address correspondence to Lorenzo Camponovo, School of Economics and Political Science, Department of Economics, University of St. Gallen, Bodanstrasse 6, CH-9000 St. Gallen, Switzerland; e-mail: Lorenzo.Camponovo@unisg.ch.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

Stambaugh (1999), Campbell and Yogo (2006), Jansson and Moreira (2006), and Polk, Thompson, and Vuolteenaho (2006), among others.

The usual inference approach adopted for testing predictability relies on the ordinary least squares (OLS) estimator of the parameter β and first-order asymptotic theory. Indeed, under some regularity conditions, including the stationarity of the predictors, the limit distribution of the t-statistic is standard normal. Unfortunately, the predictors considered in many empirical studies exhibit features that seem to violate these regularity conditions, and in particular the stationarity assumption. For instance, as pointed out in Torous, Valkanov, and Yan (2004), various economic variables assumed to predict stock returns are well approximated by nearly integrated processes. It turns out that in these cases first-order asymptotic theory may provide very poor approximations of the sampling distribution of the t-statistic, and inference based on this approach could be invalid.

Local-to-unity asymptotics may provide more accurate approximations of the sampling distribution of the t-statistic in integrated or nearly integrated settings; see e.g., Elliott and Stock (1994). Unfortunately, the implementation of these techniques still retains some important drawbacks. In particular, under the local-to-unity assumption that the predictors follow an autoregressive model with autoregressive parameter $\rho = 1 + c/n$, where $c \in \mathbb{R}$ is a fixed number, the limit distribution of the t-statistic still depends on the unknown parameter c that cannot be consistently estimated. To overcome this problem, Cavanagh, Elliott, and Stock (1995) and Campbell and Yogo (2006) use a Bonferroni method to construct confidence intervals for the parameter of interest. However, as recently shown in Phillips (2014), these procedures are invalid and may perform very poorly in predictive regression models with predictors less persistent than local-to-unity processes, such as the mildly integrated processes introduced in Phillips and Magdalinos (2007).[1]

The main contribution of this paper is to define a new class of test statistics for predictive regression models with Gaussian limit distribution that is robust to the degree of persistence of the predictors. More precisely, we show that the limit distribution of our test statistics is always standard normal, either by adopting first-order or local-to-unity asymptotic assumptions. It turns out that this desirable feature allows to construct Gaussian confidence intervals for the parameter of interest in stationary, nonstationary, and even locally explosive cases. The definition of our test statistics relies on the differencing approach put forward in Phillips and Han (2008) and Han, Phillips, and Sul (2011, 2014) in the context of autoregressive models. As shown in Han et al., differencing transformations can eliminate the nonstationarity of the observations and allow to introduce estimators of the autoregressive parameter with Gaussian limit distribution insensitive to the degree of persistence of the model. In this paper, we show that using differencing transformations we can achieve the same objective also for the parameter β in the predictive regression model (1).

After applying differencing transformations of order $l \geq 2$ to the predictors and the response variables, we prove the validity of a new set of moment conditions

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

for the parameter β. Finally, we introduce a new class of M-estimators of the parameter β based on these moment conditions.[2] Besides the limit distribution, we also study the efficiency and the rate of convergence of our estimators as a function of the differencing transformation order l. In particular, we show that in stationary cases, for $l/n \to 0$, as $l \to \infty$ and $n \to \infty$, the limit distribution of our estimators is Gaussian with rate of convergence $\sqrt{n}$. In contrast, in nonstationary settings and with the same assumptions on the differencing transformation order l, the limit distribution is still Gaussian but with rate of convergence arbitrarily close to n. This result shows that accurate Gaussian inference is attainable also in integrated or nearly integrated predictive regression models.

Recently, also Kostakis, Magdalinos, and Stamatogiannis (2012) and Phillips and Lee (2012) have shown the validity of standard Gaussian inference in integrated predictive regression models. More precisely, these studies extend the IVX (extended instrumental variable) procedure introduced in Magdalinos and Phillips (2009a) to predictive regressions and introduce test statistics with chi-squared limit distribution for stationary and nonstationary settings. It is interesting to note that our testing procedure and the IVX test statistics exhibit similar asymptotic properties in integrated and nearly integrated settings. Indeed, also the Monte Carlo experiments presented in Section 5 highlight some analogies between our method and the IVX approach. However, our procedure is substantially different from the IVX approach adopted in Kostakis et al. (2012) and Phillips and Lee (2012), while it is more in line with the differencing method proposed in Han et al. (2011).

The only practical issue in the implementation of our test statistics is the selection of the differencing transformation order l. Indeed, the condition $l/n \to 0$, as $l \to \infty$ and $n \to \infty$ does not provide a valid guideline for the finite sample selection of l. However, to overcome this problem we propose a data-driven method based on a calibration approach in the spirit of Loh (1987). Monte Carlo simulations in different stationary and nonstationary settings confirm the reliability and accuracy of our test statistics combined with this data-driven method.

The rest of the paper is organized as follows. In Section 2, we introduce the model. In Section 3, we define the new class of estimators and study their efficiency and rate of convergence. In Section 4, we present the data-driven method for the selection of the differencing transformation order l. In Section 5, we analyze through Monte Carlo simulations the accuracy of our approach. Finally, Section 6 concludes.

2. THE MODEL

Consider the predictive regression model (1). Furthermore, for $t = 1, \ldots, n$, we also assume that

$$x_t = \mu + z_t, \tag{2}$$

$$z_t = \rho z_{t-1} + v_t, \tag{3}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

where $\mu \in \mathbb{R}$ is the unknown intercept of the autoregressive model, ρ is the unknown degree of persistence in the variable x_t, $u_t \sim iid(0, \sigma_u^2)$ and $v_t \sim iid(0, \sigma_v^2)$, with $E[u_t v_t] = \sigma_{uv}$, and the autoregressive process (3) is initialized at some random quantity z_0, with $E[z_0^2] < \infty$.

Suppose that we are interested in testing the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$, for some given $\beta_0 \in \mathbb{R}$ (for instance, for the hypothesis of no predictability, $\beta_0 = 0$). Then, the usual econometric approach relies on the OLS estimator $\hat{\beta}_n$ of the parameter β and standard first-order asymptotic theory. Indeed, by further assuming that $|\rho| < 1$, the limit distribution of the t-statistic

$$t(\beta_0) = \frac{\hat{\beta}_n - \beta_0}{\sigma_u \left(\sum_{t=1}^n (x_{t-1} - \bar{x})^2 \right)^{-1/2}} \tag{4}$$

is standard normal, where $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_{t-1}$. Consequently, the critical values of the test can be easily computed using the standard normal distribution. However, note that for $\sigma_{uv} \neq 0$, the limit distribution of the t-statistic (4) is discontinuous at $\rho = 1$. Therefore, first-order asymptotic theory may provide a very poor approximation of the sampling distribution of $t(\beta_0)$ for values of ρ close to the unity.

In nearly integrated settings, local-to-unity asymptotics are more appropriate and reliable than first-order asymptotic theory. In particular, note that under the local-to-unity assumption that $\rho = 1 + c/n$, where $c \in \mathbb{R}$ is a fixed number, the limit distribution of $t(\beta_0)$ is instead continuous at $\rho = 1$. Unfortunately, the implementation of these techniques still retains some important drawbacks. First, the limit distribution of the t-statistic (4) still depends on the unknown parameter c that cannot be consistently estimated. Moreover, this approach is instead invalid for predictors that are less persistent than local-to-unity processes.

In the next section, we define a new class of estimators of the parameter β with limit distribution that is continuous as ρ passes through unity, and robust to the degree of persistence of the predictors. Moreover, we also show that the limit distribution is always Gaussian, either by adopting first-order or local-to-unity asymptotic assumptions.

3. DIFFERENCING TRANSFORMATIONS AND ROBUST INFERENCE

To introduce the new class of estimators of the parameter β, we extend the differencing approach proposed in Phillips and Han (2008) and Han et al. (2011) for autoregressive models to the predictive regression models (1)–(3). As shown in Han et al. (2011) in relation to the autoregressive model defined in (2) and (3), differencing transformations can eliminate the nonstationarity of the observations, and allow to introduce estimators of the parameter ρ with continuous Gaussian limit distribution as ρ passes through the unity. In this section, we show that using differencing transformations we can achieve the same objective also for the parameter β. In Section 3.1 we introduce the new class of estimators, while in Section 3.2 we study their rate of convergence.

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

3.1. The Estimators

Consider first the autoregressive process defined in (2) and (3), and the differenced observations $\Delta x_{t-l}^{t} := x_t - x_{t-l}$, where $t = l+1, \ldots, n$, and $l \geq 2$ is a fixed number. In this context, Han et al. (2011) prove the validity of the following moment conditions,

$$E[\Delta x_{t-l}^{t-1} \Delta x_{t-l-1}^{t} - \rho(\Delta x_{t-l}^{t-1})^2] = 0. \tag{5}$$

Based on equation (5), they introduce new classes of estimators of the parameter ρ. Finally, they show the suitable asymptotic properties of these estimators in stationary, nonstationary, and even locally explosive settings. Using the same approach, we introduce a new class of estimators of the parameter β with similar asymptotic properties.

Besides differenced predictors $\Delta x_{t-l}^{t} := x_t - x_{t-l}$, we start our analysis by considering also differenced response variables $\Delta y_{t-l}^{t} := y_t - y_{t-l}$, where $t = l+1, \ldots, n$, and $l \geq 2$. To define the new class of estimators of the parameter β, we have first to introduce moment conditions based on these differenced observations. In particular, we are interested in determining stationary instruments w_t, $t = l+1, \ldots, n$, strongly correlated with the stationary differenced predictors Δx_{t-l-1}^{t-1}, that satisfy the moment conditions

$$E\left[\left(\Delta y_{t-l}^{t} - \beta \Delta x_{t-l-1}^{t-1}\right) w_t\right] = 0. \tag{6}$$

Note that for $w_t := \Delta x_{t-l-1}^{t-1}$ we have

$$\begin{aligned} E\left[\left(\Delta y_{t-l}^{t} - \beta \Delta x_{t-l-1}^{t-1}\right) \Delta x_{t-l-1}^{t-1}\right] &= E\left[(u_t - u_{t-l}) \Delta x_{t-l-1}^{t-1}\right] \\ &= E\left[(u_t - u_{t-l})\left(\sum_{j=1}^{l} \rho^{j-1} v_{t-j} + (\rho^l - 1) x_{t-l-1}\right)\right] \\ &= \rho^{l-1} \sigma_{uv} \\ &\neq 0, \end{aligned}$$

unless $\rho = 0$ or $\sigma_{uv} = 0$. Therefore, Δx_{t-l-1}^{t-1}, $t = l+1, \ldots, n$, are not valid instruments. However, with slight modifications and based on the following moment equalities

$$\begin{aligned} &E\left[\left(\Delta y_{t-l}^{t} - \beta \Delta x_{t-l-1}^{t-1}\right) \Delta x_{t-l-1}^{t-l}\right] = \sigma_{uv}, \\ &E\left[\left(\Delta y_{t-l}^{t} - \beta \Delta x_{t-l-1}^{t-1}\right) \Delta x_{t-l}^{t-1}\right] = (\rho^{l-1} - 1)\sigma_{uv}, \end{aligned}$$

we can prove that the instruments $w_t := \Delta x_{t-l}^{t-1} + (1 - \rho^{l-1}) \Delta x_{t-l-1}^{t-l}$, $t = l+1, \ldots, n$, satisfy the moment conditions (6), i.e.,

$$E\left[\left(\Delta y_{t-l}^{t} - \beta \Delta x_{t-l-1}^{t-1}\right)\left(\Delta x_{t-l}^{t-1} + (1 - \rho^{l-1}) \Delta x_{t-l-1}^{t-l}\right)\right] = 0. \tag{7}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

Therefore, based on the moment conditions (7), for a fixed value of $l \geq 2$, we define the new class of estimators $\hat{\beta}_{n,l,\rho}$ of the parameter β as

$$\hat{\beta}_{n,l,\rho} = \frac{\sum_{t=l+1}^{n} \Delta y_{t-l}^{t} \left(\Delta x_{t-l}^{t-1} + (1-\rho^{l-1}) \Delta x_{t-l-1}^{t-l} \right)}{\sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1} \left(\Delta x_{t-l}^{t-1} + (1-\rho^{l-1}) \Delta x_{t-l-1}^{t-l} \right)}. \tag{8}$$

The asymptotic properties of these estimators are presented in the next theorem.

THEOREM 1. *Consider the predictive regression models (1)–(3). Let* $S_{l-1}(\rho) = \sum_{j=1}^{l-1} \rho^{j-1}$. *Moreover, for* $|\rho| < 1$, *let* $z_0 \sim N(0, \sigma_v^2/(1-\rho^2))$, *otherwise let* $z_0 = \sum_{j=0}^{\kappa} v_{-j}$, *for some fixed* κ. *Then,*

(i) *For* $\rho \in (-1, 1]$,

$$\sqrt{n} \left(\hat{\beta}_{n,l,\rho} - \beta \right) \Rightarrow N\left(0, \frac{V_{l,\rho}}{J_{l,\rho}^2} \right), \tag{9}$$

where $J_{l,\rho} = \left(\frac{1-\rho^l}{1+\rho} S_{l-1}(\rho) + S_{l-1}(\rho^2) \right) \sigma_v^2$ *and* $V_{l,\rho} = V_{l,\rho}^V - V_{l,\rho}^C$ *with*

$$V_{l,\rho}^V = \left(S_{l-1}(\rho^2) \left(1 + \left(2 - \rho^{l-1} \right)^2 \right) + \left(1 - \rho^{l-1} \right)^2 \left(\rho^{2(l-1)} + \frac{1-\rho^l}{1+\rho} S_l(\rho) \right) \right) \sigma_u^2 \sigma_v^2,$$

$$V_{l,\rho}^C = \left(2(l-1)\rho^{l-2}(2-\rho^{l-1}) \right) \sigma_{uv}^2.$$

(ii) *For* $\rho = 1 + c/n$, *where* $c \in \mathbb{R}$ *is a fixed number,*

$$\sqrt{n} \left(\hat{\beta}_{n,l,\rho} - \beta \right) \Rightarrow N\left(0, \frac{V_{l,1}}{J_{l,1}^2} \right), \tag{10}$$

where $J_{l,1} = (l-1)\sigma_v^2$ *and* $V_{l,1} = 2(l-1)(\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2)$.

Theorem 1 shows that the new class of estimators defined in (8) has a uniform continuous Gaussian limit distribution with rate of convergence $\sqrt{n}$ as ρ passes through the unity. In particular, in statement (ii) we observe that our estimators have still Gaussian limit distributions for the neighborhood of local-to-unity roots $\rho = 1 + c/n$, $c \in \mathbb{R}$. It turns out that this desirable feature allows to use Gaussian limit theory for constructing confidence intervals for the parameter β in stationary, nonstationary, and even locally explosive cases.

Note that the new class of estimators and the limit distributions in Theorem 1 depend on the unknown parameter ρ. However, ρ can be consistently estimated for instance by the OLS estimator $\hat{\rho}_n$ of the autoregressive model defined in (2) and (3). Moreover, the replacement of the unknown parameter ρ by $\hat{\rho}_n$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

still preserves the asymptotic results established in Theorem 1, as shown in the proof of Theorem 2.[3] Therefore, to test the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$, we introduce the statistic

$$T_{n,l}(\beta_0) = \sqrt{n}\left(\frac{\sqrt{V_{l,\hat{\rho}_n}}}{J_{l,\hat{\rho}_n}}\right)^{-1}\left(\hat{\beta}_{n,l,\hat{\rho}_n} - \beta_0\right). \tag{11}$$

The asymptotic properties of $T_{n,l}(\beta_0)$ are derived in the next theorem.

THEOREM 2. *Consider the setup of Theorem 1. Then, for* $\rho \in (-1, 1]$*, under* $\mathcal{H}_0 : \beta = \beta_0$*,*

$$T_{n,l}(\beta_0) \Rightarrow N(0, 1). \tag{12}$$

Using the result in Theorem 2, inference based on the test statistic (11) is straightforward. Indeed, we can construct Gaussian confidence intervals for the parameter β independently on the degree of persistence of the predictors.

3.2. Rate of Convergence

The new class of estimators has the desirable feature to be insensitive to the degree of persistence of the predictors. Unfortunately, for fixed values of the differencing transformation order l, the estimators are highly inefficient. In particular, in the nonstationary case, our estimators are infinite deficient having smaller rate of convergence than the usual OLS estimator $\hat{\beta}_n$ of the parameter β. Nevertheless, in statement (ii) of Theorem 1, we observe that the efficiency of the estimators increases as l increases. More precisely, for $\rho = 1$ the asymptotic variance of $\hat{\beta}_{n,l,1}$ is given by

$$\frac{V_{l,1}}{J_{l,1}^2} = \frac{2\left(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2\right)}{(l-1)\sigma_v^4}. \tag{13}$$

To increase the efficiency of the estimators, this result suggests to use large values of l. On the other hand, to eliminate the persistence of the predictors, l cannot increase at the same rate of the sample size n. Therefore, it is natural to determine conditions on the rate of the differencing transformation order l, which may provide efficiency gains, but still preserving the desirable continuous Gaussian limit distributions. These conditions are established in the next theorem.

THEOREM 3. *Consider the setup of Theorem 1. Then,*

(i) *For* $|\rho| < 1$*, and* $l/n \to 0$*, as* $l \to \infty$ *and* $n \to \infty$*,*

$$\sqrt{n}\left(\hat{\beta}_{n,l,\rho} - \beta\right) \Rightarrow N\left(0, (1-\rho^2)\frac{3\sigma_u^2}{2\sigma_v^2}\right). \tag{14}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

(ii) *For $\rho = 1 + c/n$, where $c \in \mathbb{R}$ is a fixed number, and $l/n \to 0$, as $l \to \infty$ and $n \to \infty$,*

$$\sqrt{nl}\left(\hat{\beta}_{n,l,\rho} - \beta\right) \Rightarrow N\left(0, \frac{2\left(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2\right)}{\sigma_v^4}\right). \tag{15}$$

(iii) *For $\rho = 1 + b/n^d$, where $b < 0$ is a fixed number, $d \in (0,1)$, and $l/n^d \to 0$, as $l \to \infty$ and $n \to \infty$,*

$$\sqrt{nl}\left(\hat{\beta}_{n,l,\rho} - \beta\right) \Rightarrow N\left(0, \frac{2\left(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2\right)}{\sigma_v^4}\right). \tag{16}$$

The results in Theorem 3 show that as $l \to \infty$, we may have Gaussian limit distributions, but with important efficiency gains. In particular, for the neighborhood of local-to-unity roots $\rho = 1 + c/n$, $c \in \mathbb{R}$ (statement (ii) of Theorem 3), for $l/n \to 0$, as $l \to \infty$ and $n \to \infty$, we observe that our estimators have still Gaussian limit distribution but with larger rate of convergence. More precisely, since $l = n^\gamma$, with $\gamma < 1$, satisfies the condition of statement (ii), it turns out that the rate of convergence $\sqrt{nl}$ can be arbitrarily close to n. Furthermore, also for mildly integrated regressors with roots of the form $\rho = 1 + b/n^d$, $b < 0$, $d \in (0,1)$ (statement (iii) of Theorem 3), for $l/n^d \to 0$, as $l \to \infty$ and $n \to \infty$, we observe that the rate of convergence of our estimators is $\sqrt{nl}$. Note that in this case, since $l = n^{d+\gamma}$, with $\gamma < 0$, satisfies the condition of statement (iii), it turns out that the rate of convergence $\sqrt{nl}$ can be arbitrarily close to $\sqrt{n^{1+d}}$. In the stationary case (statement (i) of Theorem 3), our estimators are slightly less efficient than the usual OLS estimator $\hat{\beta}_n$ of the parameter β having larger asymptotic variance.

Finally, note that for $l/n \to 0$, as $l \to \infty$ and $n \to \infty$, the limit distribution of the test statistic $T_{n,l}(\beta_0)$ defined in (11) is still standard normal as established in the following theorem.

THEOREM 4. *Consider the setup of Theorem 1. Then, for $\rho \in (-1, 1]$, and $l/n \to 0$, as $l \to \infty$ and $n \to \infty$, under $\mathcal{H}_0 : \beta = \beta_0$,*

$$T_{n,l}(\beta_0) \Rightarrow N(0,1). \tag{17}$$

In Theorem 4, we note again that the limit distribution of $T_{n,l}(\beta_0)$ is robust to the degree of persistence of the predictors. It turns out that Theorem 4 provides a very simple but powerful statistical tool for testing the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$.

The only practical issue in the implementation of our testing procedure is the selection of the differencing transformation order l. Indeed, the condition $l/n \to 0$, as $l \to \infty$ and $n \to \infty$, does not provide a valid guideline for the finite sample selection of l. The accuracy of our test statistics in stationary settings is typically less sensitive to the selection of the differencing transformation order (see also

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

the Monte Carlo analysis presented in Section 5 for more details). On the other hand, the selection of l in nonstationary settings requires some care. In statement (ii) of Theorem 3 we note that the rate of convergence of our estimators with nonstationary regressors is $\sqrt{nl}$. Therefore, in these settings the power of our test statistics tends to increase as l increases. However, when l is large the finite sample size of our test statistics can be quite distorted, as highlighted in Section 5. In the next section, we propose a data-driven method for the selection of l that aims to combine accurate finite sample size and power properties.

4. DATA-DRIVEN SELECTION OF L

In this section, we propose a data-driven method for the selection of the differencing transformation order based on a calibration approach in the spirit of Loh (1987).[4] The basic idea behind our approach is to select the largest value of l that still implies accurate confidence intervals with finite sample coverage probabilities close enough to the nominal coverage probability. The main problem in the implementation of this idea is to compute a valid approximation of the finite sample coverage probability of the confidence intervals. To this end, suppose first that the true data-generating process is known. Then, in this case it is possible to simulate the finite sample coverage probability. Indeed, using the true generating mechanism, we can first generate many random samples with the same size of the original sample. Moreover, for each random sample we can construct confidence intervals for the parameter β using the results in Theorem 2. Finally, we can approximate the finite sample coverage probability as the proportion of confidence intervals that contain the true value.

Unfortunately, the true data-generating process is unknown, and consequently it is not possible to simulate the finite sample coverage probability. However, we can replace the true generating mechanism by a consistent time series bootstrap procedure. It turns out that using a time series bootstrap, we can at least provide an estimate of the finite sample coverage probability. Therefore, we propose to select the order of the differencing transformation as the largest value of l that still implies an estimated finite sample coverage probability closes enough to the nominal coverage probability. In the next algorithm, we describe formally our approach.

ALGORITHM 5. *Let* $X_{(n)} := ((y_1, x_0), \ldots, (y_n, x_{n-1}))$ *denote a sample of size n generated according to the predictive regression models (1)–(3). Let* $L = \{l_1, \ldots, l_j\}$ *denote the set of admissible values of the differencing transformation order l.*

(1) *Compute the OLS estimates* $\hat{\alpha}_n$, $\hat{\beta}_n$, $\hat{\mu}_n$, *and* $\hat{\rho}_n$ *of the parameters* α, β, μ, *and* ρ.

(2) *Using the OLS estimates* $\hat{\alpha}_n$, $\hat{\beta}_n$, $\hat{\mu}_n$, *and* $\hat{\rho}_n$ *compute the fitted residuals* $(\hat{e}_1, \ldots, \hat{e}_n)$ *according to models (1)–(3), where* $\hat{e}_i = (\hat{u}_i, \hat{v}_i)$, $i = 1, \ldots, n$.

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

(3) *Generate K random bootstrap samples $X^*_{(n),k}$, $k = 1, \ldots, K$, according to models (1)–(3) with the OLS estimates $\hat{\alpha}_n$, $\hat{\beta}_n$, $\hat{\mu}_n$, and $\hat{\rho}_n$ and the bootstrap residuals $(\hat{e}^*_1, \ldots, \hat{e}^*_n)$.*

(4) *For each random bootstrap sample $X^*_{(n),k}$, and for each $l \in L$, construct a confidence interval $CI_{(k,l),q}$ for the parameter β using the results in Theorem 2, where $0 < q < 1$ denotes the nominal coverage probability.*

(5) *For each $l \in L$, compute $h(l) = \sharp\{\hat{\beta}_n \in CI_{(k,l),q}\}/K$.*

(6) *Let $\epsilon \geq 0$. Select the differencing transformation order l^* defined as*

$$l^* := sup\{l \in L : |h(l) - q| \leq \epsilon\}.$$

By definition, l^* is the largest value of the differencing transformation order l such that the difference between the estimated finite sample coverage probability and the nominal coverage probability is smaller than ϵ, where $\epsilon \geq 0$ may depend on the purpose of the empirical study.

5. SIMULATIONS

In this section, we study through Monte Carlo simulations the accuracy of the testing procedure introduced in Section 3. Furthermore, we provide also a comparison of our testing approach with the IVX procedure introduced in Kostakis et al. (2012) and the refined Bonferroni Q-test presented in Campbell and Yogo (2006).

In the first exercise, we study the finite sample size of our testing approach. In particular, we consider test statistics with fixed values of the differencing transformation order l, and also combined with the data-driven method presented in Section 4. To this end, we generate $N = 5,000$ samples of size $n = 500$ according to models (1)–(3), with $u_t \sim N(0,1)$, $v_t \sim N(0,1)$, $\alpha = \mu = 0$, $\sigma_{uv} \in \{-0.95, -0.9, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 0.9, 0.95\}$, $\rho \in \{0.4, 0.6, 0.8, 0.9, 0.95, 1\}$, and $\beta = 0$. We study the finite sample coverage of confidence intervals implied by Theorem 2 for fixed values of $l = 5, 50, 100$, where the parameter ρ is estimated with the OLS estimator $\hat{\rho}_n$. Furthermore, we also consider the data-driven method introduced in Section 4, with $\epsilon = 0.01$ and $L = \{5, 25, 50, 75, 100\}$. The nominal coverage probability is 95%. Tables 1–4 report the numerical results.

Tables 1 and 2 report the empirical coverages for the stationary settings $\rho = 0.4$, $0.6, 0.8$. Table 1 considers negative values of σ_{uv}, while Table 2 reports the empirical coverages for positive values of σ_{uv}. The empirical findings in both tables are qualitatively very similar. It is interesting to note that the empirical coverages for $l = 50, 100$ are always very close to the nominal coverage probability 95%. For instance, in Table 1 for $\rho = 0.8$ and $\sigma_{uv} = -0.95$ the empirical coverages for $l = 50$ and $l = 100$ are 0.9404 and 0.9350, respectively. Furthermore, in Tables 1 and 2 we can observe that the difference between the nominal coverage probability and the empirical coverages for $l = 50, 100$ is always smaller than 2%.

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

TABLE 1. Empirical Coverage. Simulated empirical coverage in the predictive regression model with $\beta = 0$, $\rho = 0.4, 0.6, 0.8$, and $\sigma_{uv} = -0.95, -0.9, -0.75, -0.5, -0.25, 0$. The differencing transformation order is $l = 5, 50, 100$. The data-driven procedure for the selection of the differencing transformation order is denoted by CM. The nominal coverage probability is 95%, while the sample size is $n = 500$. In brackets, we give the mean of the length of the confidence intervals. The number of replications is 5,000

$\sigma_{uv} = 0$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9468 (0.1970)	0.9444 (0.1767)	0.9450 (0.1512)
$l = 50$	0.9426 (0.1964)	0.9408 (0.1715)	0.9394 (0.1290)
$l = 100$	0.9308 (0.1963)	0.9326 (0.1715)	0.9356 (0.1291)
CM	0.9406 (0.1966)	0.9442 (0.1734)	0.9452 (0.1329)
$\sigma_{uv} = -0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9454 (0.1962)	0.9410 (0.1745)	0.9416 (0.1476)
$l = 50$	0.9396 (0.1964)	0.9402 (0.1715)	0.9398 (0.1291)
$l = 100$	0.9312 (0.1964)	0.9320 (0.1715)	0.9358 (0.1291)
CM	0.9400 (0.1963)	0.9418 (0.1726)	0.9422 (0.1324)
$\sigma_{uv} = -0.50$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9412 (0.1934)	0.9398 (0.1676)	0.9346 (0.1364)
$l = 50$	0.9388 (0.1963)	0.9396 (0.1716)	0.9390 (0.1291)
$l = 100$	0.9308 (0.1964)	0.9304 (0.1715)	0.9378 (0.1290)
CM	0.9402 (0.1953)	0.9360 (0.1710)	0.9388 (0.1304)
$\sigma_{uv} = -0.75$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9342 (0.1888)	0.9174 (0.1554)	0.9048 (0.1152)
$l = 50$	0.9358 (0.1964)	0.9384 (0.1715)	0.9380 (0.1291)
$l = 100$	0.9328 (0.1964)	0.9302 (0.1715)	0.9352 (0.1291)
CM	0.9376 (0.1939)	0.9394 (0.1686)	0.9398 (0.1275)
$\sigma_{uv} = -0.90$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9314 (0.1850)	0.8998 (0.1449)	0.8464 (0.0951)
$l = 50$	0.9382 (0.1964)	0.9366 (0.1715)	0.9414 (0.1290)
$l = 100$	0.9326 (0.1964)	0.9336 (0.1716)	0.9356 (0.1291)
CM	0.9404 (0.1948)	0.9382 (0.1695)	0.9390 (0.1283)
$\sigma_{uv} = -0.95$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9272 (0.1835)	0.8924 (0.1408)	0.8126 (0.0863)
$l = 50$	0.9366 (0.1963)	0.9370 (0.1715)	0.9404 (0.1291)
$l = 100$	0.9320 (0.1964)	0.9338 (0.1716)	0.9350 (0.1290)
CM	0.9394 (0.1919)	0.9380 (0.1700)	0.9354 (0.1284)

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

TABLE 2. Empirical Coverage. Simulated empirical coverage in the predictive regression model with $\beta = 0$, $\rho = 0.4, 0.6, 0.8$, and $\sigma_{uv} = 0, 0.25, 0.5, 0.75, 0.9, 0.95$. The differencing transformation order is $l = 5, 50, 100$. The data-driven procedure for the selection of the differencing transformation order is denoted by CM. The nominal coverage probability is 95%, while the sample size is $n = 500$. In brackets, we give the mean of the length of the confidence intervals. The number of replications is 5,000

$\sigma_{uv} = 0$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9468 (0.1970)	0.9444 (0.1767)	0.9450 (0.1512)
$l = 50$	0.9426 (0.1964)	0.9408 (0.1715)	0.9394 (0.1290)
$l = 100$	0.9308 (0.1963)	0.9326 (0.1715)	0.9356 (0.1291)
CM	0.9406 (0.1966)	0.9442 (0.1734)	0.9452 (0.1329)
$\sigma_{uv} = 0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9454 (0.1961)	0.9464 (0.1744)	0.9442 (0.1476)
$l = 50$	0.9428 (0.1963)	0.9420 (0.1715)	0.9390 (0.1290)
$l = 100$	0.9302 (0.1964)	0.9304 (0.1716)	0.9334 (0.1290)
CM	0.9420 (0.1962)	0.9430 (0.1726)	0.9406 (0.1359)
$\sigma_{uv} = 0.50$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9440 (0.1934)	0.9384 (0.1675)	0.9366 (0.1363)
$l = 50$	0.9394 (0.1964)	0.9406 (0.1715)	0.9396 (0.1291)
$l = 100$	0.9304 (0.1964)	0.9306 (0.1715)	0.9330 (0.1291)
CM	0.9408 (0.1953)	0.9396 (0.1702)	0.9406 (0.1312)
$\sigma_{uv} = 0.75$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9372 (0.1887)	0.9224 (0.1553)	0.9010 (0.1151)
$l = 50$	0.9386 (0.1963)	0.9412 (0.1715)	0.9402 (0.1291)
$l = 100$	0.9310 (0.1964)	0.9314 (0.1715)	0.9326 (0.1291)
CM	0.9416 (0.1940)	0.9392 (0.1685)	0.9396 (0.1275)
$\sigma_{uv} = 0.90$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9330 (0.1849)	0.9002 (0.1448)	0.8472 (0.0958)
$l = 50$	0.9382 (0.1964)	0.9388 (0.1715)	0.9376 (0.1290)
$l = 100$	0.9322 (0.1964)	0.9306 (0.1715)	0.9342 (0.1291)
CM	0.9398 (0.1933)	0.9376 (0.1694)	0.9382 (0.1286)
$\sigma_{uv} = 0.95$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$l = 5$	0.9308 (0.1835)	0.8902 (0.1407)	0.8188 (0.0863)
$l = 50$	0.9364 (0.1964)	0.9390 (0.1716)	0.9372 (0.1291)
$l = 100$	0.9318 (0.1963)	0.9296 (0.1715)	0.9340 (0.1291)
CM	0.9386 (0.1932)	0.9366 (0.1698)	0.9372 (0.1284)

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

TABLE 3. Empirical Coverage. Simulated empirical coverage in the predictive regression model with $\beta = 0$, $\rho = 0.9, 0.95, 1$, and $\sigma_{uv} = -0.95, -0.9, -0.75, -0.5, -0.25, 0$. The differencing transformation order is $l = 5, 50, 100$. The data-driven procedure for the selection of the differencing transformation order is denoted by CM. The nominal coverage probability is 95%, while the sample size is $n = 500$. In brackets, we give the mean of the length of the confidence intervals. The number of replications is 5,000

$\sigma_{uv} = 0$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9462 (0.1376)	0.9474 (0.1311)	0.9728 (0.1286)
$l = 50$	0.9364 (0.0946)	0.9370 (0.0704)	0.9372 (0.0381)
$l = 100$	0.9322 (0.0944)	0.9268 (0.0687)	0.9122 (0.0285)
CM	0.9452 (0.1125)	0.9492 (0.0981)	0.9496 (0.0787)
$\sigma_{uv} = -0.25$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9426 (0.1337)	0.9426 (0.1271)	0.9734 (0.1245)
$l = 50$	0.9370 (0.0945)	0.9342 (0.0699)	0.9340 (0.0370)
$l = 100$	0.9316 (0.0945)	0.9290 (0.0687)	0.9054 (0.0278)
CM	0.9434 (0.1105)	0.9422 (0.0965)	0.9520 (0.0796)
$\sigma_{uv} = -0.50$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9346 (0.1210)	0.9374 (0.1142)	0.9700 (0.1113)
$l = 50$	0.9356 (0.0942)	0.9276 (0.0681)	0.9210 (0.0332)
$l = 100$	0.9318 (0.0945)	0.9242 (0.0684)	0.8874 (0.0252)
CM	0.9374 (0.1037)	0.9376 (0.0905)	0.9500 (0.0757)
$\sigma_{uv} = -0.75$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9070 (0.0963)	0.9282 (0.0884)	0.9658 (0.0850)
$l = 50$	0.9332 (0.0935)	0.9136 (0.0649)	0.8822 (0.0259)
$l = 100$	0.9322 (0.0945)	0.9286 (0.0678)	0.8402 (0.0203)
CM	0.9312 (0.0935)	0.9302 (0.0759)	0.9480 (0.0690)
$\sigma_{uv} = -0.90$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.8402 (0.0707)	0.8672 (0.0608)	0.9532 (0.0561)
$l = 50$	0.9324 (0.0930)	0.9136 (0.0622)	0.8058 (0.0181)
$l = 100$	0.9318 (0.0944)	0.9296 (0.0674)	0.7614 (0.0153)
CM	0.9310 (0.0924)	0.9280 (0.0655)	0.9448 (0.0536)
$\sigma_{uv} = -0.95$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.7848 (0.0583)	0.8060 (0.0463)	0.9284 (0.0403)
$l = 50$	0.9328 (0.0928)	0.9086 (0.0611)	0.7240 (0.0140)
$l = 100$	0.9324 (0.0944)	0.9226 (0.0672)	0.6896 (0.0127)
CM	0.9300 (0.0925)	0.9214 (0.0651)	0.9254 (0.0397)

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

TABLE 4. Empirical Coverage. Simulated empirical coverage in the predictive regression model with $\beta = 0$, $\rho = 0.9, 0.95, 1$, and $\sigma_{uv} = 0, 0.25, 0.5, 0.75, 0.9, 0.95$. The differencing transformation order is $l = 5, 50, 100$. The data-driven procedure for the selection of the differencing transformation order is denoted by CM. The nominal coverage probability is 95%, while the sample size is $n = 500$. In brackets, we give the mean of the length of the confidence intervals. The number of replications is 5,000

$\sigma_{uv} = 0$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9462 (0.1376)	0.9474 (0.1311)	0.9728 (0.1286)
$l = 50$	0.9364 (0.0946)	0.9370 (0.0704)	0.9372 (0.0381)
$l = 100$	0.9322 (0.0944)	0.9268 (0.0687)	0.9122 (0.0285)
CM	0.9452 (0.1125)	0.9492 (0.0981)	0.9496 (0.0787)
$\sigma_{uv} = 0.25$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9474 (0.1336)	0.9474 (0.1271)	0.9746 (0.1245)
$l = 50$	0.9348 (0.0945)	0.9354 (0.0698)	0.9342 (0.0369)
$l = 100$	0.9298 (0.0944)	0.9284 (0.0686)	0.9042 (0.0277)
CM	0.9468 (0.1108)	0.9458 (0.0963)	0.9522 (0.0780)
$\sigma_{uv} = 0.50$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9376 (0.1210)	0.9436 (0.1141)	0.9736 (0.1113)
$l = 50$	0.9346 (0.0941)	0.9260 (0.0680)	0.9198 (0.0332)
$l = 100$	0.9326 (0.0944)	0.9242 (0.0683)	0.8854 (0.0252)
CM	0.9396 (0.1039)	0.9366 (0.0900)	0.9470 (0.0754)
$\sigma_{uv} = 0.75$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.9064 (0.0963)	0.9174 (0.0884)	0.9630 (0.0850)
$l = 50$	0.9338 (0.0934)	0.9158 (0.0649)	0.8862 (0.0260)
$l = 100$	0.9308 (0.0944)	0.9288 (0.0679)	0.8456 (0.0205)
CM	0.9306 (0.0936)	0.9298 (0.0762)	0.9452 (0.0693)
$\sigma_{uv} = 0.90$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.8400 (0.0707)	0.8638 (0.0608)	0.9476 (0.0561)
$l = 50$	0.9314 (0.0930)	0.9022 (0.0623)	0.8112 (0.0182)
$l = 100$	0.9350 (0.0944)	0.9232 (0.0675)	0.7644 (0.0154)
CM	0.9300 (0.0924)	0.9242 (0.0655)	0.9466 (0.0536)
$\sigma_{uv} = 0.95$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 1$
$l = 5$	0.7834 (0.0584)	0.8026 (0.0464)	0.9286 (0.0403)
$l = 50$	0.9274 (0.0928)	0.8982 (0.0613)	0.7318 (0.0141)
$l = 100$	0.9348 (0.0944)	0.9228 (0.0673)	0.6934 (0.0128)
CM	0.9310 (0.0926)	0.9218 (0.0651)	0.9264 (0.0399)

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

When the error terms are weakly correlated ($|\sigma_{uv}| \leq 0.5$), then also the empirical coverages for $l = 5$ are quite close to the nominal coverage probability. In contrast, for highly correlated error terms ($|\sigma_{uv}| \geq 0.75$) the empirical coverages for $l = 5$ can be quite far from 95%. For instance, in Table 1 for $\rho = 0.8$ and $\sigma_{uv} = -0.95$, the empirical coverage for $l = 5$ is 0.8126.

In Tables 1 and 2 it is interesting to note that for $l = 50, 100$ the mean of the length of the confidence intervals is almost identical. Indeed, in statement (i) of Theorem 3 we can observe that in stationary settings the asymptotic variance of our estimators converges to $(1 - \rho^2)3\sigma_u^2/2\sigma_v^2$, as $l \to \infty$, $n \to \infty$, and $l/n \to 0$. Therefore, the length of the confidence intervals depends asymptotically only on ρ, σ_u^2 and σ_v^2, but does not depend on σ_{uv} and l. Furthermore, in Tables 1 and 2 we can also note that when the error terms are highly correlated ($|\sigma_{uv}| \geq 0.75$), then the confidence intervals for $l = 5$ are typically smaller than those for $l = 50, 100$. In contrast, they are typically larger when the error terms are weakly correlated.

Finally, Tables 1 and 2 highlight the desirable accuracy of our data-driven method for the selection of the differencing transformation order l. Indeed, we can observe that using our data-driven method the empirical coverages are always very close to the nominal coverage probability 95%. In particular, in these cases the difference between the nominal coverage probability and the empirical coverages is always smaller than 1.5%.

Tables 3 and 4 report the empirical coverages for nearly integrated and integrated settings $\rho = 0.9, 0.95, 1$. Table 3 considers negative values of σ_{uv}, while Table 4 reports the empirical coverages for positive values of σ_{uv}. The empirical findings in both tables are qualitatively very similar. As expected, in the nonstationary case for $\rho = 1$, we can observe that the empirical coverages and the length of the confidence intervals decrease as l increases. Indeed, as highlighted in statement (ii) of Theorem 1, the asymptotic variance of our estimators is decreasing in l. For instance, in Table 3 for $\sigma_{uv} = -0.75$ and $\rho = 1$, the empirical coverage is 0.9658, 0.8822, and 0.8402 for $l = 5$, $l = 50$, and $l = 100$, respectively. Moreover, in the same setting the mean of the length of the confidence intervals is 0.0850, 0.0259, and 0.0203 for $l = 5$, $l = 50$, and $l = 100$, respectively.

On the other hand, for $\rho = 0.9, 0.95$ and weakly correlated error terms ($|\sigma_{uv}| \leq 0.5$), we observe that the accuracy of our testing approach is less sensitive to the selection of l. For instance, in Table 4 for $\sigma_{uv} = 0.25$ and $\rho = 0.95$, the empirical coverage is 0.9474, 0.9354, and 0.9284 for $l = 5$, $l = 50$, and $l = 100$, respectively. Moreover, in the same setting the mean of the length of the confidence intervals is 0.1271, 0.0698, and 0.0686 for $l = 5$, $l = 50$, and $l = 100$, respectively. In contrast, as highlighted also in Tables 1 and 2, the empirical coverages for small values of l and highly correlated error terms ($|\sigma_{uv}| \geq 0.75$) can be quite far from the nominal coverage probability. For instance, in Table 4 for $\rho = 0.9$ and $\sigma_{uv} = 0.95$, the empirical coverage for $l = 5$ is 0.7834. Finally, also Tables 3 and 4 confirm the desirable accuracy of our data-driven method for the selection of the differencing transformation order l. Indeed, we can observe the

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

difference between the nominal coverage probability and the empirical coverages with our data-driven method is always smaller than 2.5%.

In the second exercise, we compare the accuracy of our testing approach with the IVX procedure introduced in Kostakis et al. (2012) and the refined Bonferroni Q-test presented in Campbell and Yogo (2006). To this end, we generate $N = 5,000$ samples of size $n = 400, 800$ according to models (1)–(3), with $u_t \sim N(0,1)$, $v_t \sim N(0,1)$, $\alpha = \mu = 0$, $\sigma_{uv} \in \{-0.75, -0.5, -0.25, 0.25, 0.5, 0.75\}$, $\rho = 0.95$ (i.e, $\rho = 1 + c/n$, with $c = -20$ and $c = -40$, for $n = 400$ and $n = 800$, respectively), $\beta \in \{0, 0.02, 0.04, 0.06, 0.08\}$ when $n = 400$, and $\beta \in \{0, 0.01, 0.02, 0.03, 0.04\}$ when $n = 800$. We study the finite sample properties of tests of the null hypothesis $H_0 : \beta = 0$ versus the alternative $H_1 : \beta > 0$. We implement our testing procedure using the results in Theorem 2, where the parameter ρ is estimated with the OLS estimator $\hat{\rho}_n$, and the order of the differencing transformation is selected according to the data-driven method introduced in Section 4 with $\epsilon = 0.01$, $L = \{20, 40, 60\}$ and $L = \{40, 80, 120\}$ for $n = 400$ and $n = 800$, respectively. Tables 5 and 6 report the empirical rejection frequencies of the null hypothesis H_0 for $n = 400$ and $n = 800$, respectively. The significance level is 5%.

In Table 5, we can note that for $n = 400$ and $\beta = 0$ the testing procedures under investigation imply empirical rejection frequencies of the null hypothesis quite close to the significance level 5%. In particular, the difference between the significance level and the empirical rejection frequencies is always smaller than 2.5%. As expected, for $\beta > 0$ the empirical rejection frequencies of H_0 increase. For instance, in Table 5 we can note that when $\beta = 0.08$ the empirical rejection frequencies are always larger than 90%. The refined Bonferroni Q-test dominates our testing procedure and also the IVX approach, in particular when the error terms are negative correlated. Indeed, when $\beta > 0$ the empirical rejection frequencies of the Q-test are larger than those implied by our method and the IVX procedure. Finally, it is interesting to note that our approach and the IVX method imply very similar empirical rejection frequencies. Indeed, the difference between the empirical rejection frequencies of our procedure and the IVX approach is always smaller than 5%.

In Table 6, we note that also for $n = 800$ and $\beta = 0$ our testing procedure and the IVX approach provide empirical rejection frequencies very close to the significance level. In contrast, the refined Bonferroni Q-test tends slightly to underreject the null hypothesis. As expected, for $\beta > 0$ the empirical rejection frequencies of H_0 increase. Also in this setting it is interesting to note that our procedure and the IVX approach imply very similar empirical rejection frequencies. Indeed, the difference between the empirical rejection frequencies of our approach and the IVX method remains smaller than 5%. Also in this case, the refined Bonferroni Q-test tends to dominate our testing procedure and the IVX approach. However, the improvements using the Q-test are less pronounced than those observed in Table 5. Indeed, it is important to highlight that the Bonferroni approach introduced in Campbell and Yogo (2006) is valid only in nearly integrated settings,

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

TABLE 5. Empirical Rejection Frequencies of $H_0 : \beta = 0$. We report the empirical rejection frequencies of the null hypothesis $H_0 : \beta = 0$ versus the alternative $H_1 : \beta > 0$. We consider the predictive regression model with $\beta \in \{0, 0.02, 0.04, 0.06, 0.08\}$, $\rho = 0.95$, and $\sigma_{uv} \in \{-0.75, -0.5, -0.25, 0.25, 0.5, 0.75\}$. We consider the IVX procedure proposed in Kostakis et al. (2012) (denoted by IVX), the refined Bonferroni Q-test presented in Campbell and Yogo (2006) (denoted by CY), and our testing procedure combined with the data-driven method for the selection of l (denoted by CM). The significance level is 5%, while the sample size is $n = 400$. The number of replications is 5,000

$n = 400, \sigma_{uv} = -0.75$	$\beta = 0$	$\beta = 0.02$	$\beta = 0.04$	$\beta = 0.06$	$\beta = 0.08$
IVX	0.0432	0.1724	0.6014	0.9554	1.0000
CY	0.0538	0.3308	0.8532	0.9984	1.0000
CM	0.0708	0.2002	0.6250	0.9604	1.0000
$n = 400, \sigma_{uv} = -0.5$	$\beta = 0$	$\beta = 0.02$	$\beta = 0.04$	$\beta = 0.06$	$\beta = 0.08$
IVX	0.0451	0.1758	0.5840	0.9138	0.9932
CY	0.0584	0.3508	0.8296	0.9876	1.0000
CM	0.0672	0.1988	0.5520	0.8770	0.9822
$n = 400, \sigma_{uv} = -0.25$	$\beta = 0$	$\beta = 0.02$	$\beta = 0.04$	$\beta = 0.06$	$\beta = 0.08$
IVX	0.0482	0.1804	0.5624	0.8806	0.9814
CY	0.0616	0.3658	0.8162	0.9760	1.0000
CM	0.0558	0.1688	0.5226	0.8420	0.9616
$n = 400, \sigma_{uv} = 0.25$	$\beta = 0$	$\beta = 0.02$	$\beta = 0.04$	$\beta = 0.06$	$\beta = 0.08$
IVX	0.0490	0.1982	0.5196	0.8230	0.9532
CY	0.0398	0.2702	0.6792	0.9168	0.9894
CM	0.0594	0.1750	0.5122	0.8026	0.9350
$n = 400, \sigma_{uv} = 0.5$	$\beta = 0$	$\beta = 0.02$	$\beta = 0.04$	$\beta = 0.06$	$\beta = 0.08$
IVX	0.0548	0.1870	0.5154	0.8028	0.9356
CY	0.0294	0.1938	0.5780	0.8570	0.9702
CM	0.0682	0.1890	0.5098	0.7926	0.9306
$n = 400, \sigma_{uv} = 0.75$	$\beta = 0$	$\beta = 0.02$	$\beta = 0.04$	$\beta = 0.06$	$\beta = 0.08$
IVX	0.0446	0.1916	0.5062	0.7776	0.9258
CY	0.0254	0.1628	0.5212	0.8086	0.9462
CM	0.0720	0.2128	0.5066	0.7814	0.9282

while it is inconsistent in predictive regression models with predictors less persistent than local-to-unity processes (see also Phillips (2014) for more details on the size and power distortions of the Bonferroni Q-test in stationary settings). On the other hand, our testing procedure (and also the IVX approach) ensures reliable inference both in stationary and nonstationary settings.[5]

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

TABLE 6. Empirical Rejection Frequencies of $H_0 : \beta = 0$. We report the empirical rejection frequencies of the null hypothesis $H_0 : \beta = 0$ versus the alternative $H_1 : \beta > 0$. We consider the predictive regression model with $\beta \in \{0, 0.01, 0.02, 0.03, 0.04\}$, $\rho = 0.95$, and $\sigma_{uv} \in \{-0.75, -0.5, -0.25, 0.25, 0.5, 0.75\}$. We consider the IVX procedure proposed in Kostakis et al. (2012) (denoted by IVX), the refined Bonferroni Q-test presented in Campbell and Yogo (2006) (denoted by CY), and our testing procedure combined with the data-driven method for the selection of l (denoted by CM). The significance level is 5%, while the sample size is $n = 800$. The number of replications is 5,000

$n = 800, \sigma_{uv} = -0.75$	$\beta = 0$	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$
IVX	0.0484	0.1160	0.3502	0.7060	0.9350
CY	0.0152	0.1286	0.4924	0.8706	0.9912
CM	0.0647	0.1408	0.3616	0.7098	0.9342
$n = 800, \sigma_{uv} = -0.5$	$\beta = 0$	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$
IVX	0.0528	0.1150	0.3530	0.6826	0.9120
CY	0.0302	0.1742	0.5200	0.8446	0.9754
CM	0.0615	0.1356	0.3386	0.6508	0.8812
$n = 800, \sigma_{uv} = -0.25$	$\beta = 0$	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$
IVX	0.0526	0.1214	0.3502	0.6610	0.8822
CY	0.0384	0.2142	0.5484	0.8460	0.9684
CM	0.0498	0.1170	0.3308	0.6324	0.8512
$n = 800, \sigma_{uv} = 0.25$	$\beta = 0$	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$
IVX	0.0506	0.1262	0.3470	0.6228	0.8358
CY	0.0280	0.1796	0.4644	0.7726	0.9320
CM	0.0526	0.1226	0.3134	0.5942	0.8084
$n = 800, \sigma_{uv} = 0.5$	$\beta = 0$	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$
IVX	0.0532	0.1282	0.3404	0.6030	0.8080
CY	0.0150	0.1220	0.3812	0.7008	0.9038
CM	0.0594	0.1150	0.3156	0.5812	0.7863
$n = 800, \sigma_{uv} = 0.75$	$\beta = 0$	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$
IVX	0.0534	0.1258	0.3326	0.5856	0.7922
CY	0.0100	0.0970	0.3248	0.6554	0.8974
CM	0.0654	0.1278	0.3128	0.5744	0.7706

6. CONCLUSIONS

Predictive regression models play an important role in many empirical economic and financial studies. The usual inference approach adopted in these studies relies on OLS estimators and first-order asymptotic theory. Indeed, under some regularity conditions, including the stationarity of the predictors, the limit

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

distribution of the t-statistic is standard normal. Unfortunately, many economic variables considered as predictors exhibit features that seem to violate the stationarity assumptions. It turns out that in these cases first-order asymptotic may provide poor approximations of the sampling distribution of the t-statistic, and inference based on this approach could be invalid. Local-to-unity asymptotics may provide more accurate approximations of the sampling distribution of the t-statistic in nearly integrated settings. However, the implementation of these techniques still preserve some important drawbacks. In particular, these procedures are invalid for predictors less persistent than local-to-unity processes.

By applying differencing transformations, we introduce a new class of estimators and test statistics for predictive regression models with limit distribution that is robust to the degree of persistence of the predictors. More precisely, the limit distribution is always Gaussian, either by adopting first-order or local-to-unity asymptotic assumptions. This desirable feature allows to construct Gaussian confidence intervals for the parameter of interest in stationary, nonstationary, and even locally explosive settings. Besides the limit distribution, we also study the accuracy and rate of convergence of our estimators. In particular, we show that in stationary cases the rate of convergence is $\sqrt{n}$, while in nonstationary settings the rate of convergence can be arbitrarily close to n, still preserving the Gaussian limit distribution. These results show that accurate Gaussian inference is attainable also in integrated or nearly integrated predictive regression models.

NOTES

1. Phillips (2014) also shows that using appropriately centered statistics we can construct confidence intervals for the autoregressive parameter that are valid uniformly for $|\rho| \leq 1$; see e.g., Phillips (1987), Giraitis and Phillips (2006), and Mikusheva (2007) for the introduction of uniform inference in autoregressive models. Nevertheless, the construction of these confidence intervals can be quite cumbersome and computationally expensive.

2. Unlike Han et al. (2011), we do not aggregate the moment conditions and do not define generalized method of moments (GMM) estimators of the parameter of interest. However, the definition of GMM estimators and the analysis of their limit distribution are currently under investigation by the author.

3. For simplicity, we define (and derive the asymptotic properties of) our test statistic (11) using the OLS estimator $\hat{\rho}_n$ of the parameter ρ. However, to improve the accuracy of our approach, we can also replace $\hat{\rho}_n$ by a more accurate estimator, as for instance the fully aggregate estimator introduced in Han et al. (2011).

4. Note that a similar procedure has also been proposed for the selection of the block size in time series bootstrap context; see, e.g., Politis, Romano, and Wolf (1999).

5. Unreported Monte Carlo experiments with less persistent regressors confirm the inconsistency of the Bonferroni Q-test in stationary settings. They also confirm the accuracy and similarity of our testing procedure and the IVX approach.

REFERENCES

Campbell, J.Y. & R.J. Shiller (1988) The dividend ratio model and small sample bias: A Monte Carlo study. *Economics Letters* 29, 325–331.

Campbell, J.Y. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

Cavanagh, C., G. Elliott, & J.H. Stock (1995) Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.
Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.
Elliott, G. & J.H. Stock (1994) Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory* 10, 672–700.
Fama, E. & K. French (1988) Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–25.
Giraitis, L. & P.C.B. Phillips (2006) Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis* 27, 51–60.
Han, C., P.C.B. Phillips, & D. Sul (2011) Uniform asymptotic normality in stationary and unit root autoregression. *Econometric Theory* 27, 1117–1151.
Han, C., P.C.B. Phillips, & D. Sul (2014) X-Differencing and dynamic panel model estimation. *Econometric Theory* 30, 201–251.
Jansson, M. & M.J. Moreira (2006) Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74, 681–714.
Kostakis, A., T. Magdalinos, & M.P. Stamatogiannis (2012) Robust Econometric Inference for Stock Return Predictability. Working paper.
Loh, W.Y. (1987) Calibrating confidence coefficients. *Journal of the American Statistical Association* 82, 155–162.
Magdalinos, T. & P.C.B. Phillips (2009a) Econometric Inference in the Vicinity of Unity. CoFie Working paper, Singapore Management University.
Magdalinos, T. & P.C.B. Phillips (2009b) Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory* 25, 482–526.
Mikusheva, A. (2007) Uniform inference in autoregressive models. *Econometrica* 75, 1411–1452.
Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
Phillips, P.C.B. (2014) On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82, 1177–1195.
Phillips, P.C.B. & C. Han (2008) Gaussian inference in AR(1) time series with or without a unit root. *Econometric Theory* 24, 631–650.
Phillips, P.C.B. & J.H. Lee (2012) Predictive Regression under Varying Degrees of Persistence and Robust Long-Horizon Regression. Working paper.
Phillips, P.C.B. & T. Magdalinos (2007) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling*. Springer.
Polk, C., S. Thompson, & T. Vuolteenaho (2006) Cross-sectional forecast of the equity premium. *Journal of Financial Economics* 81, 101–141.
Stambaugh, R.F. (1999) Predictive regressions. *Journal of Financial Economics* 54, 375–421.
Torous, W., R. Valkanov, & S. Yan (2004) On predicting stock returns with nearly integrated explanatory variables. *Journal of Business* 77, 937–966.

APPENDIX: Proofs

Proof of Theorem 1. Note that

$$\hat{\beta}_{n,l,\rho} = \beta + \frac{\sum_{t=l+1}^{n} \left(u_t - u_{t-l}\right)\left(\Delta x_{t-l}^{t-1} + (1-\rho^{l-1})\Delta x_{t-l-1}^{t-l}\right)}{\sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1}\left(\Delta x_{t-l}^{t-1} + (1-\rho^{l-1})\Delta x_{t-l-1}^{t-l}\right)}. \tag{A.1}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

Thus,

$$\sqrt{n}(\hat{\beta}_{n,l,\rho}-\beta)=\frac{\frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t-u_{t-l})\left(\Delta x_{t-l}^{t-1}+(1-\rho^{l-1})\Delta x_{t-l-1}^{t-l}\right)}{\frac{1}{n}\sum_{t=l+1}^{n}\Delta x_{t-l-1}^{t-1}\left(\Delta x_{t-l}^{t-1}+(1-\rho^{l-1})\Delta x_{t-l-1}^{t-l}\right)}. \tag{A.2}$$

Finally, also note that for $l \geq 1$, $t \geq l+1$,

$$\Delta x_{t-l}^{t}=x_t-x_{t-l}=\sum_{i=0}^{l-1}\rho^i v_{t-i}+(\rho^l-1)z_{t-l}. \tag{A.3}$$

We consider first the case $|\rho|<1$.
Using (A.3), we can write the denominator of (A.2) as

$$\begin{aligned}D_{n,l,\rho}:=&\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i}+(\rho^l-1)z_{t-l-1}\right)\\&\times\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}+(\rho^{l-1}-1)z_{t-l-1}\right).\end{aligned} \tag{A.4}$$

By applying the results in Theorem 3.7 in Phillips and Solo (1992) and Theorem 20.10 in Davidson (1994), we show that a weak law of large numbers (WLLN) applies to (A.4). In particular, note that

$$\frac{1}{n}\sum_{t=l+1}^{n}z_{t-l-1}^2\to_{pr.}\frac{\sigma_v^2}{1-\rho^2}, \tag{A.5}$$

$$\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i}\right)\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}\right)\to_{pr.}S_{l-1}(\rho^2)\sigma_v^2, \tag{A.6}$$

$$\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i}\right)z_{t-l-1}\to_{pr.}0. \tag{A.7}$$

Equation (A.5) holds by Theorem 3.7 in Phillips and Solo (1992), since l is fixed. Moreover, equation (A.6) holds by the combination of Theorem 3.7 in Phillips and Solo (1992) and Theorem 20.10 in Davidson (1994). More precisely, $\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}\right)\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}\right)\to_{pr.}S_{l-1}(\rho^2)\sigma_v^2$, by Theorem 3.7 in Phillips and Solo (1992). Furthermore,

$$\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}\right)v_{t-l}=\sum_{i=0}^{l-2}\left(\frac{1}{n}\sum_{t=l+1}^{n}\rho^i v_{t-1-i}v_{t-l}\right)\to_{pr.}0 \tag{A.8}$$

by Theorem 20.10 in Davidson (1994). Similarly,

$$\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i}\right)z_{t-l-1}=\sum_{i=0}^{l-1}\left(\frac{1}{n}\sum_{t=l+1}^{n}\rho^i v_{t-1-i}z_{t-l-1}\right)\to_{pr.}0 \tag{A.9}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

also by Theorem 20.10 in Davidson (1994). It turns out that by combining (A.5), (A.6), and (A.7), we have

$$D_{n,l,\rho} \to_{pr.} \left(S_{l-1}(\rho^2) + \frac{(\rho^l - 1)(\rho^{l-1} - 1)}{1 - \rho^2} \right) \sigma_v^2 = \left(S_{l-1}(\rho^2) + S(\rho) \frac{1 - \rho^l}{1 + \rho} \right) \sigma_v^2. \quad \textbf{(A.10)}$$

Using (A.3), we can write the numerator of (A.2) as

$$R_{n,l,\rho} := \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \left(u_t - u_{t-l} \right) \left(\Delta x_{t-l}^{t-1} + (1 - \rho^{l-1}) \Delta x_{t-l-1}^{t-l} \right) \quad \textbf{(A.11)}$$

$$= \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \left(u_t - u_{t-l} \right) \left(\sum_{i=0}^{l-2} \rho^i v_{t-1-i} + (\rho^{l-1} - 1) z_{t-l-1} \right). \quad \textbf{(A.12)}$$

By considering the same arguments of Theorem 6 in Phillips and Han (2008), we show that a central limit theorem (CLT) applies to (A.12). To this end, for $t \geq l+1$, we introduce the random variables X_t and Y_t, and the sequences $\{c_j\}_{j=0}^{\infty}$ and $\{d_j\}_{j=0}^{\infty}$ defined as

$$X_t := u_t - u_{t-l} = \sum_{j=0}^{\infty} c_j u_{t-j}, \quad \textbf{(A.13)}$$

$$Y_t := \sum_{i=0}^{l-2} \rho^i v_{t-1-i} + (\rho^{l-1} - 1) z_{t-l-1} = \sum_{j=0}^{\infty} d_j v_{t-j}. \quad \textbf{(A.14)}$$

First note that $c_0 = 1$, $c_l = -1$, while $c_j = 0$, for $j \neq 0, l$. Moreover, $d_0 = 0$, $d_l = 0$, and $\sum_{j=1}^{\infty} j d_j^2 < \infty$, since $|\rho| < 1$. Using (A.13) and (A.14), we can write (A.12) as

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} X_t Y_t = & \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} c_j d_s u_{t-j} v_{t-s} \\ & + \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} c_s d_j u_{t-s} v_{t-j} \\ & + \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \sum_{j=0}^{\infty} c_j d_j u_{t-j} v_{t-j}. \end{aligned} \quad \textbf{(A.15)}$$

Since $c_j d_j = 0$, for $j \geq 0$, it turns out that also the third term in (A.15) is zero. Moreover, since $c_0 = 1$, $c_l = -1$, while $c_j = 0$, for $j \neq 0, l$, we can write the first term in (A.15) as

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \sum_{j=0}^{\infty} \sum_{s=j+1}^{\infty} c_j d_s u_{t-j} v_{t-s} \\ & = \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \left(\sum_{s=1}^{\infty} d_s u_t v_{t-s} - \sum_{s=l+1}^{\infty} d_s u_{t-l} v_{t-s} \right) \quad \textbf{(A.16)} \\ & = \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \sum_{s=1}^{\infty} (d_s - d_{s+l}) u_t v_{t-s} + o_P(1). \quad \textbf{(A.17)} \end{aligned}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

Similarly, for the second term in (A.15) we have,

$$\frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty}c_s d_j u_{t-s}v_{t-j} = \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\sum_{j=1}^{l-1}-d_j u_{t-l}v_{t-j} \tag{A.18}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\sum_{j=1}^{l-1}-d_{l-j}u_{t-j}v_t + o_P(1). \tag{A.19}$$

Therefore,

$$R_{n,l,\rho} = \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\sum_{j=1}^{\infty}(\psi_j^{(1)}u_t v_{t-j} + \psi_j^{(2)}u_{t-j}v_t) + o_P(1) \tag{A.20}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\xi_t + o_P(1), \tag{A.21}$$

where for $j \geq 1$, $\psi_j^{(1)} = d_j - d_{j+l}$, for $1 \leq j \leq l-1$, $\psi_j^{(2)} = -d_{l-j}$, for $j \geq l$, $\psi_j^{(2)} = 0$, and $\xi_t = \sum_{j=1}^{\infty}(\psi_j^{(1)}u_t v_{t-j} + \psi_j^{(2)}u_{t-j}v_t)$. In particular, note that

$$\sum_{i=1}^{\infty}\left(\psi_i^{(1)}\right)^2 + \left(\psi_i^{(2)}\right)^2 < \infty,$$

$$\sum_{i=1}^{\infty}\psi_i^{(1)}\psi_i^{(2)} < \infty.$$

We show that the first term of (A.21) follows the martingale CLT (i.e., $\frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\xi_t \Rightarrow N\left(0, V_{l,\rho}\right)$). To this end, it is necessary to show that (i) a version of the Lindeberg condition holds and (ii) $\frac{1}{n}\sum_{t=l+1}^{n}\xi_t^2 \to_{pr.} V_{l,\rho}$. The Lindeberg condition follows directly from stationarity and integrability. Furthermore, we have

$$V_{l,\rho}^V := \left(\sum_{i=1}^{\infty}\left(\psi_i^{(1)}\right)^2 + \left(\psi_i^{(2)}\right)^2\right)\sigma_u^2\sigma_v^2$$

$$= \left(S_{l-1}(\rho^2)\left(1+\left(2-\rho^{l-1}\right)^2\right) + \left(1-\rho^{l-1}\right)^2\left(\rho^{2(l-1)} + \frac{1-\rho^l}{1+\rho}S_l(\rho)\right)\right)\sigma_u^2\sigma_v^2,$$

$$V_{l,\rho}^C := 2\left(\sum_{i=1}^{\infty}\psi_i^{(1)}\psi_i^{(2)}\right)\sigma_{uv}^2$$

$$= 2(l-1)\rho^{l-2}(2-\rho^{l-1})\sigma_{uv}^2.$$

Therefore,

$$R_{n,l,\rho} \Rightarrow N(0, V_{l,\rho}), \tag{A.22}$$

where $V_{l,\rho} = V_{l,\rho}^V - V_{l,\rho}^C$. Finally, by combining (A.10) and (A.22), statement (i) of Theorem 1 is established.

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

Consider now the case $\rho = 1 + c/n$, with $c = 0$, i.e., $\rho = 1$.
Then, the denominator of (A.2) can be written as

$$D_{n,l,1} := \frac{1}{n} \sum_{t=l+1}^{n} \left(\sum_{i=0}^{l-1} v_{t-1-i} \right) \left(\sum_{i=0}^{l-2} v_{t-1-i} \right). \tag{A.23}$$

By equation (A.6), we have $D_{n,l,1} \to_{pr.} (l-1)\sigma_v^2$. Furthermore, in this case the numerator of (A.2) can be written as

$$R_{n,l,1} := \frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} (u_t - u_{t-l}) \left(\sum_{i=0}^{l-2} v_{t-1-i} \right).$$

Using the same argument adopted for (A.22), in this case we have $R_{n,l,1} \Rightarrow N(0, V_{l,1})$, where $V_{l,1} = 2(l-1)(\sigma_u^2 \sigma_v^2 - \sigma_{uv}^2)$.

Finally, consider the case $\rho = 1 + c/n$, where $c \neq 0$.
Since l is fixed, note that $\rho^l - 1 = o(1/\sqrt{n})$. Moreover, since $z_t = O_p(\sqrt{n})$, it turns out that $(\rho^l - 1) z_t = o_P(1)$. Therefore, we can write (A.2) as

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{n,l,\rho} - \beta) &= \frac{\frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} (u_t - u_{t-l}) \left(\sum_{i=0}^{l-2} v_{t-1-i} + o_P(1) \right)}{\frac{1}{n} \sum_{t=l+1}^{n} \left(\sum_{i=0}^{l-1} v_{t-1-i} + o_P(1) \right) \left(\sum_{i=0}^{l-2} v_{t-1-i} + o_P(1) \right)} \\ &= \frac{R_{n,l,1} + o_P(1)}{D_{n,l,1} + o_P(1)}. \end{aligned} \tag{A.24}$$

This concludes the proof of Theorem 1. ■

Proof of Theorem 2. Since $\hat{\rho}_n = \rho + o_P(1)$, note that

$$\hat{\beta}_{n,l,\hat{\rho}_n} = \frac{\sum_{t=l+1}^{n} \Delta y_{t-l}^{t} \left(\Delta x_{t-l}^{t-1} + (1 - (\rho + o_P(1))^{l-1}) \Delta x_{t-l-1}^{t-l} \right)}{\sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1} \left(\Delta x_{t-l}^{t-1} + (1 - (\rho + o_P(1))^{l-1}) \Delta x_{t-l-1}^{t-l} \right)}. \tag{A.25}$$

Moreover, since l is fixed, then we have $(\rho + o_P(1))^{l-1} = \rho^{l-1} + o_P(1)$. Therefore, we can rewrite (A.25) as

$$\sqrt{n}(\hat{\beta}_{n,l,\hat{\rho}_n} - \beta) = \frac{R_{n,l,\rho} + o_P(1) \left(\frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} (u_t - u_{t-l}) \Delta x_{t-l-1}^{t-l} \right)}{D_{n,l,\rho} + o_P(1) \left(\frac{1}{n} \sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1} \Delta x_{t-l-1}^{t-l} \right)}. \tag{A.26}$$

Finally, using the same arguments considered for establishing (A.10) and (A.22), we can also show that $\frac{1}{n} \sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1} \Delta x_{t-l-1}^{t-l} = O_p(1)$ and $\frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} (u_t - u_{t-l}) \Delta x_{t-l-1}^{t-l} = O_p(1)$. This concludes the proof of Theorem 2. ■

Proof of Theorem 3. We consider first the case $|\rho| < 1$.
We show that also in this case a WLLN applies to the denominator $D_{n,l,\rho}$ of (A.2) defined in equation (A.4). More precisely, note that since $\rho^l \to 0$, as $l \to \infty$, we can write the denominator of (A.2) as

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

$$D_{n,l,\rho} = \frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i} - z_{t-l-1} + o(1)z_{t-l-1}\right) \times \left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1} + o(1)z_{t-l-1}\right) \quad \textbf{(A.27)}$$

$$= \frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i} - z_{t-l-1}\right)\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1}\right) + o_P(1), \quad \textbf{(A.28)}$$

since the results established in equations (A.5), (A.8), and (A.9) are still valid, since $l/n \to 0$, as $l \to \infty$ and $n \to \infty$. Moreover, since $l \to \infty$, in this case we have

$$\frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}\right)\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i}\right) \to_{pr.} \frac{1}{1-\rho^2}\sigma_v^2. \quad \textbf{(A.29)}$$

Therefore, by combining (A.5), (A.8), (A.9), and (A.29) we have

$$D_{n,l,\rho} \to_{pr.} \frac{2}{1-\rho^2}\sigma_v^2. \quad \textbf{(A.30)}$$

Finally, we show that a CLT still applies to the numerator $R_{n,l,\rho}$ of (A.2) defined in equation (A.12). First, note that we can write $R_{n,l,\rho}$ as

$$R_{n,l,\rho} = \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t - u_{t-l})\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1} + o(1)z_{t-l-1}\right) \quad \textbf{(A.31)}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t - u_{t-l})\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1}\right) + o_P(1), \quad \textbf{(A.32)}$$

since by considering the same arguments adopted for the results in equation (A.22), we can show that $\frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t - u_{t-l})z_{t-l-1} = O_p(1)$. Finally, still by applying the results in (A.22), after some algebra we have

$$R_{n,l,\rho} \Rightarrow N\left(0, \frac{6}{1-\rho^2}\sigma_u^2\sigma_v^2\right). \quad \textbf{(A.33)}$$

Consider now the case $\rho = 1 + c/n$, with $c = 0$, i.e., $\rho = 1$.
Moreover, consider

$$\sqrt{nl}(\hat{\beta}_{n,l,1} - \beta) = \frac{\frac{1}{\sqrt{nl}}\sum_{t=l+1}^{n}\left(u_t - u_{t-l}\right)\Delta x_{t-l}^{t-1}}{\frac{1}{nl}\sum_{t=l+1}^{n}\Delta x_{t-l-1}^{t-1}\Delta x_{t-l}^{t-1}}. \quad \textbf{(A.34)}$$

We show that the denominator and numerator of (A.34) satisfy a WLLN and a CLT, respectively. To this end, we will apply results for mixingale processes. In particular, we will show that the denominator of (A.34) satisfies a WLLN by Theorem 19.13 in Davidson (1994). Finally, we will show that the numerator of (A.34) satisfies a CLT by Theorem 19.13 and Theorem 24.3 in Davidson (1994).

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

First, we consider the denominator of (A.34)

$$\tilde{D}_{n,l,1} := \frac{1}{n}\sum_{t=l+1}^{n}\frac{1}{l}\left(\sum_{i=0}^{l-1}v_{t-1-i}\right)\left(\sum_{i=0}^{l-2}v_{t-1-i}\right). \tag{A.35}$$

Note that

$$E\left[\frac{1}{l}\left(\sum_{i=0}^{l-1}v_{t-1-i}\right)\left(\sum_{i=0}^{l-2}v_{t-1-i}\right)\right] = \frac{(l-1)\sigma_v^2}{l}.$$

Next, consider the array

$$\tilde{Y}_{nt} := \frac{1}{n}\left(\frac{1}{l}\left(\sum_{i=0}^{l-1}v_{t-1-i}\right)\left(\sum_{i=0}^{l-2}v_{t-1-i}\right) - \sigma_v^2\right). \tag{A.36}$$

Then, since $l/n \to 0$, as $l \to \infty$, $n \to \infty$, it turns out that $\tilde{Y}_{nt}$ is a L_1-mixingale with respect to the natural filtration of $\{v_t\}_{t=1}^n$ with constant array $c_{nt} = \frac{1}{n}$ and $\zeta_m = O(1/l)$, for $m > l$. Therefore, by Theorem 19.13 in Davidson (1994), we have

$$\tilde{D}_{n,l,1} \to_{pr.} \sigma_v^2. \tag{A.37}$$

Finally, consider the numerator of (A.34)

$$\tilde{R}_{n,l,1} := \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t - u_{t-l})\left(\frac{1}{\sqrt{l}}\sum_{i=0}^{l-2}v_{t-1-i}\right) \tag{A.38}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\left(u_t\left(\frac{1}{\sqrt{l}}\sum_{j=1}^{l-1}v_{t-j}\right) - v_t\left(\frac{1}{\sqrt{l}}\sum_{j=1}^{l-1}u_{t-j}\right)\right) + o_p(1) \tag{A.39}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}\tilde{\xi}_t + o_p(1), \tag{A.40}$$

where $\tilde{\xi}_t = u_t\left(\frac{1}{\sqrt{l}}\sum_{j=1}^{l-1}v_{t-j}\right) - v_t\left(\frac{1}{\sqrt{l}}\sum_{j=1}^{l-1}u_{t-j}\right)$. Note that

$$E\left[\tilde{\xi}_t^2\right] = \frac{2(l-1)(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)}{l}.$$

Next, consider the array

$$\tilde{X}_{nt} := \frac{1}{n}\left(\frac{\tilde{\xi}_t^2}{2(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)} - 1\right). \tag{A.41}$$

Then, since $l/n \to 0$, as $l \to \infty$, $n \to \infty$, it turns out that $\tilde{X}_{nt}$ is a L_1-mixingale with respect to the natural filtration of $\{(u_t, v_t)\}_{t=1}^n$ with constant array $c_{nt} = \frac{1}{n}$ and $\zeta_m = O(1/l)$, for $m > l$. Therefore, by Theorem 19.13 in Davidson (1994), we have

$$\sum_{t=l+1}^{n}\tilde{X}_{nt} \to_{pr.} 0, \tag{A.42}$$

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

i.e., $\frac{1}{n}\sum_{t=l+1}^{n}\tilde{\xi}_t^2 \to_{pr.} 2(\sigma_u^2\sigma_v^2-\sigma_{uv}^2)$, and by Theorem 24.3 in Davidson, we have

$$\tilde{R}_{n,l,1} \Rightarrow N(0, 2(\sigma_u^2\sigma_v^2-\sigma_{uv}^2)). \tag{A.43}$$

Thus, by combining (A.37) and (A.43) we have

$$\sqrt{nl}(\hat{\beta}_{n,l,1}-\beta) \Rightarrow N\left(\frac{2(\sigma_u^2\sigma_v^2-\sigma_{uv}^2)}{\sigma_v^4}\right). \tag{A.44}$$

Consider now the case $\rho = 1+c/n$, where $c \neq 0$.
Note that $\rho^l - 1 = O(l/n)$. Moreover, since $z_t = O_p(\sqrt{n})$, it turns out that for $l/n \to 0$, as $l \to \infty$, $n \to \infty$, we have $\frac{1}{\sqrt{l}}(\rho^l-1)z_t = o_P(1)$. Therefore, in this case we have

$$\begin{aligned}\sqrt{nl}(\hat{\beta}_{n,l,\rho}-\beta) &= \frac{\frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t-u_{t-l})\left(\frac{1}{\sqrt{l}}\sum_{i=0}^{l-2}v_{t-1-i}+o_P(1)\right)}{\frac{1}{n}\sum_{t=l+1}^{n}\left(\frac{1}{\sqrt{l}}\sum_{i=0}^{l-1}v_{t-1-i}+o_P(1)\right)\left(\frac{1}{\sqrt{l}}\sum_{i=0}^{l-2}v_{t-1-i}+o_P(1)\right)} \\ &= \frac{\tilde{R}_{n,l,1}+o_P(1)}{\tilde{D}_{n,l,1}+o_P(1)}. \end{aligned} \tag{A.45}$$

Finally, consider the case $\rho = 1+b/n^d$, where $b<0$ and $d \in (0,1)$.
Note that $\rho^l - 1 = O(l/n^d)$. Moreover, as established in Magdalinos and Phillips (2009b), in this case $z_t = O_p(n^{d/2})$. Therefore, since $l/n^d \to 0$, as $l \to \infty$, $n \to \infty$, also in this case we have $\frac{1}{\sqrt{l}}(\rho^l-1)z_t = o_P(1)$. This concludes the proof of Theorem 3. ■

Proof of Theorem 4. We consider first the case $|\rho|<1$.
Since $\hat{\rho}_n = \rho + O_P(1/\sqrt{n})$, it turns out that for $l/n \to 0$ as $l \to \infty$, $n \to \infty$, $\hat{\rho}_n^l = o_P(1)$. Therefore, similarly to the results in (A.28) and (A.32), for $D_{n,l,\hat{\rho}_n}$ and $R_{n,l,\hat{\rho}_n}$, we have, respectively,

$$\begin{aligned} D_{n,l,\hat{\rho}_n} &= \frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i} - z_{t-l-1} + o_P(1)z_{t-l-1}\right) \\ &\quad\times\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1} + o_P(1)z_{t-l-1}\right) \\ &= \frac{1}{n}\sum_{t=l+1}^{n}\left(\sum_{i=0}^{l-1}\rho^i v_{t-1-i} - z_{t-l-1}\right)\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1}\right) + o_P(1), \\ R_{n,l,\hat{\rho}_n} &= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t-u_{t-l})\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1} + o_P(1)z_{t-l-1}\right) \\ &= \frac{1}{\sqrt{n}}\sum_{t=l+1}^{n}(u_t-u_{t-l})\left(\sum_{i=0}^{l-2}\rho^i v_{t-1-i} - z_{t-l-1}\right) + o_P(1). \end{aligned}$$

Consider now the case $\rho = 1$.
Note that in this case $\hat{\rho}_n = 1 + O_P(1/n)$. Therefore, for $l/n \to 0$, as $l \to \infty$ and $n \to \infty$, we have $\hat{\rho}_n^l = 1 + O_P(l/n) = 1 + o_P(1)$. Thus,

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press

$$\sqrt{nl}(\hat{\beta}_{n,l,\hat{\rho}_n} - \beta) = \frac{\frac{1}{\sqrt{nl}} \sum_{t=l+1}^{n} \left(u_t - u_{t-l}\right) \left[\Delta x_{t-l}^{t-1} + o_P(1) \Delta x_{t-l-1}^{t-l}\right]}{\frac{1}{nl} \sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1} \left[\Delta x_{t-l}^{t-1} + o_P(1) \Delta x_{t-l-1}^{t-l}\right]}, \tag{A.46}$$

$$= \frac{R_{n,l,1} + o_P(1)}{D_{n,l,1} + o_P(1)}, \tag{A.47}$$

since as for the result established in (A.26), we can show that $\frac{1}{n} \sum_{t=l+1}^{n} \Delta x_{t-l-1}^{t-1} \Delta x_{t-l-1}^{t-l} = O_p(1)$ and $\frac{1}{\sqrt{n}} \sum_{t=l+1}^{n} \left(u_t - u_{t-l}\right) \Delta x_{t-l-1}^{t-l} = O_p(1)$. This concludes the proof of Theorem 4. ■

https://doi.org/10.1017/S0266466614000723 Published online by Cambridge University Press