

Crump, Richard K.; Gospodinov, Nikolaj; Lopez Gaffney, Ignacio

Working Paper

A simple diagnostic for time-series and panel-data regressions

Staff Reports, No. 1132

Provided in Cooperation with:

Federal Reserve Bank of New York

Suggested Citation: Crump, Richard K.; Gospodinov, Nikolaj; Lopez Gaffney, Ignacio (2024) : A simple diagnostic for time-series and panel-data regressions, Staff Reports, No. 1132, Federal Reserve Bank of New York, New York, NY, <https://doi.org/10.59576/sr.1132>

This Version is available at:

<https://hdl.handle.net/10419/309137>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

NO. 1132
OCTOBER 2024

A Simple Diagnostic for Time-Series and Panel-Data Regressions

Richard K. Crump | Nikolay Gospodinov | Ignacio Lopez Gaffney

A Simple Diagnostic for Time-Series and Panel-Data Regressions

Richard K. Crump, Nikolay Gospodinov, and Ignacio Lopez Gaffney

Federal Reserve Bank of New York Staff Reports, no. 1132

October 2024

<https://doi.org/10.59576/sr.1132>

Abstract

We introduce a new regression diagnostic, tailored to time-series and panel-data regressions, which characterizes the sensitivity of the OLS estimate to distinct time-series variation at different frequencies. The diagnostic is built on the novel result that the eigenvectors of a random walk asymptotically orthogonalize a wide variety of time-series processes. Our diagnostic is based on leave-one-out OLS estimation on transformed variables using these eigenvectors. We illustrate how our diagnostic allows applied researchers to scrutinize regression results and probe for underlying fragility of the sample OLS estimate. We demonstrate the utility of our approach using a variety of empirical applications.

JEL classification: C12, C13, C22, C23

Key words: leave-one-out frequency approach, regression diagnostic, relative contributions of different frequencies, high time-series persistence and spurious regressions, trigonometric basis functions, orthogonalization

Crump: Federal Reserve Bank of New York (email: richard.crump@ny.frb.org). Gospodinov: Federal Reserve Bank of Atlanta (email: nikolay.gospodinov@atl.frb.org). Gaffney: Central Bank of the Republic of Argentina (email: iml2114@columbia.edu). The authors would like to thank Martín Almuzara, Matias Cattaneo, Marco Del Negro, Keshav Dogra, David Lucca, Tassos Magdalinos, Ulrich Müller, Katerina Petrova, Mikkel Plagborg-Møller, and Mark Watson for helpful comments and discussions.

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the author(s) and do not necessarily reflect the position of the Central Bank of the Republic of Argentina, Federal Reserve Bank of Atlanta, Federal Reserve Bank of New York, or the Federal Reserve System. Any errors or omissions are the responsibility of the author(s).

To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr1132.html.

1 Introduction

Time-series models are the main tool for evaluating the dynamic effects of macroeconomic shocks and policy actions. Features of economic data, such as stochastic and deterministic trends, can give rise to spurious relationships. Moreover, economic and statistical data transformations can induce different degrees of persistence in the model variables that may result in unbalanced regressions and misleading inference, or mask genuine relationships. Because of the strong dependence in economic variables, natural robustness and diagnostic checks for time-series regressions are less prevalent than in cross-sectional models.

In a cross-sectional setting, a variety of formal and informal tools are used to scrutinize the contribution of single or groups of observations to sample-wide estimates. However, these methods do not extend naturally to time-series regressions because of the time ordering of observations. For example, applied researchers, wary of influential observations affecting the external validity of their empirical results, have, for instance, compared the OLS estimator to one that is calculated with a single observation removed (i.e., the leave-one-out estimator). In contrast, when observations are dependent – as is typically the case in time-series, panel-data or spatial regression models – this approach no longer isolates the distinct contribution from a single observation on the sample estimate. Our aim is to re-purpose this useful diagnostic approach so as to accommodate these more general setups.

In this paper, we introduce a new regression diagnostic that is tailored to the time-series setting and allows researchers to assess the robustness of any set of linear regression results. Our diagnostic relies on the same leave-one-out approach as in the cross-sectional case but we apply it to data that have been suitably *rotated*. To do so, we use orthonormal trigonometric basis functions generated by the eigenvectors of the variance matrix of a random walk. Working with these rotated data allows us to interpret leave-one-out OLS estimates as revealing the contribution from different frequencies to the overall OLS estimate. Underpinning our approach is a new result showing that our choice of basis functions orthogonalizes a wide class of time series processes which ensures that each frequency’s contribution is (asymptotically) distinct. We provide “rule of thumb” bounds to guide empirical researchers on when the contribution from a single frequency is atypical and deserves further scrutiny.

We view our new diagnostic as a complement to existing robustness checks and other tools

already available in the literature such as out-of-sample evaluation, sample splitting, and checking for conventional outliers (e.g., the presence of a financial crisis in the sample). One key benefit of our diagnostic is that it is well-suited to address longstanding problems in applied time-series analysis such as the influence of low-frequency behavior on parameter estimates. Out-of-sample and sub-sample analysis, by removing a large share of observations, are ill-equipped to analyze these types of issues. As another example, the classical spurious regression problem arises because of the “...commonality of trending mechanisms in data” (Phillips, 1998, p.1299). This suggests a high sensitivity of parameter estimates to certain frequencies which can be easily assessed using our diagnostic.

We demonstrate the appeal of our new diagnostic by revisiting five recent papers in the Economics literature: Philippon and Reshef (2012), Farber et al. (2021), Barnichon and Mesters (2020), Nakamura and Steinsson (2014), and Hazell et al. (2022). In each of these applications, our diagnostic highlights important properties of the estimated specification. In particular, we show how our diagnostic can be used to assess the influence of common trending behavior on the empirical result, can illuminate cases where a single frequency is the main driver of the empirical result, and can highlight cases where the OLS point estimate is fragile with respect to multiple different frequencies. We also show how our diagnostic is a more general and principled way to assess robustness to trending behavior than the conventional use of time polynomials.

Our paper is most related to the literature on spectral and band-pass regressions. The seminal work in this literature is Hannan (1963a,b). In the Economics literature, Engle (1980b), and Engle (1980a) use the discrete Fourier transform to perform estimation in the frequency domain and to form tests comparing parameter estimates over split sub-intervals of frequencies by partitioning the spectrum (see also Engle and Gardner, 1976; Harvey, 1978). More recently, Corbae et al. (1994) and Corbae et al. (2002) extend the spectral regression framework to settings with stochastic and deterministic trends. Our paper is also related to Phillips (1998) and a series of papers starting with Müller and Watson (2008) and summarized in Müller and Watson (2024a) which exploit the representation of a time series as weighted sum of (approximately) orthogonal trigonometric basis functions. Phillips (1998) utilizes this property as an alternative device to understand the properties of spurious regressions. Müller and Watson (2024a) make use of a similar set of basis functions as in this paper but focus only on the low-frequency portion of the spectrum, developing tools for

characterizing the low-frequency variation and co-variation in economic time series.

In this paper, we propose an alternative approach to spectral regression relying on a different class of basis functions rather than the discrete Fourier transform. One important difference is that the discrete Fourier transform fails to orthogonalize highly persistent time series (Corbae et al., 2002) whereas Crump and Gospodinov (2021) show that the basis functions we use can approximately orthogonalize a highly-persistent spatial autoregression. We generalize this result to a wide class of time series processes including classical stationary processes.

In contrast to the existing literature, our interest lies in understanding if any single frequency plays a prominent role in informing the estimate of parameters of a regression model. In that sense, our paper is related, in spirit, to the leave-one-out estimation in cross-sectional applications and their associated regression diagnostics (Belsley et al., 1980).

We formally introduce our class of basis functions and motivate the form of our regression diagnostic in Section 2. In Section 3, we separately outline our diagnostic procedure for time-series, panel-data and spatial models. In Section 4, we propose conservative bounds to guide the user in the implementation of our diagnostic and in Section 5 we illustrate its application in a number of empirical examples. Section 6 concludes. Additional technical discussion, simulation evidence, and proofs of the main results are available in an Appendix.

2 Motivation

Consider the standard linear time-series model,

$$y_t = \alpha + x_t\beta + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where x_t is a scalar. Here, the parameter β could represent the slope coefficient under the assumption of a linear conditional expectation function (i.e., $\mathbb{E}[y_t|x_t] = \alpha + x_t\beta$) or the slope of the best linear predictor under regularity conditions on the joint distribution of $\{(y_t, x_t)\}_{t=1}^T$. In either case, a direct implication of equation (1) is that the *systematic* behavior of the process y_t , at all frequencies, is pinned down by the behavior at the corresponding frequency of x_t . Said differently, the linear relation imposes tight restrictions on how y_t varies at different frequencies.

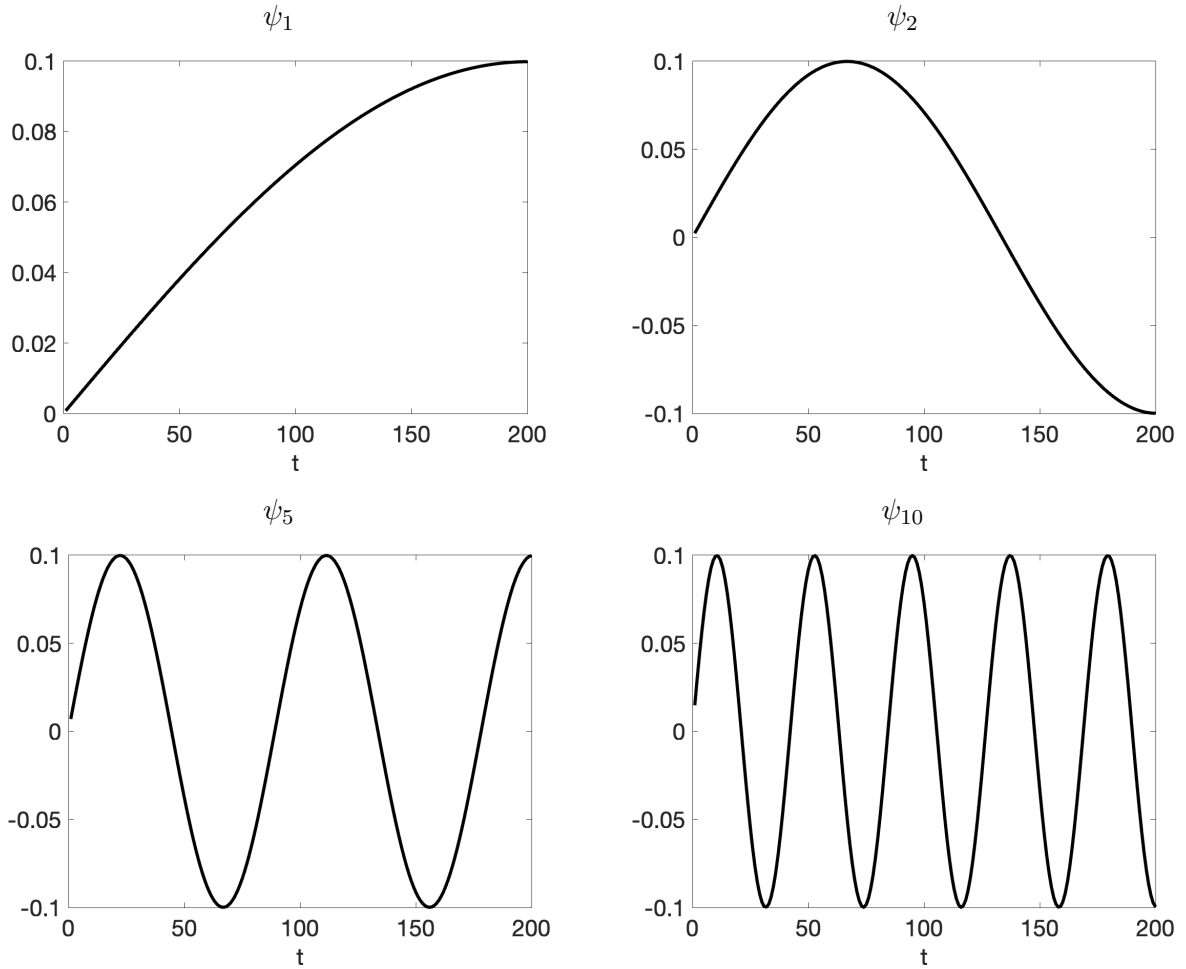
We can solidify this intuition by translating the linear model in equation (1) to the frequency

domain. Let $\Psi \equiv [\psi_1 \cdots \psi_T]$ be a $T \times T$ matrix which collects a set of trigonometric basis functions. In the next section, we will formally introduce Ψ and its associated properties, but for now it suffices to only note that Ψ is an orthonormal basis, i.e.,

$$\Psi\Psi' = \Psi'\Psi = \mathbf{I}_T. \quad (2)$$

To make things concrete, Figure 1 plots ψ_j for some selected values of j for $T = 200$.¹ Each ψ_j represents a sinusoidal function with different periodicities corresponding to a specific frequency. Clearly, the first and second basis functions (shown in the top row of Figure 1) represent the very low-frequency behavior of a series; however, as j grows the periodicity shrinks and the frequency falls. We can observe this latter property in the bottom row of Figure 1 which shows ψ_5 and ψ_{10} .

Figure 1. Examples of Basis Functions. This figure shows selected basis functions based on the choice of Ψ , as given by equation (8), for $T = 200$.



¹As will be shown in the next section, Ψ changes with T , although the shapes of the basis functions stay the same. We suppress this dependence for notational simplicity.

If we stack equation (1) as

$$y = \alpha \iota_T + X\beta + \varepsilon, \quad (3)$$

where ι_T is a $T \times 1$ vector of ones. We can then pre-multiply by Ψ' to obtain the transformed model,

$$\Psi'y = \alpha \cdot \Psi'\iota_T + \Psi'X\beta + \Psi'\varepsilon. \quad (4)$$

The model in equation (4) can be rewritten as

$$w = \alpha \cdot \zeta + Z\beta + u, \quad (5)$$

where $w = \Psi'y$, $\zeta = \Psi'\iota_T$, $Z = \Psi'X$, and $u = \Psi'\varepsilon$. To interpret these new objects, note that

$$w = \Psi'y = (\Psi'\Psi)^{-1}\Psi'y, \quad (6)$$

using equation (2). In words, w is the $(T \times 1)$ OLS coefficient obtained by regressing y on each ψ_j for $j = 1, \dots, T$. For example, w_1 is the OLS coefficient from regressing y_t on $\psi_{1,t}$ (see the top left plot in Figure 1) and represents the lowest frequency contribution to the process y_t .

Equation (5) shows how the coefficient β , just as in the time domain, connects the y_t and x_t in the frequency domain as well. However, in the frequency domain we interpret each observation in w and z differently. In particular, consider the leave-one-out (LOO) OLS estimator, $\hat{\beta}_{(-j)}$, obtained by removing the j th observation in equation (5). This LOO estimator, $\hat{\beta}_{(-j)}$, is the estimate of β when we omit information in the frequency corresponding to the j th basis function ψ_j . We can then make the comparison to $\hat{\beta}$ by inspecting the distance,

$$\hat{\beta} - \hat{\beta}_{(-j)}. \quad (7)$$

Thus, the expression in equation (7) represents the impact of information from this specified frequency on the OLS estimator. This follows because the OLS estimator of equation (3) and equation (5) are analytically equivalent (see equation (17) below); consequently, $\hat{\beta}$ can be used as the refer-

ence point. If the gap in equation (7) is “large” (by some metric), then that implies the specific frequency has an outsize influence on the OLS estimate.

2.1 A Useful Transformation

We now define our choice of Ψ . Let $x = (x_1, \dots, x_T)'$ be a Gaussian random walk with an initial condition of zero and innovation variance σ^2 . Then, it is straightforward to show that $x \sim \mathcal{N}(0, \Sigma)$, where $\Sigma = \sigma^2 \cdot L_T L_T'$ and L_T is a $T \times T$ lower triangular matrix of ones. Define the following orthogonal trigonometric basis, $\psi_j = (\psi_{1,j}, \dots, \psi_{T,j})'$, where

$$\psi_{h,j} = \frac{2}{\sqrt{2T+1}} \sin\left(\frac{h(2j-1)\pi}{2T+1}\right). \quad (8)$$

Following Tanaka (2017), for example, it can be shown that

$$\Psi' \Sigma \Psi = \Lambda, \quad (9)$$

where Λ is a diagonal matrix with j th diagonal element given by

$$\lambda_j = \frac{\sigma^2}{2 - 2 \cos\left(\frac{(2j-1)\pi}{2T+1}\right)}. \quad (10)$$

Said differently, each ψ_j is an eigenvector of the matrix Σ with an associated eigenvalue of λ_j . Let $z = \Psi' x$ and note that $\mathbb{V}(z) = \Lambda$; thus, this transformation diagonalizes the variance matrix of the random walk.

We may combine the results thus far to obtain the following characterization of x :

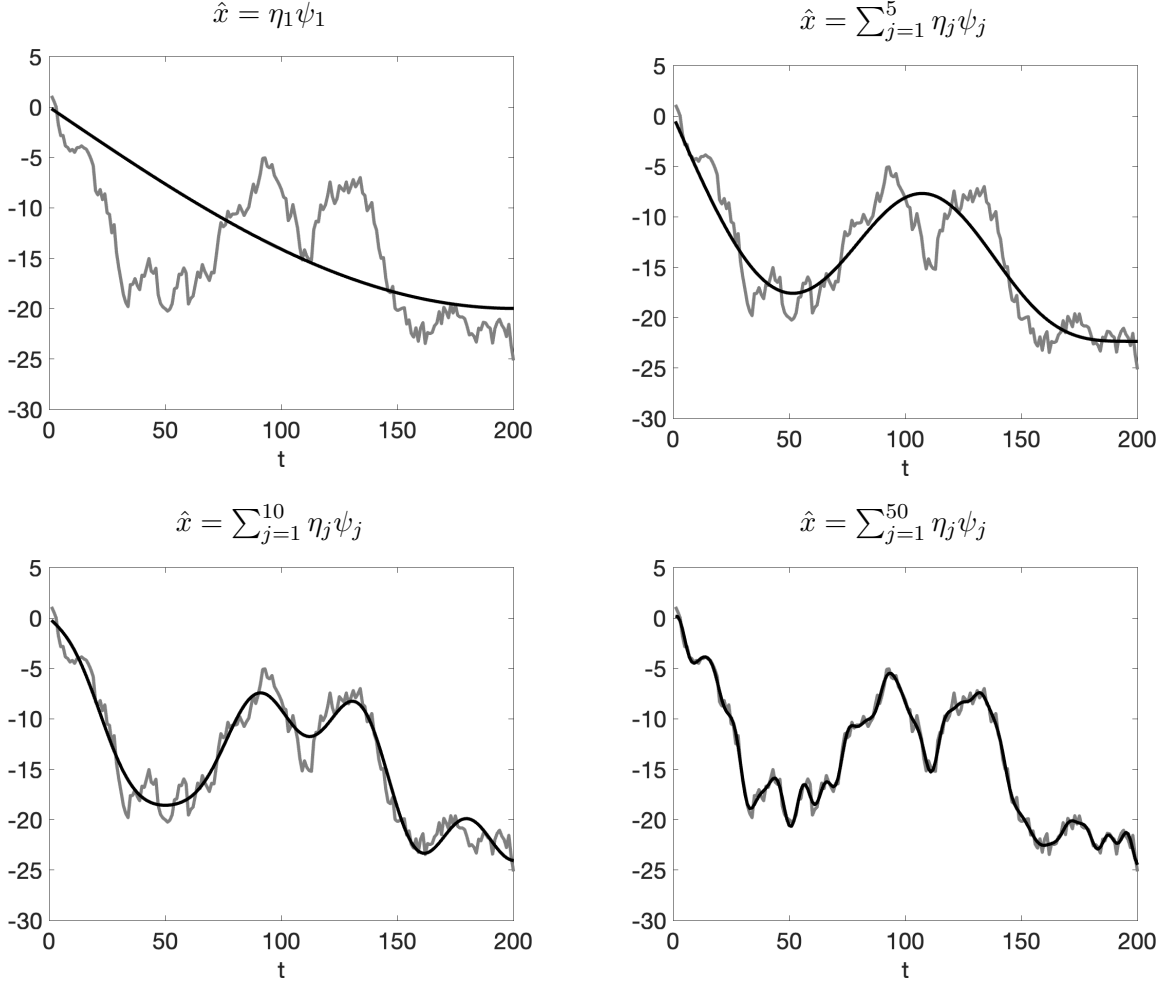
$$x = (\Psi \Psi') x = \Psi (\Psi' x) = \Psi z = \sum_{j=1}^T \psi_j z_j, \quad (11)$$

where z_j is the j th element of z . This implies an alternative representation of a Gaussian random walk of length T as

$$x =_d \sum_{j=1}^T \eta_j \psi_j, \quad (12)$$

where $\eta_j \sim \mathcal{N}(0, \lambda_j)$ and $=_d$ denotes equivalence in distribution.

Figure 2. Approximation of a Random Walk Process. This figure shows different approximations of a realized random walk process ($T = 200$) based on the representation given in equation (12). The grey line denotes the realized time series whereas the black line denotes the approximation using different choices for the number of basis functions.



Equation (12) demonstrates that we can represent a Gaussian random walk as a weighted linear combination of these (deterministic) basis functions, where the weights are independent (but not identically distributed) Gaussian variables. To cement this intuition, Figure 2 shows a particular realization of a random walk with $T = 200$ for different partial sums of the summation in equation (12). The figure illustrates how these random coefficients play the role of determining the realized properties of the time series. With just the first basis function (top left chart of Figure 2), we can trace the broad downward trend in the series.² When we add the next few basis functions, the low-frequency “cyclical” behavior becomes established (top right chart). By continuing to add

²The periodicity of each basis function can be obtained from equation (8) as $\frac{2(2T+1)}{2j-1}$. For example, when $T = 200$ and $j = 10$ (i.e., ψ_{10} from Figure 1), then the period is approximately 42. This is most easily observed by the gap between two peaks in the bottom right chart of Figure 1.

higher-frequencies, we can fully characterize the time-series behavior of the underlying process (bottom row of Figure 2).

Conversely, suppose that x now follows a white noise process with innovation variance σ^2 so that $x \sim \mathcal{N}(0, \Sigma)$, where $\Sigma = \sigma^2 \cdot I_T$. By equation (2), we have that

$$\Psi' \Sigma \Psi = \sigma^2 \cdot I_T. \quad (13)$$

Said differently, each ψ_j is also an eigenvector of the matrix Σ with an associated (repeated) eigenvalue of σ^2 . If $z = \Psi'x$, we have that $\mathbb{V}(z) = \sigma^2 I_T$; thus, this transformation also orthogonalizes the variance matrix of a white noise process. More generally, the following lemma shows that Ψ asymptotically orthogonalizes a large class of weakly-dependent time series process.

Lemma 1. *Let x be a strictly stationary time series with associated autocovariance function $\gamma(r) = \text{COV}(x_t, x_{t-r})$. Assume that $r^a \cdot \gamma(r) \rightarrow c_\gamma$ as $r \rightarrow \infty$ for $|c_\gamma|$ bounded and $a > 1$. Then, for all $j, k \in \{1, \dots, T\}$ with $j \neq k$,*

- (i) $\text{COV}(\psi'_j x, \psi'_k x) = O(T^{1-a})$ when $a \in (1, 2)$;
- (ii) $\text{COV}(\psi'_j x, \psi'_k x) = O(\ln(T)T^{-1})$ when $a = 2$;
- (iii) $\text{COV}(\psi'_j x, \psi'_k x) = O(T^{-1})$ when $a > 2$.

Proof: See Appendix A.4.

Lemma 1 shows that for any two vectors ψ_j and ψ_k ($j \neq k$), the weighted sums, $\psi'_j x$ and $\psi'_k x$, have a covariance which converges to zero with the sample size. Further, Lemma 1 characterizes how the rate of decay of the autocovariance function affects the speed of convergence at which the basis functions suppress covariation in the transformed process. When a is small, so that $\gamma(r)$ converges to zero relatively slowly as r grows, then the rate of convergence of $\text{COV}(\psi'_j x, \psi'_k x)$ to zero is slower. In contrast, for stationary ARMA processes which feature exponential decay in their autocovariance function, the convergence occurs at the rate T^{-1} . Since Lemma 1 deals with strictly stationary processes, we can immediately conclude that $\text{CORR}(\psi'_j x, \psi'_k x)$ inherits the same rates as given for $\text{COV}(\psi'_j x, \psi'_k x)$.

Lemma 1 includes all stationary processes with absolutely summable autocovariances but the assumptions rule out long-memory processes which are characterized by $\sum_r |\gamma(r)| = \infty$. However,

a direct implication of Lemma 1 is that long-memory time series generated by independent sums of granular short-memory processes will also be asymptotically orthogonalized by Ψ . For example, consider the well-known result of Granger (1980) that a sum of sufficiently many independent stationary processes exhibits long-memory. When $x_{l,t} \sim AR(1)$ with autoregressive coefficient $\rho_l \sim \text{Beta}(p, 2 - 2d)$, then $\bar{x}_t \equiv \sum_{l=1}^N x_{l,t}$ is approximately integrated of order d . When $N = o(T)$, Lemma 1 implies that $\text{COV}(\psi'_k \bar{x}, \psi'_j \bar{x}) = o(1)$.

Although Lemma 1 covers a wide range of processes, there are some time-series processes for which the asymptotic orthogonalization does not completely hold. In the Appendix, we show that autoregressive processes with a coefficient satisfying $\rho_T = 1 - c/T$ for $c > 0$ (so-called local to unity autoregression) along with fractionally-differenced processes have some pairs, $\psi'_j x$ and $\psi'_k x$ for which the correlation does not converge to zero with the sample size (see Lemma A.1 in Appendix A.2). However, we also show that the residual correlation which remains after the transformation is applied is small – a maximum of around 20% for $\psi'_1 x$ and $\psi'_2 x$, and near zero for all other pairs. In Appendix A.3, we provide a simulation exercise demonstrating that, across a range of processes, the residual correlation is small even for modest sample sizes.

There are related trigonometric basis functions that we could consider using rather than Ψ . For example, the discrete cosine transform (DCT) used by Müller and Watson (2008) is intimately related to Ψ , except that each basis function is shifted to be mean zero. Consequently, the DCT does not result in an orthonormal basis, which is a property we require in order to ensure the invariance of the OLS estimator. Conversely, the discrete Fourier transform (DFT) does form an orthonormal basis, but it cannot asymptotically orthogonalize persistent processes like a random walk (Hannan, 1973; Lahiri, 2003; Dwivedi and Subba Rao, 2011). Moreover, the DFTs do not include basis functions with pronounced low-frequency behavior as Ψ does (see Figure 1). We illustrate the favorable performance of Ψ relative to the DFT in our simulation experiments in Appendix A.3.

In sum, the transformation Ψ has the appeal of decomposing the time series into (asymptotically) orthogonal contributions from different frequencies. We will exploit this property in the next section to introduce our new diagnostic.

3 A New Regression Diagnostic

3.1 Time Series Regressions

Suppose we are interested in the coefficients $\beta \in \mathbb{R}^k$ of

$$y_t = \alpha + x_t' \beta + \varepsilon_t, \quad t = 1, \dots, T. \quad (14)$$

Again, this model can be stacked as,

$$y = \alpha \iota_T + X\beta + \varepsilon, \quad (15)$$

where ι_T is a $T \times 1$ vector of ones. As before, we can premultiply equation (15) by the basis functions Ψ' to obtain,

$$w = \alpha \cdot \zeta + Z\beta + u, \quad (16)$$

where $Z = \Psi'X$ is now a $T \times k$ matrix. By equation (2), it is immediate that the OLS estimator is invariant to this transformation of the data. Define $\mathbf{X} = [\iota_T \ X]$ and $\mathbf{Z} = [\zeta \ Z]$. Then,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = (\mathbf{X}'\Psi\Psi'\mathbf{X})^{-1}\mathbf{X}'\Psi\Psi'y = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'w. \quad (17)$$

Based on the results of the previous section, rotating the explanatory and dependent variables through Ψ renders the resultant variables approximately uncorrelated. As such, we can remove the distinct contribution of a single “observation” j from our sample and compare this LOO estimate to the full-sample OLS estimate.³

Let $w_{(-j)}$ and $\mathbf{Z}_{(-j)}$ be w and \mathbf{Z} with the j th row (w_j and \mathbf{Z}_j , respectively) removed.⁴ Then,

$$\begin{pmatrix} \hat{\alpha}_{(-j)} \\ \hat{\beta}_{(-j)} \end{pmatrix} = \left(\mathbf{Z}_{(-j)}' \mathbf{Z}_{(-j)} \right)^{-1} \mathbf{Z}_{(-j)}' w_{(-j)} = (\mathbf{Z}'\mathbf{Z} - \mathbf{Z}_j\mathbf{Z}_j')^{-1} (\mathbf{Z}'w - \mathbf{Z}_jw_j'). \quad (18)$$

³Of course, we could also remove more than one observation, removing the influence from multiple frequencies.

⁴We can use the Sherman-Morrison-Woodbury formula to write these leave-one-out estimators in a more convenient form which allows for more efficient computation. We have that $(\hat{\alpha}, \hat{\beta})' - (\hat{\alpha}_{(-j)}, \hat{\beta}_{(-j)})' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}_j\hat{u}_j/(1 - p_{jj})$, where $p_{jj} = \mathbf{Z}_j'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}_j$.

To provide additional intuition for this object, note that we can equivalently obtain $(\hat{\alpha}_{(-j)}, \hat{\beta}'_{(-j)})$ by regressing y_t on a constant, x_t , and $\psi_{t,j}$, that is, the original regressors plus the additional control variable ψ_j .

Lemma 2. *Let $\tilde{\alpha}_{(-j)}$ and $\tilde{\beta}_{(-j)}$ be the solutions to the following:*

$$(\tilde{\alpha}_{(-j)}, \tilde{\beta}'_{(-j)}, \tilde{\delta}) = \arg \min_{\alpha, \beta, \delta} \sum_{t=1}^T |y_t - \alpha - x'_t \beta - \psi_{t,j} \delta|^2. \quad (19)$$

Then, $\tilde{\alpha}_{(-j)} = \hat{\alpha}_{(-j)}$ and $\tilde{\beta}_{(-j)} = \hat{\beta}_{(-j)}$.

Proof: See Appendix A.4.

Lemma 2 shows that, by the Frisch-Waugh-Lovell Theorem, we can obtain $(\hat{\alpha}_{(-j)}, \hat{\beta}'_{(-j)})$ by residualizing y and X with respect to the j th basis function, ψ_j , and use the resultant series to calculate the LOO OLS estimate. Lemma 2 holds immediately by the partition regression formula and because Ψ forms an orthonormal basis.

Using the LOO estimate from equation (18), we can define our new regression diagnostic as

$$\text{DFQBETA}_{\ell,j} = \frac{e'_\ell (\hat{\beta} - \hat{\beta}_{(-j)})}{\sqrt{e'_\ell \hat{V}_{\hat{\beta}} e_\ell}}, \quad (20)$$

where $\hat{V}_{\hat{\beta}}$ is an estimate of the variance-covariance matrix of $\hat{\beta}$ and e_ℓ is a selection vector for the element $\ell \in \{1, \dots, k\}$.⁵ In words, our diagnostic reflects the importance of a single frequency to the estimated coefficient. We scale the difference in the numerator by the estimated standard deviation of the coefficient estimate to provide a sense of how large this difference is in terms of the estimator's variability. The diagnostic is similar in spirit to the conventional regression “difference in beta” (so-called DFBETA which is implemented in Stata as `dfbeta`; see Belsley et al., 1980, for details) except that we assess the influence from the frequency associated with the j th basis function rather than the influence from the j th observation.

We should note that there may be alternative choices of scaling $(\hat{\beta} - \hat{\beta}_{(-j)})$ depending on the specific application. For example, there may be measures of distance reflecting economic significance that would be preferred to statistical measures. As an example, in Section 5, we introduce an

⁵In practice, we may be interested in general linear combinations of the coefficients such as $a'\beta$. This case follows analogously.

application where $\hat{\beta}_{(-j)}$ for a specific choice of j , violates the underlying economic theory, suggesting that the empirical specification is particularly fragile as, without information from that frequency, the model would be rejected out of hand.

Below, we consider some natural extensions to this diagnostic beyond the time-series regression model.

3.2 Extension to Panel Data Models

Consider the following balanced panel data model,

$$y_{it} = \alpha_i + x'_{it}\beta + \epsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T,$$

where α_i is a unit-specific constant (fixed effect) and $x_{it} \in \mathbb{R}^k$ are a vector of covariates.⁶ We can stack the model as,

$$Y = \left(y_{1,1}, \dots, y_{1,T}, y_{2,1}, \dots, y_{2,T}, \dots, y_{n,1}, \dots, y_{n,T} \right)' = (\alpha \otimes \iota_T) + X\beta + \epsilon.$$

In order to apply our transformation to this model, we define an altered set of basis functions

$$\Psi_{NT} = (\mathbf{I}_N \otimes \Psi) = \begin{bmatrix} \Psi & & \\ & \ddots & \\ & & \Psi \end{bmatrix}.$$

As the basis remains orthonormal since $\Psi'_{NT}\Psi_{NT} = \mathbf{I}_{NT}$, the model and (full-sample) OLS estimate remain invariant to this transformation. Then, we can apply this transformation to obtain

$$\Psi'_{NT}Y = (\alpha \otimes \zeta_T) + \Psi'_{NT}X\beta + \Psi'_{NT}\epsilon,$$

where, for example,

$$\Psi'_{NT}Y = \left(\Psi'(y_{1,1}, \dots, y_{1,T})', \Psi'(y_{2,1}, \dots, y_{2,T})', \dots, \Psi'(y_{n,1}, \dots, y_{n,T})' \right)'. \quad (21)$$

⁶In practice, we always recommend including unit-specific constants when utilizing the diagnostic.

Equation (21) shows that applying the transformation Ψ'_{NT} is equivalent to applying Ψ' to the time series for each individual i . To calculate the diagnostic in the panel data case, we remove the j th observation from each $\Psi'(y_{i,1}, \dots, y_{i,T})'$ for $i = 1, \dots, N$. Thus, we are removing N observations reflecting the contribution from the frequency associated with the j th basis function ψ_j across all NT observations; thus, $\hat{\beta}_{(-j)}$ is based on $N(T-1)$ observations. We can then compute our influence measure $\text{DFQBETA}_{\ell,j}$ as before.

Intuitively, applying this transformation has the effect of orthogonalizing the data (Y, X) across time periods, within each unit. Consequently, we can think of omitting identifying variation at certain frequencies jointly for all units as means of assessing the fragility of our estimate.⁷

In the case of balanced panels a generalization of Lemma 2 holds and the LOO estimate, $\hat{\beta}_{(-j)}$, retains the property of being equivalent to the OLS estimate, obtained from a model with ψ_j as an additional control for each individual; i.e.,

$$(\hat{\alpha}_{(-j)}, \hat{\beta}'_{(-j)}, \hat{\gamma}_1, \dots, \hat{\gamma}_N) = \arg \min_{\alpha, \beta, \gamma_1, \dots, \gamma_N} \sum_{i=1}^N |y_i - \alpha_i \iota_T - X_i' \beta - \psi_j \gamma_i|^2,$$

where $y_i = (y_{i,1}, \dots, y_{i,T})'$ and similarly for X_i .⁸

In practice, panels may be unbalanced, that is we may not observe all time periods for all units. In order to handle this, we define a selection matrix $S \in \mathbb{R}^{\bar{N} \times (n \cdot T)}$, where \bar{N} is the number of observations in the unbalanced panel.⁹ We then redefine our basis as

$$\tilde{\Psi}_{NT} = S \Psi_{NT},$$

such that $\tilde{\Psi}_{NT} \in \mathbb{R}^{\bar{N} \times (n \cdot T)}$. Using these basis functions, we can then compute our influence measure $\text{DFQBETA}_{\ell,j}$ as before.

3.3 Extension to Instrumental Variable Models

We can also naturally extend our influence measure to instrumental variable (IV) models. Specifically, consider the case where we have a time series IV model (panel IV models are accommodated

⁷The LOO estimate here is $\hat{\beta}_{(-j)} = (X'X)^{-1}Z'_j(I_N - P_j)^{-1}\hat{u}_j$, where $P_j = Z_j(X'X)^{-1}Z'_j$.

⁸When the panel is unbalanced, this equivalence breaks down, but in most applications the difference between $\hat{\beta}_{(-j)}$ and the estimate obtained from the control approach is small.

⁹A selection matrix resembles an identity matrix with ones on the indices that are observed. Trivially, when $\bar{N} = n \cdot T$, $S = I_{nT}$.

with slight modifications, as shown above). This can be expressed as

$$y_t = \alpha + x_t' \beta + \epsilon_t$$

$$x_t = g(r_t) + v_t,$$

where r_t – the instrument vector – satisfies the standard relevance and exogeneity conditions. Let X and R be the stacked matrices collecting x_t and r_t , respectively. This model is typically estimated via two-stage least squares (2SLS); i.e.,

$$\hat{\beta}_{2SLS} = (\hat{X}' \hat{X})^{-1} \hat{X}' y,$$

where $\hat{X} = \hat{\Gamma} R$ and $\hat{\Gamma}$ collects the sample estimates of the best linear predictor of x_t given $(1, r_t)'$. We can then define $\mathbf{Z} = \Psi' \hat{X}$ and $w = \Psi' y$ and proceed as in Section 3.1. As is the case for OLS, the 2SLS estimator, $\hat{\beta}_{2SLS}$, is also invariant to the transformation using Ψ .¹⁰

3.4 Extension to Spatial Data

We can also consider forms of dependence in the data beyond the simple time series case. In fact, it is straightforward to modify our procedure to spatial series $X \in \mathbb{R}^{N^2}$ that follows a $d = 2$ Levy-Brownian motion (as in Müller and Watson, 2024b). Here, N is the cardinality of the sets $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^N$. Correspondingly, the index set $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$ has cardinality N^2 , where $\mathcal{C}(i)$ defines the location or coordinate where X_i is observed for all $i \in \{1, \dots, N^2\}$.

Given that X follows a Levy-Brownian motion, we can describe its covariance matrix $\Sigma \in \mathbb{R}^{N^2 \times N^2}$ as

$$\Sigma_{i,j} = \frac{1}{2}(|\mathcal{C}(i)| + |\mathcal{C}(j)| - |\mathcal{C}(i) - \mathcal{C}(j)|),$$

where $|x| = \sqrt{x'x}$. We can find the matrix Ψ such that

$$\Psi' \Sigma \Psi = \Lambda,$$

¹⁰While, in principle, we can consider how the coefficients of the first stage, $\hat{\Gamma}$, are affected by the omission of certain sources of variation, these coefficients are usually not of independent interest and so we do not pursue this alternative further.

where Λ is diagonal by performing a numerical eigendecomposition as the solution is not available in closed form.

More generally, we may have that $N_x \neq N_y$, or that the cardinality of the set \mathcal{X} is not equal to the cardinality of the set \mathcal{Y} (i.e., when we observe data for a rectangular gridded region). In this case, the above goes through with minimal alteration.

Even more generally, we may have that $\mathcal{C} \neq \mathcal{X} \times \mathcal{Y}$ (i.e., when we observe data that is irregularly sampled over a grid). To deal with this, we can define a set $\tilde{\mathcal{C}} \equiv \mathcal{X} \times \mathcal{Y}$ and a selection matrix $S \in \mathbb{R}^{(N_x \cdot N_y) \times \bar{N}}$, where \bar{N} is the cardinality of \mathcal{C} . Then, we can construct a matrix $\tilde{\Sigma} \in \mathbb{R}^{(N_x \cdot N_y) \times (N_x \cdot N_y)}$, perform an eigendecomposition on this matrix, and recover $\Psi = S' \tilde{\Psi}$, which provides a linear transformation that can be used to orthogonalize X .

Now, consider the spatial regression model

$$y = \alpha \iota_{\bar{N}} + X\beta + \epsilon,$$

where $y, \epsilon \in \mathbb{R}^{\bar{N}}$ and $X \in \mathbb{R}^{\bar{N} \times k}$. As before, we can apply the linear transformation $\tilde{\Psi}$ to this model and compute $\text{DFQBETA}_{\ell,j}$ for $\ell \in \{1, \dots, k\}$ and $j \in \{1, \dots, (N_x \cdot N_y)\}$.

4 Conservative Bounds for the Diagnostic

To make any diagnostic operational, some guidance on what constitutes an “unusual” value of the measure must be provided. In this section, we will motivate a “rule-of-thumb” value for DFQBETA which, when exceeded, invites further scrutiny by the researcher.

First, consider the case where (y_t, x'_t) are strictly stationary. Under regularity conditions, it can be shown that the numerator of DFQBETA , $\left(\hat{\beta} - \hat{\beta}_{(-j)}\right)$, is $O_p(T^{-1})$. Furthermore, by standard properties of OLS, the denominator is of order $O_p(T^{-1/2})$. It follows that DFQBETA is $O_p(T^{-1/2})$, which suggests that large values should become increasingly rare as T grows. Under cointegration (i.e., when x_t follows a multivariate random walk, and ϵ_t is white noise), it can be shown that both the numerator and the denominator are $O_p(T^{-1})$, so that DFQBETA is $O_p(1)$. This suggests that the degree of persistence of the data plays a key role in establishing the behavior of the DFQBETA . Moreover, it also suggests that the asymptotic results will be model dependent. In the spirit of existing regression diagnostics, we will take a conservative approach to constructing our

recommended bounds, focusing on worst-case scenarios.

Since prediction and forecasting is a primary motivation of empirical macroeconomics and finance, we will focus on the simple workhorse triangular model of the form:

$$y_t = \alpha + x_t\beta + \varepsilon_t \quad (22)$$

$$x_t = \mu + \rho x_{t-1} + \nu_t. \quad (23)$$

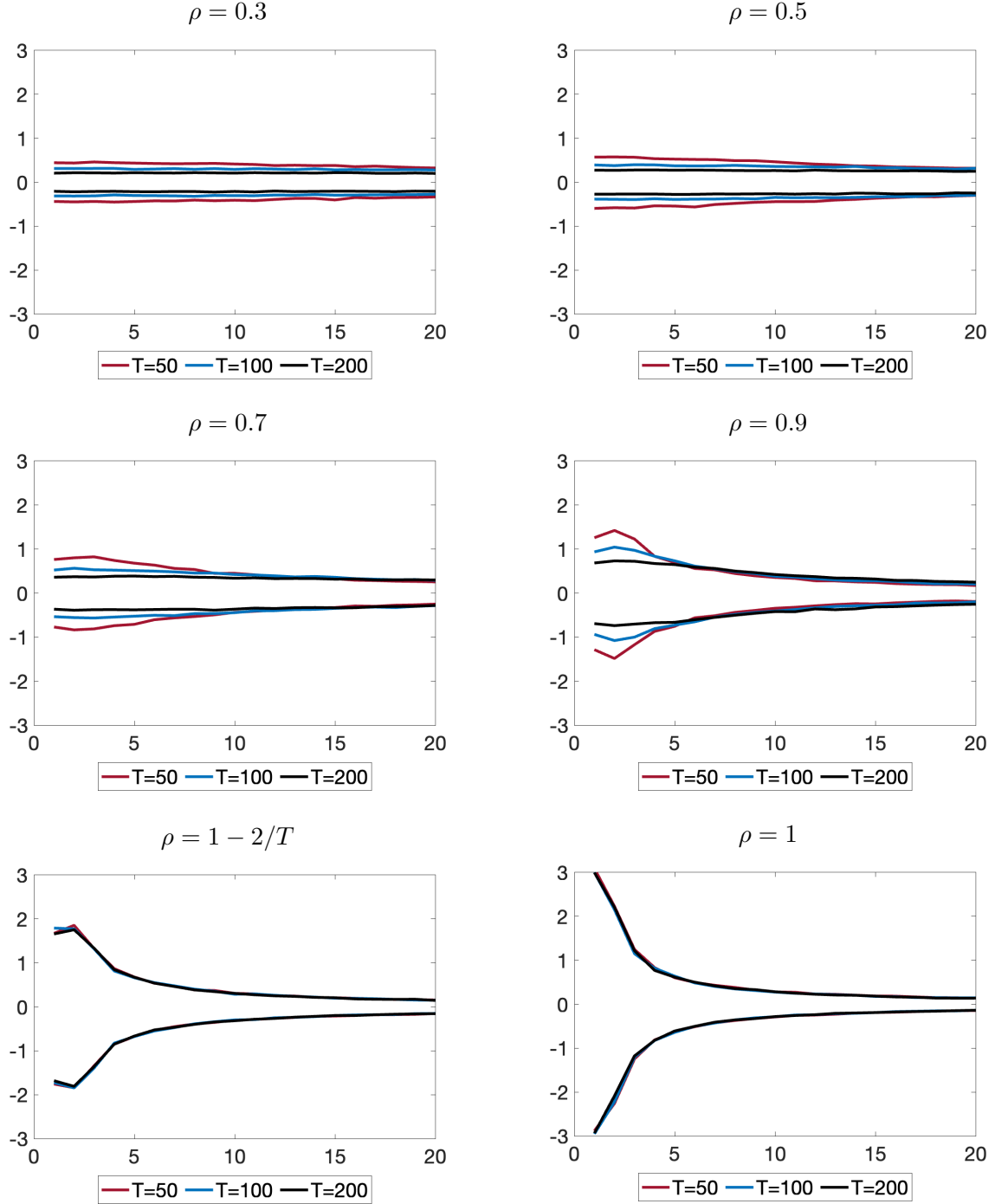
The two key parameters are β and ρ and for different choices of these parameters and the error processes, ε_t and ν_t , this setup nests the standard cointegration model, the spurious regression model, and the classical predictive regression model.

We will use this model to guide the choice of our recommended bounds. As most economic time series are dominated by their low frequency components, we will focus on the first twenty values of the DFQBETA. As we will see in the next section, this appears sufficient for economic applications.

Figure 3 presents the pointwise 2.5th and 97.5th quantiles of each of the first twenty DFQBETA values for different values of ρ . We vary the sample size T to be either 50, 100 or 200. The top row of the figure presents the results for $\rho = 0.3$ and $\rho = 0.5$. In these cases, where the dependence is weak, we can observe that the pointwise quantiles are approximately flat across the different frequencies. Furthermore, the values are comfortably below 0.5. When the degree of persistence is moderate to strong, with values of $\rho = 0.7$ and $\rho = 0.9$, then we can observe that the envelope begins to widen at the lower frequencies, especially for smaller sample sizes (see middle row of Figure 3). Finally, when the degree of persistence becomes extreme, then this widening becomes more pronounced at the lowest frequencies, peaking for the first DFQBETA at around 3 when $\rho = 1$ (see bottom row of Figure 3). We can also observe that the results are no longer dependent on the sample size for either the local-to-unity case or when $\rho = 1$.

Figure 3 shows that the case of $\rho = 1$ generates the widest range of values for the lowest frequency basis functions whereas for smaller values of ρ , the behavior is more uniform across basis functions. Our proposed bounds are informed by these observations. In Figure 4, we present the results for $T = 200$ but now grouped by the values of ρ and set against our recommended thresholds (see Appendix A.1 for the specific values). The left chart presents the results for $\rho \in \{0.3, 0.5, 0.7\}$ which we label “weak and moderate” persistence whereas the right chart presents the results for $\rho \in \{0.9, 0.99, 1\}$ which we label “strong and extreme” persistence. We also present our recommended

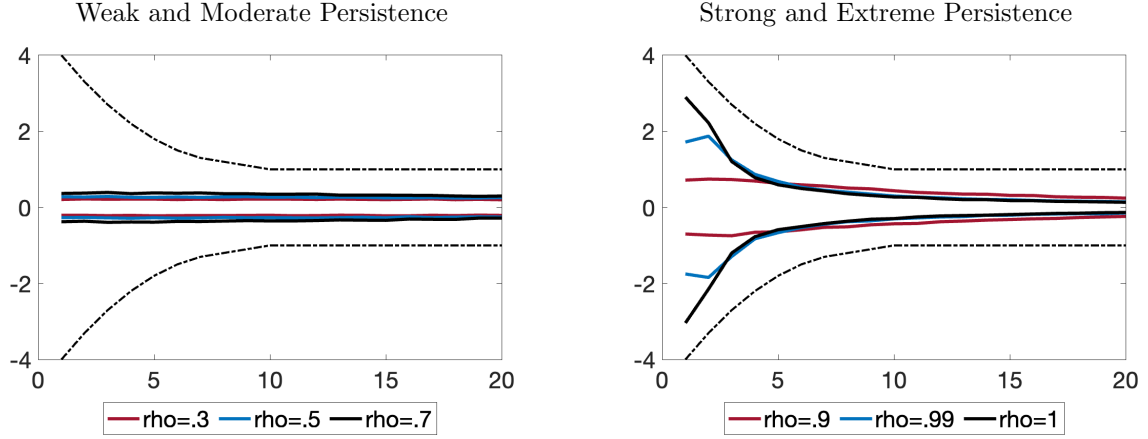
Figure 3. Marginal Quantiles of the Regression Diagnostic. This figure presents the 97.5th and 2.5th pointwise quantiles of the DFQBETA using OLS standard errors for the first twenty basis functions. The data generating process is described by equations (22)–(23) with independent standard Gaussian white noise errors and $\beta = 1$. Results are based on 10,000 simulations.



guidance bounds for which observations deserve further scrutiny from the researcher.

Comparing the bounds to the pointwise quantiles in the weak and moderate persistence case emphasizes the conservativeness of our approach as the bounds are substantially wider across all of

Figure 4. Marginal Quantiles of the Regression Diagnostic. This figure presents the 97.5th and 2.5th pointwise quantiles of the DFQBETA using OLS standard errors for the first twenty basis functions. The data generating process is described by equations (22)–(23) with independent standard Gaussian white noise errors and $\beta = 1$. The dashed lines denote our recommended thresholds which are given in Appendix A.1. Results are based on 10,000 simulations.



the twenty basis functions. But we also want to be robust to higher persistence and so in the right chart of Figure 4 we compare the bounds to the $\rho = 1$ case. The bounds are still sufficiently wide to comfortably envelop the pointwise quantiles. In fact, when the linear model holds and $T = 200$, the cutoffs are exceeded at least once in less than 5% of simulations when $\rho = 1$, 2.5% of simulations when $\rho = 0.99$ and less than 0.4% when $\rho = 0.7$.¹¹ In contrast, in the classical spurious regression setting, using the Newey-West estimator of Lazarus et al. (2018), the cutoffs are exceeded at least once 38% of the time for a sample of size $T = 200$.

Given the disparate behavior of DFQBETA across different degrees of persistence, one might ask why we do not recommend different bounds for “stationary” data. In practice, we want to remain agnostic about the true degree of dependence (since it is unknown) and so we provide bounds that can be utilized in all applications without pre-testing for unit roots or other similar assessments. We will see in the next section, that despite our conservative approach, the bounds can still be informative for studying the stability of the OLS estimator in common applications.

5 Empirical Examples

In this section, we demonstrate the usefulness of our new regression diagnostic across a number of different empirical applications from recent papers in the Economics literature. In order to

¹¹The cutoffs are exceeded more than once only (4,2,0) times out of 10,000 simulations for $\rho = (1, 0.99, 0.7)$, respectively.

operationalize the DFQBETA, we use the standard error associated with the regression coefficient of interest provided in each paper. We take this choice as given as it provides the appropriate magnitude to assess the LOO estimates in light of the original OLS results.

Philippon and Reshef (2012)

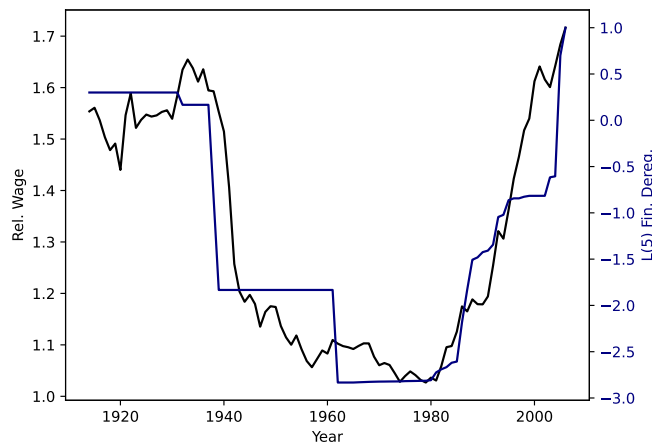
Philippon and Reshef (2012) study the relation between deregulation in the financial services industry and the increases in wages for workers in that industry. To test for this proposed relationship, the authors estimate the following time series model on US data

$$y_t^{\text{fin}} = \alpha + X_{t-5}\beta + \epsilon_t, \quad (24)$$

where y_t^{fin} is a measure of wages in the financial services industry relative to the average private wage, and X_t is an index of legislated deregulatory actions. The parameter of interest here is β , which describes the elasticity of relative wages to (lagged) deregulatory actions. Their reported value of $\hat{\beta}$ is 0.17 with a standard error estimate of 0.0096, obtained as Newey-West standard errors with a choice of five lags. Consequently, they estimate a highly statistically significant positive relationship, with a p -value of less than 1%.

What are the important features of the data that give rise to this tight positive relationship? As this specification represents a univariate time series regression, we can easily visualize the data to gain some intuition for the properties of the estimate. As can be seen in Figure 5 (and discussed in

Figure 5. Relative Wages and Financial Deregulation. This figure presents the dependent and independent variables from the regression described by equation (24) from Philippon and Reshef (2012).

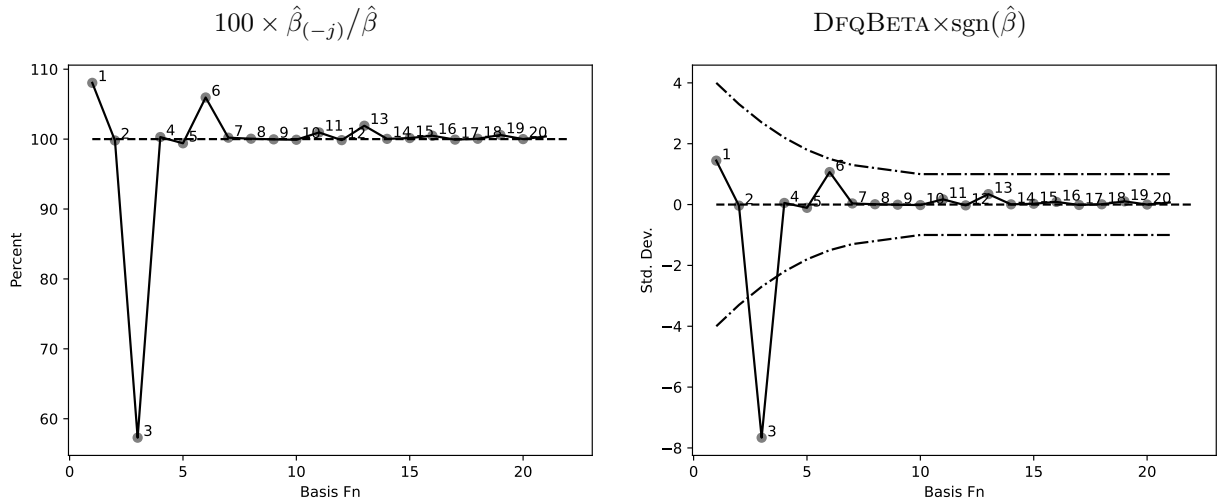


the original paper), the dependent and explanatory variables exhibit a common “U”-shape. While

this “U”-shape may well be an artifact of the true joint data-generating process, researchers are typically skeptical of regressions estimated on data with strong, seemingly deterministic trend-like behavior, as spurious correlations could arise. Consequently, we might want to understand if the results of this regression are in large part driven by this common trend-like behavior, or if in fact, the relationship holds even upon controlling for the trend.

Calculating the DFQBETA allows one to control for a variety of flexible periodic functions in a principled manner (see Lemma 2). We should expect to see movement in the 3rd basis function – which corresponds to the “U”-shape – if the estimate of $\hat{\beta}$ is in large part informed by variation at that frequency.

Figure 6. LOO Estimates. This figure presents the LOO estimates of β (see equation (24)) using the main application in Philippon and Reshef (2012). The left chart presents the ratio of the LOO estimate of β for the first twenty basis functions relative to $\hat{\beta}$ (in percentage terms). The right chart presents the the LOO estimate for the first twenty basis functions relative to the reported standard error, multiplied by the sign of the original OLS estimate. Dashed lines denote the recommended bounds.

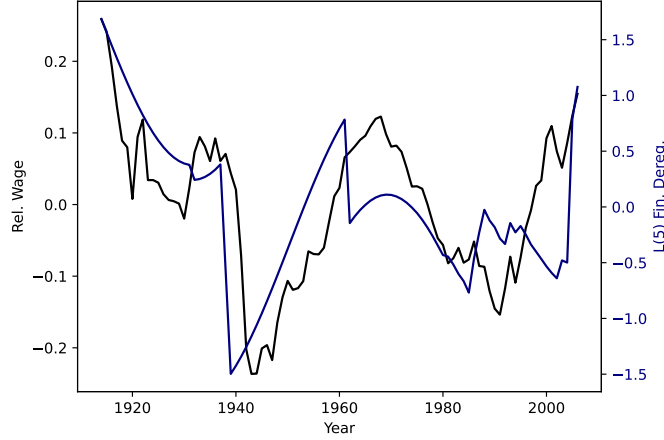


Looking at the plots generated by our diagnostic, it is clear that, in fact, the estimate of β reported in the paper is driven by the deterministic “U”-shape. If we were to control for this basis function, $\hat{\beta}$ would be 60% of the original estimate, which represents an ≈ 8 standard deviation movement (taking the original standard errors as given). This easily exceeds our recommended thresholds, which suggest that greater care should be taken in interpreting the estimate that comes out of the original specification.

That said, the results of applying the diagnostic, are, on the whole, constructive. While the estimated relationship based on the specification which includes a “U”-shaped trend is not as large

as the original, it still remains positive and statistically significant at the 1% level. We can assess the robustness of this relationship to the inclusion of the trend by plotting the residualized time series (with respect to a constant and ψ_3) of the independent and dependent variable, as we do in Figure 7.¹² We observe that even with the “U” shape removed, there is a clear commonality in the behavior of the two series.

Figure 7. Residualized Relative Wage and Financial Deregulation. This figure presents the residualized (with respect to ψ_3) dependent and independent variables from the regression utilized in Philippon and Reshef (2012) (see equation (24)).



This example shows how our diagnostic can be used to gauge the influence of common trending behavior on the empirical result.

Farber, Herbst, Kuziemko, and Naidu (2021)

We next revisit the work by Farber et al. (2021) which investigates the relation between income inequality and the decreasing share of workers that are members of a union (declining union density). They estimate the following time series regression for the United States

$$\text{LaborShare}_t = \alpha + \beta \text{UnionDensity}_t + X_t \gamma + \phi_1 t + \phi_2 t^2 + \phi_3 t^3 + \epsilon_t, \quad (25)$$

where LaborShare_t is the labor share of income, UnionDensity_t is the share of workers in a union, and X_t is a vector of controls. The last three regressors in equation (25) allow for a cubic polynomial in time, likely added to account for the trending behavior observed in some variables such as union density. The parameter of interest is β , or the elasticity of the labor share of income with respect

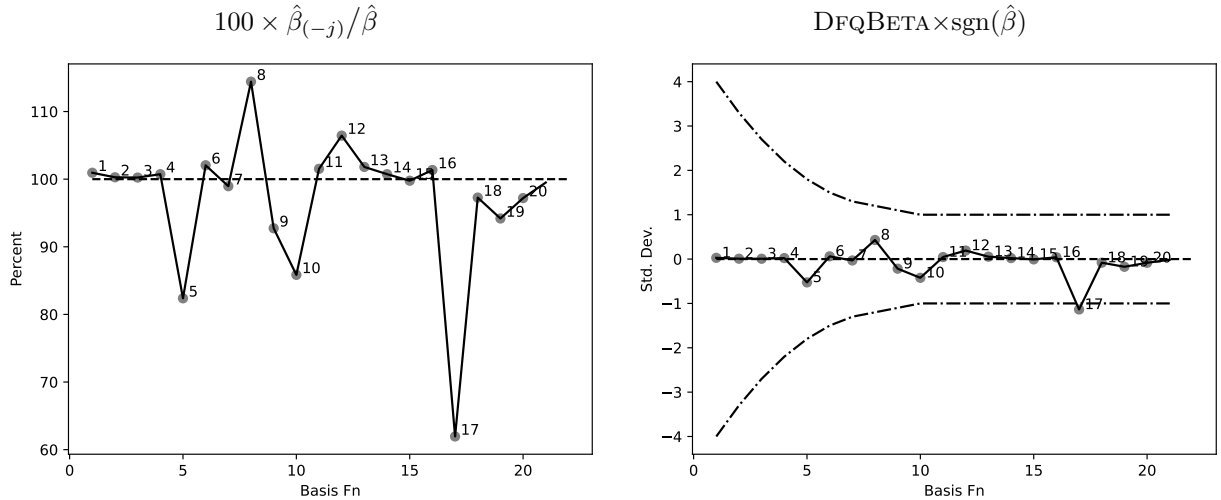
¹²A regression of these two series on each other recovers the estimate presented in DFQBETA_{1,3} by Lemma 2 and the Frisch-Waugh-Lovell theorem.

to union density. Farber et al. (2021) report a $\hat{\beta}$ of 39.43 with a standard error estimate of 13.21, robust to heteroskedasticity and serial correlation.

Figure 8 presents the results from our diagnostic to this time-series regression. We first note that the inclusion of up to cubic polynomials in time has eliminated most of the variation in the four lowest frequencies. This makes intuitive sense, as linear combinations of the first four basis functions can be made to resemble a cubic polynomial. However, this does not preclude higher frequency behavior from playing a key role in driving the value of the OLS estimate. In fact, Figure 8 shows that the sensitivity to the 17th basis function is high, with the corresponding LOO estimate about 60% of the initial OLS estimate. The right chart of Figure 8 confirms the unusual sensitivity as the DFQBETA is 1.1 which exceeds the corresponding bound. Taking the standard errors as given, this would imply that the estimate goes from being significant at the 1% level, to marginally significant at the 10% level.

How can we interpret the 17th basis function? Recall that we can obtain the periodicity of the basis function as $\frac{2(2T+1)}{2j-1}$. Here, we have $j = 17$ and $T = 75$ which corresponds to a periodicity of 9.1 years. This is approximately the average length of the business cycle in the post-war US period. Said differently, the OLS estimator appears to be disproportionately affected by information at the business-cycle frequency in this application.

Figure 8. LOO Estimates. This figure presents the LOO estimates of β (see equation (25)) using the main application in Farber et al. (2021). The left chart presents the ratio of the LOO estimate of β for the first twenty basis functions relative to $\hat{\beta}$ (in percentage terms). The right chart presents the the LOO estimate for the first twenty basis functions relative to the reported standard error, multiplied by the sign of the original OLS estimate. Dashed lines denote the recommended bounds.



This example illustrates how our diagnostic is more general and more principled than the

conventional use of time polynomials that are routinely used to control for trending behavior. Further, the example shows that higher frequency behavior may also play an influential role in OLS estimates.

Barnichon and Mesters (2020)

Our next empirical illustration is based on [Barnichon and Mesters \(2020\)](#). In this paper, the authors estimate the slope of the Phillips Curve. To do so, they employ the following standard New Keynesian Phillips Curve, estimated using quarterly US data

$$\pi_t = \gamma_b \pi_{t-1}^4 + \gamma_f \mathbb{E}[\pi_{t+4}^4] + \beta u_t^c + \epsilon_t^s, \quad (26)$$

where π_t is annualized quarter-to-quarter inflation, $\pi_{t-1}^4 = \frac{1}{4}(\pi_{t-1} + \pi_{t-2} + \pi_{t-3} + \pi_{t-4})$ is a four-quarter moving average of inflation, and u_t^c is the detrended unemployment rate.¹³ They instrument this equation with $g(\epsilon_{t:t-20}^m)$, a vector of nonlinear transformations of the [Romer and Romer \(2004\)](#) monetary policy shocks, and estimate the model via 2SLS.¹⁴ The coefficient of interest here is β , which is the slope of the Phillips curve, and, according to theory, should be negative. The authors report a value of $\hat{\beta}$ of -0.42 , with a 95% confidence band of $[-1.61, -.05]$, significantly different from zero.

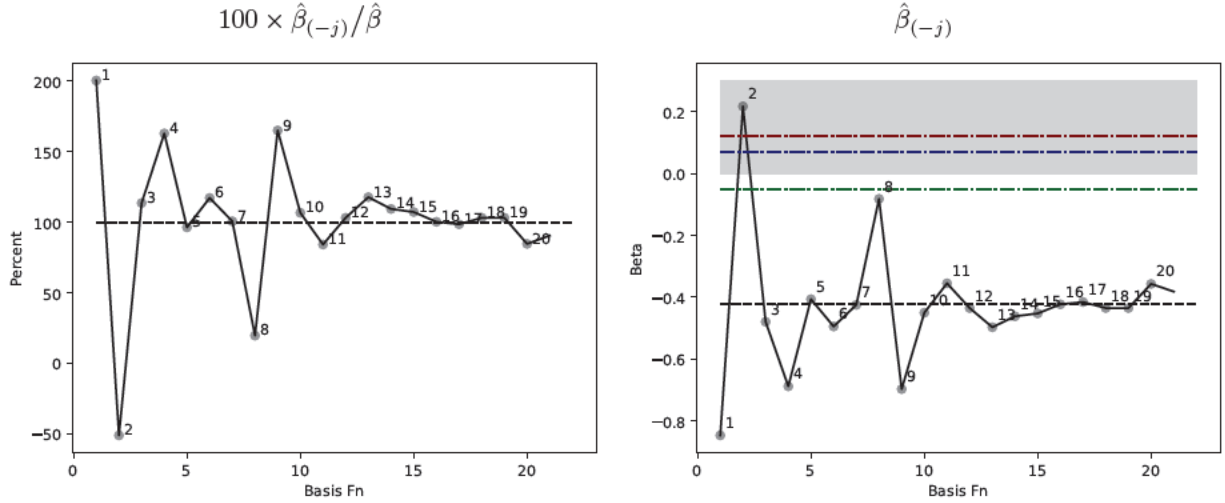
As discussed in Section 3.3, to apply our diagnostic to an instrumental variables setting, we take \hat{X} as given, and assess the sensitivity of the IV estimator of β to information from different frequencies of y and \hat{X} . Figure 9 presents the LOO estimates of β for this application. The left chart shows that the LOO estimates range from approximately double the original point estimate to half the size with the opposite sign. In fact, when the second basis function is removed, the point estimate shifts from a value of -0.42 to *positive* 0.22 ; importantly, this value directly contradicts the sign restriction implied by standard macroeconomic theory.

In the right chart, we present the LOO estimates directly. Since [Barnichon and Mesters \(2020\)](#) rely on non-standard asymptotic theory (based on an inversion of a test statistic), there is not a standard error available as would commonly be the case. Instead, the chart presents the upper

¹³In the baseline specification, [Barnichon and Mesters \(2020\)](#) use an HP-filter with parameter equal to 1600 to detrend the unemployment rate.

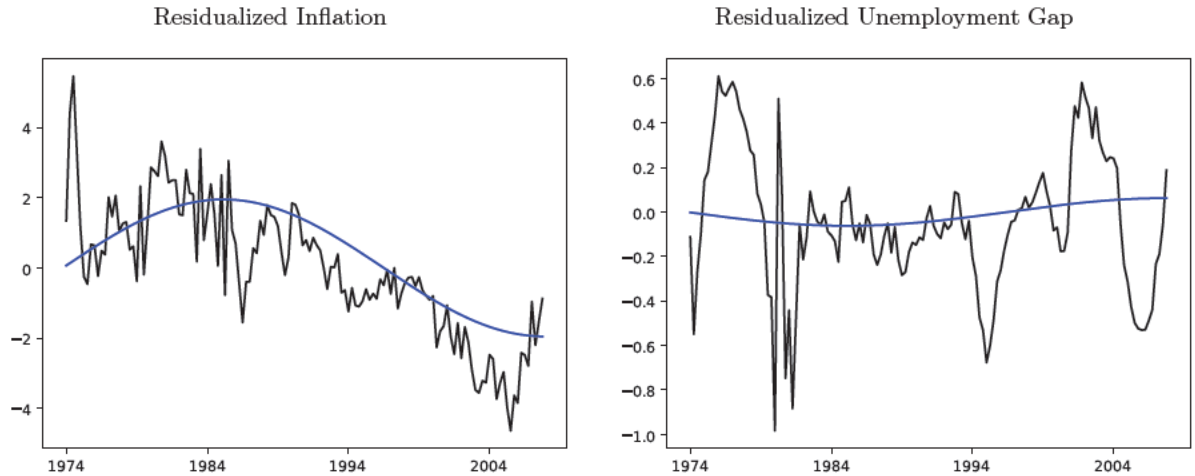
¹⁴[Barnichon and Mesters \(2020\)](#) employ the Almon parametrization which has $g(\epsilon_{t:t-20}^m) \equiv [\sum_{l=0}^{20} \epsilon_{t-l}^m, \sum_{l=0}^{20} l \epsilon_{t-l}^m, \sum_{l=0}^{20} l^2 \epsilon_{t-l}^m]$.

Figure 9. LOO Estimates. This figure presents the LOO estimates of β (see equation (26)) using the main application in [Barnichon and Mesters \(2020\)](#). The left chart presents the ratio of the LOO estimate of β for the first twenty basis functions relative to $\hat{\beta}$ (in percentage terms). The right chart presents the LOO estimate for the first twenty basis functions. The grey shaded region denotes positive values and the green, blue, and red horizontal lines represent the upper bound of the 95%, 99%, and 99.5% confidence intervals for β .



bound of their confidence intervals for nominal coverage rates of 95%, 99%, and 99.5% (dashed horizontal lines). We can observe that 0.22 comfortably exceeds the upper bound of a 99.5% confidence interval.

Figure 10. Residualized Inflation and Instrumented Unemployment Gap. This figure presents the residualized dependent (left chart) and independent variable (right chart) from the IV regression utilized in [Barnichon and Mesters \(2020\)](#). These series are plotted as black line. The blue line denotes the fitted value from a regression of either the residualized dependent or independent variable on the second basis function, ψ_2 .



The results of our diagnostic checks show that the frequency associated with the second basis function is the only source of information, making the estimate of β negative. To gain intuition for why this is the case, Figure 10 presents the residualized LHS and RHS variables from the

2SLS regression (left chart and right chart, respectively). Specifically, we residualize quarterly inflation and the instrumented detrended unemployment rate with respect to the other instrumented variables. By the Frisch-Waugh-Lovell Theorem, this residualization leaves the estimated coefficient unchanged at -0.42 . In Figure 10, we also show the second basis function, ψ_2 , scaled by the loading on each variable. The left chart shows that the second basis function represents the downward trend in inflation which occurred from the 1970s to the end of the sample. The right chart shows that the instrumented unemployment gap loads negatively on the second basis function. The positive loading on the second basis function in the left chart and the negative loading in the right chart is the source of the negative sign of the 2SLS estimate. Once removed, we obtain an estimated positive β using the information from all the other frequencies.¹⁵

This example shows how our diagnostic can highlight cases where a single frequency is the main driver of the empirical result.

Nakamura and Steinsson (2014) and Hazell, Herreño, Nakamura, and Steinsson (2022)

Last, we compare the results from the diagnostic in two related settings where the researchers use IV panel regressions relying on data from individual states. We will show that the researcher draws very different conclusions on the fragility of the estimates in these two different applications.

In the first paper, [Nakamura and Steinsson \(2014\)](#) use a panel of U.S. states to estimate the following model:

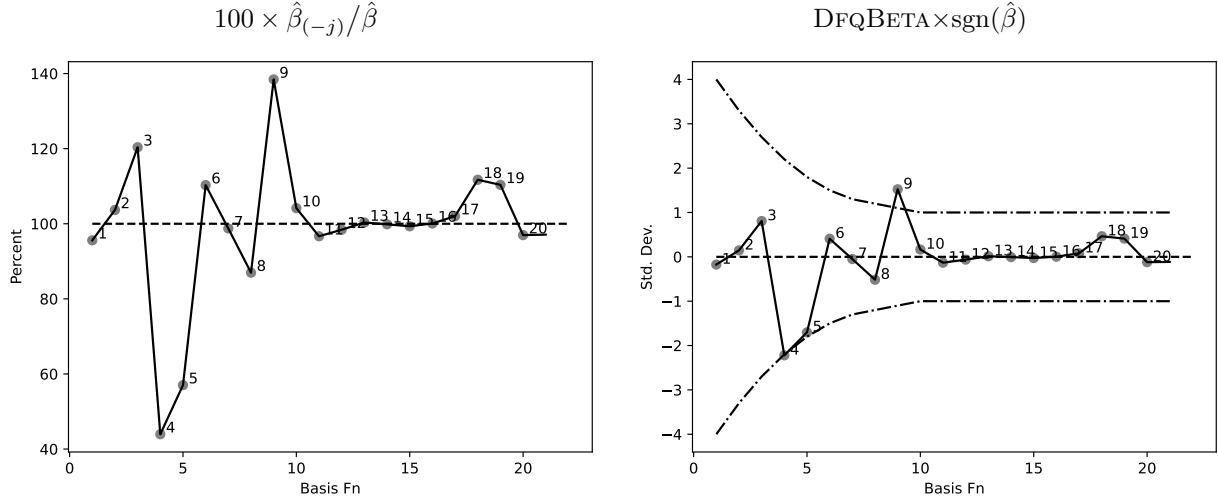
$$\frac{Y_{it} - Y_{it-2}}{Y_{it-2}} = \alpha_i + \gamma_t + \beta \frac{G_{it} - G_{it-2}}{Y_{it-2}} + \epsilon_{it}, \quad (27)$$

where Y_{it} is per capita output by state and G_{it} is per capita military procurement by state. The coefficient of interest is β which represents the fiscal multiplier. [Nakamura and Steinsson \(2014\)](#) instrument this specification with state dummies interacted with national changes in military procurement (i.e., $\frac{G_t - G_{t-2}}{Y_{t-2}}$). As before, we treat the first stage of the IV estimator as fixed and apply the diagnostic to the second-stage regression. [Nakamura and Steinsson \(2014\)](#) report an IV estimate of β of 1.43 with associated standard error of 0.36.

Figure 11 presents the leave-out estimates and diagnostic results for this application. Unlike in the previous examples, we observe that multiple frequencies are associated with a value of

¹⁵In fact, if we drop the first year of data, the estimated β moves to 0.5.

Figure 11. Leave-Out Estimates. This figure presents the leave-out estimates of β (see equation (27)) using the main application in Nakamura and Steinsson (2014). The left chart presents the ratio of the leave-out estimate of β for the first twenty basis functions relative to $\hat{\beta}$ (in percentage terms). The right chart presents the the leave-out estimate for the first twenty basis functions relative to the reported standard error, multiplied by the sign of the original OLS estimate. Dashed lines denote the recommended bounds.



DFQBETA that is near or exceeding the bound. For example, after removing the 4th basis function, the parameter of interest goes to 44% of the original estimate, a value of 0.63, corresponding to a 2.2 standard deviation change and resulting in a fiscal multiplier estimate below one. A similar result is obtained by removing the 5th basis function. In contrast, when the 9th basis function is removed, the fiscal multiplier estimate jumps to 1.97 and the associated DFQBETA exceeds the recommended bound. The OLS estimate of the fiscal multiplier appears fragile as there are multiple large DFQBETA values.

We can reinforce the conclusions in the previous application by comparing them to a paper with a similar empirical strategy. Hazell et al. (2022), using quarterly U.S. state data, estimate a Phillips Curve via the following specification:

$$\pi_{it}^N = \alpha_i + \gamma_t - \beta \sum_{j=0}^{20} \rho^j u_{i,t+j} - \lambda \sum_{j=0}^{20} \rho^j \hat{p}_{i,t+j}^N + \epsilon_{it}^N, \quad (28)$$

where π_{it}^N is the state-level inflation rate for non-tradeable goods, ρ is a discounting parameter (set to 0.99), u_{it} is the state-level unemployment rate, and \hat{p}_{it}^N are relative prices of non-tradeable goods. The parameter of interest here is β , the slope of the Phillips Curve. They estimate the model via two-sample 2SLS, instrumenting the first stage with four-quarter lagged unemployment $u_{i,t-4}$ and

four-quarter lagged relative prices of non-tradeable goods $\hat{p}_{i,t-4}^N$. In their preferred specification, they obtain a point estimate of $\hat{\beta}$ of 0.0062 and a standard error estimate of 0.0025.¹⁶

Figure 12. Leave-Out Estimates. This figure presents the leave-out estimates of β (see equation (28)) using the main application in [Hazell et al. \(2022\)](#). The left chart presents the ratio of the leave-out estimate of β for the first twenty basis functions relative to $\hat{\beta}$ (in percentage terms). The right chart presents the the leave-out estimate for the first twenty basis functions relative to the reported standard error, multiplied by the sign of the original OLS estimate. Dashed lines denote the recommended bounds.

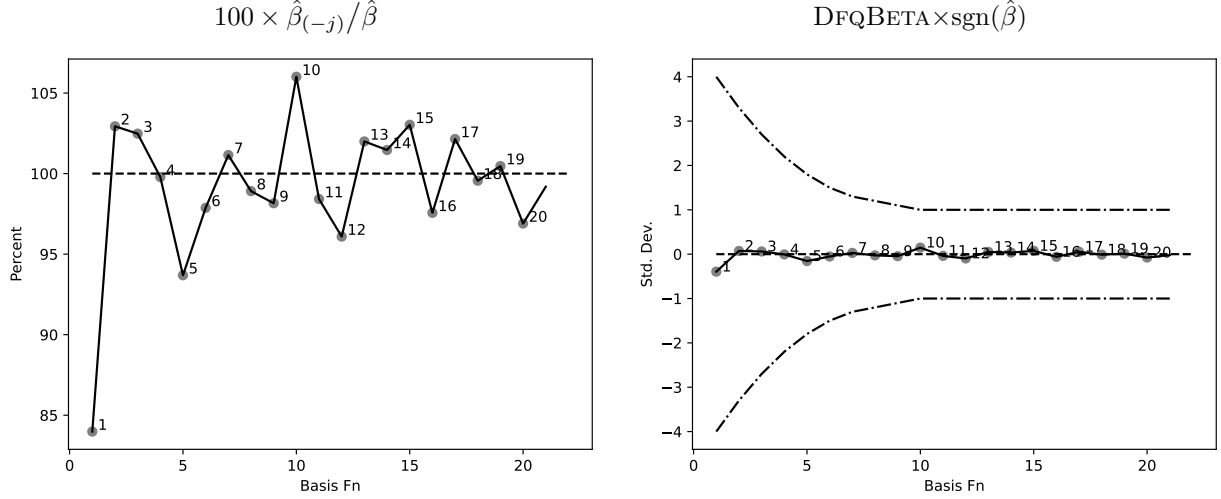


Figure 12 provides the leave-out estimates and diagnostic values for this application. In contrast to the case of [Nakamura and Steinsson \(2014\)](#), the leave-out estimates hew closely to the OLS estimate, especially when scaled by the reported standard error (see right chart). This application shows remarkable stability of the OLS estimate across frequencies.¹⁷

These examples reveal how our diagnostic can illuminate cases where the OLS point estimate is fragile with respect to multiple different frequencies.

6 Conclusion

Regression diagnostics have proved invaluable to applied researchers as a means to assess the reliability and robustness of their empirical results. Fewer diagnostic procedures are available for the challenging setting of time series, panel data, and spatial regressions where the data might feature substantial dependence across space and time. We introduce a novel approach to constructing regression diagnostics in these settings. For time-series and panel data regressions, the diagnostic

¹⁶Standard errors are adjusted for two-sample 2SLS using [Chodorow-Reich and Wieland \(2020\)](#).

¹⁷One source of the stability may be the exponential smoothing which dampens the variability of the regressors.

characterizes the contribution to the OLS estimate from different frequencies. Our proposed diagnostic relies on a new result which demonstrates the appealing orthogonalization properties of the eigenvectors of the variance matrix of a random walk and facilitates the clean interpretation of our new regression diagnostic.

To remain agnostic about the underlying properties of the true data generating process, we have proposed conservative thresholds to guide the applied researcher on which values of the diagnostic signal the need for further scrutiny of the empirical results. We demonstrate the wide applicability and usefulness of our new approach in five different empirical applications taken from recent papers in the Economics literature. In each of these applications, our diagnostic highlights important properties of the estimated specification such as the influence of common trending behavior or the disproportionate influence of information at certain frequencies.

In general economic settings, the object of interest is the regression coefficient and this has been the focus of the paper. In other settings, such as empirical asset pricing, the object of interest is the regression fitted value, which represents the estimated risk premia, or nonlinear transformations of regression coefficients, which represent the estimated price of risk. We introduce regression diagnostics tailored for these use cases in a companion paper, [Crump et al. \(2024b\)](#). The techniques which underpin the regression diagnostics can also be used to construct variance estimators for time-series, panel-data, and spatial regressions (see [Crump et al., 2024a](#)).

References

- Barnichon, R., Mesters, G., 2020. Identifying modern macro equations with old shocks. *Quarterly Journal of Economics* 135, 2255–2298.
- Belsley, D. A., Kuh, E., Welsch, R. E., 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Hoboken, New Jersey.
- Chan, N. H., Wei, C. Z., 1988. Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367–401.
- Chodorow-Reich, G., Wieland, J., 2020. Secular labor reallocation and business cycles. *Journal of Political Economy* 128, 2245–2287.
- Corbae, D., Ouliaris, S., Phillips, P. C. B., 1994. A reexamination of the consumption function using frequency domain regressions. *Empirical Economics* 19, 595–609.
- Corbae, D., Ouliaris, S., Phillips, P. C. B., 2002. Band spectral regression with trending data. *Econometrica* 70, 1067–1109.
- Crump, R. K., Gospodinov, N., 2021. On the factor structure of bond returns. *Econometrica* 9, 295–314.
- Crump, R. K., Gospodinov, N., Lopez Gaffney, I., 2024a. A jackknife variance estimator for panel regressions. Staff Report 1133, Federal Reserve Bank of New York.
- Crump, R. K., Gospodinov, N., Lopez Gaffney, I., 2024b. Robust empirical asset pricing, working paper.
- Dwivedi, Y., Subba Rao, S., 2011. A test for second-order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* 32, 68–91.
- Engle, R. F., 1980a. Exact maximum likelihood methods for dynamic regressions and band spectrum regressions. *International Economic Review* 21, 391–407.
- Engle, R. F., 1980b. Hypothesis testing in spectral regression; the lagrange multiplier test as a regression diagnostic. In: *Evaluation of Econometric Models*, Elsevier, pp. 309–321.
- Engle, R. F., Gardner, R., 1976. Some finite sample properties of spectral estimators of a linear regression. *Econometrica* 44, 149–165.
- Farber, H. S., Herbst, D., Kuziemko, I., Naidu, S., 2021. Unions and inequality over the twentieth century: New evidence from survey data. *Quarterly Journal of Economics* 136, 1325–1385.
- Granger, C. W., 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14, 227–238.
- Hannan, E., 1973. Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* .
- Hannan, E. J., 1963a. Regression for time series. In: Rosenblatt, M. (ed.), *Proceedings of the Symposium on Time Series Analysis*, John Wiley and Sons.
- Hannan, E. J., 1963b. Regression for time series with errors of measurement. *Biometrika* 50, 293–302.
- Harvey, A. C., 1978. Linear regression in the frequency domain. *International Economic Review* pp. 507–512.
- Hazell, J., Herreño, J., Nakamura, E., Steinsson, J., 2022. The slope of the Phillips curve: Evidence from U.S. states. *Quarterly Journal of Economics* 137, 1299–1344.
- Lahiri, S. N., 2003. A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Annals of Statistics* 31, 613 – 641.

- Lazarus, E., Lewis, D. J., Stock, J. H., Watson, M. W., 2018. HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics* 36, 541–559.
- Müller, U., Watson, M., 2024a. Low-frequency analysis of economic time series. In: Durlauf, S., Hansen, L., Heckman, J., Matzkin, R. (eds.), *Handbook of Econometrics*, Elsevier, vol. 7B, (forthcoming).
- Müller, U., Watson, M., 2024b. Spatial unit roots and spurious regression. *Econometrica* 92, 1661–1695.
- Müller, U. K., Watson, M. W., 2008. Testing models of low-frequency variability. *Econometrica* 76, 979–1016.
- Nakamura, E., Steinsson, J., 2014. Fiscal stimulus in a monetary union: Evidence from US regions. *American Economic Review* 104, 753–92.
- Philippon, T., Reshef, A., 2012. Wages and human capital in the U.S. finance industry: 1909–2006. *Quarterly Journal of Economics* 127, 1551–1609.
- Phillips, P. C., 1998. New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.
- Phillips, P. C., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.
- Romer, C. D., Romer, D. H., 2004. A new measure of monetary shocks: Derivation and implications. *American Economic Review* 94, 1055–1084.
- Tanaka, K., 2017. *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley, second ed.

Appendix

A.1 Bounds

The bounds introduced in Section 4, are:

j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Upper	4	3.3	2.7	2.2	1.8	1.5	1.3	1.2	1.1	1	1	1	1	1	1	1	1	1	1	1
Lower	-4	-3.3	-2.7	-2.2	-1.8	-1.5	-1.3	-1.2	-1.1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

A.2 Results for Other Time Series Processes

We consider two additional time series processes. The first one is a modified local-to-unity process (e.g., [Chan and Wei, 1988](#); [Phillips and Magdalinos, 2007](#)), where $\rho_{T,a} = 1 - \frac{c}{T^a}$ and

$$x_t = \rho_{T,a} x_{t-1} + \epsilon_t. \quad (\text{A.1})$$

When $a = 1$, this is the conventional local-to-unity process. In Lemma [A.1](#), we describe the orthogonalization properties of our basis functions for this class of time series processes.

Lemma A.1. *Let x be a modified local-to-unity process (as in equation (A.1)) with autoregressive coefficient $\rho_{T,a} = 1 - \frac{c}{T^a}$ for $c, a > 0$. Then, for all $j, k \in \{1, \dots, T\}$ with $j \neq k$,*

$$(i) \text{ COV}(\psi'_j x, \psi'_k x) = O(T^{\min(3a-1, 3-a)});$$

$$(i) \text{ CORR}(\psi'_j x, \psi'_k x) = O(T^{-|1-a|}).$$

Proof: See Appendix [A.4](#).

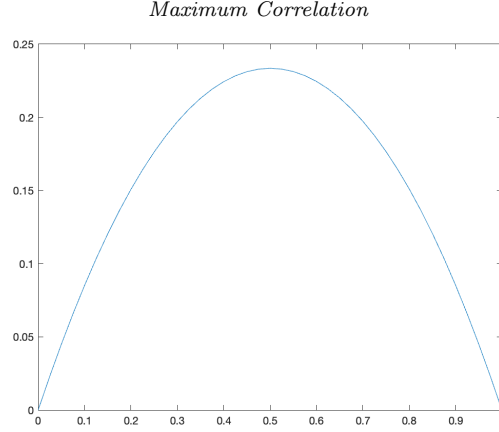
Lemma [A.1](#) readily implies that $\text{CORR}(\psi'_j x, \psi'_k x) \rightarrow 0$ as $T \rightarrow \infty$ so long as $a \neq 1$. When $a = 1$, Lemma [A.1](#) instead implies that $\text{CORR}(\psi'_j x, \psi'_k x) = O(1)$ for at least one pair j and k . Numerically, we can show that the maximum correlation across (j, k) pairs occurs at $(1, 2)$ and is highest when $c \approx 2$, peaking at a value of around 0.20. For all other pairs, the correlation is much closer to zero (see Section [A.3](#)).

The second time series process we consider is an ARFIMA(0, d , 0) model:

$$(1 - L)^d x_t = \epsilon_t, \quad (\text{A.2})$$

where $x \sim I(d)$ for $d \in \mathbb{R}$. As in the modified local-to-unity case, we have that $\text{CORR}(\psi'_j x, \psi'_k x) = O(1)$ for at least one pair j and k . Recall that for $d = 0$ or $d = 1$, $\text{CORR}(\psi'_j x, \psi'_k x) = 0$. Numerically, we can show that the maximum correlation across (j, k) pairs occurs at $(1, 2)$ and is highest when $d = .5$, peaking at a value around 0.23, as shown in Figure [A.1](#). For all other pairs, the correlation is much closer to zero.

Figure A.1. Maximum Correlation for Fractionally Integrated Models. This chart shows the maximum correlation (which occurs at $(j, k) = (1, 2)$) for different values of d for the fractionally integrated model given in equation (A.2)



A.3 Finite-Sample Evidence

In this subsection, we provide finite-sample evidence regarding the orthogonalization properties of our choice of basis functions. We consider the following time-series models to generate the data $x = (x_1, \dots, x_T)'$.

1. **AR(1)**: $x_t = \rho \cdot x_{t-1} + \epsilon_t$;
2. **AR(3)**: $x_t = \rho \cdot x_{t-1} - 0.9\rho \cdot x_{t-2} + 0.81\rho \cdot x_{t-3} + \epsilon_t$;
3. **VAR(1)**: $(x_t, y_t)' = \Phi(x_{t-1}, y_{t-1})' + (\epsilon_t^x, \epsilon_t^y)'$ with $\Phi = \begin{bmatrix} 0 & 1 - \rho \\ 0 & \rho \end{bmatrix}$;
4. **LTU** (Modified local-to-unity): $x_t = (1 - \frac{c}{Ta}) \cdot x_{t-1} + \epsilon_t$ with $a = 0.5, 1$ and 1.5 ;
5. **ARFIMA**(0, d , 0): $(1 - L)^d x_t = \epsilon_t$.

We consider two distributions for the error term: standard Gaussian or chi-square with 1 degree of freedom, properly recentered and rescaled to have mean zero and variance one. For the VAR model, the errors are drawn from a bivariate standard Gaussian or transformed independent chi-square(1) random variables with a correlation coefficient of 0.3. All processes are initialized at zero.

Tables A.1 and A.2 present results for the correlation coefficient, $\text{CORR}(\psi_j' x, \psi_k' x)$, and the two error distributions. We report the maximum correlation along with the average absolute correlation across all (j, k) pairs, calculated over 100,000 simulations. As a benchmark for comparison, we also report the results using the discrete Fourier transform (DFT) to transform the data rather than the basis functions introduced in Section 2.1.

We start by noting that there is very little difference in the results for Gaussian (Table A.1) and chi-square(1) (Table A.2) errors which suggests that the results are insensitive to innovations

with densities that deviate substantially from Gaussianity.¹⁸ For AR and VAR processes, the correlations for both DFT and Ψ basis functions tend to increase with the persistence parameter ρ although the correlations for the Ψ basis functions are generally much smaller and go to zero faster as the sample size increases. The differences between the two methods are especially large for the LTU (with $a = 1$ and $a = 1.5$) and the ARFIMA processes. The results for the Ψ basis functions are in line with the numbers reported in Section A.2 with the average absolute correlations being essentially zero for all specifications. This is in contrast to the DFT where both the maximum and average correlations remain non-zero and large in magnitude even when $T = 800$.

¹⁸Results for a Student-t distribution with 3 degrees of freedom and a GARCH(.05, .25, .7) process are also very similar to those in Tables A.1–A.2 and are not reported.

Table A.1. Orthogonalization Properties (Gaussian Innovations). This table presents the maximum absolute correlation and average absolute correlation across all pairs $\{(j, k) : j, k = 1, \dots, T\}$ where the data have been transformed using either the j th and k th discrete Fourier transform (DFT) or ψ_j and ψ_k .

	Discrete Fourier Transform (DFT)								Eigenvectors of Random Walk (Ψ)							
	Max. Abs. Corr.				Avg. Abs. Corr.				Max. Abs. Corr.				Avg. Abs. Corr.			
AR(1) $T \backslash \rho$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.04	0.08	0.23	0.34	0.02	0.03	0.09	0.17	0.03	0.06	0.16	0.2	0.01	0.01	0	0
100	0.02	0.04	0.14	0.24	0.01	0.02	0.05	0.09	0.02	0.03	0.1	0.16	0	0	0	0
200	0.02	0.03	0.07	0.14	0	0.01	0.03	0.05	0.01	0.02	0.05	0.09	0	0	0	0
400	0.01	0.02	0.04	0.07	0	0	0.01	0.03	0.02	0.02	0.03	0.06	0	0	0	0
800	0.01	0.01	0.02	0.04	0	0	0.01	0.01	0.02	0.02	0.02	0.03	0	0	0	0
AR(3) $T \backslash \rho$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.16	0.21	0.34	0.37	0.03	0.04	0.07	0.08	0.16	0.22	0.32	0.33	0.03	0.03	0.04	0.05
100	0.08	0.11	0.16	0.2	0.02	0.02	0.03	0.04	0.08	0.11	0.18	0.2	0.01	0.02	0.02	0.02
200	0.04	0.05	0.08	0.11	0.01	0.01	0.02	0.02	0.04	0.06	0.09	0.1	0.01	0.01	0.01	0.01
400	0.02	0.03	0.04	0.05	0	0.01	0.01	0.01	0.03	0.03	0.05	0.05	0	0	0.01	0.01
800	0.02	0.02	0.03	0.03	0	0	0.01	0.01	0.02	0.02	0.03	0.03	0	0	0	0
VAR(1) $T \backslash \rho$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.06	0.1	0.25	0.36	0.02	0.02	0.08	0.16	0.05	0.06	0.15	0.18	0.02	0.02	0.02	0.02
100	0.03	0.05	0.15	0.25	0.01	0.01	0.04	0.09	0.03	0.04	0.09	0.15	0.01	0.01	0.01	0.01
200	0.02	0.03	0.08	0.15	0.01	0.01	0.02	0.04	0.02	0.02	0.05	0.09	0	0.01	0	0
400	0.02	0.02	0.04	0.08	0	0	0.01	0.02	0.02	0.02	0.03	0.05	0	0	0	0
800	0.01	0.02	0.02	0.04	0	0	0.01	0.01	0.02	0.02	0.02	0.03	0	0	0	0
LTU ($a = 0.5$) $T \backslash c$	5	2	1	.5	5	2	1	.5	5	2	1	.5	5	2	1	.5
50	0.02	0.09	0.17	0.28	0.01	0.03	0.07	0.13	0.02	0.06	0.12	0.19	0	0.01	0	0
100	0.02	0.07	0.14	0.24	0.01	0.03	0.05	0.09	0.02	0.05	0.1	0.16	0	0	0	0
200	0.02	0.05	0.1	0.19	0.01	0.02	0.04	0.07	0.02	0.04	0.06	0.13	0	0	0	0
400	0.02	0.04	0.08	0.14	0.01	0.01	0.02	0.05	0.02	0.03	0.05	0.09	0	0	0	0
800	0.02	0.03	0.05	0.11	0	0.01	0.02	0.03	0.02	0.02	0.04	0.07	0	0	0	0
LTU ($a = 1$) $T \backslash c$	5	2	1	.5	5	2	1	.5	5	2	1	.5	5	2	1	.5
50	0.22	0.37	0.47	0.55	0.09	0.2	0.31	0.39	0.15	0.2	0.16	0.11	0	0	0	0
100	0.24	0.38	0.48	0.55	0.09	0.2	0.3	0.39	0.16	0.2	0.16	0.1	0	0	0	0
200	0.24	0.38	0.48	0.55	0.09	0.2	0.3	0.39	0.16	0.2	0.16	0.1	0	0	0	0
400	0.24	0.39	0.49	0.56	0.09	0.2	0.3	0.39	0.16	0.19	0.16	0.1	0	0	0	0
800	0.24	0.39	0.49	0.56	0.09	0.2	0.3	0.39	0.17	0.2	0.16	0.1	0	0	0	0
LTU ($a = 1.5$) $T \backslash c$	5	2	1	.5	5	2	1	.5	5	2	1	.5	5	2	1	.5
50	0.51	0.58	0.61	0.62	0.35	0.44	0.47	0.49	0.13	0.06	0.04	0.01	0	0	0	0
100	0.55	0.6	0.62	0.63	0.39	0.46	0.48	0.49	0.1	0.04	0.02	0.02	0	0	0	0
200	0.58	0.62	0.63	0.63	0.42	0.47	0.48	0.49	0.08	0.03	0.02	0.01	0	0	0	0
400	0.6	0.62	0.63	0.64	0.44	0.48	0.49	0.49	0.07	0.02	0.01	0.01	0	0	0	0
800	0.61	0.63	0.64	0.64	0.46	0.48	0.49	0.5	0.05	0.02	0.02	0.01	0	0	0	0
ARFIMA $T \backslash d$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.28	0.43	0.58	0.62	0.03	0.11	0.33	0.42	0.22	0.16	0.05	0.05	0.02	0.02	0.02	0.02
100	0.27	0.42	0.58	0.62	0.02	0.08	0.3	0.4	0.23	0.19	0.07	0.03	0.01	0.01	0.01	0.01
200	0.27	0.42	0.57	0.61	0.01	0.05	0.27	0.38	0.23	0.19	0.08	0.04	0.01	0.01	0	0
400	0.26	0.41	0.57	0.61	0.01	0.04	0.25	0.37	0.24	0.19	0.08	0.04	0	0	0	0
800	0.26	0.41	0.57	0.61	0	0.03	0.22	0.35	0.23	0.19	0.08	0.04	0	0	0	0

Table A.2. Orthogonalization Properties (Chi-Square Innovations). This table presents the maximum absolute correlation and average absolute correlation across all pairs $\{(j, k) : j, k = 1, \dots, T\}$ where the data have been transformed using either the j th and k th discrete Fourier transform (DFT) or ψ_j and ψ_k .

	Discrete Fourier Transform (DFT)								Eigenvectors of Random Walk (Ψ)							
	Max. Abs. Corr.				Avg. Abs. Corr.				Max. Abs. Corr.				Avg. Abs. Corr.			
AR(1) $T \backslash \rho$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.04	0.08	0.23	0.33	0.02	0.03	0.09	0.17	0.03	0.06	0.15	0.2	0.01	0.01	0	0
100	0.02	0.04	0.14	0.23	0.01	0.02	0.05	0.09	0.02	0.03	0.09	0.16	0	0	0	0
200	0.02	0.02	0.07	0.14	0	0.01	0.03	0.05	0.02	0.02	0.05	0.1	0	0	0	0
400	0.01	0.02	0.04	0.08	0	0	0.01	0.02	0.02	0.02	0.03	0.05	0	0	0	0
800	0.01	0.01	0.02	0.04	0	0	0.01	0.01	0.02	0.01	0.02	0.03	0	0	0	0
AR(3) $T \backslash \rho$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.16	0.21	0.34	0.38	0.03	0.04	0.07	0.08	0.16	0.23	0.31	0.33	0.03	0.03	0.04	0.05
100	0.08	0.11	0.16	0.2	0.02	0.02	0.04	0.04	0.08	0.12	0.18	0.2	0.01	0.02	0.02	0.02
200	0.04	0.05	0.08	0.11	0.01	0.01	0.02	0.02	0.05	0.06	0.09	0.1	0.01	0.01	0.01	0.01
400	0.02	0.03	0.05	0.05	0	0.01	0.01	0.01	0.03	0.04	0.05	0.05	0	0	0.01	0.01
800	0.02	0.02	0.02	0.03	0	0	0.01	0.01	0.02	0.02	0.03	0.03	0	0	0	0
VAR(1) $T \backslash \rho$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.06	0.1	0.25	0.36	0.02	0.02	0.08	0.16	0.06	0.06	0.14	0.18	0.02	0.02	0.02	0.02
100	0.03	0.05	0.15	0.25	0.01	0.01	0.04	0.09	0.03	0.04	0.1	0.16	0.01	0.01	0.01	0.01
200	0.02	0.03	0.08	0.15	0.01	0.01	0.02	0.04	0.02	0.02	0.05	0.1	0.01	0.01	0.01	0.01
400	0.01	0.02	0.04	0.08	0	0	0.01	0.02	0.02	0.02	0.03	0.05	0	0	0	0
800	0.01	0.02	0.03	0.04	0	0	0.01	0.01	0.02	0.02	0.02	0.03	0	0	0	0
LTU ($a = 0.5$) $T \backslash c$	5	2	1	.5	5	2	1	.5	5	2	1	.5	5	2	1	.5
50	0.02	0.08	0.17	0.28	0.01	0.03	0.07	0.13	0.02	0.06	0.12	0.18	0	0.01	0.01	0
100	0.02	0.07	0.13	0.24	0.01	0.03	0.05	0.09	0.02	0.05	0.08	0.16	0	0	0	0
200	0.02	0.05	0.11	0.19	0.01	0.02	0.03	0.07	0.02	0.04	0.07	0.13	0	0	0	0
400	0.02	0.04	0.08	0.14	0.01	0.01	0.02	0.05	0.02	0.03	0.05	0.1	0	0	0	0
800	0.02	0.03	0.06	0.11	0	0.01	0.02	0.03	0.02	0.02	0.04	0.07	0	0	0	0
LTU ($a = 1$) $T \backslash c$	5	2	1	.5	5	2	1	.5	5	2	1	.5	5	2	1	.5
50	0.22	0.37	0.48	0.54	0.09	0.2	0.31	0.39	0.16	0.19	0.16	0.11	0	0	0	0
100	0.24	0.38	0.48	0.55	0.09	0.2	0.3	0.39	0.16	0.19	0.15	0.1	0	0	0	0
200	0.24	0.38	0.48	0.56	0.09	0.2	0.3	0.39	0.16	0.2	0.16	0.1	0	0	0	0
400	0.24	0.39	0.48	0.55	0.09	0.2	0.3	0.39	0.16	0.2	0.16	0.11	0	0	0	0
800	0.24	0.39	0.49	0.56	0.09	0.2	0.3	0.39	0.16	0.2	0.16	0.1	0	0	0	0
LTU ($a = 1.5$) $T \backslash c$	5	2	1	.5	5	2	1	.5	5	2	1	.5	5	2	1	.5
50	0.51	0.58	0.61	0.62	0.35	0.44	0.47	0.49	0.13	0.06	0.03	0.02	0	0	0	0
100	0.55	0.6	0.62	0.63	0.39	0.46	0.48	0.49	0.1	0.05	0.03	0.01	0	0	0	0
200	0.58	0.61	0.63	0.63	0.42	0.46	0.48	0.49	0.08	0.04	0.02	0.01	0	0	0	0
400	0.59	0.62	0.63	0.64	0.44	0.48	0.49	0.49	0.06	0.03	0.01	0.01	0	0	0	0
800	0.61	0.63	0.64	0.64	0.46	0.48	0.49	0.5	0.04	0.02	0.02	0.02	0	0	0	0
ARFIMA $T \backslash d$.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95	.5	.7	.9	.95
50	0.28	0.43	0.58	0.62	0.03	0.11	0.33	0.41	0.21	0.18	0.05	0.05	0.02	0.02	0.02	0.02
100	0.27	0.42	0.57	0.61	0.02	0.08	0.3	0.4	0.22	0.18	0.07	0.03	0.01	0.01	0.01	0.01
200	0.27	0.42	0.58	0.61	0.01	0.05	0.27	0.38	0.23	0.19	0.07	0.04	0.01	0.01	0	0
400	0.27	0.42	0.57	0.61	0.01	0.04	0.25	0.37	0.23	0.2	0.08	0.04	0	0	0	0
800	0.26	0.42	0.57	0.61	0	0.03	0.22	0.35	0.23	0.19	0.08	0.04	0	0	0	0

A.4 Proofs

Proof of Lemma 1

Proof. By definition, we have

$$\text{COV}(\psi'_j x, \psi'_k x) = \frac{4}{2T+1} \sum_{t,s} \sin\left(\frac{\pi t(2k-1)}{2T+1}\right) \sin\left(\frac{\pi s(2j-1)}{2T+1}\right) \gamma(t-s),$$

where $\gamma(\cdot)$ is the autocovariance function of x_t . Since $\sin(A)\sin(B) = \frac{1}{2}(\cos(A-B) - \cos(A+B))$, we can write $\text{COV}(\psi'_j x, \psi'_k x)$ as

$$\begin{aligned} & \frac{2}{2T+1} \sum_{r=1-T}^{T-1} \gamma(r) \sum_{s \in I(r)} \cos\left(\frac{\pi[(2k-1)r + 2s(k-j)]}{2T+1}\right) - \cos\left(\frac{\pi[(2k-1)r + 2s(k+j-1)]}{2T+1}\right) \\ &= \frac{2}{2T+1} \sum_{r=1-T}^{T-1} \gamma(r) g(r), \end{aligned}$$

where

$$I(r) = \begin{cases} \{1, \dots, T\} & r = 0 \\ \{1, \dots, T-r\} & r > 0 \\ \{r+1, \dots, T\} & r < 0. \end{cases}$$

Using that $\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$, we can express the weight function $g(r)$ as

$$\begin{aligned} g(r) &= \sum_{s \in I(r)} \cos\left(\frac{\pi[(2k-1)r + 2s(k-j)]}{2T+1}\right) - \cos\left(\frac{\pi[(2k-1)r + 2s(k+j-1)]}{2T+1}\right) \\ &= \cos\left(\frac{\pi(2k-1)r}{2T+1}\right) \sum_{s \in I(r)} \cos\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \cos\left(\frac{\pi 2s(k+j-1)}{2T+1}\right) \\ &\quad - \sin\left(\frac{\pi(2k-1)r}{2T+1}\right) \sum_{s \in I(r)} \sin\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \sin\left(\frac{\pi 2s(k+j-1)}{2T+1}\right). \end{aligned}$$

First, when $r = 0$, $g(r)$ reduces to

$$\sum_{s=1}^T \cos\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \cos\left(\frac{\pi 2s(k+j-1)}{2T+1}\right) = 0.$$

Second, when $r > 0$, $g(r)$ is equivalent to

$$\begin{aligned} & \cos\left(\frac{\pi(2k-1)r}{2T+1}\right) \sum_{s=1}^{T-r} \cos\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \cos\left(\frac{\pi 2s(k+j-1)}{2T+1}\right) \\ & \quad - \sin\left(\frac{\pi(2k-1)r}{2T+1}\right) \sum_{s=1}^{T-r} \sin\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \sin\left(\frac{\pi 2s(k+j-1)}{2T+1}\right) \\ &= (-1)^{j-k+1} \frac{1}{2} \sin\left(\frac{\pi r(2j-1)}{2T+1}\right) \left[\csc\left(\frac{\pi(j-k)}{2T+1}\right) + \csc\left(\frac{\pi(j+k-1)}{2T+1}\right) \right] \\ & \quad + \frac{1}{2} \sin\left(\frac{\pi r(2k-1)}{2T+1}\right) \left[\cot\left(\frac{\pi(j-k)}{2T+1}\right) + \cot\left(\frac{\pi(j+k-1)}{2T+1}\right) \right]. \end{aligned}$$

Third, when $r < 0$, $g(r)$ is equivalent to

$$\begin{aligned} & \cos\left(\frac{\pi(2k-1)r}{2T+1}\right) \sum_{s=1-r}^T \cos\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \cos\left(\frac{\pi 2s(k+j-1)}{2T+1}\right) \\ & \quad - \sin\left(\frac{\pi(2k-1)r}{2T+1}\right) \sum_{s=1-r}^T \sin\left(\frac{\pi 2s(k-j)}{2T+1}\right) - \sin\left(\frac{\pi 2s(k+j-1)}{2T+1}\right) \end{aligned}$$

$$\begin{aligned}
&= (-1)^{j-k+1} \frac{1}{2} \sin\left(\frac{\pi r(2k-1)}{2T+1}\right) \left[\csc\left(\frac{\pi(j-k)}{2T+1}\right) - \csc\left(\frac{\pi(j+k-1)}{2T+1}\right) \right] \\
&\quad + \frac{1}{2} \sin\left(\frac{\pi r(2j-1)}{2T+1}\right) \left[\cot\left(\frac{\pi(j-k)}{2T+1}\right) - \cot\left(\frac{\pi(j+k-1)}{2T+1}\right) \right].
\end{aligned}$$

Returning to our object of interest, as $\gamma(r) = \gamma(-r)$ under strict stationarity, we have

$$\frac{2}{2T+1} \sum_{r=1-T}^{T-1} \gamma(r)g(r) = \frac{2}{2T+1} \sum_{r=1}^{T-1} \gamma(r)(g(r) + g(-r)) = \frac{2}{2T+1} \sum_{r=1}^{T-1} \gamma(r)w(r),$$

where $w(r)$

$$\begin{aligned}
&= \frac{1}{2} \csc\left(\frac{\pi(j-k)}{2T+1}\right) \left[(-1)^{j-k} + \cos\left(\frac{\pi(j-k)}{2T+1}\right) \right] \left[-\sin\left(\frac{\pi r(2j-1)}{2T+1}\right) + \sin\left(\frac{\pi r(2k-1)}{2T+1}\right) \right] \\
&\quad + \frac{1}{2} \csc\left(\frac{\pi(j+k-1)}{2T+1}\right) \left[(-1)^{j+k-1} + \cos\left(\frac{\pi(j+k-1)}{2T+1}\right) \right] \left[\sin\left(\frac{\pi r(2j-1)}{2T+1}\right) + \sin\left(\frac{\pi r(2k-1)}{2T+1}\right) \right].
\end{aligned}$$

We consider separately the case where j, k are fixed and then the case where j, k are a function of T .

Case 1: j, k are fixed. By a series expansion at $(y = \infty)$, we can show that

$$\begin{aligned}
\csc(y^{-1}) &= y + O(y^{-1}) = O(y) \\
\cos(y^{-1}) &= 1 - O(y^{-2}) = O(1),
\end{aligned}$$

i.e., for any functions $f_1(\cdot), f_2(\cdot)$

$$f_1(y) = O(f_2(y)) \iff |f_1(y)| \lesssim f_2(y) \quad \forall y \geq y_0.$$

This implies that, uniformly over r ,

$$\begin{aligned}
w(r) &\lesssim T \left(-\sin\left(\frac{\pi r(2j-1)}{2T+1}\right) + \sin\left(\frac{\pi r(2k-1)}{2T+1}\right) \right) + T \left(\sin\left(\frac{\pi r(2j-1)}{2T+1}\right) + \sin\left(\frac{\pi r(2k-1)}{2T+1}\right) \right) \\
&\lesssim T \sin(T^{-1}r).
\end{aligned}$$

Now, by the assumed convergence properties of $\gamma(r)$, $\forall r \geq M_\epsilon$

$$\frac{c_-}{r^a} \leq \gamma(r) \leq \frac{c_+}{r^a},$$

where $c_- = c_\gamma - \epsilon$ and $c_+ = c_\gamma + \epsilon$. Using this, we can divide our term into a component with the number of summands growing in the sample size, and another with a bounded number of summands.

$$\frac{2}{2T+1} \sum_{r=1}^{T-1} w(r)\gamma(r) = \frac{2T}{2T+1} \left[\frac{1}{T} \sum_{r=M_\epsilon}^{T-1} w(r)\gamma(r) + \frac{1}{T} \sum_{r=1}^{M_\epsilon-1} w(r)\gamma(r) \right].$$

Consider the second summand, where r is bounded by construction. Here, we have that $w(r) = O(T \sin(T^{-1}))$. We can, therefore, consider a series expansion at $(y = \infty)$ of,

$$w(y) = O(y \sin(y^{-1})) = 1 - O(x^{-2}) = O(1).$$

Further, as M_ϵ is fixed and $\gamma(r)$ is bounded, we can conclude that

$$\frac{1}{T} \sum_{r=1}^{M_\epsilon-1} w(r)\gamma(r) = O(T^{-1}).$$

Turning to the first summand, since $\gamma(r) \forall r \geq M_\epsilon$, then the summand lies in the interval

$$[F_T \cdot c_-, F_T \cdot c_+],$$

where

$$F_T = \frac{2T}{2T+1} \left(\frac{1}{T} \sum_{r=M_\epsilon}^{T-1} \left| \left(T \frac{r}{T} \right)^{-a} w\left(T \frac{r}{T} \right) \right| \right).$$

Moreover, as $w(r)$ is continuous, we can bound F_T by

$$\begin{aligned} \left(\frac{2T}{2T+1} \right) \frac{1}{T} \sum_{r=M_\epsilon}^{T-1} \left| \left(T \frac{r}{T} \right)^{-a} w \left(T \frac{r}{T} \right) \right| &\leq \frac{2T}{2T+1} \int_{M_\epsilon/T}^{1-1/T} |(Tx)^{-a} w(Tx)| dx [1 + O(T^{-1})] \\ &= \frac{2}{T^{a-1}(2T+1)} \int_{M_\epsilon/T}^{1-1/T} x^{-a} |w(Tx)| dx [1 + O(T^{-1})]. \end{aligned}$$

Next, we simplify this integral further. Inspecting our expression for $w(r)$, it is clear that it only depends on r through its dependence on $\sin(\cdot)$, so we only have to integrate over this part:

$$\begin{aligned} &\frac{2}{T^{a-1}(2T+1)} \int_{M_\epsilon/T}^{1-1/T} x^{-a} \left| \sin \left(\frac{\pi T x (2j-1)}{2T+1} \right) \right| dx \\ &= \frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T (2j-1)}{2T+1} \right)^a \int_{M_\epsilon/T}^{1-1/T} \left(\frac{\pi T x (2j-1)}{2T+1} \right)^{-a} \left| \sin \left(\frac{\pi T x (2j-1)}{2T+1} \right) \right| dx \\ &= \frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T (2j-1)}{2T+1} \right)^{a-1} \int_{\frac{\pi M_\epsilon (2j-1)}{2T+1}}^{\frac{\pi (T-1)(2j-1)}{2T+1}} u^{-a} |\sin(u)| du \\ &= \frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T (2j-1)}{2T+1} \right)^{a-1} V_a(u) \Big|_{\frac{\pi M_\epsilon (2j-1)}{2T+1}}^{\frac{\pi (T-1)(2j-1)}{2T+1}}, \end{aligned}$$

where

$$V_a(u) = \frac{-\operatorname{sgn}(\sin(u^{-1}))}{2} ((-i)^a \Gamma(1-a, -iu^{-1}) + i^a \Gamma(1-a, iu^{-1}))$$

and $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function. Note that $V_a(u)$ is real and positive. Taking a series expansion at $(y = \infty)$ we know that

$$\begin{aligned} (-i)^a \Gamma(1-a, -iy^{-1}) &= (-i)^a \Gamma(1-a) - \frac{iy^{a-1}}{a-1} + \frac{y^{a-2}}{a-2} + \frac{iy^{a-3}}{2a-6} + O(y^{a-4}) \\ i^a \Gamma(1-a, iy^{-1}) &= i^a \Gamma(1-a) + \frac{iy^{a-1}}{a-1} + \frac{y^{a-2}}{a-2} - \frac{iy^{a-3}}{2a-6} + O(y^{a-4}), \end{aligned}$$

so in the vicinity of $(u = \infty)$

$$V_a(u) = -\operatorname{sgn}(\sin(u^{-1})) \left[\cos \left(\frac{\pi a}{2} \right) \Gamma(1-a) + \frac{u^{a-2}}{a-2} + O(u^{a-4}) \right].$$

Taking a series expansion at $(y = 0)$, we know that

$$\begin{aligned} (-i)^a \Gamma(1-a, -iy^{-1}) &= \exp(-i/y) [y^a - iay^{a+1} + O(y^{a+2})] \\ i^a \Gamma(1-a, iy^{-1}) &= \exp(i/y) [y^a + iay^{a+1} + O(y^{a+2})] \end{aligned}$$

so in the vicinity of $(u = 0)$

$$V_a(u) = -\operatorname{sgn}(\sin(u^{-1})) \cos(u^{-1}) [u^a + O(u^{a+2})].$$

We now consider three subcases depending on the value of a .

Subcase 1: If $a \in (1, 2)$,

$$V_a(u) \Big|_{O(T)}^{O(1)} \leq \left| \cos \left(\frac{\pi a}{2} \right) \Gamma(1-a) \right| + O(T^{a-2}) + |\cos(O(1)) \cdot O(1)| = O(1).$$

It follows that

$$\frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T (2j-1)}{2T+1} \right)^{a-1} V_a(u) \Big|_{\frac{\pi M_\epsilon (2j-1)}{2T+1}}^{\frac{\pi (T-1)(2j-1)}{2T+1}} = O(T^{-a}) O(1) O(1) = O(T^{-a}).$$

Finally, multiplying by the terms that did not depend on r in $w(r)$ we conclude

$$F_T \leq O(T) O(T^{-a}) = O(T^{1-a}).$$

This, in turn, implies that

$$\text{Cov}(\psi'_j x, \psi'_k x) = O(T^{1-a}).$$

Next we consider the case where $a > 2$.

Subcase 2: $a \in (2, \infty)$. Suppose first that $a \in (2, \infty) \setminus \mathbb{Z}$. Then,

$$\begin{aligned} V_a(u)|_{O(T)}^{O(1)} &\leq \left| \cos\left(\frac{\pi a}{2}\right) \Gamma(1-a) \right| + O(T^{a-2}) + |\cos(O(1)) \cdot O(1)| \\ &= O(T^{a-2}). \end{aligned}$$

It follows that

$$\frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T(2j-1)}{2T+1} \right)^{a-1} V_a(u)|_{\frac{\pi(T-1)(2j-1)}{\pi M_\epsilon(2j-1)}}^{\frac{2T+1}{\pi(T-1)(2j-1)}} = O(T^{-a})O(1)O(T^{a-2}) = O(T^{-2}).$$

Finally, multiplying by the terms that do not depend on r in $w(r)$, we conclude

$$F_T \leq O(T)O(T^{-2}) = O(T^{-1}),$$

which in turn implies the result. As we can always find some $\tilde{a} < a$ where $\tilde{a} \in (2, \infty) \setminus \mathbb{Z}$, and $[F_T \cdot c_-, F_T \cdot c_+] \subset [\tilde{F}_T \cdot c_-, \tilde{F}_T \cdot c_+]$, we can then conclude that the result holds for all $a > 2$.

Subcase 3: If $a = 2$,

$$V_a(u) = \text{sgn}(\sin(u^{-1}))(\text{Ci}(u^{-1}) - u \sin(u^{-1})),$$

where $\text{Ci}(\cdot)$ is the Cosine integral, which is bounded for all positive arguments. By a series expansion at $(y = \infty)$, we know that

$$\begin{aligned} \text{Ci}(u^{-1}) &= \gamma^\circ - \ln(y) + O(y^{-2}) \\ y \sin(y^{-1}) &= 1 - O(y^{-2}), \end{aligned}$$

where γ° is the Euler-Mascheroni constant. It follows that

$$V_a(u)|_{O(T)}^{O(1)} \leq |\gamma^\circ| + |\ln(T)| + |1| + O(T^{-2}) = O(\ln(T)).$$

Thus,

$$\frac{2}{T(2T+1)} \left(\frac{\pi T(2j-1)}{2T+1} \right) V_a(u)|_{\frac{\pi M_\epsilon(2j-1)}{\pi(T-1)(2j-1)}}^{\frac{2T+1}{\pi(T-1)(2j-1)}} = O(T^{-2})O(1)O(\ln(T)) = O(\ln(T)T^{-2}).$$

Finally, multiplying by the terms that do not depend on r in $w(r)$, we conclude

$$F_T \leq O(T)O(\ln(T)T^{-2}) = O(\ln(T)T^{-1})$$

which in turn implies the result.

Case 2: j or k are $O(f(T))$, where $f'(T) > 0$ and $O(f(T)) \leq O(T)$. We know that

$$w(r) = O\left(\frac{T}{f(T)} \sin\left(\frac{f(T)r}{T}\right)\right)$$

In the second summand, where r is bounded,

$$\begin{aligned} \frac{1}{T} \sum_{r=1}^{M_\epsilon} w(r) \gamma(r) &= \frac{1}{T} \sum_{r=1}^{M_\epsilon} O\left(\frac{f(T)}{T} \cdot \frac{T}{f(T)}\right) \gamma(r) \\ &= O(T^{-1}). \end{aligned}$$

We again split into subcases depending on the value of a .

Subcase 1: For $a \in (1, 2)$,

$$V_a(u)|_{O(T/f(T))}^{O(f(T)^{-1})} \leq \left| \cos\left(\frac{\pi a}{2}\right) \Gamma(1-a) \right| + O\left(\left(\frac{T}{f(T)}\right)^{a-2}\right) + |\cos(f(T)) \cdot f(T)^{-a}| = O(1).$$

It follows that

$$\frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T(2j-1)}{2T+1} \right)^{a-1} V_a(u) \Big|_{\frac{2T+1}{\pi M_\epsilon(2j-1)}}^{\frac{2T+1}{\pi(T-1)(2j-1)}} = O(T^{-a}) O(f(T)^{a-1}) O(1) = O\left(\frac{f(T)^{a-1}}{T^a}\right).$$

Finally, multiplying by the terms that did not depend on r in $w(r)$ we conclude

$$F_T \leq O\left(\frac{T}{f(T)}\right) O\left(\frac{f(T)^{a-1}}{T^a}\right) = O\left(\frac{f(T)^{a-2}}{T^{a-1}}\right) \leq O(T^{1-a}).$$

We next consider the case where $a > 2$.

Subcase 2: $a \in (2, \infty)$. First, assume that $a \in (2, \infty) \setminus \mathbb{Z}$ so that

$$V_a(u) \Big|_{O(T/f(T))}^{O(f(T)^{-1})} \leq O\left(\left(\frac{T}{f(T)}\right)^{a-2}\right).$$

It follows that

$$\frac{2}{T^{a-1}(2T+1)} \left(\frac{\pi T(2j-1)}{2T+1} \right)^{a-1} V_a(u) \Big|_{\frac{2T+1}{\pi M_\epsilon(2j-1)}}^{\frac{2T+1}{\pi(T-1)(2j-1)}} \lesssim T^{-a} f(T)^{a-1} \left(\frac{T}{f(T)}\right)^{a-2} = f(T)^{-1} T^{-2}.$$

Finally, multiplying by the terms that do not depend on r in $w(r)$, we conclude

$$F_T \leq O\left(\frac{T}{f(T)}\right) O(f(T)^{-1} T^{-2}) = O(T^{-1}).$$

By similar arguments as above, we can conclude that the result holds for all $a > 2$.

Subcase 3: $a = 2$. In this case,

$$V_a(u) \Big|_{O(T/f(T))}^{O(f(T)^{-1})} \leq |\gamma^\circ| + \left| \ln\left(\frac{T}{f(T)}\right) \right| + |1| + O\left(\left(\frac{T}{f(T)}\right)^{-2}\right) = O\left(\ln\left(\frac{T}{f(T)}\right)\right).$$

It follows that

$$\frac{2}{T(2T+1)} \left(\frac{\pi T(2j-1)}{2T+1} \right) V_a(u) \Big|_{\frac{2T+1}{\pi M_\epsilon(2j-1)}}^{\frac{2T+1}{\pi(T-1)(2j-1)}} \lesssim T^{-2} f(T) \ln\left(\frac{T}{f(T)}\right).$$

Finally, multiplying by the terms that do not depend on r in $w(r)$, we conclude

$$F_T \lesssim \ln\left(\frac{T}{f(T)}\right) T^{-1} \lesssim \ln(T) T^{-1}.$$

Taking the two cases together, we have that the result holds uniformly over j and k such that $j \neq k$. \square

Proof of Lemma 2

Proof. As discussed in the main text, the result in Lemma 2 holds immediately by the partition regression formula and because Ψ forms an orthonormal basis. By the partition regression formula we have,

$$\begin{pmatrix} \tilde{\alpha}_{(-j)} \\ \tilde{\beta}_{(-j)} \end{pmatrix} = (\mathbf{X}'(I_T - \psi_j \psi_j') \mathbf{X})^{-1} \mathbf{X}'(I_T - \psi_j \psi_j') y = (\mathbf{X}' \mathbf{X} - \mathbf{Z}_j \mathbf{Z}_j')^{-1} (\mathbf{X}' y - \mathbf{Z}_j w_j).$$

Then, since Ψ is an orthonormal matrix we have that $\mathbf{X}' \mathbf{X} = \mathbf{Z}' \mathbf{Z}'$ and $\mathbf{X}' y = \mathbf{Z}' w$ which shows the equivalence with equation (18). \square

Proof of Lemma A.1

Proof. Let V be the variance-covariance matrix of the process, and \tilde{V} be the approximate variance-covariance matrix $\tilde{V} = \Psi \Lambda \Psi'$, where Λ is a diagonal matrix with diagonal elements

$$\tilde{\lambda}_j = \sigma^2 \left(1 + \rho_{T,a}^2 - 2\rho_{T,a} \cos\left(\frac{\pi(2n-1)}{2T+1}\right) \right)^{-1}, \quad (\text{A.3})$$

as in Lemma 2 of [Crump and Gospodinov \(2021\)](#). It follows that

$$\text{COV}(\psi'_j x, \psi'_k x) = \psi'_j V \psi_k = \psi'_j \tilde{V} \psi_k + \psi'_j (V - \tilde{V}) \psi_k = \mathfrak{R}_1 + \mathfrak{R}_2.$$

For \mathfrak{R}_1 , when $j \neq k$, then $\psi'_j \tilde{V} \psi_k = \tilde{\lambda}_j \psi'_j \psi_k = 0$. When $j = k$, $\psi'_j \tilde{V} \psi_j = \tilde{\lambda}_j \psi'_j \psi_j = \tilde{\lambda}_j$. Since $\sigma^2 \lesssim 1$, then

$$\begin{aligned} \tilde{\lambda}_j &\lesssim \left(1 + \rho_{T,a}^2 - 2\rho_{T,a} \cos \left(\frac{\pi(2n-1)}{2T+1} \right) \right)^{-1} \\ &\lesssim \left(1 + \left(1 - \frac{c}{T^a} \right)^2 - 2 \left(1 - \frac{c}{T^a} \right) (1 - T^{-2}/2) \right)^{-1} \\ &= (c^2 T^{-2a} + 2T^{-2} - 2cT^{-2-a})^{-1} \\ &\lesssim T^{\min(2a, 2)}. \end{aligned}$$

Thus, $\mathfrak{R}_1 = O(T^{\min(2a, 2)})$. Next, consider \mathfrak{R}_2 . For all j, k , we have

$$\psi'_j (V - \tilde{V}) \psi_k = -\psi'_j \nu \nu' \psi_j,$$

where

$$\psi'_j \nu \nu' \psi_k = \left(\frac{\rho_{T,a}(1 - \rho_{T,a})}{\sigma^2 + \rho_{T,a}(1 - \rho_{T,a}) \sum_{n=1}^N \tilde{\lambda}_n \psi_{T,n}^2} \right) \times (\tilde{\lambda}_j \tilde{\lambda}_k \psi_{T,j} \psi_{T,k}).$$

The second factor is

$$\tilde{\lambda}_j \tilde{\lambda}_k \psi_{T,j} \psi_{T,k} = \frac{4}{2T+1} \sin \left(\frac{T(2j-1)}{2T+1} \right) \sin \left(\frac{T(2k-1)}{2T+1} \right) \tilde{\lambda}_j \tilde{\lambda}_k \lesssim T^{\min(4a-1, 3)}.$$

For the first factor, note that

$$\sum_{n=1}^N \tilde{\lambda}_n \psi_{T,n}^2 = \frac{4}{2T+1} \sum_{n=1}^T \frac{\sigma^2}{1 - \rho_{T,a}^2 + 2\rho_{T,a} \cos \left(\frac{(2n-1)\pi}{2T+1} \right)} \cdot \sin^2 \left(\frac{T(2n-1)}{2T+1} \right) \lesssim T^{-1} \sum_{n=1}^T \tilde{\lambda}_n \lesssim T^{\min(a, 1)},$$

so that

$$\frac{\rho_{T,a}(1 - \rho_{T,a})}{\sigma^2 + \rho_{T,a}(1 - \rho_{T,a})T^{\min(a, 1)}} \lesssim \frac{c/T^a - c^2/T^{2a}}{\sigma^2 + cT^{\min(0, 1-a)} - c^2T^{\min(-a, 1-2a)}} \lesssim T^{-a}.$$

Putting this together, we have that $\psi'_j \nu \nu' \psi_k = O(T^{\min(3a-1, 3-a)})$. Thus, when $j \neq k$,

$$\text{COV}(\psi'_j x, \psi'_k x) = \psi'_j V \psi_k = O(T^{\min(3a-1, 3-a)}),$$

and when $j = k$,

$$\psi'_j V \psi_j = O(T^{\min(2a, 2)}) + O(T^{\min(3a-1, 3-a)}).$$

This implies that

$$\text{CORR}(\psi'_j x, \psi'_k x) = O(T^{-|1-a|}).$$

□