Low Frequency Cointegrating Regression with Local to Unity Regressors and Unknown Form of Serial Dependence *

Jungbin Hwang
Department of Economics,
University of Connecticut

Gonzalo Valdés
Departamento de Ingeniería Industrial y de Sistemas,
Universidad de Tarapacá

Dec, 2021

Abstract

This paper develops new t and F inferences in a low-frequency transformed triangular cointegrating regression when one may not be sure the economic variables are exact unit root processes. We first show that the low-frequency transformed and augmented OLS (TA-OLS) regression exhibits an asymptotic bias term in the limiting distribution. As a result, the size distortion of the testing cointegration vector can be substantially large for even minor deviations from the unit root regressors. We develop a method to correct the asymptotic bias for the cointegration vector. Our modified TA-OLS statistics adjust the locational bias and reflect the estimation uncertainty of the long-run endogeneity parameter in the bias correction term and lead to standard t and F critical values. Based on the modified test statistics, we provide Bonferroni-based inferences to test the cointegration vector. Monte Carlo results show that our approach has the correct size and appealing power for a wide range of local to unity parameters. Also, we find that our method has advantages to the IVX approach when the serial dependence and the long-run endogeneity in the cointegration system are important.

Keywords: Cointegration, Heteroscedasticity and autocorrelation-robust (HAR) inference, Local to unity, Low frequency transformation, t and F tests, Trasformed and augmented OLS (TA-OLS)

*Email: jungbin.hwang@uconn.edu and gvaldes@uta.cl. Correspondence to: Jungbin Hwang, 365 Fairfield Way, U-1063, Storrs, CT 06269-1063. We thank co-editor Christian Hansen, an anonymous associate editor, and anonymous referees for valuable comments. We also thank Yixiao Sun for the conversations in the early stage of this project.

1 Introduction

In economic theory, we often try to conclude a long run structural relationship among economic variables over long periods of time. When economic time series possess an exact unit root, the structural relationships between the non-stationary I(1) variables are captured by the concept of cointegration in Engle and Granger (1987). In most macroeconomic applications, however, it is arguable that fundamental economic variables follow the exact unit root process, e.g., Christiano and Eichenbaum (1990). Modeling key variables in the cointegration system using the unit roots assumption usually come in practice through a failure to reject the unit root hypothesis with a limited span of time series data (Elliott, 1998). Thus, the assumption of unit root in the cointegration model may simply represent a lack of knowledge about the economic interactions behind the common stochastic trends. See, for more details, Christiano and Eichenbaum (1990), Elliott (1998), and Müller and Watson (2008, 2013).

In the time series literature, it has been well established that several problems arise from the standard OLS procedure when the cointegration system has non-exact unit root regressors. First, the non-stationary cointegration regressors are endogenously correlated with cointegration errors. This results in a lack of mixed normality along with an unknown nuisance parameter (Park and Phillips, 1988; Phillips and Hansen, 1990). Secondly, a local-to-unity autoregressive specification for the non-unit regressor induces an uncorrectable bias in the limiting distribution, which are functions of several unknown nuisance parameters (Cavanagh et al., 1995; Elliott, 1998).

There are many studies that intend to solve these problems and search for robust inference methods. Cavanagh et al. (1995) introduce a pretest for identifying conditions under which the conventional t-test is invalid and propose a Bonferroni method as a possible solution. Campbell and Yogo (2006) further utilize the idea of the Bonferroni method by employing the fully augmented OLS (FM-OLS) approach as in Phillips and Hansen (1990). Alternatively, Jansson and Moreira (2006) suggest a conditional likelihood test that uses sufficient statistics in a Gaussian bivariate regression model with a persistent regressor. Elliott (2011) proposes a control function approach to help stabilize the non-standard limits. Phillips and Magdalinos (2009), Kostakis et al. (2015), and Phillips and Lee (2016) have developed an instrumental variable procedure, called IVX, in cointegrating regression framework.

The methods mentioned above require a consistent estimation of the long run variance of errors in the cointegration system. In time series data with an unknown form of serial dependence, it is well known that the estimation of the long run variance is exposed to severe finite-sample noises. As a result, inferences can have a large size distortion in finite-samples, e.g., Kiefer and Vogelsang (2005), Müller (2007), and Sun et al. (2008).

In this paper, we develop robust t and F inferences on the triangular cointegrated regression using a low-frequency transformation approach. To keep it general, we allow the short-run dynamics in the cointegrated system to have serial dependence of unknown forms. By transforming time series from the original time domain, the analysis is carried out the domain of frequencies, such as short-run or long run business cycles. Compared to the existing time-domain approaches, the low-frequency framework enables us to automate the estimation of long run variance parameters in the cointegrating regression.

Following Hwang and Sun (2017), we transform the original non-stationary time series data and its first differences using a K number of low-frequency basis functions. With the K number of low-frequency observations, we run a transformed and augmented ordinary least square (TA-OLS) in cointegrated regression. The triangular cointegrated system is characterized by I(1) regressors in Hwang and Sun (2017), which are endogenous within the structural relation. To account for the order of integration, this paper adopts a local-to-unity approximation of cointegration regressors in the TA-OLS framework. Instead of maintaining a strict dichotomy between integrated and non-integrated regressors, the assumption of the local-to-unity, e.g., Bobkski (1983) and Phillips (1987), allows for a smoother transition between two processes and thus can provide a more reasonable approximation to the TA-OLS methods, especially when the length of the time series is small.

We first derive the fixed-K limiting distributions of TA-OLS and show that the TA-OLS is still super consistent and share a common mixture of the normal distribution. However, due to the local-to-unity regressor, the limits of the TA-OLS estimator have an asymptotic bias term. The asymptotic bias is a product of two important characteristics in our cointegration model: the deviation from the exact unit root and the degree of long run endogeneity within the cointegration system. It is analytically shown that the limiting distributions of TA-OLS statistics are mixtures of non-central t and F distributions where the random non-centrality parameter depends on the asymptotic bias from the local-to-unity regressors. As a result, the standard t and F approximations in Hwang and Sun (2017) are no longer valid asymptotically. This result is consistent with Elliott (1998) whose approximation of the cointegration model is based on the time-domain. Our numerical results also show that the empirical size distortion of the TA-OLS method to test the cointegration vector can be large for even very small deviations from a unit root regressor. On the other hand, we find that the TA-OLS estimator of the long run endogeneity coefficient in the augmented cointegrated system is still asymptotically centered toward its true value.

To make a valid inference for the cointegration vector, we provide modified TA-OLS statistics that correct the asymptotic bias. The modified statistics not only adjust the locational bias but also correct the estimation uncertainty of the long run endogeneity parameter in the bias correction term. After we fully account for both effects on the plugged-in bias correction formula, we show that the modified statistics have the standard t and F limits.

The modified test statistics require the knowledge of the local-to-unity parameter which is not consistently estimable in general. However, there are several ways developed in the time series literature to measure the uncertainty of the local-to-unity parameter in the context of unit root testing problem. See, for example, Stock (1991), Andrews (1993), Elliott and Stock (2001), Mikusheva (2007), and Andrews and Guggenberger (2014) for constructing a confidence interval of the unknown local-to-unity parameter. All these methods, however, except Elliott and Stock (2001), require the autoregressive error to be i.i.d. or martingale difference sequence (m.d.s.), which are limited to be applied in our general cointegration setting. Therefore, we implement Elliott and Stock's (2001) approach which allows the unknown form of serial correlation by inverting a sequence of optimal tests in Gaussian autoregressions.

One concern with Elliott and Stock's (2001) CI is that it is subject to the uniformity critique in Phillips (2014b) when the true local-to-unity parameter largely deviates from zero. On the other hand, the parametric and nonparametric grid-bootstrap methods, which are proposed by Andrews (1993) and Hansen (1999), respectively, do not suffer this drawback (Mikusheva, 2007; Phillips, 2014b). However, the CIs of Andrews (1993) and Hansen (1999) are in danger of poor coverage probabilities when we ignore the serial dependence of the autoregressive error. Thus, we propose a modification of Hansen's (2001) confidence interval, which approximates the unknown dependence structure by a finite-order autoregressive process. Our modification of Hansen (1999) applies the grid-bootstrap method to an (approximated) augmented Dickey-Fuller form with reparametrized autoregressive coefficients. This allows us to overcome the uniformity critique on Elliott and Stock (2001) and construct a CI which is robust to an unknown form of serial dependence.

Using the confidence intervals of the local-to-unity parameter, we develop Bonferroni-based inferences to the modified TA-OLS. By Bonferroni's inequality, our confidence intervals for the cointegration parameter yields asymptotically correct inferences with at least nominal size. The idea of the Bonferroni confidence interval has been widely used in various contexts in statistics and econometrics. See, for example, Cavanagh et al. (1995), Campbell and Yogo (2006), and McCloskey (2017). Recent work by Franchi and Johansen (2017) and Duffy and Simons (2020) also use Bonferroni adjustments when performing inference on cointegrating parameters. Both papers impose parametric vector autoregressive (VAR) structures in the time domain to accommodate dependent cointegration errors. In contrast, our Bonferroni-based inference in frequency domain deals with unknown forms of distribution and dependence structure for cointegration errors, which relates to the literature in the semiparametric estimation of the cointegration system, e.g., Phillips and Hansen (1990), Phillips (1991a&b), and Saikkonen (1991).

Our Monte Carlo results show that the unmodified TA-OLS methods suffer from severe size distortions under the local-to-unity regressor, especially when the long-run endogeneity increases. We further show that the infeasible modified TA-OLS statistic, using the true local-to-unity parameter, successfully controls the size distortions. The feasible versions of the modified TA-OLS, using the Bonferroni-based inferences, have asymptotically correct sizes but are mildly undersized for most of the DGPs we consider. We also show that the use of Hansen's (2001) confidence interval, which ignores dependence structure, in Bonferroni-based inference can drive severe size distortions for testing cointegrating parameters.

In our simulations, we also compare our modified TA-OLS with the IVX test (Phillips and Magdalinos, 2009), which is known to be also robust in the presence of the local-to-unity regressor and serial dependence. We show that the IVX can be size-distorted in finite-samples as the serial dependence of errors increases. This is because the normal critical value in the IVX test statistics does not consider the estimation uncertainty from the nonparametric estimators of the long run variance. The size distortions of the IVX are amplified when the local-to-unity parameter and the degree of long run endogeneity increase. We also find that the IVX test can work the best when there is a low serial correlation in the errors, and the cointegration regressor is not too much deviated from the unit root.

Our paper contributes to recent literature in low-frequency econometrics (Müller and Watson; 2008, 2017). In the context of the cointegrated time series, Phillips (1991) estimates the cointegration parameter using frequency domain techniques, and Bierens (1997) proposes nonparametric tests for the number of cointegrations using a transformed time series. More recently, Phillips (2014a) develops an optimal estimation of cointegration using trend instrumental variables, and Müller and Watson (2013) use the Neyman-Pearson decision-theoretic framework to design robust and nearly optimal tests about the cointegration vectors using a fixed number of transformed data. The approach has also been used in the recent heteroskedasticity and autocorrelation robust inference (HAR) literature for time series models, e.g., Phillips (2005), Müller (2007), and Sun et al. (2008). See also Lazarus et al. (2018) for practical recommendations for HAR inference. In this paper, we develop new t and F inferences which are robust on the triangular cointegrated regression when the economic variables are not exact unit root processes and exhibit unknown form of serial dependence and long run endogeneity. In recent independent work, Sun (2020) also recovers the asymptotic t and F tests of (infeasible) TA-OLS under the local to unity regressor. While Sun (2020) uses a transformed quasi-differenced process of the TA regression, our approach recovers the standard t and F limits by correcting the asymptotic bias of the TA-OLS estimator using the estimated long-run endogeneity coefficient. This way of treating the asymptotic bias of cointegrating regression is similar in spirit to the popular Campbell and Yogo's (2006) Q-test in the predictive regression.

The rest of the paper is organized as follows. Section 2 introduces an idea of low-frequency transformed regression analysis of cointegration and the fixed-K asymptotics limits of the TA-OLS estimator and the corresponding t and F tests. Section 3 extends the low-frequency transformed cointegration system in the presence of a local-to-unity regressor. The next sections provide a method to correct the asymptotic bias of TA-OLS test statistics and suggest feasible Bonferroni-based inferences. Section 6 presents simulation evidence. The last section concludes. Supplemental Appendix provides proofs of the main results, discussions on non-linear and joint testings using the modified TA-OLS, detailed procedures for calculating the confidence sets of the local to unity parameter, and tables and figures referenced in the paper.

2 Low-Frequency Transformation of the Cointegrated System

We start by illustrating the idea of the low-frequency analysis of the triangular cointegration system.[1] We consider the following data generating process:

$$y_t = \alpha_0 + x_t'\beta_0 + u_{0t} \tag{1}$$

$$x_t = x_{t-1} + u_{xt} \text{ for } t = 1, \ldots, T, \tag{2}$$

where y_t is a scalar time series, and x_t is a $d \times 1$ vector of I(1) time series with a stationary innovation u_{xt} such that $x_0 = O_p(1)$. We assume that there exists a cointegrating relation between (y_t, x_t') with cointegrating vector $(1, \beta_0')' \in R^{d+1}$, which is the focus of interest in this paper. To keep it general, we allow the I(0) errors $u_t \equiv (u_{0t}, u_{xt}')' \in R^{d+1}$ to be weakly stationary with serial dependence of unknown forms. Let $\Omega = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}'$ denote long run variance (LRV) matrix of u_t. We partition Ω conformably with $u_t = (u_{0t}, u_{xt}')'$ as

$$\underset{(d+1)\times(d+1)}{\Omega} = \begin{pmatrix} \underset{1\times 1}{\sigma_0^2} & \underset{1\times d}{\sigma_{0x}} \\ \underset{d\times 1}{\sigma_{x0}} & \underset{d\times d}{\Omega_{xx}} \end{pmatrix}.$$

Throughout the paper, we assume that Ω_{xx} is positive definite, and hence x_t is a full-rank integrated process. Then, we can rewrite the cointegrated regression equation in (1) in the following augmented form:

$$y_t = \alpha_0 + x_t'\beta_0 + \delta_0'\Delta x_t + u_{0\cdot xt} \text{ for } t = 1, ..., T, \tag{3}$$

where $\delta_0 = \Omega_{xx}^{-1}\sigma_{x0}$, $\Delta x_t = x_t - x_{t-1}$, and $u_{0\cdot xt} = u_{0t} - \delta_0' u_{xt}$ is a long run projection of u_{0t} onto u_{xt}. Our low-frequency analysis begins with transforming (3) to the following transformed and augmented (TA) regression:

$$\mathbb{W}_{y,i} = \mathbb{W}_{x,i}'\beta_0 + \mathbb{W}_{\Delta x,i}'\delta_0 + \mathbb{W}_{0\cdot x,i} \text{ for } i = 1, ..., K, \tag{4}$$

where $\{\mathbb{W}_{y,i}, \mathbb{W}_{x,i}', \mathbb{W}_{\Delta x,i}'\}_{i=1}^{K}$ is a set of transformed data which is defined as

$$\mathbb{W}_{y,i} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t \phi_i\left(\frac{t}{T}\right), \ \mathbb{W}_{x,i} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \phi_i\left(\frac{t}{T}\right), \ \mathbb{W}_{\Delta x,i} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Delta x_t \phi_i\left(\frac{t}{T}\right), \tag{5}$$

and similarly, we define $\mathbb{W}_{0\cdot x,i} := T^{-1//2}\sum_{t=1}^{T}\phi_i(t/T)u_{0\cdot xt}$.[2] The transformation projects $\{y_t, x_t', \Delta x_t'\}_{t=1}^{T}$ onto a space spanned by K number of orthonormal basis functions, $\{\phi_i(\cdot)\}_{i=1}^{K}$, which can concentrate on the low-frequency components of the original time series data. Examples of the orthonormal

[1] Readers are also referred to Müller and Watson (2017) which overview applications of the low-frequency analysis in other econometrics models.

[2] The intercept terms, α_0, transforms to zero because $T^{-1}\sum_{t=1}^{T}\phi_i(t/T) = 0$ for $i = 1, \ldots, K$.

basis functions include Fourier basis functions considered in Sun (2013, 2014):

$$\left\{\phi_{2j-1}\left(\frac{t}{T}\right)=\sqrt{2}\cos\left(\frac{2j\pi t}{T}\right),\ \phi_{2j}=\sqrt{2}\sin\left(\frac{2j\pi t}{T}\right),\ j=1,\ldots,K/2\right\}, \tag{6}$$

and cosine basis functions in Müller and Watson (2008, 2013):

$$\left\{\phi_{j}\left(\frac{t}{T}\right)=\sqrt{2}\cos\left(\frac{j\pi(t-1/2)}{T}\right),\ j=1,\ldots,K\right\}. \tag{7}$$

The TA regression in (4) has substantive empirical content in the context of the original cointegration system in (1) and (2), which seeks a long-run relationship among economic time series. This is because the transformed data in (5) can effectively capture the long-run behaviors of the original time series (Müller and Watson, 2017). To be more specific, let

$$\Phi_i=[\phi_i(1/T),\ldots,\phi_i((T-1)/T),\phi_i(1)]'\in\mathbb{R}^T$$

denote a basis vector corresponding to the basis functions in (6) and (7), and $\Phi=[l_T,\Phi_1,\ldots,\Phi_K]\in\mathbb{R}^{T\times(K+1)}$ denote a matrix of K basis vectors including the column of ones $l_T=(1,...,1)'\in\mathbb{R}^T$. Because both (6) and (7) satisfy $T^{-1}\sum_{t=1}^{T}\phi_i(t/T)\phi_j(t/T)=1(i=j)$ and $T^{-1}\sum_{t=1}^{T}\phi_i(t/T)=0$, i.e., $\Phi'\Phi=T\cdot I_{K+1}$, the (scaled) low-frequency transformed data are equal to components of the following OLS regression coefficient:

$$(\Phi'\Phi)^{-1}\Phi'X=\frac{\Phi'X}{T}=\left(\bar{x}_T,\breve{\mathbb{W}}_{x,1}',\ldots,\breve{\mathbb{W}}_{x,K}'\right)',$$

where $X=(x_1,\ldots,x_T)'$, $\bar{x}_T=T^{-1}\sum_{j=1}^{T}x_t$, and $\breve{\mathbb{W}}_{x,i}=\mathbb{W}_{x,i}/\sqrt{T}$. Then, the low-frequency movement of the time series can be formulated by the low frequency transformed data multiplied by the non-stochastic trend predictor Φ, i.e.,

$$x_t=\bar{x}_T+\underbrace{\phi_1\left(\frac{t}{T}\right)\breve{\mathbb{W}}_{x,1}+\ldots+\phi_K\left(\frac{t}{T}\right)\breve{\mathbb{W}}_{x,K}}_{\text{low-frequency components}}+\tilde{u}_{xt}. \tag{8}$$

The low-frequency component in (8) captures the long run fluctuation of the original data with periodicity longer than $2T/j$ for $j=1,...,K$ years of cycles. A useful rule of thumb introduced in Müller (2014) and Müller and Watson (2017) suggest a choice of $K=16$ to capture the low-frequency movements of $T=65$ years of Post World War II macro data with periodicity higher than the commonly accepted business cycle period of $T/(K/2)\simeq 8$ years.

To investigate the asymptotic properties of the TA regression system, we assume the following functional central limit theorem (FCLT) for $\{u_t\}$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[T\cdot]}u_t\Rightarrow B(\cdot):=\Omega^{1/2}W(\cdot)=\begin{pmatrix}\sigma_{0\cdot x}w_0(\cdot)+\sigma_{0x}\Omega_{xx}^{-1/2}W_x(\cdot)\\ \Omega_{xx}^{1/2}W_x(\cdot)\end{pmatrix}, \tag{9}$$

where $W(\cdot):=(w_0(\cdot),W_x'(\cdot))'$ is an $d+1$-dimensional standard Brownian process, $\sigma_{0\cdot x}^2=\sigma_0^2-\sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}$, and $\Omega^{1/2}$ is a Cholesky decomposition of the LRV Ω. Primitive conditions to hold the FCLT assumption can be found in Phillips and Durlauf (1986) and Davidson (1994). With (9), we can use summation by parts, continuous mapping theorem, and integration by parts to get

$$\mathbb{W}_{\Delta x,i} \Rightarrow \Omega_{xx}^{1/2} \int_0^1 \phi_i(r)\, dW_x(r) \overset{d}{=} N(0, \Omega_{xx}), \tag{10}$$

$$\mathbb{W}_{0\cdot x,i} \Rightarrow \sigma_{0\cdot x} \int_0^1 \phi_i(r)\, dw_0(r) \overset{d}{=} N(0, \sigma_{0\cdot x}^2) \tag{11}$$

for $i = 1, ..., K$. Also, invoking the continuous mapping theorem together with (9), we have that

$$\frac{\mathbb{W}_{x,i}}{T} = \frac{1}{T^{3/2}} \sum_{s=1}^{T} \phi_i\left(\frac{s}{T}\right) x_s \Rightarrow \Omega_{xx}^{1/2} \int_0^1 \phi_i(r)\, W_x(r) dr \overset{d}{=} N(0, \Omega_{xx}^{1/2} \Sigma \Omega_{xx}^{1/2}), \tag{12}$$

where $\Sigma = \int_0^1 \int_0^1 \phi_i(r)\phi_i(s) \min(r,s) dr ds \cdot I_d$, for $i = 1, ..., K$. Since the weak convergences in (10)–(12) hold jointly, the TA regression in (4) naturally connects to the following small sample Gaussian linear regression model:

$$\mathbb{W}_{y,i} \simeq \mathbb{S}_{x,i}' \beta_{T,0} + \mathbb{S}_{\Delta x,i}' \delta_0 + \mathbb{S}_{0\cdot x,i} \text{ for } i = 1, ..., K, \tag{13}$$

where $\beta_{T,0} = T\beta_0$, $\mathbb{S}_{\Delta x,i}$, $\mathbb{S}_{0\cdot x,i}$, and $\mathbb{S}_{x,i}$ are the Gaussian weak convergence limits of $\mathbb{W}_{\Delta x,i}$, $\mathbb{W}_{0\cdot x,i}$, and $\mathbb{W}_{x,i}/T$, respectively, which are specified in (10), (11), and (12), respectively. Note that $\{\mathbb{S}_{x,i}, \mathbb{S}_{\Delta x,i}\}_{i=1}^K$ and $\{\mathbb{S}_{0\cdot x,i}\}_{i=1}^K$ are independent, because they are functionals of $W_x(\cdot)$ and $w_0(\cdot)$, respectively, which are independent stochastic processes. Also, the orthonormal property of the basis functions $\{\phi_i(\cdot)\}_{i=1}^K$ ensures the errors of regression $\{\mathbb{S}_{0\cdot x,i}\}_{i=1}^K$ are i.i.d. normal with zero mean and variance $\sigma_{0\cdot x}^2$. Therefore, standard OLS framework of the sample Gaussian linear regression model can be applied to estimate the parameters $\beta_{T,0}$ and δ_0.

Hwang and Sun (2017, HS hereafter) runs the OLS estimator for $\gamma_0 = (\beta_0', \delta_0')'$ based on (4) and defines TA-OLS estimator of γ_0 as

$$\hat{\gamma} = (\hat{\beta}', \hat{\delta}')' = (\mathbb{W}_X' \mathbb{W}_X)^{-1} \mathbb{W}_X' \mathbb{W}_y,$$

where $\mathbb{W}_X = (\mathbb{W}_x, \mathbb{W}_{\Delta x})$, $\mathbb{W}_x = (\mathbb{W}_{x,1}, \ldots, \mathbb{W}_{x,K})'$, and $\mathbb{W}_{\Delta x} = (\mathbb{W}_{\Delta x,1}, \ldots, \mathbb{W}_{\Delta x,K})'$. HS shows

$$\hat{\beta} \overset{A}{\sim} N\left[\beta_0, \sigma_{0\cdot x}^2 (\mathbb{W}_x' M_{\Delta x} \mathbb{W}_x)^{-1}\right], \tag{14}$$

and

$$\hat{\delta} \overset{A}{\sim} N\left[\delta_0, \sigma_{0\cdot x}^2 (\mathbb{W}_{\Delta x}' M_x \mathbb{W}_{\Delta x})^{-1}\right], \tag{15}$$

where $M_{\Delta x} = I_K - \mathbb{W}_{\Delta x} \left(\mathbb{W}_{\Delta x}' \mathbb{W}_{\Delta x}\right)^{-1} \mathbb{W}_{\Delta x}'$ and $M_x = I_K - \mathbb{W}_x \left(\mathbb{W}_x' \mathbb{W}_x\right)^{-1} \mathbb{W}_x'$. To test a hypothesis of

$$H_0^\beta : R_\beta \beta_0 = r_\beta \text{ vs. } H_1 : \; R_\beta \beta_0 \neq r_\beta, \tag{16}$$

where R_β is a $p_\beta \times d$ matrix, HS constructs the following (unmodified) Wald statistic and derives its limiting distribution by

$$\begin{aligned} F(\hat{\beta}) &= \frac{1}{\hat{\sigma}_{0\cdot x}^2} (R_\beta \hat{\beta} - r_\beta)' \left[R_\beta (\mathbb{W}_x' M_{\Delta x} \mathbb{W}_x)^{-1} R_\beta'\right]^{-1} (R_\beta \hat{\beta} - r_\beta)/p_\beta \\ &\Rightarrow \frac{K}{K - 2d} \cdot F_{p_\beta, K-2d}, \end{aligned} \tag{17}$$

where $F_{p_\beta, K-2d}$ is the F distribution with degrees of freedom p_β and $K - 2d$. When $p = 1$, the t-statistic can be constructed in a similar manner. Here, $\hat{\sigma}^2_{0\cdot x} = K^{-1}\sum_{i=1}^{K}\hat{\mathbb{W}}^2_{0\cdot x,i}$ is a natural variance estimate of the regression error, where $\hat{\mathbb{W}}_{0\cdot x,i} = \mathbb{W}_{y,i} - \mathbb{W}'_{x,i}\hat{\beta} - \mathbb{W}'_{\Delta x,i}\hat{\delta}$ is a residual of the small sample regression in (13).

The asymptotic variances in (14) and (15) are different with convergence orders, $(\mathbb{W}'_x M_{\Delta x}\mathbb{W}_x)^{-1} = O_p(T^{-2})$ while $(\mathbb{W}'_{\Delta x} M_x \mathbb{W}_{\Delta x})^{-1} = O_p(1)$. The different convergence rates imply different orders of convergence for estimators $\hat{\beta}$ and $\hat{\delta}$ with $T(\hat{\beta}-\beta_0) = O_p(1)$ and $\hat{\delta}-\delta_0 = O_p(1)$, respectively. The latter estimator $\hat{\delta}$ for the long run endogeneity parameter is inconsistent but yields to asymptotically valid t and F tests for

$$H_0^{\delta} : R_\delta \delta_0 = r_\delta \text{ vs. } H_1^{\delta} : \; R_\delta\delta_0 \neq r_\delta, \tag{18}$$

where R_δ is a $p_\delta \times d$ matrix. The corresponding Wald statistic and its limiting distribution are

$$\begin{aligned} F(\hat{\delta}) &= \frac{1}{\hat{\sigma}^2_{0\cdot x}}(R_\delta\hat{\delta} - r_\delta)' \left[R_\delta (\mathbb{W}'_{\Delta x} M_x \mathbb{W}_{\Delta x})^{-1} R_\delta\right]^{-1} (R_\delta\hat{\delta} - r_\delta)/p_\delta \\ &\Rightarrow \frac{K}{K-2d}\cdot F_{p_\delta, K-2d}. \end{aligned} \tag{19}$$

3 Asymptotic Behavior of TA-OLS with Local-to-Unity Regressors

The TA-OLS method is very convenient for practitioners with the standard t and F limits. However, it crucially relies on the exact unit root assumption on the cointegration regressor x_t. Once the cointegration system departs from the unit root assumption, it is questionable whether the the standard t and F tests of the TA cointegration system can be still valid. To answer this, we adopt a local-to-unity approximation of the cointegration regressor

$$x_t = \rho_T x_{t-1} + u_{xt} \text{ and } \rho_T = I_d - \frac{C_0}{T}, \tag{20}$$

where $C_0 = \mathrm{diag}(c_{0,1},\ldots,c_{0,d})$ denotes the local-to-unity coefficients in the regressor vector $x_t = (x_{1t},\ldots,x_{dt})'$. For simplicity of expositions, we assume a common local-to-unity parameter $c_{0,1} = \ldots = c_{0,d} = c_0 \geq 0$ for each components of x_{it}. A generalization to different $c_{0,i}$'s for different x_{it}'s can be made straight and is discussed later in Section 5. When $c_0 = 0$, the regressor x_t is the exact I(1) process. Modeling the cointegration regressor x_t as in (20) allows for a smooth transition between stationary but highly persistent and the "exact" I(1) non-stationary series and provides a more reasonable approximation to the TA cointegration system in (4). This is especially when the length of time series is not enough to identify the exact nature of the auto-regressive root of x_t (Elliott, 1998).

With the local-to-unity approximation of regressor x_t in (20), the differenced process Δx_t becomes

$$\Delta x_t = -\frac{c_0 x_{t-1}}{T} + u_{x,t} \text{ for } t = 1, ..., T.$$

Thus, the low-frequency transformation $\{\mathbb{W}_{\Delta x,i}\}_{i=1}^K$ is no longer the same as $\{\mathbb{W}_{u_x,i}\}_{i=1}^K$ but is now a combination of two transformed data

$$\mathbb{W}_{\Delta x,i} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{xt}\phi_i\left(\frac{t}{T}\right) - c_0\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\frac{x_{t-1}}{T}\right]\phi_i\left(\frac{t}{T}\right) \tag{21}$$

for $i = 1, ..., K$. The augmented cointegration regression in (3) is modified into

$$y_t = \alpha_0 + x_t'\beta_0 + \delta_0'\Delta x_t + \tilde{u}_{0\cdot xt} \text{ for } t = 1, ..., T,$$

where $\tilde{u}_{0\cdot xt} := u_{0\cdot xt} + c_0(\delta_0' x_{t-1}/T)$, which induces the following TA regression model:

$$\mathbb{W}_{y,i} = \mathbb{W}_{x,i}'\beta_0 + \mathbb{W}_{\Delta x,i}'\delta_0 + \tilde{\mathbb{W}}_{0\cdot x,i} \text{ for } i = 1, ..., K, \tag{22}$$

where

$$\tilde{\mathbb{W}}_{0\cdot x,i} := \mathbb{W}_{0\cdot x,i} + \frac{c_0}{T^{3/2}}\left[\sum_{t=1}^{T}\delta_0' x_{t-1}\phi_i\left(\frac{t}{T}\right)\right].$$

When compared to (4), the error term, $\tilde{\mathbb{W}}_{0\cdot x,i}$, includes an additional term, $c_0 T^{-3/2}\sum_{t=1}^{T}\delta_0' x_{t-1}\phi_i(t/T)$, which is a (scaled) low frequency transformation of x_{t-1}. Because of this extra term, we cannot guarantee the standard t and F limits, as in (17) and (19), under the local-to-unity. To formally establish the asymptotic properties of the TA-OLS estimator $\hat{\gamma} = (\hat{\beta}', \hat{\delta}')'$, we make the following assumptions.

Assumption 1 *The vector process $\{u_t = (u_{0t}, u_{xt}')'\}_{t=1}^{T}$ satisfies the FCLT in (9).*

Assumption 2 *(i) For $i = 1, \ldots, K$, each function $\phi_i(\cdot)$ is continuously differentiable; (ii) For $i = 1, \ldots, K$, each function $\phi_i(\cdot)$ satisfies $\int_0^1 \phi_i(x)\,dx = 0$; (iii) The functions $\{\phi_i(\cdot)\}_{i=1}^{K}$ are orthonormal in $L^2[0,1]$.*

Along with the local-to-unity regressors in (20), Assumption 1 of FCLT enables us to invoke the result in Phillips (1987) and get

$$\frac{1}{\sqrt{T}}x_{[Tr]} \Rightarrow \Omega_{xx}^{1/2}J_{c_0}(r), \tag{23}$$

where the Ornstein-Uhlenbeck (OU) process is defined by $J_{c_0}(r) = \int_0^r \exp(-c_0(r-s))dW_x(s)$. Since Assumption 2 holds in both (6) and (7), we can repeat the weak convergence approximations in (12) under the local-to-unity assumption in (20) as below:

$$\frac{\mathbb{W}_{x,i}}{T} \Rightarrow \Omega_{xx}^{1/2}\int_0^1 \phi_i(r)\,J_{c_0}(r)dr \overset{d}{=} N\left(0, \Omega_{xx}^{1/2}\Sigma_{c_0}\Omega_{xx}^{1/2}\right), \tag{24}$$

where $\Sigma_{c_0} = \frac{1}{2c_0}\int_0^1\int_0^1 \phi_i(r)\phi_i(s)\{\exp[-c_0|r-s|] - \exp[-c_0(r+s)]drds \cdot I_d$, for $i = 1, ..., K$. The above weak convergence shows that the local-to-unity assumption does not change the Gaussian limits but has a different asymptotic variance from (11). In the proof of Proposition 1, we show that

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\phi_i\left(\frac{t}{T}\right) = \frac{\mathbb{W}_{x,i}}{T} + O_p\left(\frac{1}{T}\right).$$

Thus, the transformed first difference $\mathbb{W}_{\Delta x,i}$ and the regression error $\tilde{\mathbb{W}}_{0\cdot x,i}$ have the following weak convergence limits of

$$\mathbb{W}_{\Delta x,i} \Rightarrow \Omega_{xx}^{1/2}\left[\int_0^1 \phi_i(r)dW_x(r) - c_0\cdot\int_0^1 \phi_i(r)J_{c_0}(r)dr\right], \tag{25}$$

$$\tilde{\mathbb{W}}_{0\cdot x,i} \Rightarrow \sigma_{0\cdot x}\int_0^1 \phi_i(r)dw_0(r) + c_0\cdot\left[\Omega_{xx}^{1/2}\delta_0\right]'\int_0^1 \phi_i(r)J_{c_0}(r)dr$$

for $i = 1, ...K$, respectively. Combining these results, the TA regression in (22) is now asymptotically equivalent to:

$$\mathbb{W}_{y,i} \simeq \mathbb{S}_{x,i}'\beta_{T,0} + \mathbb{S}_{\Delta x,i}'\delta_0 + \left[\mathbb{S}_{0\cdot x,i} + c\delta_0'\mathbb{S}_{x,i}\right] \text{ for } i = 1, ..., K,$$

where $\mathbb{S}_{x,i}$, $\mathbb{S}_{\Delta x,i}$, and $\mathbb{S}_{0\cdot x,i}$ are the Gaussian random limits of $\mathbb{W}_{x,i}/T$, $\mathbb{W}_{\Delta x,i}$, and $\mathbb{W}_{0\cdot x,i}$, respectively, which are specified in (24), (25), and (11), respectively. Then, the asymptotic behavior of the TA-OLS estimator is captured by

$$\begin{aligned} T(\hat{\beta} - \beta_0) &= \left[\frac{\mathbb{W}_x'}{T}(I_K - P_{\Delta x})\frac{\mathbb{W}_x}{T}\right]^{-1}\left[\frac{\mathbb{W}_x'}{T}(I_K - P_{\Delta x})\tilde{\mathbb{W}}_{0\cdot x}\right] \\ &\Rightarrow \left[\mathbb{S}_x'(I_K - P_{\mathbb{S}_{\Delta x}})\mathbb{S}_x\right]^{-1}\mathbb{S}_x'(I_K - P_{\mathbb{S}_{\Delta x}})\mathbb{S}_{0\cdot x} + c_0\delta_0, \end{aligned}$$

where $P_{\mathbb{S}_{\Delta x}} = \mathbb{S}_{\Delta x}(\mathbb{S}_{\Delta x}'\mathbb{S}_{\Delta x})^{-1}\mathbb{S}_{\Delta x}'$. Conditioning on $\mathbb{S}_x$ and $\mathbb{S}_{\Delta x}$, the first majorant term characterizes the weak Gaussian limit of TA-OLS estimator under the unit root regressors which is centered toward the true parameter β_0. This limit is the same as what is derived under the exact unit root regressor in HS, except for the covariance structure of the conditioning random variables $\mathbb{S}_x$ and $\mathbb{S}_{\Delta x}$. The second term $c_0\delta_0$ indicates that the asymptotic distribution of $\hat{\beta}$ possesses a bias term $c_0\delta_0$. We formally state the weak convergences result of TA-OLS estimator including $\hat{\delta}$ in the following Proposition. Define

$$\Upsilon_T = \begin{pmatrix} T \cdot I_d & \underset{d\times d}{0} \\ \underset{d\times d}{0} & I_d \end{pmatrix}. \tag{26}$$

Proposition 1 *Let $\mathbb{S}_X = [\mathbb{S}_x', \mathbb{S}_{\Delta x}']'$. Under Assumptions 1 and 2, the local-to-unity regressors in (20), and as $T \to \infty$ but holding K fixed, we have that*

$$\Upsilon_T(\hat{\gamma} - \gamma_0) = \begin{pmatrix} T(\hat{\beta} - \beta_0) \\ \hat{\delta} - \delta_0 \end{pmatrix} \Rightarrow \begin{bmatrix} c_0\delta_0 \\ 0 \end{bmatrix} + MN(0, \sigma_{0\cdot x}^2(\mathbb{S}_X'\mathbb{S}_X)^{-1}).$$

From the result of Proposition 1, we have that

$$\begin{aligned} T(\hat{\beta} - \beta_0) &\Rightarrow MN\left[c_0\delta_0, \sigma_{0\cdot x}^2(\mathbb{S}_x'M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x)^{-1}\right], \\ \hat{\delta} - \delta_0 &\Rightarrow MN\left[0, \sigma_{0\cdot x}^2(\mathbb{S}_{\Delta x}'M_{\mathbb{S}_x}\mathbb{S}_{\Delta x})^{-1}\right], \end{aligned}$$

where the convergences hold jointly. Since the local-to-unity regressor affects the limit behavior of $\hat{\beta}$ by shifting the center of the weak limit $T(\hat{\beta} - \beta_0)$ from zero to the asymptotic bias term $c\delta_0$, the TA-OLS estimator $\hat{\beta}$ is asymptotically unbiased only if i) the regressors have the exact unit root processes, i.e. $c_0 = 0$, or ii) there is no long run simultaneity between u_t and u_{xt}, i.e. $\delta_0 = 0$. Both of these cases, however, are unlikely to show up in practice. The results are similar with Elliott (1998) which finds the fragility of time-domain cointegration inference in the presence of local-to-unity regressors. Our work also shows that the same asymptotic bias terms appear in the domain of low frequencies.

The limiting distribution of the cointegration vector is affected by the local-to-unity regressor. However, the result in Proposition 1 indicates that $\hat{\delta}$ is still asymptotically centered toward δ_0 and has the same asymptotic behavior as the case of exact unit root regressors. Under the null hypotheses in (16) and (18), these results lead to

$$\begin{aligned} T(R_\beta\hat{\beta} - r_\beta) &\Rightarrow MN(c_0R_\beta\delta_0, \sigma_{0\cdot x}^2\left[R_\beta(\mathbb{S}_x'M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x)^{-1}R_\beta'\right]), \\ R_\delta\hat{\delta} - r_\delta &\Rightarrow MN(0, \sigma_{0\cdot x}^2\left[R_\delta(\mathbb{S}_{\Delta x}'M_{\mathbb{S}_x}\mathbb{S}_{\Delta x})^{-1}R_\delta'\right]). \end{aligned} \tag{27}$$

In view of the joint weak convergence results in (24) and (25), it is easy to check

$$\begin{aligned} R_\beta \left[(\mathbb{W}_x'/T) M_{\Delta x} (\mathbb{W}_x'/T)\right]^{-1} R_\beta' &\Rightarrow R_\beta \left[\mathbb{S}_x' M_{\mathbb{S}_{\Delta x}} \mathbb{S}_x\right]^{-1} R_\beta', \\ R_\delta (\mathbb{W}_{\Delta x}' M_x \mathbb{W}_{\Delta x})^{-1} R_\delta' &\Rightarrow R_\delta \left[\mathbb{S}_{\Delta x}' M_{\mathbb{S}_x} \mathbb{S}_{\Delta x}\right]^{-1} R_\delta'. \end{aligned} \tag{28}$$

Thus, if one finds an asymptotic behavior of $\hat{\sigma}^2_{0\cdot x}$ under the near-unity regressor in (20), we are able to find a weak limit of Wald and t statistics for the parameters $\gamma = (\beta_0', \delta_0')$. The results are summarized in the following Proposition.

Proposition 2 *Let Assumptions 1 and 2, and the null hypotheses in (16)-(18) hold . Define* $\theta = c_0[R_\beta \left[\mathbb{S}_x' M_{\mathbb{S}_{\Delta x}} \mathbb{S}_x\right]^{-1} R_\beta']^{-1/2}(R_\beta \delta_0/\sigma_{0\cdot x})$*. Then, under the fixed-K asymptotics, we have*

(a) $F(\hat{\beta}) \Rightarrow \frac{K}{K-2d} \cdot F_{p_\beta, K-2d}(||\theta||^2)$;
(b) $t(\hat{\beta}) \Rightarrow \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}(\theta)$;
(c) $F(\hat{\delta}) \Rightarrow \frac{K}{K-2d} \cdot F_{p_\delta, K-2d}$;
(d) $t(\hat{\delta}) \Rightarrow \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}$.

In the proof of Proposition 2, we show that the asymptotic variance estimate $\hat{\sigma}^2_{0\cdot x}$ for the long run projected variance $\sigma^2_{0\cdot x}$ weakly converges to χ^2_{K-2d} limiting distribution. Since all other components of test statistics, except the bias term $c_0 R_\beta \delta_0$, behave the same way as in the case of the exact unit root regressors, we can capture the effect of the local-to-unity regressors on the hypothesis tests of β_0 only by looking at the random non-centrality parameter $||\theta||^2$ in the limiting F and t distributions. Let $r^2 = (\sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0})/\sigma_0$ denote a squared long run correlation vector between $\{u_{0t}\}$ and $\{u_{xt}\}$. When $d = p_\beta = 1$, a simple algebra can show that the non-random part of $\|\theta\|^2$ is equal to $c_0^2 \cdot r^2/(1-r^2)$, which implies that the null rejection rate for TAOLS t-test approaches to one, as the squared long run correlation r^2 gets close to one.

The presence of non-zero $\|\theta\|^2$ implies that the hypothesis test using the Wald statistics in (17) will tend to over-reject. However, the results in Proposition 2 (c) and (d) indicate we can still perform asymptotically valid Wald and t tests about the long run endogeneity coefficient δ_0. This is expected from our previous investigation on the limit behavior of $\hat{\delta}$, which is not affected by the local-to-unity regressors. These theoretical implications are numerically supported in the Monte Carlo simulation in Section 6.

4 Bias-corrected Inferences for β_0

In this section, we provide a method to correct the asymptotic bias of TA-OLS test statistics for β_0. The modification not only adjusts the asymptotic locational bias of the TA-OLS estimator, but also fully accounts for the estimation uncertainties embodied in the bias correction term. Let $\Gamma_{c_0} = (R_\beta, -c_0 R_\beta)$ be a $p \times 2d$ matrix formed by the hypothesis matrix R_β and the local-to-unity parameter c_0. Then, under $H_0^\beta : R_\beta \beta_0 = r_\beta$,

$$\begin{aligned} \Gamma_{c_0} \Upsilon_T \left[\hat{\gamma} - \gamma_0\right] &= \begin{pmatrix} R_\beta & -c_0 R_\beta \end{pmatrix} \begin{pmatrix} T(\hat{\beta} - \beta_0) \\ \hat{\delta} - \delta_0 \end{pmatrix} \\ &= T\left[R_\beta(\hat{\beta} - c_0 \cdot \hat{\delta}/T) - r_\beta\right] + c_0 R_\beta \delta_0. \end{aligned} \tag{29}$$

Using the joint convergence result in Proposition 1 and continuous mapping theorem, we have that

$$\begin{aligned}\Gamma_{c_0}\Upsilon_T\left[\hat{\gamma}-\gamma_0\right] &= T(R_\beta(\hat{\beta}-c_0\cdot\hat{\delta}/T)-r_\beta)+c_0R_\beta\delta_0 \\ &\Rightarrow \Gamma_{c_0}\left[\begin{array}{c} c_0\delta_0 \\ 0 \end{array}\right]+\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x} \\ &\overset{d}{=} MN\left(c_0R_\beta\delta_0, \sigma_{0\cdot x}^2\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'\right).\end{aligned} \tag{30}$$

Therefore, the plugged-in estimator of $\hat{\beta}-c_0(\hat{\delta}/T)$ is able to correct the bias of $c_0(\delta_0/T)$ in the limiting distribution of $T(\hat{\beta}-\beta_0)$, because of

$$T(R_\beta(\hat{\beta}-c_0\cdot\hat{\delta}/T)-r_\beta)\Rightarrow MN\left(0,\sigma_{0\cdot x}^2\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'\right). \tag{31}$$

It is essential to point out the asymptotic variance of the plugged-in estimator $\hat{\beta}-c_0(\hat{\delta}/T)$ is no longer the same as that of $\hat{\beta}$ in (27). This is because the asymptotic variance of the plugged-in estimator has to reflect the estimation uncertainty of $\hat{\delta}$ in its limiting distribution. This motivates us to construct the following modified Wald statistic:

$$\begin{aligned}F(\hat{\beta};c_0) = & \frac{T^2}{\hat{\sigma}_{0\cdot x}^2}(R_\beta[\hat{\beta}-c_0\cdot(\hat{\delta}/T)]-r_\beta)'\left[\Gamma_{c_0}(\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1})^{-1}\Gamma_{c_0}'\right]^{-1} \\ & \times(R_\beta[\hat{\beta}-c_0\cdot\hat{\delta}/T]-r_\beta)/p.\end{aligned} \tag{32}$$

When $p=1$, we construct the modified t statistic for one-sided alternative as below:

$$t(\hat{\beta};c_0)=\frac{T(R_\beta[\hat{\beta}-c_0\cdot\hat{\delta}/T]-r_\beta)}{\sqrt{\hat{\sigma}_{0\cdot x}^2\Gamma_{c_0}(\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1})^{-1}\Gamma_{c_0}'}}. \tag{33}$$

Note that the estimations of δ_0 and $\sigma_{0\cdot x}^2$, which are necessary for the modified TA-OLS test statistics, are automated in our TA-OLS framework. In fact, practitioners just need to run a classical OLS regression with the transformed data $\{\mathbb{W}_{y,i},\mathbb{W}_{x,i}',\mathbb{W}_{\Delta x,i}'\}_{i=1}^K$ and obtain $\hat{\beta},\hat{\delta}$, and $\hat{\sigma}_{0\cdot x}^2$ at once. The theorem below establishes the limiting null distributions of $F(\hat{\beta};c_0)$ and $t(\hat{\beta};c_0)$ under the fixed-K asymptotics.

Theorem 3 *Under Assumptions 1 and 2, and $T\to\infty$ but holding K fixed, we have that*

$$F(\hat{\beta};c_0)\Rightarrow\frac{K}{K-2d}\cdot F_{p,K-2d} \text{ and } t(\hat{\beta};c_0)\Rightarrow\sqrt{\frac{K}{K-2d}}\cdot t_{K-2d}.$$

Theorem 3 indicate one can construct valid t and F tests using the modified t and Wald statistics. The modified statistics not only adjust the locational bias but also reflect the estimation uncertainty of the $\hat{\delta}$ in the bias correction term. After we fully account the effect of the plugged-in bias correction $c_0(\hat{\delta}/T)$ on the modified statistics, we obtain the exact same asymptotic F and t limits. Note that the resulting F and t limits also take into account the estimation uncertainties for the long run variance term $\sigma_{0\cdot x}^2$. Practically, the result in Theorem 3 implies that one can conveniently implement the modified test statistics, $F(\hat{\beta};c_0)$ and $t(\hat{\beta};c_0)$, using the standard t and F testing methods.

When $p_\beta=1$, Theorem 3 shows a valid $100(1-\alpha)\%$ confidence interval (CI) for the testing parameter $R\beta_0$ can be constructed as

$$CI_{R\beta_0}\left(c_0; 1-\alpha\right) = \left[r_{\beta,l}^{1-\alpha/2}(c_0), r_{\beta,h}^{1-\alpha/2}(c_0)\right], \tag{34}$$

where

$$r_{\beta,l}^{1-\alpha/2}(c_0) = R_\beta\left[\hat{\beta} - \frac{c_0\hat{\delta}}{T}\right] - \frac{1}{T}\sqrt{\hat{\sigma}_{0\cdot x}^2 \Gamma_{c_0}\left[\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1}\right]^{-1}\Gamma_{c_0}'} \cdot \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}^{1-\alpha/2}, \tag{35}$$

$$r_{\beta,h}^{1-\alpha/2}(c_0) = R_\beta\left[\hat{\beta} - \frac{c_0\hat{\delta}}{T}\right] + \frac{1}{T}\sqrt{\hat{\sigma}_{0\cdot x}^2 \Gamma_{c_0}\left[\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1}\right]^{-1}\Gamma_{c_0}'} \cdot \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}^{1-\alpha/2}, \tag{36}$$

and $t_{K-2d}^{1-\alpha/2}$ is the $1-\alpha/2$ quantile from the $t_{p,K-2d}$ distribution. With the nearly integrated regressors, the modified CI in (34) shifts the location of the interval up to $-c_0(R\hat{\delta}/T)$. With the location adjustment $-c_0(R\hat{\delta}/T)$, one may come up with the following CI:

$$R_\beta\left[\hat{\beta} - \frac{c_0\hat{\delta}}{T}\right] \pm \frac{1}{T}\sqrt{\hat{\sigma}_{0\cdot x}^2 R_\beta(\mathbb{W}_X' M_{\Delta x}\mathbb{W}_X)^{-1}R_\beta'} \cdot \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}^{1-\alpha/2}. \tag{37}$$

The common critical value $t_{K-2d}^{1-\alpha/2}$ and estimated variance terms $\hat{\sigma}_{0\cdot x}^2$ reflect the uncertainty of time series in the (un)modified confidence intervals, but there is notable difference in the margin of errors of two confidence intervals between (34) and (37). With some additional algebra, we can express the term in (34) by

$$\Gamma_{c_0}\left[\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1}\right]^{-1}\Gamma_{c_0}' = R_\beta\left[\Lambda_1(c_0)\left(\mathbb{W}_x' M_{\Delta x}\mathbb{W}_x\right)^{-1} + \Lambda_2(c_0)\left(\mathbb{W}_{\Delta x}' M_x\mathbb{W}_{\Delta x}\right)^{-1}\right]R_\beta', \tag{38}$$

where

$$\Lambda_1(c_0) = T^2(I_d + c_0 T^{-1}\left[\mathbb{W}_{\Delta x}'\mathbb{W}_{\Delta x}\right]^{-1}\mathbb{W}_{\Delta x}'\mathbb{W}_x),$$
$$\Lambda_2(c_0) = c_0^2 I_d + c_0 T\left[\mathbb{W}_x'\mathbb{W}_x\right]^{-1}\mathbb{W}_x'\mathbb{W}_{\Delta x}.$$

That is, the measure of uncertainty in (34) is a weighted average of standard error terms for $\hat{\beta}$ and $\hat{\delta}$, where weights are given by $\Lambda_1(c_0) = O_p(T^2)$ and $\Lambda_2(c_0) = O_p(1)$, respectively. The relative difference in the order of magnitude between these weights is based on the different convergence rates of the variance estimates $(\mathbb{W}_x' M_{\Delta x}\mathbb{W}_x)^{-1} = O_p(T^{-2})$ and $(\mathbb{W}_{\Delta x}' M_x\mathbb{W}_{\Delta x})^{-1} = O_p(1)$ for the estimators $\hat{\beta}$ and $\hat{\delta}$, respectively. Interestingly, the weights are functions of the OLS coefficients from the two transformed regressors, $\mathbb{W}_x$ and $\mathbb{W}_{\Delta x}$, and the local-to-unity parameter c_0.

Readers are referred to Section S.4 of Supplementary Appendix, which applies the bias-corrected inference of modified TA-OLS to test simultaneous restrictions on β_0 and δ_0 and discuss a non-linear testing hypothesis.

When $c_0 = 0$, i.e. the regressor x_t has an exact unit root, it is easy to check that the above CI of β_0 reduces to the standard form of symmetric CI,

$$R_\beta\hat{\beta} \pm \sqrt{\hat{\sigma}_{0\cdot x}^2 R_\beta(\mathbb{W}_X' M_{\Delta x}\mathbb{W}_X)^{-1}R_\beta'} \cdot \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}^{1-\alpha/2},$$

which is same as the TA-OLS tests in Section 2.

In the standard time domain framework, one can show that a popular endogeneity bias correct method such as fully modified (FM)-OLS estimator, e.g., Phillips and Hansen (1990), yields

$$T(\hat{\beta}_{\text{FM}} - \beta_0) \Rightarrow MN\left(c_0\delta_0, \sigma^2_{0\cdot x}\left[\int_0^1 B_x(r)B'_x(r)dr\right]^{-1}\right)$$

under the local-to-unity assumption (20). Using this result, Campbell and Yogo (2006) provides the feasible bias-corrected estimation and inference based on

$$\frac{T(\hat{\beta}_{\text{FM}} - c_0\ \hat{\delta}_{\text{HAC}} - \beta_0)}{\sqrt{\hat{\sigma}^2_{0\cdot x,\text{HAC}}\left[\frac{1}{T^2}\sum_{t=1}^T x_t x'_t\right]}} \Rightarrow N(0,1).$$

A key step behind the Campbell and Yogo's (2006) method is heteroscedasiticy autocorrelation consistent (HAC) estimators, including $\hat{\delta}_{\text{HAC}}$ and $\hat{\sigma}^2_{0\cdot x,,\text{HAC}}$, and the asymptotic normal critical value for the test statistics. In our cointegration setting, δ_0 and $\sigma^2_{0\cdot x}$ are functions of the long run variance matrix Ω. However, it is well known that the consistent HAC approach, e.g., Newey and West (1987), is exposed to severe finite-sample noises in time series data, e.g., Kiefer and Vogelsang (2005), Müller (2007), and Sun et al. (2008). The severity of these issues has been shown through Monte Carlo simulation in HS. The low-frequency transformed TA-OLS framework in this paper explicitly avoids these issues, as it does not need the separated estimation step for δ_0 and $\sigma^2_{0\cdot x}$. Moreover, the result in Theorem 3 that our modified TA-OLS and corresponding t and F limits successfully account for the finite-sample uncertainties embodied in $\hat{\sigma}^2_{0\cdot x}$ and $\hat{\delta}\ _0$.

5 Bonferroni-based Inference for Modified TA-OLS

5.1 Robust confidence intervals for c_0

The near-unity approximation of the modified test statistics, $F(\hat{\beta}; c_0)$ and $t(\hat{\beta}; c_0)$, requires the knowledge of the true local-to-unity parameter, c_0, which is not consistently estimable in general. However, one can construct a nontrivial and informative CI for the unknown parameter c_0 using several methods developed in the literature. Examples include Stock (1991), Andrews (1993), Hansen (1999), Elliott and Stock (2001), Mikusheva's (2007) modification of Stock (1991), and Andrews and Guggenberger (2014). However, all these methods, except Elliott and Stock (2001), have limitations to apply in (20) because they restrict the error process, $\{u_{xt}\}$, to be i.i.d. or martingale difference sequence. Therefore, we construct the CI as in Elliott and Stock (2001), which allows for an unknown form of dependence in $\{u_{xt}\}$, by inverting a sequence of asymptotically optimal tests in the Gaussian autoregressive model. In Section S.3 of Supplemental Appendix, we provide a detailed procedure for constructing the CI in Elliott and Stock (2001).

One concern with Elliott and Stock's (2001) method is that its CI is subject to the uniformity critique raised in Phillips (2014b) when the true local to unity parameter c_0 largely deviates from zero, thereby having a poor coverage rate.[3] Such drawback of Elliott and Stock (2001) can be addressed by using parametric and nonparametric grid-bootstrap methods, which are proposed by Andrews (1993) and Hansen (1999), respectively. This is because the CIs in the latter two methods are constructed by using the centered statistics on different null values of parameters in the grids, which is crucial for achieving the uniformity (Mikusheva, 2007).

However, the CIs in Andrews (1993) and Hansen (1999) are subject to severe undercoverage bias when we ignore the serial dependence of $\{u_{xt}\}$. Thus, we propose a modified version of the

[3] We thank an anonymous referee who pointed out this issue.

CI of Hansen (1999), which addresses both the bias arising from the serial dependence of and the uniformity issue on Elliott and Stock (2001). The modification of Hansen (1999) proceeds as follows. We first approximate the unknown dependence structure in $\{u_{xt}\}$ by a finite-order autoregressive process and translate (20) to an (approximated) augmented Dickey-Fuller (ADF) form. We then apply the grid-bootstrap method in Hansen (1999) to the ADF equation and construct a CI for c_0 using a reparameterization of autoregressive coefficients. The detailed implementation of this procedure is provided in Section S.3 of Supplemental Appendix.

We present some simulation evidence of the CIs that are discussed above. We generate data from (20) with $T = 200$, where u_{xt} is drawn from AR(1) and MA(1) processes, and construct CIs of Hansen (1999), Elliott and Stock (2001), and our modification of Hansen (1999) with 90% nominal coverage rate. Tables S.1 and S.2 in Section S.5 of Supplemental Appendix report the empirical coverage rates and average estimates of CIs under different degrees of dependence for the autoregressive errors and the true local-to-unity parameter, c_0. The results are summarized as below.

When c_0 is close to zero, e.g., $c_0 = 5$, the method in Elliott and Stock (2001) yields accurate and shorter CIs than other methods. However, the CIs in Elliott and Stock (2001) suffers undercoverage biases in Table S.2, varying from 53.1% to 74.7%, when c_0 grows to 20. The CI of Hansen (1999) shows accurate coverage rates for all ranges of c_0's if there is no serial dependence, i.e., $\psi = 0$. However, it is prone to poor coverage rates when ψ is non-zero. For instance, Table S.2 shows that the CI of Hansen (1999) has almost zero coverages when ψ for AR(1) error grows to 0.75. On the other hand, our modified CI of Hansen (1999) shows more accurate coverage rates for all ranges of c_0's. For example, when $c_0 = 20$, the coverage rates of modified CI of Hansen (1999), varying from 75.7% to 85.6%, significantly improve those of Elliott and Stock (2001).

In summary, we propose implementing the two methods, Elliott and Stock (2001) and the modified version of Hansen (1999), to construct robust confidence intervals in the presence of an unknown form of serial dependence. When the true autoregressive parameter is close to the unity, i.e., $c_0 \approx 0$, we check that the CI of Elliott and Stock (2001) performs well in terms of its coverage and length. When c_0 is large, however, it suffers poor coverage rates due to the lack of uniformity. We also find that our modified CI of Hansen (1999), which is robust to the serial dependence, addresses the uniformity issue on Elliott and Stock (2001) and improves the coverage probability of the CI in Elliott and Stock (2001).

5.2 Bonferroni-based confidence intervals for β_0

Let $S_T(\eta_1)$ denote a CI for c_0 with $100(1-\eta_1)\%$ asymptotic coverage rate. With $p_\beta = 1$, which is of the utmost importance in empirical research, we can construct a Bonferroni CI for $R_\beta \beta_0$ as

$$CI^{\mathrm{B}}_{R_\beta \beta_0}(\eta_1, \eta_2) = \bigcup_{c \in S_T(\eta_1)} CI_{R_\beta \beta_0}(c; 1-\eta_2) \tag{39}$$

$$= \left[\min_{c \in S_T(\eta_1)} r^{1-\eta_2/2}_{\beta,l}(c), \max_{c \in S_T(\eta_1)} r^{1-\eta_2/2}_{\beta,h}(c) \right], \tag{40}$$

where $\eta_1, \eta_2 \geq 0$ such that $\eta_1 + \eta_2 = \alpha$, and $r^{1-\eta_2/2}_{\beta,l}(c)$ and $r^{1-\eta_2/2}_{\beta,h}(c)$ are defined in (35) and (36), respectively. The idea of Bonferroni-based inference has been used in various contexts in statistics and econometrics, e.g., Cavanagh et al. (1995), Campbell and Yogo (2006), and McCloskey (2017). By Bonferroni's inequality, the above Bonferroni CI yields an asymptotic coverage rate at least $100(1-\alpha)\%$, i.e.,

$$\liminf_{T \to \infty} P\left[R_\beta \beta_0 \in CI^{\mathrm{B}}_{R_\beta \beta_0}(\eta_1, \eta_2) \right] \geq 1 - \alpha. \tag{41}$$

The infeasible CI, $[r_{\beta,l}^{1-\eta_2/2}(c), r_{\beta,h}^{1-\eta_2/2}(c)]$, depends on c only through $T^{-1}c\hat{\delta}$ and Γ_c. Thus, the computational cost of finding (40) is not too high when we search for the maximum and minimum of $r_{\beta,h}^{1-\eta_2/2}(c)$ and $r_{\beta,l}^{1-\eta_2/2}(c)$, respectively, over $S_T(\eta_1)$. In Section S.4 of Supplemental Appendix, we characterize some conditions when the lower (or upper) bound is monotone in c, making the minimization (or maximization) much simpler when computing our Bonferroni CIs.

As an alternative to the Bonferroni method in (39), we can consider Bonferroni critical values by taking maximum and minimum over the asymptotic critical values of the unmodified $t(\hat{\beta})$, as a function of c, e.g., McCloskey (2017). The result in Proposition 2-(b) shows that $t(\hat{\beta})$ has the mixed non-central t limit, $t_{K-2d}(\theta)$, where the random non-centrality parameter θ depends on c_0, δ_0, and $\sigma_{0\cdot x}^2$. This leads us to formulate the following Bonferroni critical values:

$$\min_{c\in S_T(\eta_1)}\left(\sqrt{\frac{K}{K-2d}}\cdot t_{K-2d}^{\eta_2/2}(\hat{\theta}(c))\right) \text{ and } \max_{c\in S_T(\eta_1)}\left(\sqrt{\frac{K}{K-2d}}\cdot t_{K-2d}^{1-\eta_2/2}(\hat{\theta}(c))\right),$$

where

$$\hat{\theta}(c) = \frac{c\cdot R_\beta\hat{\delta}_{\mathrm{HAC}}}{\hat{\sigma}_{0\cdot x,\mathrm{HAC}}\sqrt{R_\beta\left[(\mathbb{W}_x'/T)M_{\Delta x}(\mathbb{W}_x'/T)\right]^{-1}R_\beta'}}.$$

The corrected critical value via $\hat{\theta}(c)$ can be exposed to the estimation uncertainty of the nonparametric HAC estimators, $\hat{\delta}_{\mathrm{HAC}}$ and $\hat{\sigma}_{0\cdot x,\mathrm{HAC}}$. In contrast, the Bonferroni steps in (39) and (40) avoid the estimation issues involved in $\hat{\delta}_{\mathrm{HAC}}$ and $\hat{\sigma}_{0\cdot x,\mathrm{HAC}}$ because they make uses of the modified TA-OLS statistics, which induce the Bonferroni CI in (35) and (36).[4]

When we allow different $c_{0,i}$'s for different cointegration regressors x_{it}, the presence of multi-dimensional nuisance parameters is a potential challenge in calculating our Bonferroni intervals. However, one can follow the general Bonferroni principle with a joint confidence set, $S_T(\eta_1)\in\mathbb{R}^d$, which gives an asymptotically correct coverage for the true local-to-unity parameters, i.e.

$$\liminf_{T\to\infty} P(\mathbf{c}_0\in S_T(\eta_1))\geq 1-\eta_1,$$

where $\mathbf{c}_0 = (c_{0,1}, c_{0,2},\ldots,c_{0,d})'$ is the true local-to-unity parameters. Then, the corresponding Bonferroni CI for testing $H_0^\beta : R_\beta\beta_0$ is:

$$\begin{aligned} CI_{R\beta_0}^{\mathrm{B}}(\eta_1,\eta_2) &= \bigcup_{\mathbf{c}=(c_1,\ldots,c_d)\in S_T(\eta_1)} CI_{R\beta_0}(\mathbf{c};1-\eta_2) \\ &= \left[\min_{\mathbf{c}\in S_T(\eta_1)} r_{\beta,l}^{1-\eta_2/2}(\mathbf{c}),\ \max_{\mathbf{c}\in S_T(\eta_1)} r_{\beta,h}^{1-\eta_2/2}(\mathbf{c})\right], \end{aligned} \tag{42}$$

where $[r_{\beta,l}^{1-\eta_2/2}(\mathbf{c}), r_{\beta,h}^{1-\eta_2/2}(\mathbf{c})]$ is a generalized version of the bias-corrected CI in (34), which is defined as

$$\begin{aligned} r_{\beta,l}^{1-\eta_2/2}(\mathbf{c}) := &\left[R_\beta\hat{\beta} - \frac{R_\beta\mathrm{diag}(\mathbf{c})\hat{\delta}}{T}\right] \\ &- \frac{1}{T}\sqrt{\hat{\sigma}_{0\cdot x}^2\Gamma_{\mathbf{c}}\left[\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1}\right]^{-1}\Gamma_{\mathbf{c}}'}\cdot\sqrt{\frac{K}{K-2d}}\cdot t_{K-2d}^{1-\eta_2/2}, \end{aligned}$$

[4] We thank an anonymous referee who motivated us to clarify this subtlety in Bonferroni-based inferences using TA-OLS.

and

$$r_{\beta,h}^{1-\eta_2/2}(\mathbf{c}) := \left[R_\beta\hat{\beta} - \frac{R_\beta \text{diag}(\mathbf{c})\hat{\delta}}{T}\right] + \frac{1}{T}\sqrt{\hat{\sigma}_{0\cdot x}^2 \Gamma_{\mathbf{c}} \left[\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1}\right]^{-1}\Gamma_{\mathbf{c}}'} \cdot \sqrt{\frac{K}{K-2d}} \cdot t_{K-2d}^{1-\eta_2/2},$$

with $\Gamma_{\mathbf{c}} := (R_\beta, -R_\beta \text{diag}(\mathbf{c}))$. By construction, both $r_{\beta,l}^{1-\eta_2/2}(\mathbf{c})$ and $r_{\beta,h}^{1-\eta_2/2}(\mathbf{c})$ are functions of $R_\beta \text{diag}(\mathbf{c})$. Thus, only a subset of $\mathbf{c} = (c_1, \ldots, c_d)$ which corresponds to non-zero components in R_β is relevant. This implies that the number of specific components for β, in which researcher specifies in R_β, will set the computational complexity of constructing the Bonferroni interval. In many empirical applications, it is not unreasonable to pay attention to only a few components of β, e.g., $R_\beta = [1, 0, \ldots, 0]$ and $R_\beta = [1, -1, \ldots, 0]$, which does not add much computational burden in calculating the Bonferroni interval.

For the joint confidence set, $S_T(\eta_1)$, one feasible way is to construct a product set of $[\underline{c}_i, \bar{c}_i]$ for $i = 1, \ldots, d$, where $[\underline{c}_i, \bar{c}_i]$ is a CI for $c_{0,i}$ with $100(1-\eta_1/d)\%$ coverage rate. Then, the resulting confidence set $S_T(\eta_1) = [\underline{c}_1, \bar{c}_1] \times \ldots \times [\underline{c}_d, \bar{c}_d]$ satisfies

$$\begin{aligned}\limsup_{T\to\infty} P(\mathbf{c}_0 \notin S_T(\eta_1)) &= \limsup_{T\to\infty} P(c_{0,i} \notin [\underline{c}_i, \bar{c}_i] \text{ for some } i = 1, \ldots, d) \\ &\leq \sum_{i=1}^{d} \limsup_{T\to\infty} P(c_{0,i} \notin [\underline{c}_i, \bar{c}_i]) = \eta_1,\end{aligned}$$

which implies that the rectangular confidence set, $S_T(\eta_1)$, can formulate a valid Bonferroni inference with the vector-valued cointegrating regressor x_t. Still, the above construction of the joint confidence set is conservative. In principle, an exact joint confidence set can be constructed if we extend Proposition 1 in Elliott and Stock (2001) to a vector-valued x_t and invert their Neyman-Pearson tests for the vector-valued Gaussian autoregression model. We conjecture this way of construction yields an elliptical shape of $S_T(\eta_1)$. However, to our knowledge, there exists no existing work that generalizes Elliott and Stock (2001) to the vector-valued case. We also note the absence of a multivariate version of Hansen's (2001) grid-bootstrap method in the literature. It would be interesting to develop exact confidence set for $\mathbf{c}$ and apply it to our Bonferroni method, and we leave this for future research.

Lastly, the Bonferroni-based CIs in (39) are often too wide with a higher coverage rate than the nominal one (Cavanagh et al., 1995). To avoid the excessive conservatism conveniently, we implement a refined version of (40) which chooses a larger tuning parameter $\tilde{\eta}_1$ so that the refined $CI^{\text{B}}_{R\beta_0}(\tilde{\eta}_1, \eta_2)$ becomes a subset of the original $CI^{\text{B}}_{R\beta_0}(\eta_1, \eta_2)$. As a result, the Bonferroni inequality in (41) has less slack. This way of refinement is also implemented in Campbell and Yogo (2006) in the predictive regression model. For the choice of the Bonferroni tuning parameters, Campbell and Yogo (2006) fix $\eta_2 = \alpha$ and numerically search $\tilde{\eta}_1$ that satisfies (41) by simulating the asymptotic coverage probabilities of the Bonferroni CI. In our Bonferroni-based inference, we choose the values of $\tilde{\eta}_1 = 0.10$ and $\eta_2 = \alpha$, which shows good performances of empirical sizes over a wide range of DGPs simulated in Section 6. The detailed numerical results are summarized in Figures S.1–S.4 in Section S.5 of Supplemental Appendix. Yet, the simulation-based method might be challenging in practical applications because it requires the knowledge of several unknown model parameters such as Ω, ρ_T, and the distribution of $\{u_t\}$. There can be several ways to overcome this difficulty and select a data-adaptive value of Bonferroni tuning parameters. For example, one may simulate the (asymptotic) coverage probabilities based on parametric approximation of the true DGP and use them to find the tuning parameters of Bonferroni CI, e.g., Franchi and Johansen (2017). We leave this for future research.

6 Monte Carlo Evidence

In this section, we evaluate the performance of the modified TA-OLS methods and corresponding Bonferroni-based inferences in finite samples. We compare them with several other methods, including the unmodified TA-OLS approach in HS and the IVX test developed in Phillips and Magdalinos (2009) and Phillips and Lee (2016).

6.1 Data generation process

For a data generation process (DGP) of the cointegration regression, we consider the following triangular cointegration system as in Phillips (2014a) and HS:

$$\begin{array}{l} y_t = \alpha_0 + x_t'\beta_0 + u_{0t} \\ x_t = \rho_T x_{t-1} + u_{xt} \end{array}, u_t = \begin{pmatrix} u_{0t} \\ u_{xt} \end{pmatrix} = \Theta u_{t-1} + \epsilon_t, \tag{43}$$

with a local-to-unity coefficient $\rho_T = I_d - C_0/T$ with C_0 =diag$(c_{0,1}, \ldots, c_{0,d})$, and

$$\epsilon_t = \begin{pmatrix} \epsilon_{0t} \\ \epsilon_{xt} \end{pmatrix} \overset{i.i.d}{\sim} N(0, \Sigma), \ \Theta = \psi \cdot I_{d+1}, \ \Sigma = J_{d+1,d+1} \cdot \phi + I_{d+1} \cdot (1-\phi),$$

and $J_{d+1,d+1}$ is the $(d+1) \times (d+1)$ matrix of ones. The initial value of the error process u_t is drawn from standard normal distribution. To minimize the initialization effect, we generate a time series of length $2T$ and drop the first T observations. The parameter ψ controls the persistence of individual components in $u_t = (u_{0t}, u_{xt}')' \in \mathbb{R}^{d+1}$. We set the values of ψ as $\{0.25, 0.50, 0.75\}$, so the stationary cointegration error u_t is in the reasonable range of persistency. The parameter ϕ is a pairwise correlation coefficient between the elements of u_t and characterizes the degree of endogeneity. With some additional algebra, the squared long run correlation $r^2 = \sigma_{0x}\Omega_{xx}^{-1}\sigma_{x0}/\sigma_0^2$ is expressed by $d\phi^2/((1-\phi)+d\phi)$. Using this formula, we set ϕ to satisfy $r^2 \in \{0, 0.25, 0.50, 0.75\}$.

In our simulations, we consider $d \in \{1, 2\}$ as a dimension of the cointegration regressor x_t, and set the true regression coefficients by $\alpha_0 = 1$, and $\beta_0 = 1$ or $\beta_0 = (1, 1)'$. When $d = 1$,we take the AR(1) coefficients of x_t in $\{1, 0.975, 0.95, 0.90\}$ with sample size $T = 200$, and set the corresponding pairs of local-to-unity parameters $c_0 \in \{0, 5, 10, 20\}$. For the case of the multiple regressors, with $d = 2$, we take the same values of $c_{0,1}$ for the first regressor and set the second regressor as a unit root process with $c_{0,2} = 0$. Although we have the exact I(1) process in the second regressor, our feasible Bonferroni methods do not impose this knowledge of $c_{0,2}$ to reflect a practical empirical application. Instead, they construct the joint rectangular confidence set for $(c_{0,1}, c_{0,2})$ which is described in subsection 5.2.

6.2 Choices of tests

The null hypotheses of interests for the true parameter, $\beta_0 = (\beta_{01}, \ldots, \beta_{0d})'$, are

$$H_0^\beta : \beta_{01} = 1 \text{ vs } H_1^\beta : \beta_{01} \neq 1 \text{ with } d = 1, \tag{44}$$

$$H_0^\beta : \beta_{01} = \beta_{02} \text{ vs } H_1^\beta : \beta_{01} \neq \beta_{02} \text{ with } d = 2, \tag{45}$$

and corresponding testing matrix is $R_\beta = (1, 0)$ with $r_\beta = 1$, and $R_\beta = (1, -1)$ with $r_\beta = 0$, respectively. Also, we test the long run endogeneity parameter with the following null hypothesis of

$$H_0^\delta : \delta_0 = 0 \text{ and } H_1^\delta : \delta_0 \neq 0.$$

We consider Fourier basis functions given in (6) for our TA-OLS framework, as the same numerical evidence holds for the cosine transformation in (7). For fixed values of K, we set $K = 8$, $K = 16$, and $K = 24$ for the AR(1) parameters $\psi = 0.75$, $\psi = 0.50$, and $\psi = 0.25$, respectively. These choices of K are shown to have good finite sample performances in various literature of fixed smoothing asymptotics with extensive numerical experiments. See, for example, Müller and Watson (2013, 2017), HS, and Lazarus et al. (2018). In all of our simulations, the number of simulation replications is 10,000.

In our simulations, we consider the empirical size of five different types of TA-OLS t-tests studied in this paper at nominal size 5%. The first test is the unmodified TA-OLS test, (`TAOLS` hereafter), considered in HS. The second test is an infeasible version of the modified TA-OLS t-test in Theorem 3, (`M-TAOLS` hereafter), which treats the true local-to-unity parameter as known. The next three tests are feasible versions of the modified TA-OLS test which implement the Bonferroni-based CI in Section 5, but they use different methods to construct CIs for the local-to-unity parameter: the first one (`Bonf-M-TAOLS (Hansen)` hereafter) uses Hansen's (2001) grid-bootstrap method, the second one (`Bonf-M-TAOLS (M-Hansen)` hereafter) uses our modification of Hansen's (2001) grid-bootstrap method, and the third one (`Bonf-M-TAOLS (ES)` hereafter) uses Elliott and Stock's (2001) method. These three Bonferroni tests reject when the null hypothesized value does not fall into the Bonferroni confidence intervals. It is important to point out that `Bonf-M-TAOLS (Hansen)` ignores the temporal dependence in $\{u_{xt}\}$. On the other hand, `Bonf-M-TAOLS (M-Hansen)` reflects the dependence structure in $\{u_{xt}\}$ on the construction of CI for c by a finite-order (approximated) autoregressive process. For the choice of the Bonferroni tuning parameters, we fix $\eta_2 = 0.05$ and select the value of $\tilde{\eta}_1$ as 0.10.

As the last test in our simulation, we consider the IVX estimator in Phillips and Magdalinos (2009) and Phillips and Lee (2016). Statistical inference via the IVX estimator has been known to solve the difficulty of the cointegration regression with the near-unity regressor, which is presented in our setting. The IVX estimator for β_0 is

$$\hat{\beta}_{\text{IVX}} = \left(\sum_{t=1}^{T} \tilde{z}_t \tilde{x}_t' \right)^{-1} \left(\sum_{t=1}^{T} \tilde{z}_t \tilde{y}_t - T \hat{\Lambda}_{x0} \right),$$

where $\tilde{x}_t = x_t - T^{-1} \sum_{t=1}^{T} x_t$, $\tilde{y}_t = y_t - T^{-1} \sum_{t=1}^{T} y_t$ are demeaned observations. $\hat{\Lambda}_{x0}$ is the estimator for the one-sided long run covariance $\Lambda_{x0} = \sum_{j=0}^{\infty} E[u_{xt} u_{0t-j}]$. $\tilde{z}_{it}$'s are the self-generated instrumental variables, defined as:

$$\tilde{z}_{it} = \sum_{j=1}^{t} \left(1 - \frac{c_z}{T^\gamma} \right)^{t-j} \Delta x_j.$$

For the tuning parameters, γ and c_z, we follow Phillips and Lee (2016) and use $\gamma \in \{0.85, 0.90, 0.95\}$

and $c_z = 5$, respectively. Here, we only report results with $\gamma = 0.85$, as the quantitative results with other choices of γ are very similar. With $\hat{\beta}_{\mathrm{IVX}}$, the IVX t-test uses the asymptotic normal critical value with the following t-statistics:

$$t_{\mathrm{IVX}} = \frac{R_\beta \hat{\beta}_{\mathrm{IVX}} - r}{\sqrt{R_\beta \left\{ (X' P_z X)^{-1} \hat{\sigma}_0^2 \right\} R_\beta'}},$$

where $\hat{\sigma}_0^2$ is the long run variance estimator of $\sigma_0^2 = \sum_{j=-\infty}^{\infty} E[u_{0t} u_{0t-j}]$. To nonparametrically estimate the nuisance parameters, $\hat{\Lambda}_{x0}$ and $\hat{\sigma}_0^2$, we use Bartlett kernel with the optimal bandwidth rule in Andrews (1991). It is important to point out that this external procedure to the nonparametric long run variance estimators is required to implement the IVX test. In contrast, the TA-OLS methods developed in our paper automates the estimation of the long run nuisance parameters such as $\hat{\delta}$ and $\hat{\sigma}_{0\cdot x}^2$. In fact, our simulation results below show that the finite sample uncertainties embodied in the nonparametric long run variance estimators have crucial impacts on the performance of the IVX test in finite samples.

6.3 Results for finite sample sizes

For testing (44) in the single dimensional case, Tables S.3–S.5 and Figures S.5–S.7 report empirical sizes (Type I error) of five different TA-OLS and the IVX tests for $c_0 \in \{0, 5, 10, 20\}$. For the multi-dimensional case in (45), we only report the results for $(c_{0,1}, c_{0,2}) = (20, 0)$ in Table S.6, as the quantitative implications of other cases are quite similar.

In the unit root case, that is $c_0 = 0$, both `TAOLS` and `M-TAOLS` have empirical sizes close to the nominal size of 5%. This is not surprising given that `TAOLS` is asymptotically valid under the exact unit root assumption. The results also show that `Bonf-M-TAOLS (Hansen)`, which ignores the dependence structure in $\{u_{xt}\}$ suffers from size distortions varying from 7%–23%. The size distortion of `Bonf-M-TAOLS (Hansen)` is emphasized when the degree of serial dependence, ψ, is significant, e.g., $\psi \in \{0.50, 0.75\}$. On the other hand, the Bonferroni based methods which reflect the dependency in $\{u_{xt}\}$, `Bonf-M-TAOLS (M-Hansen)` and `Bonf-M-TAOLS (ES)` yields to correct empirical sizes, although `Bonf-M-TAOLS (M-Hansen)` is mildly undersized varying from 3.6% to 4.9%. The variations between these two Bonferroni-based methods can be explained by the different ways of constructing CIs for c_0. While `Bonf-M-TAOLS (ES)` inverts the asymptotically efficient GLS-based unit-root test, `Bonf-M-TAOLS (M-Hansen)` implements the grid-bootstrap t-test in augmented Dickey-Fuller equation.

Second, as c_0 deviates from zero, `TAOLS` suffers from severe size distortions, especially when the squared long run correlation (r^2), and the local-to-unity parameter (c_0) increase. When $r^2 = 0.75$, our numerical results in Tables S.3–S.5 show that the size distortions of `TAOLS` can be significant, e.g., 37.0%–49.3%, for even slight deviation from a unit root regressor, e.g., $c_0 = 5$. This result is consistent with our theoretical results in Proposition 2. Also, Tables S.3–S.5 show that the IVX test (`IVX`) which is known to be robust in the presence of the local-to-unity regressor, can be size-distorted in finite-samples when ψ is large. This is because the normal critical value in the IVX test statistics completely ignores the estimation uncertainty in the nonparametric estimators $\hat{\Lambda}_{x0}$ and $\hat{\sigma}_0^2$. A similar message is pointed out in HS, who finds poor performance of fully-modified (FM) cointegration in the unit root cointegration regressors. Our results also show that the size distortions of `IVX` can be amplified when the local-to-unity parameter (c_0) and the degree of the long run endogeneity (r^2) increases. We find that the IVX test can work the best when u_{0t} has a low serial correlation, e.g., $\psi = 0.25$, and the cointegration regressor x_t is not too deviated from the unit root, e.g., $c_0 = 5$.

We also find that the infeasible `M-TAOLS` has the most accurate finite sample sizes for all values of r^2 and c_0 considered in our simulations. Also, the feasible Bonferroni-based methods, `Bonf-M-TAOLS (M-Hansen)`and `Bonf-M-TAOLS (ES)`, have correct sizes, varying 1.2%–4.2% and 1.3%–7.2%, respectively. The conservatism of Bonferroni comes from the Bonferroni step in (39) and (40). While `Bonf-M-TAOLS (M-Hansen)`and `Bonf-M-TAOLS (ES)` show similar performances for $c_0 \in \{5, 10\}$, Tables S.3–S.5 indicate that `Bonf-M-TAOLS (ES)` can be size distorted when c_0 is large, e.g., $c_0 = 20$. This is because Elliott and Stock's (2001) CI is subject to uniformity critique when the true local-to-unity parameter c largely deviates from zero. In contrast, `Bonf-M-TAOLS (M-Hansen)` does not suffer from this drawback because it uses a uniform CI for c_0 and reflect the dependency in $\{u_{xt}\}$.

In summary, first, there is a large amount of size distortions for `TAOLS` in the local-to-unity case with non-zero r^2. Second, treating c_0 as known, the infeasible `M-TAOLS` successfully corrects the size distortions of `TAOLS`. When c_0 is unknown, the feasible versions of our modified TA-OLS, `Bonf-M-TAOLS (ES)`and `Bonf-M-TAOLS (M-Hansen`, have correct sizes. However, our results indicate that ignoring the dependence structure in $\{u_{xt}\}$ in `Bonf-M-TAOLS (Hansen)` drives severe size distortions in the Bonferroni-based method. Also, the valid Bonferroni-based TA-OLS methods outperform `IVX` with large margins when ψ is 0.75. Lastly, our unreported results, which are available upon requests, indicate that we can precisely perform the endogeneity test, i.e., a test of whether $\delta_0 = 0$, regardless of the local-to-unity parameters c_0. This is consistent with our fixed-K asymptotic results in Proposition 2 (c) and (d).

6.4 Results for finite sample power

Since our feasible Bonferroni-based TA-OLS methods are mildly undersized in most of our DGPs, we expect that this conservatism results in some power loss compared to other types of tests. To investigate this aspect, we simulate local power curves assuming that the true parameter of cointegration is from the local alternative hypothesis $\beta = \beta_0 + b/T$, where $b \in [-25, 25]$ measures the magnitude of the local departure. To make meaningful power comparisons between different methods, we investigate size-adjusted power curves for `M-TAOLS`, `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX`. We implement the size adjustments of `M-TAOLS` and `IVX` by computing empirical quantiles of their test statistics under $\beta = \beta_0$. For `Bonf-M-TAOLS (M-Hansen)`and `Bonf-M-TAOLS (ES)`, we adjust their empirical rejection probabilities under $\beta = \beta_0$ to be the nominal level 5% by numerically searching values of the tuning parameter $\tilde{\eta}_2$, given $\eta_1 = 0.05$. The values of $\tilde{\eta}_2$ depend on the choices of c_0, ψ, and r^2 in true DGPs. We compute the finite sample power curve of each procedure for $c_0 \in \{0, 5, 10, 20\}$ with various degrees of r^2 and ψ considered in previous subsection. To save space, we only report the results for $\psi = 0.50$ in Figures S.8–S.11, as qualitative implications for other values of ψ can be delivered in a similar way.

The results in Figures S.8–S.11 first indicate that the power of `M-TAOLS` outperforms the feasible Bonferroni-based TA-OLS and IVX tests in all cases. Thus, the cost of the lack of knowledge of c_0 is reflected on the relative power loss of the Bonferroni-based TA-OLS tests. Figures S.8–S.11 also indicate that the relative power losses increase with respect to the squared long run correlation r^2. Also, the Bonferroni-based TA-OLS tests are slightly more powerful than the IVX test when c_0 is small, e.g., $c_0 \in \{0, 5\}$. However, Figure S.11 for $c_0 = 20$ indicates that the IVX has a better power when the long-run correlations are small, e.g., $r^2 \in \{0.00, 0.25\}$. Lastly, we check that the two Bonferroni-based methods, `Bonf-M-TAOLS (ES)`and `Bonf-M-TAOLS (M-Hansen)`, have quite similar power for most cases, but `Bonf-M-TAOLS (ES)` has some power gain over `Bonf-M-TAOLS (M-Hansen)`when $c_0 = 5$ and $r^2 = 0.75$. However, the power advantage of `Bonf-M-TAOLS (ES)`over

`Bonf-M-TAOLS (M-Hansen)` disappears when c_0 increases to 20.

All in all, the feasible versions of the modified TA-OLS with the Bonferroni procedures shown in this paper have advantages over the IVX test when we consider the balance between size and power. This is because the two Bonferroni-based TA-OLS methods, `Bonf-M-TAOLS (ES)` and `Bonf-M-TAOLS (M-Hansen)`, outperforms the IVX test on a wide range of DGPs considered in our simulations. When the true c_0 is small, `Bonf-M-TAOLS (ES)` is less conservative than `Bonf-M-TAOLS (M-Hansen)` and can be more powerful. However, `Bonf-M-TAOLS (ES)` can be size distorted when c_0 largely deviates from zero. On the other hand, `Bonf-M-TAOLS (M-Hansen)` shows stable performances with correct sizes and favorable power for broad ranges of c_0.

7 Conclusion

In this paper, we develop a theory that adopts a local-to-unity approximations to a triangular cointegrated system. Our analysis is carried out on the domain of low frequencies by transforming data from the original time domain. We show that the unmodified TA-OLS in Hwang and Sun (2017) possesses an asymptotic bias term in the limiting distribution. As a result, the unmodified TA-OLS suffers from severe size distortions, especially, when the degree of long run endogeneity grows, or the cointegration regressor deviates from the exact unit root.

We develop modified TA-OLS test statistics, which yields to convenient t and F inferences for the cointegrating vector and long run endogeneity parameter. The modified TA-OLS not only adjusts for the asymptotic bias arising from the local-to-unity regressor but also corrects the uncertainty of the plugged-in bias correction term. When the local-to-unity parameter is unknown, we also provide feasible versions of modified TA-OLS, which considers a Bonferroni-based inferences. The Bonferroni-based methods require confidence intervals for the local-to-unity parameter. We note that implementation of Elliott and Stock's (2001) confidence interval can be subjective to uniformity critique in Phillips (2014b) when the true local-to-unity parameter is large. To overcome this issue, we propose a modification of Hansen's (2001) confidence interval. The corresponding Bonferroni-based method overcomes the uniformity critique and shows correct sizes and appealing power in finite samples.

Our numerical results also show that the size distortions of the existing IVX test can be amplified when the local-to-unity parameter (c_0) and the degree of the long run endogeneity (r^2) are important. Also, we find that the proposed Bonferroni-based TA-OLS tests have favorable finite sample properties compared to the IVX test when we consider the balance between size and power.

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

[2] Andrews, D.W.K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica*, 61(1), 139-165. doi:10.2307/2951781

[3] Andrews, D. W., Guggenberger, P. (2014). A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. *Review of Economics and Statistics*, 96(2), 376-381.

[4] Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics* 77(2), 379–404.

[5] Bobkoski, M. J. (1983). Hypothesis testing in nonstationary time series, Unpublished Ph.D. Thesis, Department of Statistics, University of Wisconsin.

[6] Cavanagh, C. L., Elliott, G., Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric theory*, 11(5), 1131-1147.

[7] Campbell, J. Y., Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1), 27-60.

[8] Christiano, L. J., Eichenbaum, M. (1990). Unit roots in real GNP: Do we know, and do we care?. In Carnegie-Rochester Conference Series on Public Policy (Vol. 32, pp. 7-61). North-Holland.

[9] Davidson, J. (1994). Stochastic limit theory: An introduction for econometricians. OUP Oxford.

[10] Duffy, J. A., Simons, J. R. (2020). The Cointegrated VAR without Unit Roots: Representation Theory and Asymptotics. arXiv preprint arXiv:2002.08092.

[11] Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica*, 66(1), 149-158. doi:10.2307/2998544

[12] Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of econometrics*, 164(1), 79-91.

[13] Elliott, G., Stock, J. H. (2001). Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics*, 103(1-2), 155-181.

[14] Engle, R. F., Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

[15] Franchi, M., & Johansen, S. (2017). Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles. *Econometrics*, 5(2), 25.

[16] Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics*, 81(4), 594-607.

[17] Hwang, J., Sun, Y. (2017). Simple, robust, and accurate F and t tests in cointegrated systems. *Econometric Theory*, 1-36.

[18] Jansson, M., & Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3), 681-714.

[19] Kiefer, N.M., Vogelsang, T.J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

[20] Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5), 1506-1553.

[21] Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4), 541-559.

[22] McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, 200(1), pp.17-35.

[23] Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5), 1411-1452.

[24] Müller, U.K., (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, 141(2), pp.1331-1352.

[25] Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics* 32(3), 311–322.

[26] Müller, U.K. and Watson, M.W., (2008). Testing models of low frequency variability. *Econometrica*, 76(5), pp.979-1016.

[27] Müller, U. K., Watson, M.W. (2013). low-frequency robust cointegration testing. *Journal of Econometrics* 174(2), 66–81.

[28] Müller, U., & Watson, M. (2017). low-frequency Econometrics. In B. Honoré, A. Pakes, M. Piazzesi, & L. Samuelson (Eds.), Advances in Economics and Econometrics: Eleventh World Congress (Econometric Society Monographs, pp. 53-94). Cambridge: Cambridge University Press. doi:10.1017/9781108227223.003

[29] Newey, W., West, K. (1987). A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 55, 103080.

[30] Park, J. Y., and Phillips, P.C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(3), 468-497.

[31] Phillips, P.C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3), pp.535-547.

[32] Phillips, P.C. (1991a). Spectral regression for cointegrated time series. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics.* Cambridge University Press, New York, pp. 413–435.

[33] Phillips, P. C. (1991b). Optimal inference in cointegrated systems. Econometrica: Journal of the Econometric Society, 283-306.

[34] Phillips, P.C. (2005). HAC estimation by automated regression. *Econometric Theory* 21(1), 116–142.

[35] Phillips, P.C., Magdalinos, T., (2009). Econometric inference in the vicinity of unity. Singapore Management University, CoFie Working Paper, 7.

[36] Phillips, P.C. (2014a). Optimal estimation of cointegrated systems with irrelevant instruments, *Journal of Econometrics* 178(2), 210–224.

[37] Phillips, P.C. (2014b). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3), 1177-1195.

[38] Phillips, P.C., Durlauf, S.N. (1986). Multiple regression with integrated processes. *Review of Economic Studies* 53, 473–496.

[39] Phillips, P.C., Hansen, B.E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

[40] Phillips, P. C., Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics*, 192(2), 433-450.

[41] Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. Econometric theory, 7(1), 1-21.

[42] Stock, J. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series, *Journal of Monetary Economics*, 28, issue 3, p. 435-459.

[43] Sun, Y., Phillips, P.C.B., Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76(1), 175–94.

[44] Sun, Y. (2020). Some Extensions of Asymptotic F and t Theory in Nonstationary Regressions. Working paper.

[45] Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *The Econometrics Journal*, 16(1), 1-26.

[46] Sun, Y., (2014). HAC Corrections for Strongly Autocorrelated Time Series Comment. *Journal of Business & Economic Statistics*, 32:3, 311-322.

Supplemental Appendix: "Low Frequency Cointegrating Regression with Local to Unity Regressors and Unknown Form of Serial Dependence"

Jungbin Hwang
University of Connecticut
Department of Economics

Gonzalo Valdés
Universidad de Tarapacá
Departamento de Ingeniería Industrial y de Sistemas

The first part of this supplementary appendix (Section S.1) provides proofs of the main results. The following section discusses non-linear and joint testings using the modified TA-OLS. Section S.3 presents detailed procedures for calculating the modified Hansen (1999) and Elliott and Stock's (2001) confidence intervals in our Bonferroni-based inferences. Section S.4 provides conditions that simply the computation of our Bonferroni intervals. Lastly, Section S.5 presents numerical results in Tables and Figures, which are referenced in the main text of the paper.

S.1 Proofs of Main Results

Proof of Proposition 1. We begin by showing the asymptotic equivalence between $\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_{t-1}\phi_i(\frac{t}{T})$ and the transformed regressor $\mathbb{W}_x/T$ in (25), that is,

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_{t-1}\phi_i\left(\frac{t}{T}\right) = \frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t\phi_i\left(\frac{t}{T}\right) + O_p\left(\frac{1}{T}\right).$$

The left side of the equation is

$$\frac{1}{T}\sum_{t=1}^{T}\frac{x_{t-1}}{\sqrt{T}}\phi_i\left(\frac{t}{T}\right) = \frac{1}{T}\sum_{s=0}^{T-1}\frac{x_s}{\sqrt{T}}\phi_i\left(\frac{s}{T}\right) + \frac{1}{T}\sum_{t=1}^{T}\frac{x_{t-1}}{\sqrt{T}}\left[\phi_i\left(\frac{t}{T}\right) - \phi_i\left(\frac{t-1}{T}\right)\right]. \tag{B.1}$$

By mean value theorem,

$$\phi_i\left(\frac{t}{T}\right) = \phi_i\left(\frac{t-1}{T}\right) + \phi'(r_t^*)\left(\frac{1}{T}\right) \text{ for some } r_t^* \in \left[\frac{t-1}{T}, \frac{t}{T}\right],$$

and Assumption 2 yields

$$\phi_i\left(\frac{t}{T}\right) - \phi_i\left(\frac{t-1}{T}\right) = \frac{\phi'(r_t^*)}{T} \leq \frac{M}{T}$$

for some $M > 0$ uniformly over t. Therefore, the second term in (B.1) satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\frac{x_{t-1}}{\sqrt{T}}\left[\phi_i\left(\frac{t}{T}\right) - \phi_i\left(\frac{t-1}{T}\right)\right] \leq \left(\frac{M}{T}\right)\left[\frac{1}{T}\sum_{t=0}^{T-1}\frac{x_t}{\sqrt{T}}\right] = O_p\left(\frac{1}{T}\right).$$

For the first term in (B.1),

$$\frac{1}{T}\sum_{s=0}^{T-1}\frac{x_s}{\sqrt{T}}\phi_i\left(\frac{s}{T}\right)=\frac{1}{T}\sum_{s=1}^{T}\frac{x_s}{\sqrt{T}}\phi_i\left(\frac{s}{T}\right)+\frac{x_0}{T^{3/2}}\phi_i(0)-\frac{x_T}{T^{3/2}}\phi_i(1) \tag{B.2}$$
$$=\frac{1}{T}\sum_{s=1}^{T}\frac{x_s}{\sqrt{T}}\phi_i\left(\frac{s}{T}\right)+O_p\left(\frac{1}{T}\right),$$

where the second equality follows from $x_0=O_p(1)$ and the equation (23). With this result and the weak convergences in (10), (11), and (12), we get

$$\Upsilon_T^{-1}\mathbb{W}_X=(\mathbb{W}_x/T,\mathbb{W}_{\Delta x})\Rightarrow\mathbb{S}_X=(\mathbb{S}_x,\mathbb{S}_{\Delta x}), \tag{B.3}$$
$$\tilde{\mathbb{W}}_x\Rightarrow\mathbb{S}_{0\cdot x}+c_0\cdot\mathbb{S}_x\delta_0,$$

where $\tilde{\mathbb{W}}_x=(\tilde{\mathbb{W}}_{x,1},...,\tilde{\mathbb{W}}_{x,K})'$.Then, by the definition of $\hat{\gamma}$ and Υ_T, we have

$$\begin{aligned}\Upsilon_T(\hat{\gamma}-\gamma_0)&=(\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\mathbb{W}_X'\tilde{\mathbb{W}}_{0\cdot x}\\&\Rightarrow\left(\mathbb{S}_X'\mathbb{S}_X\right)^{-1}\mathbb{S}_X'\left[\mathbb{S}_{0\cdot x}+c_0\cdot\mathbb{S}_x\delta_0\right]\\&=\left(\mathbb{S}_X'\mathbb{S}_X\right)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x}+c_0\cdot\left(\mathbb{S}_X'\mathbb{S}_X\right)^{-1}\mathbb{S}_X'\mathbb{S}_x\delta_0.\end{aligned}$$

Since $\{\mathbb{S}_{0\cdot x,i}\}_{i=1}^K$ is $i.i.d$ normal random variables with variance $\sigma_{0\cdot x}^2$ and is independent with $\mathbb{S}_X=(\mathbb{S}_x,\mathbb{S}_{\Delta x})$, the latter component can be expressed by a mixture of normal distribution

$$MN(0,\sigma_{0\cdot x}^2\left(\mathbb{S}_X'\mathbb{S}_X\right)^{-1}).$$

The second component can be written more explicitly as

$$\begin{aligned}c_0\cdot\left(\mathbb{S}_X'\mathbb{S}_X\right)^{-1}\mathbb{S}_X'\mathbb{S}_x\delta_0&=c_0\cdot\begin{pmatrix}\mathbb{S}_x'\mathbb{S}_x & \mathbb{S}_x'\mathbb{S}_{\Delta x}\\ \mathbb{S}_{\Delta x}'\mathbb{S}_x & \mathbb{S}_{\Delta x}'\mathbb{S}_{\Delta x}\end{pmatrix}^{-1}\begin{pmatrix}\mathbb{S}_x'\mathbb{S}_x\delta_0\\ \mathbb{S}_{\Delta x}\mathbb{S}_x\delta_0\end{pmatrix}\\&=\begin{pmatrix}c_0\cdot(\mathbb{S}_x'M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x)^{-1}\mathbb{S}_x'M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\delta_0\\ c_0\cdot(\mathbb{S}_{\Delta x}'M_{\mathbb{S}_x}\mathbb{S}_{\Delta x})^{-1}\mathbb{S}_{\Delta x}'M_{\mathbb{S}_x}\mathbb{S}_x\delta_0\end{pmatrix}=\begin{pmatrix}c_0\delta_0\\ 0\end{pmatrix},\end{aligned}$$

which finishes the proof. ■

Proof of Proposition 2. We prove the result for the F statistic only, as the result for t statistic can be proved in a similar manner. Note that

$$\hat{\sigma}_{0\cdot x}^2=\frac{1}{K}\sum_{i=1}^{K}\hat{\omega}_{0\cdot x,i}^2=\frac{1}{K}\mathbb{W}_Y'\left[I_K-\mathbb{W}_X(\mathbb{W}_X'\mathbb{W}_X)^{-1}\mathbb{W}_X'\right]\mathbb{W}_Y \tag{B.4}$$
$$=\frac{1}{K}\tilde{\mathbb{W}}_{0\cdot x}'\left[I_K-\mathbb{W}_X(\mathbb{W}_X'\mathbb{W}_X)^{-1}\mathbb{W}_X'\right]\tilde{\mathbb{W}}_{0\cdot x}$$
$$\Rightarrow\frac{1}{K}\left[\mathbb{S}_{0\cdot x}+c_0\cdot\mathbb{S}_x\delta_0\right]'\left[I_K-\mathbb{S}_X(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\right]\left[\mathbb{S}_{0\cdot x}+c_0\cdot\mathbb{S}_x\delta_0\right].$$

Since $P_{\mathbb{S}_X}=\mathbb{S}_X(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'$ is a projection matrix onto a space generated by $[\mathbb{S}_x,\mathbb{S}_{\Delta x}]$, it is easy to check

$$\left[I_K-\mathbb{S}_X(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\right]\left[c_0\cdot\mathbb{S}_x\delta_0\right]=0.$$

Therefore, the weak convergence limit of the estimator $\hat{\sigma}^2_{0\cdot x}$ simplifies to

$$\hat{\sigma}^2_{0\cdot x} \Rightarrow \frac{1}{K}\mathbb{S}'_{0\cdot x}M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x} \overset{d}{=} \frac{\sigma^2_{0\cdot x}}{K}\chi^2_{K-2d},$$

where $M_{\mathbb{S}_X} := I_K - \mathbb{S}_X(\mathbb{S}'_X\mathbb{S}_X)^{-1}\mathbb{S}'_X$. Combining this result with

$$T(R_\beta\hat{\beta} - r_\beta) \Rightarrow R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_{0\cdot x} + c_0 R_\beta\delta_0$$

and

$$R_\beta\left[(\mathbb{W}'_x/T)M_{\Delta x}(\mathbb{W}'_x/T)\right]^{-1}R'_\beta \Rightarrow R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}R'_\beta,$$

we have that

$$F(\hat{\beta}) \Rightarrow \frac{K}{p_\beta}\frac{\left\|\frac{Z}{\sigma_{0\cdot x}} + c_0\cdot\left[R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}R'_\beta\right]^{-1/2}\cdot\left[\frac{R_\beta\delta_0}{\sigma_{0\cdot x}}\right]\right\|^2}{\left[\frac{\mathbb{S}'_{0\cdot x}M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}}{\sigma^2_{0\cdot X}}\right]}, \tag{B.5}$$

where

$$Z = \left[R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}R_\beta\right]^{-1/2}R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_{0\cdot x} \sim N(0, \sigma^2_{0\cdot x}\cdot I_K).$$

Conditional on $\mathbb{S}_X = (\mathbb{S}_x, \mathbb{S}_{\Delta x})$, $M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}$ and $\mathbb{S}_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_{0\cdot x}$ are independent, as both $M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}$ and $\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_{0\cdot x}$ are normal and its conditional covariance is

$$cov\left(M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}, \mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_{0\cdot x}\right) = \sigma^2_{0\cdot x}\left[I_K - \mathbb{S}_X(\mathbb{S}'_X\mathbb{S}_X)^{-1}\mathbb{S}'_X\right]M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x = 0.$$

This implies that Z is independent of $\mathbb{S}'_{0\cdot x}M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}$ conditional on $\mathbb{S}_X = (\mathbb{S}_x, \mathbb{S}_{\Delta x})$, and hence

$$\begin{aligned}&\frac{K}{p_\beta}\frac{\left\|\frac{Z}{\sigma_{0\cdot x}} + c_0\cdot\left[R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}R'_\beta\right]^{-1/2}\cdot\left[\frac{R_\beta\delta_0}{\sigma_{0\cdot x}}\right]\right\|^2}{\left[\frac{\mathbb{S}'_{0\cdot x}M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}}{\sigma^2_{0\cdot X}}\right]}\\ &\overset{d}{=} \frac{K}{K-2d}F_{p_\beta, K-2d}\left(\|\theta\|^2\right),\end{aligned}$$

where

$$\theta = \left[R_\beta\left[\mathbb{S}'_x M_{\mathbb{S}_{\Delta x}}\mathbb{S}_x\right]^{-1}R'_\beta\right]^{-1/2}\cdot\left[\frac{c_0 R_\beta\delta_0}{\sigma_{0\cdot x}}\right].$$

Similarly, with $Z = \left[R_\delta\left[\mathbb{S}'_{\Delta x}M_{\mathbb{S}_X}\mathbb{S}_{\Delta x}\right]^{-1}R_\delta\right]^{-1/2}R_\delta\left[\mathbb{S}'_{\Delta x}M_{\mathbb{S}_X}\mathbb{S}_{\Delta x}\right]^{-1}\mathbb{S}'_{\Delta x}M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}$, we obtain

$$F(\hat{\delta}) \Rightarrow \frac{K}{p_\delta}\frac{\left\|\frac{Z}{\sigma_{0\cdot x}}\right\|^2}{\left[\frac{\mathbb{S}'_{0\cdot x}M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}}{\sigma^2_{0\cdot X}}\right]} \overset{d}{=} \frac{K}{K-2d}F_{p_\delta, K-2d}.$$

■

Proof of Theorem 3. We prove the result for the Wald statistics only, as the same proof goes through for the t statistics with obvious modifications. From (30) and (31), we have

$$T\left(R_\beta\left[\hat{\beta}-c_0\cdot\frac{\hat{\delta}}{T}\right]-r_\beta\right)\Rightarrow\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x},$$
$$\Gamma_{c_0}(\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1})^{-1}\Gamma_{c_0}'\Rightarrow\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'.$$

Combining these results with (B.4), we have

$$\begin{aligned}F(\hat{\beta};c_0)&=\frac{T^2}{\hat{\sigma}_{0\cdot x}^2}(R_\beta[\hat{\beta}-c_0\cdot\hat{\delta}/T]-r_\beta)'\left[\Gamma_{c_0}(\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1})^{-1}\Gamma_{c_0}'\right]^{-1}\\&\times(R_\beta[\hat{\beta}-c_0\cdot\hat{\delta}/T]-r_\beta)/p.\\&\Rightarrow\left[\frac{K}{p_\beta}\right]\frac{\left[\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x}\right]'\left[\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'\right]^{-1}\left[\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x}\right]}{\mathbb{S}_{0\cdot x}'M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}}\end{aligned}$$

Using a similar argument in the proof of Proposition 2, the conditional limit of Wald statistics $F(\hat{\beta};c_0)$ can be expressed as

$$\begin{aligned}&\left[\frac{K}{p_\beta}\right]\frac{\left[\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x}\right]'\left[\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'\right]^{-1}\left[\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\mathbb{S}_X'\mathbb{S}_{0\cdot x}\right]}{\mathbb{S}_{0\cdot x}'M_{\mathbb{S}_X}\mathbb{S}_{0\cdot x}}\\&\overset{d}{=}\frac{K}{p}\frac{\chi_{p_\beta}^2}{\chi_{K-2d}^2},\ \chi_p^2\perp\chi_{K-2d}^2\end{aligned}$$

which is invariant to the conditioning variable $\mathbb{S}_X$. Thus, it is also the unconditional distribution which proves

$$F(\hat{\beta};c_0)\Rightarrow\frac{K}{K-2d}F_{p,K-2d}.$$

■

S.2 Discussions on testing for β_0 and δ_0

In this section, we apply the bias-corrected inference of modified TA-OLS to test simultaneous restrictions on β_0 and δ_0 and discuss a non-linear testing hypothesis.[1]

For $R_\beta\in\mathbb{R}^{p_\beta\times d}$ and $R_\delta\in\mathbb{R}^{p_\delta\times d}$ such that p_β, $p_\delta\le d$, we first consider the following form of hypothesis:

$$H_0^\gamma: R_\beta\beta_0=r_\beta \text{ and } R_\delta\delta_0=r_\delta. \tag{B.6}$$

We reformulate the matrix Γ_{c_0} as

$$\Gamma_{c_0}=\begin{pmatrix}R_\beta & -c_0R_\beta\\ 0 & R_\delta\end{pmatrix}, \tag{B.7}$$

[1]We thank an anonymous referee who pointed out this issue.

and apply Proposition 1 to obtain that

$$
\begin{aligned}
\Gamma_{c_0}\Upsilon_T\left[\hat{\gamma}-\gamma_0\right] &= \begin{pmatrix} R_\beta & -c_0R_\beta \\ 0 & R_\delta \end{pmatrix}\begin{pmatrix} T(\hat{\beta}-\beta_0) \\ \hat{\delta}-\delta_0 \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} R_\beta & -c_0R_\beta \\ 0 & R_\delta \end{pmatrix}\begin{bmatrix} c_0\delta_0 \\ 0 \end{bmatrix} + MN(0,\sigma^2_{0\cdot x}\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'),
\end{aligned}
$$

where Υ_T is defined in (26). This implies that

$$
\Upsilon_T\left\{\begin{pmatrix} R_\beta\left(\hat{\beta}-\frac{c_0\hat{\delta}}{T}\right) \\ R_\delta\hat{\delta} \end{pmatrix} - \begin{pmatrix} r_\beta \\ r_\delta \end{pmatrix}\right\} \Rightarrow MN(0,\sigma^2_{0\cdot x}\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}'),
$$

which leads us to formulate the modified Wald test for (B.6) as below:

$$
\begin{aligned}
F(\hat{\gamma};c_0) &= \frac{\Upsilon_T}{\hat{\sigma}^2_{0\cdot x}}\left[\begin{pmatrix} R_\beta\left(\hat{\beta}-\frac{c_0\hat{\delta}}{T}\right) \\ R_\delta\hat{\delta} \end{pmatrix} - \begin{pmatrix} r_\beta \\ r_\delta \end{pmatrix}\right]'\left[\Gamma_{c_0}(\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1})^{-1}\Gamma_{c_0}'\right]^{-1} \\
&\times\left[\begin{pmatrix} R_\beta\left(\hat{\beta}-\frac{c_0\hat{\delta}}{T}\right) \\ R_\delta\hat{\delta} \end{pmatrix} - \begin{pmatrix} r_\beta \\ r_\delta \end{pmatrix}\right]\Upsilon_T'/p \\
&\Rightarrow \frac{K}{K-2d}\cdot F_{p,K-2d}
\end{aligned} \tag{B.8}
$$

for $p = p_\beta + p_\delta$. Note that we scale $F(\hat{\gamma};c_0)$ by the matrix Υ_T, which reflects the different rates of convergence between $\hat{\beta}$ and $\hat{\delta}$. The result also implies that the joint testing is possible, but the accuracy of testing the restriction $r_\gamma = (r_\beta', r_\delta')'$ will be different from testing the single restriction $R_\beta\beta_0 = r_\beta$. This is mainly because $\hat{\delta}$ is $O_p(1)$ and $\hat{\beta} = O_p(1/T)$ under our fixed-K asymptotics. This means that the component, r_β, is more accurately tested than the component r_δ in the joint inference.

However, the differing rates of convergences between $\hat{\beta}$ and $\hat{\delta}$ can make the joint testing problem very difficult in other types of hypothesis. For example, consider the following simultaneous restrictions on β_0 and δ_0:

$$
H_0^\gamma : R_\beta\beta_0 + R_\delta\delta_0 = r_\beta + r_\delta. \tag{B.9}
$$

In analogous to (B.7), we choose the matrix

$$
\Gamma_{c_0} = \begin{pmatrix} I_{p_\beta} & I_{p_\delta} \end{pmatrix}\begin{pmatrix} R_\beta & -c_0R_\beta \\ 0 & R_\delta \end{pmatrix},
$$

which yields the following joint convergence result:

$$
\Gamma_{c_0}\Upsilon_T\left[\hat{\gamma}-\gamma_0\right] \Rightarrow c_0R_\beta\delta_0 + MN(0,\sigma^2_{0\cdot x}\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}').
$$

Thus, with the plugged-in estimator of $\hat{\beta} - c_0(\hat{\delta}/T)$ and $\hat{\delta}$, we have that

$$
\left(TR_\beta(\hat{\beta}-c_0\hat{\delta}/T) + R_\delta\hat{\delta}\right) - (Tr_\beta + r_\delta) \Rightarrow MN(0,\sigma^2_{0\cdot x}MN(0,\sigma^2_{0\cdot x}\Gamma_{c_0}(\mathbb{S}_X'\mathbb{S}_X)^{-1}\Gamma_{c_0}').
$$

Since the second term on left-hand side, $Tr_\beta + r_\delta$, is different from $r_\beta + r_\delta$, the bias-corrected inference for testing (B.9) is not possible, as long as $r_\beta \asymp r_\delta$, i.e., r_β/r_δ and r_δ/r_β are $O(1)$. This is because the hypothesis in (B.6) consists of only one testing restriction by combining two sets of parameters, r_β and r_δ, which are estimated with different orders of uncertainties under our fixed-K asymptotics. This contrasts to the joint hypothesis in (B.6), which enables us to separate the different orders of uncertainty in $\hat{\beta}$ and $\hat{\delta}$ and make a valid joint inference via (B.9).

For the non-linear hypothesis, we consider $H_0^\beta : g_\beta(\beta_0) = r_\beta$, where the non-linear function $g_\beta(\cdot) : \mathbb{R}^d \to \mathbb{R}^p$ is continuously differentiable at β_0. Then, one can apply the Delta method to convert the non-linear restriction into the linear one in an asymptotic sense. To be more specific, we use a Taylor expansion and obtain that

$$\begin{aligned} T(g_\beta(\hat{\beta}) - r_\beta) &= T(g_\beta(\hat{\beta}) - g_\beta(\beta_0)) \\ &= \frac{\partial g_\beta(\beta_T^*)}{\partial \beta'} T(\hat{\beta} - \beta_0), \end{aligned}$$

where β_T^* lies between $\hat{\beta}$ and β_0. Since $\hat{\beta} \xrightarrow{p} \beta_0$, and $\partial g_\beta(\beta_T^*)/\partial \beta' \xrightarrow{p} \partial g_\beta(\beta_0)/\partial \beta'$, we can apply the Delta method to obtain that

$$T(g_\beta(\hat{\beta}) - r_\beta) \Rightarrow \left(\frac{\partial g_\beta(\beta_0)}{\partial \beta'}\right) \left[\mathbb{S}_x' M_{\mathbb{S}_{\Delta x}} \mathbb{S}_x\right]^{-1} \mathbb{S}_x' M_{\mathbb{S}_{\Delta x}} \mathbb{S}_{0\cdot x} + c \left(\frac{\partial g_\beta(\beta_0)}{\partial \beta'}\right) \delta_0.$$

Therefore, by simply replacing the matrix R_β with $\partial g_\beta(\beta_0)/\partial \beta' \in \mathbb{R}^{p\times d}$, we can extend Theorem 3 to test nonlinear restrictions on β_0. On the other hand, we cannot apply the Delta method to a non-linear testing for δ_0, i.e., $H_0^\delta : g_\delta(\delta_0) = r_\delta$. This is because $\partial g_\delta(\delta_T^*)/\partial \delta' \xrightarrow{p} \partial g_\delta(\delta_0)/\partial \delta'$ does not hold. Instead, we have that $\partial g_\delta(\delta_T^*)/\partial \delta' = \partial g_\delta(\hat{\delta})/\partial \delta' + o_p(1)$, and

$$\frac{\partial g_\delta(\hat{\delta})}{\partial \delta'} \Rightarrow \frac{\partial g_\delta(\delta_0 + \left[\mathbb{S}_{\Delta x}' M_{\mathbb{S}_X} \mathbb{S}_{\Delta x}\right]^{-1} \mathbb{S}_{\Delta x}' M_{\mathbb{S}_X} \mathbb{S}_{0\cdot x})}{\partial \delta'} \tag{B.10}$$

by continuous mapping theorem. Other than relying on the Delta method, one can make use of a proper simulation-based or resampling method and directly compute a critical value from the non-standard distribution implied by (B.10). Still, it is necessary to investigate further its feasibility, which deals with the unknown nuisance parameters, $\sigma_{0\cdot x}^2$ and Ω_{xx}, in the right-hand side of (B.10). This remains an interesting open question, and we leave for future research.

S.3 Construction of Confidence Interval for $\mathbf{c}_0$ under Dependent Errors

We consider the following autoregressive model:

$$x_t = \mu_x + \rho x_{t-1} + u_{xt}, \tag{B.11}$$

where x_t is a scalar time series, $|\rho| \leq 1$, and $\mu_x = 0$. The serial dependence of u_{xt} has important roles when we construct valid asymptotic confidence sets for ρ. To be specific, Elliott and Stock's (2001) approach requires a consistent estimation of the LRV, Ω_{xx}, in their test statistic, $P_T(0, \bar{c})$. Also, as shown in our numerical results in Tables S.1 and S.2, the confidence interval (CI) constructed by grid bootstrap approach in Hansen (1999) is invalid when we ignore the serial dependence of u_{xt}. To address these issues, we first approximate the unknown dependence structure of u_{xt} by the following AR(p_T) process:

$$u_{xt} = b_1 u_{xt-1} + \ldots + b_{p_T} u_{xt-p_T} + e_t, \tag{B.12}$$

where $e_t \overset{i.i.d}{\sim} (0, \sigma_e^2)$. Under suitable regular conditions, one can translate the true dependence (covariance) structure of weakly stationary process to an infinite order autoregressive (AR) process, e.g., den Haan and Levin (1998). Thus, we can justify the approximated process in (B.12) by assuming that the selected lag order grows to infinity as $T \to \infty$ such that $p_T^3/T = o(1)$ (Berk, 1974).

Using (B.12), we can reformulate (B.11) by the following (approximated) augmented Dickey–Fuller (ADF) equation:

$$\Delta x_t = \tau + (\varphi - 1)x_{t-1} + \sum_{i=1}^{p_T} \pi_i \Delta x_{t-i} + e_t, \tag{B.13}$$

where $\tau = \mu_x(1 - \sum_{i=1}^{p_T} b_i)$, $\varphi := 1 + (\rho - 1)\left(1 - \sum_{i=1}^{p_T} b_i\right)$, and $\pi_i = b_i - (1 - \rho)\sum_{j=i}^{p_T^*} b_j$ for $i = 1, \ldots, p_T$. Regarding the choice of p_T, one can use Bayesian information criterion (BIC) to the ADF equation in (B.13), e.g. Campbell and Yogo (2006). However, Ng and Perron (1995) and Lopez (1997) point out that selection rules such as BIC and Akaike information criterion (AIC) tend to select values of p_T that are generally too small for inferring the autoregressive coefficient $(\varphi - 1)$ in (B.13) with persistent data. To resolve this issue, Ng and Perron (2001) propose a modified AIC (MAIC) selection rule, which takes account of the bias in the sum of the regression coefficient for the persistent regressor x_{t-1}. The MAIC rule is implemented as below:

$$p_T^* = \arg\min_k \left\{ \log(\hat{\sigma}_k^2) + \frac{2(\tau_T(k) + k)}{T - k_{\max}} \right\},$$

where

$$\hat{\sigma}_k^2 = \frac{\sum_{t=k_{\max}+1}^{T} \hat{e}_{tk}^2}{(T - k_{\max})} \text{ and } \tau_T(k) = \frac{\hat{\lambda}^2}{\hat{\sigma}_k^2} \sum_{t=k_{\max}+1}^{T} x_{t-1}.$$

$\hat{\lambda}$ is the estimated value for the coefficient $(\varphi - 1)$, and $\hat{e}_{tk}$'s are residuals of the following ADF regressions:

$$\Delta x_t = \tau + (\varphi - 1)x_{t-1} + \sum_{i=1}^{k} \psi_i \Delta x_{t-i} + e_{tk}.$$

Modified grid bootstrap method in Hansen (1999)

Recall that, given p_T^*, we can re-formulate (B.11) by the following (approximated) augmented Dickey–Fuller (ADF) equation:

$$x_t = \tau + \varphi x_{t-1} + \sum_{i=1}^{p_T^*} \pi_i \Delta x_{t-i} + e_t, \tag{B.14}$$

Let $\hat{\varphi}$ denote the estimate of φ from OLS estimation of (B.14). Let $\{\hat{e}_t\}$ be the OLS residuals in (B.14) and s.e$(\hat{\varphi})$ be the least square standard error for $\hat{\varphi}$. The corresponding t-statistic, under φ, is then defined by $t(\hat{\varphi}; \varphi) = (\hat{\varphi} - \varphi)/\text{s.e}(\hat{\varphi})$. Also, given φ, we also compute $\hat{\tau}(\varphi)$ and $\hat{\pi}_i(\varphi)$ for $i = 1, \ldots, p$, which is the OLS estimates in the following restricted regression model:

$$x_t - \varphi x_{t-1} = \tau + \sum_{i=1}^{p_T^*} \pi_i \Delta x_{t-i} + e_t. \tag{B.15}$$

Below we describe an algorithm to compute the CI for the coefficient φ with $100 \cdot (1-\eta)\%$ coverage rate using the grid-bootstrap approach in Hansen (1999).

Step 1: Given φ in the grid, generate bootstrapped residuals $\{e_t^*\}_{t=1}^T$. According to Mikusheva (2007), there are at least two ways to perform the nonparametric grid bootstrap. The first method is drawn from the empirical distribution of the residuals $\{\hat{e}_t\}_{t=1}^T$ from (B.14). The second one is to draw from the centered residuals, $\{\hat{e}_t(\varphi) - T^{-1}\sum_{t=1}^T \hat{e}_t(\varphi)\}_{t=1}^T$, where $\hat{e}_t(\varphi)$ is obtained by OLS regression in (B.15), i.e.,

$$\hat{e}_t(\varphi) := x_t - \varphi x_{t-1} - \hat{\tau}(\varphi) - \sum_{i=1}^{p_T^*} \hat{\pi}_i(\varphi)\Delta x_{t-i}.$$

Our Monte Carlo analysis only reports the results using the first method, as the second one yields very similar outputs.

Step 2: Using $\{e_t^*\}_{t=1}^T$, generate the bootstrapped samples as below:

$$x_t^* = \hat{\tau}(\varphi) + \theta x_{t-1}^* + \sum_{i=1}^{p_T^*} \hat{\pi}_i(\varphi)\Delta x_{t-i}^* + e_t^*$$

where $(x_0^*, x_{-1}^*, \ldots, x_{-p_T^*}^*)$ is set to be zero vector, or $(x_0, x_{-1}, \ldots, x_{-p_T^*})$.

Step 3: Using the bootstrapped samples, compute the bootstrapped t-statistic under θ:

$$t^*(\hat{\varphi}^*; \varphi) = \frac{\hat{\varphi}^* - \varphi}{\text{s.e}(\hat{\varphi}^*)},$$

where $\hat{\varphi}^*$ is from the OLS estimation of (B.15) using bootstrapped samples $\{x_t^*\}_{t=1}^T$, and s.e$(\hat{\varphi}^*)$ is corresponding OLS standard error estimated from the bootstrapped samples. Given φ in the grid chosen in the Step 1, repeat Steps 1–3, B-times for some large number of B, say $B = 200$.

Step 4: Based on the B-number of the bootstrapped t-statistic, $t^*(\hat{\varphi}^*; \varphi)$, compute the $\eta/2$ and the $(1-\eta/2)$ quantiles of $t^*(\varphi)$, which are denoted $q_T^*(\eta/2, \varphi)$ and $q_T^*(1-\eta/2, \varphi)$, respectively.

Step 5: Repeat Steps 1–4 for different grid points of φ, and draw the quantile curves of $q_T^*(\eta/2, \varphi)$ and $q_T^*(1-\eta/2, \varphi)$ with respect to φ.

Step 6: Using the OLS estimate, $\hat{\varphi}$, based on the original sample $\{x_t\}_{t=1}^T$, construct the CI for the parameter φ as a set of values for which the corresponding hypothesis is not reject at $100\eta\%$, i.e., .

$$S_{T,\varphi}(\eta) = \{\varphi : q_T^*(\eta/2, \varphi) \le t(\hat{\varphi}; \varphi) \le q_T^*(1-\eta/2, \varphi)\}.$$

Step 7: After we construct the uniform confidence set of θ, we transform it to the CI for c_0 as below:

$$S_T(\eta) = \left\{ c = T(1-\rho) : \rho = 1 + \left(\frac{\varphi - 1}{1 - \sum_{i=1}^{p_T^*} \hat{b}_i} \right) \text{ such that } \varphi \in S_{T,\varphi}(\eta) \right\},$$

where $\hat{b}_i$'s are estimated by the following OLS regression

$$\hat{u}_{xt} = b_1 \hat{u}_{xt-1} + \ldots + b_{p_T^*} \hat{u}_{xt-p_T^*} + e_t$$

with $\hat{u}_{xt} = (x_t - T^{-1}\sum_{t=1}^{T} x_t) - \hat{\rho}_{\text{ols}}(x_{t-1} - T^{-1}\sum_{t=1}^{T} x_t)$.

Inversion of efficient tests in Elliott and Stock (2001)

Considering (B.11) with $\rho = 1 - c_0/T$, a confidence interval for the local-to-unity parameter c_0, proposed by Elliott and Stock (2001), builds on the idea of inverting asymptotically optimal Neyman-Pearson tests in the Gaussian autoregression model. Below we describe a procedure to compute the CI of Elliott and Stock (2001) for c_0 with $100(1-\eta)\%$ coverage rate.

Step 1: Obtain a heteroskedasticity autocorrelation robust estimator of Ω_{xx}, which is based on the approximated process in (B.14):

$$\hat{\Omega}_{xx} = \frac{\hat{\sigma}_e^2}{\left(1 - \sum_{i=1}^{p_T^*} \hat{b}_i\right)^2},$$

where $\hat{\sigma}_e^2 = T^{-1}\sum_{t=1}^{T} \hat{e}_t$ and $\hat{e}_t = \hat{u}_{xt} - (\hat{b}_1 \hat{u}_{xt-1} + \ldots + \hat{b}_p \hat{u}_{xt-p})$. Note that one can also estimate $\hat{\Omega}_{xx}$ by the nonparametric kernel type HAC estimation, e.g., Newey and West (1987), using an optimal bandwidth rule suggested by Andrews (1991).

Step 2: Following Elliott and Stock (2001, pp161), we choose $\bar{c} = 13.5$ with $\bar{\rho} = 1 - \bar{c}/T$, and construct the following test statistics:

$$P_T(0, \bar{c}) := \frac{1}{\hat{\Omega}_{xx}} \left[\sum_{t=1}^{T} (u_{\text{GLS},t}(\bar{\rho}))^2 - \bar{\rho} \sum_{t=1}^{T} (u_{\text{GLS},t}(1))^2 \right],$$

where $u_{\text{GLS},t}(\rho) = x_t(\rho) - z_t(\rho)'\beta(\rho)$ for $t = 1, ..., T$, and

$$\beta(\rho) = (Z'(\rho)Z(\rho))^{-1}Z'(\rho)X(\rho);$$

$$Z(\rho) = \begin{bmatrix} z_1(\rho) \\ z_2(\rho) \\ \vdots \\ z_T(\rho) \end{bmatrix} = \begin{bmatrix} 1 \\ 1-\rho \\ \vdots \\ 1-\rho \end{bmatrix} \text{ and } X(\rho) = \begin{bmatrix} x_1(\rho) \\ x_2(\rho) \\ \vdots \\ x_T(\rho) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_T - \rho x_{T-1} \end{bmatrix}.$$

Step 3: Let $W(\cdot)$ be a standard Wiener process, and $J_c(\cdot)$ be OU-process $J_c(r) = \int_0^r \exp(-c(r-s))dW(s)$. Given c in the grid, we obtain the following two quantities:

$$p(c, \epsilon_1) : \epsilon_1 \text{ quantile of } P(c, \bar{c});$$
$$p(c, 1-\epsilon_2) : 1-\epsilon_2 \text{ quantile of } P(c, \bar{c}),$$

where $\eta = \epsilon_1 + \epsilon_2$, and

$$P(c, \bar{c}) = \bar{c}^2 \int_0^1 (J_c(s))^2 \, ds + \bar{c} J_c^2(1).$$

For $\eta = 0.10$, we choose $\epsilon_1 = 0.06$ and $\epsilon_2 = 0.04$, which are suggested in Elliott and Stock (2001). To simulate $p(c, \epsilon_1)$ and $p(c, 1-\epsilon_2)$, we draw the random variable, $\hat{P}_{B_1}(c, \bar{c})$, with B_2

times:

$$\hat{P}_{B_1}(c,\bar{c}) := \frac{\bar{c}^2}{B_1}\sum_{b=1}^{B_1}\left(\hat{J}_c\left(\frac{b}{B_1}\right)\right)^2 + \bar{c}\left(\hat{J}_c(1)\right)^2,$$
$$\hat{J}_c\left(\frac{s}{B_1}\right) := \frac{1}{\sqrt{B_1}}\sum_{b=1}^{s}\exp\left(c\left(\frac{s-b}{B_1}\right)\right)e_b,$$

where $e_b \overset{i.i.d}{\sim} N(0,1)$, and B_1 and B_2 are large numbers, say, $B_1 = 500$ and $B_2 = 5000$. Then, $p(c,\epsilon_1)$ and $p(c, 1-\epsilon_2)$ can be obtained by ϵ_1 and $1-\epsilon_2$ quantiles of $\hat{P}_{B_1}(c,\bar{c})$, respectively.

Step 4: Construct the CI for parameter c, $S_T(\eta)$, which is a set of values for which the corresponding hypothesis, $H_0 : \rho = 1 - c/T$, is not rejected at $100\eta\%$, i.e.,

$$S_T(\eta) = \{c : p(c,\epsilon_1) \le P_T(0,\bar{c}) \le p(c, 1-\epsilon_2)\}.$$

Note that the above definition allows the possibility of disconnected sets. In this case, we use a conservative confidence interval which can be defined as the convex hull of $S_T(\eta)$.

S.4 Computation of Bonferroni Intervals

Recall that the upper and lower bounds of Bonferroni CI in (40) are maximum and minimum of

$$r_{\beta,l}^{1-\alpha/2}(c) = R_\beta\hat{\beta} - \frac{1}{T}\left(cR_\beta\hat{\delta} + \sqrt{\hat{\sigma}_{0\cdot x}^2 D(c)}\cdot\sqrt{\frac{K}{K-2d}}\cdot t_{K-2d}^{1-\alpha/2}\right), \tag{B.16}$$

and

$$r_{\beta,h}^{1-\alpha/2}(c) = R_\beta\hat{\beta} - \frac{1}{T}\left(cR_\beta\hat{\delta} - \sqrt{\hat{\sigma}_{0\cdot x}^2 D(c)}\cdot\sqrt{\frac{K}{K-2d}}\cdot t_{K-2d}^{1-\alpha/2}\right), \tag{B.17}$$

respectively, where

$$\begin{aligned} D(c) &= \Gamma_{c_0}\left[\Upsilon_T^{-1}\mathbb{W}_X'\mathbb{W}_X\Upsilon_T^{-1}\right]^{-1}\Gamma_{c_0}' \\ &= R_\beta\left[\Lambda_1(c)\left(\mathbb{W}_x' M_{\Delta x}\mathbb{W}_x\right)^{-1} + \Lambda_2(c)\left(\mathbb{W}_{\Delta x}' M_x \mathbb{W}_{\Delta x}\right)^{-1}\right]R_\beta', \end{aligned}$$

and

$$\begin{aligned} \Lambda_1(c) &= T^2(I_d + cT^{-1}\left[\mathbb{W}_{\Delta x}'\mathbb{W}_{\Delta x}\right]^{-1}\mathbb{W}_{\Delta x}'\mathbb{W}_x), \\ \Lambda_2(c) &= c^2 I_d + cT\left[\mathbb{W}_x'\mathbb{W}_x\right]^{-1}\mathbb{W}_x'\mathbb{W}_{\Delta x}. \end{aligned}$$

Let $\underline{c}$ and $\bar{c}$ denote the minimum and maximum values of $S_T(\eta)$, respectively. Below we characterize conditions that guarantee that $r_{R,l}^{1-\alpha}(c)$ ($r_{R,h}^{1-\alpha}(c)$) is monotone in c.

i) When $R_\beta\hat{\delta} \ge 0$, $(\mathbb{W}_{\Delta x}'\mathbb{W}_{\Delta x}^{-1})\mathbb{W}_{\Delta x}'\mathbb{W}_x \ge 0$, and $\left[\mathbb{W}_x'\mathbb{W}_x\right]^{-1}\mathbb{W}_x'\mathbb{W}_{\Delta x} \ge 0$:

   Both $cR_\beta\hat{\delta}$ and $\sqrt{D(c)}$ in (B.16) are increasing in c for $c \ge 0$, so $r_{\beta,l}^{1-\alpha/2}(c)$ is decreasing in

$c \geq 0$. This leads us to compute the lower bound of the Bonferroni CI as follows:

$$\min_{c \in S_T(\eta)} r_{\beta,l}^{1-\alpha/2}(c) = r_{\beta,l}^{1-\alpha/2}(\underline{c}).$$

ii) When $R_\beta \hat{\delta} \leq 0$, $(\mathbb{W}'_{\Delta x}\mathbb{W}_{\Delta x}^{-1})\mathbb{W}'_{\Delta x}\mathbb{W}_x \geq 0$, and $[\mathbb{W}'_x\mathbb{W}_x]^{-1}\mathbb{W}'_x\mathbb{W}_{\Delta x} \geq 0$:

Both $cR_\beta\hat{\delta}$ and $-\sqrt{D(c)}$ in (B.17) are decreasing in c for $c \geq 0$, so $r_{\beta,h}^{1-\alpha/2}(c)$ is increasing in $c \geq 0$. This leads us to compute the upper bound of the Bonferroni CI as follows:

$$\max_{c \in S_T(\eta)} r_{\beta,h}^{1-\alpha/2}(c) = r_{\beta,h}^{1-\alpha/2}(\overline{c}).$$

S.5 Tables and Figures

Table S.1: Empirical coverage rates and averaged estimates of 90% CIs for autoregressive parameter $\rho \in \{0.975, 0.950, 0.90\}$ using various methods with $T = 200$, AR(1) error, and $\psi \in \{0.00, 0.25, 0.50, 0.75\}$

AR(1) process for autoregressive error:						
$\rho_T = 0.975$ with $c_0 = 5$ and $T = 200$						
	`Hansen (1999)`		`Modified Hansen (1999)`		`Elliott and Stock (2001)`	
ψ	Coverage	Averaged CI	Coverage	Averaged CI	Coverage	Averaged CI
0.00	0.898	[0.930, 0.997]	0.892	[0.926, 0.997]	0.896	[0.932, 0.994]
0.25	0.777	[0.957, 1.000]	0.891	[0.928, 0.997]	0.906	[0.934, 0.995]
0.50	0.413	[0.977, 1.000]	0.891	[0.929, 0.997]	0.897	[0.933, 0.994]
0.75	0.066	[0.991, 1.000]	0.883	[0.931, 0.997]	0.895	[0.934, 0.994]
$\rho_T = 0.950$ with $c_0 = 10$ and $T = 200$						
	`Hansen (1999)`		`Modified Hansen (1999)`		`Elliott and Stock (2001)`	
ψ	Coverage	Averaged CI	Coverage	Averaged CI	Coverage	Averaged CI
0.00	0.896	[0.899, 0.990]	0.878	[0.895, 0.990]	0.855	[0.904, 0.985]
0.25	0.674	[0.937, 0.999]	0.875	[0.898, 0.990]	0.852	[0.907, 0.987]
0.50	0.191	[0.965, 1.000]	0.866	[0.902, 0.991]	0.852	[0.907, 0.986]
0.75	0.004	[0.986, 1.000]	0.841	[0.907, 0.992]	0.837	[0.909, 0.987]
$\rho_T = 0.90$ with $c_0 = 20$ and $T = 200$						
	`Hansen (1999)`		`Modified Hansen (1999)`		`Elliott and Stock (2001)`	
ψ	Coverage	Averaged CI	Coverage	Averaged CI	Coverage	Averaged CI
0.00	0.896	[0.839, 0.957]	0.856	[0.832, 0.958]	0.747	[0.854, 0.961]
0.25	0.515	[0.896, 0.992]	0.846	[0.842, 0.962]	0.727	[0.861, 0.966]
0.50	0.041	[0.939, 1.000]	0.817	[0.852, 0.967]	0.707	[0.866, 0.967]
0.75	0.000	[0.974, 1.000]	0.720	[0.869, 0.977]	0.651	[0.875, 0.972]

Table S.2: Empirical coverage rates and averaged estimates of 90% CIs for autoregressive parameter $\rho \in \{0.975, 0.950, 0.90\}$ using various methods with $T = 200$, MA(1) error, and $\psi \in \{0.00, 0.25, 0.50, 0.75\}$

MA(1) process for autoregressive error: $\rho_T = 0.975$ with $c_0 = 5$ and $T = 200$						
	`Hansen (1999)`		`Modified Hansen (1999)`		`Elliott and Stock (2001)`	
ψ	Coverage	Averaged CI	Coverage	Averaged CI	Coverage	Averaged CI
0.00	0.898	[0.930, 0.997]	0.892	[0.926, 0.997]	0.896	[0.932, 0.994]
0.25	0.832	[0.952, 1.000]	0.889	[0.926, 0.996]	0.903	[0.932, 0.994]
0.50	0.734	[0.961, 1.000]	0.886	[0.929, 0.997]	0.895	[0.936, 0.995]
0.75	0.678	[0.964, 1.000]	0.880	[0.934, 0.998]	0.904	[0.941, 0.997]
$\rho_T = 0.95$ with $c_0 = 10$ and $T = 200$						
	`Hansen (1999)`		`Modified Hansen (1999)`		`Elliott and Stock (2001)`	
ψ	Coverage	Averaged CI	Coverage	Averaged CI	Coverage	Averaged CI
0.00	0.896	[0.899, 0.990]	0.878	[0.895, 0.990]	0.855	[0.904, 0.985]
0.25	0.762	[0.929, 0.998]	0.871	[0.895, 0.989]	0.850	[0.905, 0.986]
0.50	0.600	[0.942, 1.000]	0.864	[0.900, 0.991]	0.837	[0.911, 0.988]
0.75	0.519	[0.947, 1.000]	0.839	[0.908, 0.994]	0.797	[0.921, 0.992]
$\rho_T = 0.90$ with $c_0 = 20$ and $T = 200$						
	`Hansen (1999)`		`Modified Hansen (1999)`		`Elliott and Stock (2001)`	
ψ	Coverage	Averaged CI	Coverage	Averaged CI	Coverage	Averaged CI
0.00	0.896	[0.839, 0.957]	0.856	[0.832, 0.958]	0.747	[0.854, 0.961]
0.25	0.655	[0.885, 0.987]	0.842	[0.836, 0.958]	0.736	[0.857, 0.964]
0.50	0.395	[0.905, 0.995]	0.813	[0.847, 0.966]	0.669	[0.874, 0.971]
0.75	0.285	[0.913, 0.997]	0.757	[0.861, 0.976]	0.531	[0.891, 0.980]

Table S.3: Empirical size of 5% various TA-OLS methods with $T = 200$, $K = 8$ and AR(1) error with $\psi = 0.75$ with a single regressor.

$H_0 : \beta_1 = 1$ with $c_0 = 0$, $\psi = 0.75$, and $K = 8$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.055	0.055	0.228	0.036	0.056	0.074
0.25	0.055	0.055	0.226	0.042	0.058	0.052
0.50	0.054	0.054	0.226	0.042	0.058	0.031
0.75	0.054	0.054	0.236	0.044	0.058	0.014

$H_0 : \beta_1 = 1$ with $c_0 = 5$, $\psi = 0.75$, and $K = 8$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.057	0.055	0.148	0.026	0.027	0.094
0.25	0.092	0.055	0.162	0.020	0.022	0.077
0.50	0.172	0.056	0.200	0.016	0.017	0.051
0.75	0.370	0.055	0.300	0.014	0.013	0.029

$H_0 : \beta_1 = 1$ with $c_0 = 10$, $\psi = 0.75$, and $K = 8$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.057	0.054	0.085	0.022	0.027	0.101
0.25	0.143	0.055	0.135	0.020	0.025	0.104
0.50	0.301	0.056	0.234	0.019	0.021	0.105
0.75	0.637	0.054	0.454	0.021	0.022	0.106

$H_0 : \beta_1 = 1$ with $c_0 = 20$, $\psi = 0.75$, and $K = 8$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.057	0.050	0.046	0.023	0.029	0.104
0.25	0.202	0.052	0.136	0.028	0.032	0.179
0.50	0.458	0.053	0.297	0.032	0.037	0.280
0.75	0.825	0.056	0.592	0.045	0.061	0.409

Table S.4: Empirical size of 5% various TA-OLS methods with $T = 200$, $K = 16$ and AR(1) error with $\psi = 0.50$ with a single regressor.

$H_0 : \beta_1 = 1$ with $c_0 = 0$, $\psi = 0.50$, and $K = 16$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.054	0.054	0.134	0.040	0.049	0.063
0.25	0.053	0.053	0.129	0.045	0.049	0.045
0.50	0.053	0.053	0.126	0.043	0.049	0.029
0.75	0.055	0.055	0.130	0.048	0.050	0.016
$H_0 : \beta_1 = 1$ with $c_0 = 5$, $\psi = 0.50$, and $K = 16$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.054	0.055	0.064	0.028	0.031	0.085
0.25	0.101	0.051	0.064	0.016	0.018	0.066
0.50	0.209	0.052	0.086	0.014	0.013	0.048
0.75	0.467	0.052	0.149	0.017	0.014	0.033
$H_0 : \beta_1 = 1$ with $c_0 = 10$, $\psi = 0.50$, and $K = 16$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.055	0.053	0.040	0.024	0.028	0.093
0.25	0.173	0.050	0.069	0.017	0.019	0.083
0.50	0.406	0.051	0.127	0.015	0.016	0.080
0.75	0.788	0.053	0.260	0.018	0.022	0.075
$H_0 : \beta_1 = 1$ with $c_0 = 20$, $\psi = 0.50$, and $K = 16$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.056	0.050	0.032	0.025	0.028	0.104
0.25	0.296	0.049	0.091	0.020	0.029	0.142
0.50	0.656	0.051	0.202	0.022	0.038	0.191
0.75	0.956	0.058	0.420	0.032	0.072	0.247

Table S.5: Empirical size of 5% various TA-OLS methods with $T = 200$, $K = 24$ and AR(1) error with $\psi = 0.25$ with a single regressor.

$H_0 : \beta_1 = 1$ with $c_0 = 0$, $\psi = 0.25$, and $K = 24$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.052	0.052	0.075	0.043	0.047	0.052
0.25	0.053	0.053	0.072	0.046	0.046	0.038
0.50	0.052	0.052	0.071	0.042	0.045	0.030
0.75	0.051	0.051	0.071	0.049	0.046	0.019
$H_0 : \beta_1 = 1$ with $c_0 = 5$, $\psi = 0.25$, and $K = 24$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.053	0.053	0.040	0.028	0.030	0.074
0.25	0.100	0.049	0.028	0.015	0.017	0.056
0.50	0.217	0.050	0.030	0.012	0.014	0.046
0.75	0.493	0.052	0.050	0.018	0.015	0.035
$H_0 : \beta_1 = 1$ with $c_0 = 10$, $\psi = 0.25$, and $K = 24$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.054	0.051	0.032	0.025	0.028	0.081
0.25	0.183	0.050	0.029	0.014	0.018	0.073
0.50	0.435	0.052	0.041	0.012	0.016	0.072
0.75	0.828	0.054	0.075	0.020	0.022	0.066
$H_0 : \beta_1 = 1$ with $c_0 = 20$, $\psi = 0.25$, and $K = 24$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	(M-Hansen)	(ES)	IVX
0	0.053	0.049	0.028	0.026	0.029	0.091
0.25	0.333	0.051	0.042	0.018	0.030	0.129
0.50	0.722	0.055	0.065	0.018	0.038	0.170
0.75	0.975	0.065	0.128	0.027	0.071	0.215

Table S.6: Empirical size of 5% various TA-OLS methods with $T = 200$, $K = 8, 16, 24$ and AR(1) error with $\psi \in \{0.75, 0.50, 0.25\}$ with two regressors.

$H_0 : \beta_1 = \beta_2$ with $(c_{0,1}, c_{0,2}) = (20, 0)$, $\psi = 0.75$ and $K = 8$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	Bonf-M-TAOLS (M-Hansen)	Bonf-M-TAOLS (ES)	IVX
0	0.042	0.045	0.177	0.016	0.026	0.100
0.25	0.074	0.049	0.178	0.022	0.035	0.167
0.50	0.092	0.048	0.184	0.023	0.032	0.279
0.75	0.118	0.049	0.198	0.019	0.028	0.494
$H_0 : \beta_1 = \beta_2$ with $(c_{0,1}, c_{0,2}) = (20, 0)$, $\psi = 0.50$ and $K = 16$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	Bonf-M-TAOLS (M-Hansen)	Bonf-M-TAOLS (ES)	IVX
0	0.053	0.052	0.086	0.023	0.029	0.098
0.25	0.100	0.049	0.097	0.020	0.029	0.130
0.50	0.154	0.047	0.103	0.019	0.025	0.183
0.75	0.253	0.058	0.132	0.019	0.030	0.313
$H_0 : \beta_1 = \beta_2$ with $(c_{0,1}, c_{0,2}) = (20, 0)$, $\psi = 0.25$ and $K = 24$						
r^2	TAOLS	M-TAOLS (Infeasible)	Bonf-M-TAOLS (Hansen)	Bonf-M-TAOLS (M-Hansen)	Bonf-M-TAOLS (ES)	IVX
0	0.047	0.051	0.040	0.024	0.030	0.074
0.25	0.107	0.056	0.044	0.023	0.029	0.096
0.50	0.191	0.052	0.048	0.018	0.025	0.154
0.75	0.311	0.060	0.056	0.015	0.029	0.264

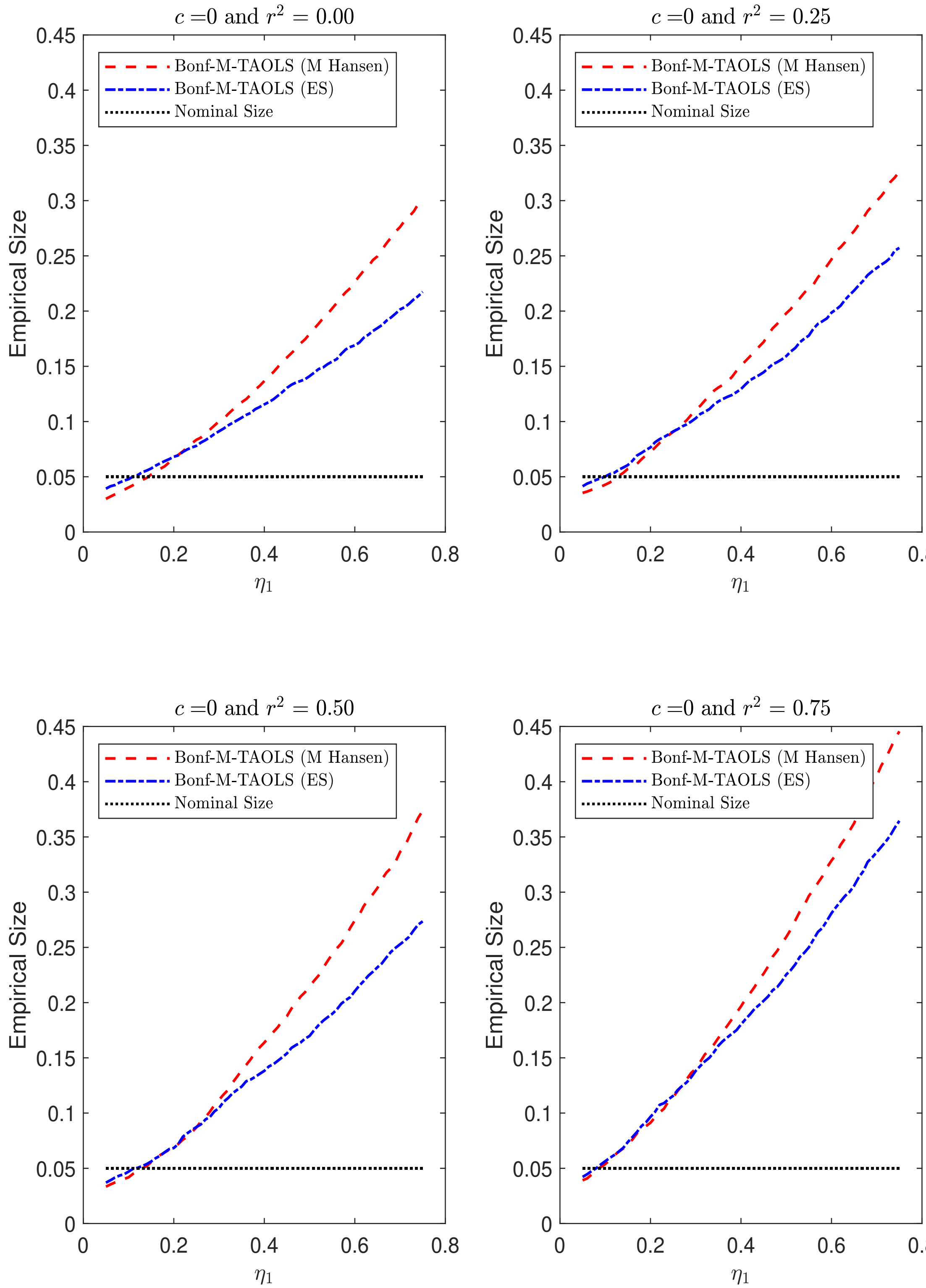

Figure S.1: Empirical sizes of `Bonf M TAOLS (M Hansen)` and `Bonf M TAOLS (ES)` for different values of tuning parameters η_1 and $\eta_2 = 0.05$ when $c_0 = 0$

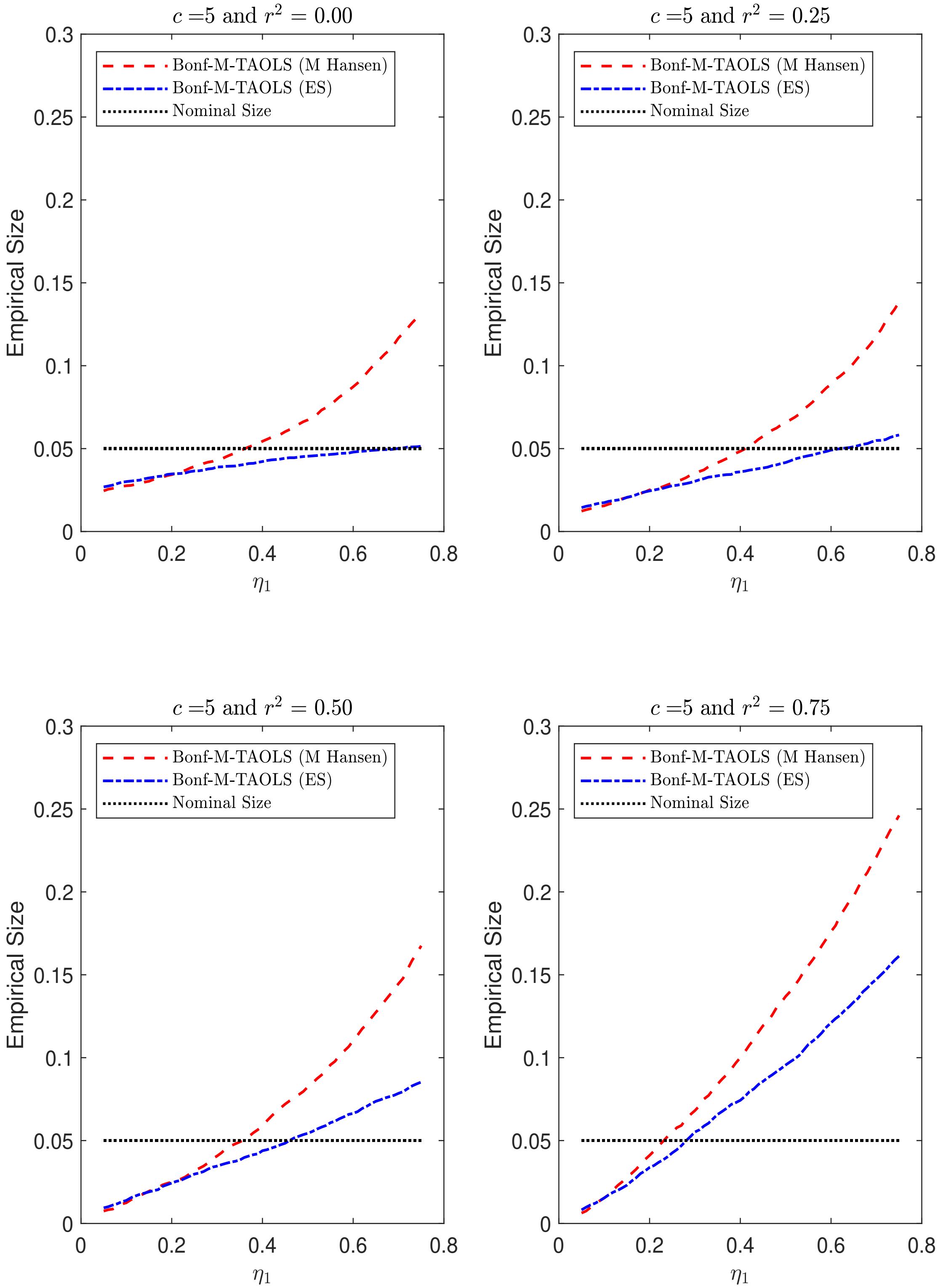

Figure S.2: Empirical sizes of `Bonf M TAOLS (M Hansen)` and `Bonf M TAOLS (ES)` for different values of tuning parameters η_1 and $\eta_2 = 0.05$ when $c_0 = 5$

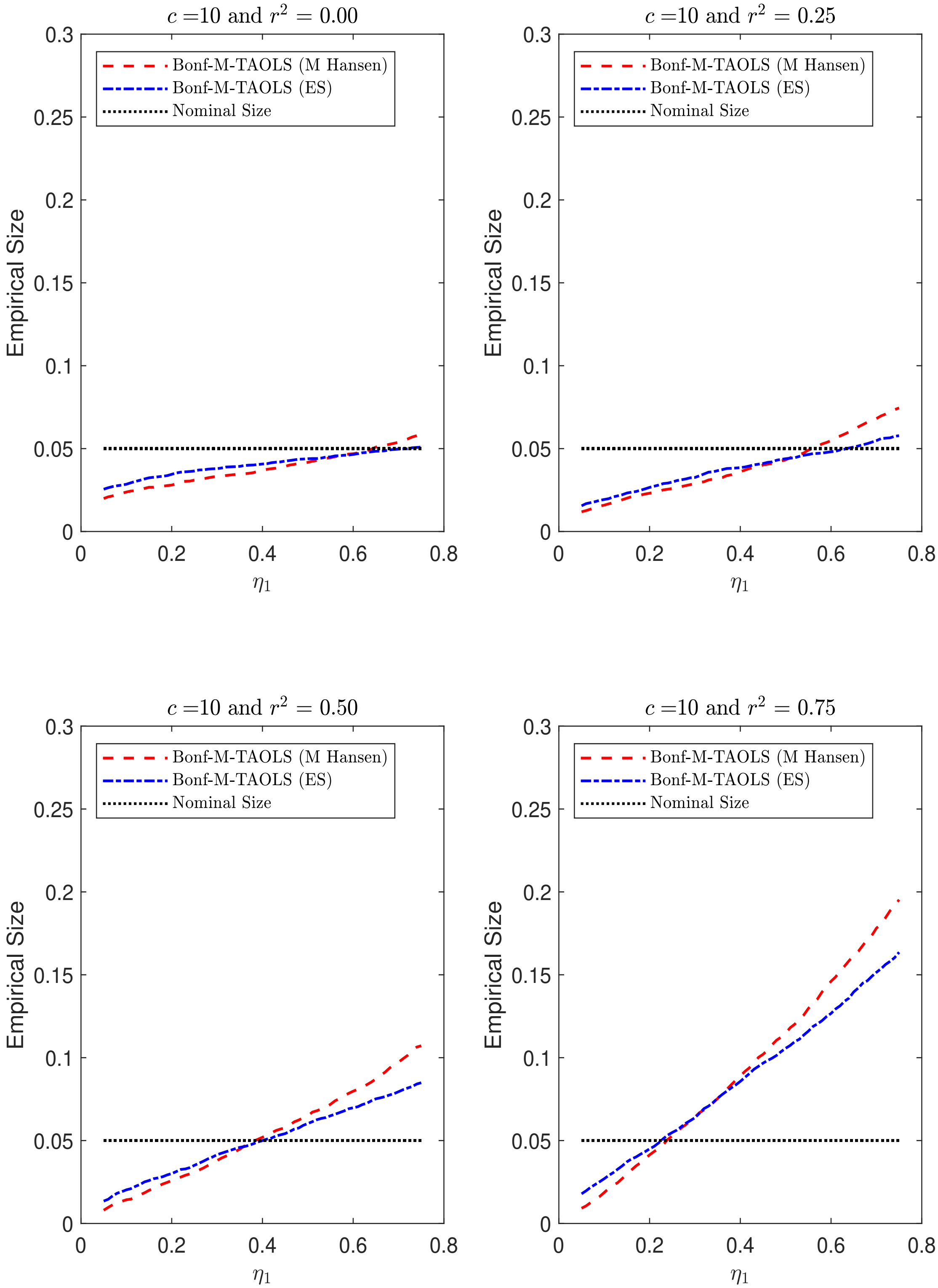

Figure S.3: Empirical sizes of `Bonf M TAOLS (M Hansen)` and `Bonf M TAOLS (ES)` for different values of tuning parameters η_1 and $\eta_2 = 0.05$ when $c_0 = 10$

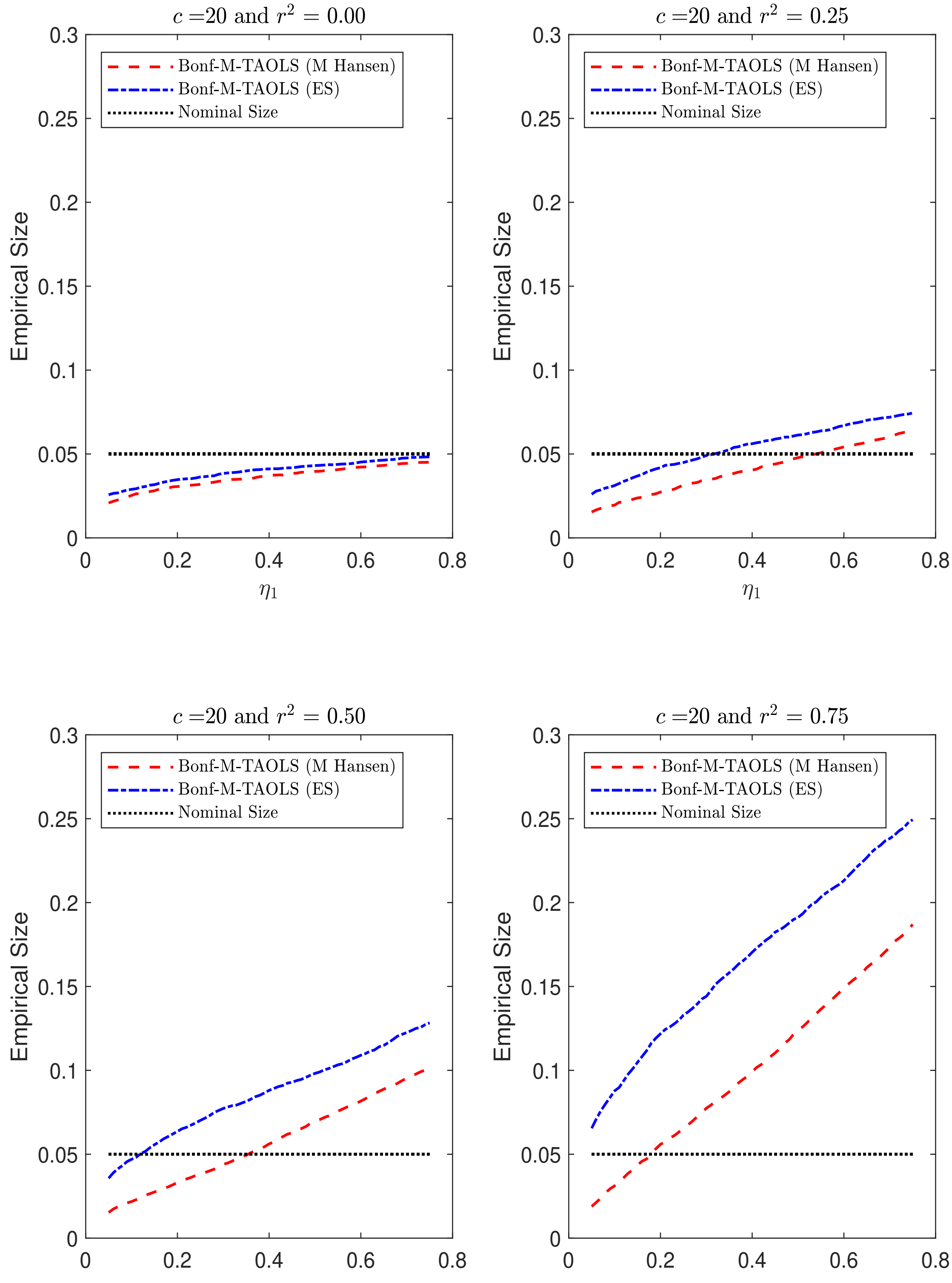

Figure S.4: Empirical sizes of `Bonf M TAOLS (M Hansen)` and `Bonf M TAOLS (ES)` for different values of tuning parameters η_1 and $\eta_2 = 0.05$ when $c_0 = 20$

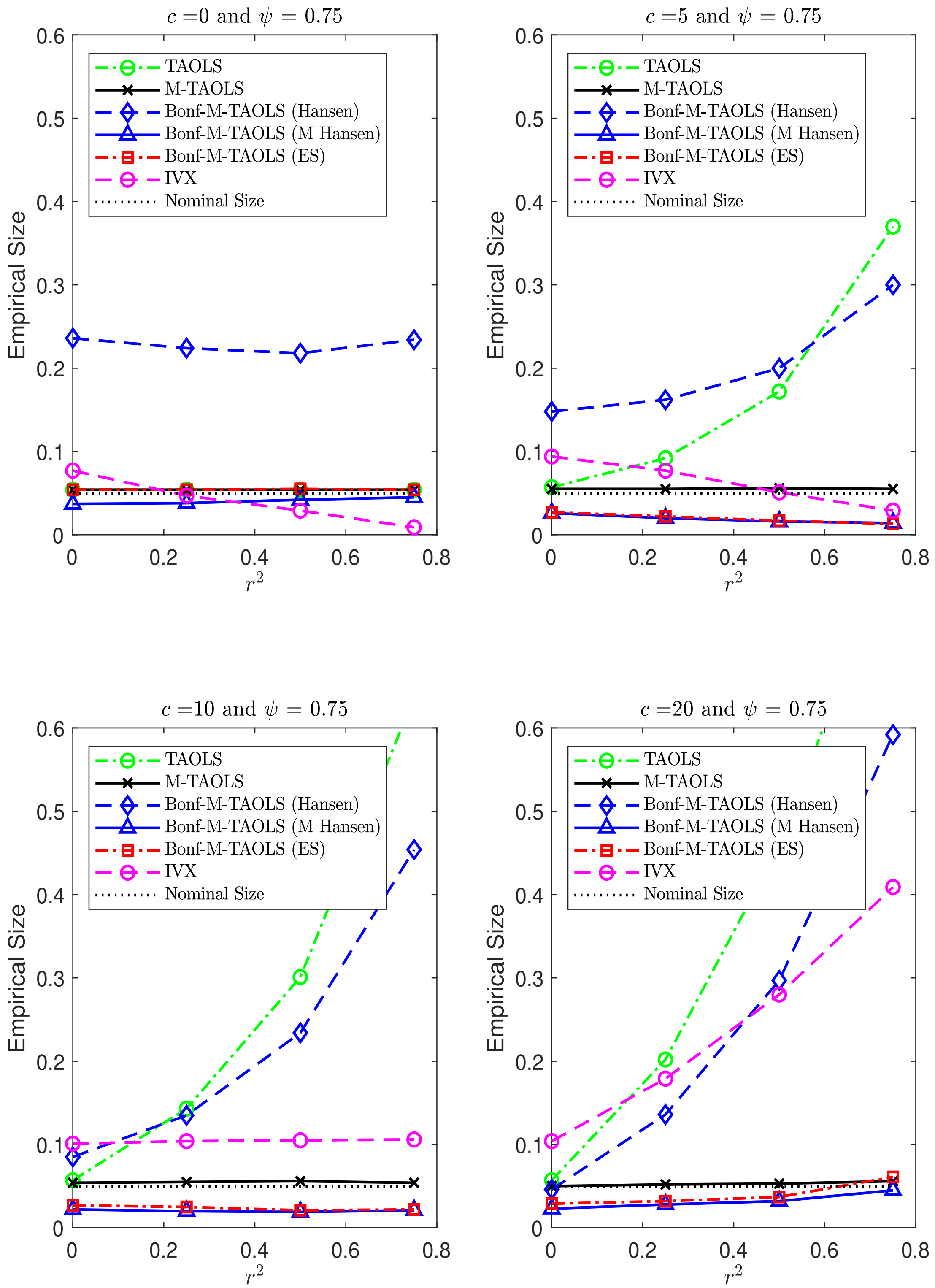

Figure S.5: Empirical sizes of `TAOLS`, `M-TAOLS`, `Bonf-M-TAOLS (Hansen)`, `Bonf-M-TAOLS (Hansen)`, `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with a single regressor, $K = 8$, and $\psi = 0.75$

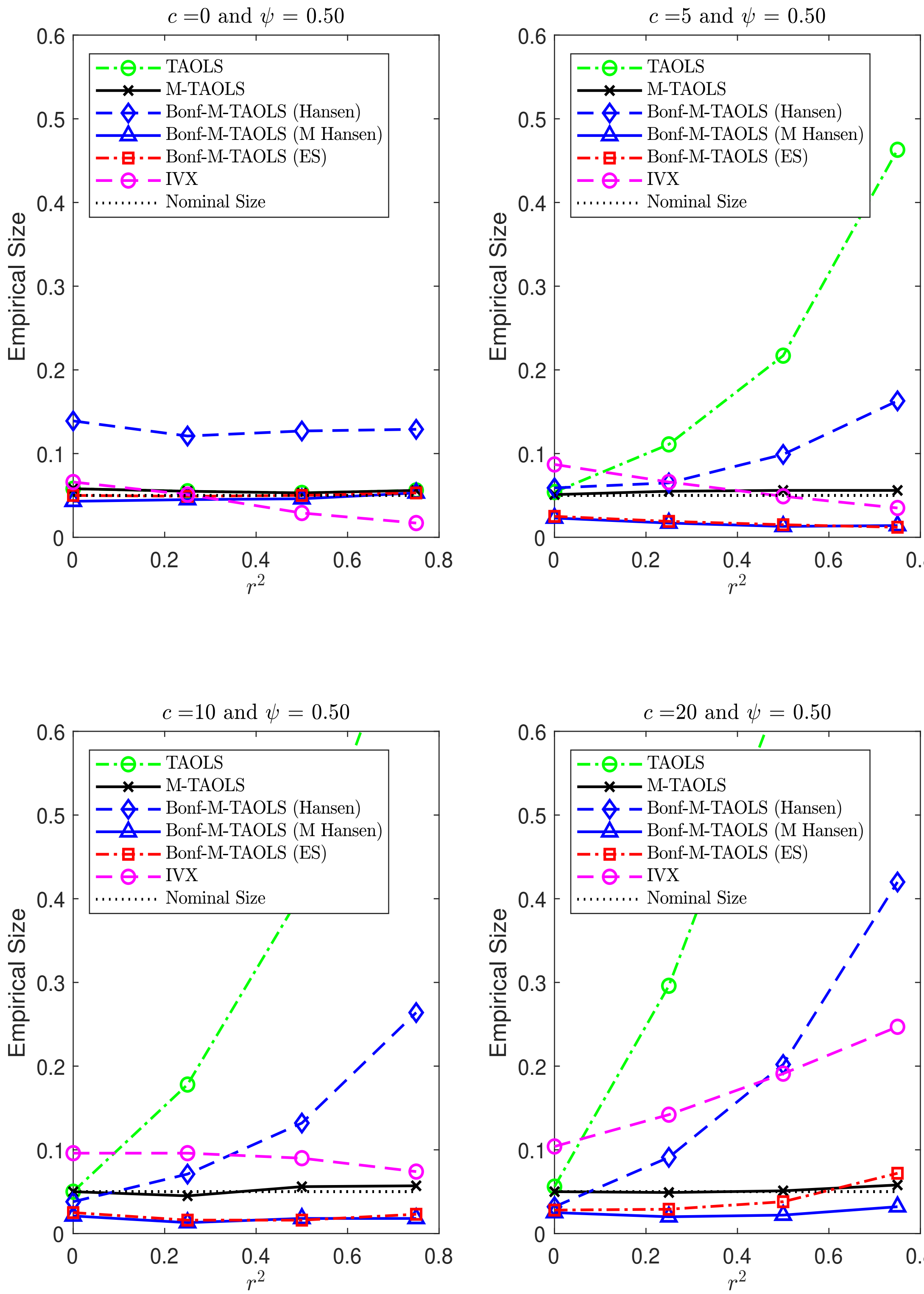

Figure S.6: Empirical sizes of `TAOLS`, `M-TAOLS`, `Bonf-M-TAOLS (Hansen)`, `Bonf-M-TAOLS (Hansen)`, `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with a single regressor, $K = 16$, and $\psi = 0.50$

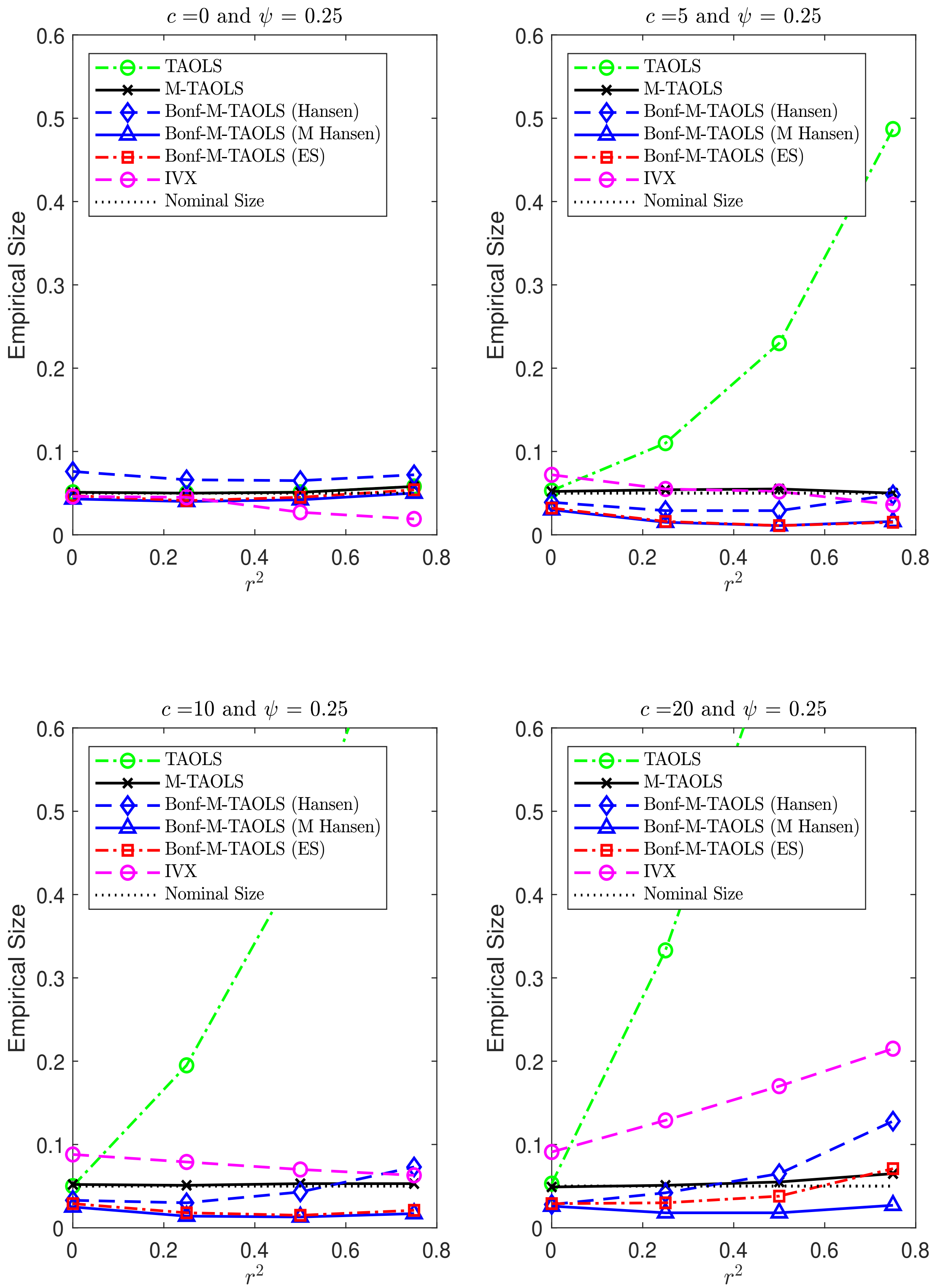

Figure S.7: Empirical sizes of `TAOLS`, `M-TAOLS`, `Bonf-M-TAOLS (Hansen)`, `Bonf-M-TAOLS (Hansen)`, `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with a single regressor, $K = 24$, and $\psi = 0.25$

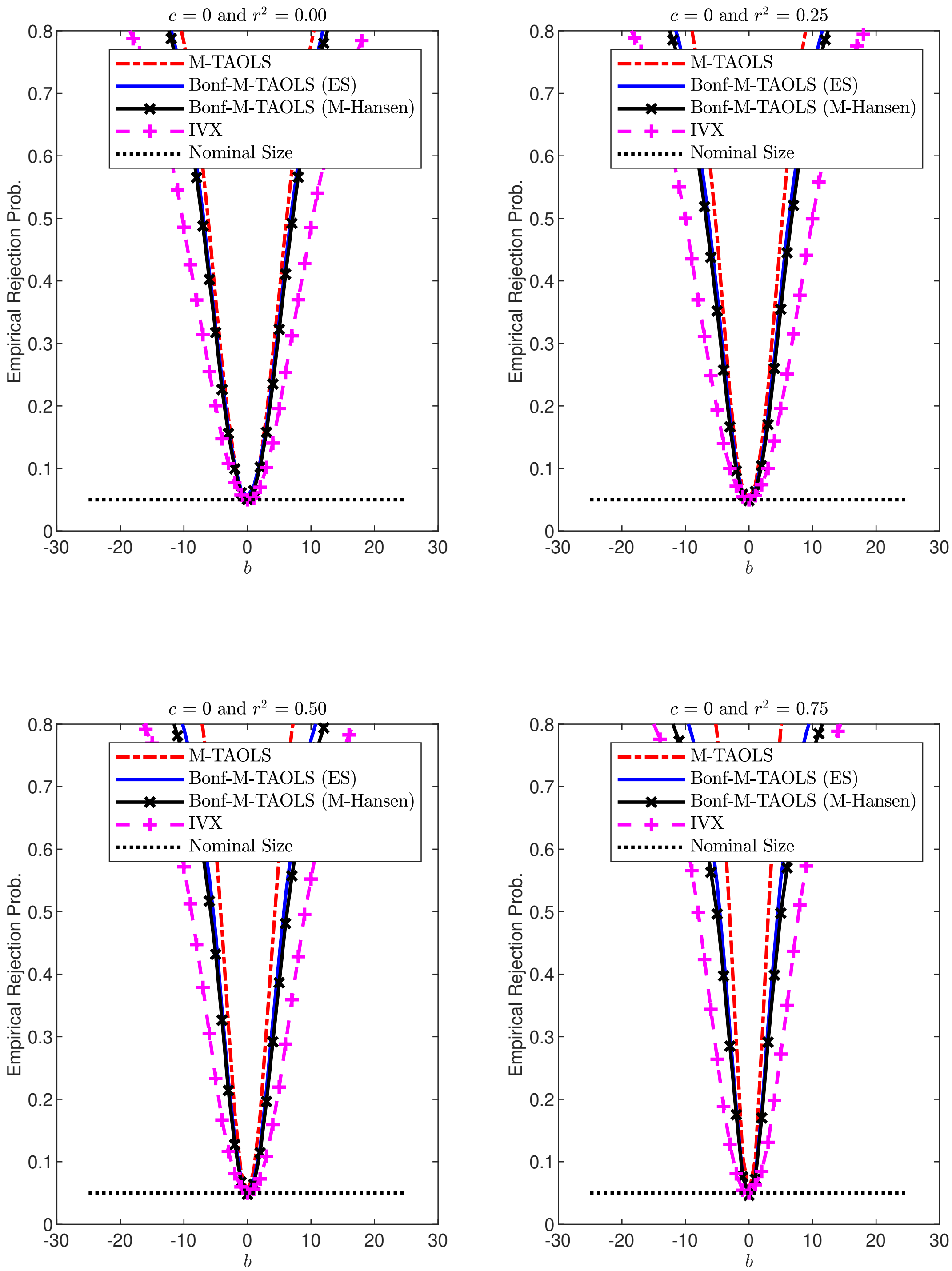

Figure S.8: Finite sample size-adjusted power curves of `M-TAOLS` (infeasible), `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with $K = 16$, $\psi = 0.50$, and $c_0 = 0$

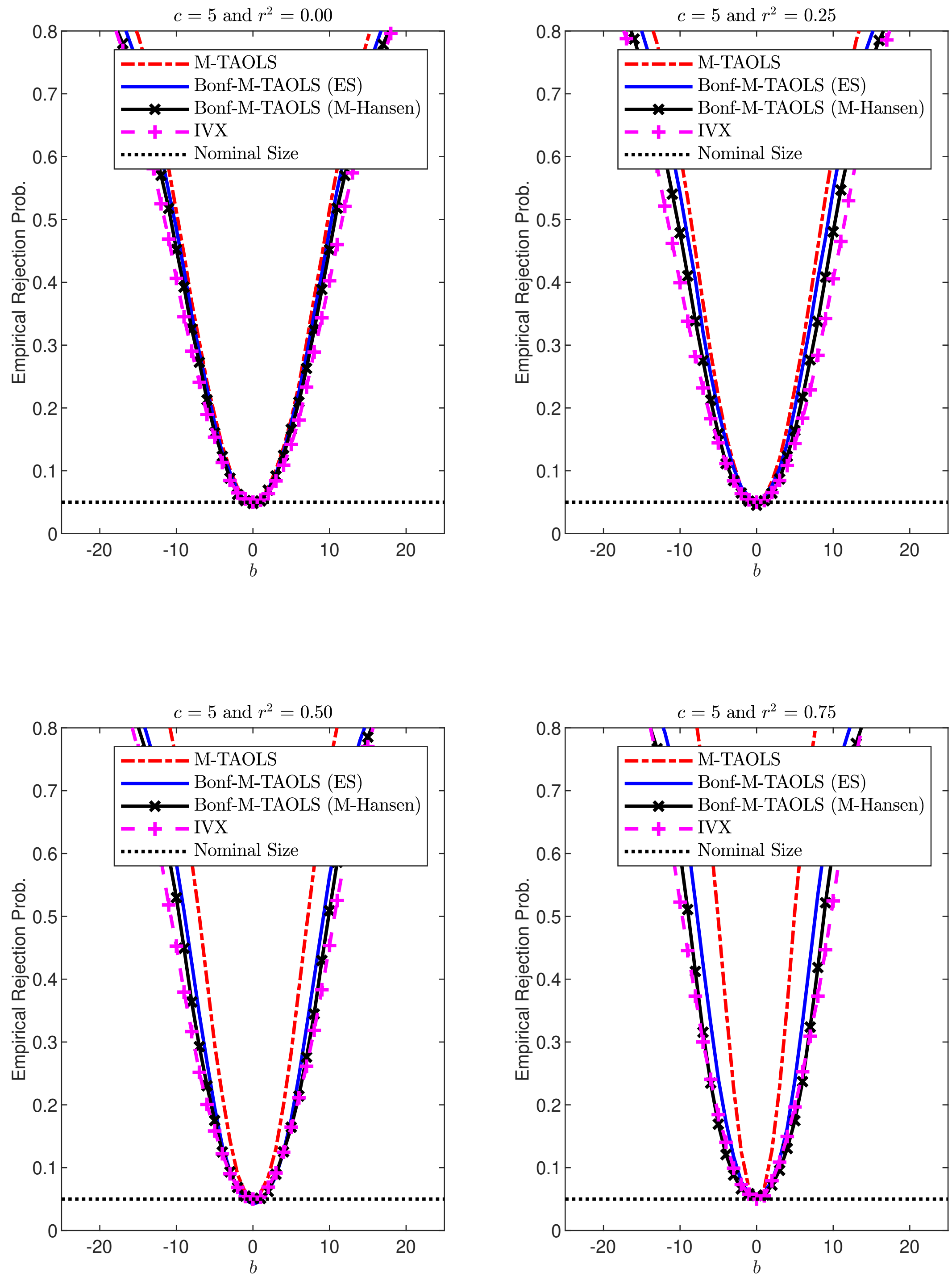

Figure S.9: Finite sample size-adjusted power curves of `M-TAOLS` (infeasible), `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with $K = 16$, $\psi = 0.50$, and $c_0 = 5$

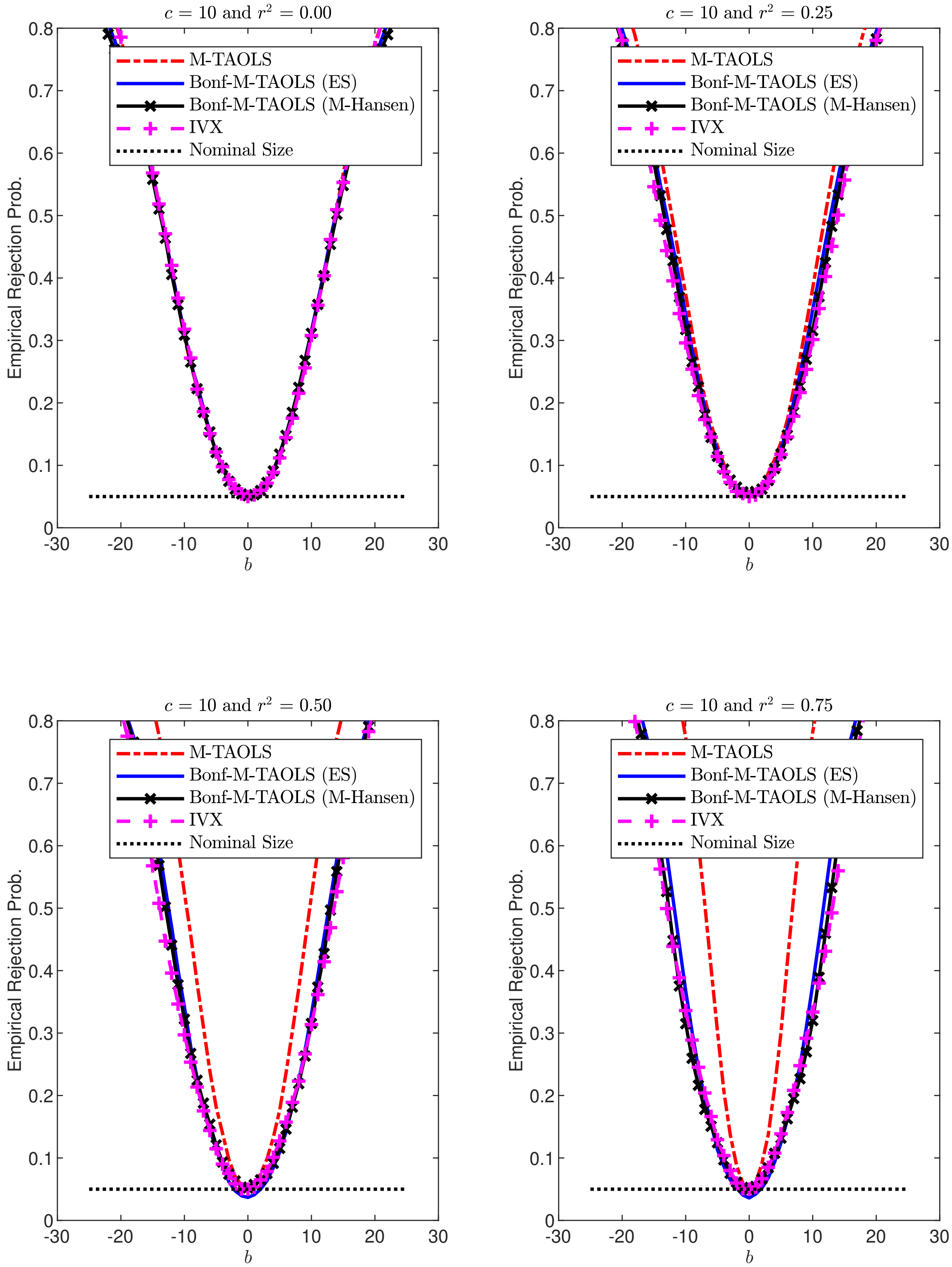

Figure S.10: Finite sample size-adjusted power curves of `M-TAOLS` (infeasible), `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with $K = 16$, $\psi = 0.50$, and $c_0 = 10$

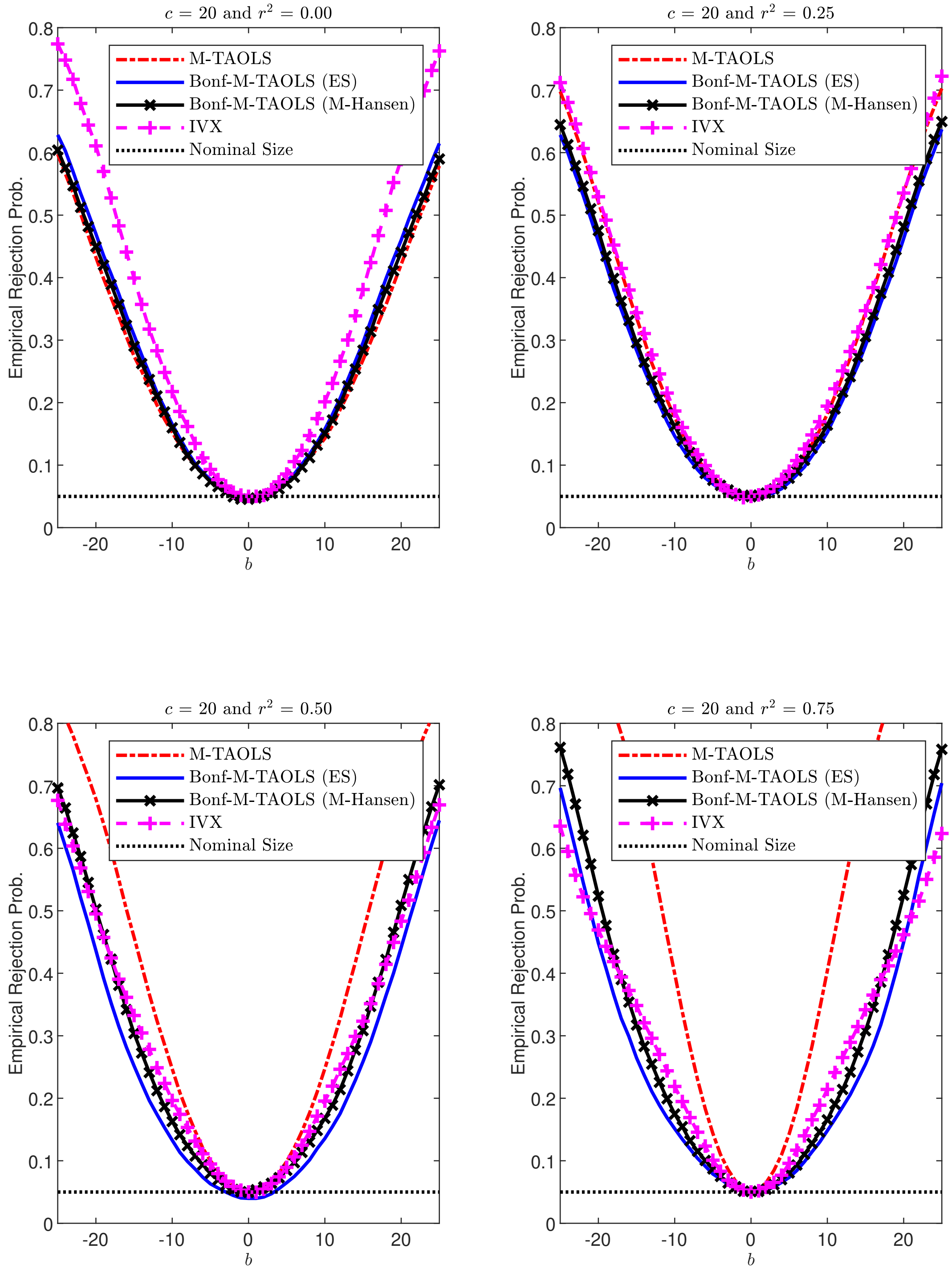

Figure S.11: Finite sample size-adjusted power curves of `M-TAOLS` (infeasible), `Bonf-M-TAOLS (M-Hansen)`, `Bonf-M-TAOLS (ES)`, and `IVX` with $K = 16$, $\psi = 0.50$, and $c_0 = 20$

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

[2] Campbell, J. Y., Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1), 27-60.

[3] Elliott, G., Stock, J. H. (2001). Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics*, 103(1-2), 155-181.

[4] Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics*, 81(4), 594-607.

[5] Lopez, J. H. (1997): "The Power of the ADF Test," *Economics Letters*, 57, 5–10.

[6] Ng, S., Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90(429), 268-281.

[7] Ng, S., Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6), 1519-1554.

[8] Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5), 1411-1452.