OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305–9049
doi: 10.1111/obes.12377

Response Surface Regressions for Critical Value Bounds and Approximate p-values in Equilibrium Correction Models*

Sebastian Kripfganz,† Daniel C. Schneider‡

†*Department of Economics, University of Exeter Business School, Streatham Court, Rennes Drive, Exeter, Devon EX4 4PU, UK (e-mail: S.Kripfganz@exeter.ac.uk)*
‡*Max Planck Institute for Demographic Research, Konrad-Zuse-Straße 1, Rostock, Mecklenburg-Vorpommern 18057, Germany (e-mail: schneider@demogr.mpg.de)*

Abstract

We consider the popular 'bounds test' for the existence of a level relationship in conditional equilibrium correction models. By estimating response surface models based on about 95 billion simulated F-statistics and 57 billion t-statistics, we improve upon and substantially extend the set of available critical values, covering the full range of possible sample sizes and lag orders, and allowing for any number of long-run forcing variables. By computing approximate P-values, we find that the bounds test can be easily oversized by more than 5 percentage points in small samples when using asymptotic critical values.

I. Introduction

The empirical analysis of time series data is often confronted with test statistics that have non-standard distributions in the presence of a unit root. While the asymptotic distributions can be characterized as functions of stochastic processes such as Brownian motions, the corresponding quantiles that are needed to compute critical values (CVs) for hypothesis testing are usually obtained with stochastic simulations. As an additional complication, the distributions of the test statistics generally depend on the specific assumptions about the data-generating process (DGP) and the specification of the estimated model, in particular whether an intercept or time trend are allowed. In a multivariable model, the dimension of the variable space and the cointegration rank matter. Importantly, the finite-sample distributions of the test statistics depend on further characteristics of the estimation. While augmenting the regression model with additional stationary variables does not affect the asymptotic distributions of unit-root and cointegration tests, their influence on the finite-sample distributions can be non-negligible. Given the vast number of empirically relevant regression specifications that lead to possibly different distributions, the tabulation of CVs

JEL Classification numbers: C12; C15; C32; C46; C63
*We are grateful for comments by Mehdi Hosseinkouchack and the anonymous referees. This study uses simulated data only. Details to reproduce the data are provided within this paper and the Supplementary Appendix.

quickly approaches space limits and is usually only done for a selected number of situations. This leaves blank areas that can be interpolated only to a limited extent.

All of these remarks apply to the Pesaran, Shin and Smith (2001) bounds test for the existence of a level relationship in an unrestricted conditional equilibrium correction model. This test is highly prominent among empirical researchers, not least because it evades the necessity of pretesting for the existence of unit roots, assuming that all variables are integrated at most of order one. The test yields conclusive evidence if the value of the test statistic falls outside of the critical value bounds established for the situations where all long-run forcing variables are purely integrated of either order zero, $I(0)$, or order one, $I(1)$.[1] Because the bounds procedure does not require that all variables are individually $I(1)$, the considered concept of a level relationship is broader than that of cointegration.

Pesaran *et al.* (2001) derive the asymptotic distributions of their test statistics under the null hypothesis of no level relationship and then use stochastic simulations to compute near-asymptotic CVs. However, the asymptotic distributions might be poor approximations of the actual distributions in small samples. Finite-sample CVs are tabulated by Mills and Pentecost (2001), Narayan and Smyth (2004), Kanioura and Turner (2005), and Narayan (2005), but they cover only a limited portion of the set of possible model specifications and sample sizes. Moreover, the precision of these CVs suffers from a relatively small number of replications in the respective simulations.

In this paper, we set out to systematically approximate the finite-sample and asymptotic distribution functions for the Pesaran *et al.* (2001) bounds test statistics. We fill the gaps regarding the CVs by estimating response surface (RS) models that predict the quantiles of the distributions as a function of the sample size, lag order and number of long-run forcing variables. The RS technique was introduced into the field of unit-root testing and cointegration analysis by MacKinnon (1991) for a range of Dickey and Fuller (1979) and Engle and Granger (1987) tests, and has since been applied numerous times.

Ericsson and MacKinnon (2002) provide RS estimates for the cointegration t-statistic in single-equation conditional error correction models that comprise the Dickey–Fuller statistic as a special case. Both asymptotic and finite-sample CVs can be obtained from these estimates.[2] As an important extension, Cheung and Lai (1995a) estimate RS models for the augmented Dickey–Fuller unit-root test, acknowledging the influence of the lag order on the finite-sample distributions.[3] As a complement to the generalized Dickey–Fuller t-statistic, Pesaran *et al.* (2001) propose a related F-statistic to test for the existence of a level relationship in a conditional equilibrium correction model.[4] So far, the only RS

[1] McNown et al. (2018) propose a bootstrap procedure for the Pesaran *et al.* (2001) test that allows for conclusive inference when the test statistic falls within the two bounds.

[2] Previously tabulated CVs for a small set of sample sizes can be found in Fuller (1976) and Dickey (1976) for the univariable and Banerjee, Dolado and Mestre (1998) for the multivariable setting.

[3] Cook (2001) compares the response surfaces from Cheung and Lai (1995a) with those from MacKinnon (1991) and concludes that adjusting for the lag order leads to a gain in power. RS estimates for finite-sample CVs of other unit-root tests are provided by Cheung and Lai (1995b), Harvey and van Dijk (2006), Otero and Smith (2012, 2017), and Otero and Baum (2017). All of them take the lag order into account. Further related applications of the RS methodology include Sephton (1995, 2008, 2017), Carrion-i-Silvestre, Sansó Rosselló and Artís Ortuño (1999), and Presno and López (2003).

[4] In the univariable model with restricted intercept or time trend, this statistic reduces to the Dickey and Fuller (1981) unit-root F-statistic.

estimates available for this F-statistic stem from Turner (2006) but they again cover only a narrow subset of the empirically relevant situations.

Our work improves and expands on the previous literature in several ways. With the stochastic simulation of about 95 billion F-statistics and 57 billion t-statistics under several scenarios regarding the deterministic model components, number of variables, sample size, and lag order, we can draw smooth density functions to illustrate how the distributions of the Pesaran *et al.* (2001) bounds test statistics change along various dimensions. Being based on these large-scale simulations, our RS estimates are both comprehensive and precise. Tabulations for selected combinations of the critical value determinants and interpolations between them become redundant. While previously reported CVs could not easily be extrapolated beyond the largest number of variables considered in the respective simulations, our modified RS approach does not impose a limit on the number of variables in the level relationship. We achieve this aim by exploiting the monotonically decreasing impact of adding another variable to the model.

Lastly, to facilitate a more informative statistical inference, we adopt the approach of MacKinnon (1994, 1996) to numerically approximate p-values and distribution functions.[5] Together with the CVs from our RS regressions, the approximate P-values can be computed with a program in the statistical software *Stata* (Kripfganz and Schneider, 2018). By comparing P-values, we can meaningfully quantify the finite-sample distortions of the bounds test. While these distortions are relatively small for the t-statistic, we find that the test based on the F-statistic at the 5% and 10% nominal levels can be easily oversized by more than 5%-points when using the asymptotic rather than the small-sample CVs. The distortions from ignoring the lag order of the variables in the regression model are less severe, but still relevant, and they can go in either direction.

II. Bounds testing for a level relationship

In this section, we provide a compact summary of the model and assumptions used by Pesaran *et al.* (2001) to derive the asymptotic distributions of their bounds testing procedure for the existence of a level relationship.

Equilibrium correction model

Let $\mathbf{z}_t$ be a column vector of $k+1$ random variables, generated by a vector-autoregressive (VAR) model of order q:

$$\mathbf{\Phi}(L)(\mathbf{z}_t - \mathbf{b}_0 - \mathbf{b}_1 t) = \boldsymbol{\epsilon}_t, \quad t = q+1, q+2, \ldots, T, \tag{1}$$

where $\mathbf{\Phi}(L) = \mathbf{I}_{k+1} - \sum_{i=1}^{q} \mathbf{\Phi}_i L^i$ is a q-th order polynomial in the lag operator L with unknown $(k+1)\times(k+1)$ coefficient matrices $\mathbf{\Phi}_i$, and $\mathbf{b}_0$ and $\mathbf{b}_1$ are $(k+1)$-dimensional vectors of unknown intercept and trend parameters. The initial observations $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_q$ are assumed to be observed. By defining the long-run multiplier matrix $\mathbf{\Pi} = \sum_{i=1}^{q} \mathbf{\Phi}_i - \mathbf{I}_{k+1}$ and the short-run coefficient matrices $\mathbf{\Gamma}_i = -\sum_{j=i+1}^{q} \mathbf{\Phi}_j$, $i = 1, 2, \ldots, q-1$, we can rewrite

[5] MacKinnon et al. (1999) proceed similarly for cointegration tests in a vector error correction model.

the above VAR(q) model in vector equilibrium correction (VEC) form:

$$\Delta \mathbf{z}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{\Pi} \mathbf{z}_{t-1} + \sum_{i=1}^{q-1} \mathbf{\Gamma}_i \Delta \mathbf{z}_{t-i} + \boldsymbol{\epsilon}_t, \tag{2}$$

where $\Delta = (1 - L)$ is the first-difference operator, $\mathbf{a}_0 = -\mathbf{\Pi}\mathbf{b}_0 + (\mathbf{\Pi} + \mathbf{\Gamma})\mathbf{b}_1$, $\mathbf{a}_1 = -\mathbf{\Pi}\mathbf{b}_1$, and $\mathbf{\Gamma} = \mathbf{I}_{k+1} - \sum_{i=1}^{q-1} \mathbf{\Gamma}_i$. Let us partition $\mathbf{z}_t = (y_t, \mathbf{x}_t')'$ and the long-run multiplier matrix conformably as

$$\mathbf{\Pi} = \begin{pmatrix} \pi_{yy} & \boldsymbol{\pi}_{yx}' \\ \boldsymbol{\pi}_{xy} & \mathbf{\Pi}_{xx} \end{pmatrix}.$$

Furthermore, partition $\mathbf{\Gamma}_i = (\boldsymbol{\gamma}_{yi}, \mathbf{\Gamma}_{xi}')'$ and $\mathbf{\Gamma} = (\boldsymbol{\gamma}_y, \mathbf{\Gamma}_x')'$.

Pesaran *et al.* (2001) impose the following assumptions:

Assumption 1 The roots of $|\mathbf{I}_{k+1} - \sum_{i=1}^{q} \mathbf{\Phi}_i z^i| = 0$ satisfy $-1 < 1/z \leqslant 1$. The DGP of $\mathbf{z}_t$ is integrated at most of order unity.[6]

Assumption 2 The vector of errors $\boldsymbol{\epsilon}_t$ is independent multivariate normally distributed, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$, with mean vector zero and positive-definite variance matrix $\mathbf{\Omega}$.

Assumption 3 The DGP of $\mathbf{x}_t$ is long-run forcing for the process of y_t, that is $\boldsymbol{\pi}_{xy} = \mathbf{0}$.

Assumption 4 The matrix $\mathbf{\Pi}_{xx}$ has rank r with $0 \leqslant r \leqslant k$.

Assumption 1 allows the individual elements of the vector $\mathbf{z}_t$ to be $I(0)$ or $I(1)$, or to be cointegrated. The cointegration order for the DGP of $\mathbf{x}_t$ is defined by Assumption 4. Consequently, the rank of the long-run multiplier matrix $\mathbf{\Pi}$ is either r or $r + 1$. Assumption 3 implies that $\mathbf{\Pi}$ being of rank r corresponds to the parameter restriction $\pi_{yy} = 0$, while the rank $r + 1$ necessitates $\pi_{yy} \neq 0$. Under Assumptions 3 and 4, we can express the long-run multiplier matrix as $\mathbf{\Pi} = \boldsymbol{\alpha}_y \boldsymbol{\beta}_y' + \mathbf{A}\mathbf{B}'$, where $\boldsymbol{\alpha}_y = (\alpha_{yy}, \mathbf{0}')'$ and $\boldsymbol{\beta}_y = (\beta_{yy}, \boldsymbol{\beta}_{yx}')'$ are $(k + 1)$-dimensional vectors, and $\mathbf{A} = (\boldsymbol{\alpha}_{yx}, \mathbf{A}_{xx}')'$ and $\mathbf{B} = (\mathbf{0}, \mathbf{B}_{xx}')'$ are $(k + 1) \times r$ matrices of full column rank respectively.[7] With the normalization $\beta_{yy} = 1$, it follows $\pi_{yy} = \alpha_{yy}$. Clearly, $\mathbf{A}\mathbf{B}' = \mathbf{0}$ if $r = 0$.

Under Assumptions 2 and 3, we can now obtain the following equilibrium correction (EC) model for y_t conditional on $\mathbf{x}_t$ and their past values $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{t-1}$:

$$\Delta y_t = c_0 + c_1 t + \boldsymbol{\pi}' \mathbf{z}_{t-1} + \sum_{i=1}^{q-1} \boldsymbol{\psi}_i' \Delta \mathbf{z}_{t-i} + \boldsymbol{\omega}' \Delta \mathbf{x}_t + u_t, \tag{3}$$

with intercept $c_0 = -\boldsymbol{\pi}'\mathbf{b}_0 + [(\boldsymbol{\gamma}_y - \mathbf{\Gamma}_x'\boldsymbol{\omega})' + \boldsymbol{\pi}']\mathbf{b}_1$ and trend coefficient $c_1 = -\boldsymbol{\pi}'\mathbf{b}_1$, and where $\boldsymbol{\pi} = (\pi_{yy}, \boldsymbol{\varphi}')'$, with $\boldsymbol{\varphi} = \boldsymbol{\pi}_{yx} - \mathbf{\Pi}_{xx}'\boldsymbol{\omega}$. Furthermore, $\boldsymbol{\psi}_i = \boldsymbol{\gamma}_{yi} - \mathbf{\Gamma}_{xi}'\boldsymbol{\omega}$ for all i. With the partition of the error term $\boldsymbol{\epsilon}_t = (\epsilon_{yt}, \boldsymbol{\epsilon}_{xt}')'$ and the conformably partitioned variance matrix

$$\mathbf{\Omega} = \begin{pmatrix} \omega_{yy} & \boldsymbol{\omega}_{xy}' \\ \boldsymbol{\omega}_{xy} & \mathbf{\Omega}_{xx} \end{pmatrix},$$

[6] See Pesaran *et al.* (2001) for a more formal statement of the last part of this assumption.

[7] This decomposition is useful for the derivation of the asymptotic distribution of the t-statistic used by Banerjee *et al.* (1998) to test whether $\pi_{yy} = 0$. See Pesaran *et al.* (2001) for details.

$\boldsymbol{\omega} = \boldsymbol{\Omega}_{xx}^{-1}\boldsymbol{\omega}_{xy}$ is obtained as the coefficient vector in the linear projection of ϵ_{yt} on $\boldsymbol{\epsilon}_{xt}$. The corresponding projection error u_t is independent normally distributed under Assumption 2, $u_t \sim \mathcal{N}(\mathbf{0}, \omega_{yy} - \boldsymbol{\omega}_{xy}'\boldsymbol{\Omega}_{xx}^{-1}\boldsymbol{\omega}_{xy})$.

A conditional level relationship between y_t and $\mathbf{x}_t$ exists if both $\pi_{yy} \neq 0$ and $\boldsymbol{\varphi} \neq \mathbf{0}$, and the data-generating processes of y_t and $\mathbf{x}_t$ are cointegrated if y_t is $I(1)$. In the opposite situation, $\boldsymbol{\pi} = \mathbf{0}$, the conditional EC model (3) only contains first-differenced terms such that no level relationship between y_t and $\mathbf{x}_t$ can exist and y_t must be $I(1)$. There are two degenerate cases. If just $\pi_{yy} = 0$, y_t is still $I(1)$ and there exists only a level relationship among the elements of $\mathbf{x}_t$ not involving y_t. If π_{yy} is the only non-zero element of $\boldsymbol{\pi}$, y_t is generated by a trend-stationary or $I(0)$ process not involving the levels of $\mathbf{x}_t$.

Bounds test

In the light of the two degenerate situations, the following testing procedure can be applied:

(1) Test the joint null hypothesis $H_0^{\pi} : \boldsymbol{\pi} = \mathbf{0}$ versus $H_1^{\pi} : \boldsymbol{\pi} \neq \mathbf{0}$.
(2) If H_0^{π} is rejected, test the single hypothesis $H_0^{\pi_{yy}} : \pi_{yy} = 0$ versus $H_1^{\pi_{yy}} : \pi_{yy} < 0$, under the additional assumption that either $r = 0$ or $\boldsymbol{\alpha}_{yx} - \mathbf{A}_{xx}'\boldsymbol{\omega} = \mathbf{0}$ if $0 < r \leqslant k$.
(3) If $H_0^{\pi_{yy}}$ is rejected, test the joint hypothesis $H_0^{\theta} : \boldsymbol{\theta} = \mathbf{0}$ versus $H_1^{\theta} : \boldsymbol{\theta} \neq \mathbf{0}$, where $\boldsymbol{\theta} = -\boldsymbol{\varphi}/\pi_{yy}$ are the long-run multipliers in the conditional level relationship between y_t and $\mathbf{x}_t$.

The reason for proceeding with steps (2) and (3) is that the alternative hypothesis H_1^{π} in step (1) does not rule out any of the two degenerate cases mentioned above. The latter are the subject of the hypothesis tests in steps (2) and (3). Only if all three null hypotheses are rejected, we can conclude that there is statistical evidence for the existence of a non-degenerate level relationship between y_t and $\mathbf{x}_t$.

As demonstrated by Pesaran *et al.* (2001), y_t is $I(1)$ under the null hypothesis in steps (1) and (2) and the respective test statistics have non-standard asymptotic distributions. The additional assumption required for step (2) implies $\boldsymbol{\varphi} = \pi_{yy}\boldsymbol{\beta}_{yx}$. Consequently, under $H_0^{\pi_{yy}}$ we have again $\boldsymbol{\pi} = \mathbf{0}$ as in step (1), but $H_1^{\pi_{yy}}$ is more informative at the cost of imposing additional structure on the DGP. Without this assumption, the asymptotic distribution of the t-statistic would depend on nuisance parameters and tabulations of CVs for general purposes would become practically infeasible.[8]

For the long-run multipliers $\boldsymbol{\theta}$ that are the subject of step (3), Pesaran and Shin (1998) and Hassler and Wolters (2006) show that the ordinary least squares (OLS) estimator is super-consistent if $\mathbf{x}_t$ contains $I(1)$ regressors, and it is asymptotically normally distributed irrespective of the order of integration. This constitutes a practical advantage over tests directly based on $\boldsymbol{\varphi}$ because the latter have non-standard distributions.[9] The remainder of this text is primarily concerned with the test statistics in steps (1) and (2).

[8] See Pesaran *et al.* (2001) for a discussion. Banerjee *et al.* (1998) assume $r = 0$ and briefly argue that the CVs obtained under this assumption will lead to a conservative test if it is violated.

[9] McNown et al. (2018) propose a bootstrap procedure for the inference on the coefficients $\boldsymbol{\varphi}$ of the level regressors. Following the procedure of Pesaran *et al.* (2001) and Narayan (2005), Sam, McNown and Goh (2018) tabulate CVs for a Wald test of joint insignificance of up to seven long-run forcing variables in the level relationship.

The restricted VAR formulation (1) imposes constraints on the coefficients c_0 and c_1 in the conditional EC model (3) that ensure that the cointegration rank r does not affect the deterministic trending behaviour.[10] Pesaran *et al.* (2001) distinguish five cases, depending on which deterministic components are included in the model specification and whether we disregard the implied restrictions on their coefficients or not:

(i) No intercept and no trend are included, $c_0 = c_1 = 0$,
(ii) A restricted intercept is included but no trend, $c_0 = -\boldsymbol{\pi}'\mathbf{b}_0$ and $c_1 = 0$,
(iii) An unrestricted intercept is included but no trend, $c_0 \neq 0$ and $c_1 = 0$,
(iv) An unrestricted intercept and a restricted trend are included, $c_0 \neq 0$ and $c_1 = -\boldsymbol{\pi}'\mathbf{b}_1$,
(v) An unrestricted intercept and an unrestricted trend are included, $c_0 \neq 0$ and $c_1 \neq 0$.

As emphasized by Pesaran *et al.* (2001), the data-generating processes under cases (ii) and (iii) are identical, and similarly for cases (iv) and (v), but the Wald test statistics in step (1) and their asymptotic distributions differ under the null hypothesis H_0^{π}. For the single-hypothesis test in step (2), the restrictions can be ignored.

Pesaran *et al.* (2001) argue that the CVs for the two polar cases of $\mathbf{x}_t$ being purely $I(0)$ or purely $I(1)$ provide lower and upper bounds, respectively, when the orders of integration and the cointegration rank r are unknown. They derive the asymptotic distributions of the Wald test statistic in step (1) and the t-statistic in step (2) respectively. Both statistics are functions of standard Brownian motions, de-meaned and de-trended where necessary, and depend on the cointegration rank r.[11]

III. Critical values and approximate *P*-values

Pesaran *et al.* (2001) use stochastic simulations to obtain near-asymptotic critical value bounds based on a sample size of 1000 time periods for the F-statistic under H_0^{π} in step (1) and the t-statistic under $H_0^{\pi_{yy}}$ in step (2).[12] They tabulate the CVs for the range of $k \in [0, 10]$ long-run forcing variables. Several other authors provide finite-sample CVs for a subset of the relevant situations. We summarize the existing literature in Table 1.[13] A number of authors tabulated CVs that require interpolations between the reported sample sizes. Accordingly, they are unanimously superseded by the estimates from RS regressions, whenever the latter are available and sufficiently precise.

Although unit-root tests are not the primary focus of our work, the Dickey–Fuller test statistics result as a special case in the univariable setting, $k = 0$. When there is no need for a lag augmentation, the RS estimates of MacKinnon (1996, 2010) and Ericsson and MacKinnon (2002) are the primary source for accurate finite-sample CVs, as far as the t-statistic is concerned. In many situations, however, serial error correlation threatens to undermine the validity of the test. A remedy is the augmented Dickey–Fuller test based

[10] See Pesaran, Shin and Smith (2000) for details.

[11] See Theorems 3.1 and 3.2 in Pesaran *et al.* (2001).

[12] The F-statistic is obtained by dividing the Wald statistic by $k + 1$ in cases (i), (iii) and (v), and by $k + 2$ in cases (ii) and (iv).

[13] The distributions of the cointegration test statistics resulting from the Engle and Granger (1987) two-stage procedure differ from those considered in the Pesaran *et al.* (2001) framework. Corresponding RS estimates can be found in MacKinnon (1991, 1996, 2010).

TABLE 1

Critical value tabulations in the previous literature

					Deterministics cases‡	
	$N(T,q)$	q	k	$I(d)$	F	t
Fuller (1976)	$25, 50, 100, 250, 500, \infty$	1	0	–	–	(i), (iii), (v)
Dickey (1976)	$25, 50, 100, 250, 500, 750, \infty$	1	0	–	–	(i), (iii), (v)
Dickey and Fuller (1981)	$25, 50, 100, 250, 500, \infty$	1	0	–	(ii), (iv)	–
MacKinnon (1991, 2010)	RS	1	0	–	–	(i), (iii), (v)
Cheung and Lai (1995a)	RS	$\geqslant 1$	0	–	–	(i), (iii), (v)
MacKinnon (1996)*	RS	1	0	–	–	(i), (iii), (v)
Banerjee *et al.* (1998)	$25, 50, 100, 500, \infty$	1	$[1, 5]$	1	–	(iii), (v)
Pesaran *et al.* (2001)	1000	0	$[0, 10]$	0, 1	(i)–(v)	(i), (iii), (v)
Mills and Pentecost (2001)	22, 26	1	3	0, 1	(i)–(v)	(i), (iii), (v)
Ericsson and MacKinnon (2002)*	RS	1	$[0, 11]$	1	–	(i), (iii), (v)
Narayan and Smyth (2004)	$22, 25, 30, 37$	0	2	0, 1	(ii)	–
Kanioura and Turner (2005)†	$50, 100, 200, 500$	0/1	$[1, 3]$	1	(iii)	(i)
Narayan (2005)	30–80 in steps of 5	0	$[0, 7]$	0, 1	(ii)–(v)	–
Turner (2006)	RS	1	$[1, 3]$	0, 1	(iii), (v)	–

Notes: The regression model used to compute the F-statistics and t-statistics can be written as in equation (6) with q lags and k long-run forcing variables that are integrated of order d. For the unit-root tests, that is, $k = 0$, the specifications are equivalent for $q = 0$ and $q = 1$. $N(T, q) = T - \max(q, 1)$ denotes the effective sample size.

*MacKinnon (1996) and Ericsson and MacKinnon (2002) provide computer programs that compute the CVs and approximate p-values.

†Kanioura and Turner (2005) compute their test statistics from different regression specifications. Their F-statistic is based on $q = 1$ and their t-statistic on $q = 0$. The latter is only tabulated for $k = 1$.

‡MacKinnon (1996, 2010) and Ericsson and MacKinnon (2002) furthermore consider the t-statistic in the presence of a quadratic trend.

on a higher-order autoregressive model. The test statistic remains the same, and Said and Dickey (1984) prove that its asymptotic distribution is unaffected as well. However, the degrees-of-freedom reduction affects the finite-sample distributions. The RS from Cheung and Lai (1995a) provides more accurate CVs in that situation. For the unit-root F-statistic, we are the first to provide comprehensive RS estimates.

In the multivariable setting, the lag order dependence of finite-sample CVs has been neglected completely so far. A stronger emphasis has been put on the number of variables in the level relationship. The RS estimates from Ericsson and MacKinnon (2002) cover the cointegration t-statistic for up to 11 long-run forcing variables that are purely $I(1)$. For the F-statistic, the coverage is much thinner. To date, only Turner (2006) provides such RS estimates, but merely for cases (iii) and (v) and a small number of up to three long-run forcing variables.

Monte Carlo simulations

To improve upon and substantially expand existing critical value tabulations via RS regressions, we start by computing empirical distribution functions (EDFs) for the F- and t-statistic under a variety of scenarios. The respective quantiles from these EDFs will be used in the subsequent RS analysis. For each replication in our Monte Carlo simulations,

we generate the data according to the following processes that satisfy H_0^{π} and $H_0^{\pi_{yy}}$:

$$y_t = y_{t-1} + \epsilon_{yt}, \tag{4}$$

$$\mathbf{x}_t = \mathbf{P}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_{xt}, \tag{5}$$

for $t = 1, 2, \ldots, T + 50$ and with the initializations $y_0 = 0$ and $\mathbf{x}_0 = \mathbf{0}$. The first 50 observations are discarded. The elements of the vector of shocks $\boldsymbol{\epsilon}_t$ are independently drawn from the standard normal distribution. The matrix $\mathbf{P}$ equals either the zero or the identity matrix, depending on whether $\mathbf{x}_t$ is supposed to be purely $I(0)$ or $I(1)$.[14]

The test statistics are constructed from the unrestricted regression coefficients in a reparameterization of equation (3):

$$\Delta y_t = c_0 + c_1 t + \pi_{yy} y_{t-1} + \boldsymbol{\varphi}' \mathbf{x}_t + \sum_{i=1}^{q-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \boldsymbol{\psi}_{xi}' \Delta \mathbf{x}_{t-i} + u_t, \tag{6}$$

where $(\psi_{yi}, \boldsymbol{\psi}_{xi}')' = \boldsymbol{\psi}_i$ for all $i = 1, 2, \ldots, q-1$. The use of the contemporaneous $\mathbf{x}_t$ instead of the lagged $\mathbf{x}_{t-1}$ is advocated by Pesaran and Shin (1998). It has the advantage that the short-run coefficients $\boldsymbol{\psi}_{xi}$ can be treated as unrestricted for all lag orders q, while in the representation (3) the presence of the term $\boldsymbol{\omega}' \Delta \mathbf{x}_t$ induces an overparameterization when $q = 0$.[15] In cases (i), (iii) and (v), under the null hypothesis H_0^{π}, the F-statistic is used to test for joint insignificance of the level regressors y_{t-1} and $\mathbf{x}_t$ in equation (6). In cases (ii) and (iv), the respective exclusion restriction on the intercept c_0 or trend coefficient c_1 is added. Under $H_0^{\pi_{yy}}$, the t-statistic is computed for π_{yy}.

For each of the two integration orders and five deterministic model component cases, we run separate simulations for all combinations of $k \in [0, 10]$,

$$T \in \{18, 20, 22, 25, 28, 30, 32, 36, 40, 50, 60, 80, 100, 150, 200, 300, 400, 500, 1000\},$$

and $q \in \{0, 1, 2, 3, 4, 6, 8, 12\}$, subject to the restriction that there are at least twice as many observations as coefficients in equation (6) to ensure a sufficient number of degrees of freedom.[16] This yields a total of 9,498 simulation designs.[17] For each design, we run 100,000 replications and then repeat the entire procedure 100 times, which we refer to as 'meta replications'.[18] We thus compute a total number of 9.498×10^{10} F-statistics and 5.726×10^{10} t-statistics. To reduce the storage memory requirements for such a large number of test statistics, we first round the statistics to three digits after the decimal point and then apply a reversible transformation in terms of first differences of sorted statistics and occurrence counts. This procedure allows us to change the storage type of the data

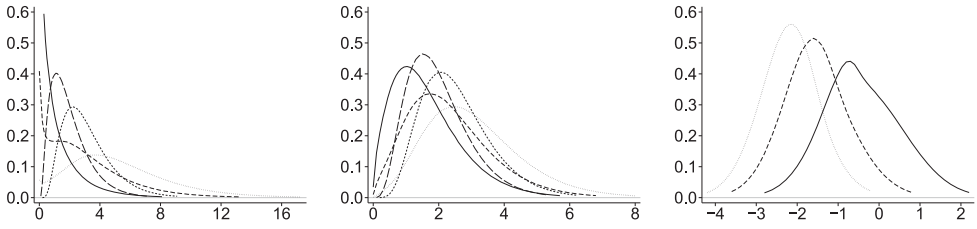
[14] The DGP is identical to the one used by Pesaran *et al.* (2001), besides the discarded observations.

[15] The lag specification $q = 0$ can be obtained from the VAR(1) model in equation (1) by imposing the restriction $\boldsymbol{\omega} = \boldsymbol{\varphi}$.

[16] That is, $\max(q, 1) + k(q + 1) + \mathcal{I}(c_0 \neq 0) + \mathcal{I}(c_1 \neq 0) \leqslant (T - \max(q, 1))/2$, where $\mathcal{I}(\cdot)$ is an indicator function that equals unity if the respective deterministic component is included and zero otherwise. The effective sample size is $T - \max(q, 1)$. The distinction between $q = 0$ and $q = 1$ is irrelevant when $k = 0$.

[17] There are 1,954 simulation designs for case (i), 1,904 designs for cases (ii) and (iii) each, and 1,868 designs for cases (iv) and (v) respectively.

[18] There is no longer a computational reason as in MacKinnon (1996) for the use of meta replications instead of a single experiment with 10 million replications. His second argument, that meta replications provide an easy way to evaluate the experimental randomness, survives.



(a) F-statistic, $k = 0$ (b) F-statistic, $k = 2$ (c) t-statistic, $k = 0$

Figure 1 Probability density functions obtained from the 10^7 simulated test statistics for cases (i)-(v) with sample size $T = 1000$ and lag order $q = 1$. With increasing case number, the curves have shorter dashes. For $k = 2$, the upper-bound densities are shown.

from 4-byte floating-point numbers to 2-byte integers. Furthermore, the resulting bit-level repeating patterns increase the efficiency of conventional compression algorithms.[19]

The 10 million statistics for each configuration are sufficiently many to draw reasonably smooth probability density functions. With a bin width of 0.1, Figure 1 is obtained by connecting the points that result from counting the number of simulated test statistics for each bin (divided by the total number of test statistics and the bin width).[20] In particular for the F-statistic, the shape of the distributions varies quite a bit depending on the deterministic model components. This is illustrated in Figure 1 for a sample size of $T = 1000$ that was considered by Pesaran *et al.* (2001) in their simulation of near-asymptotic CVs.

In the univariable situation, $k = 0$, we observe unimodal densities in cases (ii) and (iv) with a restricted intercept or trend. In case (i) without any deterministic component, the density is zeromodal. The density in case (iii) with an unrestricted intercept looks similar in that it is downward sloping almost everywhere, but with a saddle point or tiny mode after the initial steep descent. In the unrestricted trend case (v), we observe a local minimum close to the origin. In the multivariable designs, all densities have the expected unimodal shape with positive skewness. For the t-statistic, the densities resemble a normal distribution but are slightly asymmetric and not centred around zero. Note that the dent to the right of the mode in case (i) is consistent with the asymptotic density obtained by MacKinnon (1994, figure 3) and the theoretical asymptotic density derived by Abadir (1995, figure 1). With increasing case number, the mode moves further away from zero and the dispersion becomes smaller. In the following, we restrict the discussion primarily to the empirically most often applied case (iii).

Figures 2 and 3 highlight the variation of the densities across the number of variables k, separately for different sample sizes. For the F-statistic, the probability mass around the mode is increasing in both k and T but the mode itself remains fairly stable. The shape of the distributions is quite similar when all long-run forcing variables are $I(1)$ compared to when

[19] Rounding calculated statistics may not always be innocuous. Rather, the choice of the rounding precision was tailored specifically to our compression procedure and to our particular application to ensure that it does not influence the results. Details on the compression procedure, including a robustness check regarding the effect of rounding, as well as other computational aspects are relegated to the Supplementary Appendix.

[20] We restrict the plots of density and distribution functions to the quantiles $P \leqslant 0.995$ for the F-statistic and $0.005 \leqslant P \leqslant 0.995$ for the t-statistic. Larger and coloured versions of some of our figures can be found in the Supplementary Appendix. Rendered at a larger size, it can be seen that a bit of additional smoothing would still be helpful, in particular around the mode of the densities.

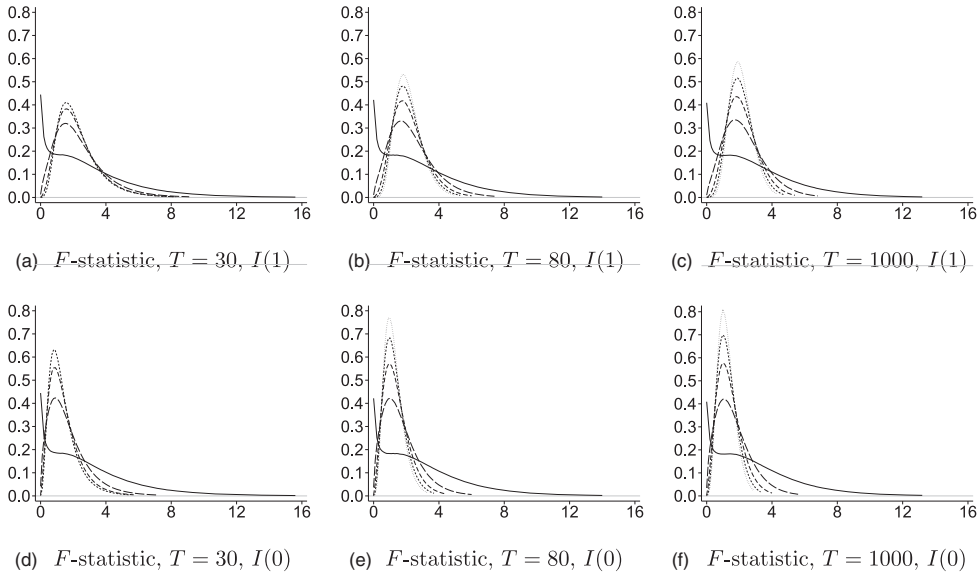


Figure 2 Upper-bound and lower-bound probability density functions obtained from the 10^7 simulated F-statistics in case (iii) with $k \in \{0, 2, 4, 6, 8\}$ variables and lag order $q = 1$. The solid curve refers to $k = 0$. With increasing k, the curves have shorter dashes.

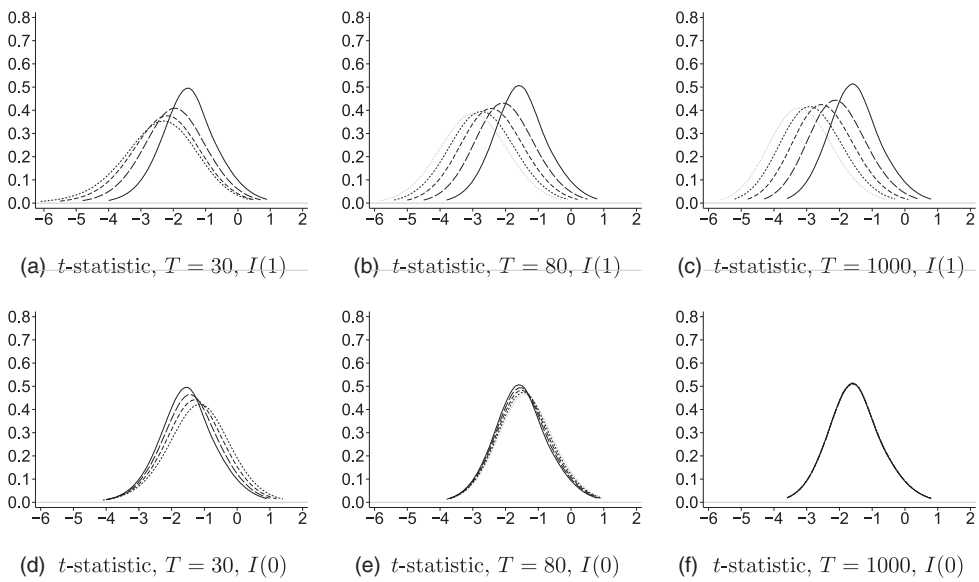


Figure 3 Upper-bound and lower-bound probability density functions obtained from the 10^7 simulated t-statistics in case (iii) with $k \in \{0, 2, 4, 6, 8\}$ variables and lag order $q = 1$. The solid curve refers to $k = 0$. With increasing k, the curves have shorter dashes.

they are $I(0)$. For obvious reasons, the corresponding quantiles are found closer to zero for the lower-bound distributions.[21] For the t-statistic, some differences arise. The densities are as well less dispersed with larger sample size but more dispersed with increasing number of

[21] For $k = 0$, the upper-bound and lower-bound densities coincide.

variables. While the upper-bound densities become more distinct with increasing sample size and their quantiles grow with k, the opposite is true for the lower bound. As formally shown by Pesaran *et al.* (2001), the distributions of the t-statistic asymptotically no longer depend on the number of $\mathbf{x}_t$ variables when all of them are $I(0)$.[22]

We can construct such probability density functions for any of our simulation designs. By sorting the 10^7 simulated test statistics in ascending order, it is straightforward to obtain the corresponding quantiles of interest. For example, in case (iii), the 95th percentile of the F-statistic with $k=2$ long-run forcing variables that are $I(1)$, $T=1000$ observations, and a lag order of $q=1$ is found to be 4.81. Pesaran *et al.* (2001) report a critical value of 4.85 for this setup. The difference between these two numbers is within the range of the simulation uncertainty that can be measured by the variation across the 100 meta-replication EDFs, each of them based on 10^5 replications instead of the 10^7 replications used to construct the aggregate EDFs. For our example, the observed quantiles fall into the interval $[4.77, 4.86]$ with a coefficient of variation of 0.29%. This number is close to the average of 0.30% over all simulation designs for the F-statistic. The further we go into the tail of the distribution, the more noisy the quantile estimates are. For the 99th percentile, the average coefficient of variation is 0.51%. In the Supplementary Appendix, we show that the variation tends to shrink with larger T and larger k, and that it is larger for the lower than for the upper bound. For the t-statistic, the coefficient of variation is a bit smaller in absolute terms, on average 0.21% for the 95th percentile and 0.33% for the 99th percentile.

Due to the independence of the replications, we can infer statements about the precision of the aggregate EDFs. Since their number of replications exceeds that of the meta replications by factor 100, the respective coefficient of variation is an order of magnitude smaller than for a single meta replication. In the above example, this implies a coefficient of variation of 0.03% for the 95th percentile of the F-statistic. By contrast, for 40,000 replications, as performed by Pesaran *et al.* (2001), it would be about 0.46% which is still a non-negligible amount of variation. This is best seen by noting that their tabulated critical value of 4.85 corresponds to a P-value of 0.048 rather than 0.05 when we use our aggregate EDF as the reference distribution. Similar arguments apply to the finite-sample CVs tabulated by Narayan (2005) that do not comply with the monotonic decline of the finite sample towards the asymptotic quantiles due to the experimental randomness.

Our approach is not without caveat. In a strict sense, our simulations and RS regressions generate accurate finite-sample distributions and CVs only if the actual DGP coincides with the simulated one. The data-generating processes (4) and (5) are a restricted version of the general VAR(q) model (1). As long as Assumptions 1 to 4 are respected, the restrictions imposed on the coefficient matrices $\mathbf{\Phi}_i, i=1,2,\ldots,q$, and the variance matrix $\mathbf{\Omega}$ are innocuous for the asymptotic distribution of the test statistic (Pesaran *et al.*, 2001). However, this may no longer be the case for finite samples. While our simulations can be easily modified to accommodate alternative restrictions or additional nuisance parameters, the true DGP is unknown in practice. As a consequence, our finite-sample distributions and corresponding CVs should in general be regarded as approximations of the true but unknown

[22] When $T=1000$, the upper-bound densities for the t-statistic look very similar to the asymptotic density functions plotted by Ericsson and MacKinnon (2002, figure 2).

distributions.[23] To gain some insights into the quality of our approximations, we analyse the size distortions that result from using our CVs when the true DGP has autoregressive or moving-average disturbances ϵ_t. For most scenarios, we find that the bounds test with our critical values overrejects only modestly under these modified processes, provided that the regression model is augmented with sufficiently many lags to capture the serial correlation.[24] Given the variety of possible data-generating processes and the significant computational cost associated with the simulations, we focus in the following on the basic DGP in equations (4) and (5).

Response surface regressions

Even with our restricted DGP, the tabulation of all empirically relevant CVs would be cumbersome since it would stretch dozens of pages. Moreover, even though we have obtained EDFs from 9,498 simulation designs, they still do not cover the whole spectrum of sample sizes, lag orders and variable counts. In the following, we thus estimate RS models that allow us to predict CVs for any point in this three-dimensional space.

For each meta replication and simulation design, we compute the quantiles of interest from the EDFs of both test statistics. In the previous literature, the most relevant quantiles have either been tabulated or used in RS regressions for a given number of k long-run forcing variables. The RS models are usually estimated by regressing the simulated quantiles on a polynomial in the reciprocal of the sample size. To account for the increasing relevance of the lag order in smaller samples, Cheung and Lai (1995a) have added a polynomial in the lag order divided by the sample size.

In the Supplementary Appendix to this paper, we proceed similarly by estimating RS regressions for each quadruplet $\{c,k,d,p\}$, where c is the case regarding the deterministic model components, k is the number of long-run forcing variables with integration order d, and p is the level of the quantile:

$$Q_k(T,q)=\sum_{j=0}^{m}\sum_{l=0}^{n}\theta_{j,l}[N(T,q)]^{-j}[H(q,k)]^{l}+u, \tag{7}$$

where $Q_k(T,q)$ refers to the quantiles from the meta-replication EDFs for a given k, $N(T,q)=T-\max(q,1)$ is the effective sample size, $H(q,k)=\max(q-1,0)+kq$ denotes the number of unrestricted short-run coefficients in equation (6), and u is the regression error. Because the presence of stationary first-differenced terms in equation (6) does not affect the asymptotic properties of the distributions, we can restrict $\theta_{0,l}=0$ for all $l>0$. The intercept can then be interpreted as the asymptotic quantile when $T\to\infty$.

For the limited number of congruent scenarios, the estimated RS hardly differs from those of Turner (2006) for the F-statistic and MacKinnon (2010) and Ericsson and MacK-

[23] We thank an anonymous referee for raising this issue. The same qualification applies to critical value tabulations and RS regressions in the previous literature. A bootstrap procedure could be used to circumvent this concern. See Palm, Smeekes and Urbain (2010) and McNown, Sam and Goh (2018) and the references therein.

[24] See the Supplementary Appendix for a detailed discussion. Similarly, for the augmented Dickey–Fuller test with sufficiently large lag order q, Cheung and Lai (1995a) report that the size distortions can be kept small for a DGP with autoregressive or moving-average disturbances. Also note that a DGP with moving-average shocks violates Assumption 2 for any finite q.

innon (2002) for the t-statistic. Yet, their CVs are no longer ideal for higher lag orders in equation (6). For most sample sizes, they are too conservative, to such an extent that even the asymptotic CVs would provide a better approximation. The Cheung and Lai (1995a) RS addresses this problem but is slightly skewed towards zero compared to ours.[25]

Carrying out RS estimations separately for each k has two shortcomings. First, this approach does not allow to obtain CVs if the actual number of long-run forcing variables has not been considered in the simulations. Second, any attempt to cover a larger range of k inflates the number of regression results that need to be tabulated or stored in a computer program. In the following, we overcome this problem by directly modelling the RS as a function of k. A close look at either the existing RS estimates or those from our Supplementary Appendix reveals that the marginal differences between the quantiles become smaller with increasing k. This suggests to model this diminishing slope with negative powers in the total number of variables $1+k$. Thus, for each triplet $\{c,d,p\}$, we consider the following regression:

$$Q(k,T,q)=\sum_{i=0}^{r}\sum_{j=0}^{m}\sum_{l=0}^{n}\theta_{i,j,l}(1+k)^{-i}[N(T,q)]^{-j}[H(q,k)]^{l}+v. \tag{8}$$

The lag order q is uninformative for the asymptotic quantiles which implies the restrictions $\theta_{i,0,l}=0$ for all $l>0$. The intercept $\theta_{0,0,0}$ has the interpretation as the asymptotic quantile when both $T\to\infty$ and $k\to\infty$. For a given k, the respective asymptotic quantile can be computed from the coefficients $\theta_{i,0,0}$. When $k=0$, it is $\sum_{i=0}^{r}\theta_{i,0,0}$.

Given the 100 meta replications for each feasible combination of k, T, and q, taking into account the restriction on the degrees of freedom, our regressions are performed on 97,700 observations for case (i), 95,200 observations for cases (ii) and (iii), and 93,400 observations for cases (iv) and (v). While these large numbers of observations imply that the estimation uncertainty conditional on the chosen model becomes practically irrelevant, the uncertainty about the correct specification of the RS remains.[26] Regarding the choice of the polynomial orders r, m, and n, there is no clear guidance and the optimal order possibly differs across the many regressions. As emphasized by MacKinnon (1996), it is important to choose the same specification across quantiles in order to avoid discontinuities in the distributions that are inferred from the predicted values. After extensive experimentation, we found that the polynomial orders $r=4$, $m=3$ and $n=1$ yield satisfactory regression fits, as indicated by the adjusted R-squared or the root mean square error (RMSE). In addition, the coefficients of the interaction terms of the variable count with the inverse sample size are often statistically insignificant when the latter is raised to a higher power. We thus set $\theta_{i,j,l}=0$ when both $i>0$ and $j>1$ to obtain a more parsimonious model. Incorporating all the restrictions, equation (8) becomes

[25] See our Supplementary Appendix for a graphical comparison. Merely adjusting the sample size for the number of estimated coefficients, as done by Ericsson and MacKinnon (2002), does not prove to be a successful strategy.

[26] The variance of the regression errors is a decreasing function in the effective sample size $N(T,q)$ which could be taken into account with a generalized least squares procedure (MacKinnon, 1991) or a generalized method of moments estimator (MacKinnon, 1994, 1996). However, the numerical differences in the predictions are negligible, in particular in the light of the remaining model uncertainty.

$$Q(k,T,q)=\theta_{0,0,0}+\sum_{i=1}^{4}\theta_{i,0,0}\frac{1}{(1+k)^i}+\sum_{j=1}^{3}\theta_{0,j,0}\frac{1}{[N(T,q)]^j}+\sum_{i=1}^{4}\theta_{i,1,0}\frac{1}{(1+k)^iN(T,q)}$$
$$+\sum_{j=1}^{3}\theta_{0,j,1}\frac{H(q,k)}{[N(T,q)]^j}+\sum_{i=1}^{4}\theta_{i,1,1}\frac{H(q,k)}{(1+k)^iN(T,q)}+v. \tag{9}$$

For the t-statistic, as shown by Pesaran *et al.* (2001), the asymptotic distribution does not depend on k when all variables are $I(0)$. Hence, we further restrict $\theta_{i,0,0}=0$ for all $i>0$ in this situation.

The OLS estimates are presented in Tables 2–5 for the quantiles corresponding to a nominal size of 1%, 5% and 10%.[27] For any given k, the fit from equation (9) is expected to be worse than from the tailored regressions in the Supplementary Appendix. However, Figure 4 illustrates that the use of the joint RS model is justified since the differences to the separate RS estimates for each k and the simulated quantiles from our aggregate EDFs are negligible. By contrast, the simple 'meta response surface' estimated by Ericsson and MacKinnon (2002) for the asymptotic quantiles as an affine-linear function of k and the number of deterministic model components is only useful as a crude approximation. It does not readily extend to larger models because it ignores the diminishing slope of the RS with increasing k. Also notice that for the t-statistic in case (i), the predicted asymptotic CVs for $k=0$ (i.e. $\theta_{0,0,0}$ in the $I(0)$ columns of Table 2) differ only marginally at the third decimal place from the exact CVs obtained with the closed-form formulae of Abadir (1995, table 1).

The joint RS model, equation (9), allows us to present the estimates in a more compact way compared to the separate regressions, and to compute the finite-sample CVs for any number k of long-run forcing regressors, effective sample size $N(T,q)$ and number of short-run coefficients $H(q,k)$, as long as there are sufficiently many degrees of freedom. Figure 4 illustrates that for small sample sizes this degrees-of-freedom restriction is often binding. For $T=30$ and $q=1$, the EC model can accommodate at most $k=6$ long-run forcing variables. For larger sample sizes, for example $T=80$, our procedure allows us to predict CVs beyond the maximum k considered in our simulations and the previous literature.

Figure 5 highlights the variation of the RS over the sample size and lag order for selected variable counts. For the F-statistic, the differences across lag orders are more pronounced for the lower-bound CVs that exhibit a slower convergence rate to the respective asymptotic critical value than the upper bounds. Moreover, the convexity of the RS increases with the lag order. While the slope of the RS is negative in q for larger sample sizes, it can become positive for relatively small sample sizes, increasingly so the more long-run forcing variables are in the model. The inconclusive area between the lower and the upper bound widens with increasing lag order. The picture is slightly different for the t-statistic. A larger lag order pulls the CVs closer to zero almost everywhere for both the lower and the upper bound. As seen in Figure 4 before and backed by the asymptotic distributions derived by Pesaran *et al.* (2001), the lower-bound CVs are fairly stable with respect to the number of variables k.

[27] The coefficient estimates for other quantiles are available upon request.

TABLE 2

Response surface estimates, unrestricted deterministic terms

	F-statistic, case (i)						*t-statistic, case (i)*					
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
	I(0)	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*
$\theta_{0,0,0}$	1.397	2.474	1.304	2.381	1.243	2.307	−2.563	−7.324	−1.940	−6.699	−1.617	−6.355
$\theta_{1,0,0}$	10.238	13.318	6.085	7.670	4.244	5.185	–	28.175	–	28.086	–	27.977
$\theta_{2,0,0}$	−11.917	−23.153	−8.183	−14.736	−6.301	−11.122	–	−83.519	–	−83.176	–	−82.848
$\theta_{3,0,0}$	12.636	27.402	8.432	16.224	6.432	11.880	–	114.040	–	113.744	–	113.359
$\theta_{4,0,0}$	−5.334	−12.910	−3.475	−7.333	−2.618	−5.227	–	−53.974	–	−53.904	–	−53.751
$\theta_{0,1,0}$	37.85	78.85	20.23	43.01	14.09	29.50	−7.32	−5.65	−1.52	16.54	1.03	25.36
$\theta_{1,1,0}$	−202.68	−566.21	−105.12	−305.70	−71.41	−211.17	41.31	−61.70	5.34	−232.05	−10.32	−299.83
$\theta_{2,1,0}$	529.54	1681.57	258.87	900.63	167.83	619.63	−115.08	408.01	−10.20	929.59	37.33	1139.10
$\theta_{3,1,0}$	−623.38	−2266.16	−298.28	−1212.82	−187.28	−831.79	144.85	−731.71	7.79	−1442.96	−55.78	−1729.88
$\theta_{4,1,0}$	266.30	1062.87	126.25	570.13	77.54	390.58	−64.61	392.80	−1.78	728.37	27.73	864.15
$\theta_{0,2,0}$	478.0	927.6	197.9	384.5	106.7	221.6	−80.1	−113.4	−18.0	19.4	0.0	64.2
$\theta_{0,3,0}$	−2025	−5155	−1082	−2419	−601	−1440	390	398	68	−474	−41	−811
$\theta_{0,1,1}$	−0.47	0.09	−0.47	−0.03	−0.43	−0.04	0.07	1.41	0.06	1.62	0.09	1.74
$\theta_{1,1,1}$	−3.03	−3.55	−1.07	−0.20	−0.57	0.34	0.48	−7.81	0.47	−8.88	0.44	−9.26
$\theta_{2,1,1}$	12.36	29.99	1.92	4.23	0.17	−0.58	−3.46	22.92	−2.23	27.39	−1.89	28.66
$\theta_{3,1,1}$	−17.22	−68.94	1.03	−14.43	3.50	−2.96	7.90	−26.09	4.49	−35.17	3.45	−37.81
$\theta_{4,1,1}$	7.23	40.55	−2.02	9.61	−3.02	2.81	−4.57	10.58	−2.51	15.79	−1.86	17.34
$\theta_{0,2,1}$	32.4	61.4	24.6	33.9	18.8	22.0	−11.9	−35.9	−4.5	−25.9	−1.9	−23.2
$\theta_{0,3,1}$	327	101	−67	−146	−100	−128	39	258	8	225	−8	218
$\bar{R}^2$	0.9981	0.9936	0.9983	0.9930	0.9979	0.9903	0.9716	0.9988	0.9247	0.9993	0.7756	0.9993
RMSE	0.0752	0.1128	0.0332	0.0445	0.0222	0.0265	0.0165	0.0303	0.0077	0.0217	0.0061	0.0212

Notes: The RS regression model is equation (9). The dependent variable is the simulated α-quantile of the test statistic. $\bar{R}^2$ denotes the adjusted coefficient of determination, and RMSE the root mean square error.

TABLE 3

Response surface estimates, unrestricted deterministic terms (continued)

	F-statistic, case (iii)						*t-statistic, case (iii)*					
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
	I(0)	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*
$\theta_{0,0,0}$	1.384	2.520	1.288	2.393	1.229	2.312	−3.434	−7.477	−2.864	−6.846	−2.569	−6.502
$\theta_{1,0,0}$	12.900	14.953	8.712	9.771	6.749	7.350	–	26.842	–	26.502	–	26.296
$\theta_{2,0,0}$	−6.762	−15.945	−5.249	−11.255	−4.114	−8.781	–	−81.948	–	−80.784	–	−80.231
$\theta_{3,0,0}$	7.395	19.062	6.138	13.343	4.877	10.419	–	112.456	–	110.837	–	110.120
$\theta_{4,0,0}$	−3.091	−8.643	−2.669	−5.986	−2.131	−4.664	–	−53.344	–	−52.583	–	−52.254
$\theta_{0,1,0}$	38.27	73.40	20.44	41.00	14.23	28.51	−3.96	−3.16	4.55	18.92	8.31	28.05
$\theta_{1,1,0}$	−148.82	−424.75	−93.50	−252.76	−70.06	−183.42	−12.72	−116.29	−57.18	−270.22	−78.54	−334.82
$\theta_{2,1,0}$	314.16	1119.40	209.31	697.62	156.16	514.42	88.74	622.79	204.07	1064.84	265.00	1257.22
$\theta_{3,1,0}$	−218.15	−1316.01	−194.25	−874.58	−153.94	−659.30	−162.45	−1064.80	−300.29	−1639.88	−378.60	−1896.22
$\theta_{4,1,0}$	43.00	558.05	65.85	389.49	56.19	298.33	88.17	560.67	147.57	823.86	183.01	942.91
$\theta_{0,2,0}$	510.4	1022.5	268.5	473.3	168.5	288.2	−144.2	−134.0	−54.6	23.6	−26.3	76.3
$\theta_{0,3,0}$	−694	−4506	−1282	−2889	−941	−1918	781	421	340	−563	178	−941
$\theta_{0,1,1}$	−0.21	0.30	−0.34	0.06	−0.36	0.02	0.42	1.49	0.50	1.72	0.55	1.86
$\theta_{1,1,1}$	−7.07	−5.73	−4.08	−1.44	−2.97	−0.60	0.64	−7.59	0.08	−8.81	−0.07	−9.22
$\theta_{2,1,1}$	28.63	43.45	12.44	11.42	7.54	4.49	−7.55	22.11	−3.19	28.00	−2.14	29.74
$\theta_{3,1,1}$	−49.65	−96.97	−21.01	−31.84	−12.49	−16.69	16.64	−24.51	7.54	−36.21	5.13	−39.79
$\theta_{4,1,1}$	25.46	55.11	10.29	18.68	5.83	10.11	−9.57	9.69	−4.40	16.35	−2.98	18.44
$\theta_{0,2,1}$	−4.4	37.5	11.1	28.6	11.4	20.9	−10.3	−36.9	−3.4	−27.9	−0.9	−25.7
$\theta_{0,3,1}$	1667	1335	409	272	148	77	−91	162	−60	189	−53	206
$\bar{R}^2$	0.9992	0.9981	0.9995	0.9989	0.9996	0.9990	0.9811	0.9978	0.9737	0.9986	0.9770	0.9986
RMSE	0.0945	0.1240	0.0438	0.0545	0.0295	0.0354	0.0214	0.0318	0.0109	0.0236	0.0086	0.0233

Notes: The RS regression model is equation (9). The dependent variable is the simulated α-quantile of the test statistic. $\bar{R}^2$ denotes the adjusted coefficient of determination, and RMSE the root mean square error.

TABLE 4

Response surface estimates, unrestricted deterministic terms (continued)

	F-statistic, case (v)						*t-statistic, case (v)*					
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
	I(0)	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*
$\theta_{0,0,0}$	1.358	2.528	1.269	2.392	1.213	2.311	−3.963	−7.617	−3.413	−6.988	−3.130	−6.643
$\theta_{1,0,0}$	16.141	17.284	11.723	12.130	9.622	9.629	–	25.318	–	24.871	–	24.584
$\theta_{2,0,0}$	−4.693	−11.350	−4.045	−8.373	−3.285	−6.314	–	−78.667	–	−77.229	–	−76.459
$\theta_{3,0,0}$	4.487	12.928	4.678	9.992	3.980	7.560	–	108.693	–	106.716	–	105.734
$\theta_{4,0,0}$	−1.616	−5.589	−1.977	−4.444	−1.733	−3.361	–	−51.727	–	−50.792	–	−50.346
$\theta_{0,1,0}$	37.61	69.10	19.91	38.94	13.97	27.23	−1.09	−1.78	8.64	20.57	12.97	29.91
$\theta_{1,1,0}$	−95.04	−310.29	−73.40	−202.99	−60.04	−150.69	−59.46	−152.92	−104.20	−298.15	−126.01	−359.54
$\theta_{2,1,0}$	120.76	716.30	144.89	534.57	125.36	405.58	264.58	753.94	368.07	1156.62	425.66	1333.02
$\theta_{3,1,0}$	182.91	−617.72	−53.29	−597.22	−81.19	−475.09	−429.46	−1258.81	−539.99	−1770.30	−609.66	−1999.80
$\theta_{4,1,0}$	−189.35	181.46	−17.79	240.22	11.78	199.13	221.86	656.73	265.10	887.16	295.28	992.12
$\theta_{0,2,0}$	560.7	1070.2	356.3	554.4	239.2	350.4	−189.1	−151.9	−72.0	28.1	−32.9	91.2
$\theta_{0,3,0}$	1892	−2099	−1231	−2908	−1206	−2219	919	296	414	−712	210	−1134
$\theta_{0,1,1}$	0.23	0.69	−0.13	0.22	−0.21	0.11	0.62	1.50	0.75	1.79	0.83	1.94
$\theta_{1,1,1}$	−14.90	−11.07	−8.92	−4.08	−6.87	−2.29	1.41	−6.02	0.13	−8.05	−0.30	−8.64
$\theta_{2,1,1}$	60.23	66.28	29.78	22.00	20.74	10.59	−12.13	16.59	−4.51	26.27	−2.15	28.97
$\theta_{3,1,1}$	−100.40	−134.12	−48.85	−50.95	−34.13	−28.91	24.69	−16.41	10.27	−34.22	5.66	−39.38
$\theta_{4,1,1}$	50.84	72.95	23.90	27.95	16.43	16.04	−13.83	5.75	−5.91	15.52	−3.34	18.41
$\theta_{0,2,1}$	−26.7	23.4	4.5	27.8	8.2	21.9	−13.7	−42.2	−5.0	−32.3	−1.5	−29.6
$\theta_{0,3,1}$	3049	2726	867	737	374	321	−150	127	−102	178	−92	203
$\bar{R}^2$	0.9990	0.9986	0.9994	0.9991	0.9995	0.9993	0.9837	0.9968	0.9781	0.9977	0.9800	0.9976
RMSE	0.1456	0.1597	0.0746	0.0796	0.0529	0.0558	0.0262	0.0340	0.0149	0.0260	0.0120	0.0256

Notes: The RS regression model is equation (9). The dependent variable is the simulated α-quantile of the test statistic. $\bar{R}^2$ denotes the adjusted coefficient of determination, and RMSE the root mean square error.

TABLE 5

Response surface estimates, restricted deterministic terms

	F-statistic, case (ii)						*F-statistic, case (iv)*					
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
	I(0)	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*	*I(0)*	*I(1)*
$\theta_{0,0,0}$	1.434	2.572	1.317	2.424	1.250	2.335	1.396	2.567	1.289	2.414	1.228	2.326
$\theta_{1,0,0}$	11.747	12.303	7.907	7.601	6.128	5.436	15.146	14.899	10.997	10.123	9.041	7.847
$\theta_{2,0,0}$	−14.731	−22.340	−10.431	−14.840	−8.145	−11.162	−17.183	−22.057	−12.912	−15.385	−10.589	−11.823
$\theta_{3,0,0}$	13.058	25.062	9.598	16.873	7.527	12.888	14.291	22.486	11.219	15.874	9.221	12.104
$\theta_{4,0,0}$	−5.001	−10.980	−3.753	−7.376	−2.947	−5.658	−5.280	−9.415	−4.291	−6.678	−3.538	−5.065
$\theta_{0,1,0}$	34.17	71.58	19.36	41.42	13.69	29.24	34.59	67.71	19.96	40.00	14.26	28.50
$\theta_{1,1,0}$	−125.18	−431.04	−79.23	−258.19	−58.05	−186.25	−87.30	−324.25	−69.90	−212.97	−55.65	−157.96
$\theta_{2,1,0}$	219.96	1097.36	160.93	702.06	118.94	518.05	92.98	721.10	137.90	548.31	117.15	420.96
$\theta_{3,1,0}$	−143.85	−1318.47	−146.77	−889.69	−111.84	−667.42	94.58	−720.70	−93.31	−647.10	−96.30	−512.54
$\theta_{4,1,0}$	24.02	572.61	49.45	401.13	39.31	304.22	−112.20	261.93	15.51	275.05	26.65	223.16
$\theta_{0,2,0}$	606.1	1082.1	273.6	476.0	162.7	287.2	674.2	1155.1	349.2	549.8	224.1	343.8
$\theta_{0,3,0}$	−1951	−5420	−1305	−2849	−843	−1821	−87	−3793	−1193	−2826	−989	−2008
$\theta_{0,1,1}$	−0.44	0.14	−0.46	−0.01	−0.44	−0.02	0.01	0.54	−0.24	0.17	−0.28	0.09
$\theta_{1,1,1}$	−5.19	−4.19	−2.94	−0.61	−2.25	−0.04	−13.74	−10.46	−8.06	−3.71	−6.27	−2.07
$\theta_{2,1,1}$	20.27	34.78	7.34	7.54	4.14	2.18	59.09	63.94	29.48	22.01	21.13	11.62
$\theta_{3,1,1}$	−32.29	−78.92	−8.93	−22.49	−3.40	−10.12	−93.34	−124.76	−43.12	−45.80	−29.57	−25.60
$\theta_{4,1,1}$	15.94	45.66	3.62	13.84	0.78	6.65	46.77	68.47	20.64	25.48	13.80	14.41
$\theta_{0,2,1}$	20.0	58.6	18.6	35.0	15.1	24.0	7.0	53.8	13.8	36.2	12.0	25.7
$\theta_{0,3,1}$	751	398	124	−27	14	−68	1643	1260	462	292	209	117
$\bar{R}^2$	0.9980	0.9934	0.9989	0.9954	0.9992	0.9956	0.9987	0.9968	0.9993	0.9981	0.9995	0.9985
RMSE	0.0718	0.1097	0.0307	0.0445	0.0200	0.0270	0.0843	0.1142	0.0395	0.0506	0.0273	0.0334

Notes: The RS regression model is equation (9). The dependent variable is the simulated α-quantile of the test statistic. $\bar{R}^2$ denotes the adjusted coefficient of determination, and RMSE the root mean square error.

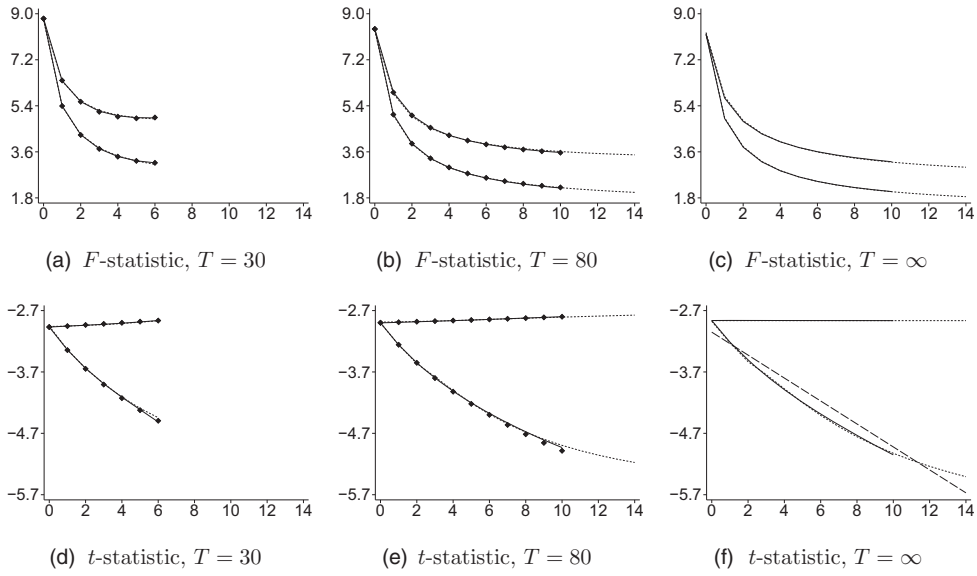


Figure 4 RS for the F- and t-statistic in case (iii) at the 5% significance level over a range of variable numbers k with lag order $q = 1$. The solid curves are the combination of the separate RS estimates for each k for the lower bound (closer to zero) and the upper bound, respectively, and the short-dashed curves are the joint RS estimates from equation (9). The diamonds are the CVs directly computed from our aggregate EDFs. The long-dashed line is the 'meta response surface' from Ericsson and MacKinnon (2002) for the asymptotic upper-bound CVs.

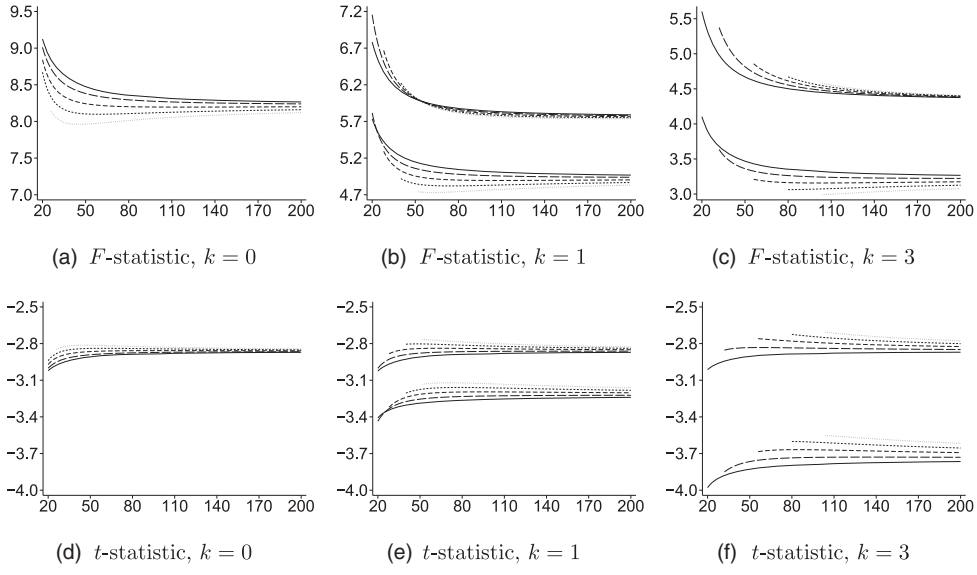


Figure 5 RS from equation (9) for the F- and t-statistic in case (iii) at the 5% significance level over a range of effective sample sizes $N(T, q)$. The solid curves represent the lower bound (closer to zero) and the upper bound for $q = 0$. With increasing lag order, $q \in \{0, 3, 6, 9, 12\}$, the curves have shorter dashes.

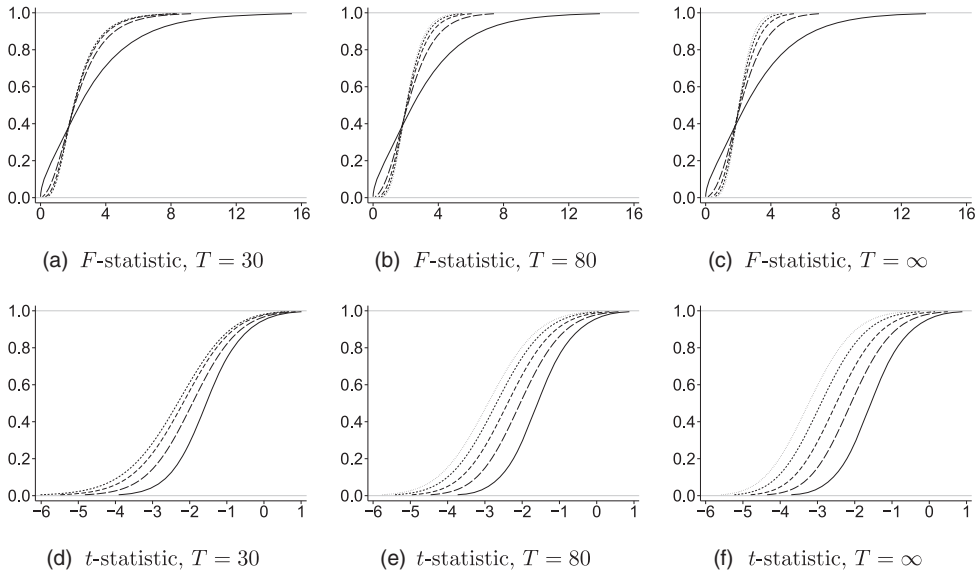


Figure 6 Implied upper-bound cumulative distribution functions from equation (9) for the F- and t-statistic in case (iii) with $k \in \{0, 2, 4, 6, 8\}$ variables and lag order $q = 1$. The solid curve refers to $k = 0$. With increasing k, the curves have shorter dashes.

Approximate *P*-values

With the RS regressions from above for a fine grid of quantiles, we can already describe the shape of the finite-sample and asymptotic distributions quite well. To obtain a P-value corresponding to any given value of the test statistic, we still need to interpolate between the two nearest quantiles for which we have obtained predictions. We follow MacKinnon (1996) and Ericsson and MacKinnon (2002) regarding the choice of 221 quantiles that we compute for each test statistic:

$$P \in \{0.0001, 0.0002, 0.0005, 0.001, \ldots, 0.01, 0.015, \\ \ldots, 0.99, 0.991, \ldots, 0.999, 0.9995, 0.9998, 0.9999\}.$$

Some of the resulting cumulative distribution functions are shown in Figure 6. It is apparent again that the differences diminish with increasing number of long-run forcing variables, and that the shape of the distributions varies with the sample size.

To obtain P-values, MacKinnon (1994, 1996) suggests a local approximation strategy. Consider the following regression model:

$$F^{-1}(p) = \sum_{i=0}^{n} \phi_i \left[\hat{Q}(p)\right]^i + e, \tag{10}$$

where $F^{-1}(p)$ is the inverse cumulative distribution function of the test statistic that would apply under standard asymptotics,[28] and $\hat{Q}(p)$ is the predicted p-quantile from equation (9)

[28] We use the F-distribution with appropriate degrees of freedom to approximate the shape of the distribution for the F-statistic and the t-distribution for the t-statistic.

for a given combination of k, T and q.[29] If the distributional assumption was correct, then model (10) would be correctly specified with $\phi_1 = 1$ and all other coefficients being zero. $\phi_0 \neq 0$ allows for a shift in the mean and $\phi_1 \neq 1$ for a different variance. Since in our case this regression only serves as an approximation of the unknown shape of the distribution, the higher-order terms potentially help to improve the fit. It turns out that for our purpose a second-order polynomial, $n = 2$, works sufficiently well.

Equation (10) is then estimated for the nine predicted quantiles that are nearest to the observed value of the test statistic. MacKinnon (1994, 1996) notices that an OLS estimation ignores heteroskedasticity and pairwise correlation of the quantiles, and he suggests to estimate equation (10) by generalized least squares (GLS). However, we do not find that a GLS estimation uniformly improves the fit. For practical purposes, a feasible GLS estimation requires estimates of the variances of the respective quantiles. While the variance estimates can in principle be obtained from the RS regressions, this would require to supply the variance–covariance matrices from all estimations together with the computer program that computes the approximate p-values. From our perspective, it seems worth to trade off minor efficiency gains for the convenience of not having to store this larger amount of data, again emphasizing that such efficiency gains are negligible in the light of the remaining model uncertainty.

The approximate P-value corresponding to the observed value of the test statistic τ is finally computed as

$$\hat{p} = F\left(\sum_{i=0}^{n} \hat{\phi}_i \tau^i\right), \tag{11}$$

where $\hat{\phi}_i$ are the coefficient estimates from equation (10). This procedure to approximate P-values, as well as the critical value predictions from equation (9), is implemented in the *Stata* program described by Kripfganz and Schneider (2018) for both the F-statistic and the t-statistic. Figure 7 illustrates the resulting P-value curves for the right tail of the F-distribution. These P-values can help us to shed some light on the relevance of the differences between the finite-sample and the asymptotic CVs. When we compute a finite-sample P-value for a test statistic τ that equals the asymptotic critical value, we can interpret this P-value as the finite-sample size of the asymptotic test.

For example, consider a situation with $k = 4$ variables, $T = 30$ data points, $q = 1$ lag for each variable, and an unrestricted intercept. Our RS regressions predict an asymptotic upper-bound critical value of 3.994 at a significance level of 5%. The finite-sample upper-bound P-value that corresponds to this value is 0.112 such that we do not even reject the null hypothesis at the 10% significance level. The asymptotic test is substantially oversized in such a small sample. If we ignored the presence of the short-run coefficients, the P-value would slip back by almost 1%-point to 0.103. These differences can be quite relevant in empirical work. With a larger sample size, the asymptotic CVs obviously become better approximations. When we move to $T = 80$ in our example, the correct finite-sample p-value falls to 0.067 which still implies that the asymptotic test is oversized by a practically relevant magnitude. Because the number of short-run coefficients is now small relative to

[29] For convenience, we are suppressing the arguments k, T, q in favour of p that is variable in this regression.

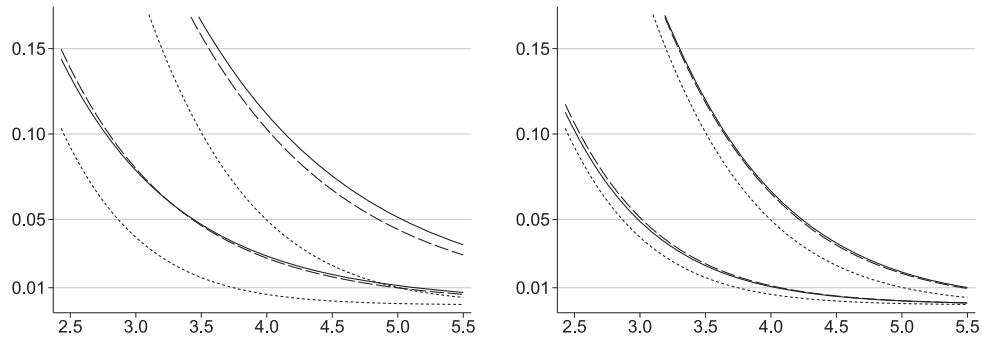


Figure 7 Approximate lower-bound and upper-bound p-value curves from equation (11) for the F-statistic in case (iii) with $k = 4$ variables. The solid curve is obtained accounting for the lag order $q = 1$. The long-dashed curve ignores the presence of the short-run coefficients by setting $q = 0$, and the short-dashed curve relates to the asymptotic distribution.

the sample size, the lag order no longer plays much of a role. For higher lag orders, the P-value curves would still be visibly distinct even for moderately large sample sizes.

For the F-statistic, size distortions of more than 5%-points are not uncommon, in particular in models with a large number of long-run forcing variables. Furthermore, the distortions tend to be stronger in cases with restricted rather than unrestricted deterministic model components. For the t-statistic, we find less reasons to be overly concerned about the use of the asymptotic CVs. The expected size distortions remain mostly below 2%-points. This is in line with our earlier observation in Figure 5 that the RS for the t-statistic is much flatter than for the F-statistic. More detailed information on the finite-sample size distortions can be found in our Supplementary Appendix.

To provide an example of the practical relevance of our RS estimates, we finish with a reassessment of the empirical results of Mills and Pentecost (2001) regarding the effects of the real exchange rate on the gross domestic product in Hungary, controlling for the effects of the real money supply and real wages. Thus, $k = 3$. The time span is very short with only $N(T, q) = 22$ effective observations. Mills and Pentecost (2001) report F-statistics for the cases (iv) and (v) with restricted and unrestricted trend. They are 4.55 and 5.68 respectively. The asymptotic P-values between the lower and upper bounds span the range 0.006–0.030 under case (iv) and 0.005–0.023 under case (v), indicating conclusive evidence in favour of a long-run relationship at the 5% significance level. With our finite-sample CVs, ignoring the presence of any short-run coefficients, the P-values increase to 0.046–0.104 in case (iv) and 0.033–0.081 in case (v), turning the test results inconclusive at the 5% significance level, and even at the 10% level in case (iv). Assuming a total of $H(q, k) = 5$ short-run coefficients, the P-value bounds shift further upward to 0.051–0.125 and 0.040–0.099 respectively.[30] In case (iv), the F-test is now conclusive against a long-run relationship at the 5% significance level. In case (v), the bounds test only marginally escapes becoming

[30] Mills and Pentecost (2001) chose $q = 2$ but then excluded an unreported number of insignificant short-run terms.

inconclusive at the 10% level.[31] Eventually, we can conclude not to reject the null hypothesis of no long-run relationship because the value of the t-statistic, -2.94, remains closer to zero than the 10% lower bound, both under asymptotic and finite-sample CVs.

IV. Conclusion

The Pesaran *et al.* (2001) bounds test for the existence of a level relationship is widely applied in the empirical practice. The current paper provides RS estimates for the respective lower-bound and upper-bound CVs, corresponding to the situations where all long-run forcing variables are either $I(0)$ or $I(1)$, respectively. Finite-sample and asymptotic CVs for various cases of unrestricted or restricted deterministic model components and any number of long-run forcing variables can be computed directly from the regression tables. While such CVs have been reported previously in the literature, they often only cover a rather small subset of the possible model specifications and sample sizes, and they are typically less precise due to a smaller number of replications in the respective Monte Carlo simulations.

With the exception of Cheung and Lai (1995a) for the augmented Dickey–Fuller test that results as a special case of the framework considered here, the previously obtained response surfaces do not account for the lag augmentation in the underlying regression model. With our RS estimates, finite-sample critical value bounds can be obtained for any number of short-run coefficients. In practice, the correct lag order is usually unknown and possibly different across variables. For the purpose of efficient estimation of the model coefficients, an optimal lag order is often obtained with model selection criteria such as the Akaike or Schwarz information criterion. However, as stressed by Pesaran *et al.* (2001) and underlined by our simulation results, for testing purposes it is of primary concern that the error term is free of serial correlation. As long as there are enough degrees of freedom available, additional lags of the variables can help to achieve this aim. Once a conclusion from the test is drawn, a more parsimonious model can be estimated along the lines of the Pesaran and Shin (1998) autoregressive distributed lag (ARDL) modelling approach. In the statistical software *Stata*, the ARDL and EC models can be estimated with the same program that computes the CVs and approximate P-values for the bounds test (Kripfganz and Schneider, 2018).

The finite-sample CVs in this paper are obtained by simulating a VAR(1) process with independent and identically normally distributed disturbances. Under more general data-generating processes, our CVs are still useful as approximations in the absence of exact results. Our simulation approach can be easily adapted to specific deviations from the DGP considered here. An alternative inferential approach is to bootstrap the finite-sample distributions of the test statistics. Bootstrapping avoids the parametric assumptions of our simulations at the cost of additional computational burden. McNown *et al.* (2018) outline a residual-based bootstrap procedure for the test statistics considered in this paper. They use a modification of the sieve bootstrap algorithm proposed by Palm, Smeekes and Urbain (2010) for the cointegration Wald statistic in the conditional error correction

[31] Besides Hungary, Mills and Pentecost (2001) test their hypothesis for the Czech Republic, Poland and Slovakia. For these three countries, the test results are unambiguous irrespective of the chosen CVs.

model.[32] By leaving the process for the long-run forcing regressors unrestricted, McNown *et al.* (2018) avoid the potential problem of inconclusive evidence when the integration order of the regressors is *a priori* unknown. A shortcoming of the bootstrap procedure is that residuals for each variable in the underlying VAR model need to be estimated and resampled, not just for the single-equation EC model of interest. Eventually, there is still a lack of comprehensive evidence about the relative performance of the bootstrap, in particular when there are many variables in the model and the sample size is small.

Final Manuscript Received: April 2020

References

Abadir, K. M. (1995). 'The limiting distribution of the t ratio under a unit root', *Econometric Theory*, Vol. 11, pp. 775–793.

Banerjee, A., Dolado, J. J. and Mestre, R. (1998). 'Error-correction mechanism tests for cointegration in a single-equation framework', *Journal of Time Series Analysis*, Vol. 19, pp. 267–283.

Carrion-i-Silvestre, J. L., Sansó Rosselló, A. and Artís Ortuño, M. (1999). 'Response surface estimates for the Dickey-Fuller unit root test with structural breaks', *Economics Letters*, Vol. 63, pp. 279–283.

Chang, Y. and Park, J. Y. (2003). 'A sieve bootstrap for the test of a unit root', *Journal of Time Series Analysis*, Vol. 24, pp. 379–400.

Cheung, Y.-W. and Lai, K. S. (1995a). 'Lag order and critical values of the augmented Dickey-Fuller test', *Journal of Business & Economic Statistics*, Vol. 13, pp. 277–280.

Cheung, Y.-W. and Lai, K. S. (1995b). 'Lag order and critical values of a modified Dickey- Fuller test', *Oxford Bulletin of Economics and Statistics*, Vol. 57, pp. 411–419.

Cook, S. (2001). 'Finite-sample critical values of the augmented Dickey-Fuller statistic: a note on lag order', *Economic Issues*, Vol. 6, pp. 31–38.

Dickey, D. A. (1976). *Estimation and Hypothesis Testing in Nonstationary Time Series*. Ph. D. thesis, Iowa State University.

Dickey, D. A. and Fuller, W. A. (1979). 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, Vol. 74, pp. 427–431.

Dickey, D. A. and Fuller, W. A. (1981). 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica*, Vol. 49, pp. 1057–1072.

Engle, R. F. and Granger, C. W. J. (1987). 'Co-integration and error correction: representation, estimation, and testing', *Econometrica*, Vol. 55, pp. 251–276.

Ericsson, N. R. and MacKinnon, J. G. (2002). 'Distributions of error correction tests for cointegration', *Econometrics Journal*, Vol. 5, pp. 285–318.

Ferretti, N. and Romo, J. (1996). 'Unit root bootstrap tests for AR(1) models', *Biometrika*, Vol. 83, pp. 849–860.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

Harvey, D. I. and van Dijk, D. (2006). 'Sample size, lag order and critical values of seasonal unit root tests', *Computational Statistics & Data Analysis*, Vol. 50, pp. 2734–2751.

Hassler, U. and Wolters, J. (2006). 'Autoregressive distributed lag models and cointegration', *Allgemeines Statistisches Archiv*, Vol. 90, pp. 59–74.

Kanioura, A. and Turner, P. (2005). 'Critical values for an F-test for cointegration in a multivariate model', *Applied Economics*, Vol. 37, pp. 265–270.

Kripfganz, S. and Schneider, D. C. (2018). *ardl: Estimating Autoregressive Distributed Lag and Equilibrium Correction Models*. Proceedings of the 2018 London Stata Conference.

[32] Earlier bootstrap contributions for the special case of the (augmented) Dickey–Fuller unit-root *t*-statistic include Ferretti and Romo (1996), Psaradakis (2001), Park (2002, 2003), Chang and Park (2003), and Swensen (2003).

MacKinnon, J. G. (1991). 'Critical values for cointegration tests', in Engle R. F. and Granger C. W. J. (eds), *Long-Run Economic Relationships: Readings in Cointegration*, Chapter 13, Oxford: Oxford University Press, pp. 267–276.
MacKinnon, J. G. (1994). 'Approximate asymptotic distribution functions for unit-root and cointegration tests', *Journal of Business & Economic Statistics*, Vol. 12, pp. 167–176.
MacKinnon, J. G. (1996). 'Numerical distribution functions for unit root and cointegration tests', *Journal of Applied Econometrics*, Vol. 11, pp. 601–618.
MacKinnon, J. G. (2010). *Critical values for cointegration tests*. QED Working Paper 1227, Queen's University, Department of Economics.
MacKinnon, J. G., Haug, A. A. and Michelis, L. (1999). 'Numerical distribution functions of likelihood ratio tests for cointegration', *Journal of Applied Econometrics*, Vol. 14, pp. 563–577.
McNown, R., Sam, C. Y. and Goh, S. K. (2018). 'Bootstrapping the autoregressive distributed lag test for cointegration', *Applied Economics*, Vol. 50, pp. 1509–1521.
Mills, T. C. and Pentecost, E. J. (2001). 'The real exchange rate and the output response in four EU accession countries', *Emerging Markets Review*, Vol. 2, pp. 418–430.
Narayan, P. K. (2005). 'The saving and investment nexus for China: evidence from cointegration tests', *Applied Economics*, Vol. 37, pp. 1979–1990.
Narayan, P. K. and Smyth, R. (2004). 'Crime rates, male youth unemployment and real income in Australia: evidence from Granger causality tests', *Applied Economics*, Vol. 36, pp. 2079–2095.
Otero, J. and Baum, C. F. (2017). 'Response surface models for the Elliott, Rothenberg, and Stock unit-root test', *Stata Journal*, Vol. 17, pp. 985–1002.
Otero, J. and Smith, J. (2012). 'Response surface models for the Leybourne unit root tests and lag order dependence', *Computational Statistics*, Vol. 27, pp. 473–486.
Otero, J. and Smith, J. (2017). 'Response surface models for OLS and GLS detrendingbased unit-root tests in nonlinear ESTAR models', *Stata Journal*, Vol. 17, pp. 704–722.
Palm, F. C., Smeekes, S. and Urbain, J.-P. (2010). 'A sieve bootstrap test for cointegration in a conditional error correction model', *Econometric Theory*, Vol. 26, pp. 647–681.
Park, J. Y. (2002). 'An invariance principle for sieve bootstrap in time series', *Econometric Theory*, Vol. 18, pp. 469–490.
Park, J. Y. (2003). 'Bootstrap unit root tests', *Econometrica*, Vol. 71, pp. 1845–1895.
Pesaran, M. H. and Shin, Y. (1998). 'An autoregressive distributed-lag modelling approach to cointegration analysis', in Strøm S. (ed), *Econometrics and Economic Theory in the 20th Century*. The Ragnar Frisch Centennial Symposium, Chapter 11, Cambridge: Cambridge University Press, pp. 371–413.
Pesaran, M. H., Shin, Y. and Smith, R. J. (2000). 'Structural analysis of vector error correction models with exogenous I(1) variables', *Journal of Econometrics*, Vol. 97, pp. 293–343.
Pesaran, M. H., Shin, Y. and Smith, R. J. (2001). 'Bounds testing approaches to the analysis of level relationships', *Journal of Applied Econometrics*, Vol. 16, pp. 289–326.
Presno, M. J. and López, A. J. (2003). 'Response surface estimates of stationarity tests with a structural break', *Economics Letters*, Vol. 78, pp. 395–399.
Psaradakis, Z. (2001). 'Bootstrap tests for an autoregressive unit root process in the presence of weakly dependent errors', *Journal of Time Series Analysis*, Vol. 22, pp. 577–594.
Said, S. E. and Dickey, D. A. (1984). 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika*, Vol. 71, pp. 599–607.
Sam, C. Y., McNown, R. and Goh, S. K. (2018). 'An augmented autoregressive distributed lag bounds test for cointegration', *Economic Modelling forthcoming*, pp. 1–12.
Sephton, P. S. (1995). 'Response surface estimates of the KPSS stationarity test', *Economics Letters*, Vol. 47, pp. 255–261.
Sephton, P. S. (2008). 'Critical values of the augmented fractional Dickey-Fuller test', *Empirical Economics*, Vol. 35, pp. 437–450.
Sephton, P. S. (2017). 'Finite sample critical values of the generalized KPSS stationarity test', *Computational Economics*, Vol. 50, pp. 161–172.
Swensen, A. R. (2003). 'Bootstrapping unit root tests for integrated processes', *Journal of Time Series Analysis*, Vol. 24, pp. 99–126.

Turner, P. (2006). 'Response surfaces for an F-test for cointegration', *Applied Economics Letters*, Vol. 13, pp. 479–482.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix A: Details on the computational methods
Appendix B: Alternative data-generating processes
Appendix C: Separate response surface regressions
Appendix D: Critical values and approximate *P*-values