Response surface regressions for critical value bounds and approximate p-values in equilibrium correction models*

Sebastian Kripfganz† Daniel C. Schneider‡

January 29, 2019

Abstract

Single-equation conditional equilibrium correction models can be used to test for the existence of a level relationship among the variables of interest. The distributions of the respective test statistics are nonstandard under the null hypothesis of no such relationship and critical values need to be obtained with stochastic simulations. We compute more than 95 billion F-statistics and 57 billion t-statistics for a large number of specifications of the Pesaran, Shin, and Smith (2001, *Journal of Applied Econometrics* 16: 289–326) bounds test. Our large-scale simulations enable us to draw smooth density functions and to estimate response surface models that improve upon and substantially extend the set of available critical values for the bounds test. Besides covering the full range of possible sample sizes and lag orders, our approach notably allows for any number of variables in the long-run level relationship by exploiting the diminishing effect on the distributions of adding another variable to the model. The computation of approximate p-values enables a fine-grained statistical inference and allows us to quantify the finite-sample distortions from using asymptotic critical values. We find that the bounds test can be easily oversized by more than 5 percentage points in small samples.

Keywords: Bounds test; Cointegration; Error correction model; Generalized Dickey-Fuller regression; Level relationship; Unit roots

JEL Classification: C12; C15; C32; C46; C63

*We are grateful for comments by Mehdi Hosseinkouchack.

†Corresponding author: University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter, EX4 4PU, UK. Tel.: +44-1392-722110; E-mail: S.Kripfganz@exeter.ac.uk

‡Max Planck Institute for Demographic Research, Konrad-Zuse-Straße 1, 18057 Rostock, Germany. Tel.: +49-381-2081245; E-mail: schneider@demogr.mpg.de

1 Introduction

The empirical analysis of time series data is often confronted with test statistics that have nonstandard 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 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 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 nonnegligible. Given the vast number of empirically relevant regression specifications that lead to possibly different distributions, the tabulation of critical values 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 et al. (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 critical values. However, the asymptotic distributions might be poor

[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.

approximations of the actual distributions in small samples. Finite-sample critical values 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 critical values 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 critical values 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 critical values 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 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 more than 95 billion F-statistics and 57 billion t-statistics under several scenarios regarding the deterministic model components, number of variables, sam-

[2]Previously tabulated critical values for a small set of sample sizes can be found in Fuller (1976) and Dickey (1976) for the univariable and Banerjee et al. (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 critical values 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 et al. (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.

ple 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 critical values 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 critical values 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 percentage points when using the asymptotic rather than the small-sample critical values. 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.

2 Bounds testing for the existence of 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.

2.1 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:

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

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

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 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 $-1 < 1/z \leq 1$. The data-generating process 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 data-generating process 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 $0 \leq r \leq k$.

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 data-generating process 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

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

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

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

$\boldsymbol{\omega} = \mathbf{\Omega}_{xx}^{-1} \boldsymbol{\omega}_{xy}$ 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, $u_t \sim \mathcal{N}(\mathbf{0}, \omega_{yy} - \boldsymbol{\omega}_{xy}' \mathbf{\Omega}_{xx}^{-1} \boldsymbol{\omega}_{xy})$.

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 nonzero element of $\boldsymbol{\pi}$, y_t is generated by a trend-stationary or $I(0)$ process not involving the levels of $\mathbf{x}_t$.

2.2 Bounds test

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

(1) Test the joint null hypothesis $H_0^{\boldsymbol{\pi}} : \boldsymbol{\pi} = \mathbf{0}$ versus the alternative hypothesis $H_1^{\boldsymbol{\pi}} : \boldsymbol{\pi} \neq \mathbf{0}$.

[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.

(2) If $H_0^{\boldsymbol{\pi}}$ 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 $0 < r \leq k$.

(3) If $H_0^{\pi_{yy}}$ is rejected, test the joint hypothesis $H_0^{\boldsymbol{\theta}} : \boldsymbol{\theta} = \mathbf{0}$ versus $H_1^{\boldsymbol{\theta}} : \boldsymbol{\theta} \neq \mathbf{0}$, 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 $H_1^{\boldsymbol{\pi}}$ 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 nondegenerate 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 nonstandard 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 data-generating process. Without this assumption, the asymptotic distribution of the t-statistic would depend on nuisance parameters and tabulations of critical values 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 nonstandard distributions.[9] The remainder of this text is primarily concerned with the test statistics in steps (1) and (2).

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 behavior.[10] Pesaran et al. (2001) distinguish five cases, depending

[8] See Pesaran et al. (2001) for a discussion. Banerjee et al. (1998) assume $r = 0$ and briefly argue that the critical values 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 et al. (2018) tabulate critical values for a Wald test of joint insignificance of up to 7 long-run forcing variables in the level relationship.

[10] See Pesaran et al. (2000) for details.

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 case (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 $H_0^{\boldsymbol{\pi}}$. For the single-hypothesis test in step (2), the restrictions can be ignored.

Pesaran et al. (2001) argue that the critical values 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]

3 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 $H_0^{\boldsymbol{\pi}}$ in step (1) and the t-statistic under $H_0^{\pi_{yy}}$ in step (2).[12] They tabulate the critical values for the range of $k \in [0, 10]$ long-run forcing variables. Several other authors provide finite-sample critical values for a subset of the relevant situations. We summarize the existing literature in Table 1.[13] A number of authors tabulated critical values that require interpolations

[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

	$T-q$	q	k	$I(d)$	deterministics cases[+] 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	≥ 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)	–

Note: The regression model used by these authors 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, i.e. $k = 0$, the specifications are equivalent for $q = 0$ and $q = 1$.
*MacKinnon (1996) and Ericsson and MacKinnon (2002) provide computer programs that compute the critical values 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.

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 critical values, 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 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 critical values in that situation. For the unit-root F-statistic, we are the first to provide comprehensive RS estimates.[14]

[14]Dickey and Fuller (1981) tabulate a few critical values for the restricted intercept or trend cases (ii) and (iv). While the F-statistic in the unrestricted cases (i), (iii), and (v) equals the square of the t-statistic, this is not true for the quantiles of the corresponding distributions. Consequently, separate critical values

In the multivariable setting, the lag order dependence of finite-sample critical values 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 3 long-run forcing variables.

3.1 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 $H_0^{\boldsymbol{\pi}}$ 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 $I(1)$.[15]

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

need to be obtained.

[15] The data-generating process is identical to the one used by Pesaran et al. (2001), besides the discarded observations.

representation (3) the presence of the term $\boldsymbol{\omega}'\Delta\mathbf{x}_t$ induces an overparameterization when $q = 0$.[16] In cases (i), (iii), and (v), under the null hypothesis $H_0^{\boldsymbol{\pi}}$, 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 2 integration orders and 5 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.[17] This yields a total of 9,528 simulation designs.[18] For each design, we run 100,000 replications and then repeat the entire procedure 100 times, which we refer to as 'meta replications'. We thus compute a total number of 9.528×10^{10} F-statistics and 5.744×10^{10} t-statistics.[19] 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.[20] The effect of rounding on the RS regressions is absolutely negligible.

The 10 million statistics for each configuration are sufficiently many to draw smooth probability density functions without the need for sophisticated kernel density estimators. With a bin width of 0.1, Figures 1 and 2 are 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). In particular for the F-statistic, the shape

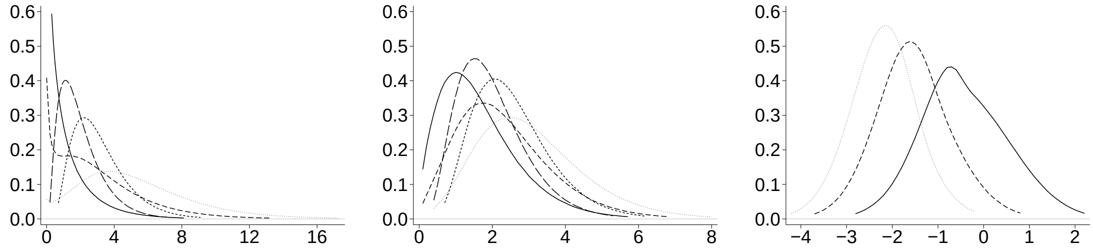
[16]The lag specification $q = 0$ can be obtained from the VAR(1) model in equation (1) by imposing the restriction $\boldsymbol{\omega} = \boldsymbol{\varphi}$.

[17]That is $\max(1, q) + k(q + 1) + \mathcal{I}(c_0 \neq 0) + \mathcal{I}(c_1 \neq 0) \leq (T - \max(q, 1))/2$, 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 $k = 0$.

[18]There are 1,960 simulation designs for case (i), 1,910 designs for cases (ii) and (iii) each, and 1,874 designs for cases (iv) and (v), respectively.

[19]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.

[20]Details on the compression procedure as well as other computational aspects are relegated to the Supplementary Appendix.



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

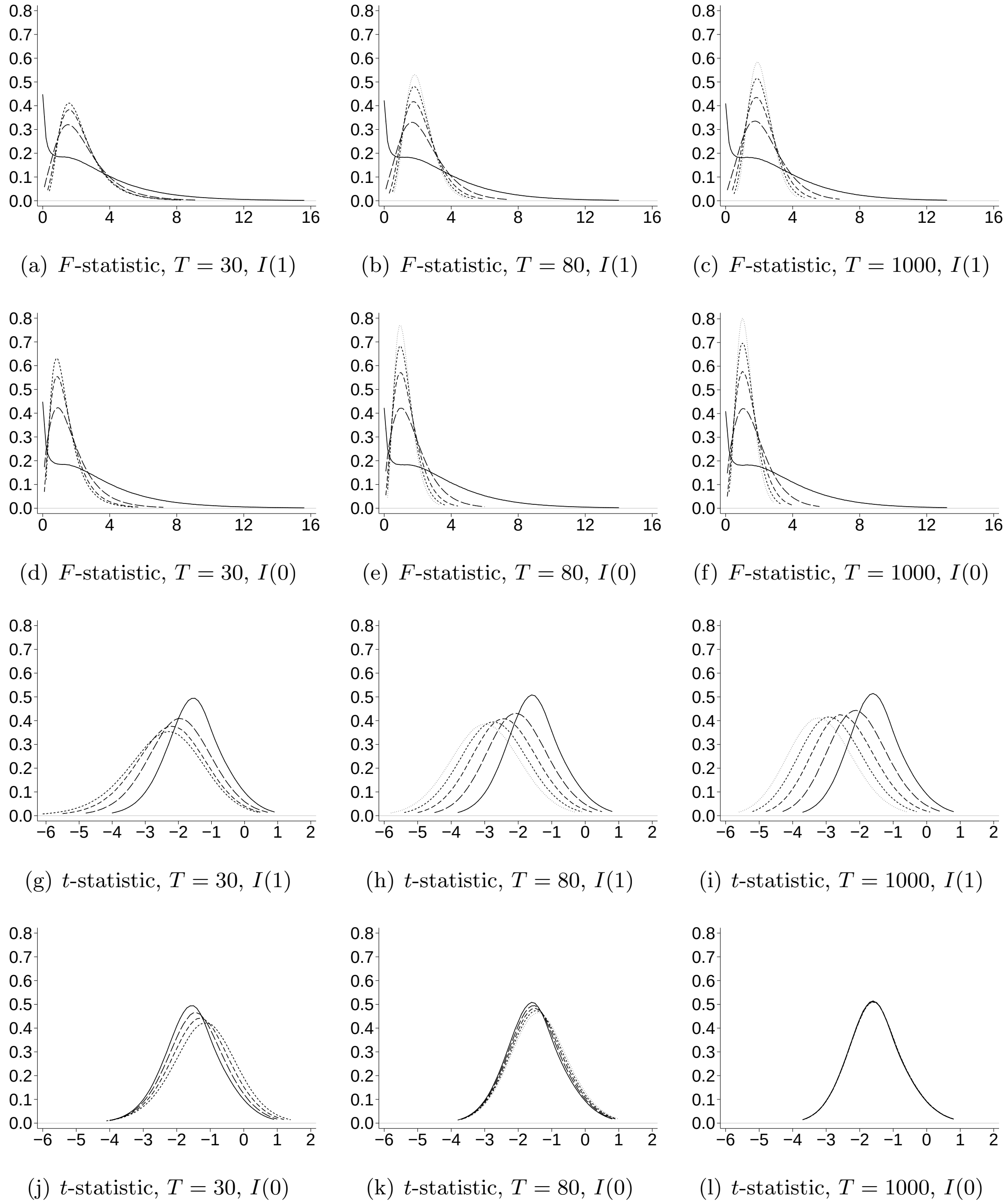
Figure 1: Probability density functions obtained from the 10^7 simulated F-statistics for each of the five cases and the t-statistics for cases (i), (iii), and (v) with sample size $T = 1000$ and lag order $q = 1$. With increasing case number, the curves have shorter dashes. For $k = 2$, the upper-bound densities are shown. We restrict the plots to the quantile interval $p \in [0.005, 0.995]$.

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 critical values.

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 have the familiar bell shape but are not centered around zero. 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).

Figure 2 highlights 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 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.

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



(a) F-statistic, $T = 30$, $I(1)$ (b) F-statistic, $T = 80$, $I(1)$ (c) F-statistic, $T = 1000$, $I(1)$

(d) F-statistic, $T = 30$, $I(0)$ (e) F-statistic, $T = 80$, $I(0)$ (f) F-statistic, $T = 1000$, $I(0)$

(g) t-statistic, $T = 30$, $I(1)$ (h) t-statistic, $T = 80$, $I(1)$ (i) t-statistic, $T = 1000$, $I(1)$

(j) t-statistic, $T = 30$, $I(0)$ (k) t-statistic, $T = 80$, $I(0)$ (l) t-statistic, $T = 1000$, $I(0)$

Figure 2: Upper-bound and lower-bound probability density functions obtained from the 10^7 simulated F- and t-statistics, respectively, in case (iii) with $k \in \{0, 2, 4, 6, 8\}$ variables for different sample sizes T and with lag order $q = 1$. The solid curve refers to $k = 0$. With increasing k, the curves have shorter dashes. We restrict the plots to the quantile interval $p \in [0.005, 0.995]$.

The densities are as well less dispersed with larger sample size but more dispersed with increasing number of 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 95-th percentile of the F-statistic with $k = 2$ long-run forcing variables that are $I(1)$, $T = 1000$ observations, and a lag order of $q = 2$ is found to be 4.81. Pesaran et al. (2001) report a critical value of 4.85 for the same 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 99-th 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 95-th percentile and 0.33% for the 99-th 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 95-th 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 nonnegligible 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 critical

[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).

values tabulated by Narayan (2005) that do not comply with the monotonic decline of the finite-sample toward the asymptotic quantiles due to the experimental randomness.

3.2 Response surface regressions

The tabulation of all empirically relevant critical values would be cumbersome since it would stretch dozens of pages. Moreover, even though we have obtained EDFs from 9,528 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 critical values 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. The intercept in such a regression can be interpreted as the quantile of the asymptotic distribution.

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. 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 MacKinnon (2002) for the t-statistic. Yet, their critical values 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 critical values would provide a better approximation. The Cheung and Lai (1995a) RS addresses this problem but is slightly skewed towards zero compared to ours.[23]

Carrying out RS estimations separately for each k has two shortcomings. First, this approach does not allow to obtain critical values if the actual number of long-run forcing

[23]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.

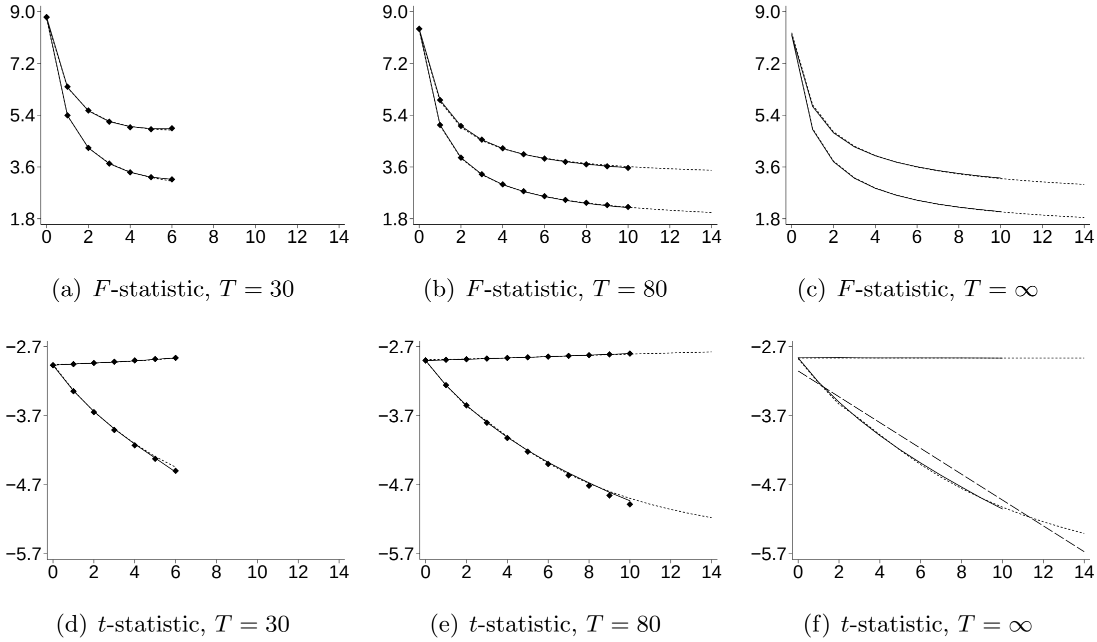
variables has not been considered in the simulations. Second, any attempt to cover a larger range of k inflates the number of regression results that need to be tabulated or stored in a computer program. In the following, we overcome this problem by directly modeling the RS as a function of k. A close 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:

$$Q(k, T, q) = \sum_{i=0}^{r} \sum_{j=0}^{m} \sum_{l=0}^{n} \theta_{i,j,l} (1 + k)^{-i} [N(T, q)]^{-j} [H(q, k)]^{l} + \nu, \tag{7}$$

where $Q(k, T, q)$ refers to the quantiles from the meta-replication EDFs, $N(T, q) = T - \max(q, 1)$ 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 ν is the regression error. 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 98,000 observations for case (i), 95,500 observations for cases (ii) and (iii), and 93,700 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.[24] 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 sat-

[24] 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.



(a) F-statistic, $T = 30$ (b) F-statistic, $T = 80$ (c) F-statistic, $T = \infty$

(d) t-statistic, $T = 30$ (e) t-statistic, $T = 80$ (f) t-statistic, $T = \infty$

Figure 3: RS for the F- and t-statistic in case (iii) at the 5% significance level over a range of variable numbers k for different sample sizes T and 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 (8). Panels (a)–(b) and (d)–(e) show finite-sample results. The diamonds are the critical values directly computed from our aggregate EDFs. Panels (c) and (f) show the asymptotic results. The long-dashed line is the "meta response surface" from Ericsson and MacKinnon (2002) for the asymptotic upper-bound critical values.

isfactory 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 (7) becomes

$$Q(k,T,q) = \theta_{0,0,0} + \sum_{i=1}^{4} \theta_{i,0,0} \frac{1}{(1+k)^i} + \sum_{j=1}^{3} \theta_{0,j,0} \frac{1}{[N(T,q)]^j} + \sum_{i=1}^{4} \theta_{i,1,0} \frac{1}{(1+k)^i N(T,q)} + \sum_{j=1}^{3} \theta_{0,j,1} \frac{H(q,k)}{[N(T,q)]^j} + \sum_{i=1}^{4} \theta_{i,1,1} \frac{H(q,k)}{(1+k)^i N(T,q)} + \nu. \quad (8)$$

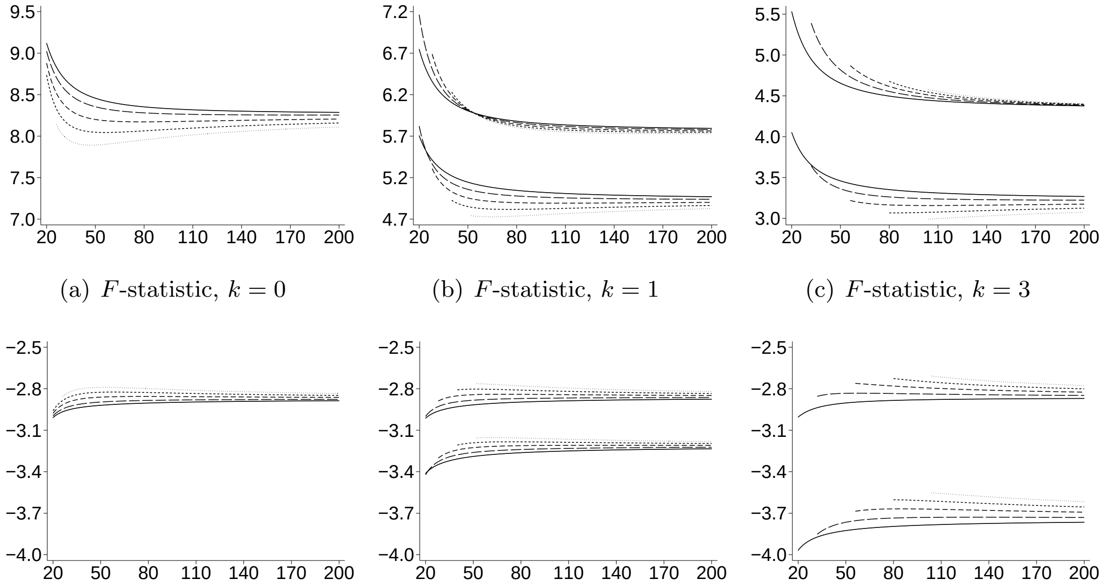
For the t-statistic, as shown by Pesaran et al. (2001), the asymptotic distribution does not depend on k when all variables are $I(0)$. Hence, we further restrict $\theta_{i,0,0} = 0$ for all $i > 0$ in this situation.

The OLS estimates are presented in Tables 2 to 4 in Appendix A for the quantiles corresponding to a nominal size of 1%, 5%, and 10%.[25] For any given k, the fit from equation (8) is expected to be worse than from the tailored regressions in the Supplementary Appendix. However, Figure 3 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.

The joint RS model, equation (8), allows us to present the estimates in a more compact way compared to the separate regressions, and to compute the finite-sample critical values 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 3 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 critical values beyond the maximum k considered in our simulations and the previous literature.

Figure 4 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 critical values 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 critical values closer to zero almost everywhere for both the lower and the upper bound. As seen in Figure 3 before and backed by the asymptotic distributions derived by Pesaran et al. (2001), the lower-bound critical values

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



(a) F-statistic, $k = 0$ (b) F-statistic, $k = 1$ (c) F-statistic, $k = 3$

(d) t-statistic, $k = 0$ (e) t-statistic, $k = 1$ (f) t-statistic, $k = 3$

Figure 4: RS from equation (8) 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.

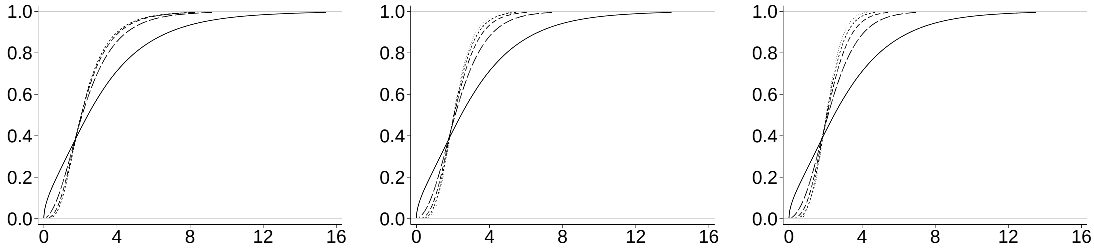
are fairly stable with respect to the number of variables k.

3.3 Approximate p-values

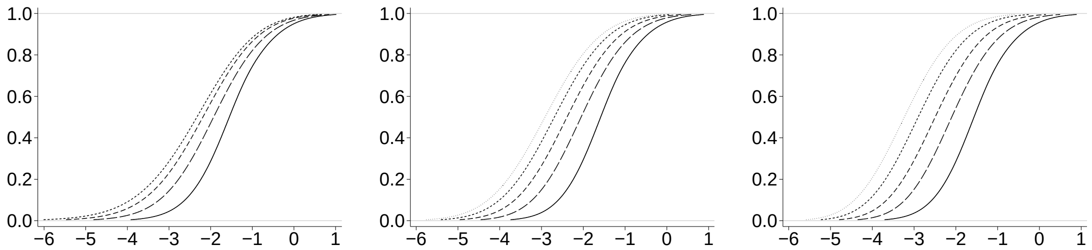
With the RS regressions from Section 3.2 for a fine grid of quantiles, we can already describe the shape of the finite-sample and asymptotic distributions quite well. To obtain a p-value corresponding to any given value of the test statistic, we still need to interpolate between the two nearest quantiles 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 5. 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.



(a) F-statistic, $T = 30$ (b) F-statistic, $T = 80$ (c) F-statistic, $T = \infty$



(d) t-statistic, $T = 30$ (e) t-statistic, $T = 80$ (f) t-statistic, $T = \infty$

Figure 5: Implied upper-bound cumulative distribution functions from equation (8) for the F- and t-statistic in case (iii) with $k \in \{0, 2, 4, 6, 8\}$ variables for different sample sizes T and with lag order $q = 1$. The solid curve refers to $k = 0$. With increasing k, the curves have shorter dashes. We restrict the plots to the quantile interval $p \in [0.005, 0.995]$.

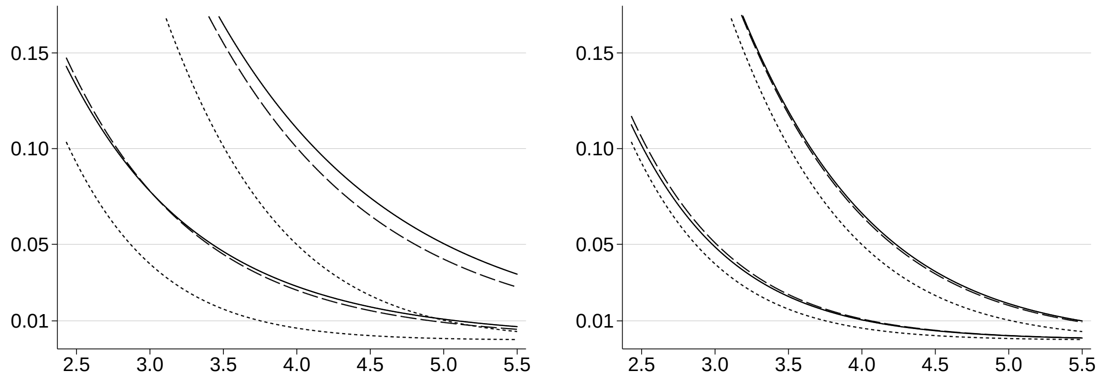
To obtain p-values, MacKinnon (1994, 1996) suggests a local approximation strategy. Consider the following regression model:

$$F^{-1}(p) = \sum_{i=0}^{n} \phi_i \left[\hat{Q}(p)\right]^i + e, \tag{9}$$

where $F^{-1}(p)$ is the inverse cumulative distribution function of the test statistic that would apply under standard asymptotics,[26] and $\hat{Q}(p)$ is the predicted p-quantile from equation (8) for a given combination of k, T, and q.[27] If the distributional assumption was correct, then model (9) 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.

[26] 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.

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



(a) $T = 30$

(b) $T = 80$

Figure 6: Approximate lower-bound and upper-bound p-value curves from equation (10) for the F-statistic in case (iii) with $k = 4$ variables. The solid curve is obtained taking the lag order $q = 1$ into account. 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.

Equation (9) is then estimated for the 9 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 (9) 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

$$\hat{p} = F\left(\sum_{i=0}^{n} \hat{\phi}_i \tau^i\right), \tag{10}$$

where $\hat{\phi}_i$ are the coefficient estimates from equation (9). This procedure to approximate p-values, as well as the critical-value predictions from equation (8), is implemented in the *Stata* program described by Kripfganz and Schneider (2018) for both the F-statistic and the t-statistic. Figure 6 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 critical values. 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 4.00 at a significance level of 5%. The finite-sample upper-bound p-value that corresponds to this value is 0.111 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 more than one percentage point to 0.100. These differences can be quite relevant in empirical work. With a larger sample size, the asymptotic critical values 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 test is oversized by a practically relevant magnitude. Because the number of short-run coefficients is now small relative to the sample size, the lag order no longer plays a big 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 percentage 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 critical values. The expected size distortions remain mostly below two percentage points. This is in line with our earlier observation in Figure 4 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.

4 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 response surface estimates for the respective lower-bound and upper-bound critical values, corresponding to the situations where all long-run forcing variables are either $I(0)$ or $I(1)$, respectively. Precise finite-sample and asymptotic critical values 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 critical values 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 response surface estimates, accurate 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), 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 critical values and approximate p-values for the bounds test (Kripfganz and Schneider, 2018).

References

Banerjee, A., J. J. Dolado, and R. Mestre (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis 19*(3), 267–283.

Carrion-i-Silvestre, J. L., A. Sansó Rosselló, and M. Artís Ortuño (1999). Response surface estimates for the Dickey-Fuller unit root test with structural breaks. *Economics Letters 63*(3), 279–283.

Cheung, Y.-W. and K. S. Lai (1995a). Lag order and critical values of the augmented Dickey-Fuller test. *Journal of Business & Economic Statistics 13*(3), 277–280.

Cheung, Y.-W. and K. S. Lai (1995b). Lag order and critical values of a modified Dickey-Fuller test. *Oxford Bulleting of Economics and Statistics 57*(3), 411–419.

Cook, S. (2001). Finite-sample critical values of the augmented Dickey-Fuller statistic: A note on lag order. *Economic Issues 6*(2), 31–38.

Dickey, D. A. (1976). *Estimation and hypothesis testing in nonstationary time series.* Ph. D. thesis, Iowa State University.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366), 427–431.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica 49*(4), 1057–1072.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica 55*(2), 251–276.

Ericsson, N. R. and J. G. MacKinnon (2002). Distributions of error correction tests for cointegration. *Econometrics Journal 5*(2), 285–318.

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

Harvey, D. I. and D. van Dijk (2006). Sample size, lag order and critical values of seasonal unit root tests. *Computational Statistics & Data Analysis 50*(10), 2734–2751.

Hassler, U. and J. Wolters (2006). Autoregressive distributed lag models and cointegration. *Allgemeines Statistisches Archiv 90*(1), 59–74.

Kanioura, A. and P. Turner (2005). Critical values for an F-test for cointegration in a multivariate model. *Applied Economics 37*(3), 265–270.

Kripfganz, S. and D. C. Schneider (2018). ardl: Estimating autoregressive distributed lag and equilibrium correction models. Proceedings of the 2018 London Stata Conference.

MacKinnon, J. G. (1991). Critical values for cointegration tests. In R. F. Engle and C. W. J. Granger (Eds.), *Long-Run Economic Relationships: Readings in Cointegration*, Chapter 13, pp. 267–276. Oxford: Oxford University Press.

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics 12*(2), 167–176.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics 11*(6), 601–618.

MacKinnon, J. G. (2010). Critical values for cointegration tests. QED Working Paper 1227, Queen's University, Department of Economics.

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics 14*(5), 563–577.

McNown, R., C. Y. Sam, and S. K. Goh (2018). Bootstrapping the autoregressive distributed lag test for cointegration. *Applied Economics 50*(13), 1509–1521.

Mills, T. C. and E. J. Pentecost (2001). The real exchange rate and the output response in four EU accession countries. *Emerging Markets Review 2*(4), 418–430.

Narayan, P. K. (2005). The saving and investment nexus for China: evidence from cointegration tests. *Applied Economics 37*(17), 1979–1990.

Narayan, P. K. and R. Smyth (2004). Crime rates, male youth unemployment and real income in Australia: evidence from Granger causality tests. *Applied Economics 36*(18), 2079–2095.

Otero, J. and C. F. Baum (2017). Response surface models for the Elliott, Rothenberg, and Stock unit-root test. *Stata Journal 17*(4), 985–1002.

Otero, J. and J. Smith (2012). Response surface models for the Leybourne unit root tests and lag order dependence. *Computational Statistics 27*(3), 473–486.

Otero, J. and J. Smith (2017). Response surface models for OLS and GLS detrending-based unit-root tests in nonlinear ESTAR models. *Stata Journal 17*(3), 704–722.

Pesaran, M. H. and Y. Shin (1998). An autoregressive distributed-lag modelling approach to cointegration analysis. In S. Strøm (Ed.), *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium*, Chapter 11, pp. 371–413. Cambridge: Cambridge University Press.

Pesaran, M. H., Y. Shin, and R. J. Smith (2000). Structural analysis of vector error correction models with exogenous $I(1)$ variables. *Journal of Econometrics 97*(2), 293–343.

Pesaran, M. H., Y. Shin, and R. J. Smith (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics 16*(3), 289–326.

Presno, M. J. and A. J. López (2003). Response surface estimates of stationarity tests with a structural break. *Economics Letters 78*(3), 395–399.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*(3), 599–607.

Sam, C. Y., R. McNown, and S. K. Goh (2018). An augmented autoregressive distributed lag bounds test for cointegration. *Economic Modelling forthcoming*, 1–12.

Sephton, P. S. (1995). Response surface estimates of the KPSS stationarity test. *Economics Letters 47*(3–4), 255–261.

Sephton, P. S. (2008). Critical values of the augmented fractional Dickey-Fuller test. *Empirical Economics 35*(3), 437–450.

Sephton, P. S. (2017). Finite sample critical values of the generalized KPSS stationarity test. *Computational Economics 50*(1), 161–172.

Turner, P. (2006). Response surfaces for an F-test for cointegration. *Applied Economics Letters 13*(8), 479–482.

Appendix A Tables for Section 3.2

Table 2: Response surface estimates, F-statistic, unrestricted deterministic terms

		α = 1%		α = 5%		α = 10%	
		$I(0)$	$I(1)$	$I(0)$	$I(1)$	$I(0)$	$I(1)$
Case (i)							
	$\theta_{0,0,0}$	1.3696	2.4281	1.2942	2.3622	1.2370	2.2960
	$\theta_{1,0,0}$	10.6537	14.0067	6.2409	7.9543	4.3366	5.3600
	$\theta_{2,0,0}$	-13.6561	-26.0058	-8.8317	-15.9274	-6.7007	-11.8646
	$\theta_{3,0,0}$	15.4161	31.9173	9.4594	18.1120	7.0777	13.0640
	$\theta_{4,0,0}$	-6.7631	-15.2134	-3.9997	-8.2957	-2.9522	-5.8334
	$\theta_{0,1,0}$	43.832	89.090	22.341	46.696	15.234	31.655
	$\theta_{1,1,0}$	-300.208	-733.557	-141.425	-368.295	-91.727	-247.906
	$\theta_{2,1,0}$	974.321	2450.577	426.737	1192.902	263.505	792.198
	$\theta_{3,1,0}$	-1361.786	-3547.985	-577.890	-1701.901	-347.942	-1121.248
	$\theta_{4,1,0}$	652.692	1734.590	272.626	826.596	161.970	542.515
	$\theta_{0,2,0}$	452.19	878.73	186.33	360.06	98.79	205.25
	$\theta_{0,3,0}$	-2057.4	-4987.2	-1060.6	-2279.6	-573.0	-1327.0
	$\theta_{0,1,1}$	-0.753	-0.407	-0.572	-0.223	-0.495	-0.146
	$\theta_{1,1,1}$	1.199	3.901	0.494	2.658	0.350	1.939
	$\theta_{2,1,1}$	-9.034	-7.868	-6.079	-10.421	-4.562	-8.815
	$\theta_{3,1,1}$	19.607	-3.609	14.778	10.738	11.604	11.147
	$\theta_{4,1,1}$	-12.451	5.657	-9.339	-3.762	-7.324	-4.659
	$\theta_{0,2,1}$	39.31	74.26	27.88	39.99	20.90	25.73
	$\theta_{0,3,1}$	331.8	56.3	-76.3	-179.4	-107.6	-152.4
	$\bar{R}^2$	0.9980	0.9934	0.9982	0.9927	0.9977	0.9898
	RMSE	0.0769	0.1146	0.0344	0.0453	0.0230	0.0271
Case (iii)							
	$\theta_{0,0,0}$	1.3503	2.4703	1.2769	2.3748	1.2232	2.3014
	$\theta_{1,0,0}$	13.3980	15.6809	8.8811	10.0441	6.8420	7.5100
	$\theta_{2,0,0}$	-8.8477	-18.9414	-5.9499	-12.3812	-4.5039	-9.4560
	$\theta_{3,0,0}$	10.7169	23.7481	7.2420	15.1088	5.4957	11.4995
	$\theta_{4,0,0}$	-4.7944	-11.0137	-3.2305	-6.8804	-2.4466	-5.2185
	$\theta_{0,1,0}$	44.026	82.066	22.411	43.843	15.248	30.049
	$\theta_{1,1,0}$	-237.098	-559.081	-125.009	-299.888	-86.715	-209.389
	$\theta_{2,1,0}$	705.899	1722.883	353.204	914.499	233.623	636.496
	$\theta_{3,1,0}$	-852.779	-2300.592	-430.577	-1230.482	-281.745	-861.454
	$\theta_{4,1,0}$	370.539	1067.711	188.684	573.944	122.676	403.538
	$\theta_{0,2,0}$	458.05	937.48	243.44	434.66	151.87	263.02
	$\theta_{0,3,0}$	-569.4	-4085.7	-1161.5	-2636.4	-843.0	-1730.9
	$\theta_{0,1,1}$	-0.377	0.014	-0.413	-0.041	-0.397	-0.045
	$\theta_{1,1,1}$	-4.906	-1.843	-3.105	-0.185	-2.442	0.186
	$\theta_{2,1,1}$	17.922	23.522	7.352	4.744	4.819	0.253
	$\theta_{3,1,1}$	-32.520	-63.725	-12.473	-20.859	-8.010	-9.674
	$\theta_{4,1,1}$	16.604	37.596	5.760	12.971	3.483	6.457
	$\theta_{0,2,1}$	3.30	51.23	14.71	35.04	13.60	24.82
	$\theta_{0,3,1}$	1722.7	1352.0	423.0	264.2	154.4	69.5
	$\bar{R}^2$	0.9992	0.9981	0.9995	0.9988	0.9996	0.9990
	RMSE	0.0917	0.1235	0.0433	0.0548	0.0294	0.0357
Case (v)							
	$\theta_{0,0,0}$	1.3230	2.4837	1.2588	2.3775	1.2077	2.3022
	$\theta_{1,0,0}$	16.6413	17.8970	11.8758	12.3407	9.7073	9.7565
	$\theta_{2,0,0}$	-6.7467	-13.7784	-4.6466	-9.1796	-3.6210	-6.8118
	$\theta_{3,0,0}$	7.7086	16.6029	5.5895	11.1665	4.4870	8.2963
	$\theta_{4,0,0}$	-3.2515	-7.4078	-2.4278	-5.0074	-1.9835	-3.7182
	$\theta_{0,1,0}$	42.712	75.204	21.326	40.536	14.657	27.962
	$\theta_{1,1,0}$	-166.956	-396.087	-96.367	-228.627	-72.165	-163.748
	$\theta_{2,1,0}$	427.486	1081.487	243.336	645.539	178.211	463.610
	$\theta_{3,1,0}$	-296.668	-1179.408	-204.692	-764.883	-162.466	-563.067
	$\theta_{4,1,0}$	53.384	461.642	57.774	322.476	52.225	242.258
	$\theta_{0,2,0}$	492.69	959.84	329.78	507.05	222.76	321.51
	$\theta_{0,3,0}$	1789.0	-1775.4	-1224.8	-2686.6	-1172.6	-2048.6
	$\theta_{0,1,1}$	0.160	0.605	-0.155	0.202	-0.226	0.108
	$\theta_{1,1,1}$	-14.178	-10.448	-8.744	-4.124	-6.794	-2.399
	$\theta_{2,1,1}$	56.969	63.238	28.919	21.946	20.278	10.891
	$\theta_{3,1,1}$	-96.582	-131.545	-48.211	-51.837	-33.755	-29.984
	$\theta_{4,1,1}$	49.063	72.154	23.735	28.658	16.316	16.757
	$\theta_{0,2,1}$	-22.42	33.34	8.11	33.83	10.69	25.79
	$\theta_{0,3,1}$	3272.7	2911.1	923.9	774.2	399.5	334.1
	$\bar{R}^2$	0.9991	0.9986	0.9994	0.9991	0.9996	0.9993
	RMSE	0.1414	0.1584	0.0730	0.0794	0.0518	0.0558

Note: The RS regression model is equation (8). 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, F-statistic, restricted deterministic terms

		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
		$I(0)$	$I(1)$	$I(0)$	$I(1)$	$I(0)$	$I(1)$
Case (ii)							
	$\theta_{0,0,0}$	1.4019	2.5281	1.3055	2.4053	1.2438	2.3247
	$\theta_{1,0,0}$	12.2206	12.9344	8.0776	7.8819	6.2268	5.5885
	$\theta_{2,0,0}$	-16.7218	-24.9204	-11.1424	-16.0068	-8.5678	-11.8062
	$\theta_{3,0,0}$	16.2339	29.0801	10.7213	18.7103	8.2050	13.9175
	$\theta_{4,0,0}$	-6.6301	-13.0084	-4.3245	-8.3088	-3.2954	-6.1863
	$\theta_{0,1,0}$	39.756	79.524	21.222	44.059	14.650	30.617
	$\theta_{1,1,0}$	-211.809	-555.426	-109.353	-302.872	-74.456	-210.728
	$\theta_{2,1,0}$	606.821	1660.558	298.970	909.712	195.959	634.363
	$\theta_{3,1,0}$	-772.210	-2238.762	-372.888	-1231.818	-239.173	-860.622
	$\theta_{4,1,0}$	348.517	1048.841	166.626	578.749	105.581	404.890
	$\theta_{0,2,0}$	555.39	997.78	249.93	439.08	147.87	263.95
	$\theta_{0,3,0}$	-1810.3	-4949.6	-1216.5	-2627.9	-778.9	-1670.9
	$\theta_{0,1,1}$	-0.612	-0.156	-0.532	-0.110	-0.482	-0.084
	$\theta_{1,1,1}$	-2.983	-0.255	-2.052	0.670	-1.745	0.735
	$\theta_{2,1,1}$	9.329	14.773	2.659	0.623	1.487	-2.007
	$\theta_{3,1,1}$	-14.823	-45.998	-1.203	-10.996	0.963	-3.229
	$\theta_{4,1,1}$	6.966	28.541	-0.427	7.841	-1.494	3.082
	$\theta_{0,2,1}$	27.98	72.06	22.15	41.23	17.36	28.00
	$\theta_{0,3,1}$	792.8	399.7	139.3	-32.8	20.5	-75.2
	$\bar{R}^2$	0.9979	0.9931	0.9989	0.9951	0.9991	0.9954
	RMSE	0.0728	0.1116	0.0315	0.0455	0.0206	0.0277
Case (iv)							
	$\theta_{0,0,0}$	1.3614	2.5266	1.2800	2.4010	1.2237	2.3182
	$\theta_{1,0,0}$	15.6352	15.4478	11.1327	10.3089	9.1079	7.9667
	$\theta_{2,0,0}$	-19.2144	-24.2083	-13.4541	-16.0919	-10.8478	-12.2982
	$\theta_{3,0,0}$	17.4958	25.7188	12.0425	16.8948	9.5992	12.8153
	$\theta_{4,0,0}$	-6.9111	-11.0076	-4.7001	-7.1658	-3.7211	-5.4127
	$\theta_{0,1,0}$	39.351	73.260	21.358	41.489	14.974	29.179
	$\theta_{1,1,0}$	-153.902	-402.409	-91.075	-236.859	-67.028	-170.087
	$\theta_{2,1,0}$	376.096	1053.617	230.114	654.846	167.876	477.117
	$\theta_{3,1,0}$	-343.581	-1227.709	-235.226	-810.125	-174.785	-598.955
	$\theta_{4,1,0}$	107.545	512.691	86.203	355.536	65.796	265.809
	$\theta_{0,2,0}$	605.48	1043.38	317.08	499.07	203.89	311.82
	$\theta_{0,3,0}$	-83.0	-3304.2	-1119.2	-2548.4	-926.5	-1816.6
	$\theta_{0,1,1}$	-0.041	0.448	-0.265	0.143	-0.305	0.083
	$\theta_{1,1,1}$	-13.289	-9.709	-7.844	-3.565	-6.091	-2.124
	$\theta_{2,1,1}$	57.446	60.349	28.529	21.045	20.218	11.650
	$\theta_{3,1,1}$	-92.826	-121.561	-42.514	-45.218	-28.516	-26.291
	$\theta_{4,1,1}$	47.007	67.559	20.570	25.470	13.356	14.957
	$\theta_{0,2,1}$	11.42	64.46	16.86	42.00	14.18	29.49
	$\theta_{0,3,1}$	1837.1	1399.5	525.7	329.7	242.1	135.2
	$\bar{R}^2$	0.9987	0.9967	0.9993	0.9981	0.9995	0.9984
	RMSE	0.0816	0.1145	0.0381	0.0506	0.0264	0.0333

Note: The RS regression model is equation (8). 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, t-statistic

		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 10\%$	
		$I(0)$	$I(1)$	$I(0)$	$I(1)$	$I(0)$	$I(1)$
Case (i)							
	$\theta_{0,0,0}$	-2.5640	-7.3169	-1.9398	-6.6952	-1.6169	-6.3527
	$\theta_{1,0,0}$	–	28.0719	–	28.0268	–	27.9488
	$\theta_{2,0,0}$	–	-83.1077	–	-82.9353	–	-82.7398
	$\theta_{3,0,0}$	–	113.4083	–	113.3623	–	113.1899
	$\theta_{4,0,0}$	–	-53.6575	–	-53.7088	–	-53.6652
	$\theta_{0,1,0}$	-8.304	-7.718	-1.813	15.466	0.836	24.494
	$\theta_{1,1,0}$	58.111	-27.385	10.675	-216.221	-7.296	-288.605
	$\theta_{2,1,0}$	-190.469	251.065	-34.105	862.473	24.671	1095.707
	$\theta_{3,1,0}$	268.585	-471.426	47.185	-1335.576	-35.494	-1663.496
	$\theta_{4,1,0}$	-128.897	256.894	-22.303	673.336	17.319	830.904
	$\theta_{0,2,0}$	-77.82	-104.71	-17.93	19.25	-1.31	60.12
	$\theta_{0,3,0}$	408.5	368.5	71.3	-462.5	-30.7	-769.2
	$\theta_{0,1,1}$	0.141	1.526	0.088	1.654	0.103	1.761
	$\theta_{1,1,1}$	-0.554	-9.560	0.139	-9.373	0.261	-9.517
	$\theta_{2,1,1}$	1.248	31.426	-0.692	29.576	-1.169	29.520
	$\theta_{3,1,1}$	0.206	-40.412	1.904	-38.664	2.305	-39.009
	$\theta_{4,1,1}$	-0.573	18.108	-1.144	17.578	-1.279	17.913
	$\theta_{0,2,1}$	-12.91	-38.58	-4.86	-26.41	-1.81	-22.92
	$\theta_{0,3,1}$	38.2	269.6	10.0	226.4	-8.7	213.5
	$\bar{R}^2$	0.9716	0.9987	0.9249	0.9993	0.7784	0.9993
	RMSE	0.0164	0.0313	0.0077	0.0218	0.0060	0.0210
Case (iii)							
	$\theta_{0,0,0}$	-3.4345	-7.4681	-2.8642	-6.8423	-2.5692	-6.4989
	$\theta_{1,0,0}$	–	26.6999	–	26.4474	–	26.2648
	$\theta_{2,0,0}$	–	-81.3601	–	-80.5666	–	-80.1186
	$\theta_{3,0,0}$	–	111.5262	–	110.4990	–	109.9530
	$\theta_{4,0,0}$	–	-52.8701	–	-52.4117	–	-52.1717
	$\theta_{0,1,0}$	-5.169	-4.975	3.978	17.930	7.895	27.135
	$\theta_{1,1,0}$	6.222	-87.411	-48.810	-256.411	-73.151	-323.489
	$\theta_{2,1,0}$	4.981	491.885	169.065	1008.331	244.286	1215.153
	$\theta_{3,1,0}$	-27.267	-849.868	-245.128	-1551.623	-347.306	-1833.769
	$\theta_{4,1,0}$	18.680	449.038	119.545	779.224	167.469	912.197
	$\theta_{0,2,0}$	-132.33	-116.85	-51.83	24.99	-26.52	72.38
	$\theta_{0,3,0}$	698.9	321.3	319.2	-552.9	178.4	-894.6
	$\theta_{0,1,1}$	0.493	1.564	0.527	1.740	0.571	1.862
	$\theta_{1,1,1}$	-0.271	-8.609	-0.261	-9.025	-0.236	-9.286
	$\theta_{2,1,1}$	-3.465	27.142	-1.827	28.843	-1.595	29.721
	$\theta_{3,1,1}$	10.250	-32.807	5.502	-37.386	4.401	-39.509
	$\theta_{4,1,1}$	-6.340	14.024	-3.395	16.914	-2.643	18.232
	$\theta_{0,2,1}$	-12.39	-39.97	-3.93	-28.44	-0.86	-25.33
	$\theta_{0,3,1}$	-87.5	165.9	-60.4	183.6	-55.5	197.3
	$\bar{R}^2$	0.9812	0.9977	0.9733	0.9986	0.9767	0.9986
	RMSE	0.0211	0.0328	0.0109	0.0239	0.0086	0.0232
Case (v)							
	$\theta_{0,0,0}$	-3.9636	-7.6120	-3.4137	-6.9848	-3.1299	-6.6410
	$\theta_{1,0,0}$	–	25.2386	–	24.8309	–	24.5586
	$\theta_{2,0,0}$	–	-78.3522	–	-77.0823	–	-76.3657
	$\theta_{3,0,0}$	–	108.2191	–	106.5084	–	105.5947
	$\theta_{4,0,0}$	–	-51.4929	–	-50.6949	–	-50.2777
	$\theta_{0,1,0}$	-1.976	-2.965	8.094	19.655	12.510	29.063
	$\theta_{1,1,0}$	-47.416	-135.028	-96.698	-285.958	-120.306	-349.458
	$\theta_{2,1,0}$	214.910	675.344	340.159	1109.102	406.410	1297.359
	$\theta_{3,1,0}$	-354.745	-1135.951	-499.934	-1699.376	-583.357	-1948.866
	$\theta_{4,1,0}$	185.122	595.013	245.932	852.345	283.037	967.675
	$\theta_{0,2,0}$	-171.00	-129.71	-70.36	30.19	-35.26	86.35
	$\theta_{0,3,0}$	815.8	208.9	424.6	-677.7	238.6	-1055.7
	$\theta_{0,1,1}$	0.658	1.539	0.767	1.804	0.838	1.949
	$\theta_{1,1,1}$	1.012	-6.526	-0.032	-8.204	-0.378	-8.747
	$\theta_{2,1,1}$	-10.677	18.823	-4.107	26.704	-2.151	29.025
	$\theta_{3,1,1}$	23.051	-19.457	10.038	-34.568	6.008	-39.147
	$\theta_{4,1,1}$	-13.179	7.158	-5.906	15.612	-3.606	18.214
	$\theta_{0,2,1}$	-15.77	-44.36	-5.61	-32.39	-1.55	-28.80
	$\theta_{0,3,1}$	-164.2	95.8	-108.0	158.9	-96.4	184.5
	$\bar{R}^2$	0.9836	0.9966	0.9777	0.9976	0.9796	0.9976
	RMSE	0.0261	0.0350	0.0149	0.0263	0.0121	0.0255

Note: The RS regression model is equation (8). 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.