NBER WORKING PAPER SERIES

LOW-FREQUENCY ROBUST COINTEGRATION TESTING

Ulrich Müller
Mark W. Watson

Working Paper 15292
http://www.nber.org/papers/w15292

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2009

We thank participants of the Cowles Econometrics Conference, the NBER Summer Institute and the Greater New York Metropolitan Area Econometrics Colloquium, and of seminars at Chicago, Cornell, Northwestern, NYU, Rutgers, and UCSD for helpful discussions. Support was provided by the National Science Foundation through grants SES-0518036 and SES-0617811. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Low-Frequency Robust Cointegration Testing
Ulrich Müller and Mark W. Watson
NBER Working Paper No. 15292
August 2009
JEL No. C32,E32

ABSTRACT

Standard inference in cointegrating models is fragile because it relies on an assumption of an I(1) model for the common stochastic trends, which may not accurately describe the data's persistence. This paper discusses efficient low-frequency inference about cointegrating vectors that is robust to this potential misspecification. A simple test motivated by the analysis in Wright (2000) is developed and shown to be approximately optimal in the case of a single cointegrating vector.

Ulrich Müller
Department of Economics
Princeton University
Princeton, NJ 08544-1013
umueller@princeton.edu

Mark W. Watson
Department of Economics
Princeton University
Princeton, NJ 08544-1013
and NBER
mwatson@princeton.edu

1 Introduction

The fundamental insight of cointegration is that while economic time series may be individually highly persistent, some linear combinations are much less persistent. Accordingly, a suite of practical methods have been developed for conducting inference about cointegrating vectors, the coefficients that lead to this reduction in persistence. In their standard form, these methods assume that the persistence is the result of common $I(1)$ stochastic trends,[1] and their statistical properties crucially depend on particular characteristics of $I(1)$ processes. But in many applications there is uncertainty about the correct model for the persistence which cannot be resolved by examination of the data, rendering standard inference potentially fragile. This paper studies efficient inference methods for cointegrating vectors that is robust to this fragility.

We do this using a transformation of the data that focuses on low-frequency variability and covariability. This transformation has two distinct advantages. First, as we have argued elsewhere (Müller and Watson (2008)), persistence ("trending behavior") and lack of persistence ("non-trending, $I(0)$ behavior") are low-frequency characteristics, and attempts to utilize high-frequency variability to learn about low-frequency variability are fraught with their own fragilities.[2] Low-frequency transformations eliminate these fragilities by focusing attention on the features of the data that are of direct interest for questions relating to persistence. The second advantage is an important by-product of discarding high frequency variability. The major technical challenge when conducting robust inference about cointegrating vectors is to control size over the range of plausible processes characterizing the model's stochastic common trends. Restricting attention to low frequencies greatly reduces the dimensionality of this challenge.

The inference problem studied in this paper has a long history. Elliott (1998) provides a dramatic demonstration of the fragility of standard cointegration methods by showing that they fail to control size when the common stochastic trends are not $I(1)$, but rather are "local-to-unity" in the sense of Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987) and Phillips (1987).[3] In a bivariate model, Cavanagh, Elliott, and Stock (1995) propose several

[1] See, for instance, Johansen (1988), Phillips and Hansen (1990), Saikkonen (1991), Park (1992) and Stock and Watson (1993).

[2] Perhaps the most well-known example of this fragility involves estimation of HAC standard errors, see Newey and West (1987), Andrews (1991), den Haan and Levin (1997), Kiefer, Vogelsang, and Bunzel (2000), Kiefer and Vogelsang (2005), Müller (2007) and Sun, Phillips, and Jin (2008).

[3] Also see Elliott and Stock (1994) and Jeganathan (1997).

procedures to adjust critical values from standard tests to control size over a range of values of the local-to-unity parameter, and their general approach has been used by several other researchers; Campbell and Yogo (2006) provides a recent example. Stock and Watson (1996) and Jansson and Moreira (2006) go further and develop inference procedures with specific optimality properties in the local-to-unity model. In the fractional cointegration literature, the common stochastic trends are modelled as fractionally integrated, although the problem is different from the local-to-unity case as the fractional parameter can be consistently estimated under standard asymptotics. Yet, Müller and Watson (2008) demonstrate that, at least based on below business cycle variation, it is a hopeless endeavor to try to consistently discriminate between, say, local-to-unity and fractionally integrated stochastic data spanning 50 years.[4] While local-to-unity and fractional processes generalize the assumption of $I(1)$ trends, they do so in a very specific way, leading to worries about the potential fragility of these methods to alternative specifications of the stochastic trend.

As demonstrated by Wright (2000), it is nevertheless possible to conduct inference about a cointegrating vector without knowledge about the precise nature of the common stochastic trends. Wright's idea is to use the $I(0)$ property of the error correction term as the identifying property of the true cointegrating vector, so that a stationarity test of the model's putative error correction term is used to conduct inference about the value of the cointegrating vectors. Because the common stochastic trends drop out under the null hypothesis, Wright's procedure is robust in the sense that it controls size under any model for the common stochastic trend. But the procedure ignores the data beyond the putative error correction term, and is thus potentially quite inefficient.

Section 2 of this paper provides a formulation of the cointegrated model in which the common stochastic trends follow a flexible limiting Gaussian process that includes the $I(1)$, local-to-unity, and fractional/long-memory models as special cases. Section 3 discusses the low-frequency transformation of the cointegrated model. Throughout the paper, inference procedures are studied in the context of this general formulation of the cointegrated model. The price to pay for this generality is that it introduces a potentially large number of nuisance parameters that characterize the properties of the stochastic trends and the relationship between the stochastic trends and the model's $I(0)$ components, and cannot be estimated consistently in our framework. The main challenge of this paper is to study efficient tests

[4] Granger's Frank Paish Lecture (1993) discusses a wide range of possible data generating processes beyond the $I(1)$ model and argues, sensibly in our opinion, that it is fruitless to attempt to identify the exact nature of the persistence using the limited information in typical macro time series.

in the presence of nuisance parameters under the null hypothesis, and Sections 4–6 address this issue.

By definition, a valid test must control size for any possible value of the nuisance parameter. In our application, the nuisance parameter is often highly dimensional, to the order of 200×1 even in a bivariate system. This makes it extremely hard to directly construct good, let alone efficient tests. Our strategy is thus rather to indirectly learn about the quality of potential tests by deriving bounds on their performance. In particular, Section 4 presents two general results for hypothesis tests in the presence of nuisance parameters under the null. The first is an upper bound for the power of *any* test that controls size.[5] The second result provides a lower bound on size under a more general, "auxiliary" null hypothesis of *any* test that satisfies a lower bound on power and controls size under the original null hypothesis. These bounds provide limits on the performance characteristics of tests, and they can be computed without ever determining a test that is known to control size. Section 5 implements these bounds for tests concerning the value of cointegrating vectors in our low-frequency framework, and discusses numerical techniques to obtain low upper power bounds (approximate "least upper power bounds") and high lower bounds on size under the auxiliary null hypothesis. The power bounds provide a benchmark for the efficiency of any valid test, and differences in the power bounds (interpreted as differences in least upper bounds) associated with restrictions on the trend process (for example, restricting the general stochastic trend process to be $I(1)$) quantify the restriction's information about the value of the cointegrating vector. We find that restrictions can be very informative in the sense of allowing for more powerful tests, but whenever this is the case, any test that were to successfully exploit this information would suffer from large size distortions under a less restrictive trend process. Our analysis using bounds thus quantifies the intuitive notion that extracting information from the assumption of a particular trend process (*e.g.,* $I(1)$ or local-to-unity) makes inference fragile relative to this assumption.

Section 6 builds on Wright's (2000) suggestion and derives a low-frequency test for the value of the cointegration vectors based on an $I(0)$ test for the putative error correction term. Specifically, we derive a low-frequency version of a multivariate point-optimal scale and rotation invariant test against the alternative in which the common trends are $I(1)$. Similar to Wright's (2000) original suggestion, while simple, this low-frequency test for inference about cointegrating vectors is potentially quite inefficient, as it ignores the data beyond the

[5]The same insight about upper bounds on power was noted independently by Andrews, Moreira, and Stock (2008) and used for inference in IV models with potentially weak instruments.

putative error correction term. But the null rejection probability of this test is unaffected by the properties of the common stochastic trend, so its power constitutes an easily achievable lower bound on the power of efficient tests. As it turns out, when attention is focused on a single cointegrating vector, and regardless of the number of common trends, the power of this test essentially coincides with the upper bound for an unrestricted version of the common trend process under the null hypothesis, and is close to the bound for several restricted, but still flexible common trend processes. Thus in this case, the low-frequency version of Wright's test—that is, ignoring the data beyond the putative error correction term—yields an essentially efficient test in the absence of strong *a priori* knowledge about the nature of the persistence.

The implication for applied work is that, at least in the model with a single cointegrating vector, approximately efficient and robust inference may be carried out using the simple test described in Section 6.2. The test is robust in two ways. First, it is robust to arbitrary autocorrelation properties in the error correction term above the pre-specified low-frequency band. Second, it is robust to the precise nature of persistence, as its rejection probability under the null hypothesis does not depend on the nature of the stochastic trend. As in Wright (2000), confidence sets for the cointegrating vector can easily be obtained be inverting the test. We present a brief empirical illustration in Section 7.

2 Model

Let p_t, $t = 1, ..., T$ denote the $n \times 1$ vector of variables under study. This section outlines a time domain representation of the cointegrated model for p_t in terms of canonical variables representing a set of common trends and $I(0)$ error correction terms. The common trends are allowed to follow a flexible process that includes $I(1)$, local-to-unity, and fractional models as special cases, but aside from this generalization, the cointegrated model for p_t is standard.

To begin, p_t is transformed into two components, where one component is $I(0)$ under the null hypothesis and the other component contains elements that are not cointegrated. Let β denote an $n \times r$ matrix whose linearly independent columns are the cointegrating vectors, let β_0 denote the value of β under the null, and $y_t = \beta_0' p_t$. The elements in y_t are the model's error correction terms under the null hypothesis. Let $x_t = \delta' p_t$ where δ is $n \times k$ with $k = n - r$, and where the linearly independent columns of δ are linearly independent of the columns of β_0, so that the elements of x_t are not cointegrated under the null. Because the cointegrated model only determines the column space of the matrix of cointegrating vectors, the variables

y_t and x_t are determined up to transformations $(y_t, x_t) \to (A_{yy}y_t, A_{xx}x_t + A_{xy}y_t)$, where A_{yy} and A_{xx} are non-singular. Most extant inference procedures are invariant (or asymptotically invariant) to these transformations, and, as discussed in detail below, our analysis will also focus on invariant tests.

2.1 Canonical Variable Representation of y_t and x_t

We will represent y_t and x_t in terms of a common stochastic trend vector v_t and an $I(0)$ vector z_t

$$\begin{aligned} y_t &= \Gamma_{yz} z_t + \Gamma_{yv} v_t \\ x_t &= \Gamma_{xz} z_t + \Gamma_{xv} v_t, \end{aligned} \tag{1}$$

where z_t is $r \times 1$, v_t is $k \times 1$, and Γ_{yz} and Γ_{xv} have full rank. In this representation, the restriction that y_t is $I(0)$ corresponds to the restriction $\Gamma_{yv} = 0$. All of the test statistics discussed in this paper are invariant to adding constants to the observations, so that constant terms are suppressed in (1). As a technical matter, we think of $\{z_t, v_t\}_{t=1}^T$ (and thus also $\{x_t, y_t\}_{t=1}^T$) as being generated from a triangular array; we omit the additional dependence on T to ease notation. Also, we write $\lfloor x \rfloor$ for the integer part of $x \in \mathbb{R}$, $||A|| = \sqrt{\text{tr}\, A'A}$ for any real matrix A, $x \vee y$ for the maximum of $x, y \in \mathbb{R}$, '$\otimes$' for the usual Kronecker product and '$\Rightarrow$' to indicate weak convergence.

Let $W(\cdot)$ denote a $n \times 1$ standard Wiener process. The vector z_t is a canonical $I(0)$ vector in the sense that its partial sums converge to a $r \times 1$ Wiener process

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} z_t \Rightarrow S_z W(s) = W_z(s), \text{ where } S_z S_z' = I_r. \tag{2}$$

The vector v_t is a common trend in the sense that scaled versions of its level converge to a stochastic integral with respect to $W(\cdot)$. For example, in the standard $I(1)$ model, $T^{-1/2} v_{\lfloor sT \rfloor} \Rightarrow \int_0^s H dW(t)$, where H is a $k \times n$ matrix and (H', S_z') has full rank. More general trend processes, such as the local-to-unity formulation, allow the matrix H to depend on s and t. The general representation for the common trends used in this paper is

$$T^{-1/2} v_{\lfloor sT \rfloor} \Rightarrow \int_{-\infty}^{s} H(s,t) dW(t) \tag{3}$$

where $H(s,t)$ is sufficiently well behaved to ensure that there exists a cadlag version of the process $\int_{-\infty}^{s} H(s,t) dW(t)$.[6]

[6]The common scale $T^{-1/2}$ for the $k \times 1$ vector v_t in (3) is assumed for convenience; with an appropriate

2.2 Invariance and Reparameterization

As discussed above, because cointegration only identifies the column space of β, attention is restricted to tests that are invariant to the group of transformations

$$(y_t, x_t) \to (A_{yy} y_t, A_{xx} x_t + A_{xy} y_t) \tag{4}$$

where A_{yy} and A_{xx} are non-singular, but (A_{yy}, A_{xx}, A_{xy}) are otherwise unrestricted real matrices.

The restriction to invariant tests allows a simplification of notation: because the test statistics are invariant to the transformations in (4), there is no loss of generality setting $\Gamma_{yz} = I_r$, $\Gamma_{xv} = I_k$, and $\Gamma_{xz} = 0$. With these values, the model is

$$\begin{aligned} y_t &= z_t + \Gamma_{yv} v_t \\ x_t &= v_t. \end{aligned} \tag{5}$$

2.3 Restricted Versions of the Trend Model

We will refer to the general trend specification in (3) as the "unrestricted" stochastic trend model throughout the remainder of the paper. The existing literature on efficient tests relies on restricted forms of the trend process (3) such as $I(1)$ or local-to-unity processes, and we compute the potential power gains associated with these and other *a priori* restrictions on $H(s,t)$ below. Here we describe five restricted versions of the stochastic trend.

The first model, which we will refer to as the G-model, restricts $H(s,t)$ to satisfy

$$H(s,t) = G(s,t) S_v, \tag{6}$$

where $G(s,t)$ is $k \times k$ and S_v is $k \times n$ with $S_v S_v' = I_k$ and (S_z', S_v') nonsingular. In this model, the common trend depends on $W(\cdot)$ only through the $k \times 1$ standard Wiener process $W_v(\cdot) = S_v W(\cdot)$, and this restricts the way that v_t and z_t interact. In this model

$$T^{-1/2} v_{\lfloor sT \rfloor} \Rightarrow \int_{-\infty}^{s} G(s,t) dW_v(t), \tag{7}$$

and the covariance between the Wiener process characterizing the partial sums of z_t, W_z, and W_v is equal to the $r \times k$ matrix $R = S_z S_v'$. Standard $I(1)$ and local-to-unity formulations of cointegration satisfy this restriction and impose additional parametric restrictions on $G(s,t)$.

definition of local alternatives, the invariance (4) ensures that one would obtain the same results for any scaling of v_t. For example, for an $I(2)$ stochastic trend scaled by $T^{-3/2}$, set $H(s,t) = \mathbf{1}[t \geq 0](s-t)H$, with the $k \times n$ matrix H as in the $I(1)$ case.

The second model further restricts (7) so that $G(s,t)$ is diagonal:

$$G(s,t) = \text{diag}(g_1(s,t), \cdots, g_k(s,t)). \tag{8}$$

An interpretation of this model is that the k common trends evolve independently of one another (recall that W_v has identity covariance matrix), where each trend is allowed to follow a different process characterized by the functions $g_i(s,t)$.

The third model further restricts the diagonal-G model so that the k stochastic trends converge weakly to a stationary continuous time process. We thus impose

$$g_i(s,t) = g_i^S(s-t),\ i = 1, \cdots, k. \tag{9}$$

The stationary local-to-unity model (with an initial condition drawn from the unconditional distribution), for instance, satisfies this restriction.

Finally, we consider two parametric restrictions of G:

$$G(s,t) = \mathbf{1}[t > 0] I_k \tag{10}$$

which is the $I(1)$ model, and

$$G(s,t) = \mathbf{1}[t > 0] e^{C(s-t)} \tag{11}$$

which is the multivariate local-to-unity model, where C is the $k \times k$ diffusion matrix of the limiting Ornstein-Uhlenbeck process (with zero initial condition).[7]

2.4 Testing Problem and Local Alternatives

The goal of the paper is to derive asymptotically efficient tests for the value of the cointegrating vectors with controlled rejection probability under the null hypothesis for a range of stochastic trend specifications. The different orders of magnitude of z_t and v_t in (2) and (3) suggest a local embedding of this null hypothesis against alternatives where $\Gamma_{yv} = T^{-1}B$ for B a constant $r \times k$ matrix, so that in model (5),

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} y_t \Rightarrow S_z W(s) + B \int_0^s \int_{-\infty}^u H(u,t) dW(t) du.$$

[7] The $I(1)$ specification in (10) is the same as the $I(1)$ specification given below (2) because the invariance in (4) implies that the trend models are unaffected by premultiplication of $H(s,t)$ (or $G(s,t)$) by an arbitrary non-singular $k \times k$ matrix.

In this parametrization, the null hypothesis becomes

$$\mathrm{H}_0 : B = 0,\ H(s,t) \in \mathcal{H}_0 \tag{12}$$

where $H(s,t)$ is restricted to a set of functions $\mathcal{H}_0$, that, in the unrestricted trend model includes functions sufficiently well behaved to ensure that there exists a cadlag version of the process $\int_{-\infty}^{s} H(s,t)dW(t)$, or more restricted versions of $H(s,t)$ as in (6), (8), (9), (10), or (11).

Since our goal is to consider efficient tests of the null hypothesis (12), we also need to specify the alternative hypothesis. Our results below are general enough to allow for the derivation of efficient tests against any particular alternative with specified $B = B_1$ and stochastic trend process $H(s,t) = H_1(s,t)$,

$$\mathrm{H}_1 : B = B_1,\ H(s,t) = H_1(s,t) \tag{13}$$

or, more generally, for tests that are efficient in the sense of maximizing weighted average power against a set of values for B_1 and stochastic trend models $H_1(s,t)$.

Our numerical results, however, focus on alternatives in which the stochastic trend v_t is $I(1)$, so that $H_1(s,t)$ satisfies (6) and (10). This is partly out of practical considerations: while there is a wide range of potentially interesting trend specification, the computations for any particular specification are involved, and these computational complications limit the number of alternatives we can usefully consider.[8] At the same time, one might consider the classical $I(1)$ model as an important benchmark against which it is useful to maximize power—not necessarily because this is the only plausible model under the alternative, but because a test that performs well against this alternative presumably has reasonable power properties for a range of empirically relevant models. We stress that despite this focus on the $I(1)$ stochastic trend model for the alternative hypothesis (13), we restrict attention to tests that control size for a range of models under the null hypothesis (12). The idea is to control the frequency of rejections under the null hypothesis for any stochastic trend model in $\mathcal{H}_0$, so that the rejection of a set of cointegrating vectors cannot simply be explained by the stochastic trends not being exactly $I(1)$. In this sense, our approach is one of "robust" cointegration testing, with the degree of robustness governed by the size of the set $\mathcal{H}_0$.

[8] If the non-cointegrated components are modelled as $I(1)$ with a deterministic linear time trend, one could choose $H_1(s,t)$ as the sum of (10) and $st\Sigma_\tau$ to obtain tests that maximize weighted average power for v_t that is $I(1)$ with a linear trend of slope β_τ, with a weighting function $\beta_\tau \sim \mathcal{N}(0, \Sigma_\tau/T^2)$. We do not pursue this further, though.

2.5 Summary

To summarize, this section has introduced the time domain representation of the cointegrated model with a focus on the problem of inference about the space of cointegrating vectors. In all respects except one, the representation is the standard one: the data are expressed as a linear function of a canonical vector or common trends and a vector of $I(0)$ components. Under the null, certain linear combinations of the data do not involve the common trends. Because the null only restricts the column space of the matrix of cointegrating vectors, attention is restricted to invariant tests. The goal is to construct tests with best power for an alternative value for the matrix of cointegrating vectors under a particular model for the trend (or best weighted average power for a collection of B_1 and $H_1(s,t)$). The formulation differs from the standard one only in that it allows the model for trend under the null to be less restrictive than the standard formulation. Said differently, because of potential uncertainty about the specific form of the trend process, the formulation restricts attention to tests that control size for a range of different trend processes. This generalization complicates the problem of constructing efficient tests by introducing a potentially large number of nuisance parameters (associated with the trend process) under the null hypothesis.

3 Low-Frequency Representation of the Model

Cointegration is a restriction on the low-frequency behavior of time series, and as discussed in the introduction, we therefore focus on the low-frequency behavior of (y_t, x_t). This low-frequency variability is summarized by a small number, q, of weighted averages of the data. In this section we discuss these weighted averages and derive their limiting behavior under the null and alternative hypotheses.

3.1 Low-Frequency Weighted Averages

We use weights associated with the cosine transform, where the j'th weight is given by $\Psi_j(s) = \sqrt{2}\cos(j\pi s)$. For any sequence $\{a_t\}_{t=1}^T$, the j'th weighted average will be denoted by

$$A_T(j) = \int_0^1 \Psi_j(s) a_{\lfloor sT \rfloor + 1} ds = \iota_{jT} T^{-1} \sum_{t=1}^{T} \Psi_j(\tfrac{t-1/2}{T}) a_t \tag{14}$$

where $\iota_{jT} = (2T/j\pi)\sin(j\pi/2T) \to 1$ for all fixed j. As demonstrated by Müller and Watson (2008), the weighted averages $A_T(j)$, $j = 0, \cdots, q$, essentially capture the variability in the

sequence corresponding to frequencies below $q\pi/T$.

We use the following notation: with a_t a $h\times 1$ vector time series, let $\Psi(s) = (\Psi_1(s), \Psi_2(s), \cdots, \Psi_q(s))'$ denote the $q \times 1$ vector of weighting functions, and $A_T = \int_0^1 \Psi(s) a'_{\lfloor sT \rfloor + 1} ds$ the $q \times h$ matrix of weighted averages of the elements of a_t, where $\Psi_0(s)$ is excluded to make the results invariant to adding constants to the data. Using this notation, the $q \times r$ matrix Y_T and the $q \times k$ matrix X_T summarize the variability in the data corresponding to frequencies lower than $q\pi/T$. With $q = 12$, (Y_T, X_T) capture variability lower than the business cycle (periodicities greater than 8 years) for time series that span 50 years (postwar data) regardless of the sampling frequency (months, quarters, weeks, etc.). This motivates us to consider the behavior of these matrices as $T \to \infty$, but with q held fixed.

The large-sample behavior of X_T and Y_T follows from the behavior of Z_T and V_T. Using the assumed limits (2) and (3), the continuous mapping theorem, and integration by parts for the terms involves Z_T, one obtains

$$\begin{bmatrix} T^{1/2} Z_T \\ T^{-1/2} V_T \end{bmatrix} \Rightarrow \begin{bmatrix} Z \\ V \end{bmatrix} \tag{15}$$

where

$$\begin{bmatrix} \operatorname{vec} Z \\ \operatorname{vec} V \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} I_{rq} & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix}\right) \tag{16}$$

with

$$\begin{aligned} \Sigma_{VZ} &= \int_0^1 \left(\int_{t \vee 0}^1 [H(s,t) \otimes \Psi(s)] ds \right) [S_z \otimes \Psi(t)]' dt \\ \Sigma_{VV} &= \int_{-\infty}^1 \left(\int_{t \vee 0}^1 [H(s,t) \otimes \Psi(s)] ds \right) \left(\int_{t \vee 0}^1 [H(s,t) \otimes \Psi(s)] ds \right)' dt. \end{aligned} \tag{17}$$

The relative scarcity of low-frequency information is thus formally captured by considering the weak limits (2) and (3) as pertinent only for the subspace spanned by the weight function $\Psi(\cdot)$, yielding (15) as a complete characterization of the relevant properties of the error correction term z_t and the common stochastic trend v_t.

Using $\Gamma_{yv} = T^{-1}B$, equation (5) implies that $Y_T = Z_T + T^{-1} V_T B'$ and $X_T = V_T$. Thus,

$$\begin{bmatrix} T^{1/2} Y_T \\ T^{-1/2} X_T \end{bmatrix} \Rightarrow \begin{bmatrix} Y \\ X \end{bmatrix} = \begin{bmatrix} Z + VB' \\ V \end{bmatrix} \tag{18}$$

where

$$\begin{bmatrix} \operatorname{vec} Y \\ \operatorname{vec} X \end{bmatrix} \sim \mathcal{N}\left(0, \Sigma_{(Y,X)}\right) \tag{19}$$

with

$$\Sigma_{(Y,X)} = \begin{bmatrix} I_r \otimes I_q & B \otimes I_q \\ 0 & I_k \otimes I_q \end{bmatrix} \begin{bmatrix} I_r \otimes I_q & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix} \begin{bmatrix} I_r \otimes I_q & 0 \\ B' \otimes I_q & I_k \otimes I_q \end{bmatrix}. \tag{20}$$

3.2 "Best" Low-Frequency Hypothesis Tests

We consider invariant tests of H_0 against H_1 given in (12) and (13) based on the data $\{y_t, x_t\}_{t=1}^T$. Because we are concerned with the model's implications for the low-frequency variability of the data, we restrict attention to tests that control asymptotic size for all models that satisfy (18)-(20). Our goal is to find an invariant test that maximizes power subject to this restriction, and for brevity we will refer to such a test as a "best" test. Müller (2008) considers the general problem of constructing asymptotically most powerful tests subject to asymptotic size control over a class of models such as ours. In our context, his results imply that asymptotically best tests correspond to the most powerful invariant tests associated with the limiting distribution (19).

Thus, the relevant testing problem has a simple form: $\mathrm{vec}(Y, X)$ has a normal distribution with mean zero and covariance matrix that depends on B. Under the null $B = 0$, while under the alternative $B \neq 0$. Tests are restricted to be invariant to the group of transformations

$$(Y, X) \to (YA'_{yy}, XA'_{xx} + YA'_{xy}) \tag{21}$$

where A_{yy} and A_{xx} are nonsingular, and A_{yy}, A_{xx}, and A_{xy} are otherwise unrestricted. Thus, the hypothesis testing problem becomes the problem of using an invariant procedure to test a restriction on the covariance matrix of a multivariate normal vector.

4 Bounds on Power and Size

The general version of the hypothesis testing problem we are facing is a familiar one: Let U denote a single observation of dimension $m \times 1$. (In our problem, U corresponds to the maximal invariant for (Y, X)). Under the null hypothesis U has probability density $f_\theta(u)$ with respect to some measure μ, where $\theta \in \Theta$ is a vector of nuisance parameters. (In our problem, the vector θ describes the stochastic trend process under the null hypothesis and determines $\Sigma_{(Y,X)}$ via (17) and (20)). Under the alternative, U has known density $h(u)$. (Choices for $h(u)$ for our problem will be discussed in Subsection 5.2.1.) Thus, the null and

alternative hypothesis are

$$\begin{aligned} &\mathrm{H}_0 : \text{The density of } U \text{ is } f_\theta(u),\ \theta \in \Theta \\ &\mathrm{H}_1 : \text{The density of } U \text{ is } h(u), \end{aligned} \tag{22}$$

and possibly randomized tests are (measurable) functions $\varphi : \mathbb{R}^m \mapsto [0,1]$, where $\varphi(u)$ is the probability of rejecting the null hypothesis when observing $U = u$, so that size and power are given by $\sup_{\theta\in\Theta} \int \varphi f_\theta d\mu$ and $\int \varphi h d\mu$, respectively.

This section presents two results on power and size in this general problem. The first result is an upper bound on the power of any valid test of H_0 versus H_1. This bound will be useful for our specific problem because, as we show below, it can be computed numerically and we will construct a feasible test that (essentially) achieves the bound when there is a single cointegrating vector. This shows that the feasible test that we construct is efficient when $r = 1$. The second result provides a lower bound on the size under a null hypotheses other than H_0 for any test that is powerful against H_1. To see why this is useful in our context suppose that H_0 specifies that the stochastic trend follows an $I(1)$ process, and consider a test that exploits features of the $I(1)$ process to increase power. Uncertainty about the trend process means that it is useful know something about the rejection frequency of tests under null hypotheses that allow for more general trends, such as the unrestricted trend model (3) or other less restricted versions described above. The second result provides a lower bound on this rejection frequency, where a large value of this lower bound highlights the fragility of tests that exploit a particular H_0 to obtain more powerful inference.

4.1 An Upper Bound on Power

A standard device for problems such as (22) is to consider a Neyman-Pearson test for a related problem in which the null hypothesis is replaced with a mixture

$$\mathrm{H}_\Lambda : \text{The density of } U \text{ is } \int f_\theta d\Lambda(\theta)$$

where Λ is a probability distribution for θ with support in Θ. The following lemma shows that the power of the Neyman-Pearson test of H_Λ versus H_1 provides an upper power bound for tests of H_0 versus H_1.

Lemma 1 *Let φ_Λ be the best level α test of H_Λ against H_1. Then for any level α test φ of H_0 against H_1, $\int \varphi_\Lambda h d\mu \geq \int \varphi h d\mu$.*

Proof. *Since φ is a level α test of* H_0, $\int \varphi f_\theta d\mu \leq \alpha$ *for all* $\theta \in \Theta$. *Therefore,* $\int\int \varphi f_\theta d\mu d\Lambda(\theta) = \int\int \varphi f_\theta d\Lambda(\theta) d\mu \leq \alpha$ *(where the change in the order of integration is allowed by Fubini's Theorem), so that φ is also a level α test of* H_Λ *against* H_1. *The result follows by the definition of a best test.* ■

This result is closely related to Theorem 3.8.1 of Lehmann and Romano (2005) which provides conditions under which a least upper bound on the power for tests H_0 versus H_1 is associated with a "least favorable distribution" for θ, and that using this distribution for Λ produces the least upper power bound. The least favorable distribution Λ^* has the characteristic that the resulting φ_{Λ^*} is a level α test for testing H_0 versus H_1. Said differently, if φ_{Λ^*} is the best level α test of H_{Λ^*} against H_1 *and* is also a level α test for testing H_0 versus H_1, then $\varphi^* = \varphi_{\Lambda^*}$, that is φ_{Λ^*} is the most powerful level α test of H_0 versus H_1. Unfortunately, while the test associated with the least favorable distribution solves the testing problem (22), there is no general and constructive method for finding the least favorable distribution Λ^* (and it does not always exist).

With this in mind, Lemma 1 is stated so that Λ is not necessarily the least favorable distribution. That is, the bound in Lemma 1 holds for any probability distribution Λ. The goal of the numerical analysis carried out below is to choose Λ to approximate the least upper bound. Importantly, even if one cannot identify the least favorable distribution, Lemma 1 shows that the power of φ_Λ provides a valid bound for the power of any test of H_0 versus H_1, for any Λ.

4.2 A Lower Bound on Size under an Auxiliary Null

Now consider the "Larger" auxiliary null hypothesis

$$\mathrm{H}_L : \text{The density of } U \text{ is } f_\theta(u),\ \theta \in \Theta_L$$

with an associated mixture

$$\mathrm{H}_{\Lambda_L} : \text{The density of } U \text{ is } \int f_\theta d\Lambda_L(\theta)$$

where Λ_L has support in $\Theta_{L:}$. (In our problem H_L will be a null hypothesis that allows for a less restricted trend process than under H_0. Thus if H_0 allows only for an $I(1)$ trend, H_L might allow for a local-to-unity trend or one of the more general trends processes discussed in Section 2.)

Consider any test φ of level α_0 under H_0 with power of at least β. The following lemma provides a lower bound on the rejection frequency under the auxiliary null H_L.

Lemma 2 *Suppose the test* $\varphi^* = \mathbf{1}[\int \varphi h d\mu \geq \lambda_1 \int \varphi \int f_\theta d\Lambda_0(\theta) d\mu + \lambda_2 \int \varphi \int f_\theta d\Lambda_L(\theta) d\mu]$ *with* $\lambda_1, \lambda_2 \geq 0$ *has rejection probability* $\alpha_0 = \int \int \varphi^* f_\theta d\Lambda_0(\theta) d\mu$ *under* H_{Λ_0}, $\alpha_L = \int \int \varphi^* f_\theta d\Lambda_L(\theta) d\mu$ *under* H_{Λ_L} *and power* $\beta = \int \varphi^* h d\mu$. *Then for any test* φ *of level* α_0 *under* H_0 *and power of at least* β, $\sup_{\theta \in \Theta_L} \int \varphi f_\theta d\mu \geq \alpha_L$.

Proof. By a variant of the generalized Neyman-Pearson Lemma (Theorem 3.6.1 in Lehmann and Romano (2005)), the test φ^* solves the program

$$\begin{aligned} & \min_{\varphi} \int \varphi \int f_\theta d\Lambda_L(\theta) d\mu \\ \text{s.t.} \int \varphi \int f_\theta d\Lambda_0(\theta) d\mu \;&\leq\; \alpha_0 \;\text{ and }\; \int \varphi h d\mu \geq \beta. \end{aligned}$$

Since φ is a level α_0 test of H_0, $\int \varphi f_\theta d\mu \leq \alpha$ for all $\theta \in \Theta$, and $\int \int \varphi f_\theta d\mu d\Lambda_0(\theta) = \int \int \varphi f_\theta d\Lambda_0(\theta) d\mu \leq \alpha$ (where the change in the order of integration is allowed by Fubini's Theorem). Also, by assumption, the power of the test φ is at least β, $\int \varphi h d\mu \geq \beta$, so that φ satisfies the two constraints in the program above. Thus, $\sup_{\theta \in \Theta_L} \int \varphi f_\theta d\mu \geq \int \varphi \int f_\theta d\Lambda_L(\theta) d\mu \geq \int \varphi^* \int f_\theta d\Lambda_L(\theta) d\mu = \alpha_L$. ■

This lemma is particularly useful in conjunction with Lemma 1: Suppose application of Lemma 1 yields that no 5% level test of a relatively restricted H_0 can have power of more than, say, 70%. This suggests that there could indeed exist a 5% level test φ with power, say, 67%, and one might want to learn about the size properties of such tests under the more general null hypothesis H_L. Lemma 2 provides a way of computing a lower bound on this size that is valid for *any* test with power of at least 67%. So if this size distortion is large, then without having to determine the class of 5% level tests of H_0 with power of at least 67%, one can already conclude that all such tests will be fragile. In the numerical section below, we discuss how to determine a suitable Λ_0 and Λ_L to obtain a large lower bound α_L (λ_1 and λ_2 are determined through the two constraints on φ^*).

5 Computing Bounds

In this section we compute the power and size bounds from the last section. The analysis proceeds in four steps. First, we derive the density of the maximal invariant of (Y, X); this density forms the basis of the likelihood ratio. Second, since the density of the maximal invariant depends on the covariance matrix of (Y, X), we discuss the parameterization of $\Sigma_{(Y,X)}$ under the null and alternative hypotheses. In the third step we describe how the

mixing distributions Λ, Λ_0 and Λ_L are chosen to yield tight bounds. Finally, we present numerical values for the bounds.

5.1 Density of a Maximal Invariant

Recall that we are considering tests that are invariant to the group of transformations $(Y, X) \to (YA'_{yy}, XA'_{xx} + YA'_{xy})$ where A_{yy} and A_{xx} are nonsingular, and A_{yy}, A_{xx}, and A_{xy} are otherwise unrestricted. Any invariant test can be written as a function of a maximal invariant (Theorem 6.2.1 in Lehmann and Romano (2005)), so that by the Neyman-Pearson lemma, the most powerful invariant test rejects for large values of the likelihood ratio statistic of a maximal invariant. The remaining challenge is the computation of the density of a maximal invariant, and this is addressed in the following theorem.

Theorem 1 *If* $\operatorname{vec}(Y, X) \sim \mathcal{N}(0, \Sigma_{(Y,X)})$ *with positive definite* $\Sigma_{(Y,X)}$ *and* $q > r + k$*, the density of a maximal invariant of (21) has the form*

$$c(\det \Sigma_{(Y,X)})^{-1/2}(\det V'_{0Y}\Sigma^{-1}_{(Y,X)}V_{0Y})^{-1/2}(\det \Omega)^{-1/2}E_\omega[|\det(\omega_Y)|^{q-r}|\det(\omega_X)|^{q-r-k}]$$

where c *does not depend on* $\Sigma_{(Y,X)}$*,* ω_Y *and* ω_X *are random* $r \times r$ *and* $k \times k$ *matrices, respectively, with* $(\operatorname{vec}\omega'_y, \operatorname{vec}\omega'_x)' \sim \mathcal{N}(0, \Omega^{-1})$*,*

$$\Omega = D'_{YX}\Sigma^{-1}_{(Y,X)}D_{YX} - D'_{YX}\Sigma^{-1}_{(Y,X)}V_{0Y}(V'_{0Y}\Sigma^{-1}_{(Y,X)}V_{0Y})^{-1}V'_{0Y}\Sigma^{-1}_{(Y,X)}D_{YX},$$

$D_{YX} = \operatorname{diag}(I_r \otimes Y, I_k \otimes X)$*,* $V_{0Y} = (0'_{rq\times rk}, I_k \otimes Y')'$*, and* E_ω *denotes integration with respect to* ω_Y *and* ω_X*, conditional on* (Y, X)*.*

Theorem 1 shows that the density of a maximal invariant can be expressed in terms of absolute moments of determinants of jointly normally distributed random matrices, whose covariance matrix depends on (Y, X). We do not know of a useful and general closed-form solution for this expectation; for $r = k = 1$, however, Nabeya's (1951) results for the absolute moments of a bivariate normal yields an expression in terms of elementary functions, which we omit for brevity. When $r + k > 2$, the moments can be computed via Monte Carlo integration. However, computing accurate approximations is difficult when r and k are large, and the numerical analysis reported below is therefore limited to small values of r and k.

5.2 Parameterization of $\Sigma_{(Y,X)}$

Since the density of the maximal invariant of Theorem 1 depends on $\Sigma_{(Y,X)}$, the derivation of efficient invariant tests requires specification of $\Sigma_{(Y,X)}$ under the alternative and null hypothesis. We discuss each of these in turn.

5.2.1 Specification of $\Sigma_{(Y,X)}$ under the Alternative Hypothesis

As discussed above, we focus on the alternative where the stochastic trends follow an $I(1)$ process, so that $H(s,t)$ satisfies (6) and (10). There remains the issue of the value of B (the coefficients that show how the trends affect Y) and R (the correlation of the Wiener processes describing the $I(0)$ variables, z_t, and the common trends, v_t). For these parameters, we consider point-valued alternatives with $B = B_1$ and $R = R_1$; the power bounds derived below then serve as bounds on the asymptotic power envelope over these values of B and R. Invariance reduces the effective dimension of B and R somewhat, and this will be discussed in the context of the numerical results presented below.

5.2.2 Parameterization of $\Sigma_{(Y,X)}$ under the Null Hypothesis

From (20), under the null hypothesis with $B = 0$, the covariance matrix $\Sigma_{(Y,X)}$ satisfies

$$\Sigma_{(Y,X)} = \begin{bmatrix} I_{rq} & \Sigma_{ZV} \\ \Sigma_{VZ} & \Sigma_{VV} \end{bmatrix}.$$

The null model's specification of the stochastic trend determines the $rq \times kq$ matrix Σ_{ZV} and the $kq \times kq$ matrix Σ_{VV} by the formulae given in (17). Since these matrices contain a finite number of elements, it is clear that even for nonparametric specifications of $H(s,t)$, the effective parameter space for low-frequency tests based on (Y,X) is finite dimensional. We collect these nuisance parameters in a vector $\theta \in \Theta$.

Section 2 discussed several trend processes, beginning with the general process given in (3) with an unrestricted version of $H(s,t)$, and then five restricted models: (i) the "G" model in (6), (ii) the "Diagonal" model (8), (iii) the "Stationary" model (9), (iv) the local-to-unity model (11), and (v) the $I(1)$ model (10). The appendix discusses convenient parameterization for $\Sigma_{(Y,X)}$ for these five restricted models, and the following lemma provides the basis for parameterizing $\Sigma_{(Y,X)}$ when $H(s,t)$ is unrestricted.

Lemma 3 *(a) For any $(r+k)q \times (r+k)q$ positive definite matrix Σ^* with upper left $rq \times rq$ block equal to I_{rq}, there exists an unrestricted trend model with $H(s,t) = 0$ for $t < 0$ such that $\Sigma^* = E[\text{vec}(Z,V)(\text{vec}(Z,V))']$.*

(b) If $r \leq k$, this $H(s,t)$ can be chosen of the form $H(s,t) = G(s,t)S_v$, where (S_z', S_v') has full rank.

The lemma shows that when $H(s,t)$ is unrestricted or $r \leq k$ and $H(s,t) = G(s,t)S_v$, the only restriction that the null hypothesis imposes on $\Sigma_{(Y,X)}$ is that $\Sigma_{YY} = I_{rq}$.[9] In other words, since Σ_{ZV} and Σ_{VV} have $rkq^2 + kq(kq+1)/2$ distinct elements, an appropriately chosen θ of that dimension determines $\Sigma_{(Y,X)}$ under the null hypothesis in the unrestricted model, and in the model where $H(s,t) = G(s,t)S_v$ for $r \leq k$.

5.3 Approximating the Least Upper Power and Greatest Lower Size Bound

We develop two methods to approximate the power bound associated with the least favorable distribution from Section 4, and use the second method also to determine a large lower size bound. First, we develop an algorithm that simultaneously determines a low upper bound on power, and a level α test whose power is close to that bound. This algorithm is entirely generic in the sense that it does not exploit any specificities of the low-frequency robust cointegration testing problem; in practice, it only requires that the densities f_θ and h can be quickly evaluated numerically. The computational complexity is such, however, that it can only be applied when θ is low-dimensional; as such, it is useful for our problem only in the $I(1)$ and local-to-unity stochastic trend model for $r = k = 1$. Second, when the dimension of θ is large we choose Λ (and Λ_0 and Λ_L for Lemma 2) so the null and alternative distributions are close in some numerically convenient metric. Two numerical results suggest that this second method produces a reasonably accurate estimate of the lower bound: the method produces power bounds only marginally higher than the first method (when the first method is feasible), and when $r = 1$ we find that the method produces a power bound that can be achieved by a feasible test that we present in the next section.

We discuss the two methods in turn.

[9] Without the invariance restriction (21), this observation would lead to an analytic least favorable distribution result: Factor the density of (Y,X) under the alternative into the product of the density of Y, and the density of X given Y. By choosing Σ_{VZ} and Σ_{VV} under the null hypothesis appropriately, the latter term cancels, and the Neyman-Pearson test is a function of Y only.

5.3.1 Low Dimensional Nuisance Parameter

Suppose that $\text{LR}_\Lambda = h(U)/\int f_\theta(U)d\Lambda(\theta)$ is a continuous random variable for any Λ, so that by the Neyman-Pearson Lemma, φ_Λ is of the form $\varphi_\Lambda = \mathbf{1}[\text{LR}_\Lambda > \text{cv}_\Lambda]$, where the critical value cv_Λ is chosen to satisfy the size constraint $\int\int \varphi_\Lambda f_\theta d\mu d\Lambda(\theta) = \alpha$. Then by Lemma 1, the power of φ_Λ, $\beta_\Lambda = \int \varphi_\Lambda h d\mu$, is an upper bound on the power of any test that is level α under H_0. If Λ is not the least favorable distribution, then φ_Λ is not of size α under H_0, i.e. $\sup_{\theta\in\Theta}\int \varphi_\Lambda f_\theta d\mu > \alpha$. Now consider a version of φ_Λ with a size-corrected critical value $\text{cv}^c_\Lambda > \text{cv}_\Lambda$, that is $\varphi^c_\Lambda = \mathbf{1}[\text{LR}_\Lambda > \text{cv}^c_\Lambda]$ with cv^c_Λ chosen to satisfy the size constraint $\sup_{\theta\in\Theta}\int \varphi^c_\Lambda f_\theta d\mu = \alpha$. Because the size adjusted test φ^c_Λ is of level α under H_0, the least upper power bound must be between β^c_Λ and β_Λ. Thus, if β^c_Λ is close to β_Λ, then β^c_Λ serves as a good approximation to the least upper bound.

The challenge is to find an appropriate Λ. This is difficult because, in general, no closed form solutions are available for the size and power of tests, so that these must be approximated by Monte Carlo integration. Brute force searches for an appropriate Λ are not computationally feasible. We exploit numerical advantages of discrete distributions for Λ, that have point masses at only N points, and smooth out the Monte Carlo integration estimates of size and power, so that gradient methods can be employed. The suggested algorithm is related to, but distinct from those developed in Nelson (1966), Kempthorne (1987) and Srianantha kumar and King (2006), and is described in detail in the appendix.

5.3.2 High Dimensional Nuisance Parameter

The dimension of θ can be very large in our problem: even when $r = k = 1$, the model with unrestricted stochastic trend leads to θ of dimension $q^2 + q(q+1)/2$ so that θ contains 222 elements when $q = 12$. Approximating the least upper power bound directly then becomes numerically intractable. This motivates a computationally practical method for computing a *low* (as oppose to *least)* upper power bound.

The method restricts Λ so that it is degenerate with all mass on a single point, say θ^*, which is chosen so that the null distribution of the maximal invariant of Theorem 1 is close to its distribution under the alternative. Intuitively, this should make it difficult to distinguish the null from the alternative hypothesis, and thus lead to a low power bound. Also, this choice of θ^* ensures that draws from the null model look empirically reasonable, as they are nontrivial to distinguish from draws of the alternative with an $I(1)$ stochastic trend.

Since the density of the maximal invariant is quite involved, θ^* is usefully approximated

by a choice that makes the multivariate normal distribution of $\text{vec}(Y, X)$ under the null close to its distribution under the alternative, as measured by a convenient metric. We choose θ^* to minimize the Kullback-Leibler divergence (KLIC) between the null and alternative distributions. Since the bounds from Lemmas 1 and 2 are valid for any mixture, numerical errors in the KLIC minimization do not invalidate the resulting bound. Details are provided in the appendix.

5.4 Numerical Bounds

Table 1 shows numerical results for power and size bounds for 5% level tests with $q = 12$. Results are shown for $r = k = 1$ (panel A), $r = 1$ and $k \geq 2$ (panel B), and $r = 2$ and $k = 1$ (Panel C).[10] Numerical results for larger values of $n = r + k$ are not reported because of the large number of calculations required to evaluate the density in large models.

Power depends on the values of B and R under the alternative, and results are presented for various values of these parameters. Because of invariance, when $r = 1$ (as in panels A and B), or $k = 1$ (as in panel C), the distribution of the maximal invariant depends on B and R only through $||B||$, $||R||$, and, if $||R|| > 0$, on $\text{tr}(R'B)/(||B|| \cdot ||R||)$. Thus, in panel A, where $r = k = 1$, results are shown for two values of $|B|$, three values of $|R|$ and for $R \cdot B < 0$ and $R \cdot B > 0$, while panels B and C show results for three values of $\omega = \text{tr}(R'B)/(||B|| \cdot ||R||)$ when $||R|| > 0$. All of the results in Table 1 use the KLIC minimized values of θ as described in the last subsection. Table 2 compares this KLIC-based bounds to to the numerical least upper power bounds when the parameter space is sufficiently small to allow calculation of the numerical least upper bounds.

To understand the formatting of Table 1, look at panel A. The panel contains italicized and non-italicized numerical entries. The non-italicized numbers are power bounds, and the italicized numbers are size bounds. The first column in the table shows the trend specification allowed under H_0. The first entry, labelled "*unr*" corresponds to the unrestricted trend specification in (3) and the other entries correspond to the restricted trend processes discussed in Section 2. Because $r = k = 1$, there are no restrictions imposed by the assumption that $H(s,t) = G(s,t)S_v$ or that G is diagonal, so these models are not listed in panel A. Stationarity ($G(s,t) = G(s-t)$) is a restriction, and this is the second entry in the first column. The final two entries correspond to the local-to-unity ("*LTU*") and $I(1)$

[10] The results shown in panel B were computed using the KLIC minimized value of θ for the model with $r = 1$ and $k = 2$. The appendix shows the resulting bounds are valid for $k \geq 2$.

restrictions. The numerical entries shown in the rows corresponding to these trend models are the power bounds. For example, the non-italicized entries in the first numerical column show power bounds for $|R| = 0$ and $|B| = 7$, which are 0.36 for the unrestricted null, 0.41 when the trend is restricted to be stationary, 0.50 when the trend is restricted to follow a local-to-unity process, and 0.50 when the trend is further restricted to follow an $I(1)$ process.

The second column of panel A shows the auxiliary null hypotheses H_L, corresponding to the null hypothesis H_0, shown in the first column. The entries under H_L represent less restrictive models than H_0. For example, when H_0 restricts the null to be stationary, an unrestricted trend process ("*unr*") is shown for H_L, while when H_0 restricts the trend to be $I(1)$, the less restrictive local-to-unity, stationary, and unrestricted nulls are listed under H_L. The numerical entries for these rows (shown in italics in the table) are the lower size bounds for H_L for 5% level tests under H_0 and with power that is 3 percentage points less than the corresponding power bound shown in the table. For example, from the first numerical column of panel A, the power bound for the $I(1)$ version of H_0 is 0.50. For any test with size no larger than 5% under this null and with power of at least 0.47 ($= 0.50 - 0.03$), the size under a null that allows an unrestricted trend ("*unr*" under H_L) is at least 12%, the size under a null that restricts the trend to be stationary is at least 8%, and the size under a null that restricts the trend to follow a local-to-unity process is at least 4%.

Looking at the entries in Panel A, two results stand out. First, and not surprisingly, restricting tests so that they control size for the unrestricted trend process leads to a non-negligible reduction in power. For example, when $|B| = 7$, and $R = 0$, the power bound is 0.36, for tests that control size for unrestricted trends, the bound increases to 0.41 for tests that control size for stationary trends, and increases to 0.50 for tests that control size for local-to-unity or $I(1)$ trend processes. Second, whenever there is a substantial increase in power associated with restricting the trend process, there are large size distortions under the null hypothesis without this restriction. For example, Elliott's (1998) observation that efficient tests under the $I(1)$ trend have large size distortions under a local-to-unity process is evident in the table. From the table, when $|B| = 7$, $|R| = 0.9$, and $R \cdot B > 0$, the power bound for the null with an $I(1)$ trend is 0.95, but any test that controls size for this null and has power of at least 0.92 will have size that is greater than 0.50 when the trend is allowed to follow a local-to-unity process. However, addressing Elliott's (1998) concern by controlling for size in the local-to-unity model, as in the analysis of Stock and Watson (1996) or Jansson and Moreira (2006) does not eliminate the fragility of the test. For example, with the same values of B and R, the power bound for the null that allows for a local-to-unity trend is

0.67, but any test that controls size for this null and has a size of at least 0.64 will have a size greater than 0.32 when the trend is unrestricted.

Panels B ($r = 1$ and $k = 2$) and C ($r = 2$ and $k = 1$) show qualitatively similar results. Indeed these panels show even more fragility of tests that do not allow for general trends. For example, the lower size bound for the unrestricted trend null exceeds 0.50 in several cases for tests that restrict trends to be $I(1)$, local-to-unity, or stationary.

When $r = k = 1$, it is feasible to approximate the least upper power bound for the $I(1)$ and local-to-unity trend restrictions using the method described in subsection 5.3.1. As described in the appendix, the approximate least upper bounds (LUB) in Table 2 are no more than 2.5 percentage points above the actual least upper bound, apart from Monte Carlo error. The differences with the KLIC minimized power bounds are small, suggesting that the bounds in Table 1 are reasonably tight.

6 Efficient Y-Only Tests

The primary obstacle for constructing efficient tests of the null hypothesis that $B = 0$ is the large number of nuisance parameters associated with the stochastic trend (the parameters that determine $H(s,t)$). These parameters govern the values of Σ_{ZV} and Σ_{VV}, which in turn determine Σ_{YX} and Σ_{XX}. Any valid test must control size over all values of these nuisance parameters. Wright (2000) notes that this obstacle can be avoided by ignoring the x_t data and basing inference only on y_t, since under the null hypothesis, $y_t = z_t$. This section takes up Wright's suggestion and discusses efficient low-frequency "Y-only" tests.[11]

We have two related goals. The first is to study the power properties of these tests relative to the power bounds computed in the last section. As it turns out, when $r = 1$ (so there is only a single cointegrating vector), this Y-only test essentially achieves the power bound, so the test efficiently uses all of the information in Y and X. Given the efficiency property of the Y-only test, the second goal is to develop simple formulae for implementing the test. We discuss these in reverse order, first deriving a convenient formulae for the test

[11] Wright (2000) implements this idea using a "stationarity" test of the $I(0)$ null proposed by Saikkonen and Luukonen (1993), using a robust covariance matrix as in Kwiatkowski, Phillips, Schmidt, and Shin (1992) for the test proposed in Nyblom (1989). This test relies on a consistent estimator of the spectral density matrix of z_t at frequency zero. But consistent estimation requires a lot of pertinent low frequency information, and lack thereof leads to well-known size control problems (see for example, Kwiatkowski, Phillips, Schmidt, and Shin (1992), Caner and Kilian (2001), and Müller (2005)). These problems are avoided by using the low-frequency components of y_t only; see Müller and Watson (2008) for further discussion.

statistic and then studying the power of the resulting test.

6.1 Efficient Tests against General Alternatives

The distribution of $\operatorname{vec} Y \sim \mathcal{N}(0, \Sigma_{YY})$ follows from the derivations in Section 3: Under the null hypothesis, $\Sigma_{YY} = I_{rq}$, and under the alternative, Σ_{YY} depends on the local alternative B, the properties of the stochastic trend and its relationship with the error correction term Z. For a particular choice of alternative, the testing problem thus becomes $\mathrm{H}_0 : \Sigma_{YY} = I$ against $\mathrm{H}_1 : \Sigma_{YY} = \Sigma_{YY1}$, and the invariance requirement (21) becomes

$$Y \to Y A_{yy}' \text{ for arbitrary nonsingular } r \times r \text{ matrices } A_{yy}. \tag{23}$$

The density of the maximal invariant is given in the following theorem.

Theorem 2 *(a) If* $\operatorname{vec} Y \sim \mathcal{N}(0, \Sigma_{YY})$ *with positive definite* Σ_{YY} *and* $q > r$*, the density of a maximal invariant to (23) has the form*

$$c_1 (\det \Sigma_{YY})^{-1/2} (\det \Omega_Y)^{-1/2} E_{\omega_Y}[|\det(\omega_Y)|^{q-r}]$$

where c_1 *does not depend on* Σ_{YY}*,* ω_Y *is an* $r \times r$ *random matrix with* $\operatorname{vec} \omega_Y \sim \mathcal{N}(0, \Omega_Y^{-1})$*,* $\Omega_Y = (I_r \otimes Y)' \Sigma_{YY}^{-1} (I_r \otimes Y)$*, and* E_{ω_Y} *denotes integration with respect to the distribution of* ω_Y *(conditional on* Y*).*

(b) If in addition, $\Sigma_{YY} = \tilde{V}_{YY} \otimes \tilde{\Sigma}_{YY}$*, then the density simplifies to*

$$c_2 (\det \tilde{\Sigma}_{YY})^{-r/2} \det(Y' \tilde{\Sigma}_{YY}^{-1} Y)^{-q/2}$$

where c_2 *does not depend on* Σ_{YY}*.*

As in Theorem 1, part (a) of this theorem provides a formula for the density of a maximal invariant in terms of absolute moments of the determinant of a multivariate normal matrix with a covariance matrix that depends on the data. Part (b) provides an explicit and simple formula when the covariance matrix is of a specific Kronecker form. This form arises under the null hypothesis with $\Sigma_{YY} = I_{rk}$, and under alternatives where each of the r putative error correction terms in y_t have the same low-frequency covariance matrix. For a simple alternative hypothesis with $\Sigma_{YY1} = \tilde{V}_{YY1} \otimes \tilde{\Sigma}_{YY1}$, the best test then rejects for large values of $\det(Y'Y)/\det(Y' \tilde{\Sigma}_{YY1}^{-1} Y)$. The form of weighted average power maximizing tests over a set of alternative covariance matrices Σ_{YY1} are also easily deduced from Theorem 2 parts (a) and (b).

6.2 Efficient Tests against $I(1)$ Alternative

As discussed above, the numerical results in this paper focus on the benchmark alternative where the stochastic trend follows an $I(1)$. Under this alternative, y_t follows a multivariate "local level model" (cf. Harvey (1989)), which is the alternative underlying well-known "stationarity" tests such as Nyblom and Mäkeläinen (1983), Nyblom (1989), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Nyblom and Harvey (2000), Jansson (2004), and others. Thus, suppose that the stochastic trend satisfies (6) and (10), so that

$$T^{-1/2}\sum_{t=1}^{\lfloor sT\rfloor} y_t \Rightarrow W_z(s) + B\int_0^s W_v(t)dt. \tag{24}$$

The optimal test depends on the value of B under the alternative, and it is convenient to assume

$$B = bS, \tag{25}$$

where b is a scalar and S is the $r \times k$ selection matrix equal to $S = [I_r, 0_{k-r}]$ when $r \leq k$ and $S = [I_k, 0_{k-r}]'$ when $r > k$. The invariance requirement (4) implies that (25) is without loss of generality whenever there exist orthonormal $r \times r$ and $k \times k$ matrices P_y and P_x such that $P_y B P_x = ||B||S$, which is always the case when $\min(r, k) = 1$. In the formulation (25), when $r \leq k$ (so that the number of linearly independent cointegrating vectors does not exceed the number of common trends), each element of y_t is the sum of an $I(0)$ component and an $I(1)$ component, where the common relative magnitude of the two components is determined by b. When $r > k$, there are fewer trends than cointegrating vectors, so that y_t can be rotated such that the trends load on only a subset of the variables in y_t. This is the "reduced rank" formulation used, for example, in the multivariate stationarity test proposed in Eliasz, Stock, and Watson (2004).

In this parameterization, the covariance matrix of Y depends on b and $R = S_z S_v' = E[W_z(1)W_v(1)']$, the correlation between the Wiener processes describing z_t and v_t. A straightforward calculation shows that Σ_{YY} can be written as

$$\Sigma_{YY} = (I_r \otimes I_q) + b^2(SS' \otimes D) + b(SR' \otimes F) + b(RS' \otimes F') \tag{26}$$

where F and D are $q \times q$ matrices, where D is a diagonal matrix with i'th diagonal element equal to $(\pi i)^{-2}$ and $F = [f_{ij}]$, with $f_{ij} = 0$ if i and j are both even or odd, and $f_{ij} = 4/[\pi^2(i^2 - j^2)]$ otherwise. (The simple diagonal form of D is due to the particular choice of the weighting functions Ψ in (14); see Section 2.3 in Müller and Watson (2008)).

Examination of (26) suggests three simplifications of the testing problem. First, because $F = -F'$, the final two terms cancel when SR' is symmetric. (When $r \leq k$, SR' is symmetric if $R_{ij} = R_{ji}$ for $i, j \leq r$, and when $r > k$, symmetry requires $R_{ij} = R_{ji}$ for $i, j \leq k$ and $R_{ij} = 0$ for $i > k$ and all j.) Thus, when SR' is symmetric, Σ_{YY} does not depend on R, which implies that the efficient test constructed using $R = 0$ is uniformly most powerful for all values of R with SR' symmetric. Second, when $r \leq k$, $SS' = I_r$, so in this case when SR' is symmetric, $\Sigma_{YY} = I_r \otimes (I_q + b^2 D)$, and from part (b) of Theorem 2, the optimal test rejects for large values of $\det(Y'Y)/\det(Y'(I_q + b^2 D)^{-1}Y)$. (This statistic with $r = 1$ was labeled "$LFST$" in Müller and Watson (2008) and we continue to use that label here.) Finally, when SR' is symmetric, but $r > k$, a calculation based on Theorem 2 (a) produces an expression for the best test. These simplifications are summarized in the following corollary.

Corollary 1 *For the alternative (24) and (25), the Neyman-Pearson test constructed with $R = 0$ is uniformly most powerful over all values of R with SR' symmetric, and rejects for large values of*

$$LFST(b) = \det(Y'Y)/\det(Y'(I_q + b^2 D)^{-1}Y)$$

when $r \leq k$, and for large values of

$$\xi(b) = \det(Y'Y)^{(q+k-r)/2} \det(Y'(I_q + b^2 D)^{-1}Y)^{-k/2} E_{\omega_Y}[|\det(\omega_Y)|^{q-r}]$$

when $r > k$, where $\operatorname{vec} \omega_Y \sim \mathcal{N}(0, \Omega_Y^{-1})$ *and* $\Omega_Y = \operatorname{diag}(I_k \otimes Y'(I_q + b^2 D)^{-1}Y, I_{r-k} \otimes Y'Y)$.

The corollary shows that when $r \leq k$ the best test is based on $LFST$, but when $r > k$, the optimal test statistic is ξ given in part (b) of the corollary.[12]

Table 3 presents 10%, 5%, and 1% critical values for the point-optimal $LFST(10/\sqrt{r})$ test for various values of r and q, where the alternative is chosen so that 5% test has approximately 50% power for $b = 10/\sqrt{r}$.

[12]This test statistic is more difficult to calculate than $LFST$ because ξ depends on the term $E_{\omega_Y}[|\det(\omega_Y)|^{q-r}]$, which requires evaluating absolute moments of order $q - r$ from an r^2-dimensional multivariate normal distribution. In an earlier version of this paper we compared the power of $\xi(b)$ and $LFST(b)$ for $(r, k) = (2, 1)$, $(3, 1)$, and $(3, 2)$ in models with SR' symmetric for $q = 12$. When $r = 2$ and $k = 1$, the power of the $LFST(b)$ statistic is within 3% of the power of $\xi(b)$ when the power of $\xi(b)$ is less than 50%, but the difference increases to nearly 10% when the power of $\xi(b)$ exceeds 80%. The differences are more substantial when $r = 3$ and $k = 1$, where the power difference is approximately 7% when power is 50%; the power differences are negligible when $r = 3$ and $k = 2$. Eliasz, Stock, and Watson (2004) report similar power differences in a related testing problem.

6.3 Power of Efficient Y-only Tests

Table 4 shows the power of the point-optimal $LFST$ test and the corresponding power envelope of the Y-only test for $r = 1$ in panel A, and $r = 2$ and $k = 1$ in panel B. In panel A, the power envelope is given by the $LFST$ test evaluated at the value of B under the alternative, while the point-optimal test $LFST$ is evaluated at $B = 10$. The power of the point-optimal test is very close to the Y-only power envelope. A more interesting comparison involves the power of the point-optimal $LFST$ test with the (Y, X)-power bound computed in the last section. Because the Y-only tests control size for any trend process, the relevant comparison is the unrestricted $H(s, t)$ bound shown in panels A and B of Table 1. The power of the point-optimal $LFST$ test differs from the Table 1 power bound by no more than 0.01 when $|B| = 7$ and no more that 0.03 when $|B| = 14$. Thus, for all practical purposes, the point-optimal $LFST$ test corresponds to an efficient test in models with a single cointegrating vector ($r = 1$).

The results in panel B of Table 4 are somewhat different. Here, because $r > k$, the Y-only power envelope is given by the ξ test of Corollary 1 and not by the $LFST$ test. The numerical results show that the relative power of the $LFST$ test depends on both B and R, and the loss can be large when B, and R are large and orthogonal ($\omega = 0$). Comparing results in panel B of Table 4 to the corresponding results in panel C of Table 1, also show that in this case there are potentially additional power gains associated with using data on both Y and X. For example, when $||B|| = 20$, $||R|| = 0.9$, and $\omega = 0$, the Y-only power envelope is 0.86, while the (Y, X) power bound in 0.94.

7 Wages, Prices, Employment and Output

The long-run relationship between wages (W), prices (P), employment (N) and real output (Y) has been of long-standing interest to economists. Labor's share of income is given by WN/YP, or (with lower case letters denoting logarithms) $w + n - y - p$, and is one of the "Great Ratios of Economics" investigated by Klein and Kosobud (1961). The average value of this ratio plays a key role for calibrating the aggregate function in macroeconomic business cycle models (see King, Plosser, and Rebelo (1988) and Cooley and Prescott (1995)). Differences between prices (p) and unit labor costs ($w+n-y$) or between real wages ($w-p$) and labor productivity ($y + n$) have been used as "cost-push" or error correction terms in

wage inflation or price inflation Phillips curve equations.[13] All of these suggest that w, p, y, and n are cointegrated, with cointegrating relationship

$$w - \beta_p p - \beta_y y - \beta_n n \tag{27}$$

where $\beta_p = 1$, $\beta_y = 1$, and $\beta_n = -1$.[14] In this section we investigate this hypothesis using post-war data for the United States.

We use data for the non-farm business and non-financial corporate sectors of the U.S. economy. Restricting attention to the non-farm business sector eliminates measurement issues associated with the government sector and with rental income from owner-occupied housing. Restricting attention even further to the non-financial corporate sector eliminates problems associated with the allocation of proprietors' income to wages and capital returns and price index problems in the financial sector.[15] Wages are measured by total labor compensation per hour (which includes employer-paid fringe benefits), prices and output are measured by the sector-specific price deflator and real output index, and employment is measured by total employee hours. Data are available quarterly from 1947 for the non-farm business sector and from 1958 for the non-financial corporate sector.[16]

In the standard model, (w, p, y, n) are driven by two real common trends representing labor supply and productivity, and by a nominal common trend that affects prices and wages. The real common trends are typically modeled as $I(1)$ processes (with drift) and the nominal common trend is often modeled as an $I(2)$ process. Of course, there is substantial uncertainty associated with these $I(1)/I(2)$ specifications, and so it is interesting to compare inference from standard cointegration methods that rely critically on these specifications with methods that allow for more general common trend processes.

The top panel of Table 5 shows estimates of the cointegrating coefficients and standard errors computed using a standard $I(1)/I(2)$ estimator (here, Stock and Watson's (1993) DOLS estimator). The estimates are somewhat close to their null values of $\beta_p = 1$, $\beta_y = 1$, and $\beta_n = -1$, but the Wald statistic soundly rejects this null for both data sets. The

[13] Sargan (1964) is a classic reference (although he used a smooth function of time as a proxy for productivity). More recent examples include Gordon (1985, 1998), Blanchard and Katz (1997), Brayton, Roberts, and Williams (1999), Mehra (2000) and Staiger, Stock, and Watson (2001).

[14] Hall (1986) seems to be the earliest paper to use formal I(1) cointegration methods to investigate this relationship.

[15] Gomme and Rupert (2004) discuss how these measurement issues affect inference about labor's share.

[16] The data are from the DRI Economics Database. The series used are LBCPU and LCPB (wages for the non-farm business and non-financial corporate sector), LBGDPU and LGDPB (prices), LBIPU and LIPB (output), and LBMNU and LMNB (employee hours).

final row of the table shows results using the $LFST$ statistic of the last section.[17] This test rejects the null for the non-farm business sector (p-value = 0.004) but not for the non-financial corporate sector (p-value = 0.28). Thus, standard $I(1)/I(2)$ inference and robust low-frequency inference coincide for the non-farm business sector, but not for the non-financial corporate sector.

Figure 1 plots the putative error correction term $w - p - y + n$ for each sector. The data for the non-farm business sector exhibits a negative trend over sample period, which readily explains the rejection of the null hypothesis for this sector. In contrast, the data for the non-financial corporate sector do not exhibit an obvious trend and (to our eyes) the plot appears to be consistent with an $I(0)$ series. Inference based on the $LFST$ statistic is consistent with this conclusion. The puzzle is then why the null is rejected so dramatically using standard cointegration methods.

A plausible explanation is that the common stochastic trends are not exactly $I(1)/I(2)$, as is assumed by the standard method. Uncertainty about the nature of the common trends is evident in confidence sets for largest autoregressive roots (local-to-unity parameters), long-memory parameters, or other nesting of the $I(1)$ and $I(2)$ models. This statistical evidence is reinforced by introspection about low-frequency changes in macroeconomy such as the productivity slowdown of the 1970s, the productivity rebound in the 1990s, changing demographics of the labor force, and shifts in monetary policy over the past five decades.. These phenomena are consistent with a range of common trend processes beyond the standard $I(1)/I(2)$ model. Yet, standard inference is based on specific characteristics of the $I(1)/I(2)$ trend model, which sharpens inference when the common trends follow these processes, but can lead to mistaken rejections otherwise. This can be seen in Figure 2 which plots confidence sets for $LFST$ and for the $I(1)/I(2)$ model. (The confidence sets impose the constraint that $\beta_y = -\beta_n$ so that we can show the plot in two dimensions.) The confidence ellipse constructed using the $I(1)/I(2)$ model is markedly smaller than the ellipse constructed from $LFST$ and does not contain the null value (indicated by the symbol "+" in the figure). In contrast, the $LFST$ confidence set, while larger, approximately efficiently exploits the available low-frequency information about the cointegrating vector in absence of specific assumptions about the common trends, and does not reject $\beta_p = 1$, $\beta_y = 1$, and $\beta_n = -1$.

[17] The LFST statistic was computed using the Ψ_j transformation in (14) where q was chosen to isolate frequencies associated with periods longer than eight years. The non-farm business data set contains 62 years of data so that $q = 15$, while the non-financial corporated data set contains 51 years of data, so that $q = 12$. Letting Y_T denote the $q \times 1$ vector of obsevations, the test statistic is $LFST = \left(\sum_{l=1}^{q} Y_{Tl}^2\right) / \left(\sum_{l=1}^{q} \frac{Y_{Tl}^2}{1+b^2/(\pi l)^2}\right)$, with $b = 10$.

8 Conclusion

This paper studies inference about the cointegrating vector in a framework in which the common stochastic trends are modelled in a flexible way beyond the standard $I(1)$ framework. The problem is studied with the low-frequency transformation approach suggested by Müller and Watson (2008). The paper derives bounds on the power of tests that control size over flexible stochastic trend specifications, and which maximize power against alternatives with the usual $I(1)$ trend. We find that a low-frequency version of Wright's (2000) test ($LFST$) essentially achieves the upper power bound in the model with $r = 1$ cointegrating vectors, making it an attractive choice for applied work.

The construction of efficient tests for the value of the cointegrating vector that control size for an unrestricted trend model when $r > 1$ remains an open question. However, the power bounds computed here provide a useful check for the efficiency of *ad hoc* tests that might be suggested for this problem, and the $LFST$ test remains a practically useful valid method also if $r > 1$.

If the stochastic trend model is tightly parametrized, the size control issue becomes muss less severe, as the dimension of the nuisance parameter is then small. Our algorithm for computing the approximate least upper bound on power for such models also yields an approximately power maximizing, feasible test that controls size. The paper thus also suggests a way to approximately efficiently exploit strong *a priori* knowledge about the stochastic trend.

The suggested method is generic in the sense that it computes an approximately efficient test in the presence of a low dimensional nuisance parameters under the null hypothesis. This type of problem arises naturally in nonstandard testing problems, so we would expect the method to be useful also in other contexts. For instance, the recent analysis of Elliott and Müller (2009) of inference about the pre and post break parameter value builds on this algorithm.

A Appendix

A.1 Proof of Theorem 1

Write $Y = (Y_1', Y_2', Y_3')'$ and $X = (X_1', X_2', X_3')'$, where Y_1 and X_1 have r rows, and Y_2 and X_2 have k rows. Consider the one-to-one mapping $h : \mathbb{R}^{q\times n} \mapsto \mathbb{R}^{q\times n}$ with

$$h(Y,X) = Q = \begin{pmatrix} Q_{Y1} & Q_{X1} \\ Q_{Y2} & Q_{X2} \\ Q_{Y3} & Q_{X3} \end{pmatrix} = \begin{pmatrix} Y_1 & Y_1^{-1}X_1 \\ Y_2(Y_1)^{-1} & X_2 - Y_2Y_1^{-1}X_1 \\ Y_3(Y_1)^{-1} & (X_3 - Y_3Y_1^{-1}X_1)(X_2 - Y_2Y_1^{-1}X_1)^{-1} \end{pmatrix}.$$

A straightforward calculation shows that $(\operatorname{vec} Q_{Y2}', \operatorname{vec} Q_{Y3}', \operatorname{vec} Q_{X3}')$ is a maximal invariant to (21). The inverse of h is given by

$$h^{-1}(Q) = \begin{pmatrix} Q_{Y1} & Q_{Y1}Q_{X1} \\ Q_{Y2}Q_{Y1} & Q_{X2} + Q_{Y2}Q_{Y1}Q_{X1} \\ Q_{Y3}Q_{Y1} & Q_{X3}Q_{X2} + Q_{Y3}Q_{Y1}Q_{X1} \end{pmatrix}.$$

Using matrix differentials (cf. Chapter 9 of Magnus and Neudecker (1988.)), a calculation shows that the Jacobian determinant of h^{-1} is equal to $(\det Q_{Y1})^{q-r+k}(\det Q_{X2})^{q-k-r}$. The density of Q is thus given by

$$(2\pi)^{-qn/2}(\det \Sigma_{(Y,X)})^{-1/2}|\det Q_{Y1}|^{q-r+k}|\det Q_{X2}|^{q-k-r}\exp[-\tfrac{1}{2}(\operatorname{vec} h^{-1}(Q))'\Sigma_{(Y,X)}^{-1}(\operatorname{vec} h^{-1}(Q))]$$

and we are left to integrate out Q_{Y1}, Q_{X1} and Q_{X2} to determine the density of the maximal invariant.

Now consider the change of variables from Q_{Y1}, Q_{X1}, Q_{X2} to ω_Y, ω_X and ω_{YX}

$$\begin{aligned} Q_{Y1} &= Y_1\omega_Y \\ Q_{X1} &= \omega_Y^{-1}Y_1^{-1}X_1\omega_X - \omega_Y^{-1}\omega_{YX} \\ Q_{X2} &= (X_2 - Y_2Y_1^{-1}X_1)\omega_X \end{aligned}$$

with Jacobian determinant $(\det Y_1)^r(\det(X_2 - Y_2Y_1^{-1}X_1))^k \det(-\omega_Y)^{-k}$. Noting that with this change, $h^{-1}(Q) = (Y\omega_Y, X\omega_X - Y\omega_{YX})$, we find that the density of the maximal invariant is equal to

$$\int (2\pi)^{-qn/2}(\det \Sigma_{(Y,X)})^{-1/2}|\det Y_1|^{q+k}|\det(X_2 - Y_2Y_1^{-1}X_1)|^{q-r}|\det \omega_Y|^{q-r}|\det \omega_X|^{q-k-r}$$
$$\cdot \exp[-\tfrac{1}{2}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX})'\Sigma_{(Y,X)}^{-1}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX})]d(\operatorname{vec}\omega_Y', \operatorname{vec}\omega_X', \operatorname{vec}\omega_{YX}')'.$$

Since $\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX}) = D_{YX}\operatorname{vec}(\omega_Y, \omega_X) - V_{0Y}\operatorname{vec}(\omega_{YX})$, we have

$$\begin{aligned}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX})'\Sigma^{-1}_{(Y,X)}\operatorname{vec}(Y\omega_Y, X\omega_X - Y\omega_{YX}) \\ = \operatorname{vec}(\omega_Y, \omega_X)'D'_{YX}\Sigma^{-1}_{(Y,X)}D_{YX}\operatorname{vec}(\omega_Y, \omega_X) \\ - 2\operatorname{vec}(\omega_Y, \omega_X)D'_{YX}\Sigma^{-1}_{(Y,X)}V_{0Y}\operatorname{vec}(\omega_{YX}) + \operatorname{vec}(\omega_{YX})'V'_{0Y}\Sigma^{-1}V_{0Y}\operatorname{vec}(\omega_{YX}).\end{aligned}$$

The result now follows from integrating out ω_{YX} by 'completing the square'.

A.2 Proof of Theorem 2

The proof to part (a) mimics the proof to Theorem 1 and is omitted. To prove part (b), note that because of invariance, we can set $\tilde{V}_{YY} = I_r$ without loss of generality, so that $\det\Sigma_{YY} = (\det\tilde{\Sigma}_{YY})^r$, $\Omega_Y = (I_r \otimes Y'\tilde{\Sigma}^{-1}_{YY}Y)$ and $(\det\Omega_Y)^{-1/2} = \det(Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-r/2}$. Since $(\operatorname{vec}\omega_Y)'\Omega_Y(\operatorname{vec}\omega_Y) = \operatorname{tr}(\omega'_Y Y'\tilde{\Sigma}^{-1}_{YY}Y\omega_Y)$, the density in part (a) of the Theorem becomes proportional to

$$(\det\tilde{\Sigma}_{YY})^{-r/2}\det(Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-r/2}\int|\det\omega_Y|^{q-r}\exp[-\tfrac{1}{2}\operatorname{tr}(\omega'_Y Y'\tilde{\Sigma}^{-1}_{YY}Y\omega_Y)]d(\operatorname{vec}\omega_Y).$$

Let $\tilde{\omega}_Y = (Y'\tilde{\Sigma}^{-1}_{YY}Y)^{1/2}\omega_Y$, so that $|\det\omega_Y|^{q-r} = \det(Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-(q-r)/2}|\det\tilde{\omega}_Y|^{q-r}$ and $\operatorname{vec}\omega_Y = (I_r \otimes (Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-1/2})\operatorname{vec}\tilde{\omega}_Y$, and the Jacobian determinant of the transformation from ω_Y to $\tilde{\omega}_Y$ is $\det(I_r \otimes (Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-1/2}) = (Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-r/2}$. Thus, the density is proportional to

$$(\det\tilde{\Sigma}_{YY})^{-r/2}\det(Y'\tilde{\Sigma}^{-1}_{YY}Y)^{-q/2}\int|\det\tilde{\omega}_Y|^{q-r}\exp[-\tfrac{1}{2}\operatorname{tr}(\tilde{\omega}'_Y\tilde{\omega}_Y)]d(\operatorname{vec}\tilde{\omega}_Y),$$

and the result follows.

A.3 Proof of Lemma 3

We first establish a preliminary result.

Lemma 4 *For any $t > 0$ and integer κ, the functions $\Psi_l : [0,t] \mapsto \mathbb{R}$ with $\Psi_l(s) = \sqrt{2}\cos(\pi l s)$, $l = 1, \cdots, \kappa$ are linearly independent.*

Proof. Choose any real constants c_j, $j = 1, \cdots, \kappa$, so that $\sum_{j=1}^{\kappa} c_j\Psi_j(s) = 0$ for all $s \in [0,t]$. Then also $\sum_{j=1}^{\kappa} c_j\Psi_j^{(i)}(0) = 0$ for all $i > 0$, where $\Psi_j^{(i)}(0)$ is the ith (right) derivative of Ψ_j at $s = 0$. A direct calculation shows $\Psi_j^{(i)}(0) = (-1)^{i/2}\sqrt{2}(\pi j)^i$ for even i. It is not hard to see that the $\kappa \times \kappa$ matrix with j,ith element $(-1)^{i/2}(\pi j)^i$ is nonsingular, so that $\sum_{j=1}^{\kappa} c_j\Psi_j^{(i)}(0) = 0$ for $i = 2, 4, \cdots, 2\kappa$ can only hold for $c_j = 0$, $j = 1, \cdots, \kappa$. ■

For the proof Lemma 3 we construct $H(s,t)$ such $\operatorname{vec} Z = \int_0^1 (I_r \otimes \Psi(t)) S_z dW(t)$ and $\operatorname{vec} V = \int_0^1 \int_t^1 (H(s,t) \otimes \Psi(s)) ds\ dW(t)$ have the specified covariance matrix. The proof of the slightly more difficult part (b), where $H(s,t) = G(s,t)S_v$, is based on the following observations:

(i) Ignoring the restriction on the form of $\operatorname{vec} V$, it is straightforward to construct an appropriate multivariate normal vector $\operatorname{vec} V$ from a linear combination of $\operatorname{vec} Z$ and ζ, where $\zeta \sim \mathcal{N}(0, I_{kq \times kq})$ independent of Z.

(ii) Suppose that $R = S$ was allowed, where $S = (I_r, 0_{r \times (k-r)})$. Then $S_z = SS_v$, $\operatorname{vec} Z \sim \int_0^1 F_Z(t) S_v dW(t)$ for $F_Z(t) = S \otimes \Psi(t)$, and one can also easily construct ζ as in (i) via $\zeta = \int_0^1 F_\zeta(t) S_v dW(t)$ by an appropriate choice of F_ζ. Since Ito-Integrals are linear, one could thus write $\operatorname{vec} V = \int_0^1 F(t) S_v dW(t)$ with F a linear combination of F_Z and F_ζ, using observation (i).

(iii) For any matrix function $F : [0,1] \mapsto \mathbb{R}^{kq \times k}$ that is equal to zero on the interval $(1-\varepsilon, 1]$ for some $\varepsilon > 0$, one can set $G(s,t) = (I_k \otimes \Psi(s)' J(t)^{-1}) F(t)$, where $J(t) = \int_t^1 \Psi(s)\Psi(s)' ds$ and obtain $\int_0^1 \int_t^1 (G(s,t) \otimes \Psi(s)) ds\ S_v dW(t) = \int_0^1 F(t) S_v dW(t)$, since for any matrix A with k rows and vector v, $A \otimes v = (I_k \otimes v)A$.

The following proof follows this outline, but three complications are addressed: $R = S$ is not allowed; the matrix function F needs to be zero on the interval $(1-\varepsilon, 1]$, which does not happen automatically in the construction in (ii); one must verify that the process $\int_0^s G(s,t) S_v dW(t)$ admits a cadlag version.

Set S_z to be the first r rows of I_n. Since $\Psi_l(1-s) = (-1)^l \Psi_l(s)$ for all $l \geq 1$, Lemma 4 implies that $J(t) = \int_t^1 \Psi(s)\Psi(s)' ds$ and $I_q - J(t)$ are nonsingular for any $t \in (0,1)$. The $rq \times 1$ random vector $\operatorname{vec} Z_\varepsilon = \int_0^{1-\varepsilon} (S_z \otimes \Psi(s)) dW(s)$ thus has nonsingular covariance matrix $I_r \otimes \Sigma_q^\varepsilon$, where $\Sigma_q^\varepsilon = I_q - J(1-\varepsilon)$. Also, since

$$\Sigma^* = \begin{pmatrix} I_r \otimes I_q & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is positive definite, so is $I_{rq} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, so that we can choose $0 < \varepsilon < 1$ such that $I_r \otimes \Sigma_q^\varepsilon - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is positive definite. With that choice of ε, also $\Sigma_{22} - \Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12}$ is positive definite.

For part (a) of the lemma, define the $[0,1] \mapsto \mathbb{R}^{kq \times n}$ function $F_a(t) = A_a(I_n \otimes \Psi(t))$, where $A_a = (A_{a1}, A_{a2})$ with $A_{a1} = \Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}$ and $A_{a2} = (\Sigma_{22} - \Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12})^{1/2}(I_k \otimes (\Sigma_q^\varepsilon)^{-1/2})$.

For part (b) of the lemma, choose $0 < \rho < 1$ so that $\Sigma_{22} - \rho^{-2}\Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12}$ is positive definite. Choose S_v to be the first k rows of I_n multiplied by ρ, so that $R = S_z S_v' = \rho S$. Let $\tilde{\Psi}_1(s)$ be scaled residuals of a continuous time projection of $\mathbf{1}[s \leq 1-\varepsilon]\Psi_{q+1}(s)$ on $\{\mathbf{1}[s \leq 1-\varepsilon]\Psi_l(s)\}_{l=1}^q$ on the unit interval, and let $\tilde{\Psi}_j(s)$, $j = 2, \cdots, kq$ be the scaled residuals of continuous time projection

of $\mathbf{1}[s \le 1-\varepsilon]\Psi_{q+j}(s)$ on $\{\mathbf{1}[s \le 1-\varepsilon]\Psi_l(s)\}_{l=1}^q$ and $\{\mathbf{1}[s \le 1-\varepsilon]\tilde{\Psi}_l(s)\}_{l=1}^{j-1}$. By Lemma 4, $\tilde{\Psi}_j(s)$, $j = 1, \cdots, kq$, are not identically zero, and we can choose their scale to make them orthonormal. Let $\tilde{\Psi}(s) = (\tilde{\Psi}_1(s), \cdots, \tilde{\Psi}_{kq}(s))'$, the $k \times 1$ vector $\iota_k = (1, 0, \cdots, 0)'$, and $A_b = (A_{b1}, A_{b2})$ with $A_{b1} = \rho^{-1}\Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}$ and $A_{b2} = (\Sigma_{22} - \rho^{-2}\Sigma_{21}(I_r \otimes \Sigma_q^\varepsilon)^{-1}\Sigma_{12})^{1/2}$. Now define the $[0,1] \mapsto \mathbb{R}^{kq \times n}$ function

$$F_b(t) = A_b \begin{pmatrix} S \otimes \Psi(t) \\ \iota_k' \otimes \tilde{\Psi}(t) \end{pmatrix} S_v.$$

For both parts, that is for $i \in \{a, b\}$, set

$$H_i(s,t) = (I_k \otimes \Psi(s)' J(t)^{-1}) F_i(t) \text{ for } t \in [0, 1-\varepsilon]$$

and $H_i(s,t) = 0$ otherwise. With this choice

$$\begin{aligned} \operatorname{vec} V_i &= \int_0^1 \int_t^1 (H_i(s,t) \otimes \Psi(s)) ds \; dW(t) \\ &= \int_0^{1-\varepsilon} \int_t^1 ((I_k \otimes \Psi(s)\Psi(s)' J(t)^{-1}) F_i(t)) ds \; dW(t) \\ &= \int_0^{1-\varepsilon} F_i(t) dW(t). \end{aligned}$$

Thus

$$\begin{aligned} E[(\operatorname{vec} V_i)(\operatorname{vec} V_i)'] &= \int_0^{1-\varepsilon} F_i(t) F_i(t)' dt \\ E[(\operatorname{vec} V_i)(\operatorname{vec} Z)'] &= \int_0^{1-\varepsilon} F_i(t)(S_z \otimes \Psi(t))' dt \end{aligned}$$

since $\operatorname{vec}(Z - Z_\varepsilon) = \int_{1-\varepsilon}^1 (I_r \otimes \Psi(t)) S_z dW(t)$ is independent of $\operatorname{vec} V_i$. A direct calculation now shows that $\int_0^{1-\varepsilon} F_a(t) F_a(t)' dt = A_a(I_n \otimes \Sigma_q^\varepsilon) A_a'$, $\int_0^{1-\varepsilon} F_a(t)(S_z \otimes \Psi(t))' dt = A_a(S_z' \otimes \Sigma_q^\varepsilon)$, $\int_0^{1-\varepsilon} F_b(t) F_b(t)' dt = A_b \operatorname{diag}(I_r \otimes \Sigma_q^\varepsilon, I_{kq}) A_b'$ and $\int_0^{1-\varepsilon} F_b(t)(S_z \otimes \Psi(t))' dt = \rho A_b(S_z' \otimes \Sigma_q^\varepsilon)$, so that from the definitions of A_i, $E[(\operatorname{vec} V_i)(\operatorname{vec} V_i)'] = \Sigma_{22}$ and $E[(\operatorname{vec} V_i)(\operatorname{vec} Z)'] = \Sigma_{21}$, as required.

It thus remains to show that the processes $\int_0^s H_i(s,t) dW(t)$, $i \in \{a, b\}$, admit a cadlag version.

Recall that $||A||$ is the Frobenius norm of the real matrix A, $||A|| = \sqrt{\operatorname{tr} A'A}$, which is submultiplicative. If $v \sim \mathcal{N}(0, \Sigma)$, then $E[||v||^4] = E[(v'v)^2] = 2\operatorname{tr}(\Sigma^2) + (\operatorname{tr}\Sigma)^2 \le 3(\operatorname{tr}\Sigma)^2$, so that with $\int_s^t H_i(u,\lambda) dW(\lambda) \sim \mathcal{N}(0, \int_s^t H_i(u,\lambda) H_i(u,\lambda)' d\lambda)$, we find

$$\begin{aligned} E[|| \int_s^t H_i(u,\lambda) dW(\lambda) ||^4] &\le 3(\operatorname{tr} \int_s^t H_i(u,\lambda) H_i(u,\lambda)' d\lambda)^2 \\ &\le 3(\int_s^t ||H_i(u,\lambda)||^2 d\lambda)^2. \end{aligned}$$

Thus, for $0 \leq s < t \leq 1$, we have with $\psi(s) = d\Psi(s)/ds$

$$\begin{aligned}
& E[|| \int_0^t H_i(t,\lambda)dW(\lambda) - \int_0^s H_i(s,\lambda)dW(\lambda)||^4] \\
= \; & E[|| \int_0^s (H_i(t,\lambda) - H_i(s,\lambda))dW(\lambda) + \int_s^t H_i(t,\lambda)dW(\lambda)||^4] \\
\leq \; & 3[\int_0^s ||H_i(t,\lambda) - H_i(s,\lambda)||^2 d\lambda + \int_s^t ||H_i(t,\lambda)||^2 d\lambda]^2 \\
\leq \; & 3[k^2(\sup_{0\leq\lambda\leq 1-\varepsilon} ||J(\lambda)^{-1}||^2 ||F_i(\lambda)||^2)(||\Psi(s) - \Psi(t)||^2 + (t-s)\sup_{0\leq\lambda\leq 1} ||\Psi(\lambda)||^2)]^2 \\
\leq \; & 3k^4(\sup_{0\leq\lambda\leq 1-\varepsilon} ||J(\lambda)^{-1}||^4 ||F_i(\lambda)||^4)(\sup_{0\leq\lambda\leq 1} ||\psi(\lambda)||^2 + \sup_{0\leq\lambda\leq 1} ||\Psi(\lambda)||^2)^2 (t-s)^2
\end{aligned}$$

where the last inequality follows from $\Psi(t) - \Psi(s) = (t-s)\int_0^1 \psi(s + \lambda(t-s))d\lambda$, so that by Kolmogorov's continuity theorem (p. 14 of Oksendal (2000)), there exist continuous (and thus cadlag) versions of the stochastic processes $\int_0^s H_i(s,t)dW(t)$, $i \in \{a,b\}$.

A.4 Parameterization of $\Sigma_{(Y,X)}$ under H_0 in the restricted trend models

G-model with $r > k$: Because of invariance, it is without loss of generality to assume that the first $r-k$ rows of R are equal to zero, so that the first $r-k$ columns of Z are independent of V. The joint distribution of V and the last k rows of Z are then just as in the model with $r = k$, so that Lemma 3 implies that in the G-model with $r > k$, $\Sigma_{(Y,X)}$ is of the form $\Sigma_{(Y,X)} = \text{diag}(I_{r-k} \otimes I_q, \Sigma_k^*)$ under the null hypothesis, where Σ_k^* is any positive definite $k^2q \times k^2q$ matrix with upper left $kq \times kq$ block equal to the identity matrix. The nuisance parameter θ is thus of dimension $k^2q^2 + kq(kq+1)/2$

Diagonal G-model: Let Z_V and ζ be $q \times k$ random matrices with $\text{vec}(Z, Z_V, \zeta) \sim \mathcal{N}(0, \Sigma_{(Z,Z_V,\zeta)})$, where

$$\Sigma_{(Z,Z_V,\zeta)} = \text{diag}\left(\begin{bmatrix} I_r & R \\ R' & I_k \end{bmatrix}, I_k\right) \otimes I_q.$$

A construction as in the proof of Lemma 3 implies that the j'th column of V can be chosen as an arbitrary linear combination of the j'th column of Z_V and the j'th column of ζ, $j = 1, \cdots, k$ (subject to the constraint that the resulting matrix is positive definite). Thus, $\Sigma_{(Y,X)}$ may be parametrized as $\Sigma_{(Y,X)} = A_{(Z,Z_V,\zeta)}\Sigma_{(Z,Z_V,\zeta)}A'_{(Z,Z_V,\zeta)}$, where

$$A_{(Z,Z_V,\zeta)} = \begin{pmatrix} I_{rq} & 0 & 0 \\ 0 & \text{diag}(A_{V,1}, A_{V,2}, \cdots, A_{V,k}) & \text{diag}(L_{\zeta,1}, L_{\zeta,2}, \cdots, L_{\zeta,k}) \end{pmatrix},$$

$A_{V,j}$ are arbitrary $q \times q$ matrices and $L_{\zeta,j}$ are arbitrary lower diagonal $q \times q$ matrices. Including R, θ is thus of dimension $rk + kq^2 + kq(q+1)/2$.

In the stationary diagonal model where $G(s,t) = \text{diag}(g_1^S(s-t), \cdots, g_k^S(s-t))$, we set g_j^S to be step functions

$$g_j^S(x) = \sum_{i=1}^{n_g} \xi_{j,i} \mathbf{1}\left[\frac{i-1}{n_g+1} \leq \frac{x}{1+x} < \frac{i}{n_g+1}\right] \tag{28}$$

for $n_g = 40$ and some scalar parameters $\xi_{j,i}$, $j = 1, \cdots, k$, $i = 1, \cdots, n_g$. The steps occur at the points $i/(n_g + 1 - i)$, so that more flexibility is allowed for small values of x (half of the points are associated with values of x less than 1, for example). The values of Σ_{ZV} and Σ_{VV} then follow from (17). In this specification θ contains the kn_g coefficients $\xi_{j,i}$ and the rk coefficients in the correlation matrix R. While the specification (28) only captures a subset of all possible covariance matrices $\Sigma_{(Y,X)}$ in the (nonparametric) stationary model, any test that controls size for all functions $H(s,t)$ of the form $H(s,t) = \text{diag}(g_1^S(s-t), \cdots, g_k^S(s-t))S_v$ *a fortiori* has to control size for the specification (28). The upper bounds on power of tests that control size for all values of θ under (28) are therefore also upper bounds for tests that control size for the generic stationary stochastic trend model.

In the local-to-unity model, a straightforward (but tedious) calculation determines the value of $\Sigma_{(Y,X)}$ as function of the $k \times k$ matrix C and the $r \times k$ correlation matrix R, so that θ is of dimension $k^2 + rk$. Finally, the $I(1)$ model is a special case of the local-to-unity model with $C = 0$.

A.5 Kullback-Leibler Divergence Problem of Section 5.3.2

Let Σ_1 denote the covariance matrix of $\text{vec}(Y, X)$ under a specific $I(1)$ alternative as described in Subsection 5.2.1(that is, for specific values of $B = B_1$ and $R = R_1$), let $\Sigma_0(\theta)$ with $\theta \in \Theta$ be the covariance matrix of $\text{vec}(Y, X)$ under the null for the relevant restrictions on the stochastic trend, and define the $nq \times nq$ matrix (recall that $n = r + k$)

$$A(\gamma) = \begin{bmatrix} \gamma_{yz} \otimes I_q & 0 \\ \gamma_{xz} \otimes I_q & \gamma_{xv} \otimes I_q \end{bmatrix}$$

where γ_{yz} is $r \times r$, γ_{xz} is $k \times r$, and γ_{xv} is $k \times k$. This yields $A(\gamma)\,\text{vec}(Y, X) \sim \mathcal{N}(0, A(\gamma)\Sigma_0(\theta)A(\gamma)')$. Denote the Kullback-Leibler divergence between the $nq \times 1$ distributions $\mathcal{N}(0, \Sigma_1)$ and $\mathcal{N}(0, \Sigma_0)$ as $K(\Sigma_1, \Sigma_0) = \frac{1}{2}\ln(\det \Sigma_1 / \det \Sigma_0) + \frac{1}{2}\,\text{tr}(\Sigma_0^{-1}\Sigma_1) - nq$. The value of θ^* is chosen to numerically solve

$$\min_{\gamma \in \mathbb{R}^{r^2+k^2+kr}} K(\Sigma_1, A(\gamma)\Sigma_0(\theta^*)A(\gamma)') = \min_{\theta \in \Theta, \gamma \in \mathbb{R}^{r^2+k^2+kr}} K(\Sigma_1, A(\gamma)\Sigma_0(\theta)A(\gamma)'), \tag{29}$$

that is, θ^* numerically minimizes the Kullback-Leibler divergence (or KLIC) between the null and alternative densities of (Y, X), allowing for transformations as described by $A(\gamma)$ under the null. While these transformations do not affect the implied distribution of the maximal invariant, they do in general lead to a different θ^*, which we found to yield a slightly lower upper bound. The

minimization problem is over a high dimensional parameter, but the objective function is quickly computed and well behaved, so that numerical minimization is feasible.

A.6 Algorithm for Approximating the Least Upper Power Bound and Optimal Test

A computationally more convenient variation of the size adjustment idea described in the main text is as follows: Starting from the level α test $\varphi_\Lambda = \mathbf{1}[\mathrm{LR}_\Lambda > \mathrm{cv}_\Lambda]$ of H_Λ against H_1, for some small $\varepsilon > 0$, let $\mathrm{cv}_\Lambda^\varepsilon$ be an adjusted critical value so that the resulting test $\varphi_\Lambda^\varepsilon = \mathbf{1}[\mathrm{LR}_\Lambda > \mathrm{cv}_\Lambda^\varepsilon]$ (with $\mathrm{cv}_\Lambda^\varepsilon > \mathrm{cv}_\Lambda$) has only slightly lower power than φ_Λ, i.e. $\int \varphi_\Lambda^\varepsilon h d\mu = \beta_\Lambda - \varepsilon$. Now *if* $\varphi_\Lambda^\varepsilon$ is of level α under H_0, i.e. $\sup_{\theta\in\Theta} \int \varphi_\Lambda^\varepsilon f_\theta d\mu < \alpha$, then we have a level α test of H_0 against H_1 with power that is only ε below β_Λ, and the least upper bound is again sandwiched between β_Λ and $\beta_\Lambda - \varepsilon$. The advantage of this method over the direct size adjustment discussed in the text is that the size adjustment is costly to compute, while Λ can often be quickly dismissed by checking its size control for a small number of values of θ under H_0.

Now consider discrete distributions for Λ: Let $\Theta^N = \{\theta_1, \cdots, \theta_N\} \subset \Theta$ for some $N > 1$ and consider the restricted null hypothesis H_N :The density of U is f_θ, $\theta \in \Theta^N$. In this restricted problem, the least favorable distribution is fully described by the point masses p_i^* on θ_i, $i = 1, \cdots, N$, where $\sum_{i=1}^N p_i^* = 1$. The resulting test φ_N^* is thus of the form $\varphi_N^* = \mathbf{1}[\sum_{i=1}^N p_i^* f_{\theta_i}/h < 1/\mathrm{cv}_N]$. Note that by construction, the test φ_N^* is of level α on $\Theta^N \subset \Theta$. The central idea of the algorithm is to identify a (hopefully not too large) set of points Θ^N so that corresponding adjusted test $\varphi_N^{*\varepsilon}$ is of level α on the whole set Θ.

Introduce the notation $\varphi(\bar\theta, \bar p, \mathrm{cv})(u)$ for the test $\varphi = \mathbf{1}[\sum_{i=1}^N p_i f_{\theta_i}/h < 1/\mathrm{cv}]$ evaluated at u, with $\bar\theta = (\theta_1, \cdots, \theta_N)'$ and $\bar p = (p_1, \cdots, p_N)'$, and $\sum_{i=1}^N p_i = 1$ (but $\bar p$ is not necessarily the least favorable distribution on Θ^N). The rejection probability of φ under the alternative is $\Pi_1(\bar\theta, \bar p, \mathrm{cv}) = \int \varphi(\bar\theta, \bar p, \mathrm{cv})(u) h(u) d\mu(u)$, and it is $\Pi_0(\bar\theta, \bar p, \mathrm{cv}; \theta)(u) = \int \varphi(\bar\theta, \bar p, \mathrm{cv}) f_\theta(u) d\mu(u)$ under the null hypothesis with $\theta \in \Theta$. We numerically approximate $\Pi_0(\bar\theta, \bar p, \mathrm{cv}; \theta)$ by

$$\hat\Pi_0(\bar\theta, \bar p, \mathrm{cv}; \theta) = \frac{1}{m} \sum_{j=1}^m \Upsilon_L \left(\frac{\sum_{i=1}^N p_i f_{\theta_i}(u_j)}{h(u_j)}, \frac{1}{\mathrm{cv}} \right) \tag{30}$$

where for some real $L > 0$, $\Upsilon_L : \mathbb{R}^2 \mapsto \mathbb{R}$ is defined as $\Upsilon_L(x, y) = y^L/(y^L + x^L)$. The pseudo random variables u_j, $j = 1, \cdots, m$ have density f_θ and are obtained by suitably transforming some underlying pseudo random variables ξ_j, $u_j = g_\theta(\xi_j)$, $j = 1, \cdots, m$. The variables ξ_j are drawn only once in the evaluation of $\hat\Pi_0$ at different arguments (so the transformation g_θ depends on θ). Define $\hat\Pi_1(\bar\theta, \bar p, \mathrm{cv})$ analogously, with u_j given by $g_h(\xi_j)$. Note that as $L \to \infty$, $\Upsilon_L(x, y) \to \mathbf{1}[x < y] + \frac{1}{2}\mathbf{1}[x = y]$, so that for L large, $\hat\Pi_0(\bar\theta, \bar p, \mathrm{cv}; \theta)$ approximates the standard Monte Carlo integration for the rejection probability of $\varphi(\bar\theta, \bar p, \mathrm{cv})$. The advantage of choosing $L < \infty$ is that

$\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}; \theta)$ and $\hat{\Pi}_1(\bar{\theta}, \bar{p}, \text{cv})$ become smooth and differentiable functions of their arguments, which greatly simplifies numerical optimizations. The computations in this paper were performed with $m = 25,000$ and $L = 25$.

The algorithm consists of three subroutines SR1, SR2 and SR3.

SR1 The routine takes $\bar{\theta} = (\theta_1, \cdots, \theta_N)$ as given, and returns an estimate of the least favorable distribution on Θ^N, as described by $\bar{p} = (p_1, \cdots, p_N)$. By Theorem 3.8.1 of Lehmann and Romano (2005), the least favorable distribution $\bar{p}^* = (\bar{p}_1^*, \cdots, \bar{p}_N^*)'$ has the two properties (i) $\int \varphi(\bar{\theta}, \bar{p}^*, \text{cv}) f_{\theta_l} d\mu \leq \alpha$ for $l = 1, \cdots, N$; and (ii) $\int \varphi(\bar{\theta}, \bar{p}^*, \text{cv}) f_{\theta_l} d\mu < \alpha$ only if $p_l = 0$. This motivates the joint determination of $\bar{p}$ and cv as numerical solutions to

$$\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}; \theta_l) \leq \alpha \text{ and } p_l(\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}; \theta_l) - \alpha) = 0 \text{ for } l = 1, \cdots, N. \tag{31}$$

Specifically, we determine appropriate $\bar{p}$ and cv by minimizing the objective function

$$\sum_{l=1}^{N} (a_0 p_l + \exp[a_1(\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}; \theta_l) - \alpha)])(\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}; \theta_l) - \alpha)^2 \tag{32}$$

where $a_0 = 100$ and $a_1 = 2000$. As a function of $\bar{p}$ and cv, (32) is continuous and with known first derivative, so that a standard quasi-Newton optimizer can be employed. Also, the $N^2 m$ numbers $f_{\theta_i}(g_{\theta_l}(\xi_j))/h(g_{\theta_l}(\xi_j))$ for $i = 1, \cdots, N$, $l = 1, \cdots, N$ and $j = 1, \cdots, m$ can be computed and stored once to speed up the the evaluation of $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}; \theta_i)$ and its partial derivatives.

SR2 The routine takes $(\bar{\theta}, \bar{p})$ as inputs and returns vectors $(\bar{\theta}_1, \bar{p}_1)$ of length $N_1 \leq N$ by eliminating pairs of values (θ_j, p_j) with p_j approximately equal to zero.

SR3 The routine takes $(\bar{\theta}, \bar{p})$ as given and either identifies $(\bar{\theta}, \bar{p})$ as yielding a sufficiently precise approximation to the least favorable distribution, or it returns a parameter value $\hat{\theta} \in \Theta$ that needs to be included in the set of points Θ^N. Specifically, the routine consists of three steps:

(a) Solve for the real number cv_Λ that satisfies $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \text{cv}_\Lambda; \theta_l) \leq \alpha$ for all $l = 1, \cdots, N$, so that the test $\varphi(\bar{\theta}, \bar{p}, \text{cv}_\Lambda)$ is the (approximate) level α likelihood ratio test of $H_\Lambda : U$ has density $\sum_{l=1}^{N} p_l f_{\theta_l}$ against H_1.

(b) Compute $\beta_\Lambda = \hat{\Pi}_1(\bar{\theta}, \bar{p}, \text{cv}_\Lambda)$, and numerically solve for $\text{cv}_\Lambda^\varepsilon \geq \text{cv}_\Lambda$ such that $\hat{\Pi}_1(\bar{\theta}, \bar{p}, \text{cv}_\Lambda) - \hat{\Pi}_1(\bar{\theta}, \bar{p}, \text{cv}_\Lambda^\varepsilon) = \varepsilon$. By Lemma 1, β_Λ is (an estimate of) a power bound on level α tests of H_0. As described above, the size adjustment as a function of the power implies that *if* $\varphi(\bar{\theta}, \bar{p}, \text{cv}_\Lambda^\varepsilon)$ is of level α under H_0, then we have found a test whose power is within ε of the bound. The computations in this paper use $\varepsilon = 2.5\%$.

(c) Check on a grid of values $\Theta_G \subset \Theta$ whether $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ controls size, i.e. evaluate $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_j)$ for all $\theta \in \Theta_G$ in the finite set Θ_G. If $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_j) > \alpha$ for some j, return $\hat{\theta} = \theta_j$. Otherwise, preliminarily accept β_Λ as the approximate least upper bound, and $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ as an approximately efficient test. As a practical matter, it makes sense to return $\hat{\theta}_j = \theta_j$ even if $\hat{\Pi}_0(\bar{\theta}, \bar{p}, \mathrm{cv}; \theta_j)$ is below, but very close to α. We use a threshold of 4.8% for $\alpha = 5\%$.

Overall the algorithm iterates between the subroutines as follows:

1. Initialize $\bar{\theta}$ with $N = 25$ values of θ_j that are spread out over the grid Θ_G, and call SR1.
2. Call SR2 to obtain a new N and $(\bar{\theta}, \bar{p})$ pair.
3. While SR3 returns $\hat{\theta}$:

   Add $\hat{\theta}$ to $\bar{\theta}$ (so that N is increased by one) and call SR1.
4. Call SR2 to obtain a new N and $(\bar{\theta}, \bar{p})$ pair. Repeat Step 3.
5. Perform a final check on whether $\varphi(\bar{\theta}, \bar{p}, \mathrm{cv}_\Lambda^\varepsilon)$ is a level α test by evaluating its rejection probability over a fine grid of values for θ_j, using a different set of draws of pseudo-random variables ξ_j in (30). For the correlation R, we use the grid $R = -0.99, -0.96, \cdots, 0.99$ in the $I(1)$ model, and in the local-to-unity model, a square grid of the same values of R, and $C = -3, -2.5, \cdots, -.5, 0, e^{-1}, e^{-0.8}, \cdots, e^{6.8}, e^7$.

The advantage of the thinning operation in SR2 is that it accelerates the computation of the test statistic, and it facilitates the numerical minimization of (32). We do not call SR2 after each call of SR1, though, because doing so can result in cycles, so that Step 3 above could potentially result in an infinite loop. Step 4 "cleans" the feasible test and is skipped when only an estimator of the least upper bound is required.

A.7 Validity of $r = 1, k = 2$ bounds for $r = 1, k > 2$

Here we show that the power bounds computed using $r = 1$ and $k = 2$ also serve as power bounds for models with $r = 1$ and all values of $k > 2$.

To see why, first consider the alternative $I(1)$ model as described in subsection 5.2.1, $Y = Z + VB'$ and $X = V$. Let P be a $k \times k$ orthonormal matrix whose last $k-2$ rows are orthogonal to R and B, and whose second row is orthogonal to R. Partition $X = (X_{12}, X_{3k})$, where X_{12} contains the first two columns of X and X_{3k} contains the remaining $k-2$ columns. By invariance, there is no loss in generality in setting $X = \tilde{X}P = (\tilde{X}_{12}, \tilde{X}_{3k})P$, so that $Y = Z + \tilde{X}PB' = Z + \tilde{X}_{12}B_{12}'$,

where $\tilde{X}_{12}$ and B_{12} are the first two columns of $\tilde{X}$ and B, respectively, and the last $k-1$ columns of $\tilde{X}$ (and thus $\tilde{X}_{3k}$) are independent of Z. The group of transformations

$$(Y, \tilde{X}_{12}, \tilde{X}_{3k}) \to (YA_{yy}, \tilde{X}_{12}\tilde{A}_{xx} + YA_{xy}, \tilde{X}_{3k}) \tag{33}$$

for nonsingular A_{yy} and $\tilde{A}_{xx}$ is a subset of the transformations $(Y, \tilde{X}) \to (YA_{yy}, \tilde{X}A_{xx} + YA_{xy})$, so the best invariant test to (33) is as least as powerful as the best invariant test to (21). Let $\tilde{Q}_{12}$ be a maximal invariant to $(Y, \tilde{X}_{12}) \to (YA_{yy}, \tilde{X}_{12}\tilde{A}_{xx} + YA_{xy})$, such that $\{\tilde{Q}_{12}, \tilde{X}_{3k}\}$ is a maximal invariant to (33). Since $\tilde{X}_{3k}$ is independent of $(Y, \tilde{X}_{12})$, the density of $\{\tilde{Q}_{12}, \tilde{X}_{3k}\}$ under the alternative factors as $f_{a,\tilde{Q}_{12}} \cdot f_{a,\tilde{X}_{3k}}$.

For all null models discussed in subsection 5.2.2, it is possible to choose $X = (X_{12}, X_{3k}) = V$ in a way such that X_{3k} is independent of X_{12} with marginal distribution $f_{0,X_{3k}} = f_{a,\tilde{X}_{3k}}$, (i.e. it corresponds to the $I(1)$ model) and the possibilities for X_{12} and its relationship with Z are the same as in the version of the model with $k = 2$. Thus, with this choice, the term $f_{a,\tilde{X}_{3k}}$ cancels in the likelihood ratio test of the maximal invariant to (33), and the testing problem corresponds precisely to the model with $k = 2$.[18] An upper bound for the model with $r = 1$ and $k = 2$ is therefore also an upper bound for the model with $r = 1$ and $k > 2$.

[18] This is not strictly true for the stationary G-model, which excludes $I(1)$ stochastic trends. But the low-frequency transformation of the suitably scaled stationary local-to-unity model converges in mean squared to the $I(1)$ model as the local-to-unity parameter approaches zero (cf. Section 2.4 of Müller and Watson (2008)), so that the additional discriminatory power from X_{3k} can be made arbitrarily small, and the conclusion continues to hold.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 146, 241–254.

Blanchard, O. J., and L. F. Katz (1997): "What We Know and Do Not Know About the Natural Rate of Unemployment," *Journal of Economic Perspectives*, 11, 51–72.

Bobkoski, M. J. (1983): "Hypothesis Testing in Nonstationary Time Series," *unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin*.

Brayton, F., F. M. Roberts, and J. C. Williams (1999): "What's Happened to the Phillips Curve?," *Board of Governors of the Federal Reserve System Finance and Discussion Series No. 49*.

Campbell, J. Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60.

Caner, M., and L. Kilian (2001): "Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate," *Journal of International Money and Finance*, 20, 639–657.

Cavanagh, C. L. (1985): "Roots Local To Unity," *Working Paper, Harvard University*.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147.

Chan, N. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Cooley, T., and E. C. Prescott (1995): "Economic Growth and Business Cycles," in *Frontiers of Business Cycle Research*, ed. by T. Cooley, pp. 1–38, Princeton, NJ. Princeton University Press.

den Haan, W. J., and A. T. Levin (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics 15*, ed. by G. S. Maddala, and C. R. Rao, pp. 309–327. Elsevier, Amsterdam.

Eliasz, P., J. H. Stock, and M. W. Watson (2004): "Optimal Tests for Reduced Rank Time Variation in Regression Coefficients and for Level Variation in the Multivariate Local Level Model," *Working paper, Princeton University.*

Elliott, G. (1998): "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158.

Elliott, G., and U. K. Müller (2009): "Pre and Post Break Parameter Inference," *Working Paper, Princeton University.*

Elliott, G., and J. H. Stock (1994): "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown," *Econometric Theory*, 10, 672–700.

Gomme, P., and P. Rupert (2004): "Measuring Labor's Share of Income," *Federal Reserve Bank of Cleveland Policy Dicussion Paper No. 7.*

Gordon, R. J. (1985): "Understanding the Inflation in the 1980s," *Brookings Papers on Economic Activity*, 1, 263–299.

——— (1998): "Foundations of Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU," *Brookings Papers on Economic Activity*, 2, 297–346.

Granger, C. W. J. (1993): "What Are We Learning About the Long-Run?," *The Economic Journal*, 103, 307–317.

Hall, S. G. (1986): "An Application of the Granger and Engle Two-Step Estimation Procedure to United Kingdom Aggregate Wage Data," *Oxford Bulletin of Economics and Statistics*, 48, 229–239.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Jansson, M. (2004): "Stationarity Testing with Covariates," *Econometric Theory*, 20, 56–94.

Jansson, M., and M. J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–714.

Jeganathan, P. (1997): "On Asymptotic Inference in Linear Cointegrated Time Series Systems," *Econometric Theory*, 13, 692–745.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

Kempthorne, P. J. (1987): "Numerical Specification of Discrete Least Favorable Distributions," *SIAM Journal on Scientific and Statistical Computing*, 8, 171–184.

Kiefer, N., and T. J. Vogelsang (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988): "Production, Growth and Business Cycles," *Journal of Monetary Economics*, 21, 195–232.

Klein, L. R., and R. F. Kosobud (1961): "Some Econometrics of Growth: Great Ratios of Economics," *Quarterly Journal of Economics*, LXXV, 173–198.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159–178.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer, New York.

Magnus, J. R., and H. Neudecker (1988.): *Matrix Differential Calculus*. Wiley, New York.

Mehra, R. (2000): "Wage-Price Dynamics: Are They Consistent with Cost Push," *Federal Bank of Richmond Economic Quarterly*, 86, 27–43.

MÜLLER, U. K. (2005): "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series," *Journal of Econometrics*, 128, 195–213.

——— (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352.

——— (2008): "Efficient Tests under a Weak Convergence Assumption," *Working Paper, Princeton University*.

MÜLLER, U. K., AND M. W. WATSON (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016.

NABEYA, S. (1951): "Absolute Moments in 2-Dimensional Normal Distribution," *Annals of the Institute of Statistical Mathematics*, 3, 2–6.

NELSON, W. (1966): "Minimax Solution of Statistical Decision Problems by Iteration," *The Annals of Mathematical Statistics*, 37, 1643–1657.

NEWEY, W. K., AND K. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NYBLOM, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

NYBLOM, J., AND A. HARVEY (2000): "Tests of Common Stochastic Trends," *Econometric Theory*, 16, 176–199.

NYBLOM, J., AND T. MÄKELÄINEN (1983): "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association*, 78, 856–864.

OKSENDAL, B. (2000): *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 5 edn.

PARK, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–143.

PHILLIPS, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

SAIKKONEN, P. (1991): "Estimation and Testing of Cointegrating Regressions," *Econometric Theory*, 7, 1–27.

SAIKKONEN, P., AND R. LUUKONEN (1993): "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, 343–362.

SARGAN, J. D. (1964): "Wages and Prices in the UK: A Study in Econometric Methodology," in *Econometric Analysis for National Planning*, ed. by P. E. Hart, G. Mills, and J. K. Whitaker, pp. 25–26, London. Butterworths.

SRIANANTHAKUMAR, S., AND M. L. KING (2006): "A New Approximate Point Optimal Test of a Composite Null Hypothesis," *Journal of Econometrics*, 130, 101–122.

STAIGER, D., J. H. STOCK, AND M. W. WATSON (2001): "Prices, Wages, and the U.S. NAIRU in the 1990s," in *The Roaring Nineties*, ed. by A. B. Krueger, and R. Solow, pp. 3–60, New York. The Russell Sage Foundation.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

——— (1996): "Confidence Sets in Regressions with Highly Serially Correlated Regressors," *Working Paper, Harvard University.*

SUN, Y., P. C. B. PHILLIPS, AND S. JIN (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–794.

WRIGHT, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222.

Table 1: Power and Size Bounds for 5% Tests (q = 12)

A. $r = k = 1$

		$\|B\|$ = 7					$\|B\|$ = 14				
H_0	H_B	$\|R\|$ = 0	$\|R\|$ = 0.5		$\|R\|$ = 0.9		$\|R\|$ = 0	$\|R\|$ = 0.5		$\|R\|$ = 0.9	
			RB<0	*RB*>0	*RB*<0	*RB*>0		*RB*<0	*RB*>0	*RB*<0	*RB*>0
unr		0.36	0.36	0.36	0.36	0.36	0.64	0.65	0.65	0.66	0.66
stat		0.41	0.52	0.40	0.89	0.44	0.70	0.80	0.68	0.98	0.72
	unr	*0.06*	*0.14*	*0.05*	*0.65*	*0.07*	*0.07*	*0.17*	*0.05*	*0.47*	*0.06*
LTU		0.50	0.65	0.59	0.95	0.67	0.81	0.92	0.81	1.00	0.87
	unr	*0.13*	*0.28*	*0.21*	*0.79*	*0.32*	*0.15*	*0.42*	*0.17*	*0.49*	*0.20*
	stat	*0.08*	*0.09*	*0.15*	*0.07*	*0.20*	*0.11*	*0.13*	*0.13*	*0.03*	*0.14*
I(1)		0.50	0.65	0.65	0.95	0.95	0.82	0.91	0.91	1.00	1.00
	unr	*0.12*	*0.28*	*0.26*	*0.76*	*0.80*	*0.19*	*0.37*	*0.34*	*0.48*	*0.48*
	stat	*0.08*	*0.09*	*0.21*	*0.07*	*0.71*	*0.12*	*0.12*	*0.31*	*0.03*	*0.42*
	LTU	*0.04*	*0.04*	*0.06*	*0.04*	*0.50*	*0.04*	*0.03*	*0.16*	*0.00*	*0.25*

B. $r = 1$ and $k \geq 2$

		$\|\|B\|\|$ = 7							$\|\|B\|\|$ = 14						
H_0	H_B	$\|\|R\|\|$=0	$\|\|R\|\|$ = 0.5			$\|\|R\|\|$ = 0.9			$\|\|R\|\|$=0	$\|\|R\|\|$ = 0.5			$\|\|R\|\|$ = 0.9		
			ω=-1	ω=0	ω=1	ω=–1	ω=0	ω=1		ω=–1	ω=0	ω=1	ω=–1	ω=0	ω=1
unr		0.36	0.37	0.36	0.36	0.36	0.36	0.36	0.63	0.64	0.64	0.65	0.65	0.65	0.66
diag		0.36	0.37	0.37	0.36	0.36	0.42	0.36	0.63	0.64	0.64	0.65	0.65	0.74	0.66
	unr	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.06*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.08*	*0.04*
stat		0.40	0.48	0.42	0.39	0.83	0.59	0.44	0.68	0.77	0.71	0.67	0.97	0.84	0.71
	unr	*0.05*	*0.10*	*0.07*	*0.05*	*0.55*	*0.19*	*0.07*	*0.06*	*0.12*	*0.08*	*0.05*	*0.45*	*0.20*	*0.06*
	diag	*0.05*	*0.10*	*0.06*	*0.05*	*0.55*	*0.15*	*0.07*	*0.06*	*0.12*	*0.07*	*0.05*	*0.45*	*0.14*	*0.06*
LTU		0.46	0.57	0.49	0.53	0.92	0.70	0.64	0.74	0.84	0.77	0.76	1.00	0.85	0.85
	unr	*0.09*	*0.18*	*0.10*	*0.15*	*0.71*	*0.34*	*0.27*	*0.09*	*0.22*	*0.12*	*0.12*	*0.50*	*0.20*	*0.18*
	diag	*0.09*	*0.18*	*0.09*	*0.15*	*0.71*	*0.31*	*0.27*	*0.09*	*0.22*	*0.13*	*0.12*	*0.50*	*0.16*	*0.18*
	stat	*0.07*	*0.07*	*0.06*	*0.12*	*0.08*	*0.10*	*0.15*	*0.06*	*0.08*	*0.06*	*0.10*	*0.05*	*0.05*	*0.12*
I(1)		0.46	0.57	0.48	0.57	0.91	0.86	0.91	0.75	0.83	0.78	0.83	0.99	0.98	0.99
	unr	*0.09*	*0.18*	*0.11*	*0.17*	*0.71*	*0.70*	*0.71*	*0.10*	*0.23*	*0.15*	*0.21*	*0.50*	*0.46*	*0.49*
	diag	*0.09*	*0.18*	*0.11*	*0.17*	*0.71*	*0.53*	*0.71*	*0.10*	*0.23*	*0.13*	*0.21*	*0.50*	*0.37*	*0.49*
	stat	*0.07*	*0.07*	*0.07*	*0.17*	*0.08*	*0.23*	*0.67*	*0.08*	*0.08*	*0.08*	*0.18*	*0.05*	*0.20*	*0.45*
	LTU	*0.04*	*0.04*	*0.04*	*0.05*	*0.04*	*0.22*	*0.46*	*0.04*	*0.03*	*0.07*	*0.10*	*0.01*	*0.22*	*0.28*

Table 1 (Continued)

C. $r = 2$ and $k = 1$

		$\|\|B\|\| = 10$							$\|\|B\|\| = 20$						
H_0	H_B	$\|\|R\|\|=0$	$\|\|R\|\| = 0.5$			$\|\|R\|\| = 0.9$			$\|\|R\|\|=0$	$\|\|R\|\| = 0.5$			$\|\|R\|\| = 0.9$		
			ω=-1	ω=0	ω=1	ω=–1	ω=0	ω=1		ω=–1	ω=0	ω=1	ω=–1	ω=0	ω=1
unr		0.45	0.47	0.50	0.48	0.50	0.69	0.51	0.69	0.73	0.75	0.73	0.79	0.94	0.79
G		0.45	0.47	0.55	0.48	0.50	0.74	0.51	0.69	0.73	0.76	0.73	0.79	0.97	0.79
	unr	*0.04*	*0.04*	*0.06*	*0.04*	*0.04*	*0.06*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.05*	*0.04*
stat		0.50	0.61	0.61	0.51	0.94	0.74	0.57	0.72	0.81	0.78	0.74	0.98	0.97	0.82
	unr	*0.06*	*0.13*	*0.11*	*0.05*	*0.61*	*0.06*	*0.07*	*0.05*	*0.10*	*0.05*	*0.05*	*0.29*	*0.04*	*0.05*
	G	*0.06*	*0.13*	*0.07*	*0.05*	*0.61*	*0.04*	*0.07*	*0.05*	*0.10*	*0.05*	*0.05*	*0.29*	*0.02*	*0.05*
LTU		0.55	0.72	0.63	0.66	0.99	0.74	0.76	0.79	0.92	0.87	0.81	1.00	0.97	0.89
	unr	*0.10*	*0.25*	*0.11*	*0.17*	*0.70*	*0.06*	*0.29*	*0.12*	*0.26*	*0.13*	*0.09*	*0.32*	*0.05*	*0.12*
	G	*0.10*	*0.25*	*0.08*	*0.17*	*0.70*	*0.04*	*0.29*	*0.12*	*0.26*	*0.13*	*0.09*	*0.32*	*0.02*	*0.12*
	stat	*0.06*	*0.09*	*0.05*	*0.15*	*0.07*	*0.04*	*0.19*	*0.07*	*0.13*	*0.10*	*0.08*	*0.03*	*0.03*	*0.09*
I(1)		0.55	0.70	0.63	0.70	0.97	0.74	0.97	0.84	0.91	0.88	0.91	1.00	0.97	1.00
	unr	*0.10*	*0.22*	*0.11*	*0.24*	*0.69*	*0.06*	*0.69*	*0.16*	*0.26*	*0.15*	*0.29*	*0.32*	*0.05*	*0.32*
	G	*0.10*	*0.22*	*0.08*	*0.24*	*0.69*	*0.04*	*0.69*	*0.16*	*0.26*	*0.14*	*0.29*	*0.32*	*0.02*	*0.32*
	stat	*0.06*	*0.07*	*0.05*	*0.19*	*0.05*	*0.04*	*0.65*	*0.12*	*0.14*	*0.11*	*0.26*	*0.03*	*0.03*	*0.31*
	LTU	*0.04*	*0.04*	*0.04*	*0.06*	*0.02*	*0.04*	*0.45*	*0.10*	*0.03*	*0.04*	*0.18*	*0.00*	*0.02*	*0.21*

Notes: Non-italicized numerical entries are upper bounds on power for 5% tests for the null hypothesis restricting the trend process as shown in the first column. The italicized entries are lower bounds on size for auxiliary trend processes given in columns for 5% level tests with power greater than the power bound less 3 percentage points. Abbreviations for the trend models in columns 1 and 2 are: (i) "*unr*" is the unrestricted model given in (3), (ii) "*G*" is the *G*-model in (6), (iii) "*diag*" is the diagonal *G*-model in (8), (iv) "*stat*" is the stationary *G* model given in (9), (v) "*LTU*" is the local-to-unity model given in (11), and (vi) "*I(1)*" is I(1) model given in (10). In panels B and C, $\omega = \text{tr}(R'B) / (\|R\| \cdot \|B\|)$. Results are based on 20,000 Monte Carlo replications.

Table 2: Comparison of KLIC Minimized and Approximate Least Upper Power Bounds
($r = k = 1$, $q = 12$)

		|B| = 7					|B| = 14				
H_0	Bound	|R| = 0	|R| = 0.5		|R| = 0.9		|R| = 0	|R| = 0.5		|R| = 0.9	
			RB<0	*RB*>0	*RB*<0	*RB*>0		*RB*<0	*RB*>0	*RB*<0	*RB*>0
LTU	KLIC	0.51	0.66	0.59	0.95	0.66	0.81	0.93	0.80	1.00	0.86
	LUB	0.50	0.66	0.58	0.93	0.65	0.78	0.88	0.78	1.00	0.82
I(1)	KLIC	0.51	0.65	0.65	0.95	0.95	0.82	0.92	0.91	1.00	1.00
	LUB	0.50	0.65	0.65	0.94	0.94	0.81	0.90	0.90	1.00	1.00

Notes: The entries labeled "KLIC" are computed using the KLIC minimization discussed in Section 5.3.2. The entries labeled "LUB" are computed using the approximate least upper algorithm discussed in Section 5.3.1, and are by construction no more than 2.5 percentage points above the actual least upper bound. Results are based on 20,000 Monte Carlo replications.

Table 3: 1%, 5%, and 10% Critical Values for the *LFST* Statistic

q	***r***				
	1	**2**	**3**	**4**	**5**
6	5.25 3.62 3.08	6.76 5.16 4.39	7.25 6.09 5.43	7.14 6.46 6.02	6.54 6.33 6.16
7	4.33 3.08 2.68	5.52 4.20 3.63	6.05 4.95 4.37	6.15 5.35 4.92	5.89 5.48 5.20
8	3.68 2.73 2.39	4.65 3.54 3.08	5.17 4.16 3.68	5.29 4.55 4.12	5.26 4.73 4.42
9	3.21 2.46 2.18	4.02 3.09 2.73	4.46 3.58 3.19	4.63 3.93 3.56	4.66 4.12 3.83
10	2.86 2.25 2.02	3.56 2.79 2.48	3.94 3.17 2.84	4.10 3.47 3.15	4.18 3.66 3.38
11	2.62 2.10 1.90	3.16 2.54 2.29	3.53 2.87 2.59	3.71 3.12 2.84	3.78 3.30 3.03
12	2.46 1.98 1.81	2.89 2.35 2.13	3.18 2.64 2.39	3.38 2.84 2.60	3.48 3.02 2.78
13	2.29 1.88 1.73	2.68 2.21 2.01	2.92 2.44 2.23	3.13 2.63 2.42	3.20 2.77 2.57
14	2.16 1.80 1.67	2.50 2.09 1.92	2.74 2.31 2.11	2.91 2.47 2.27	2.97 2.59 2.40
15	2.07 1.74 1.61	2.36 1.99 1.84	2.56 2.18 2.01	2.69 2.32 2.15	2.80 2.44 2.27
16	1.97 1.67 1.56	2.24 1.91 1.77	2.44 2.08 1.92	2.55 2.21 2.05	2.64 2.30 2.15
17	1.89 1.62 1.52	2.15 1.84 1.71	2.32 1.99 1.85	2.43 2.11 1.96	2.50 2.20 2.05
18	1.82 1.58 1.49	2.07 1.78 1.66	2.21 1.92 1.79	2.32 2.02 1.89	2.39 2.10 1.98

Notes: The table shows asymptotic critical for the $LFST(b)$ statistic computed using $b = 10/\sqrt{r}$, where $LFST(b) = \det(Y'Y)/\det(Y'(I + b^2D)^{-1}Y)$, with $D = \text{diag}(d_1,\ldots,d_k)$ and $d_i = (i\pi)^{-2}$. Results are based on 50,000 Monte Carlo replications.

Table 4: Power of *Y*-only 5% Tests ($q = 12$)

A. $r = 1$

	$\|B\| = 7$	$\|B\| = 14$
LFST(10)	0.36	0.63
Y-Only Power Envelope	0.36	0.64

B. $r = 2$ and $k = 1$

$\|\|B\|\| = 10$							$\|\|B\|\| = 20$						
Power of *LFST*$(10/\sqrt{2})$													
0.39							0.58						
Y-Only Power Envelope													
$\|\|R\|\|=0$	$\|\|R\|\| = 0.5$			$\|\|R\|\| = 0.9$			$\|\|R\|\|=0$	$\|\|R\|\| = 0.5$			$\|\|R\|\| = 0.9$		
	ω=-1	ω=0	ω=1	ω=–1	ω=0	ω=1		ω=–1	ω=0	ω=1	ω=–1	ω=0	ω=1
0.41	0.41	0.44	0.41	0.41	0.55	0.41	0.68	0.64	0.68	0.64	0.64	0.86	0.64

Notes: The power envelope is computed using the test statistic *LFST*($|B|$) in panel A and by $\xi(B)$ in panel B. In panel B , $\omega = \text{tr}(R'B) / (\|R\| \cdot \|B\|)$. Results are based on 20,000 Monte Carlo replications.

Table 5: Cointegrating Coefficients and Tests

$$w_t = \beta_0 + \beta_p p_t + \beta_y y_t + \beta_n n_t$$

	Non-Farm Business	Non-Financial Corporations
a. DOLS Coefficients (SEs)		
P	1.046 (0.020)	1.003 (0.018)
Y	1.028 (0.037)	0.852 (0.023)
N	−1.227 (0.106)	−0.723 (0.057)
b. Joint Test $\beta_p = 1$, $\beta_y = 1$, $\beta_n = -1$		
DOLS Wald Statistic (*p*-value)	35.82 (<0.001)	45.41 (<0.001)
LFST *p*-value	0.004	0.282

Notes: The top panel shows DOLS estimated coefficients and standard errors. The bottom panel shows the DOLS Wald statistic (which is distributed χ^2_3 under the null) and associated *p*-value, and the LFST *p*-value. The DOLS estimator uses 6 leads and lags of $\Delta^2 p_t$, Δy_t, and Δn_t, and the DOLS standard errors are computed using Newey-West HAC estimators with 6 lags.

Figure 1: $w_t - p_t - y_t - n_t$
Non-Farm Business (thin black) and Non-Financial Corporations (thick blue)

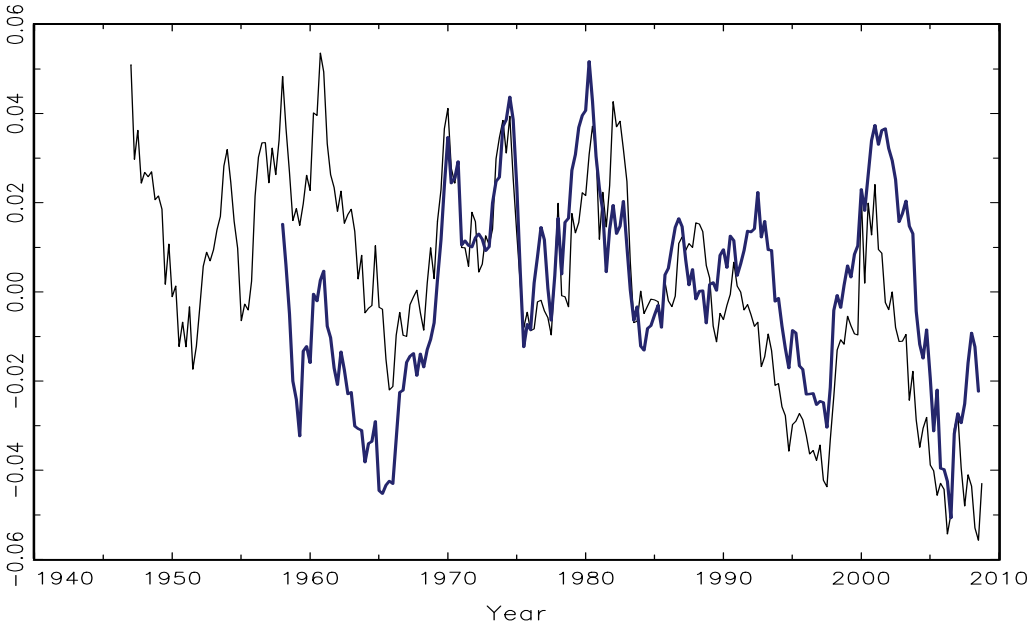

Figure 2: 95% Confidence Sets for Cointegrating Coefficients, Non-Financial Corporations
LFST (thin black) and DOLS (thick blue)