IOWA STATE UNIVERSITY

Fixed Bandwidth Asymptotics in Single Equation Models of Cointegration with an Application to Money Demand

Helle Bunzel

August 2003

Working Paper # 03024

Department of Economics
Working Papers Series

Ames, Iowa 50011

Fixed Bandwidth Asymptotics in Single Equation Models of Cointegration with an Application to Money Demand[*]

Helle Bunzel[†]

October, 2004

Abstract

This paper provides a new approach to testing cointegration parameters in a single-equation cointegration environment. The novelty is in improving over the well-known heteroscedasticity and autocorrelation consistent (HAC) robust standard errors using fixed bandwidth (fixed-b) asymptotic theory and adapting it to the cointegration environment. It is shown that the standard tests still have asymptotic distributions free of serial correlation nuisance parameters regardless of the bandwidth or kernel used, even if the regressors in the cointegration relationship are endogenous. Using asymptotic power and finite sample size simulation experiments, a specific kernel and bandwidth choice is recommended. Finite sample simulations comparing the size and power of the test using the fixed-b asymptotics to some of the currently popular tests are performed. These simulations confirm that the well-known size distortion of the standard tests can be greatly reduced. Finally, the newly developed test is employed to investigate the standard money-demand relationship for US data.

1　Introduction

In this paper, we consider univariate models where time series data are generated by unit root processes and the variables captured by these data series are cointegrated. Among the many applications of such models are estimation of money demand, testing of the Purchasing Power Parity hypothesis, and examination of the expectations hypothesis governing the term structure of interest rates.[1] It is well known that heteroscedasticity and serial correlation are almost always present in the aforementioned type of data, and unless properly dealt with, impair the ability of the researcher to conduct proper statistical inference.[2]

The standard single-equation approach used to deal with heteroscedasticity and serial correlation is to estimate the correlation structure of the error terms using non-parametric heteroscedasticity and autocorrelation consistent (HAC) estimators.[3] These estimators furnish consistent estimates of the correlation structure, allowing inference on the cointegrating vector to be carried out using conventional tests. Inference conducted in this manner leads to pivotal tests, and is robust to heteroscedasticity and serial correlation of unknown form. Even though tests that use HAC estimators are valid *asymptotically*, they typically display substantial size distortions.[4] Assuming that size distortions are a problem in non-stationary models as well, it is apparent that there may be significant benefits from improving upon this procedure.

Recent efforts have been made to improve upon the HAC approach in standard (stationary) regression models. The first paper in this literature was Kiefer, Vogelsang and Bunzel (2000) where a new test based on the Bartlett kernel with bandwidth equal to sample size was developed. Continuing this line of research, Bunzel, Kiefer and Vogelsang (2001) extended the theory to non-linear, stationary regression models and Kiefer and Vogelsang (2002) developed the new fixed bandwidth (fixed-b) asymptotic theory. In the case of a cointegration relationship with exogenous regressors, applying fixed-b asymptotics would have been a straightforward extension of the theory in the standard regression model. However, when the regressors are allowed to be endogenous, the task is non-trivial.

2

The principle behind the fixed-b theory is to let $b = M/T$ where T is the sample size and M is the truncation lag or bandwidth used in the HAC estimator. The standard assumptions would require that $b \to 0$, but fixed-b asymptotics instead assumes that the truncation lag is a fixed proportion of the sample, i.e., that b is fixed. This approach has several advantages. First, it improves the asymptotic approximation, resulting in reduced size distortions. Second it provides an asymptotic distribution which depends on the bandwidth and kernel, thus providing us with better tools for choosing these parameters. We show that tests based on the Daniell kernel with $b = 0.2$ provides excellent finite sample size while sacrificing the least possible power.

To demonstrate the properties of the selected test statistics, we carry out a set of finite sample simulations, which compare the size and power of the test using the fixed-b asymptotics to some of the currently used tests in the literature. These comparisons document that the new asymptotic theory can provide us with tests that have vastly improved size properties even in small samples, although the size improvements come at the cost of some power.

In the empirical application we use the recommended test statistic to investigate the long-run money-demand relationship for the US. A thorough examination was provided by Stock and Watson (1993). Their results were subsequently re-examined and rejected by Ball (2001), who used simulations to attempt to counter the usual size inflation. Our results confirm that Ball's simulation results were fairly accurate, and that the results obtained by Stock and Watson can be rejected. Thus, while Ball was able to reach a conclusion only after extensive simulations which required parametric modelling and estimation of the error terms, the test we recommend can provide the same conclusion in one step using a standard software package.

The rest of the paper is organized as follows. In Section 2, we describe the model and the basic assumptions. The asymptotic distributions are derived and described in Section 3, in Section 4 we perform simulations to determining which kernel and bandwidth should

3

be used and in Section 5 we compare the performance of the chosen test statistics to those commonly used in the literature. In Section 6 we re-consider the money-demand estimation performed in Ball (2001) using the new test and Section 7 concludes.

2   Preliminaries

Consider the following model containing a single cointegrating relationship as well as some deterministic variables:

$$y_t = f(t)' \alpha + X_t' \beta + u_{1,t}, \quad t = 1, ..., T, \tag{1}$$

$$X_t = X_{t-1} + u_{2,t},$$

where $f(t)$ denotes a $(k_1 \times 1)$ vector of trend functions, X_t is a $(\mathbf{k} \times 1)$ vector of regressors, and α and β are $(k_1 \times 1)$ and $(\mathbf{k} \times 1)$ vectors of parameters respectively. Let $'$ denote the transpose, except when it is used in conjunction with the kernel function, where it will denote the derivative. The following assumptions will be maintained throughout the paper. Conditional on X_t, $u_{1,t}$ is a scalar, mean zero random process. The sequence $\{u_t\} = \{(u_{1,t}, u_{2,t}')'\}$ does not contain unit roots, but may exhibit serial correlation or heteroscedasticity.[5]

At times, it will be useful to stack the first equation in (1) and rewrite it as

$$y = \mathbf{f}(T) \alpha + X\beta + u_1. \tag{2}$$

Here $\mathbf{f}(T)$ is the $(T \times k_1)$ stacked vector of trend functions, and X is the $(T \times \mathbf{k})$ matrix of regressors. The following notation is required before we state the main assumptions of the paper. Denote $S_{ct} = \sum_{j=1}^{t} u_{c,j}$, $c = 1, 2$, $S_t = \sum_{j=1}^{t} u_j$, $\Gamma(j) = E\left(u_t u_{t+j}'\right)$, $\Gamma_{22}(j) = E\left(u_{2,t} u_{2,t+j}'\right)$, let $w_j(r)$ be a j-vector of independent Wiener processes, and $[rT]$ the integer part of rT, where $r \in [0, 1]$. " $\Rightarrow$ " is used to denote weak convergence.

The first assumption, which follows Vogelsang (1998), is made to rule out ill-behaved

4

trend functions, and to provide some useful notation for deriving and stating the asymptotic distributions.

Assumption 1 : There exists a $(k_1 \times k_1)$ diagonal matrix τ_T and a vector of functions F, such that $\tau_T f(t) = F\left(\frac{t}{T}\right) + o(1)$, $\int_0^1 F_i(s)\,ds < \infty$, $i = 1, ..., k_1$, and $\det\left[\int_0^1 F(s)\,F(s)'\,ds\right] > 0$. In addition, $f(t)$ includes a constant term.

Assumption 1 can be relaxed, but as it stands, is sufficiently general to cover most commonly used models. For later use, let $\mathbf{F}(T)$ be the matrix of the stacked $F(t/T)$ functions.

The next assumption provides us with the necessary invariance principles, and ensures that we can estimate (1) consistently, even when the regressors are endogenous.

Assumption 2 : $\{u_t\}_{t=1}^{\infty}$ satisfies the following conditions

(a) $E(u_t) = 0$ for all t.

(b) $E\left(T^{-1} S_T S_T'\right) \to \Omega$, a positive definite matrix, as $T \to \infty$.

(c) $\sup_t\left(E\|u_t\|^{\kappa}\right) < \infty$ for some κ where $2 < \kappa < \infty$.

(d) $E\left(T^{-1}\left(S_{j+T} - S_j\right)\left(S_{j+T} - S_j\right)'\right) \to \Omega$, as $\min(j, T) \to \infty$.

(e) $\{u_t\}_{t=1}^{\infty}$ is $\alpha-$ mixing with coefficient $-\kappa/(\kappa - 2)$.

(f) $g_{uu}(\lambda) \geq \vartheta I_{\mathbf{k}+1}$, where $\vartheta > 0$, $\lambda \in [0; \pi]$ and $g_{uu}(\lambda)$ is the spectral density matrix of u.[6]

(g) $\sum_{j=-\infty}^{\infty}\|\Gamma(j)\| < \infty$

(h) $\sum_{m_1,m_2=-\infty}^{\infty} |kum_{ijkl}(m_1, m_2)| < \infty$, where $kum_{ijkl}(m_1, m_2)$ denotes the fourth order cumulant of u_t.

Assumption 2 (a)-(e) has been used extensively in the literature on non-parametric covariance matrix estimation to ensure that the relevant multivariate invariance principles hold. These conditions are sufficient to provide the asymptotic distribution of the OLS estimates of (1) if the regressors are exogenous. Assumption 2 (f)-(h) is made to allow us

to deal with endogenous regressors in the manner suggested by Saikkonen (1991), Phillips and Loretan (1991), Stock and Watson (1993) and Wooldridge (1991). A direct implication (see Saikkonen (1991)) is that we can write $u_{1,t}$ as

$$u_{1,t} = \sum_{j=-\infty}^{\infty} \gamma_j u_{2,t-j} + v_t, \tag{3}$$

where $\sum_{j=-\infty}^{\infty} \|\gamma_j\| < \infty$ and v_t is a stationary process such that $E\left(u_{2,t} v'_{t+l}\right) = E\left(\left(X_t - X_{t-1}\right) v'_{t+l}\right) = 0$, $l = 0, \pm 1, \pm 2, ...$ Following standard procedure, we can thus estimate the model using Dynamic Ordinary Least Squares (henceforth DOLS), i.e., we estimate

$$y_t = f(t)' \alpha + X_t' \beta + \sum_{s=-p}^{p} \Delta X'_{t-s} \gamma_s + \dot{v}_t, \quad t = p+1, ..., T-p, \tag{4}$$

where $\Delta X_t = X_t - X_{t-1}$, and $\dot{v}_t = v_t + \sum_{|j|>p} \gamma_j u_{2t-j}$. We are now ready to make the third and final assumption:

Assumption 3 : Let $p \to \infty$ such that $p^3/T \to 0$ and $T^{\frac{1}{2}} \sum_{|j|>p} \|\gamma_j\| \to 0$.

Saikkonen (1991) shows that if Assumption 3 holds, then (4) is asymptotically equivalent to[7]

$$y_t = f(t)' \alpha + X_t' \beta + v_t,$$

and under Assumptions 1-3, the asymptotic distributions of the least squares estimates of α and β are well known.[8]

To conduct inference on β using the DOLS estimates, the standard procedure is to estimate the asymptotic covariance matrix using non-parametric covariance matrix (HAC) estimates and then form Wald or $t-$type tests. Specifically, if we denote the Cholesky composition of Ω by $\Omega^{\frac{1}{2}} = \begin{bmatrix} \sigma & \sigma_{12} \\ \sigma_{21} & \Lambda \end{bmatrix}$, an estimate of σ is required. The HAC estimators of σ take the general form

$$\hat{\sigma}^2 = \sum_{j=-(T-p-1)}^{T-p-1} k\left(j/M\right) \hat{\Gamma}_j, \text{ where } \hat{\Gamma}_j = \begin{cases} \frac{1}{N} \sum_{t=j+1}^{T-p} \hat{v}_t \hat{v}_{t-j}, \text{ for } j \geq 0 \\ \frac{1}{N} \sum_{t=-j+1}^{T-p} \hat{v}_{t+j} \hat{v}_t, \text{ for } j < 0 \end{cases}. \tag{5}$$

Here $N = T - (2p + 1)$, $\hat{v}_t$ are the residuals from (4), M is called the bandwidth or the truncation lag, and $k(x)$ is a kernel function satisfying $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, $k(x)$ continuous at $x = 0$ and $\int_0^1 k^2(x) dx < \infty$. For $\hat{\sigma}^2$ to be consistent, it is necessary that $M \to \infty$ and $M/T \to 0$ as $T \to \infty$. Since the standard asymptotic tests are based on consistent estimates of σ^2, the choices of kernel and bandwidth do not enter the asymptotic distribution of the test statistic. While this may be convenient, it has been well documented in the literature that these choices affect the finite sample behavior of $\hat{\sigma}^2$ and hence the finite sample performance of the test statistic. In the next section, we will apply the fixed bandwidth (fixed-b henceforth) asymptotic theory introduced by Kiefer and Vogelsang (2002). Using this asymptotic theory, the asymptotic distribution of $\hat{\sigma}^2$ depends directly on the choice of bandwidth and kernel. This dependence can help provide some guidance to the applied researcher regarding the choices of kernel and bandwidth through local asymptotic power comparisons.

3 Fixed-b asymptotics.

Although the standard methods of asymptotic testing make use of a consistent estimator of σ, this is not required to carry out valid testing. In its place, any stochastic variable with an asymptotic distribution proportional to σ can be utilized to obtain a pivotal statistic. Thus, the assumption that $M/T \to 0$, which is required for consistency, can be relaxed. Following Kiefer and Vogelsang (2002), we instead assume that M is directly proportional to T, such that $M = [bT]$ and develop this asymptotic theory for the cointegration model. The limiting distribution of $\hat{\sigma}^2$ will depend on the specific bandwidth (now fully determined by the parameter b) and kernel used to construct the estimator. This dependence improves the asymptotic approximation, which shows up in the simulations in the form of smaller size distortions than tests where $b \to 0$.

To proceed we provide the following definition, which describes two different types of kernels.

Definition A kernel is labelled Type 1 if $k(x)$ is twice continuously differentiable everywhere and as a Type 2 kernel if $k(x)$ is continuous, $k(x) = 0$ for $|x| \geq 1$ and $k(x)$ is twice continuously differentiable everywhere except at $|x| = 1$.

In addition, we will consider the Bartlett kernel separately. The following lemma provides the asymptotic distribution of $\hat{\sigma}^2$ under fixed-b asymptotics and for various choices of kernels. To state the asymptotic distributions, we define

$$
\begin{aligned}
V(r) &= w_1(r) - \int_0^r F(s)' \, ds \left(\int_0^1 F^X(s) F^X(s)' \, ds \right)^{-1} \int_0^1 F^X(s) \, dw_1(s) \\
&\quad - \int_0^r w_{\mathbf{k}}(s)' \, ds \left(\int_0^1 w_{\mathbf{k}}^F(s) w_{\mathbf{k}}^F(s)' \, ds \right)^{-1} \int_0^1 w_{\mathbf{k}}^F(s) \, dw_1,
\end{aligned}
$$

where $w_{\mathbf{k}}^F(s)$ is defined as the residual from the projection of $w_{\mathbf{k}}(s)$ on the subspace generated by $F(s)$ in the Hilbert space of square integrable functions on [0,1] with the inner product $(f, g) = \int_0^1 fg$. Correspondingly, $F(s)^X$ is the residual from the projection of $F(s)$ onto the space generated by $w_{\mathbf{k}}(s)$.

Lemma 1 *If k is Type 1,*

$\hat{\sigma}^2 \Rightarrow -\sigma^2 \int_0^1 \int_0^1 k^{*\prime\prime}(r-s) V(r) V(s) \, dr ds$.

If k is Type 2

$\hat{\sigma}^2 \Rightarrow \sigma^2 \left(\int \int_{|r-s|<b} -k^{*\prime\prime}(r-s) V(r) V(s) \, dr ds + 2k_-^{*\prime}(b) \int_0^{1-b} V(r+b) V(r) \, dr \right)$,

where $k^*(x) = k\left(\frac{x}{b}\right)$, and $k_-^{*\prime}(b)$ is the derivative of $k^*(x)$ from below at b.

If k is the Bartlett kernel,

$\hat{\sigma}^2 \Rightarrow \sigma^2 \cdot \frac{2}{b} \left[\int_0^1 V(r)^2 \, dr - \int_0^{1-b} V(r) V(r+b) \, dr \right]$.

The proof of Lemma 1 follows that of Kiefer and Vogelsang (2002), but with the added complication that endogenous regressors are present, and therefore some additional work is required to determine the asymptotic distribution of the partial sums of the residuals. The asymptotic distribution of $\hat{\sigma}^2$ is proportional to σ^2 and depends on the bandwidth and kernel as expected.

8

Using Lemma 1, hypotheses of the form $H_0 : R\hat{\beta} = \beta_0$, can be tested using the standard Wald test.[9] In what follows R is a non-stochastic restriction matrix of dimension $q \times \mathbf{k}$ and rank q. The Wald test for H_0 is defined as

$$
\begin{aligned}
W &= T\left(R\hat{\beta} - \beta_0\right)' \left[\hat{\sigma}^2 R Q_{fX}^{-1} R'\right]^{-1} \left(R\hat{\beta} - \beta_0\right), \text{ where} \\
Q_{fX} &= \left(\frac{1}{T} \begin{bmatrix} \mathbf{f}(T)' \mathbf{f}(T) & \mathbf{f}(T)' X \\ X'\mathbf{f}(T) & X'X \end{bmatrix}\right).
\end{aligned}
$$

The corresponding one-dimensional $t-$test can be obtained in the usual manner. Theorem 2 below states the asymptotic distribution of W under fixed-b asymptotics.

Theorem 2 *Suppose Assumptions 1, 2, and 3 hold. Then, under H_0,*

$$
W \Rightarrow \begin{cases} \left(-\int_0^1 \int_0^1 k^{*\prime\prime}(r-s) V^F(r) V^F(s)\, drds\right)^{-1} \tilde{W}_X & \text{if } k \text{ is type 1} \\ U^F(k,b)^{-1} \tilde{W}_X & \text{if } k \text{ is type 2} \\ \left(\frac{2}{b}\left[\int_0^1 V^F(r)^2\, dr - \int_0^{1-b} V^F(r) V^F(r+b)\, dr\right]\right)^{-1} \tilde{W}_X & \text{if } k \text{ is Bartlett} \end{cases}
$$

where

$$
\tilde{W}_X = \int_0^1 \hat{w}_q^F(s)'\, d\hat{w}_1^F(s) \left(\int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)'\, ds\right)^{-1} \int_0^1 \hat{w}_q^F(s)\, d\hat{w}_1^F(s),
$$

$$
V^F(r) = \hat{w}_1^F(r) - \int_0^r \hat{w}_q^F(s)'\, ds \left(\int_0^1 \hat{w}_q^F(s) \hat{w}_q^F(s)'\, ds\right)^{-1} \int_0^1 \hat{w}_q^F(s)\, d\hat{w}_1^F(s),
$$

and

$$
U^F(k,b) = \int \int_{|r-s|<b} -k^{*\prime\prime}(r-s) V^F(r) V^F(s)\, drds + 2k_-^{*\prime}(b) \int_0^{1-b} V^F(r+b) V^F(r)\, dr.
$$

This theorem demonstrates that it is possible to obtain pivotal test statistics with the fixed-b assumption. The asymptotic distribution of the Wald test depends on the kernel and bandwidth, and through $\hat{w}_q^F(s)$, it also depends on the number of restrictions being tested, the number of regressors in the model, and the trends included, where standard $b \to 0$ asymptotics would have resulted in an asymptotic χ^2 distribution. The fact that the limiting distribution of the test statistic depends upon the choice of bandwidth and kernel allows us to carry out asymptotic simulations to determine how the bandwidth affects performance of the test-statistic, something which isn't possible under the standard χ^2

9

distribution. These simulations will guide the choices of kernel and bandwidth and are implemented in Section 4.

4 Choice of Kernel and Bandwidth.

In this section we use simulation experiments to analyze how the performance of the test statistic varies with the choice of bandwidth and kernel. Ideally, this analysis should be performed by examining the higher order expansions for the test statistic, but at this time, these have not been developed for fixed-b asymptotics. The only theory comparing the fixed-b approach to the standard asymptotics is developed in Jansson (2002), where it is shown that the fixed-b asymptotics provides a smaller error in rejection probability than the standard approach in a simple location model with normally distributed errors. Lacking the required theory at this point in time, we therefore perform the analysis using simulations. This exercise will culminate in the recommendation of specific choices of both kernel and bandwidth. Initially, we examine how local asymptotic power of the test statistic in a simple model varies depending on the choice of kernel and bandwidth. In addition, we consider finite sample coverage probabilities, since there is likely to be the usual trade-off between size and power.

4.1 Local Asymptotic Power

Local asymptotic power will be examined in the following simple model:

$$
\begin{aligned}
y_t &= \alpha + \beta x_t + u_{1,t}, \\
x_t &= x_{t-1} + u_{2,t}, \ t = 1, .., T,
\end{aligned}
\tag{6}
$$

where x is exogenous. The first set of simulations determines the asymptotic power of the $t-$test for the hypothesis $H_0 : \beta = 0$. For the asymptotic power analysis, the local alternative is given by $H_A : \beta = T^{-1}c$. All tests will be carried out at the 5% level.

10

To obtain local asymptotic power, we need the distribution of the $t-$statistic both under the null and under the alternative. The distribution under H_0 follows directly from Theorem 2. Under H_A the asymptotic distribution of the numerator of the $t-$statistic is given by:

$$T\hat{\beta} \Rightarrow c + \sigma\Lambda^{-1}\left(\int_0^1 w^F(s)\,w^F(s)'\,ds\right)^{-1}\left(\int_0^1 w^F(s)\,dw_1(s)\right).$$

Inserting this expression as well as the asymptotic distributions of $\hat{\sigma}^2$ from Lemma 1, the asymptotic distribution of t under H_A is

$$t \stackrel{H_A}{\Rightarrow} \begin{cases} \dfrac{\Lambda c/\sigma + \left(\int_0^1 w^F(s)^2 ds\right)^{-1}\int_0^1 w^F(s)dw_1(s)}{\sqrt{-\int_0^1\int_0^1 k''(r-s)V^F(r)V^F(s)drds\left(\int_0^1 w_k^F(s)w_k^F(s)'ds\right)^{-1}}} & \text{if } k \text{ is type 1} \\[2em] \dfrac{\Lambda c/\sigma + \left(\int_0^1 w^F(s)^2 ds\right)^{-1}\int_0^1 w^F(s)dw_1(s)}{\sqrt{U^F(k,b)\left(\int_0^1 w_k^F(s)w_k^F(s)'ds\right)^{-1}}} & \text{if } k \text{ is type 2} \\[2em] \dfrac{\Lambda c/\sigma + \left(\int_0^1 w^F(s)^2 ds\right)^{-1}\int_0^1 w^F(s)dw_1(s)}{\sqrt{\left(\frac{2}{b}\left[\int_0^1 V^F(r)^2 dr - \int_0^{1-b} V^F(r)V^F(r+b)dr\right]\right)\left(\int_0^1 w_k^F(s)w_k^F(s)'ds\right)^{-1}}} & \text{if } k \text{ is Bartlett} \end{cases} \tag{7}$$

All the simulations in this section were performed using sums of $N(0,1)$ i.i.d. random variables to approximate the Wiener processes in the distributions. In each case, the programming was performed using GAUSS and $50,000$ replications were used. The integrals were computed as averages over 1500 simulated observation points. The distance from the null hypothesis, c, was allowed to vary from 2 to 14, and the power was calculated for $b = 0.02, 0.04, ..., 1$. Figure 1 depicts the asymptotic power of the Daniell kernel as a function of b for different values of c. From this figure, it is immediately clear that smaller b provides better power. The corresponding figures for the Quadratic Spectral (QS), the Parzen, the Bartlett and the Bohman kernels differ only in the magnitude of the power loss when b increases, and are hence omitted.

In Table 1, we report the power of all five kernels when $b = 0.02$. From these numbers, it is immediately clear that when $b = 0.02$ the power across kernels is virtually identical. Intuitively, this is not surprising: The $b = 0.02$ case is very similar to the asymptotic results when $b \to 0$, where asymptotic power is identical across kernels and bandwidths. From Figure 1 and Table 1 it is clear that were we to recommend a kernel and a bandwidth based

11

on local asymptotic power alone, we would recommend as small a bandwidth as the data allows and any convenient kernel.

4.2   Finite Sample Size

Next, we consider how the choice of bandwidth and kernel affects finite sample size. Again the simulations are based on (6). The errors are generated according to $u_{1t} = \rho u_{1t-1} + e_t + \lambda e_{t-1}$, where $\{e_t\}$ and $\{u_{1t}\}$ are $i.i.d.$ $N(0,1)$, with $\rho = 0, 0.8, 0.9$ and for $\lambda = -0.8, -0.4, 0, 0.4, 0.8$. We report results for $T = 50$. Figure 2 depicts the size of the test using the Daniell kernel as a function of the bandwidth, where each curve corresponds to a different DGP (for reasons of visual clarity only a selection of processes are presented). The graphs for the other kernels are qualitatively similar and therefore not reported. From Figure 2, it is clear that the size improves as b increases.[10] This implies that we will face a trade-off between size and power, with the size distortion minimized at $b = 1$, but the highest power at $b = 0.02$.

Before we proceed to recommend a bandwidth choice, we again consider the question of which kernel to use. Table 2 provides the size for the five kernels when $b = 1$ across various data generating processes. From here it is apparent that the Bartlett kernel performs significant worse than the other four kernels while the Daniell kernel provides marginally less size distortion than the rest. As such we should recommend the Daniell kernel with $b = 1$ if the choice were to be made solely based on finite sample size considerations. Unfortunately, the power loss resulting from using $b = 1$ as opposed to $b = 0.02$ is considerable, as can be seen from Figure 3, which depicts the local asymptotic power of the test when we use the Daniell kernel with bandwidths $b = 0.02$ and $b = 1$ respectively. While we cannot completely avoid weighing power and size considerations when choosing a bandwidth and a kernel, we do not actually have to go to the extreme of comparing $b = 0.02$ and $b = 1$. Returning our attention to Figure 2, we see that while size does improve as b increases, the curve is very flat for most values of b. Specifically, the increased size distortion when

moving from $b = 1$ to $b = 0.2$ is minimal. Similarly the QS kernel sees little differences in size between $b = 0.2$ and $b = 1$, while the Parzen and Bohman kernels have size curves that remain flat for $b \geq 0.4$. From Table 3, we note that among the QS-0.2, Dan-0.2, Parzen-0.4 and Bohman-0.4, the Dan-0.2 still has marginally better size than the other tests.[11] Figure 4 in turn demonstrates that these four test have basically the same asymptotic power. Because of the marginally better size properties, we choose the Dan-0.2 test statistic, but it is clear that the four statistics we are considering have virtually identical properties. We are now ready to provide the critical values for the Dan-0.2 test.

4.3 Critical Values

All the critical values in this paper are calculated using sums of $N(0,1)$ i.i.d. random variables to approximate the Wiener processes in the distributions. In each case, $50,000$ replications were used, and the integrals were computed as averages over $1,000$ equally spaced points. Table 4 reports critical values for the Daniell kernel with $b = 0.2$ for $f(t) = \alpha$ and $f(t) = \alpha_0 + \alpha_1 t$. The critical values correspond to the Wald version of the tests, and as usual critical values for the t version of the test (applicable when $q = 1$) are calculated by taking the square root. The table provided here allows for up to six regressors.

5 Monte-Carlo Comparisons

In this section, we will compare the performance of the Dan-0.2 test with some of the standard HAC tests currently employed. We will do this by finite sample simulations, where we compare the size and power of the Dan-0.2 with two standard tests. These are a) the test statistic using the HAC estimator recommended by Andrews (1991), which utilizes the quadratic spectral kernel and an automatic data-dependent bandwidth selection procedure, b) the same HAC estimator, but pre-whitened based on a AR(1) model, as suggested by Andrews and Monahan (1992). These tests are labelled HAC and HAC-pw respectively.

The simulations are carried out for the model

$$y_t = \alpha + \beta x_t + u_{1,t}, \tag{8}$$

$$x_t = x_{t-1} + u_{2,t}, \ t = 1, .., T,$$

where u_t is generated according to $\Phi(L) u_t = e_t$, where $\{e_t\}$ is $i.i.d.$ $N(0, \Sigma_e)$. As a benchmark, we report results for the case where the errors are independent and $T = 50$. In addition we report results for $T = 50, 100$ in a case where there is endogeneity present. The data is generated with $\alpha = \beta = 0$, since the results are invariant to this normalization. The power and size are given for the hypothesis $H_0 : \beta = 0$, which is performed as a two-sided test with a nominal level of 5%, and the alternative used is given by $H_A : \beta = c$.

The case with independent errors corresponds to $\Phi = 0$ and $\Sigma_e = I_2$. In this case (8) is estimated using standard OLS, and the standard t-test, which is the optimal test for this DGP, is included as well. The tests are carried out at a nominal size of 5%, which coincides with the actual size of the the Dan-0.2 test, but the finite sample sizes of the t, HAC and the HAC-pw tests are 6%, 7% and 9% respectively. Figure 5 depicts the finite sample power of the four tests. From here it is clear that the HAC and the HAC-pw have virtually identical power, which dominates the power of the Dan-0.2 test statistic. The HAC tests also slightly dominate the t-test with regards to power. With this benchmark in mind, we will now add endogeneity to the model.

We model endogeneity following Stock and Watson (1993), where $\Phi = \begin{bmatrix} -.103 & -.039 \\ -.062 & .643 \end{bmatrix}$ and $\Sigma_e = 0.01 \times \begin{bmatrix} .951 & .499 \\ .499 & 1.374 \end{bmatrix}$ and (8) is estimated with DOLS. When $T = 50$, we report results for $p = 1, 2, 3$ and when $T = 100$, we report results for $p = 1, .., 5$. Tables 5 provides the finite sample sizes of the three test-statistics for the hypothesis $H_0 : \beta = 0$, which is performed as a two-sided test with a nominal level of 5%. From this table it is clear that using Dan-0.2 strictly dominates any of the other tests in size. In fact, when $T = 50$, 0.18 is the *highest* rejection probability obtained by Dan-0.2, while 0.30 is the *lowest* rejection

14

probability obtained by the other two test statistics! It is also worth noting that HAC-pw strictly dominates the HAC test. In conclusion, the Dan-0.2 test should be chosen for superior size performance.

An additional point which deserves mention is the fact that choosing the best test statistic seems far more important that choosing the "right" value of p. For a given test statistic, size changes by no more than 0.09 depending on the choice of p, while for a given value of p, size changes by as much as 0.21 depending on the choice of test statistic.

The next step, then, is to compare the power of these three test statistics. Tables 6 and 7 provide the finite sample power for all three statistics for $T = 50$, $p = 1, 2, 3$ and $T = 100$, $p = 1, .., 5$ respectively.

In terms of power, the HAC dominates the HAC-pw, which dominates the Dan-0.2 test. This result is not particularly surprising as the HAC and HAC-pw tests have actual sizes of 50% and 30% respectively with a nominal level of 5%. So while the Dan-0.2 has excellent size properties, the power performance is weaker than that of the standard HAC test, and the size power trade-off is still evident, also between the HAC and the HAC-pw tests. It is evident that the differences between the test statistics are much more pronounced when there is endogeneity in the data: The size distortion of the standard HAC tests skyrockets, but their power advantage increases. Again, it is worth noting that the choice of p has very little impact on the power of the test statistics compared to the choice of the test statistic. In fact no choice of p maximizes power uniformly for all alternative hypotheses.

In conclusion, the new test statistic provides a test which has approximately correct size for even very small sample sizes, while still maintaining good power.

6   Money-Demand Estimation

In this section, we will re-examine the long run money demand relationship for the United States. It is generally accepted that the long-run demand for money is functionally related to interest rates and national income. The model and data for this application is the same

as that of Ball (2001), who in turn based his analysis on the seminal work of Lucas (1988) and Stock and Watson (1993). Ball used the same econometric methods as Stock and Watson (1993), but extended the data by 9 years and got drastically different results. We are interested in examining the robustness of the results obtained by Ball.

The model of interest is the following canonical money-demand function:

$$m - p = \alpha + \theta_y y + \theta_r r + \varepsilon, \tag{9}$$

where m, p, and y are the logs of the money stock, the price level and real output. r is the nominal interest rate. m is measured as $M1$, output as NNP, the price level is the NNP-deflator and r is the commercial paper rate. Details can be found in Stock and Watson (1993). Both Ball (2001) and Stock and Watson (1993) have independently verified that (9) is a valid cointegration relationship. This allows us to proceed with the methods described in earlier sections of this paper. In this application, we focus on the post-war era only, utilizing data from 1946 through 1996.

Stock and Watson (1993) had reached the conclusion that the income elasticity (θ_y) was near one and the interest semi-elasticity (θ_r) was approximately -0.1. Ball (2001), on the other hand, obtains estimates of 0.5 and -0.05 respectively, and with much tighter standard errors, thus he rejects the values $\theta_y = 1$ and $\theta_r = -0.1$ obtained by Stock and Watson (1993).[12]

The hypotheses tested by Ball (2001), which we wish to re-visit are: $H_{joint} : \theta_y = 1$ and $\theta_r = -0.1$, $H_y : \theta_y = 1$, and $H_r : \theta_r = -0.1$. Estimating (9) using DOLS and routine procedures to obtain the standard errors, Ball rejects H_{joint} with a $p-$value less that 10^{-14}. He then estimates the error structure parametrically and uses these estimates to perform simulations and obtains more realistic $p-$values. Through this procedure, he obtains $p-$values for H_y and H_r of 0.001 and 0.02 respectively, and for H_{joint}, he ultimately finds a $p-$value of 0.004. These results indicate that the values obtained by Stock and Watson, can be rejected and that the usual standard errors are grossly over-estimated. The

16

exaggerated standard errors under the null correspond to what we would expect according to the simulations in earlier sections. Table 8 provides the parameter estimates from the estimation of (9) with 2 leads and lags of y and r included as regressors as well as the $W-$statistics for H_{joint}, H_y and H_r. Using Dan-0.2 the $p-$values are 0.0001 and 0.0006 for H_y and H_r and 0.0003 for H_{joint}. These values are quite similar to the simulation results reported in Ball (2001), and certainly more realistic than those obtained by standard methods. If the $p-$values obtained by the Dan-0.2 test are accurate, the suggestion is that Ball ultimately underestimates the standard errors of the parameter estimates.

The results obtained in this section should provide confidence in the rejection of the elasticities obtained by Stock and Watson. While there is no way to determine which $p-$values are correct, this application in conjunction with the simulations in earlier section of the paper indicates that realistic $p-$values can be quickly obtained without explicit parametric modeling of the unknown error structure and without the time-consuming simulations performed in Ball (2001).

7   Conclusion

In this paper we have proposed a new test for hypotheses regarding the cointegration vector in a single-equation cointegration model. The new test is based on the standard OLS HAC robust covariance estimators and does not require knowledge of the form of serial correlation in the data The tests we analyze are . We extend the fixed-b asymptotic framework for HAC robust tests recently proposed by Kiefer and Vogelsang (2002). This allows us to analyze the power properties of the new test with regards to bandwidth and kernel choices. We address the traditionally difficult issue of HAC bandwidth choice using fixed-b asymptotics in conjunction with local to unity asymptotics. Our analysis shows that among popular kernels, the Daniell kernel with bandwidth $0.2T$ delivers tests with size close to the nominal size while retaining good power and hence the Dan-0.2 test is recommended in practice. The test provides a new tool for investigation of single equation

cointegration models and it can be computed using any standard package.

We perform finite sample simulation experiments to verify the performance of the test statistic, and finally we apply the newly developed test to investigate the well known money-demand relationship.

The test introduced in this paper provides a solution to the size inflation and choice of bandwidth introduced by the serial correlation in the errors. To estimate the parameters when the regressors are endogenous, DOLS is used. This, however introduces some issues that are very similar to the choice of bandwidth. While simulations indicate that the problems introduced by serial correlation are more severe, a goal for future research should be to deal with both issues at once. Currently the most promising avenue to achieve this is to use the Fully Modified estimator introduced by Phillips and Hansen (1990) to estimate the model and the apply the fixed-b theory to these estimators directly.

References

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

Andrews, D. W. K. and Monahan, J. C.: (1992), An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* **60**, 953–966.

Ball, L.: (2001), Another look at long-run money demand, *Journal of Monetary Economics* **47**, 31–44.

Bunzel, H., Kiefer, N. M. and Vogelsang, T. J.: (2001), Simple robust testing of hypotheses in non-linear models, *Journal of American Statistical Association* **96**, 1088–1098.

De Brouwer, G. and Ericsson, N.: (1998), Modeling inflation in Australia, *Journal of Business and Economic Statistics* **16**, 433–449.

de Jong, R. M.: 2002, A note on "convergence rates and asymptotic normality for series estimators": Uniform convergence rates, *Journal of Econometrics* **111**, 1–9.

de Jong, R. M. and Davidson, J.: (2000), Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices, *Econometrica* **68**, 407–424.

den Haan, W. J. and Levin, A.: (1997), A practictioner's guide to robust covariance matrix estimation, *in* G. Maddala and C. Rao (eds), *Handbook of Statistics: Robust Inference*, Volume 15, Elsevier, New York, pp. 291–341.

Ericsson, N.R., D. H. and Mizon, G.: (1998), Exogeneity, cointegration, and economic policy analysis, *Journal of Business and Economic Statistics* **16**, 370–387.

Hansen, B. E.: (1992), Consistent covariance matrix estimation for dependent heterogenous processes, *Econometrica* **60**, 967–972.

Jansson, M.: (2002), Autocorrelation robust tests with good size and power, *Working Paper, Department of Economics, U.C. Berkeley* .

Johansen, S.: (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**, 231–254.

Johansen, S.: (1991), Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models, *Econometrica* **59**, 1551–80.

Johansen, S. and Juselius, K.: (1990), Maximum likelihood estimation and inference on cointegration: With applications to the demand for money, *Oxford Bulletin of Economics and Statistics* **52**, 169–210.

Johansen, S. and Juselius, K.: (1992), Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK, *Journal of Econometrics* **53**, 211–44.

Kiefer, N. M. and Vogelsang, T. J.: (2002), A new asymptotic theory for heteroskedasiticy-autocorrelation robust tests. Working Paper, Center for Analytic Economics, Cornell University.

Kiefer, N. M., Vogelsang, T. J. and Bunzel, H.: (2000), Simple robust testing of regression hypotheses, *Econometrica* **68**, 695–714.

Lewis, R. and Reinsel, G. C.: (1985), Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**, 393–411.

Lucas, R.: (1988), Money demand in the United States: a quantitive review, *Carnegie-Rochester Conference Series on Public Policy* **29**, 137–168.

Metin, K.: (1998), The relationship between inflation and the budget deficit in turkey, *Journal of Business and Economic Statistics* **16**, 412–422.

Newey, W. K. and West, K. D.: (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

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

Phillips, P. C. B. and Loretan, M.: (1991), Estimating long run economic equilibria, *Review of Economic Studies* **58**, 407–436.

Robinson, P.: (1991), Automatic frequency domain inference on semiparametric and non-parametric models, **59**, 1329–1363.

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

Saikkonen, P.: (1991), Asymptotically efficient estimation of cointegrating regressions, *Econometric Theory* **7**, 1–21.

Stock, J. H. and Watson, M. W.: (1993), A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**, 783–820.

Vogelsang, T. J.: (1998), Trend function hypothesis testing in the presence of serial correlation correlation parameters, *Econometrica* **65**, 123–148.

White, H.: (1984), *Asymptotic Theory for Econometricians*, Academic Press, New York.

Wooldridge, J. M.: (1991), Notes on regression with difference-stationary data. Michigan State University, Mimeo.

Appendix

A Proof of Lemma 1.

Following Kiefer and Vogelsang (2002), we define

$$\Delta^2 \kappa_{ij} = \left\{ k\left(\frac{i-j}{[bN]}\right) - k\left(\frac{i-j-1}{[bN]}\right) \right\} - \left\{ k\left(\frac{i-j+1}{[bN]}\right) - k\left(\frac{i-j}{[bN]}\right) \right\},$$

and use this expression to rewrite $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = N^{-1} \sum_{l=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2 \Delta^2 \kappa_{il} \left(N^{-1/2} \widehat{S}_i \right) \left(N^{-1/2} \widehat{S}_l \right), \tag{10}$$

where $\widehat{S}_{[rN]} = \sum_{t=1}^{[rN]} \hat{v}_{t+p}$ and $\{\hat{v}_t\}_{t=p+1}^{T-p}$ are the residuals from (4). Note that for (10) to be valid it must be the case that the residuals sum to zero. Therefore, for the results which follow to be valid $f(t)$ must include a constant term as assumed in Assumption 1. To establish the asymptotic distribution of $\hat{\sigma}^2$, it is necessary first to determine the asymptotic distribution of $\hat{S}_{[rN]}$.

Lemma 3 *Under Assumptions 1-3, $N^{-\frac{1}{2}} \hat{S}_{[rN]} \Rightarrow \sigma V(r)$.*

Proof. First define $\gamma = [\gamma'_{-p}, ..., \gamma'_p]'$, which is a $(\mathtt{k}(2p+1) \times 1)$ vector of parameters and let $\Delta Z_{t+p} = \left[\Delta X'_{t-p}, \Delta X'_{t-p+1,...,} \Delta X'_{t+p} \right]'$ be the corresponding vector of regressors. Simple matrix manipulations yield:

$$
\begin{aligned}
N^{-\frac{1}{2}} \hat{S}_{[rN]} &= N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \left(\dot{v}_{t+p} - f(t+p)'(\hat{\alpha} - \alpha) - X'_{t+p}\left(\hat{\beta} - \beta\right) - \Delta \mathbf{Z}'_{t+p}(\hat{\gamma} - \gamma) \right) \\
&= N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \dot{v}_{t+p} - \left[\begin{array}{c} N^{-1} \sum_{t=1}^{[rN]} f(t+p) \\ N^{-\frac{3}{2}} \sum_{t=1}^{[rN]} X_{t+p} \end{array} \right]' \left[\begin{array}{c} N^{\frac{1}{2}}(\hat{\alpha} - \alpha) \\ N\left(\hat{\beta} - \beta\right) \end{array} \right] \\
&\quad - N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \Delta \mathbf{Z}'_{t+p}(\hat{\gamma} - \gamma)
\end{aligned}
\tag{11}
$$

21

In what follows, we will show that the last term in the expression for $\hat{S}_{[rN]}$, (11), is $O_P\left(\frac{p^2}{T}\right)$ and therefore does not affect the asymptotic distribution of $\hat{S}_{[rN]}$. First rewrite the expression.

$$
\begin{aligned}
N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \mathbf{\Delta Z}'_{t+p} \left(\hat{\gamma} - \gamma\right) &= N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} \sum_{s=-p}^{p} u'_{2,t-s} \left(\hat{\gamma}_s - \gamma_s\right) \\
&= N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} \sum_{s=-p}^{p} \left(\hat{\gamma}_s - \gamma_s\right)' u_{2,t-s} \\
&= \sum_{s=-p}^{p} \left(\hat{\gamma}_s - \gamma_s\right)' \left(N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} u_{2,t-s}\right)
\end{aligned}
$$

Considering the norm of this expression,

$$
\begin{aligned}
E \left\| N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \mathbf{\Delta Z}'_{t+p} \left(\hat{\gamma} - \gamma\right) \right\| &\leq \sum_{s=-p}^{p} E \left(\left\| \left(\hat{\gamma}_s - \gamma_s\right) \right\| \left\| N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} u_{2,t-s} \right\| \right) \\
&\leq \sum_{s=-p}^{p} \left(E \left\| \left(\hat{\gamma}_s - \gamma_s\right) \right\|^2 E \left\| N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} u_{2,t-s} \right\|^2 \right)^{\frac{1}{2}}.
\end{aligned}
$$

Saikkonen (1991) proved that under Assumption 2, $\left\| \left(\hat{\gamma}_s - \gamma_s\right) \right\|^2 = O_P\left(\frac{1}{T}\right)$, thus we concentrate on the last term in the product above.

$$
\begin{aligned}
E \left\| N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} u_{2,t-s} \right\|^2 &= N^{-1} \sum_{t=p+1}^{[r(T-p)]} \sum_{m=p+1}^{[r(T-p)]} E\left(u'_{2,t-s} u_{2,m-s}\right) \\
&= N^{-1} \sum_{t=p+1}^{[r(T-p)]} \sum_{m=p+1}^{[r(T-p)]} tr\Gamma_{22}\left(t - m\right) \\
&= N^{-1} \sum_{t=p+1}^{[r(T-p)]} \sum_{q=p+1-t}^{[r(T-p)]-t} tr\Gamma_{22}\left(q\right)
\end{aligned}
$$

By Assumption 2 g) $\sum_{q=p+1-t}^{[r(T-p)]-t} tr\Gamma_{22}\left(q\right) \leq C < \infty$, implying that $N^{-1} \sum_{t=p+1}^{[r(T-p)]} \sum_{q=p+1-t}^{[r(T-p)]-t} tr\Gamma_{22}\left(q\right) = O_P\left(1\right)$. We can thus conclude that

$$
E \left\| N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \mathbf{\Delta Z}'_{t+p} \left(\hat{\gamma} - \gamma\right) \right\| \leq \sum_{s=-p}^{p} \left(E \left\| \left(\hat{\gamma}_s - \gamma_s\right) \right\|^2 E \left\| N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} u_{2,t-s} \right\|^2 \right)^{\frac{1}{2}} = O_P\left(\frac{p}{N^{\frac{1}{2}}}\right)
$$

22

and by Jensen's inequality, it follows that $E \left\| N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \Delta \mathbf{Z}'_{t+p} (\hat{\gamma} - \gamma) \right\|^2 = O_P \left(\frac{p^2}{N} \right).$
Now, since $plim \left(\frac{T}{N} \right) = 1$, $N^{-\frac{1}{2}} \sum_{t=1}^{[rN]} \Delta \mathbf{Z}'_{t+p} (\hat{\gamma} - \gamma) = O_P \left(\frac{p^2}{T} \right)$ as desired.

We can now determine the asymptotic distribution of $T^{-\frac{1}{2}} \hat{S}_{[rN]}$ from the first two terms of (11). By Assumptions 1-3 we know from Saikkonen (1991) and Phillips and Hansen (1990) that

$$
\begin{bmatrix} T^{\frac{1}{2}} \tau_T^{-1} (\hat{\alpha} - \alpha) \\ T \left(\hat{\beta} - \beta \right) \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma \left(\int_0^1 F^X (s) F^X (s)' ds \right)^{-1} \int_0^1 F^X (s) dw_1 (s) \\ \sigma (\Lambda')^{-1} \left(\int_0^1 w_k^F (s) w_k^F (s)' ds \right)^{-1} \int_0^1 w_k^F (s) dw_1 (s) \end{bmatrix},
$$

and

$$
\begin{bmatrix} T^{-1} \sum_{t=1}^{[rN]} \tau_T f (t + p) \\ T^{-\frac{3}{2}} \sum_{t=1}^{[rN]} X_{t+p} \end{bmatrix} \Rightarrow \begin{bmatrix} \int_0^r F^X (s) ds \\ \Lambda \int_0^r w_k^F (s) ds \end{bmatrix}.
$$

It also follows directly from Assumption 2 that $N^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} \dot{v}_t \Rightarrow \sigma w_1 (r)$. Since $\frac{N}{T} \to 1$, it will also be the case that $T^{-\frac{1}{2}} \sum_{t=p+1}^{[r(T-p)]} \dot{v}_t \Rightarrow \sigma w_1 (r)$. So it is now established that

$$
\begin{aligned}
T^{-\frac{1}{2}} \hat{S}_{[rN]} &\Rightarrow \sigma w_1 (r) - \sigma \int_0^r F (s)' ds \left(\int_0^1 F^X (s) F^X (s)' ds \right)^{-1} \int_0^1 F^X (s) dw_1 (s) \\
&\quad - \sigma \int_0^r w_k (s)' ds \left(\int_0^1 w_k^F (s) w_k^F (s)' ds \right)^{-1} \int_0^1 w_k^F (s) dw_1 (s) \\
&= \sigma V (r).
\end{aligned}
$$

∎

The rest of the proof is split into three cases, corresponding to Type 1, Type 2 and the Bartlett kernels. It follows directly from Kiefer and Vogelsang (2002) and Lemma 3.

Case 1: $k (x)$ is a Type 1 kernel. By definition of the second derivative, $T^2 \Delta^2 \kappa_{il} - \left(-k^{*\prime\prime} \left(\frac{i-l}{N} \right) \right) \to 0$, and using Lemma 3 it follows easily that

$$
\begin{aligned}
\hat{\sigma}^2 &= N^{-1} \sum_{l=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2 \Delta^2 \kappa_{il} N^{-1/2} \hat{S}_i N^{-1/2} \hat{S}_l \\
&\Rightarrow \sigma^2 \int_0^1 \int_0^1 -k^{*\prime\prime} (r - s) V (s) V (r) dr ds.
\end{aligned}
$$

Case 2: $k (x)$ is a Type 2 kernel. Following Kiefer and Vogelsang (2002), we use simple algebra and the definition of $\Delta^2 \kappa_{ij}$ to establish that when $|i - j| > [bN]$, $\Delta^2 \kappa_{ij} = 0$, and

23

when $|i - j| = [bN]$, $\Delta^2 \kappa_{ij} = -k \left(\frac{[bN]-1}{[bN]} \right)$. Also recall that when $|i - j| < [bN]$ $k(x)$ is twice continuously differentiable. We split up the expression of $\hat{\sigma}^2$ as follows:

$$
\begin{aligned}
\hat{\sigma}^2 &= N^{-1} \sum_{l=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} 1_{\{|i-j|<[bN]\}} N^2 \Delta^2 \kappa_{il} N^{-1/2} \widehat{S}_i N^{-1/2} \widehat{S}_l \\
&\quad + 2N^{-2} \sum_{l=1}^{N-1-[bN]} N^2 k \left(\frac{[bN]-1}{[bN]} \right) N^{-1/2} \widehat{S}_i N^{-1/2} \widehat{S}_{i+[bN]} \\
&= N^{-1} \sum_{l=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} 1_{\{|i-j|<[bN]\}} N^2 \Delta^2 \kappa_{il} N^{-1/2} \widehat{S}_i N^{-1/2} \widehat{S}_l \\
&\quad - 2k \left(1 - \frac{1}{[bN]} \right) \sum_{l=1}^{N-1-[bN]} N^{-1/2} \widehat{S}_i N^{-1/2} \widehat{S}_{i+[bN]} \\
&\Rightarrow \sigma^2 \left(\int \int_{|r-s|<b} -k^{*''}(r-s) V(r) V(s) \, dr ds + 2k_-^{*'}(b) \int_0^{1-b} V(r+b) V(r) \, dr \right),
\end{aligned}
$$

where the asymptotic distribution follows directly from Lemma 3 and Kiefer and Vogelsang (2002).

Case 3: $k(x)$ is the Bartlett Kernel. Here again following Kiefer and Vogelsang (2002), it can be verified that when $|i - j| = 0$, $\Delta^2 \kappa_{ij} = \frac{2}{[bN]}$, and when $|i - j| = [bN]$, $\Delta^2 \kappa_{ij} = -\frac{1}{[bN]}$. Using these expressions and Lemma (3) in (10), we obtain the following limiting distribution:

$$
\begin{aligned}
\hat{\sigma}^2 &= N^{-1} \sum_{l=1}^{N-1} N^{-1} \sum_{i=1}^{N-1} N^2 \Delta^2 \kappa_{il} N^{-1/2} \widehat{S}_i N^{-1/2} \widehat{S}_l \\
&= \frac{2}{[bN]} \sum_{i=1}^{N-1} \left(N^{-1/2} \widehat{S}_i \right)^2 - \frac{2}{[bN]} \sum_{i=1}^{N-1-[bN]} N^{-1/2} \widehat{S}_i N^{-1/2} \widehat{S}_{i+[bN]} \\
&\Rightarrow \sigma^2 \left(\frac{2}{b} \int_0^1 V(r)^2 \, dr - \frac{2}{b} \int_0^{1-b} V(r+b) V(r) \, dr \right).
\end{aligned}
$$

B   Proof of Theorem 2.

The initial step of the proof will be to re-write the model, projecting out all regressors which are not related to the hypothesis in question. Then we will prove that the statistic is numerically unchanged if it is calculated from the re-written model. Finally the expression of W obtained from the re-written model will be used to derive the asymptotic distribution of the statistic.

24

To re-write the model, let $L = \begin{bmatrix} R \\ D \end{bmatrix}$, where D is chosen such that L has full rank (k),

and define $\begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 \end{bmatrix} = XL^{-1}$ and $\tilde{Z} = Z \cdot (L^{-1} \otimes I_{2p+1})$. Using these definitions, Model (4) can be rewritten in the following manner:

$$
\begin{aligned}
y &= f(T)\alpha + (XL^{-1})(L\beta) + (\Delta Z \cdot (L^{-1} \otimes I_{2p+1}))(L \otimes I_{2p+1})\gamma + \dot{v} \\
&= f(T)\alpha + \begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + \Delta\tilde{Z}\tilde{\gamma} + \dot{v} \\
&= f(T)\alpha + \tilde{X}_1\beta_1^* + \tilde{X}_2\beta_2^* + \Delta\tilde{Z}\tilde{\gamma} + \dot{v}
\end{aligned}
$$

Since $\tilde{X}_1$ and $\tilde{X}_2$ are linear combinations of X, they too contain unit root processes as long as the original assumption of just one cointegration relationship is maintained. Furthermore $\Delta\tilde{Z}$ contains the leads and lags of the differenced $\tilde{X}$ variables. We will now show that testing, $H_0 : R\beta = \beta_0$, is equivalent to testing the hypothesis $\tilde{H}_0 : \beta_1^* = \beta_0$ in the model

$$
y^* = X_1^*\beta_1^* + \Delta Z^*\gamma + \dot{v}^*, \tag{E1.1}
$$

where for any matrix G, $M_G = I - G(G'G)^{-1}G$, $\tilde{X}_2^F = M_f\tilde{X}_2$, $X_1^* = M_{\tilde{X}_2^F}M_f\tilde{X}_1$, $Z^* = M_{\tilde{X}_2^F}M_f\tilde{Z}$, $\dot{v}^* = M_{\tilde{X}_2^F}M_f\dot{v}$, and $y^* = M_{\tilde{X}_2^F}M_f y$.

Lemma 4 *The statistic for testing $\tilde{H}_0 : \beta_1^* = \beta_0$ from (E1.1) is numerically identical to the statistic for testing $H_0 : R\beta = \beta_0$ from (4).*

Proof. The statistic for testing $\tilde{H}_0$ from (E1.1) takes the form

$$
W^* = T(\beta_1^* - \beta_0)' \left[\hat{\sigma}^{2*} (T^{-1}(X_1^*)'X_1^*)^{-1}\right]^{-1} (\beta_1^* - \beta_0),
$$

and the statistic for testing $H_0 : R\beta = \beta_0$ from (4) can be written as

$$
W = T(R\beta - \beta_0)' \left[\hat{\sigma}^2 R Q_{fX}^{-1} R'\right]^{-1} (R\beta - \beta_0).
$$

Since $\hat{\sigma}^2$ and $\hat{\sigma}^{2*}$ are calculated from just the residuals, we know from the Frisch-Waugh-Lovell Theorem that they will be identical. In addition, since β_1^* is equal to $R\beta$ by definition, we know that $R\beta - \beta_0 = \beta_1^* - \beta_0$. We then need to demonstrate that

$$
(T^{-1}(X_1^*)'X_1^*)^{-1} = R Q_{fX}^{-1} R'. \tag{12}
$$

By definition,

$$\left(T^{-1}\left(X_1^*\right)'X_1^*\right)^{-1} = \left(T^{-1}\tilde{X}_1'M_f M_{\tilde{X}_2^F} M_f \tilde{X}_1\right)^{-1}, \text{ and}$$

$$M_f M_{\tilde{X}_2^F} M_f = M_f \left(I - M_f \tilde{X}_2 \left(\tilde{X}_2' M_f \tilde{X}_2\right)^{-1} \tilde{X}_2' M_f\right) M_f$$

$$= \left(M_f - M_f \tilde{X}_2 \left(\tilde{X}_2' M_f \tilde{X}_2\right)^{-1} \tilde{X}_2' M_f\right),$$

such that we can write

$\left(T^{-1}\left(X_1^*\right)'X_1^*\right)^{-1} = \left(T^{-1}\tilde{X}_1'\left(M_f - M_f \tilde{X}_2 \left(\tilde{X}_2' M_f \tilde{X}_2\right)^{-1} \tilde{X}_2' M_f\right)\tilde{X}_1\right)^{-1}$. Now look at
$RQ_{fX}^{-1}R'$.

$$RQ_{fX}^{-1}R' = \begin{bmatrix} 0 & R \end{bmatrix} \left(\frac{1}{T}\begin{bmatrix} \mathbf{f}(T)'\mathbf{f}(T) & \mathbf{f}(T)'X \\ X'\mathbf{f}(T) & X'X \end{bmatrix}\right)^{-1}\begin{bmatrix} 0 \\ (R)' \end{bmatrix}.$$

By the formula for the inverse of partitioned matrices, this simplifies to

$$RQ_{fX}^{-1}R' = R\left(T^{-1}X'M_f X\right)^{-1}(R)' = R\left(T^{-1}\begin{bmatrix} X_1'M_f X_1 & X_1'M_f X_2 \\ X_2'M_f X_1 & X_2'M_f X_2 \end{bmatrix}\right)^{-1}(R)'.$$

We will now write R as

$$R = [I\ 0]\begin{bmatrix} R \\ D \end{bmatrix} = [I\ 0]\,L.$$

This expression for R along with the fact that L is invertible, makes it possible to obtain the following expression:

$$RQ_{fX}^{-1}R' = \hat{\sigma}^2\begin{bmatrix} I & 0 \end{bmatrix}\left(T^{-1}(L')^{-1}X'M_f X L^{-1}\right)^{-1}\begin{bmatrix} I \\ 0 \end{bmatrix}$$

$$= \hat{\sigma}^2\begin{bmatrix} I & 0 \end{bmatrix}\left(T^{-1}\begin{bmatrix} \left(M_f \tilde{X}_1\right)' \\ \left(M_f \tilde{X}_2\right)' \end{bmatrix}\begin{bmatrix} M_f \tilde{X}_1 & M_f \tilde{X}_2 \end{bmatrix}\right)^{-1}\begin{bmatrix} I \\ 0 \end{bmatrix}.$$

Using the inverse matrix formula yet again, we get

$$RQ_{fX}^{-1}R' = \left(T^{-1}\tilde{X}_1'M_f\left(I - M_f \tilde{X}_2 \left(\tilde{X}_2' M_f \tilde{X}_2\right)^{-1} M_f \tilde{X}_2\right)M_f \tilde{X}_1\right)^{-1}$$

$$= \left(T^{-1}\tilde{X}_1'\left(M_f - M_f \tilde{X}_2 \left(\tilde{X}_2' M_f \tilde{X}_2\right)^{-1} M_f \tilde{X}_2 M_f\right)\tilde{X}_1\right)^{-1},$$

26

which is exactly identical to the expression for $\left(T^{-2}\left(X_1^*\right)' X_1^*\right)^{-1}$, proving that the two test statistics are numerically identical. ∎

To complete the proof of **(a)**, we thus need to determine the asymptotic distribution of

$$
\begin{aligned}
W^* &= T\left(\beta_1^* - \beta_0\right)'\left[\hat{\sigma}^{2*}\left(T^{-1}\left(X_1^*\right)' X_1^*\right)^{-1}\right]^{-1}\left(\beta_1^* - \beta_0\right) \\
&= T\left(\beta_1^* - \beta_0\right)'\left[\hat{\sigma}^{2*}\left(T^{-2}\left(X_1^*\right)' X_1^*\right)^{-1}\right]^{-1} T\left(\beta_1^* - \beta_0\right).
\end{aligned}
$$

Since $\hat{\sigma}^{2*} = N^{-1}\sum_{l=1}^{N-1} N^{-1}\sum_{i=1}^{N--1} N^2\Delta^2\kappa_{il}\left(N^{-1/2}\widehat{S}_i^*\right)\left(N^{-1/2}\widehat{S}_l^*\right)$, where , $\hat{S}_t^*$ is defined as $\hat{S}_t$, but for the model in (E1.1), we know

$$
\begin{aligned}
N^{-\frac{1}{2}}\hat{S}_{[rN]}^* &= N^{-\frac{1}{2}}\sum_{t=1}^{[rN]}\dot{v}_{t+p}^* - N^{-\frac{3}{2}}\sum_{t=1}^{[rN]}X_{1,t+p}^*\left(N\left(\hat{\beta}_1^* - \beta_1^*\right)\right) - N^{-1}\sum_{t=1}^{[rN]}\Delta\mathbf{Z}_{t+p}^{*'}N^{\frac{1}{2}}\left(\hat{\gamma} - \gamma\right) \\
&\Rightarrow \sigma^*\hat{w}_1^F(r) - \sigma^*\int_0^r\hat{w}_q^F(s)'\,ds\left(\int_0^1\hat{w}_q^F(s)\hat{w}_q^F(s)'\,ds\right)^{-1}\int_0^1\hat{w}_q^F(s)\,d\hat{w}_1^F(s) \\
&= \sigma^*V^F(r),
\end{aligned}
$$

and therefore,

$\hat{\sigma}^{*2} \Rightarrow -\sigma^{*2}\int_0^1\int_0^1 k^{*''}(r-s)V^F(r)V^F(s)\,dr\,ds$ if k is type 1

$\hat{\sigma}^{*2} \Rightarrow \sigma^{*2}(\int\int_{|r-s|<b} -k^{*''}(r-s)V^F(r)V^F(s)\,dr\,ds + 2k_-^{*'}(b)\int_0^{1-b}V^F(r+b)V^F(r)\,dr)$ if k is type 2

$\hat{\sigma}^{*2} \Rightarrow \sigma^{*2}\frac{2}{b}\left[\int_0^1 V^F(r)^2\,dr - \int_0^{1-b}V^F(r)V^F(r+b)\,dr\right]$ if k is Bartlett.

By the definition of X_1^*,

$$
\left(T^{-2}\left(X_1^*\right)' X_1^*\right)^{-1} \Rightarrow \left(\Lambda^*\int_0^1\hat{w}_q^F(s)\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1}.
$$

The distribution of W^* can now be obtained.

If k is Type 1,

$$
W^* = T\left(\beta_1^* - \beta_0\right)'\left[\hat{\sigma}^{2*}\left(T^{-2}\left(X_1^*\right)' X_1^*\right)^{-1}\right]^{-1} T\left(\beta_1^* - \beta_0\right)
$$

$$
\Rightarrow \left(-\int_0^1\int_0^1 k^{*''}(r-s)V^F(s)V^F(r)\,dr\,ds\right)^{-1}
$$

$$
\int_0^1\hat{w}_q^F(s)'\,d\hat{w}_1^F(s)\left(\int_0^1\hat{w}_q^F(s)\hat{w}_q^F(s)'\,ds\right)^{-1}\int_0^1\hat{w}_q^F(s)\,d\hat{w}_1^F(s)
$$

If k is Type 2,

$$W^* \;=\; T\,(\beta_1^* - \beta_0)' \left[\hat{\sigma}^{2*}\left(T^{-2}\,(X_1^*)'\,X_1^*\right)^{-1}\right]^{-1} T\,(\beta_1^* - \beta_0)$$

$$\Rightarrow \;\; \sigma^* \int_0^1 \hat{w}_q^F(s)'\,d\hat{w}_1^F(s)\,(\Lambda^*)' \left(\Lambda^* \int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1}$$

$$\left\{\sigma^{*2}\left(\int\!\!\int_{|r-s|<b} -k^{*\prime\prime}(r-s)\,V^F(r)\,V^F(s)\,dr\,ds + 2k_-^{*\prime}(b)\int_0^{1-b} V^F(r+b)\,V^F(r)\,dr\right)\right.$$

$$\left.\left(\Lambda^* \int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1}\right\}^{-1}$$

$$\sigma^*\left(\Lambda^* \int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1} \Lambda^* \int_0^1 \hat{w}_q^F(s)\,d\hat{w}_1^F(s)$$

$$=\;\; \left(\int\!\!\int_{|r-s|<b} -k^{*\prime\prime}(r-s)\,V^F(r)\,V^F(s)\,dr\,ds + 2k_-^{*\prime}(b)\int_0^{1-b} V^F(r+b)\,V^F(r)\,dr\right)^{-1}$$

$$\int_0^1 \hat{w}_q^F(s)'\,d\hat{w}_1^F(s)\left(\int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\right)^{-1}\int_0^1 \hat{w}_q^F(s)\,d\hat{w}_1^F(s).$$

If k is Bartlett,

$$W^* \;=\; T\,(\beta_1^* - \beta_0)' \left[\hat{\sigma}^{2*}\left(T^{-2}\,(X_1^*)'\,X_1^*\right)^{-1}\right]^{-1} T\,(\beta_1^* - \beta_0)$$

$$\Rightarrow \;\; \sigma^* \int_0^1 \hat{w}_q^F(s)'\,d\hat{w}_1^F(s)\,(\Lambda^*)' \left(\Lambda^* \int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1}$$

$$\left\{\frac{2}{b}\,(\sigma^*)^2\left(\Lambda^* \int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1}\left[\int_0^1 V^F(r)^2\,dr - \int_0^{1-b} V^F(r)\,V^F(r+b)\,dr\right]\right\}^{-1}$$

$$\sigma^*\left(\Lambda^* \int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\,(\Lambda^*)'\right)^{-1} \Lambda^* \int_0^1 \hat{w}_q^F(s)\,d\hat{w}_1^F(s)$$

$$=\;\; \frac{b}{2}\left(\int_0^1 V^F(r)^2\,dr - \int_0^{1-b} V^F(r)\,V^F(r+b)\,dr\right)^{-1}$$

$$\int_0^1 \hat{w}_q^F(s)'\,d\hat{w}_1^F(s)\left(\int_0^1 \hat{w}_q^F(s)\,\hat{w}_q^F(s)'\,ds\right)^{-1}\int_0^1 \hat{w}_q^F(s)\,d\hat{w}_1^F(s).\;\blacksquare$$

C Kernels

$$\text{Bartlett} \quad k(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Parzen} \quad k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } |x| \leq \frac{1}{2} \\ 2(1 - |x|)^3 & \text{for } \frac{1}{2} \leq |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Quadratic Spectral (QS)} \quad k(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right)$$

$$\text{Daniell} \quad k(x) = \frac{\sin(\pi x)}{\pi x}$$

$$\text{Bohman} \quad k(x) = \begin{cases} (1 - |x|)\cos(\pi x) + \sin(\pi |x|)/\pi & \text{for } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Notes

[1]See for example De Brouwer and Ericsson (1998), Metin (1998) or Ericsson and Mizon (1998).

[2]There are currently two competing frameworks used to estimate such models in ways that account for serial correlation and heteroscedasticity: the single equation framework (which forms the basis of this paper), and the systems framework. The systems approach to estimating cointegrating systems, applies the full information maximum likelihood (FIML) approach developed in Johansen (1988), Johansen (1991), Johansen and Juselius (1990) and Johansen and Juselius (1992). While the systems framework does account for serial correlation and heteroscedasticity, it is somewhat orthogonal to the framework used in this paper; as such, detailed comparisons between the FIML approach and the one presented here are beyond the scope of this paper.

[3]HAC estimators have been thoroughly examined in the literature. Among the major contributions are Andrews (1991), Andrews and Monahan (1992), Hansen (1992), de Jong 2002, de Jong and Davidson (2000), Newey and West (1987), Robinson (1991) and White (1984).

[4]This has been documented through simulations for stationary models in, for example, Andrews (1991), Andrews and Monahan (1992) and den Haan and Levin (1997).

[5]Although the model as it is characterized in (1) does not allow for trends in the regressors, the asymptotic results derived in this paper remain valid for hypotheses on β if the trends in the regressors are included in $f(t)$. This stems from the fact that the test statistic is invariant to projections of subsets of regressors in linear models.

[6]Note that $g_{uu}(0) = \Omega$.

[7]The bounds on p are similar to those used by Berk (1974), Lewis and Reinsel (1985), Said and Dickey (1984) and Saikkonen (1991).

[8]These results were developed by Saikkonen (1991), Phillips and Loretan (1991), Stock and Watson (1993) and Wooldridge (1991). Phillips and Hansen (1990) provide the results

when trends are included in the model.

[9]Hypotheses of the form $H_0 : R\alpha = \alpha_0$ can be dealt with in a similar manner.

[10]A possible exception is the Bartlett kernel, which might have a slight non-monotonicity in the size/b relationship.

[11]QS-0.2 denotes the test using the QS kernel with $b = 0.2$, Dan-0.2 is the Daniell kernel with $b = 0.2$, Parzen-0.4 is the Parzen kernel with $b = 0.4$ and Bohman-0.4 is the Bohman kernel with $b = 0.4$.

[12]These estimates have serious real world significance. For example, it is important for the implementation of monetary policy whether the income elasticity is unity or not, because a less than unity elasticity implies the money stock must grow more slowly than output if price stability is desired.

Table 1: Local Asymptotic Power, b = 0.02

c	QS	Daniell	Bohman	Parzen	Bartlett
0	0.05	0.05	0.05	0.05	0.05
1	0.09	0.09	0.09	0.09	0.09
2	0.18	0.18	0.18	0.19	0.19
3	0.34	0.34	0.35	0.35	0.34
4	0.55	0.56	0.58	0.59	0.57
5	0.77	0.77	0.79	0.80	0.78
6	0.90	0.89	0.91	0.91	0.90
7	0.96	0.96	0.96	0.96	0.96
8	0.98	0.98	0.99	0.99	0.99
9	0.99	0.99	0.99	0.99	0.99
10	1.00	1.00	1.00	1.00	1.00

Table 2: Finite Sample Size, nominal level 5% , b = 1♣

(λ, ρ)♣	Bartlett	Parzen	Bohman	Daniell	QS
$-0.8, 0.0$	0.007	0.035	0.035	0.043	0.043
$-0.8, 0.8$	0.057	0.057	0.055	0.054	0.054
$-0.8, 0.9$	0.122	0.082	0.080	0.070	0.070
$-0.4, 0.0$	0.037	0.053	0.052	0.051	0.051
$-0.4, 0.8$	0.113	0.078	0.076	0.067	0.068
$-0.4, 0.9$	0.203	0.121	0.118	0.096	0.096
$0.0, 0.0$	0.054	0.055	0.054	0.052	0.053
$0.0, 0.8$	0.124	0.084	0.082	0.071	0.071
$0.0, 0.9$	0.220	0.130	0.125	0.101	0.102
$0.4, 0.0$	0.059	0.056	0.057	0.054	0.054
$0.4, 0.8$	0.131	0.086	0.084	0.074	0.075
$0.4, 0.9$	0.225	0.132	0.128	0.103	0.104
$0.8, 0.0$	0.062	0.057	0.058	0.054	0.054
$0.8, 0.8$	0.133	0.087	0.085	0.073	0.074
$0.8, 0.9$	0.227	0.135	0.131	0.108	0.109

♣ *The data is generated according to $y_t = \alpha + \beta x_t + u_{1,t}$, $x_t = x_{t-1} + u_{2,t}$, $t = 1, .., 50$, where the errors are $u_{1,t} = \rho u_{1,t-1} + e_t + \lambda e_{t-1}$, and $\{e_t\}$ and $\{u_{2,t}\}$ are i.i.d. $N(0,1)$.*

Table 3: Finite Sample Size Comparison, T=50, nominal size 5%▲

λ	ρ	Dan-0.2	QS-0.2	Bohman-0.4	Parzen-0.4
-0.8	0	0.056	0.054	0.045	0.047
-0.8	0.2	0.054	0.051	0.046	0.049
-0.8	0.4	0.051	0.050	0.046	0.046
-0.8	0.8	0.060	0.060	0.059	0.060
-0.8	0.9	0.083	0.083	0.081	0.084
-0.4	0	0.056	0.056	0.053	0.055
-0.4	0.2	0.053	0.054	0.054	0.054
-0.4	0.4	0.052	0.052	0.052	0.052
-0.4	0.8	0.061	0.061	0.062	0.065
-0.4	0.9	0.080	0.081	0.078	0.084
0	0	0.052	0.052	0.052	0.052
0	0.2	0.049	0.050	0.051	0.051
0	0.4	0.051	0.051	0.052	0.052
0	0.8	0.060	0.061	0.062	0.064
0	0.9	0.078	0.080	0.079	0.084
0.4	0	0.050	0.052	0.051	0.051
0.4	0.2	0.049	0.051	0.051	0.051
0.4	0.4	0.049	0.051	0.051	0.052
0.4	0.8	0.061	0.063	0.063	0.065
0.4	0.9	0.079	0.081	0.078	0.083
0.8	0	0.050	0.051	0.051	0.051
0.8	0.2	0.049	0.051	0.051	0.051
0.8	0.4	0.050	0.049	0.050	0.051
0.8	0.8	0.061	0.063	0.063	0.066
0.8	0.9	0.079	0.081	0.078	0.085

▲ *The data is generated according to $y_t = \alpha + \beta x_t + u_{1,t}$, $x_t = x_{t-1} + u_{2,t}$, $t = 1,..,T$, where the errors are $u_{1,t} = \rho u_{1,t-1} + e_t + \lambda e_{t-1}$, and $\{e_t\}$ and $\{u_{2,t}\}$ are i.i.d. $N(0,1)$.*

Table 4: Asymptotic critical values for W, Daniell kernel, $b = 0.2$[✠]

		$f(t) = \alpha$					$f(t) = \alpha_0 + \alpha_1 t$		
k	q	90%	95%	99%	k	q	90%	95%	99%
1	1	6.55	10.87	27.22	1	1	9.15	15.66	41.69
2	1	8.33	13.93	36.34	2	1	11.41	19.45	51.21
2	2	16.08	25.22	60.35	2	2	21.95	34.90	87.10
3	1	10.52	17.86	45.66	3	1	14.08	23.90	63.52
3	2	20.60	32.27	75.44	3	2	27.09	44.06	108.15
3	3	29.75	45.63	101.72	3	3	39.70	61.91	149.02
4	1	13.58	23.03	59.88	4	1	17.76	30.67	78.51
4	2	25.60	40.79	96.74	4	2	33.88	53.68	135.34
4	3	36.88	56.64	132.37	4	3	48.97	75.94	182.90
4	4	48.26	72.67	172.21	4	4	64.19	97.55	230.02
5	1	16.01	27.42	68.28	5	1	20.88	35.57	95.07
5	2	30.82	49.26	116.74	5	2	39.38	63.07	158.02
5	3	45.33	69.01	161.09	5	3	58.32	92.10	220.74
5	4	59.11	89.58	196.10	5	4	76.47	117.58	281.99
5	5	72.90	109.51	240.71	5	5	94.45	143.81	345.34
6	1	19.20	32.98	85.75	6	1	24.32	41.19	109.66
6	2	37.02	58.84	141.50	6	2	46.99	75.57	182.81
6	3	54.08	83.63	188.15	6	3	68.50	106.90	253.19
6	4	71.80	108.73	236.97	6	4	89.74	140.09	319.27
6	5	88.43	133.89	286.73	6	5	112.50	172.59	384.05
6	6	104.74	159.14	335.96	6	6	133.69	204.95	455.74

[✠] These are critical calues for testing the hypothesis $H_0 : R\beta = \beta_0$, where $rank(R) = q$ with the Wald test using the Daniell kernel in the model: $y = f(t) + X\beta + u$, where $X : T \times$ k is integrated of order 1. The critical values were calculated using $N(0,1)$ i.i.d. random variables to approximate the Wiener processes in the distributions, $50,000$ replications were used, and the integrals were computed as averages over $1,000$ equally spaced points.

Table 5: Finite Sample Size Comparison♣

	$T = 50$			$T = 100$		
p	Dan 0.2	HAC	HAC-pw	Dan 0.2	HAC	HAC-pw
1	0.11	0.46	0.30	0.07	0.42	0.18
2	0.15	0.50	0.34	0.08	0.43	0.19
3	0.18	0.54	0.39	0.10	0.45	0.21
4	-	-	-	0.11	0.47	0.24
5	-	-	-	0.14	0.49	0.26

♣These are 5% rejection probabilities under the null for the hypothesis $H_0 : R\beta = \beta_0$ in the model: $y = \alpha + \beta X + u$, where $X : T \times 1$ and $\Phi(L) u_t = e_t$, where $\{e_t\}$ is $i.i.d.$ $N(0, \Sigma_e)$,
$$\Phi = \begin{bmatrix} -.103 & -.039 \\ -.062 & .643 \end{bmatrix} \text{ and } \Sigma_e = 0.01 \times \begin{bmatrix} .951 & .499 \\ .499 & 1.374 \end{bmatrix}.$$

Table 6: Finite Sample Power, T=50[§]

c\p	HAC 1	HAC 2	HAC 3	HAC-pw 1	HAC-pw 2	HAC-pw 3	Dan-0.2 1	Dan-0.2 2	Dan-0.2 3
0	0.46	0.50	0.54	0.30	0.34	0.39	0.11	0.15	0.18
2	0.50	0.52	0.55	0.33	0.36	0.40	0.13	0.16	0.20
4	0.55	0.56	0.58	0.38	0.39	0.43	0.16	0.18	0.22
6	0.60	0.60	0.61	0.43	0.44	0.47	0.19	0.22	0.25
8	0.66	0.65	0.65	0.50	0.49	0.51	0.24	0.26	0.30
10	0.72	0.70	0.70	0.57	0.55	0.56	0.30	0.31	0.33
12	0.77	0.75	0.74	0.63	0.61	0.61	0.35	0.36	0.38
14	0.82	0.78	0.77	0.69	0.67	0.66	0.41	0.41	0.43
16	0.85	0.82	0.80	0.74	0.71	0.70	0.46	0.47	0.48
18	0.87	0.85	0.83	0.78	0.75	0.73	0.52	0.52	0.52
20	0.90	0.88	0.86	0.81	0.79	0.77	0.57	0.56	0.56
22	0.92	0.89	0.88	0.85	0.82	0.80	0.62	0.61	0.60
24	0.93	0.91	0.89	0.87	0.85	0.83	0.66	0.65	0.64
26	0.95	0.93	0.91	0.89	0.87	0.85	0.70	0.68	0.68
28	0.96	0.94	0.92	0.91	0.89	0.87	0.73	0.72	0.71
30	0.97	0.95	0.93	0.93	0.91	0.89	0.77	0.75	0.74
32	0.97	0.96	0.94	0.94	0.92	0.90	0.79	0.78	0.77
34	0.98	0.96	0.95	0.95	0.93	0.92	0.82	0.80	0.79
36	0.98	0.97	0.95	0.96	0.95	0.93	0.84	0.83	0.81
38	0.99	0.97	0.96	0.97	0.95	0.94	0.86	0.85	0.83
40	0.99	0.98	0.97	0.97	0.96	0.95	0.88	0.86	0.85
42	0.99	0.98	0.97	0.98	0.97	0.95	0.89	0.88	0.87
44	0.99	0.98	0.97	0.98	0.97	0.96	0.91	0.89	0.88
46	1.00	0.99	0.98	0.99	0.98	0.97	0.92	0.90	0.89
48	1.00	0.99	0.98	0.99	0.98	0.97	0.93	0.91	0.90
50	1.00	0.99	0.98	0.99	0.98	0.97	0.94	0.92	0.91

[§]These are 5% rejection probabilities under $H_a : R\beta = c$ for the hypothesis $H_0 : R\beta = 0$ in the model: $y = \alpha + \beta X + u$, where $X : 50 \times 1$ and $\Phi(L) u_t = e_t$, where $\{e_t\}$ is $i.i.d.$ $N(0, \Sigma_e)$ $t = 1, .., 50$, $\Phi = \begin{bmatrix} -.103 & -.039 \\ -.062 & .643 \end{bmatrix}$ and $\Sigma_e = 0.01 \times \begin{bmatrix} .951 & .499 \\ .499 & 1.374 \end{bmatrix}$.

Table 7: Finite Sample Power, T=100[★]

	HAC					HAC-pw					Dan-2				
c\p	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0	.42	.43	.45	.47	.49	.18	.19	.21	.23	.26	.08	.09	.10	.11	.13
2	.45	.45	.46	.48	.49	.21	.21	.23	.24	.26	.09	.10	.11	.12	.14
4	.50	.50	.49	.51	.52	.25	.25	.25	.27	.28	.11	.11	.13	.14	.15
6	.56	.55	.54	.54	.55	.31	.29	.30	.31	.32	.15	.15	.15	.16	.18
8	.62	.60	.59	.59	.59	.37	.35	.35	.36	.36	.19	.18	.19	.19	.21
10	.68	.65	.64	.63	.63	.44	.41	.40	.40	.41	.23	.23	.23	.23	.24
12	.74	.70	.69	.68	.68	.51	.48	.46	.46	.46	.29	.28	.28	.28	.29
14	.79	.75	.73	.72	.71	.57	.54	.52	.51	.51	.34	.33	.33	.33	.33
16	.83	.80	.78	.76	.75	.63	.60	.58	.56	.56	.40	.38	.38	.38	.38
18	.87	.83	.81	.80	.79	.68	.65	.63	.62	.61	.45	.43	.43	.42	.42
20	.89	.87	.84	.83	.82	.73	.70	.68	.66	.65	.50	.48	.47	.47	.47
22	.92	.89	.87	.85	.84	.77	.74	.73	.71	.70	.55	.53	.52	.52	.51
24	.93	.91	.89	.88	.87	.81	.78	.76	.75	.73	.59	.57	.56	.56	.55
26	.95	.93	.91	.90	.89	.84	.82	.80	.78	.77	.63	.61	.60	.60	.59
28	.96	.95	.93	.92	.91	.87	.84	.83	.81	.79	.67	.65	.64	.63	.63
30	.97	.96	.94	.93	.92	.89	.87	.85	.83	.82	.71	.69	.68	.67	.66
32	.98	.97	.95	.94	.93	.91	.89	.87	.86	.84	.75	.73	.71	.70	.70
34	.98	.97	.96	.95	.94	.93	.91	.89	.88	.86	.78	.76	.74	.73	.72
36	.99	.98	.97	.96	.95	.94	.93	.91	.90	.88	.80	.78	.77	.76	.75
38	.99	.98	.97	.96	.96	.96	.94	.93	.91	.90	.83	.81	.79	.79	.78
40	.99	.99	.98	.97	.96	.97	.95	.94	.93	.92	.85	.83	.82	.81	.80
42	.99	.99	.98	.98	.97	.97	.96	.95	.94	.93	.87	.85	.84	.83	.82
44	1.0	.99	.99	.98	.97	.98	.97	.96	.95	.94	.88	.87	.86	.85	.84
46	1.0	.99	.99	.98	.98	.98	.97	.97	.96	.95	.90	.88	.87	.86	.85
48	1.0	.99	.99	.99	.98	.99	.98	.97	.96	.95	.91	.90	.89	.88	.87
50	1.0	.99	.99	.99	.98	.99	.98	.98	.97	.96	.93	.91	.90	.89	.88

[★]These are 5% rejection probabilities under $H_a : R\beta = c$ for the hypothesis $H_0 : R\beta = 0$ in the model: $y = \alpha + \beta X + u$, where $X : 100 \times 1$ and $\Phi(L) u_t = e_t$, where $\{e_t\}$ is $i.i.d.$ $N(0, \Sigma_e)$ $t = 1, .., 100$, $\Phi = \begin{bmatrix} -.103 & -.039 \\ -.062 & .643 \end{bmatrix}$ and $\Sigma_e = 0.01 \times \begin{bmatrix} .951 & .499 \\ .499 & 1.374 \end{bmatrix}$.

Table 8: Money-demand estimation.

	θ_y	θ_r
$DOLS$ estimates	0.427	-0.045
$W(\theta_y = 1$ or $\theta_r = -0.05)$, Dan-0.02	271.6	144.2
$W(\theta_y = 1$ and $\theta_r = -0.05)$, Dan-0.02	271.6	

Figure 1: Local Asymptotic Power Daniell kernel

Figure 2: Finite Sample Size, T = 50, Daniell Kernel, 5% nominal size

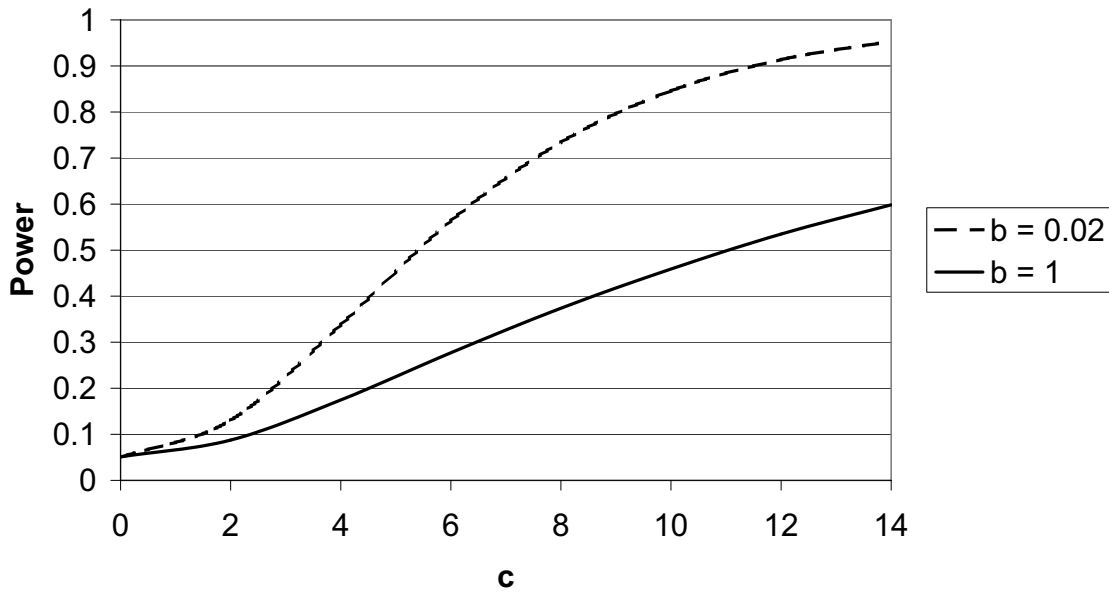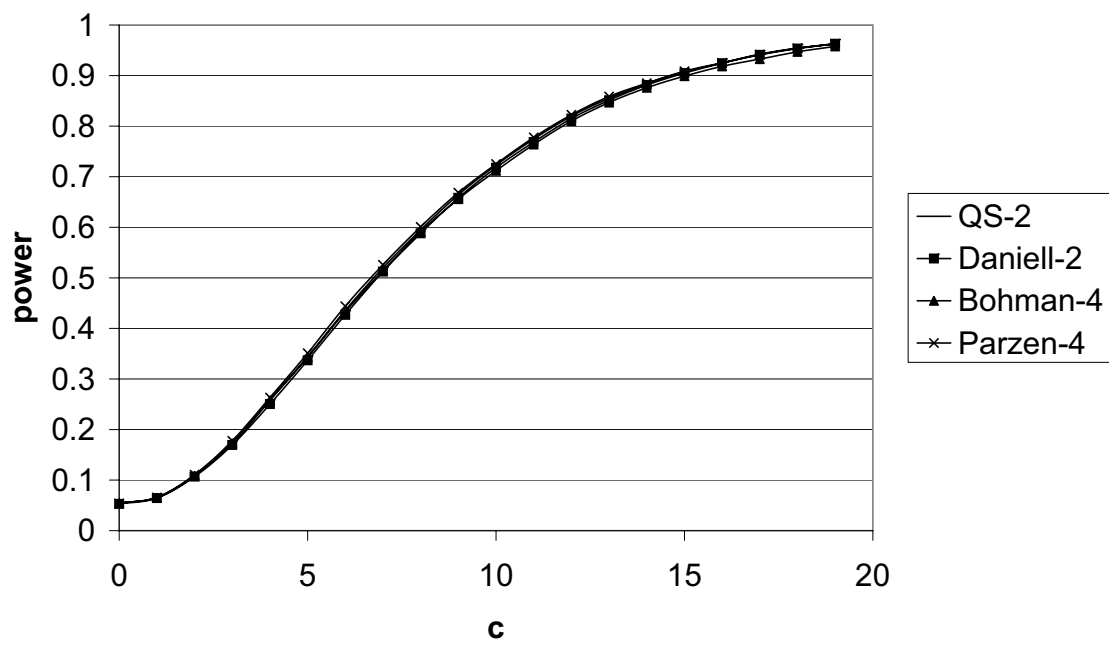**Figure 3: Local Asymptotic Power Daniell kernel**

Figure 4: Local Asymptotic Power

Figure 5: Finite Sample Power iid errors, T=50