Consistent Monitoring of Cointegrating Relationships: The US Housing Market and the Subprime Crisis

Martin Wagner & Dominik Wied*

Faculty of Statistics, Technical University Dortmund

Institute for Advanced Studies, Vienna

Bank of Slovenia, Ljubljana

&

Institute for Econometrics and Statistics, University of Cologne

Faculty of Statistics, Technical University Dortmund

Abstract

We propose a consistent monitoring procedure to detect a structural change from a cointegrating relationship to a spurious relationship. The procedure is based on residuals from modified least squares estimation, using either Fully Modified, Dynamic or Integrated Modified OLS. It is inspired by Chu et al. (1996) in that it is based on parameter estimation on a pre-break "calibration" period only, rather than being based on sequential estimation over the full sample. We investigate the asymptotic behavior of the procedure under the null, for (fixed and local) alternatives and in case of parameter changes. We finally use our approach to monitor two fundamentals driven US housing prices cointegrating relationships over the period 1976:Q1–2010:Q4 using the same data as Anundsen (2015). Depending upon relationship considered and estimation method used, a breakpoint is detected as early as 2003:Q2, i.e., well before US housing prices started to fall in 2007.

Keywords: Cointegration, Monitoring, Structural Change, US Housing Market

JEL Codes: C22, C32, C52, R30

*Technical University Dortmund, Faculty of Statistics, Vogelpothsweg 87, D-44227 Dortmund, Germany. E-Mail: mwagner@statistik.tu-dortmund.de, Phone: +49 231 755 3174 (M. Wagner). E-Mail: wied@statistik.tu-dortmund.de, Phone: +49 231 755 5419 (D. Wied).

1. Introduction

It is common practice in time series econometrics to investigate the cointegration properties of time series and a plethora of tests for cointegration is available. In relation to this practice, however, it may well be reasonable to investigate whether the cointegration behavior of time series changes over time. In particular, a cointegrating relationship between several time series may break down and turn into a spurious relationship.[1] Examples where one may be concerned about this type of structural change include deviations from purchasing power parity after a period of international economic stability, asset price bubbles during times of financial crises (see, e.g., Phillips et al., 2011 and Astill et al., 2015) or deviations from economic theory based equilibrium relationships between housing prices and fundamentals during the US subprime crisis (see Anundsen, 2015). In many of these situations, the date of a potential breakpoint is not known a priori and it is desirable to have a statistically valid procedure that discovers changes well without that information.[2]

Our monitoring procedure is inspired by the monitoring procedure for linear regression models of Chu et al. (1996) in that parameter estimation, for estimating trend, slope and long-run variance parameters, is based solely on a "calibration" period at the beginning of the sample that is known or assumed to be free of structural change.[3] At each time point after the calibration period, one merely needs data up to that time point for calculating the detector. Based on the parameter estimates, the residuals of an estimated cointegrating relationship are the key ingredient for the monitoring procedure. The monitoring procedure is based on sequentially computing the differences of scaled

[1] As discussed below, the approach also allows to monitor whether the coefficients of a cointegrating relationship change over time.

[2] In this paper we revisit the US housing data investigated in detail by Anundsen (2015), who uses recursive VAR estimation and cointegration testing to determine whether the housing market at some point enters a bubble-type period. Whilst this approach is, of course, subject to all problems related with recursive estimation and subsequent multiple testing, our approach allows for asymptotically consistent monitoring of potential structural change in the relationships away from cointegration.

[3] This approach to monitoring, based on estimation in a pre-break sample period only, has been extended to the multivariate linear regression case by Groen et al. (2013) and has been applied to monitor changes in the correlation structure by Wied and Galeano (2013).

partial sums of squared residuals over the growing monitoring period and the calibration period.[4] The idea of considering partial sums of squared residuals to test for stationarity or cointegration goes back to the KPSS-test (see Kwiatkowski et al., 1992 and Shin, 1994). The detection time, defined formally in the following section, serves as an immediately available estimate of the break-point. In order to obtain nuisance parameter free limiting distributions of the test statistics when applying the principle to monitor cointegrating relationships, parameter estimation on the calibration sample is performed using any of the available modified least squares estimators that lead to nuisance parameter free limiting distributions of the parameters of the cointegrating relationship. In particular we consider here Fully Modified OLS (FM-OLS) of Phillips and Hansen (1990), Dynamic OLS (D-OLS) of Saikkonen (1991), Phillips and Loretan (1993) and Stock and Watson (1993), and Integrated Modified OLS (IM-OLS) of Vogelsang and Wagner (2014).

The asymptotic properties of the monitoring procedure are derived under both the null as well as under (fixed and local) alternatives and for the case of breaks in trend or slope parameters. In all of these cases, our procedure is consistent for fixed alternatives. Based on the asymptotic results, the performance of the proposed methods is investigated in more detail by means of local asymptotic power (LAP) analysis. Furthermore, with finite sample simulations we consider empirical size and power for a variety of scenarios. In addition to studying power against the alternative of integrated behavior, we also assess the performance in case of parameter changes in the trend and/or slope coefficients. We also assess the estimated detection times in the finite sample simulations.

The literature does not offer many solutions for detecting structural change away from a cointegrating relationship or parameter changes in a cointegrating relationship. There are some papers on testing for cointegration (on the full sample) in the presence of breaks in, e.g., the deterministic trend parameters (Johansen et al., 2000) or in the unconditional variances of the errors (Cavaliere and Taylor, 2006). Moreover, there is

[4]This, of course, immediately implies consistency of the procedure against any "more explosive" alternative, like higher order integration or explosive behavior. Furthermore, one can also easily show consistency against fractionally integrated alternatives, compare Wagner and Wied (2015, Remark 4).

a small literature on testing for cointegration breakdown (see, e.g., Andrews and Kim, 2006). However, as the test statistics in these approaches are calculated based on the full sample, respectively, the procedures cannot be used for monitoring purposes. Chen et al. (2009) and Wang et al. (2014) use cointegration methods for monitoring non-stationary processes in industrial applications; both of these approaches are ad-hoc procedures based on well-known statistical tests but without well understood asymptotic properties. One procedure that is more closely related to ours is Steland and Weidauer (2013), see also Steland (2007), who consider monitoring the null hypothesis of a spurious regression against the alternative of a cointegrating relationship from some point onwards. Thus, they consider the opposite pair of null and alternative hypothesis compared to the present paper. It has to be noted (see Chen et al., 2010) that in such a case a consistent monitoring procedure has to be based on moving residuals rather than on expanding sets of residuals. However, the procedure in Steland and Weidauer (2013) is not based on moving residuals and is therefore not consistent. For consistently monitoring a change from spurious regression to cointegration see Sakarya et al. (2015).

We revisit the application from Anundsen (2015), who considers two (slightly different) cointegrating relationships between housing prices and fundamentals that arise from equilibrium and no-arbitrage considerations. By using recursive estimation and cointegration testing in a VAR framework, Anundsen (2015) finds evidence for a breakdown of these relationships as early as 2000:Q4 for one of the two relationships. We, in some sense, confirm his results and make them at the same time more precise by our asymptotically founded monitoring procedure for cointegration. Depending upon specification of the equilibrium relationship and estimation method used, we detect a breakpoint between 2003:Q2 and 2007:Q3. Clearly, our detected break-points are later than what is found by Anundsen (2015), but delay is the price to be paid for having an asymptotically consistent procedure.[5]

[5]Anundsen (2015), of course, contains also other interesting aspects in the empirical analysis. E.g., he constructs a simple bubble-indicator and analyzes the relationships of this indicator to other real and financial economic series. Our approach adds to this type of analysis by offering a consistent approach of detecting a structural change in the cointegrating dimension.

2. Monitoring Cointegration

Throughout the paper, we consider – under the null hypothesis – the setup

$$y_t = D_t'\theta_D + X_t'\theta_X + u_t \quad (1)$$

$$X_t = X_{t-1} + v_t, \quad (2)$$

for $t = 1, \dots, T$, where y_t is scalar, $D_t \in \mathbb{R}^p$ is a deterministic trend function and X_t is a k-dimensional vector. The case of $\dim(X_t) = 0$ leads to a procedure for monitoring stationarity, which is elaborated on in detail in our more extensive working paper Wagner and Wied (2015), but is ignored in the present paper for brevity. For the trend function, we make the following assumption:

Assumption 1. *There exists a sequence of $p \times p$ scaling matrices G_D and a p-dimensional vector of functions $D(z)$, with $0 < \int_0^s D(z)D(z)'dz < \infty$ for $0 \leq s \leq 1$, such that for $0 \leq s \leq 1$*

$$\lim_{T\to\infty} \sqrt{T} G_D^{-1} D_{[sT]} = D(s), \quad (3)$$

with $[sT]$ denoting the integer part of sT.

If, e.g., $D_t = (1, t, t^2, ..., t^{p-1})'$, then $G_D = \text{diag}(T^{1/2}, T^{3/2}, T^{5/2}, \dots, T^{p-1/2})$ and $D(z) = (1, z, z^2, ..., z^{p-1})'$.

With respect to the joint error process $\{\eta_t\}_{t\in\mathbb{Z}}$, with $\eta_t = (u_t, v_t')'$, we simply posit the required functional central limit theorems (FCLTs):

Assumption 2.

(a) The stationary process $\{\eta_t\}_{t\in\mathbb{Z}}$ fulfills

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \eta_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \Rightarrow \Omega^{1/2} W(s) = B(s), \quad (4)$$

with $W(s) = [W_{u\cdot v}(s), W_v(s)']'$ a (k+1)-dimensional vector of standard Brownian *motions and* $0 < \Omega < \infty$*, with*

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} := \sum_{j=-\infty}^{\infty} \mathbb{E}(\eta_t \eta_{t-j}'). \tag{5}$$

(b) Denoting with $S_t^\eta = \sum_{j=1}^{t} \eta_j$ *it holds that*

$$\frac{1}{T} \sum_{t=1}^{[sT]} S_t^\eta \eta_t' \Rightarrow \int_0^s B(r) dB(r)' + \Delta, \tag{6}$$

with

$$\Delta := \sum_{j=0}^{\infty} \mathbb{E}(\eta_{t-j} \eta_t') \tag{7}$$

partitioned similarly as Ω*.*

(c) The convergence results in (a) and (b) hold jointly.

The results posited in Assumption 2 are standard and typical in the cointegration literature and are implied by a variety of underlying primitive assumptions (for some early contributions see, e.g., Phillips and Hansen, 1990; Phillips and Durlauf, 1986; Stock, 1987). For our purposes it is convenient to use

$$\Omega^{1/2} = \begin{bmatrix} \omega_{u \cdot v} & \lambda_{uv} \\ \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix}, \tag{8}$$

where $\omega_{u \cdot v}^2 := \Omega_{uu} - \Omega_{uv} \Omega_{vv}^{-1} \Omega_{vu}$ and $\lambda_{uv} := \Omega_{uv} (\Omega_{vv}^{1/2})^{-1}$.

The assumption $\Omega_{vv} > 0$ excludes cointegration amongst the regressors and is typically assumed for the modified OLS estimation techniques available, including Fully Modified OLS (FM-OLS) of Phillips and Hansen (1990), Dynamic OLS (D-OLS) of Saikkonen (1991), and Integrated Modified OLS (IM-OLS) of Vogelsang and Wagner (2014).

In formal terms, our null and alternative hypotheses are given by

$$H_0: \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} u_t \;\Rightarrow\; \omega W(s), \;\text{ for all }\; 0 \le s \le 1 \text{ and } 0 < \omega < \infty \tag{9}$$

$$H_1: \frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]} u_t \;=\; O_p(T) \text{ and } \frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]} u_t \neq o_p(T) \tag{10}$$
$$\text{for some }\; 0 < m \le r < 1 \text{ and for all } r < s.$$

This means that, under the alternative, we consider the situation that there exists some time point $[rT]$ such that the process $\{u_t\}_{t\in\mathbb{Z}}$ behaves like an I(1) process from $[rT]+1$ onwards. Thus, under the alternative, Assumption 2 is violated from $[rT]+1$ onwards in a specific way. We define a process $\{x_t\}_{t\in\mathbb{Z}}$ to be an I(1) process, in accordance with our I(0) definition, if

$$\frac{1}{\sqrt{T}} x_{[sT]} \Rightarrow \omega W(s), \tag{11}$$

for $0 \le s \le 1$, some $0 < \omega < \infty$ and where $W(s)$ denotes again standard Brownian motion. It is clear that the (partial) sum process of an I(0) process is an I(1) process.

The formulation of the test problem is to be understood in the sense that also under the alternative, the process $\{\eta_t\}_{t\in\mathbb{Z}}$ fulfills Assumption 2 up to $[rT] \ge [mT]$. Note that we want to detect a change from I(0) to I(1) behavior in $\{u_t\}_{t\in\mathbb{Z}}$ under the alternative that occurs at time point $[rT]$ with $m \le r < 1$, i.e., a change that occurs only *after some pre-break sample fraction* of length $[mT]$ with $0 < m < 1$.

As we shall see below, we need such a pre-break sample fraction m in particular in order to consistently estimate several quantities required to obtain a null limiting distribution of our detector that is a function only of the included deterministic components and standard Brownian motions and for which consequently critical values can be simulated. These quantities include the long-run variance $\omega^2_{u\cdot v}$, the trend parameters θ_D and the slope parameters θ_X.

The idea underlying our monitoring procedure is to consider appropriate estimates of the errors u_t and see whether they become "too large" over time. It is well-known that OLS estimation of $\theta = [\theta_D', \theta_X']'$ in (1) is consistent in case of cointegration, but that in general the limiting distribution of the OLS estimator depends on second order bias terms, which render asymptotic standard inference based on the OLS estimator infeasible. This problem occurs in particular when the regressors are not strictly exogenous, i.e., when the matrix Ω is not block-diagonal.[6] The mentioned modified OLS estimators lead to limiting distributions of the parameters that are proportional to functionals of standard Brownian motions (which depend upon D_t and the number of integrated regressors k) also in case of endogeneity. For brevity we abstain from explaining these well-known procedures here and just consider the residuals obtained from these estimation procedures as input in our monitoring procedure.[7]

We first consider the residuals of FM-OLS estimation, denoting the dependent variable used in FM-OLS by $y_t^+ := y_t - \Delta X_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}$:[8]

$$\begin{aligned} \hat{u}_{t,m}^+ &:= y_t^+ - D_t'\hat{\theta}_{D,m} - X_t'\hat{\theta}_{X,m} \\ &= y_t - \Delta X_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} - D_t'\hat{\theta}_{D,m} - X_t'\hat{\theta}_{X,m} \\ &= u_t - v_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} - D_t' \left(\hat{\theta}_{D,m} - \theta_D \right) - X_t' \left(\hat{\theta}_{X,m} - \theta_X \right), \end{aligned} \tag{12}$$

where $\hat{\theta}_{D,m}$ and $\hat{\theta}_{X,m}$ denote the FM-OLS coefficient estimates and $\hat{\Omega}(= \hat{\Omega}_m)$ denotes the long-run variance estimate, all computed from the pre-break sample $1, \ldots, [mT]$.[9]

[6] In case of strict exogeneity, asymptotically valid inference can be based on the OLS estimates if serial correlation in $\{u_t\}_{t \in \mathbb{Z}}$ is handled appropriately using consistent long-run variance estimation.

[7] Only the less well-known IM-OLS estimator is briefly discussed below.

[8] Clearly, the construction of y_t^+ using ΔX_t implies that the sample size available is $t = 2, \ldots, T$. To simplify notational flow we nevertheless consider the sums below starting at $t = 1$, by setting $\hat{u}_{1,m}^+ = 0$, with $\hat{u}_{t,m}^+$ defined in the next equation. In the same way, D-OLS estimation leads to a reduced effective sample size due to the usage of leads and lags.

[9] In case the procedure is implemented using the D-OLS estimator, the residuals are defined (using the same notation for the residuals and coefficient estimates) as $\hat{u}_{t,m}^+ := y_t - D_t'\hat{\theta}_{D,m} - X_t'\hat{\theta}_{X,m} - \sum_{j=-k_1}^{k_2} \Delta X_{t-j}' \hat{\Theta}_{j,m}$, or equivalently $\hat{u}_{t,m}^+ = u_t - D_t' \left(\hat{\theta}_{D,m} - \theta_D \right) - X_t' \left(\hat{\theta}_{X,m} - \theta_X \right) - \sum_{j=-k_1}^{k_2} \Delta X_{t-j}' \hat{\Theta}_{j,m}$, with $\hat{\theta}_{D,m}$, $\hat{\theta}_{X,m}$ and $\hat{\Theta}_{j,m}$ being the OLS estimates from the regression $y_t = D_t'\theta_D + X_t'\theta_X + \sum_{j=-k_1}^{k_2} \Delta X_{t-j}' \Theta_j + u_t$ estimated using observations $1, \ldots, [mT]$. Whereas in FM-OLS estimation bandwidth and kernel have to be chosen, D-OLS estimation requires choosing the number of leads k_1 and lags k_2. Under appropriate assumptions concerning the asymptotic behavior of lag/lead

By using consistent estimators of the long-run variances, ensured, e.g., by assuming to be in the framework covered by Jansson (2002), we obtain a FCLT for the FM-OLS residuals $\hat{u}^+_{t,m}$. Analogously, if D-OLS is used, we assume lag/lead length choices to be such that estimation is consistent (see, e.g., Kejriwal and Perron, 2008). Throughout the rest of the paper we always assume corresponding choices of bandwidths or leads and lags and do not mention this explicitly.

Lemma 1. *Let the data be generated by* (1) *and* (2) *with Assumptions 1 and 2 in place. Then it holds under the null hypothesis and for* $m \leq s \leq 1$ *for* $T \to \infty$ *for FM-OLS and D-OLS that*

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \hat{u}^+_{t,m} &\Rightarrow \omega_{u\cdot v} \left(W_{u\cdot v}(s) - \int_0^s J(z)' dz \left(\int_0^m J(z) J(z)' dz \right)^{-1} \int_0^m J(z) dW_{u\cdot v}(z) \right) \\ &=: \omega_{u\cdot v} \widehat{W}_{u\cdot v}(s) \end{aligned} \quad (13)$$

with $J(s) := [D(s)', W_v(s)']'$.

Clearly, the process $\widehat{W}_{u\cdot v}(s)$ depends upon D_t, the number of integrated regressors k and the pre-break fraction m, with these dependencies neglected for notational brevity henceforth.

Given the FCLT (13) for the partial sum process of the modified residuals, the detector for cointegration, using either the FM-OLS or the D-OLS residuals, is defined by

$$\widehat{H}^{m,+}(s) := \frac{1}{\hat{\omega}^2_{u\cdot v}} \left(\frac{1}{T} \sum_{i=[mT]+1}^{[sT]} \left(\frac{1}{\sqrt{T}} \widehat{S}^+_i \right)^2 - \frac{1}{T} \sum_{i=1}^{[mT]} \left(\frac{1}{\sqrt{T}} \widehat{S}^+_i \right)^2 \right). \quad (14)$$

Here, $\widehat{S}^+_i = \sum_{t=1}^{i} \hat{u}^+_{t,m}$ and the scaling factor is now a consistent estimator $\hat{\omega}^2_{u\cdot v} = \hat{\Omega}_{uu} - \hat{\Omega}_{uv} \hat{\Omega}^{-1}_{vv} \hat{\Omega}_{vu}$ of the conditional long-run variance $\omega^2_{u\cdot v}$.[10]

choices, the D-OLS residuals fulfill the same FCLT as the FM-OLS residuals. Asymptotically, therefore the usage of either estimator leads to the same monitoring procedure.

[10]We have also investigated other variants of our detector such as a ratio statistic, $\frac{\sum_{i=[mT]+1}^{[sT]} \left(\frac{1}{\sqrt{T}} \widehat{S}^+_i \right)^2}{\sum_{i=1}^{[mT]} \left(\frac{1}{\sqrt{T}} \widehat{S}^+_i \right)^2}$, a detector based on moving windows, $\frac{1}{\hat{\omega}^2_{u\cdot v}} \frac{1}{T} \sum_{i=[sT]-[\epsilon T]}^{[sT]} \left(\frac{1}{\sqrt{T}} \widehat{S}^+_i \right)^2$, a detector based on recursive residuals

All long-run variances and covariances required in the procedure (both for modified OLS parameter estimation as well as for scaling the detector) are based on the OLS residuals $\hat{u}_{t,m}$ stacked on top of the first differences of the regressors, i.e., upon $\hat{\eta}_t = [\hat{u}_{t,m}, v_t']'$. Again, the OLS estimation from which the parameter estimates and long-run variance estimates are computed uses observations $t = 1, \ldots, [mT]$ only.

Given the definition of the detector for cointegration (14), the first result to be established is the asymptotic distribution of the detector under the null hypothesis. Its straightforward proof (as well as the other subsequent proofs in the paper concerning the null hypothesis) follows directly from known convergence results and can be found in detail in Wagner and Wied (2015, Appendix A).

Lemma 2. *Let the data be generated by* (1) *and* (2) *with Assumptions 1 and 2 in place and let $\hat{\omega}^2_{u\cdot v}$ denote a consistent long-run variance estimator. Then it holds under the null hypothesis and for $m \leq s \leq 1$ for $T \to \infty$ for FM-OLS and D-OLS that*

$$\widehat{H}^{m,+}(s) \;\Rightarrow\; \int_m^s \widehat{W}^2_{u\cdot v}(z)dz - \int_0^m \widehat{W}^2_{u\cdot v}(z)dz =: \widehat{\mathcal{H}}^{m,+}(s) \tag{15}$$

We define the detection time $\tau_m(H^m, g, c)$, often only written as τ_m if the context is clear, as

$$\tau_m \;:=\; \min_{s:[mT]+1\leq[sT]\leq T} \left\{ \left| \frac{H^m(s)}{g(s)} \right| > c \right\}, \tag{16}$$

i.e., the null hypothesis is declared rejected when the standardized detector, $\frac{H^m(s)}{g(s)}$, exceeds a critical value c in absolute value for the first time. In case that $\left|\frac{H^m(s)}{g(s)}\right| \leq c$ for all $m \leq s \leq 1$ we write $\tau_m = \infty$. Thus, a finite value of τ_m indicates a rejection of the null and at the same time gives information about the potential break point.

The properties of such a monitoring procedure hinge, by construction, upon the threshold function $g(s)$ and the constant c, which itself depends upon the function $g(s)$. For sim-

or a combination of all. Finite sample simulations revealed no considerable differences in terms of size and power, however.

plicity we only consider weighting functions that are continuous, bounded and positive throughout this paper, compare Aue et al. (2012, Assumption 3.6). Note, for completeness, that Chu et al. (1996) consider more general weighting functions in a much simpler setting. Weighting function and critical value have to be chosen to ensure that under the null hypothesis

$$\begin{aligned}
\lim_{T\to\infty} \mathbb{P}(\tau_m < \infty) &= \lim_{T\to\infty} \mathbb{P}\left(\min_{s:[mT]+1\leq[sT]\leq T} \left\{ \left| \frac{H^m(s)}{g(s)} \right| > c \right\} < \infty \right) \\
&= \lim_{T\to\infty} \mathbb{P}\left(\sup_{s:[mT]+1\leq[sT]\leq T} \left| \frac{H^m(s)}{g(s)} \right| > c \right) \\
&= \mathbb{P}\left(\sup_{m\leq s\leq 1} \left| \frac{\mathcal{H}^m(s)}{g(s)} \right| > c \right) = \alpha,
\end{aligned} \tag{17}$$

with α denoting the chosen significance level. Deriving such a result can be based on a functional central limit theorem for $H^m(s)$, since we consider only continuous $g(s)$. The choice of $g(s)$ is in the words of Chu et al. (1996, p. 1052) "often dictated by mathematical convenience rather than optimality, since crossing probabilities for an arbitrary boundary are analytically intractable in general". The problem is typically transformed into delivering simple stopping times. This is due to the fact, already used in (17), that the event $\{\inf_{m\leq s\leq 1}\{|\mathcal{H}^m(s)/g(s)| > c\} \neq \infty\}$ is equal to the event $\{\sup_{m\leq s\leq 1}\{|\mathcal{H}^m(s)/g(s)|\} > c\}$. Therefore, the sup-functional in the context of *testing* can be considered as the natural equivalent to the stopping time based on the inf-functional in the context of *monitoring*. This is also the approach pursued in Chu et al. (1996) for parameter change in linear regression models. Clearly, the choice of the weighting function impacts the performance of the monitoring procedure and has to *combine* two opposing goals of a monitoring procedure: (a) small size distortions under the null and (b) small delays under the alternative, i.e., detection of a break soon after the break. The discussion in Chu et al. (1996, Section 3) makes clear that it will in general be impossible to derive analytically tractable optimal weighting functions, e.g., with respect to minimal expected delay whilst asymptotically controlling size.[11]

Alternatively to FM- and D-OLS, one can also base the cointegration monitoring procedure on the residuals of the recently proposed Integrated Modified OLS (IM-OLS) estimator of Vogelsang and Wagner (2014). A potential advantage of the IM-OLS estimator compared to FM-OLS and D-OLS is that for parameter estimation no kernel and bandwidth or lead and lag choices are required. The IM regression is given by

$$S_t^y = S_t^{D\prime}\theta_D + S_t^{X\prime}\theta_X + X_t'\varphi + S_t^u, \tag{18}$$

with $S_t^y = \sum_{j=1}^t y_j$ denoting the partial sums, and similar definitions of S_t^D, S_t^X and S_t^u. We denote the corresponding OLS residuals, with estimation based upon the pre-break sample $1,\ldots,[mT]$ by (using the same notation for the coefficient estimates as before)

$$\begin{aligned} \hat{S}_{t,m}^u &:= S_t^y - S_t^{D\prime}\hat{\theta}_{D,m} - S_t^{X\prime}\hat{\theta}_{X,m} - X_t'\hat{\varphi}_m \\ &= S_t^u - X_t'\hat{\varphi}_m - S_t^{D\prime}(\hat{\theta}_{D,m} - \theta_D) - S_t^{X\prime}(\hat{\theta}_{X,m} - \theta_X) \end{aligned} \tag{19}$$

Under the assumptions stated the following FCLT holds:

Lemma 3. *Let the data be generated by* (1) *and* (2) *with Assumptions 1 and 2 in place. Then it holds for* $T \to \infty$ *that*

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=2}^{[sT]} \Delta\hat{S}_{t,m}^u &\Rightarrow \omega_{u\cdot v}\left(W_{u\cdot v}(s) - f(s)'\left(\int_0^m f(z)f(z)'dz\right)^{-1}\int_0^m [F(m)-F(z)]\,dW_{u\cdot v}(z)\right) \\ &=: \omega_{u\cdot v}\tilde{P}_m(s), \end{aligned} \tag{20}$$

where $f(s) := [\int_0^s D(z)'dz, \int_0^s W_v(z)'dz, W_v(s)']'$ *and* $F(s) := \int_0^s f(z)dz$.

[11] Aue et al. (2009) derive the limiting distribution of the delay time for a one-time parameter change in a linear regression model with stationary regressors for a simple class of weighting functions depending on a single tuning parameter. To the best of the authors' knowledge, no results of this kind are available in a unit root or cointegration setting.

Based upon the above result, the IM-OLS based detector is defined as:

$$\hat{I}^m(s) \ := \ \frac{1}{\hat{\omega}^2_{u\cdot v}}\left(\frac{1}{T}\sum_{i=[mT]+1}^{[sT]}\left(\frac{1}{\sqrt{T}}\hat{S}^u_{i,m}\right)^2 - \frac{1}{T}\sum_{i=1}^{[mT]}\left(\frac{1}{\sqrt{T}}\hat{S}^u_{i,m}\right)^2\right), \tag{21}$$

where the scaling is, as for the other detectors, based on a consistent estimator of $\omega^2_{u\cdot v}$. Note that the same estimator of $\omega^2_{u\cdot v}$ as for FM-OLS or D-OLS is used, i.e., the estimator based on the OLS residuals $\hat{u}_{t,m}$ stacked on top of the first differences of the regressors. The asymptotic null behavior of the IM-OLS detector is given next.

Lemma 4. *Let the data be generated by* (1) *and* (2) *with Assumptions 1 and 2 in place and let $\hat{\omega}^2_{u\cdot v}$ denote a consistent long-run variance estimator. Then it holds under the null hypothesis and for $m \leq s \leq 1$ for $T \to \infty$ for IM-OLS that*

$$\hat{I}^m(s) \ \Rightarrow \ \int_m^s \tilde{P}_m(z)^2 dz - \int_0^m \tilde{P}_m(z)^2 dz =: \mathcal{I}^m(s) \tag{22}$$

It can be shown for all three detectors that under the null hypothesis for given weighting function $g(s)$, there exist critical values $c = c(\alpha, g(s))$, such that the detection time is finite with probability equal to the pre-specified level α.

Proposition 1. *Let the data be generated by* (1) *and* (2) *with Assumptions 1 and 2 in place, let long-run variance estimation be carried out consistently and assume again that $g(s)$ is continuous with $0 < g(s) < \infty$. Then, under the null hypothesis there exist for any given $0 < \alpha < 1$ critical values $c = c(\alpha, g)$, depending upon estimation method, such that*

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{H}^{m,+}, g, c(\alpha, g)) < \infty) = \alpha, \tag{23}$$

in case that FM-OLS or D-OLS is used, respectively

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{I}^m, g, c(\alpha, g)) < \infty) = \alpha, \tag{24}$$

in case that IM-OLS is used.

Note that for given weighting function $g(s)$ the critical values are identical for the FM-OLS and D-OLS based detectors, but are different for the IM-OLS based detector.

One way of choosing $g(s)$ is a function that is related to $\mathbb{E}(\widehat{\mathcal{H}}^m)$. Doing this, one can show that $g(s) = s^3$ in the intercept case $D_t = 1$ and $g(s) = s^5$ in the linear trend case $D_t = (1, t)'$ is a suitable choice. For these weighting functions and for both FM-/D- and IM-OLS, critical values are provided in Wagner and Wied (2015, Appendix B). They depend upon $0 < m < 1$ and are obtained by simulations based on 1,000,000 replications. To be precise, the functionals of Brownian motions are approximated by the corresponding functions of random walks of length 1,000 generated from i.i.d. standard normal random variables.

It remains to establish the behavior of the monitoring procedure under alternatives. In fact there are three dimensions of structural change against which the procedure has power. First, changes in the behavior of $\{u_t\}_{t\in\mathbb{Z}}$. Second, breaks in the parameters corresponding to the deterministic component. Third, breaks in the slope coefficients corresponding to the integrated regressors. For all three cases we consider fixed and local alternatives.

As fixed alternative for the first case we consider the situation that $\{u_t\}_{t\in\mathbb{Z}}$ changes its behavior from I(0) to I(1) at some point after $[mT]$, i.e., that H_1 as given above holds. To understand the properties of our procedure in more detail we also consider local alternatives of the following form (inspired by Cappuccio and Lubian, 2005). There exists an r, with $m \leq r < 1$ such that for all $t \leq [rT]$ we have $u_t = u_t^0$, while for all $t > [rT]$ it holds that

$$u_t = u_t^0 + \frac{\delta}{T} \sum_{i=[rT]+1}^{t} \xi_i, \tag{25}$$

with $\{u_t^0\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$ independent processes both fulfilling Assumption 2, with long-run variances ω^2 and ω_ξ^2, and $\delta > 0$.[12] I.e., under the considered local alternatives the process $\{u_t\}_{t\in\mathbb{Z}}$ is, from time point $[rT]+1$ onwards, the sum of an I(0) process and an independent I(1) process divided by the sample size. For example, in the case of FM-/D-OLS, the local alternatives imply (for $m \leq r < s \leq 1$):

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \hat{u}_{t,m}^{+} \;\Rightarrow\; \omega_{u\cdot v}\widehat{W}_{u\cdot v}(s) + \delta\omega_\xi \int_r^s (W_\xi(z) - W_\xi(r))dz. \tag{26}$$

Here, integrals (and sums) with the lower boundary larger than the upper are defined to be equal to 0 and $W_\xi(s)$ is standard Brownian motion independent of $W(s)$.

Proposition 2. (Consistency and Local Asymptotic Power)
Let the data be generated by (1) *and* (2) *with Assumption 1 in place and* $\{\eta_t\}_{t\in\mathbb{Z}}$ *fulfilling Assumption 2 until* $[rT]$*, with* $m \leq r < 1$*. Furthermore, assume that long-run variance estimation is performed consistently using observations* $1, \ldots, [mT]$ *and assume again that* $g(s)$ *is continuous with* $0 < g(s) < \infty$*. Denote with* $\widehat{F}^m(s)$ *either* $\widehat{H}^{m,+}(s)$ *or* $\widehat{I}^m(s)$*.*

(a) Let $\{u_t\}_{t\in\mathbb{Z}}$ *be an I(1) process (as specified in* H_1*) from* $[rT]+1$ *onwards. Then the monitoring procedure is consistent, i.e., for any* $c > 0$ *it holds that*

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{F}^m, g, c(\alpha, g)) < \infty) = 1. \tag{27}$$

(b) Let $\{u_t\}_{t\in\mathbb{Z}}$ *be as specified in* (25) *from* $[rT]+1$ *onwards. Then the monitoring procedure has non-trivial local power. This means, for any* $0 < \epsilon \leq 1-\alpha$ *and the critical value* $0 < c = c(\alpha, g) < \infty$ *from Proposition 1 there exists a* $0 < \delta = \delta(c, g) < \infty$ *such that*

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{F}^m, g, c(\alpha, g)) < \infty) \geq 1 - \epsilon. \tag{28}$$

[12] Note that it is sufficient to consider $\{u_t^0\}_{t\in\mathbb{Z}}$ and $\{\xi_t\}_{t\in\mathbb{Z}}$ independent, as asymptotic independence between the two components can always be achieved by redefining the two quantities correspondingly after "orthogonalization".

Proof. We provide the proof of a) in detail; the proof of b) as well as subsequent proofs are analogous and are given in detail in Wagner and Wied (2015, Appendix A). The key to the result is the limiting behavior of the partial sum process of the residuals. We have to distinguish between two cases, FM-OLS and D-OLS estimation on the one hand and IM-OLS estimation on the other.

For FM-OLS, the partial sum process of the residuals is given by (again for $m \leq r < s \leq 1$):

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]}\hat{u}_{t,m}^{+} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\hat{u}_{t,m}^{+} + \frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]}\hat{u}_{t,m}^{+} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]}\hat{u}_{t,m}^{+} + \frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]}u_t - \frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]}v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} - \\ &\quad -\frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]}D_t'\left(\hat{\theta}_{D,m}-\theta_D\right) - \frac{1}{\sqrt{T}}\sum_{t=[rT]+1}^{[sT]}X_t'\left(\hat{\theta}_{X,m}-\theta_X\right). \end{aligned} \tag{29}$$

The first term above converges to $\omega_{u\cdot v}\widehat{W}_{u\cdot v}(r)$, according to Lemma 1 and the second term is $O_p(T)$, since for the sample period considered u_t is an I(1) process. The remaining three terms converge in distribution. Thus, the partial sum process is in this case $O_p(T)$. Similar arguments apply for D-OLS. The argument is analogous for the IM-OLS residual process, i.e., for $\hat{S}_{t,m}^u$, where of course no partial summation is required given that the equation itself is partial summed. □

Proposition 3. (Behavior in Case of Trend Breaks)

Let the data be generated by (1) *and* (2) *with Assumptions 1 and 2 in place. Furthermore, assume that long-run variance estimation is performed consistently using observations* $1, \ldots, [mT]$ *and and assume again that* $g(s)$ *is continuous with* $0 < g(s) < \infty$*. Denote with* $\widehat{F}^m(s)$ *either* $\widehat{H}^{m,+}(s)$ *or* $\widehat{I}^m(s)$*.*

(a) (Fixed Alternative) Let $\theta_D = \theta_{D,1}$ *for* $t = 1, \ldots, [rT]$ *and* $\theta_D = \theta_{D,2}$*, with* $\theta_{D,1} \neq$

$\theta_{D,2}$, from $t = [rT] + 1$ onwards, with

$$\lim_{T\to\infty} \frac{1}{\sqrt{T}} \sum_{t=[rT]+1}^{T} D_t'(\theta_{D,1} - \theta_{D,2}) = \pm\infty, \tag{30}$$

then it holds for any $c > 0$ that

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{F}^m, g, c) < \infty) = 1. \tag{31}$$

(b) *(Local Alternative) Let $\theta_D = \theta_{D,1}$ for $t = 1, \ldots, [rT]$ and $\theta_D = \theta_{D,1} + G_D^{-1\prime}\Delta_\theta$ from $t = [rT] + 1$ onwards, with*

$$\int_r^1 D(z)' dz \Delta_\theta \neq 0, \tag{32}$$

then the monitoring procedure has non-trivial local power. This means, for any $0 < \epsilon \leq 1 - \alpha$ and the critical value $0 < c = c(\alpha, g) < \infty$ from Proposition 1 there exists a $\Delta_\theta = \Delta_\theta(c, g)$ fulfilling (32) such that

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{F}^m, g, c(\alpha, g)) < \infty) \geq 1 - \epsilon. \tag{33}$$

There is the additional possibility of breaks in the slope coefficients θ_X, which are in a sense equivalent to changes in the behavior of $\{u_t\}_{t\in\mathbb{Z}}$. Consider for simplicity the case $\theta_X = \theta_{X,1}$ for $t = 1, \ldots, [rT]$ and $\theta_X = \theta_{X,2}$, with $\theta_{X,1} \neq \theta_{X,2}$, for $t = [rT] + 1, \ldots, T$. In this case consider:

$$\begin{aligned} y_t &= D_t'\theta_D + X_t'\theta_{X,2} + u_t \qquad (34) \\ &= D_t'\theta_D + X_t'\theta_{X,1} + X_t'(\theta_{X,2} - \theta_{X,1}) + u_t. \end{aligned}$$

Clearly, this implies that in the residual process $\hat{u}_{t,m}^+$, starting from $[rT]$ onwards an integrated process given by $X_t'(\theta_{X,2} - \hat{\theta}_{X,1,m})$ is present. This component remains present as an I(1) process also in the limit due to consistency of $\hat{\theta}_{X,1,m} \to \theta_{X,1} \neq \theta_{X,2}$. Consequently, in case of a break in the slope parameters, the residual process is an I(1) process. Therefore, the asymptotic behavior in case of slope breaks is similar to the case discussed in Proposition 2. We therefore have a very similar result, where local alternatives are now of the form $\theta_{X,2} = \theta_{X,1} + \frac{1}{T}\Delta_\theta$.

Proposition 4. (Behavior in Case of Slope Breaks)
Let the data be generated by (1) *and* (2) *with Assumptions 1 and 2 in place. Furthermore, assume that long-run variance estimation is performed consistently using observations* $1, \ldots, [mT]$ *and and assume again that* $g(s)$ *is continuous with* $0 < g(s) < \infty$*. Denote with* $\widehat{F}^m(s)$ *either* $\widehat{H}^{m,+}(s)$ *or* $\widehat{I}^m(s)$*.*

(a) Considering fixed alternatives of the form $\theta_X = \theta_{X,1}$ *for* $t = 1, \ldots, [rT]$ *and* $\theta_X = \theta_{X,2}$ *for* $t = [rT] + 1, \ldots, T$*, with* $\theta_{X,1} \neq \theta_{X,2}$ *leads to a similar result as in part (a) of Proposition 2.*

(b) Considering local alternatives of the form $\theta_X = \theta_{X,1}$ *for* $t = 1, \ldots, [rT]$ *and* $\theta_X = \theta_{X,1} + \frac{1}{T}\Delta_\theta$*, with* $\Delta_\theta \neq 0$*, for* $t = [rT] + 1, \ldots, T$ *leads to a similar result as in part (b) of Proposition 2.*

We close this section by considering local asymptotic power. Here, the number of replications is 10,000 and the time series considered are of length 1,000. All random variables are i.i.d. standard normal. The limiting distribution for LAP as discussed in Proposition 2(b) is based on the FCLTs under local alternatives given in (26) for FM-OLS and D-OLS and in Wagner and Wied (2015, eq. (61)) for IM-OLS. Also local asymptotic power against trend and slope breaks (Propositions 3 and 4) is simulated in the same way with the FCLTs given in Wagner and Wied (2015, eqs. (64), (65), (68), (69)). Considering these results in detail leads to the observation that LAP depends on the product of δ and the "signal-to-noise" ratio ω_ξ/ω. As expected, ω enters with negative powers, i.e., a larger error variance decreases local asymptotic power and similarly a larger variance of the additional I(1) component increases local asymptotic power. Moreover, it turns out that LAP against slope breaks increases with the variance of the regressors. We set all signal-to-noise ratios equal to one. It is also clear that in addition to the dependence upon the deterministic component, LAP also depends upon the number of integrated regressors, as illustrated in Figure 1. As expected, LAP decreases with an increasing number of regressors. Keeping this observation in mind, all other results displayed are for the case of only one integrated regressor for brevity.

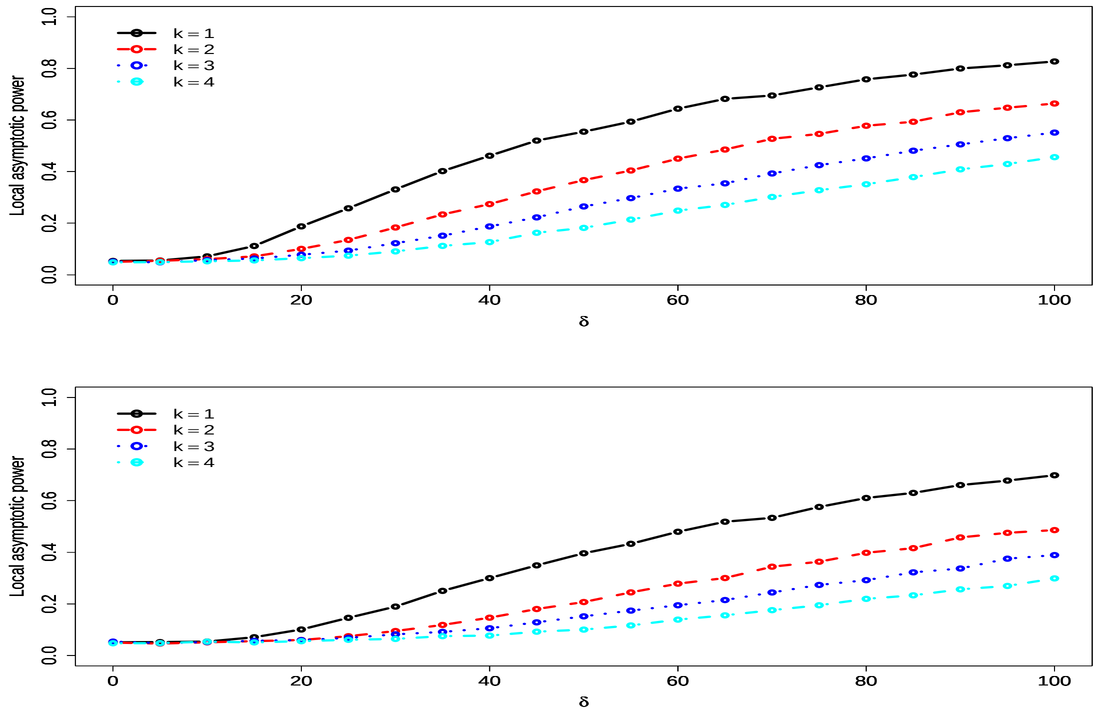

Figure 1: Local asymptotic power for $k = 1, \ldots, 4$ regressors for monitoring cointegration for the case with intercept. The upper plot corresponds to FM-OLS & D-OLS and the lower plot to IM-OLS. The plots show results for $m = r = 0.25$.

In Figure 2 we display LAP against local I(1) alternatives in case of intercept and linear trend included in the model. The upper two plots correspond to FM-OLS and D-OLS and the lower two plots correspond to IM-OLS. We consider several combinations of m and r, the upper sub-plots display the cases $m = 0.25$ and $r = 0.25, 0.5, 0.75$, i.e., the situation with fixed calibration period and equal or later break points; the lower sub-plots display the cases $m = r = 0.25, 0.5, 0.75$, i.e., the situation with increasing calibration period and break immediately at the end of the calibration period. The results show that LAP is lower for IM-OLS, which is as expected given the results of Vogelsang and Wagner (2014) concerning the relative conditional efficiency of FM-OLS over IM-OLS.[13] The practical usefulness of IM-OLS stems from the lower finite sample size distortions that it implies compared to FM-OLS and D-OLS, as illustrated in the following section where we consider finite sample simulations. With respect to changing values of m and r, the LAP ranking is typically similar across methods.

In Figure 3 we display local asymptotic power results against breaks in the intercept, with the same structure of the figure as in Figure 2. The ordering of LAP as a function of $m = r$ (in the second and fourth plot) differs between FM-OLS/D-OLS and IM-OLS. For IM-OLS LAP increases with increasing $m = r$, whereas for FM-OLS and D-OLS LAP is highest for $m = r = 0.5$.

3. Finite Sample Performance

In this section we investigate the finite sample performance of the monitoring procedure by means of a small simulation study. We consider a data generating process similar to Vogelsang and Wagner (2014), i.e., we consider (under the null hypothesis):

$$\begin{aligned} y_t &= \mu + \gamma t + x_{1t}\beta_1 + x_{2t}\beta_2 + u_t, \\ x_{it} &= x_{i,t-1} + v_{it}, \quad x_{i0} = 0, \quad i = 1, 2 \end{aligned} \tag{35}$$

where

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \varepsilon_t + \rho_2(e_{1t} + e_{2t}), \quad u_0 = 0, \\ v_{it} &= e_{it} + 0.5 e_{i,t-1}, \qquad i = 1, 2, \end{aligned}$$

[13] The results are similar, with the differences smaller, in the intercept only case. Additional results, including also results for breaks in the slope parameter, are available upon request. The findings are, as expected, very similar to the ones for local I(1) alternatives.

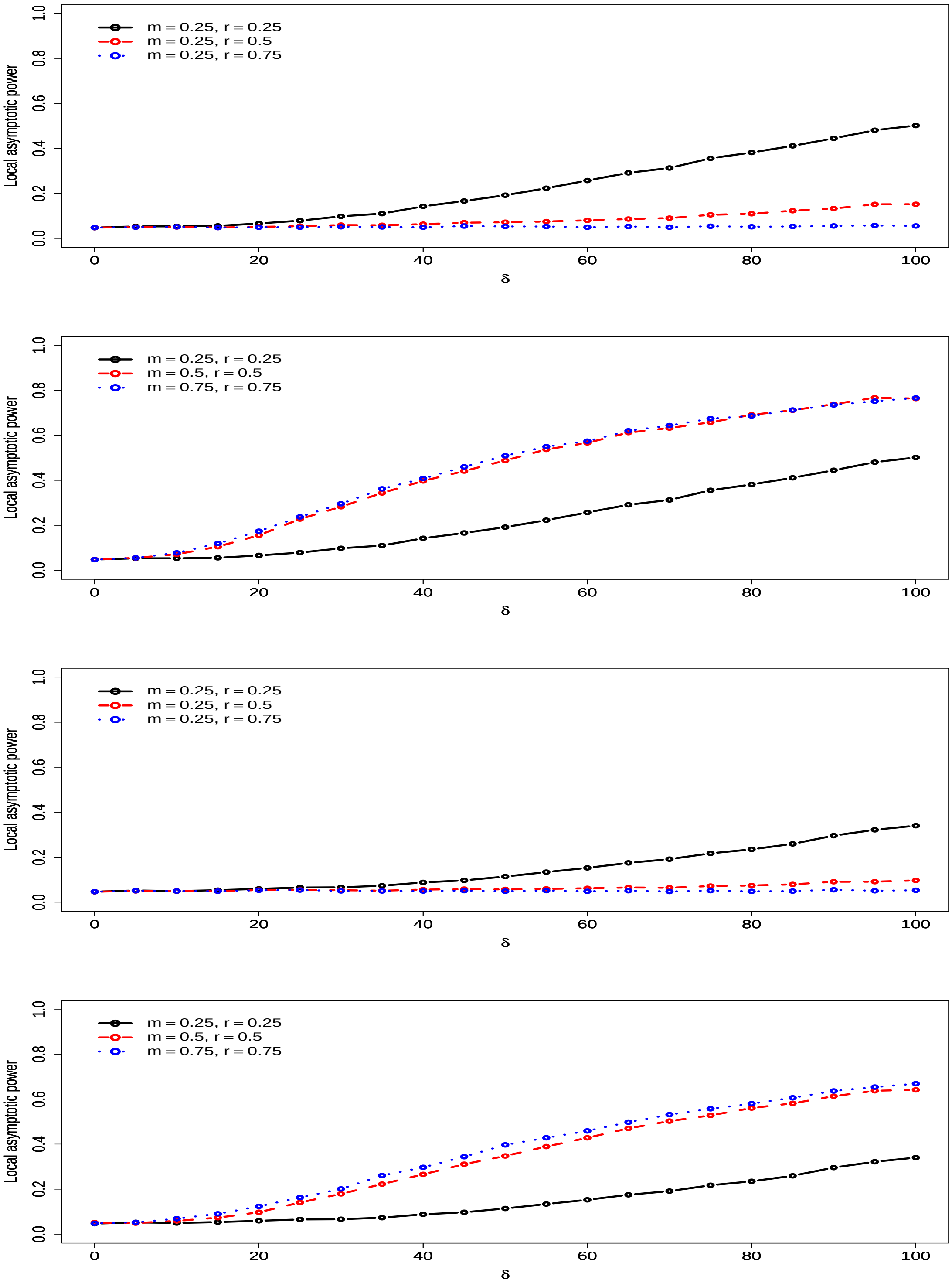

Figure 2: Local asymptotic power for monitoring cointegration for the case with intercept and linear trend. The upper two plots correspond to FM-OLS & D-OLS and the lower two plots to IM-OLS. The plots show results for different combinations of m and r.

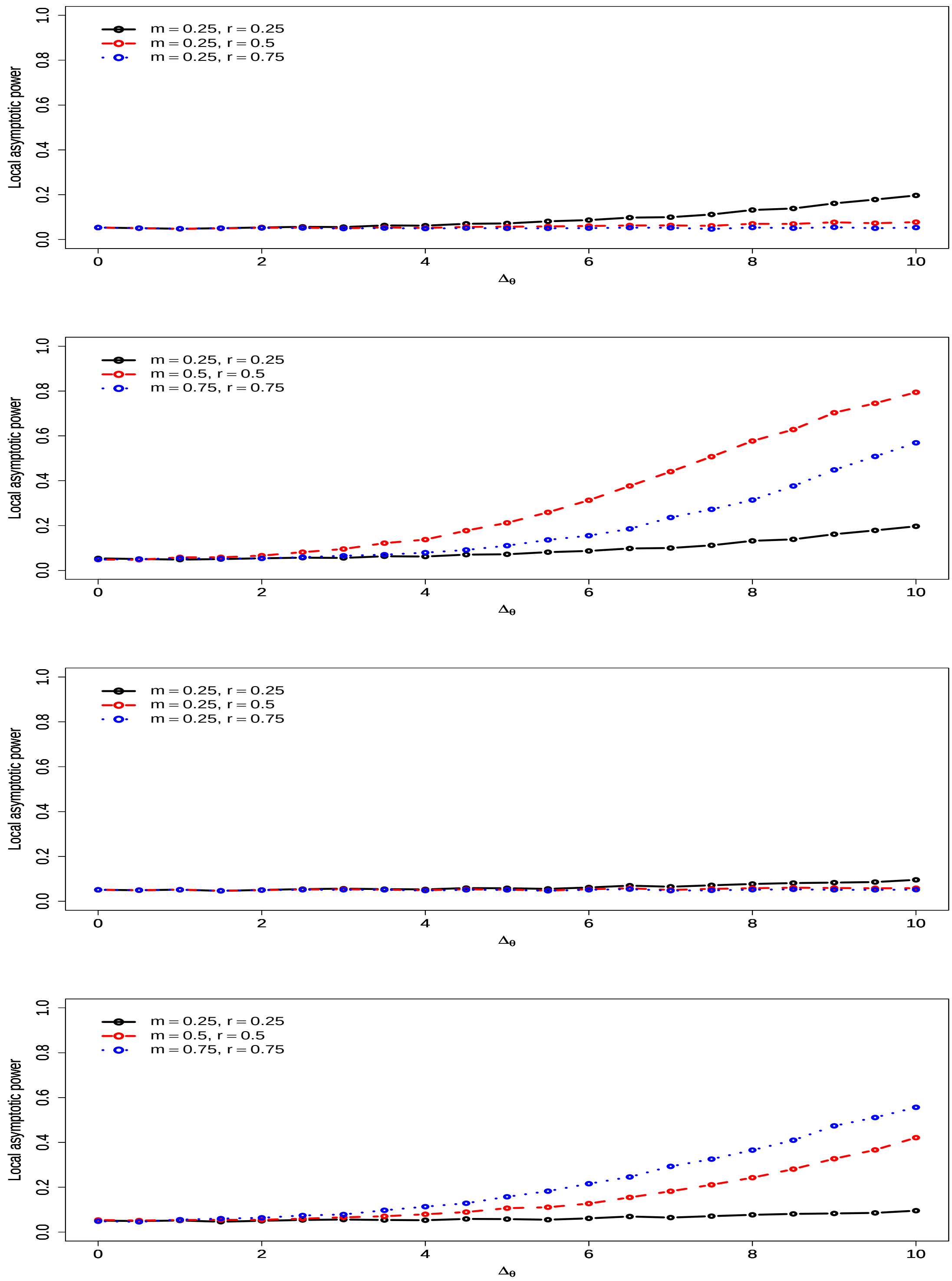

Figure 3: Local asymptotic power against break in intercept for monitoring cointegration for the case with intercept. The upper two plots correspond to FM-OLS & D-OLS and the lower two plots to IM-OLS. The plots show results for different combinations of m and r.

where ε_t, e_{1t} and e_{2t} are i.i.d. standard normal random variables independent of each other. The parameter values chosen are $\mu = 3$, $\beta_1, \beta_2, \gamma = 1$. Due to space limitations we report the results for the intercept and linear trend case only and only briefly comment on the corresponding results for the intercept only case ($\gamma = 0$), with the corresponding results available upon requests. The values for ρ_1 and ρ_2 are chosen from the set $\{0.0, 0.3, 0.6, 0.9\}$. The parameter ρ_1 controls serial correlation in the regression error and is set to $\rho_1 = 1$ under the alternative of I(1) errors, whereas the parameter ρ_2 controls whether (and to which extent) the regressors are endogenous ($\rho_2 \neq 0$) or not ($\rho_2 = 0$). Both, parameter estimation as well as the computation of the detector require the choice of kernel and bandwidth for long-run variance estimation. We use the data dependent bandwidth rule of Andrews (1991) and the Bartlett kernel. The D-OLS estimator is implemented using the information criterion based lead and lag length choice developed in Kejriwal and Perron (2008), where we use the more flexible version discussed in Choi and Kurozumi (2012) in which the numbers of leads and lags included are not restricted to be equal. The considered sample sizes are $T = 200, 500$ and the number of replications is 10,000. Monitoring is performed at the 5% nominal level.

We start by considering empirical null rejection probabilities for a grid of 81 values given by $m = 0.1, 0.11, \ldots, 0.9$ for $\rho_1, \rho_2 = 0, 0.3, 0.6, 0.9$ and $T = 200$ in Figure 4 and $T = 500$ in Figure 5. Several main patterns in line with expectations emerge: First, size distortions decrease with increasing m and increasing sample size T. The exception to this general pattern is D-OLS and $T = 200$, where size distortions are very high and partly increasing for small values of m before they start to decrease. Second, larger values of ρ_1, ρ_2 lead to increasing size distortions and the larger ρ_1, ρ_2 are, the more beneficial is a larger value of m to mitigate the size distortions. Third, the largest size distortions occur for D-OLS and typically the smallest ones for IM-OLS, with the largest differences for small values of m. The larger ρ_1, ρ_2, the bigger is the performance advantage of IM-OLS over the other two methods. Grosso modo the observed size distortions are comparable to the size distortions observed for parameter tests in Vogelsang and Wagner (2014). Let us note for completeness that size distortions are throughout smaller when considering the intercept only case.

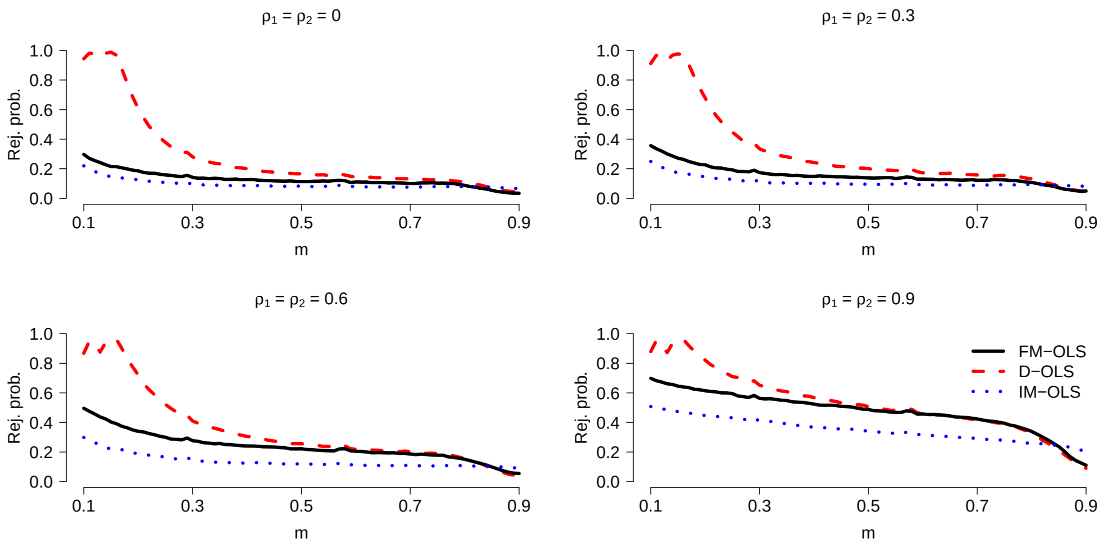

Figure 4: Empirical null rejection probabilities for monitoring cointegration for the case with intercept and linear trend for a grid of values of m and $T = 200$.

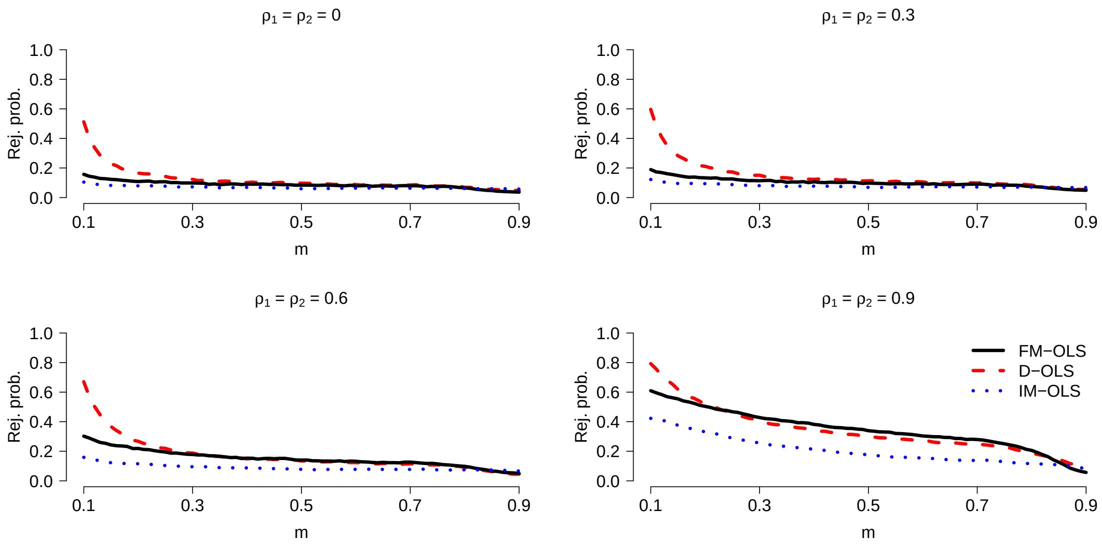

Figure 5: Empirical null rejection probabilities for monitoring cointegration for the case with intercept and linear trend for a grid of values of m and $T = 500$.

Consider next power against I(1) errors from sample fraction $[rT]$ onwards for calibration sample $1, \dots, [mT]$. Note that all power results reported and discussed in the paper refer to *size-corrected* power, which is for brevity simply referred to as power in the text. We consider for both m and r all three values $0.25, 0.5, 0.75$. Thus, we also include cases where $r < m$, i.e., where the break occurs in the calibration period. It can be shown, most easily for stationarity monitoring with $\dim(X_t) = 0$, that in case $r < m$, the order of the detector is given by $O_p(\frac{T}{b_T})$, with b_T denoting the bandwidth chosen. Given that in case of I(1) errors, i.e., under the alternative, typically large bandwidths b_T are chosen, one can expect a low divergence rate of the detector in the $r < m$ case and thus low power in finite samples. This is exactly what happens. The results are shown for $T = 200$ in Table 1 and for $T = 500$ in Table 2.

		$m = 0.25$			$m = 0.5$			$m = 0.75$		
$\rho_1 = \rho_2$	r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
	FM	0.60	0.52	0.43	0.21	0.87	0.56	0.06	0.34	0.89
0	D	0.20	0.49	0.44	0.07	0.82	0.57	0.05	0.27	0.87
	IM	0.51	0.47	0.43	0.17	0.80	0.54	0.06	0.25	0.82
	FM	0.45	0.47	0.39	0.14	0.79	0.51	0.05	0.23	0.83
0.3	D	0.13	0.42	0.39	0.06	0.71	0.51	0.05	0.16	0.80
	IM	0.39	0.43	0.40	0.12	0.71	0.50	0.05	0.17	0.75
	FM	0.21	0.31	0.29	0.07	0.56	0.38	0.05	0.10	0.65
0.6	D	0.08	0.34	0.33	0.05	0.53	0.42	0.05	0.09	0.65
	IM	0.23	0.35	0.37	0.08	0.53	0.43	0.05	0.09	0.60
	FM	0.06	0.07	0.09	0.05	0.09	0.10	0.05	0.05	0.15
0.9	D	0.05	0.08	0.12	0.05	0.09	0.13	0.05	0.05	0.15
	IM	0.06	0.09	0.14	0.05	0.09	0.14	0.05	0.05	0.13

Table 1: Empirical rejection probabilities for monitoring cointegration for the case with intercept and linear trend for $T = 200$: Size corrected power against I(1) breaks.

In addition to the above observation, two more (to be expected) findings are that power increases with the sample size T and decreases with ρ_1, ρ_2. For any given constellation of T and ρ_1, ρ_2, power is largest for $m = r$, compared to situations where $m \neq r$. For $m = r$, power is the larger the larger are m and r. In case that $r < m$, i.e., when the break occurs in the calibration period, power also increases with the sample size. In this case, however, power decreases in the difference between m and r. This finding reflects the fact that asymptotically the term that is subtracted in the numerator of the detector is – under the alternative – a function of the difference $m - r$, rather than of m as under

		$m = 0.25$			$m = 0.5$			$m = 0.75$		
$\rho_1 = \rho_2$	r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
	FM	0.94	0.57	0.47	0.59	0.99	0.61	0.11	0.72	0.99
0	D	0.91	0.56	0.48	0.54	0.98	0.63	0.09	0.70	0.99
	IM	0.89	0.52	0.48	0.49	0.97	0.60	0.08	0.63	0.98
	FM	0.86	0.53	0.45	0.47	0.97	0.58	0.08	0.61	0.98
0.3	D	0.82	0.52	0.45	0.40	0.96	0.58	0.07	0.59	0.98
	IM	0.80	0.49	0.45	0.38	0.95	0.56	0.07	0.51	0.95
	FM	0.64	0.44	0.38	0.24	0.90	0.50	0.06	0.39	0.93
0.6	D	0.62	0.48	0.41	0.22	0.91	0.52	0.06	0.41	0.93
	IM	0.63	0.45	0.43	0.23	0.87	0.53	0.06	0.34	0.89
	FM	0.09	0.17	0.19	0.05	0.27	0.24	0.05	0.06	0.41
0.9	D	0.09	0.21	0.24	0.06	0.32	0.30	0.05	0.06	0.46
	IM	0.10	0.25	0.32	0.06	0.32	0.37	0.05	0.06	0.43

Table 2: Empirical rejection probabilities for monitoring cointegration for the case with intercept and linear trend for $T = 500$: Size corrected power against I(1) breaks.

the null. In case that $m \leq r$ power decreases with increasing r, which merely reflects the fact that a smaller sub-sample is available under the alternative for detecting the structural change from I(0) to I(1) behavior.

The differences in power between FM-OLS and IM-OLS are minor, whereas power of D-OLS is markedly smaller in case $r = 0.25$. Clearly, this latter finding cannot be predicted by LAP, which is by construction identical between FM-OLS and D-OLS. This finding reflects the well-known fact that D-OLS often leads to coefficient estimators with larger variances in small samples. Comparing FM-OLS and IM-OLS shows that occasionally power is marginally lower for IM-OLS than for FM-OLS, and in case that r is large marginally lower than for D-OLS, which is in line with the LAP analysis. In case of large values of ρ_1, ρ_2, however, power is sometimes higher for IM-OLS. These findings concerning power have to be seen in conjunction with the partly substantially lower size distortions when using IM-OLS compared to FM-OLS or in particular D-OLS. All findings hold analogously for the intercept only case, with power being typically higher in the intercept only case compared to the intercept and linear trend case.

We next consider power against breaks in the trend or slope parameters. For brevity we only display results for the case $\rho_1, \rho_2 = 0.9$, i.e., the worst results from our set of values for ρ_1, ρ_2. All figures have the same structure and display in the left column the graphs for $m = 0.25$ and $r \in \{0.25, 0.5, 0.75\}$ and in the right column the graphs for $m = r \in \{0.25, 0.5, 0.75\}$.

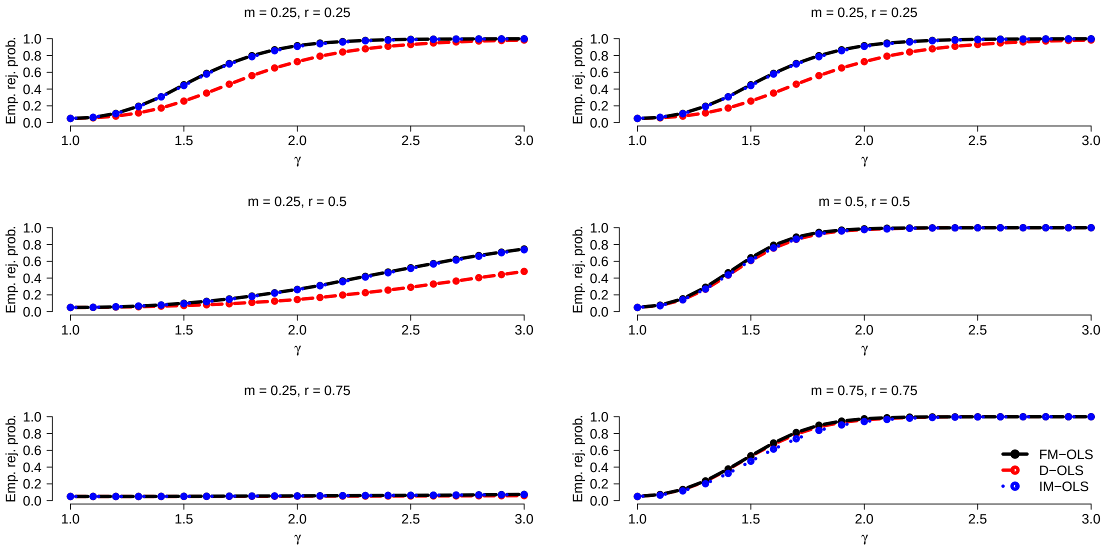

Figure 6: Empirical rejection probabilities for monitoring cointegration for the case with intercept and linear trend for $T = 200$, $\rho_1 = \rho_2 = 0.9$: Size corrected power against trend breaks.

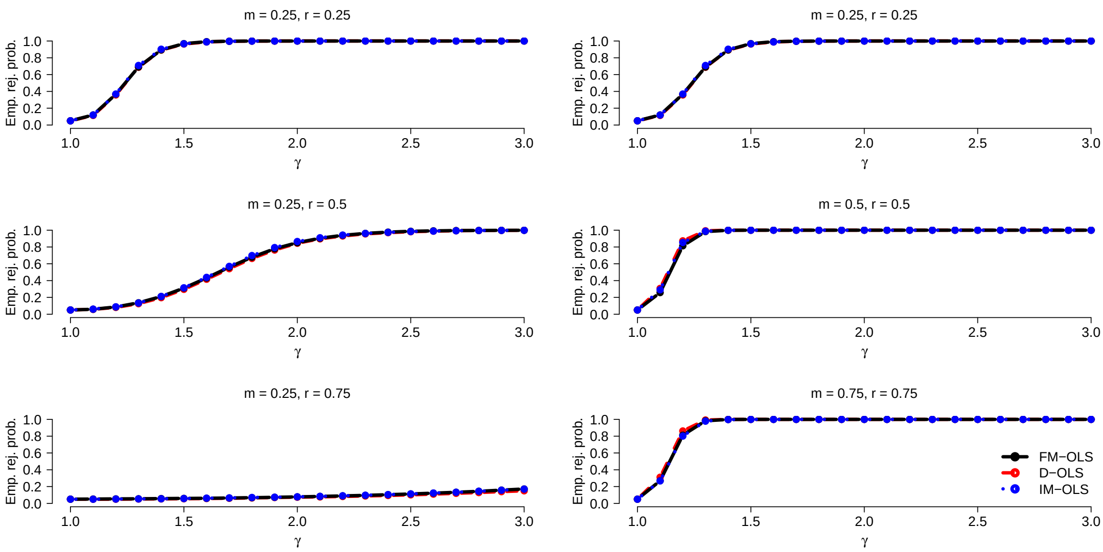

Figure 7: Empirical rejection probabilities for monitoring cointegration for the case with intercept and linear trend for $T = 500$, $\rho_1 = \rho_2 = 0.9$: Size corrected power against trend breaks.

For trend breaks we consider a grid with 21 points and mesh 0.1 for the trend slope

$\gamma \in [1, 3]$ (including the null) displayed on the horizontal axis in Figure 6 for $T = 200$ and in Figure 7 for $T = 500$. Three main observations emerge: First, power decreases with increasing r for given m (left column). Second, power increases with increasing $m = r$ (right column). Third, power is very similar for FM-OLS and IM-OLS and is in some cases lower for D-OLS. The differences across estimation methods get smaller for increasing sample size.

For slope breaks we consider the case that β_1, β_2 both change their value from 1 to some value larger than 1, where we consider the same grid of 21 points as for trend breaks above, i.e., $\beta_i \in [1, 3]$ (including the null). The results for $T = 200$ are displayed in Figure 8 and for $T = 500$ in Figure 9. The results are qualitatively very similar to the findings in case of a trend break. Note for completeness that power against slope breaks (of similar magnitude) is slightly higher in the intercept only case compared to the intercept and linear trend case. Note also that the similarities and differences between LAP and finite sample simulations discussed above, when considering power against I(1) errors, exist as well for the case of trend and slope breaks.

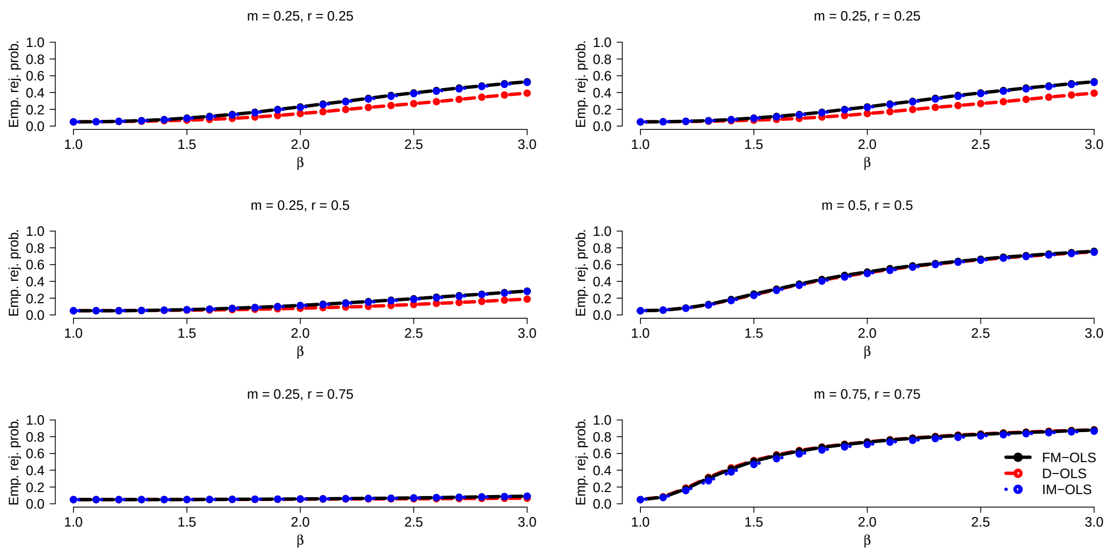

Figure 8: Empirical rejection probabilities for monitoring cointegration for the case with intercept and linear trend for $T = 200$, $\rho_1 = \rho_2 = 0.9$: Size corrected power against slope breaks.

Finally we investigate the estimated detection times against I(1) breaks. These are shown

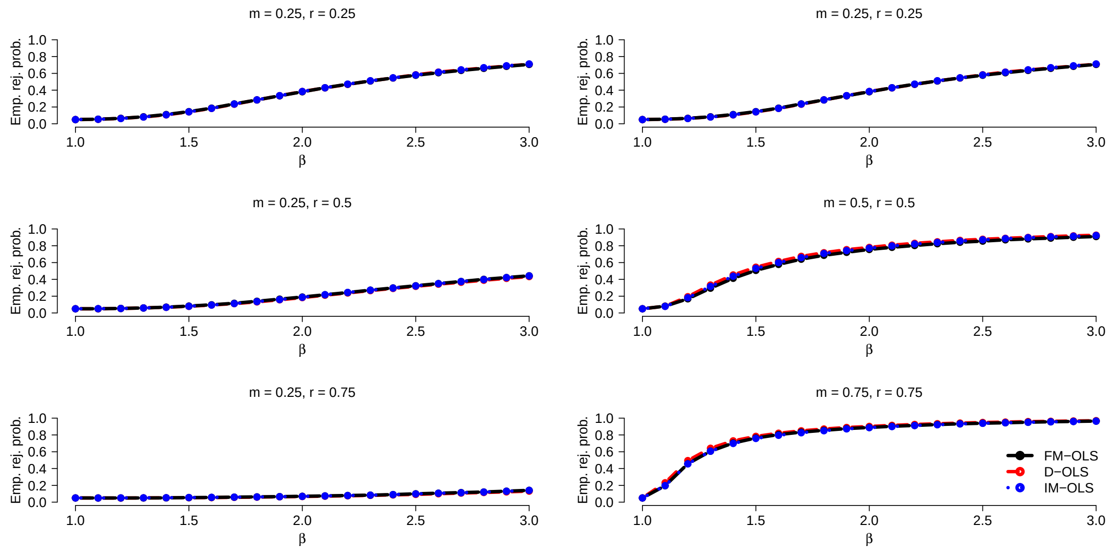

Figure 9: Empirical rejection probabilities for monitoring cointegration for the case with intercept and linear trend for $T = 500$, $\rho_1 = \rho_2 = 0.9$: Size corrected power against slope breaks.

in Figures 10 ($\rho_1, \rho_2 = 0$) and 11 ($\rho_1, \rho_2 = 0.9$) for $T = 200$ and in Figures 12 ($\rho_1, \rho_2 = 0$) and 13 ($\rho_1, \rho_2 = 0.9$) for $T = 500$. We display the detection times in the form of Box-Whiskers plots. The numbers below the abbreviated method names indicate the null rejection probabilities given in Tables 1 and 2. Thus, the different Box-Whiskers plots are based on different *numbers of replications* because of a different number of rejections across different methods, sample sizes and ρ-parameters.[14]

The structure of the six graphs within one figure corresponds to the structure of the power figures with respect to the combinations of m and r displayed. By construction, detection occurs *typically* with delay. An increasing sample size leads to a – ceteris paribus – more concentrated distribution of the estimated detection times (based on a larger number of observations), but does not necessarily lead to smaller average delays. As expected, increasing endogeneity and error serial correlation lead to increasing detection times, i.e., bigger delays. For fixed $m = 0.25$ increasing values of r lead to decreasing delays. For $m = 0.25$ and $r = 0.75$ detection often occurs already *prior* to the structural change.

[14]The detection times are related to the so-called average run lengths often considered in the control chart literature, e.g., the median average run length is given by $T(\bar{\tau}_m - m)$, with $\bar{\tau}_m$ denoting the median detection time.

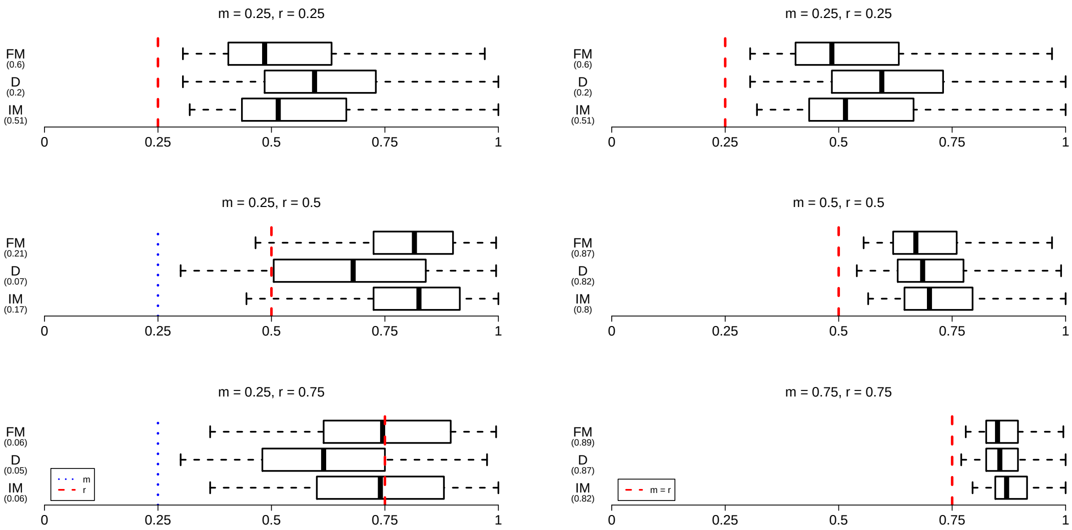

Figure 10: Detection times when monitoring cointegration for the case with intercept and linear trend for $T = 200$, $\rho_1 = \rho_2 = 0$: Detection times for I(1) breaks.

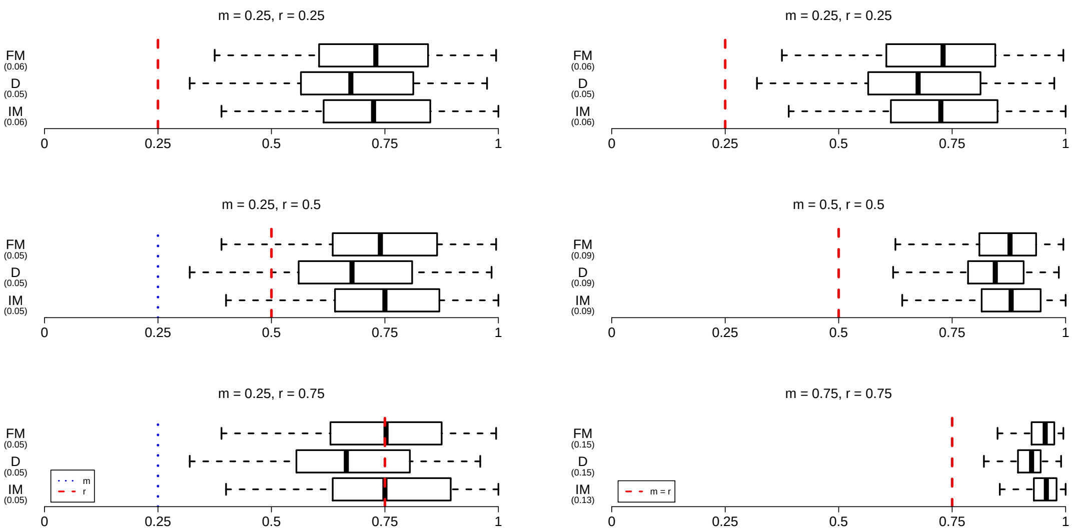

Figure 11: Detection times when monitoring cointegration for the case with intercept and linear trend for $T = 200$, $\rho_1 = \rho_2 = 0.9$: Detection times for I(1) breaks.

Increasing values of $m = r$ also lead to smaller delays of detecting the structural change. When comparing the three methods, there is no clear ranking with respect to delay. FM-OLS and IM-OLS lead to very similar detection time distributions, whereas the D-OLS detection times look rather different from the other two. These differences are more

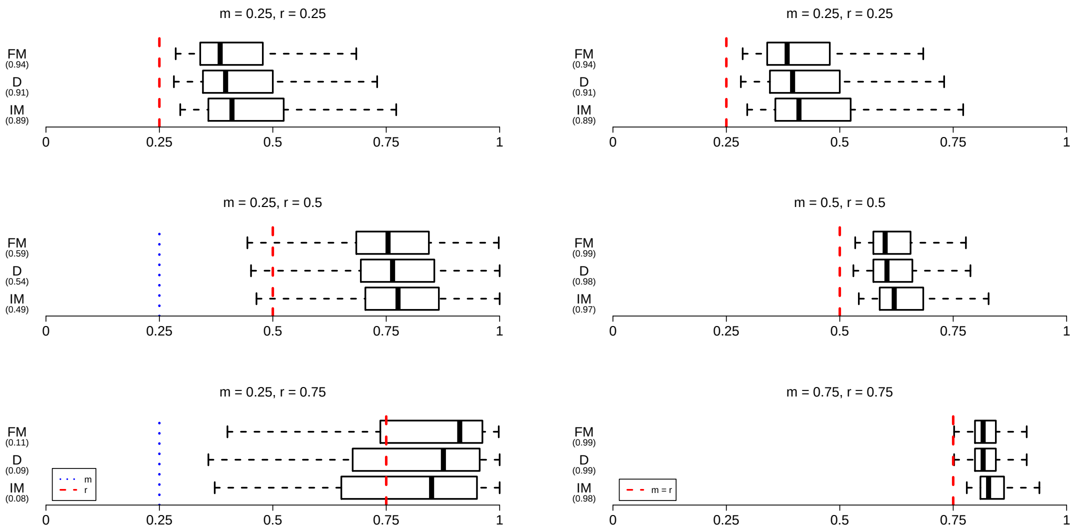

Figure 12: Detection times when monitoring cointegration for the case with intercept and linear trend for $T = 500$, $\rho_1 = \rho_2 = 0$: Detection times for I(1) breaks.

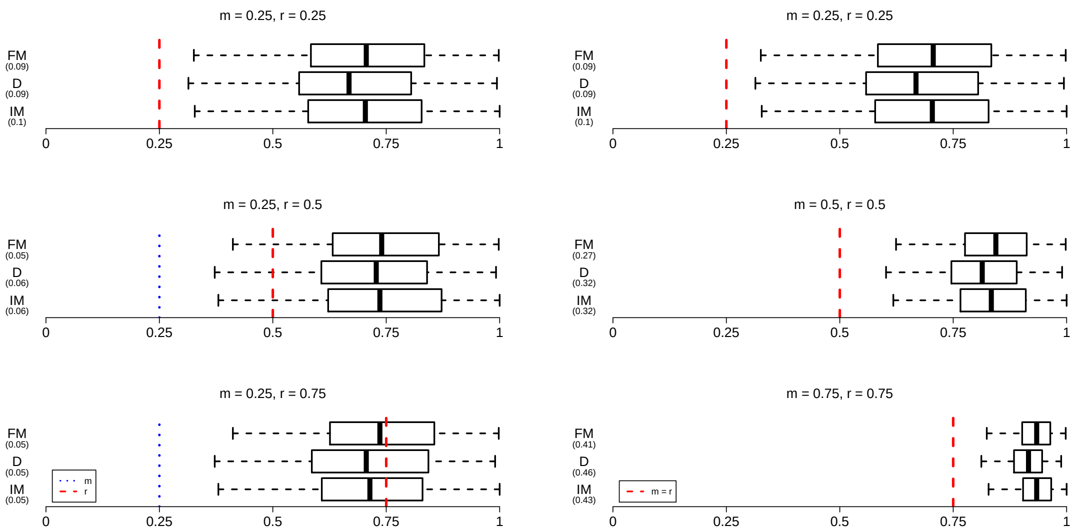

Figure 13: Detection times when monitoring cointegration for the case with intercept and linear trend for $T = 500$, $\rho_1 = \rho_2 = 0.9$: Detection times for I(1) breaks.

marked in cases where the rejection probabilities (i.e., powers) of D-OLS are sizeably smaller than for the other two methods. The delays are partly substantial, in particular for $\rho_1, \rho_2 = 0.9$. Obtaining a better understanding of the impact of the weighting function on the expected delays is consequently a topic of future research, notwithstanding the complications outlined in Section 2.

4. Empirical Application

In this section we apply our monitoring procedure to the question and data analyzed in Anundsen (2015), who studies the potential breakdown of fundamentals driven housing price cointegrating relationships before the outburst of the US subprime crisis.[15]

In particular Anundsen (2015, Section 3) considers two related relationships, both based on the life-cycle model and a no arbitrage condition on the housing market. The first relationship, henceforth referred to as price-to-rent model, stems from the equilibrium equality of rents and user costs of similar units of housing.[16] This leads to the following approximate equilibrium relationship (ignoring deterministic components and stochastic errors at this point):

$$p_t = \theta_r r_t + \theta_{UC} UC_t, \tag{36}$$

with p_t the logarithm of real housing prices (in period t), r_t the logarithm of real rents and UC_t the real direct user costs of housing. Here, as in Anundsen (2015), lower case letters indicate logarithms of variables, with the user costs considered in levels since they assume also negative values over the sample period.

The second relationship, henceforth referred to as inverted demand model, considers as a starting point (imputed) rents as a function of both income and the housing stock. Combining this with the equilibrium considerations alluded to above leads to the following approximate equilibrium relationship (again ignoring deterministic components and errors

[15] Anoother illustrative application of our procedure is contained in the extensive working paper Wagner and Wied (2015), where we consider the stationarity of CDS spreads around the Lehman crisis. In a short note, Aschersleben et al. (2015), we analyze the stationarity of Euro area real exchange rates.

[16] The real user cost of housing is given by $UC_t = (1 - \tau_t^y)(i + \tau_t^p) - \pi_t + \delta_t + \frac{\dot{P}_t}{P_t}$, with i_t the nominal interest rate, τ_t^p the property tax rate, τ_t^y capturing tax deductions, π_t the overall price inflation, δ_t the housing depreciation rate, and $\frac{\dot{P}_t}{P_t}$ expected real housing price inflation. The underlying no-arbitrage relationship is given by $Q_t = P_t UC_t$, with Q_t being the real imputed rent, for the empirical analysis replaced by observed real rents.

here):

$$p_t = \theta_y y_t + \theta_h h_t + \theta_{UC} UC_t, \tag{37}$$

with y_t the logarithm of real (per capita disposable) income and h_t the logarithm of the (per capita) housing stock in period t. Full details concerning the data, the sources and the construction of the variables are contained in Anundsen (2015, Section 4). The quarterly data available for 1976:Q1 – 2010:Q4 have been downloaded from the archive of the *Journal of Applied Econometrics*.

Since for all considered variables present in (36) and (37), the null hypothesis of a unit root cannot be rejected, the empirical econometric counterpart of the above *error-free* relationships is that of a cointegrating relationship between these variables, leading to (including deterministic components and errors now):[17]

$$p_t = \theta_c + \theta_t t + \theta_r r_t + \theta_{UC} UC_t + u_t \tag{38}$$

$$p_t = \theta_c + \theta_t t + \theta_y y_t + \theta_h h_t + \theta_{UC} UC_t + u_t, \tag{39}$$

with u_t a stationary error term in case of cointegration. The absence or breakdown of a cointegrating relationship is then interpreted as an indication that housing prices are not anymore driven by fundamentals, which is interpreted as a housing price bubble by Anundsen (2015).

Anundsen (2015) collects the variables appearing in the two equations in vector autoregressive models and performs recursive cointegration analysis over expanding samples. The first sample considered ranges from 1976:Q1 to 1995:Q4. Then the sample is extended by one year per step, i.e. by four observations, until the full sample range up to 2010:Q4 is exhausted. In the price-to-rent model the variables modelled by a VAR are real housing prices p_t, real rents r_t and the real direct user cost UC_t. In the inverted demand

[17]A small caveat mentioned by Anundsen (2015) is that the log real housing stock might be I(2). Similarly to Anundsen (2015) we nevertheless consider the housing stock as an I(1) process.

model the variables are again real housing prices p_t, real per capita disposable income y_t, the real direct user cost UC_t and as an exogenous variable the real per capita housing stock h_t. Furthermore, a constant and three centered seasonal dummies are included, as well as a linear trend that is restricted to be in the cointegrating space. Note that, since we do not find evidence for seasonality in the data, we do not include the seasonal dummies in our analysis. Anundsen (2015) finds a one-dimensional cointegrating space until the end of 2001 for the price-to-rent approach and until the end of 2000 for the inverted demand approach. Thereafter the recursive analysis does not find evidence for a cointegrating relationship.

These are, of course, highly interesting results, but they are prone to all problems related with multiple testing, like uncontrolled size. Furthermore, recursive testing is here performed both before and after a structural break has been found, which makes the interpretation of the results even more complicated. Exactly for this type of problem our monitoring procedure can be applied, since it is a procedure with controlled asymptotic size properties that overcomes the problems inherent in a (multiple) recursive cointegration testing setting. Consequently, we apply our monitoring procedure to the two relationships given in (38) and (39), with calibration period 1976:Q1–1995:Q4, i.e., $m = \frac{4}{7}$. The equations are estimated over the calibration period with all three methods mentioned in Section 2. Since the linear trend is significant for both models for at least one estimator, we only report the results here for the case with intercept and linear trend included.[18] The estimation results are given in Table 3 for the price-to-rent model and in Table 4 for the inverted demand model.

For the price-to-rent model, if one takes the underlying theory at face value, the coefficients should equal $\theta_r = 1$ and $\theta_{UC} = -1$. The coefficient estimates given in Table 3 are close to and not significantly different from these values, as are those in Anundsen (2015), for both the FM-OLS and D-OLS estimates, but are, surprisingly, not close for the IM-OLS estimates. There are no systematic differences across estimation methods

[18] The results for the intercept only case are available upon request. They do not differ substantially.

Method	Coefficient	Estimate	Std. Error	t-value	p-value
FM-OLS	θ_c	0.1910	0.0136	14.0341	0.0000
	θ_t	0.0007	0.0004	1.9540	0.0544
	θ_r	0.8905	0.1941	4.5873	0.0000
	θ_{UC}	-1.0398	0.2374	-4.3791	0.0000
D-OLS	θ_c	0.2026	0.0093	21.9005	0.0000
	θ_t	0.0004	0.0003	1.2497	0.2152
	θ_r	0.9123	0.1406	6.4862	0.0000
	θ_{UC}	-0.9175	0.2711	-3.3842	0.0011
IM-OLS	θ_c	0.1618	0.0157	10.3406	0.0000
	θ_t	0.0004	0.0005	0.7564	0.4518
	θ_r	0.4814	0.2346	2.0522	0.0436
	θ_{UC}	0.3776	0.4936	0.7650	0.4467

Table 3: Estimation results for the price-to-rent model with intercept and linear trend.

Method	Coefficient	Estimate	Std. Error	t-value	p-value
FM-OLS	θ_c	9.5136	0.4972	19.1352	0.0000
	θ_t	-0.0065	0.0004	-15.5857	0.0000
	θ_y	0.5646	0.0473	11.9456	0.0000
	θ_h	1.7969	0.1466	12.2535	0.0000
	θ_{UC}	-0.2152	0.0674	-3.1939	0.0021
D-OLS	θ_c	14.2474	0.4156	34.2797	0.0000
	θ_t	-0.0098	0.0003	-31.3090	0.0000
	θ_y	0.2262	0.0567	3.9924	0.0002
	θ_h	3.4772	0.1601	21.7170	0.0000
	θ_{UC}	-0.0743	0.0739	-1.0051	0.3181
IM-OLS	θ_c	10.4304	0.5974	17.4598	0.0000
	θ_t	-0.0072	0.0005	-14.5568	0.0000
	θ_y	0.5222	0.0653	7.9902	0.0000
	θ_h	2.0926	0.1923	10.8796	0.0000
	θ_{UC}	-0.2132	0.0927	-2.3003	0.0242

Table 4: Estimation results for the inverted demand model with intercept and linear trend.

	FM-OLS	D-OLS	IM-OLS
Price-to-rent model	2007:Q3	2006:Q4	2006:Q4
Inverted-demand-model	2003:Q2	2007:Q3	2004:Q2

Table 5: Detected break-points for the price-to-rent and the inverted-demand models.

for the inverted demand model, compare Table 4.

The detected break-points are reported in Table 5. The results show some variation in the detection times across methods for both models, with different rankings across methods for both models. This is in line with the simulation findings with respect to detection times reported in Section 3, where also no clear ranking for the detection times has emerged. For both models, however, except for the latest detector, the detection times are before the collapse of house prices that started at the beginning of 2007.[19] The earliest detection occurs already in 2003:Q2 for the inverted demand model using the FM-OLS residuals. For the price-to-rent model the earliest detection, however, occurs only in 2006:Q4 for both the D-OLS and the IM-OLS residuals, i.e., just before house prices started to fall. By construction, compare also the simulations in Section 3, detection occurs with a delay when using a monitoring procedure. In contrast, the break-points found by Anundsen (2015) are much earlier. On the one hand, our monitoring procedure suffers from delays, but is asymptotically both size controlled under the null and consistent under alternatives. On the other hand, the recursive testing approach signals an early break, but has no asymptotic justification. In our view, the delay is the price that has to be paid for asymptotic validity.

Let us close this section by looking at the residuals in Figure 14 and the detectors in Figure 15. In these two figures the left graph corresponds to the price-to-rent model and the right graph to the inverted demand model. Note that the residuals displayed in Figure 14 are obtained from parameter estimation only over the calibration period that ranges until 1995:Q4. Thus, by construction until the end of 1995 the graphs display

[19] Note for completeness that for the price-to-rent model the imposition of the theoretical restrictions $\theta_r = 1$ and $\theta_{UC} = -1$ – and thereby using the series $p_t - r_t + UC_t$ for stationary monitoring (including a constant and a linear trend) – leads to a detected break-point in 1996:Q3.

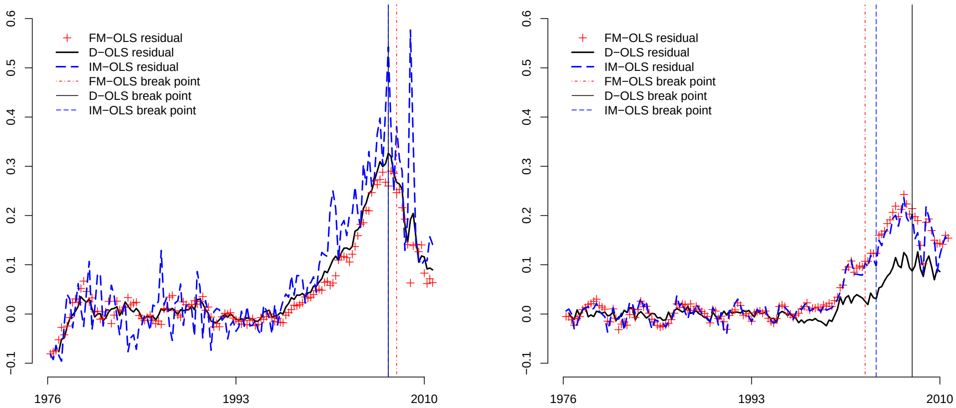

Figure 14: Residuals of the price-to-rent (left graph) and inverted demand (right graph) models. The vertical lines indicate the detected break points for the three used estimation methods. For IM-OLS we display the first differences of the residuals of the partial summed regression used in IM-OLS estimation.

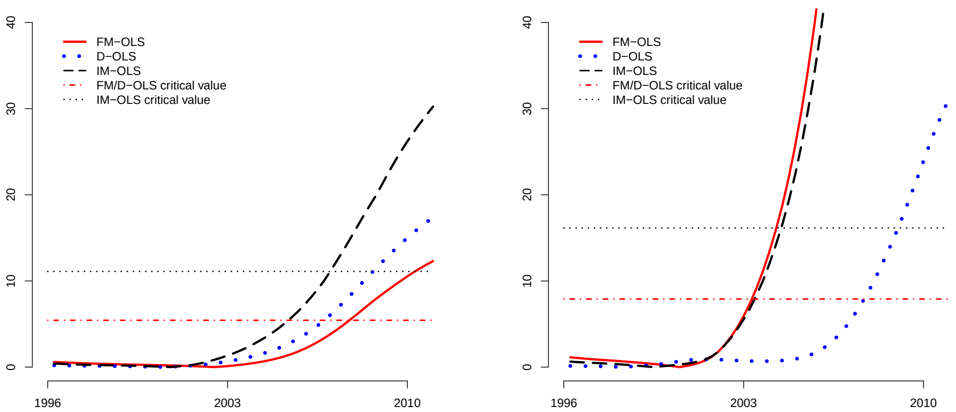

Figure 15: Detectors for the price-to-rent (left graph) and inverted demand (right graph) models. The two horizontal lines indicate the critical values for the three versions of the monitoring procedure.

actual modified least squares estimation residuals. Both graphs show a tendency of the residuals to become larger after the end of the calibration period until almost the onset of the house price collapse before they become smaller again towards the end of the sample, at least for the price-to-rent model. The vertical lines indicate when the residuals have become "large enough for long enough" for the detectors to signal a structural break. Looking at Figure 14, this happens relatively late and happens later for the residual series with larger variations (in particular IM-OLS in the price-to-rent model). Figure 15 shows information also contained in the previous graph by plotting the detectors. The intersection of the detectors with the corresponding critical values occurs exactly at the break-points displayed by means of vertical lines in Figure 14.

5. Summary and Conclusions

We have proposed a consistent monitoring procedure for cointegration based on parameter estimation on a pre-break calibration period. The detector is inspired by Chu et al. (1996) as well as by the cointegration test of Kwiatkowski et al. (1992). The key ingredients of the detector are properly scaled partial sums of residuals that are compared between the calibration and the successively increasing monitoring period. In order to accommodate error serial correlation as well as regressor endogeneity modified least squares estimators, to be precise FM-OLS, D-OLS and IM-OLS, are used, which allow to obtain nuisance parameter free limiting distributions. The procedure is consistent against I(1), explosive or fractionally integrated alternatives and breaks in the trend and/or slope parameters.
The performance of the procedure has been investigated both in terms of local asymptotic power as well as by means of finite sample simulations. The effects of sample size, regressor endogeneity and error serial correlation are as typically found in the unit root and cointegration literature. Local asymptotic power and – in line with the asymptotic findings – finite sample (size corrected) power are in some configurations slightly lower for IM-OLS than for the other two procedures. However, IM-OLS has partly substantially lower size distortions under the null, which makes IM-OLS a good choice given that

size corrections cannot be performed in actual applications. Detection occurs in many circumstances with delays, which are partly substantial. A longer calibration period is, as expected, often beneficial in lessening the delays, as is a late break point.

We have applied the procedure to reinvestigate the question and data analyzed by Anundsen (2015). Using a consistent monitoring procedure adds to his analysis by overcoming the problems associated with multiple testing in recursive VAR cointegration testing. The price to be paid for consistency is the delay of our monitoring procedure, compared to the Anundsen (2015) break-points.

Several extensions of our approach are conceivable in addition to flipping null and alternative hypotheses, a question that is – as mentioned in the introduction – analyzed in Sakarya et al. (2015). First, it may be relevant to consider multivariate monitoring procedures to investigate situations with higher-dimensional cointegrating spaces. Second, especially important for financial applications with higher frequency data, the effects of non-constant (conditional or unconditional) variances need to be understood. Finally, the impact of the weighting functions needs to be understood better. Here it has to be noted that an advantage of a moving monitoring scheme is the potential to avoid weighting altogether.

Acknowledgements

Financial support from Deutsche Forschungsgemeinschaft via the Collaborative Research Center 823: *Statistical Modelling of Nonlinear Dynamic Processes* (Projects A1, A3 and A4) is gratefully acknowledged. The first author furthermore acknowledges financial support from the Jubiläumfonds of the Oesterreichische Nationalbank (Grant No. 15334). We are grateful to Philipp Aschersleben and Fang Duan for excellent research assistance. The usual disclaimer applies.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–854.

Andrews, D. W. K. and J.-Y. Kim (2006): "Tests for Cointegration Breakdown Over a Short Time Period," *Journal of Business and Economic Statistics*, 24, 379–394.

Anundsen, A. (2015): "Econometric Regime Shifts and the US Subprime Bubble," *Journal of Applied Econometrics*, 30, 145–169.

Aschersleben, P., M. Wagner, and D. Wied (2015): "Monitoring Euro Area Real Exchange Rates," in *Stochastic Models, Statistics and Their Applications*, ed. by A. Steland, E. Rafajłowicz, and K. Szajowski, Springer, vol. 122 of *Springer Proceedings in Mathematics and Statistics*, 363–370.

Astill, S., D. Harvey, S. Leybourne, and A. Taylor (2015): "Tests for an End-of-Sample Bubble in Financial Time Series," *Discussion Paper (University of Nottingham).*

Aue, A., S. Hörmann, L. Horváth, M. Hušková, and J. Steinebach (2012): "Sequential Testing for the Stability of High Frequency Portfolio Betas," *Econometric Theory*, 28, 804–837.

Aue, A., L. Horváth, and M. L. Reimherr (2009): "Delay Times of Sequential Procedures for Multiple Time Series Regression Models," *Journal of Econometrics*, 149, 174–190.

Cappuccio, N. and D. Lubian (2005): "Local Asymptotic Distributions of Stationarity Tests," *Journal of Time Series Analysis*, 27, 323–345.

Cavaliere, G. and A. M. R. Taylor (2006): "Testing the Null of Cointegration in the Presence of Variance Breaks," *Journal of Time Series Analysis*, 27, 613–636.

Chen, Q., U. Kruger, and A. Leung (2009): "Cointegration Testing Method for Monitoring Nonstationary Processes," *Industrial and Engineering Chemical Research*, 48, 3533–3543.

Chen, Z., Z. Tian, and Y. Wei (2010): "Monitoring Change in Persistence in Linear Time Series," *Statistics and Probability Letters*, 19-20, 1520–1527.

Choi, I. and E. Kurozumi (2012): "Model Selection Criteria for the Leads-and-Lags Cointegrating Regression," *Journal of Econometrics*, 169, 224–238.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996): "Monitoring Structural Change," *Econometrica*, 64, 1045–1065.

Groen, J. J. J., G. Kapetanios, and S. Price (2013): "Multivariate Methods for Monitoring Structural Change," *Journal of Applied Econometrics*, 28, 250–274.

Jansson, M. (2002): "Consistent Covariance Matrix Estimation for Linear Processes," *Econometric Theory*, 18, 1449–1459.

Johansen, S., R. Mosconi, and B. Nielsen (2000): "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend," *Econometrics Journal*, 3, 216–249.

Kejriwal, M. and P. Perron (2008): "Data Dependent Rules for Selection of the Number of Leads and Lags in the Dynamic OLS Cointegrating Regressions," *Econometric Theory*, 24, 1425–1441.

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.

Phillips, P. and S. Durlauf (1986): "Multiple Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–496.

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

Phillips, P. C. B. and M. Loretan (1993): "Estimating Long Run Economic Equilibria," *Review of Economic Studies*, 58, 407–436.

Phillips, P. C. B., Y. Wu, and J. Yu (2011): "Explosive Behavior in the 1990s Nasdaq: When did Exuberance Escalate Asset Values?" *International Economic Review*, 52, 201–226.

Saikkonen, P. (1991): "Asymptotically Efficient Estimation of Cointegrating Regressions," *Econometric Theory*, 7, 1–21.

Sakarya, N., M. Wagner, and D. Wied (2015): "Monitoring a Change from Spurious Regression to Cointegration," *Mimeo*.

Shin, Y. (1994): "A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration," *Econometric Theory*, 10, 91–115.

Steland, A. (2007): "Monitoring Procedures to Detect Unit Roots and Stationarity," *Econometric Theory*, 23, 1108–1135.

Steland, A. and S. Weidauer (2013): "Detection of Stationary Errors in Multiple Regressions with Integrated Regressors and Cointegration," *Sequential Analysis*, 32, 319–349.

Stock, J. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–1056.

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

Vogelsang, T. J. and M. Wagner (2014): "Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions," *Journal of Econometrics*, 148, 741–760.

Wagner, M. and D. Wied (2015): "Monitoring Stationarity and Cointegration," *SSRN Research Paper (Nr. 2624657).*

Wang, G., C. Liu, Y. Cui, and X. Feng (2014): "Tool Wear Monitoring Based on Cointegration Modelling of Multisensory Information," *International Journal of Computer Integrated Manufacturing*, 27, 479–487.

Wied, D. and P. Galeano (2013): "Monitoring Correlation Change in a Sequence of Random Variables," *Journal of Statistical Planning and Inference*, 143, 186–196.