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 for structural change in a cointegrating relationship. The procedure is inspired by Chu et al. (1996) by being based on parameter estimation on a pre-break "calibration" period. We use three modified least squares estimators to obtain nuisance parameter free limiting distributions. We study the asymptotic and finite sample properties of the procedures and finally apply the approach to monitor two fundamentals driven US housing prices cointegrating relationships over the period 1976:Q1–2010:Q4 using the data of Anundsen (2015). Depending upon the relationship considered and estimation method used, a break-point 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). 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, Fang Duan and Maximilian Grupe for excellent research assistance. Finally we thank the editor as well as two anonymous referees for their helpful comments. The usual disclaimer applies.

1. Introduction

This paper presents a residual based monitoring procedure for structural change in cointegrating relationships. Hereby structural change – at an unknown point in time – can occur in two facets, as also discussed in Andrews and Kim (2006). First, the cointegrating relationship may turn into a spurious relationship and second, there may be a structural change in the trend and/or slope parameters, with the details given in Section 2. A cointegration monitoring procedure requires two ingredients: (i) parameter estimates and (ii) a test or monitoring statistic. Similar to Chu et al. (1996) for linear regression models, we perform parameter estimation on a *calibration* period at the beginning of the sample, rather than recursive or full sample estimation, to obtain residuals as input into the procedure. The calibration period has to be known or at least assumed to be free of structural change. The monitoring procedure then combines a properly redefined cointegration test statistic calculated from the residuals over expanding or moving windows. The test statistic underlying our monitoring procedure is the Shin (1994) test for the null hypothesis of cointegration and we focus on calculating the *detector* over expanding windows.[1]

In order to obtain a nuisance parameter free null limiting distributions of our detectors, we need to use parameter estimation methods that lead to mean zero Gaussian mixture limiting distributions in cointegrating regressions. In this respect we consider three modified least squares estimators: 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 former two variants lead to the same null limiting distribution, whereas IM-OLS leads to a different one. As usual in residual based cointegration analysis, the limiting distribution depends upon the specification of the deterministic

[1]See Footnote 4 for a discussion of other variants – including moving windows – considered. In the context of monitoring (trend-)stationarity, Homm and Breitung (2012) also discuss some variants of detectors based on several underlying test statistics, some of which may more easily be extended to cointegration monitoring than others.

component and the number of integrated regressors included. Additionally, the critical values now depend upon the length of the calibration period as a fraction of the – actual or maximal – sample considered. This in turn implies that the number of periods on which monitoring can maximally be performed has to be decided upon ex ante, i.e., we propose a *closed-end* monitoring procedure. Details on this are given in Section 2, with this issue also discussed in some detail in Homm and Breitung (2012, Section 3).

We derive consistency of our procedures against fixed alternatives and also study local asymptotic power (presented in Supplementary Appendix B due to space constraints). At the end of the theory Section 2 we briefly discuss a few relevant properties of our procedures in case of additional forms of structural changes not considered in full detail in the paper. The theoretical results are complemented by finite sample simulations to study the rejection probabilities under both the null as well as under alternatives; the latter being reported in the form of size-corrected power. We also include the end-of-sample break-point test for cointegrating relationships of Andrews and Kim (2006) in the simulations. Inference for this test is based on sub-sampling and thus general forms of error serial correlation and regressor endogeneity can be accommodated whilst asymptotically controlling size. This wide applicability is a similarity to our approach and is not given for other procedures available in the literature, e.g., Chen et al. (2009), Steland and Weidauer (2013) or Wang et al. (2014). This makes the Andrews-Kim test, despite its different focus, a natural candidate for comparison. Finally, given their importance in practice, we investigate also the *detection times*, which serve as natural estimates of a possible break-point, of our monitoring procedures in the simulations. Note that the Andrews and Kim (2006) test, being a single retrospective test, by construction cannot deliver estimated break-points.

In our application we revisit one aspect of Anundsen (2015), who tests the stability of two related cointegrating relationships between housing prices and fundamentals that arise from equilibrium and no-arbitrage conditions. Using recursive estimation and cointegration testing in a VAR framework, Anundsen (2015) finds evidence for a breakdown of one of these two relationships already in 2000:Q4. This recursive testing approach is by construction not (asymptotically) size-controlled.[2] To overcome this problem we apply our monitoring procedure to these relationships. Depending upon specification of the equilibrium relationship and estimation method used, we detect a break-point between 2003:Q2 and 2007:Q3. While our detected break-points are later, they are still before the collapse of US housing prices started.

The paper is organized as follows: Section 2 presents the theory. Section 3 presents finite sample simulation results and the empirical application is presented in Section 4. Section 5 briefly summarizes and concludes. Three supplementary appendices are available: Supplementary Appendix A contains all proofs, Supplementary Appendix B provides further simulation results and Supplementary Appendix C consists of tables with critical values for a variety of specifications.

2. Monitoring Cointegration

We consider monitoring a potential structural change in a cointegrating regression (with the precise assumptions given below) of the form:

$$y_t = \begin{cases} D_t'\theta_D + X_t'\theta_X + u_t, & t = 1, \ldots, [rT] \\ D_t'\theta_{D,1} + X_t'\theta_{X,1} + u_t, & t = [rT]+1, \ldots, T \end{cases} \quad (1)$$

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

with scalar y_t, $D_t \in \mathbb{R}^p$ the deterministic trend function, X_t a k-dimensional vector random walk regressor and $0 < m \leq r < 1$. Under the null hypothesis no structural change occurs, i.e., $\theta_1 = [\theta_{D,1}', \theta_{X,1}']' = [\theta_D', \theta_X']' = \theta$ and u_t is an I(0) process throughout. Under the alternative either the parameters change or the relation turns spurious (or

[2] Anundsen (2015) contains many other interesting aspects on top of the recursive cointegration testing. He constructs, e.g., a simple bubble-indicator and analyzes the relationship of this indicator to other real and financial series.

both) at a sample fraction $[rT]$ larger than $[mT]$. Thus,

$$H_0 : \begin{cases} \theta_1 = \theta \quad \text{for all } m \leq r < 1 \\ \text{and } u_t, \ t = 1, \ldots, T \text{ is I(0)} \end{cases} \tag{3}$$

and

$$H_1 : \begin{cases} \theta_1 \neq \theta \quad \text{for some } m \leq r < 1 \text{ or} \\ u_t, \ t = 1, \ldots, [rT] \text{ is I(0) and} \\ u_t, \ t = [rT]+1, \ldots, T \text{ is I(1)} \quad \text{for some } m \leq r < 1 \end{cases} \tag{4}$$

Remark 1. The case $\dim(X_t) = 0$ is considered in full detail in the longer working paper Wagner and Wied (2015). In this case it suffices to consider monitoring based on the OLS residuals rather than upon modified OLS residuals. In analogy to cointegration monitoring we refer to this case as stationarity monitoring.

With respect to the trend function we assume:

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 $0 \leq s \leq 1$*

$$\lim_{T \to \infty} \sqrt{T} G_D^{-1} D_{[sT]} = D(s), \tag{5}$$

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

If, e.g., $D_t = (1, t, t^2, ..., t^{p-1})'$, then $G_D = \text{diag}(T^{1/2}, T^{3/2}, T^{5/2}, \ldots, T^{p-1/2})$ and $D(z) = (1, z, z^2, ..., z^{p-1})'$. With respect to the stacked error process $\eta_t := [u_t, v_t']'$ we make the following "high-level" assumption under the null hypothesis:

Assumption 2.

(a) The stationary process $\{\eta_t\}$ fulfills

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \eta_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \Rightarrow \Omega^{1/2} W(s) = B(s), \tag{6}$$

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*

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} := \sum_{j=-\infty}^{\infty} \mathbb{E}(\eta_t \eta_{t-j}'). \tag{7}$$

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

$$\frac{1}{T}\sum_{t=1}^{[sT]} S_t^\eta \eta_t' \Rightarrow \int_0^s B(r)dB(r)' + \Delta, \tag{8}$$

with $\Delta := \sum_{j=0}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$ *partitioned similarly as* Ω*.*

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

The results posited in Assumption 2 are standard in the cointegration literature. They 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

$$\Omega^{1/2} = \begin{bmatrix} \omega_{u\cdot v} & \lambda_{uv} \\ \mathbf{0} & \Omega_{vv}^{1/2} \end{bmatrix}, \tag{9}$$

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 and used in this paper.

Given Assumption 2 we can formalize our notions of I(0) and I(1) processes. A process $\{u_t\}$ is an I(0) process if it holds that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} u_t \Rightarrow \omega W(s)$, with $0 < \omega < \infty$ and $W(s)$ standard Brownian motion. Accordingly an I(1) process, i.e., a summed up I(0) process, fulfills $\frac{1}{\sqrt{T}} u_{[sT]} \Rightarrow \omega W(s)$ with ω and $W(s)$ as before. Thus, under the corresponding alternative we assume that $\{u_t\}$ is an I(0) process fulfilling Assumption 2 until $[rT]$ and an I(1) process from $[rT]+1$ onwards.

An important feature of our procedure, similar to Chu et al. (1996), is that structural breaks are allowed to occur only after a *calibration period* of length $[mT]$, with $0 < m < 1$. This calibration period is required for (consistent) estimation of the parameter vector θ and the conditional long-run variance $\omega^2_{u\cdot v}$. In this respect modified least squares estimators are required to obtain scaled residual partial sum processes that are asymptotically nuisance parameter free (up to a scalar variance parameter), which in turn allows to simulate critical values. We consider three modified estimation procedures, 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). Given space limitations we assume that these methods are known to the reader.

The monitoring procedure considers the residuals and assesses whether they become "too large" over time as an indication or structural change. Consider specifically the case of FM-OLS, with $y_t^+ := y_t - \Delta X_t' \hat{\Omega}_{vv,m}^{-1} \hat{\Omega}_{uv,m}$. The fully modified residuals are given by:

$$\begin{aligned} \hat{u}_{t,m}^+ &:= y_t^+ - D_t'\hat{\theta}_{D,m} - X_t'\hat{\theta}_{X,m} \\ &= u_t - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} - D_t'\left(\hat{\theta}_{D,m} - \theta_D\right) - X_t'\left(\hat{\theta}_{X,m} - \theta_X\right), \end{aligned} \tag{10}$$

where $\hat{\theta}_{D,m}$ and $\hat{\theta}_{X,m}$ denote the FM-OLS coefficient estimates and $\hat{\Omega}_m$ denotes the long-run variance estimate, all computed from the pre-break sample $1, \ldots, [mT]$ only. Long-run variance estimation is performed using the stacked error process $\hat{\eta} := [\hat{u}_{t,m}, v_t']'$ for $t = 1, \ldots, [mT]$ and with $\hat{u}_{t,m}$ denoting the OLS residuals obtained form estimation over the calibration period. The corresponding scaled partial sum process is given by:

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \hat{u}_{t,m}^+ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} u_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} v_t'\hat{\Omega}_{vv,m}^{-1}\hat{\Omega}_{vu} - \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} D_t'\left(\hat{\theta}_{D,m} - \theta_D\right) \\ &\quad - \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} X_t'\left(\hat{\theta}_{X,m} - \theta_X\right) \end{aligned} \tag{11}$$

The above equation (11) is informative for understanding why the detector defined below works. First, under the null – if long run variances and lead/lag choices are performed to ensure consistent estimation – the scaled partial sum residual process converges to a function of Brownion motions. Under the alternative the scaled partial sum residual process diverges: In case that $\{u_t\}$ changes its behavior to I(1) from $[rT]+1$ onwards, the first term diverges for $s > r$. A similar reasoning holds for parameter changes, since due to estimation only in the calibration period $\hat{\theta}_m \to \theta$, whereas under parameter change $\theta_1 \neq \theta$ from $[rT]+1$ onwards. This leads to divergence of the third and/or fourth time in (11).

Using the notation $\widehat{S}_i^+ := \sum_{t=1}^{i} \hat{u}_{t,m}^+$, the detector, using either FM-OLS or D-OLS[3], is defined by

$$\widehat{H}^{m,+}(s) \quad := \quad \frac{1}{\hat{\omega}_{u\cdot v,m}^2}\left(\frac{1}{T}\sum_{i=[mT]+1}^{[sT]}\left(\frac{1}{\sqrt{T}}\widehat{S}_i^+\right)^2\right), \tag{12}$$

with $\hat{\omega}_{u\cdot v,m}^2 := \hat{\Omega}_{uu,m} - \hat{\Omega}_{uv,m}\hat{\Omega}_{vv,m}^{-1}\hat{\Omega}_{vu,m}$. Clearly, setting $m = 0$ and $s = 1$ leads to the Shin (1994) statistic.[4]

[3]In case the monitoring 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 for FM-OLS 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 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.

[4]In the working paper Wagner and Wied (2015) we consider, directly inspired by Chu et al. (1996), the detector

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

We have furthermore also considered a "self-normalized" version

$$\widehat{H}_{sn}^{m,+}(s) \quad := \quad \frac{\sum_{i=[mT]+1}^{[sT]}\left(\widehat{S}_i^+\right)^2}{\sum_{i=1}^{[mT]}\left(\widehat{S}_i^+\right)^2},$$

a detector based on moving windows, $\widehat{H}_{mov}^{m,+}(s) := \frac{1}{\hat{\omega}_{u\cdot v,m}^2}\frac{1}{T}\sum_{i=[sT]-[nT]}^{[sT]}\left(\frac{1}{\sqrt{T}}\widehat{S}_i^+\right)^2$ or detectors based on recursive residuals. These variants perform relatively similar and are in some (of the simulation) cases

Throughout this paper we implicitly assume that long-run variances are consistently estimated, ensured, e.g., by assuming to be in the framework covered by Jansson (2002); and that lead/lag choices are such that D-OLS estimation is consistent (see, e.g., Kejriwal and Perron, 2008 or Choi and Kurozumi, 2012). This leads to the first proposition covering the null behavior of the FM-OLS or D-OLS based detectors:

Lemma 1. *Let the data be generated by* (1) *and* (2) *with Assumptions 1 and 2 in place. Define* $J(s) := [D(s)', W_v(s)']'$. *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*

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

This implies that

$$\widehat{H}^{m,+}(s) \Rightarrow \int_m^s \widehat{W}_{u\cdot v}^2(z)dz =: \mathcal{H}^{m,+}(s). \tag{14}$$

Note that the process $\widehat{W}_{u\cdot v}(s)$ depends upon D_t, the number of integrated regressors k and the pre-break fraction m. Consequently also the critical values depend upon these quantities. We neglect these dependencies for notational brevity in the paper.

Let us now turn to the detector based upon the IM-OLS estimator of Vogelsang and Wagner (2014), which has the advantage that for parameter estimation no kernel and bandwidth or lead and lag choices are required. The IM regression is given by

$$S_t^y = S_t^{D\prime}\theta_D + S_t^{X\prime}\theta_X + X_t'\varphi + S_t^u, \tag{15}$$

with $S_t^y = \sum_{j=1}^t y_j$ denoting the partial sums of y_t, and similar definitions of S_t^D, S_t^X and S_t^u. We denote the corresponding OLS residuals, with estimation again performed on the calibration period $1, \ldots, [mT]$, using the same notation for the coefficient estimates as

considered outperformed by the version considered here.

before, by

$$\hat{S}^u_{t,m} := S^y_t - S^{D\prime}_t\hat{\theta}_{D,m} - S^{X\prime}_t\hat{\theta}_{X,m} - X_t'\hat{\varphi}_m \tag{16}$$
$$= S^u_t - X_t'\hat{\varphi}_m - S^{D\prime}_t(\hat{\theta}_{D,m} - \theta_D) - S^{X\prime}_t(\hat{\theta}_{X,m} - \theta_X).$$

Given that IM-OLS estimation is performed on the partial summed regression, in the detector the corresponding residuals need not be partial summed anymore, leading to:[5]

$$\hat{I}^m(s) := \frac{1}{\hat{\omega}^2_{u\cdot v,m}}\left(\frac{1}{T}\sum_{i=[mT]+1}^{[sT]}\left(\frac{1}{\sqrt{T}}\hat{S}^u_{i,m}\right)^2\right), \tag{17}$$

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 null behavior of this detector follows immediately.

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

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{[sT]}\Delta\hat{S}^u_{t,m} \Rightarrow \omega_{u\cdot v}\left(W_{u\cdot v}(s) - f(s)'\left(\int_0^m f(z)f(z)'dz\right)^{-1}\int_0^m [F(m)-F(z)]\,dW_{u\cdot v}(z)\right)$$
$$=: \omega_{u\cdot v}\tilde{P}_m(s), \tag{18}$$

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

$$\hat{I}^m(s) \Rightarrow \int_m^s \tilde{P}_m(z)^2 dz =: \mathcal{I}^m(s). \tag{19}$$

Use $\widehat{F}^m(s)$ to denote either $\widehat{H}^{m,+}(s)$ or $\widehat{I}^m(s)$. The null hypothesis is declared rejected, if the *weighted* detector, $\frac{\widehat{F}^m(s)}{g(s)}$, where $g(s)$ is a weighting function to be chosen, exceeds

[5] The discussion in Footnote 4 applies analogously to the IM-OLS-based detectors.

a critical value for the first time, with this time point referred to as detection time, i.e.

$$\tau_m := \min_{s:[mT]+1\leq[sT]\leq T} \left\{ \frac{\widehat{F}^m(s)}{g(s)} > c \right\}. \tag{20}$$

In case that $\frac{\widehat{F}^m(s)}{g(s)} \leq c$ for all $m \leq s \leq 1$ we set $\tau_m = \infty$. Thus, a finite value τ_m indicates a rejection of the null hypothesis and at the same time gives information about the location of the potential break-point.

The weighting function and critical value have to be chosen such that under the null hypothesis it holds that

$$\begin{aligned} \lim_{T\to\infty} \mathbb{P}(\tau_m < \infty) &= \lim_{T\to\infty} \mathbb{P}\left(\min_{s:[mT]+1\leq[sT]\leq T} \left\{ \frac{F^m(s)}{g(s)} > c \right\} < \infty \right) \\ &= \lim_{T\to\infty} \mathbb{P}\left(\sup_{s:[mT]+1\leq[sT]\leq T} \frac{F^m(s)}{g(s)} > c \right) \\ &= \mathbb{P}\left(\sup_{m\leq s\leq 1} \frac{\mathcal{F}^m(s)}{g(s)} > c \right) = \alpha, \end{aligned} \tag{21}$$

with α denoting the chosen significance level and $\mathcal{F}^m(s)$ the limit of $\widehat{F}^m(s)$. For simplicity we only consider continuous positive and bounded weighting functions, see also Aue et al. (2012, Assumption 3.6). This allows to derive the above result using the established functional central limit theorems for $\widehat{F}^m(s)$ and the continuous mapping theorem. In other words, for bounded and positive weighting functions it can be shown for the detectors that there exist – depending upon the chosen $g(s)$ – critical values 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 and assume 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, FM-OLS or D-OLS on the one hand and IM-OLS on the other, such* *that*

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{F}^{m,+}, g, c(\alpha, g)) < \infty) = \alpha. \qquad (22)$$

Clearly, the choice of the weighting function $g(s)$ 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 is in general impossible to derive analytically tractable optimal weighting functions, e.g., with respect to minimal expected delay whilst asymptotically controlling size.[6] Here we base our weighting function choice on the expected value function of the (limiting distribution of the) detector, which leads – most easily seen in case of stationarity monitoring – to $g(s) = s^3$ in the intercept only case and $g(s) = s^5$ in the linear trend case, as is outlined in more detail in the longer working paper. We henceforth use these weighting functions only and also the critical values available in Supplementary Appendix C correspond to this choice of weighting function. The critical values are given for $k = 0, 1, \ldots, 4$ stochastic regressors for the intercept only and intercept and linear trend cases. The detectors discussed in this paper are implemented in the R package `cointmonitoR` (Aschersleben et al., 2016) that contains several variants of the monitoring procedure. It accesses the `cointReg` R package of Aschersleben and Wagner (2016) for parameter estimation and inference for all three modified OLS approaches. Both our finite sample simulations as well as the empirical application are performed with these packages.

It is important to understand the meaning of m and T for our procedure, since the choice of m has important consequences for the properties of the procedure. It is most convenient

[6] 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". 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; and they appear hard to obtain.

to interpret T as the sample size including the out-of-sample monitoring period. Thus, denote the last actually available observation by T_0, and $T = T_0 + H$. Using this notation $H > 0$ means that one intends to use the procedure for out-of-sample monitoring, whereas $H = 0$ corresponds to the special case where monitoring takes place on historic data only (as in the application in Section 4).[7] The fact that the critical values depend upon m means that a decision has to be made about both the length of the calibration period and of an out-of-sample monitoring period H prior to the analysis. The latter necessity renders our procedure a *closed-end* monitoring procedure. The calibration period will be chosen as large as possible (as a sub-sample $1, \ldots, T_C$ of $1, \ldots, T_0$) in order to increase the precision of the parameter estimates whilst avoiding the risk of having a structural break in the calibration period. Now, m is given by $m = \frac{T_C}{T_0+H}$. Thus, choosing H larger implies that m is smaller, which in turn implies that the critical values are larger (since they are decreasing in m). This decreases *ceteris paribus*, despite asymptotic size-control, the empirical rejection probabilities under both the null and the alternative. Consequently one should choose the monitoring period as short as possible by using as large a calibration period as possible and an out-of-sample monitoring period as short as possible.

It remains to establish the behavior of the detectors under the alternative. With respect to the alternative there are three "different types" of deviations from the null, which need to be analyzed separately. First, changes in $\{u_t\}$ from I(0) to I(1) behavior. Second, breaks in the trend parameters θ_D and third breaks in the slope parameters θ_X. The third case is in some sense equivalent to the first. Consider the case that $\theta_{X,1} \neq \theta_X$ for $t = [rT] + 1, \ldots, T$. In this case we can write for $t > [rT]$

$$\begin{aligned} y_t &= D_t'\theta_D + X_t'\theta_{X,1} + u_t \\ &= D_t'\theta_D + X_t'\theta_X + X_t'(\theta_{X,1} - \theta_X) + u_t. \end{aligned} \tag{23}$$

[7] It is to a certain extent a semantic question whether this should be called monitoring or structural break testing. The procedures in any case have an online monitoring flavor as for the calculation of the detectors at a certain time point only observations up to that point in time are required.

Clearly, this implies that in the residual process starting from $[rT]$ onwards an integrated process given by $X_t'(\theta_{X,1} - \hat{\theta}_{X,m})$ is present. This component remains present as an I(1) process also in the limit due to consistency of $\hat{\theta}_{X,m} \to \theta_X \neq \theta_{X,1}$. 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 where $\{u_t\}$ changes its behavior from I(0) to I(1).

For all three cases we consider both fixed and local alternatives. The fixed alternative for the first case is, obviously, given by the situation that $\{u_t\}$ changes its behavior from I(0) to I(1) from some point $[rT] > [mT]$ onwards. As local alternatives we consider (inspired by Cappuccio and Lubian, 2005) the situation that 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

$$u_t = u_t^0 + \frac{\delta}{T} \sum_{i=[rT]+1}^{t} \xi_i, \tag{24}$$

with $\{u_t^0\}$ and $\{\xi_t\}$ independent processes both fulfilling Assumption 2, with long-run variances ω^2 and ω_ξ^2, and $\delta > 0$.[8] I.e., under the considered local alternatives the process $\{u_t\}$ 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 $m \leq r < s \leq 1$):

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \hat{u}_{t,m}^{+} \Rightarrow \omega_{u \cdot v} \widehat{W}_{u \cdot v}(s) + \delta \omega_\xi \int_r^s (W_\xi(z) - W_\xi(r)) dz. \tag{25}$$

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

For breaks in the parameter vector the fixed alternative is clearly $\theta_1 \neq \theta$. For local alternatives we have to differentiate between the trend parameters θ_D and the slope

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

parameters θ_X due to the generally different convergent rates of the components of θ_D and super-consistency of the estimation of θ_X. In particular, for consistency against fixed alternatives in the trend slopes we need the condition

$$\lim_{T\to\infty} \frac{1}{\sqrt{T}} \sum_{t=[rT]+1}^{T} D_t'(\theta_D - \theta_{D,1}) = \pm\infty. \tag{26}$$

For local alternatives with respect to the trend slope θ_D the additional constraint to be fulfilled is given by

$$\int_r^1 D(z)'dz\Delta_{\theta_D} \neq 0, \tag{27}$$

to have non-trivial local asymptotic power, with Δ_{θ_D} put in context below in the proposition.

Proposition 2. (Consistency and Local Asymptotic Power)
Let the data be generated by (1) *and* (2) *with Assumption 1 in place and* $\{\eta_t\}$ *fulfilling Assumption 2 until* $[rT]$, *with* $m \leq r < 1$. *Assume again that* $g(s)$ *is continuous with* $0 < g(s) < \infty$.

(a) Let (i) $\{u_t\}$ *be an I(1) process (as specified in* H_1*) from* $[rT]+1$ *onwards; or (ii)* $\theta_1 \neq \theta$, *with* (26) *fulfilled, from* $[rT]+1$ *onwards. Then the monitoring procedure is consistent, i.e., for any* $0 < c < \infty$ *it holds that*

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

(b) Let (i) $\{u_t\}$ *be as specified in* (24) *from* $[rT]+1$ *onwards; (ii)* $\theta_{X,1} = \theta_X + \frac{1}{T}\Delta_{\theta_X}$ *from* $[rT]+1$ *onwards with* $\Delta_{\theta_X} \neq 0$; *or (iii)* $\theta_{D,1} = \theta_D + G_D^{-1}\Delta_{\theta_D}$ *from* $[rT]+1$ *onwards with* G_D *as in Assumption 1 and* Δ_{θ_D} *fulfilling* (27). *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* $\delta = \delta(c, g)$,

$\Delta_{\theta_X} = \Delta_{\theta_X}(c,g)$ *or* $\Delta_{\theta_D} = \Delta_{\theta_D}(c,g)$ *fulfilling* (27) *such that*

$$\lim_{T\to\infty} \mathbb{P}(\tau_m(\widehat{F}^m, g, c(\alpha,g)) < \infty) \geq 1-\epsilon. \tag{29}$$

Clearly, the results in part (b) stem from the convergence results for $\widehat{F}^m(s)$ under the considered local alternatives. Consider, e.g., case (i) and in relation to (25) FM/D-OLS based monitoring, where for $1 \geq s \geq r \geq m$ it holds that:

$$\begin{aligned} \widehat{H}^m(s) \quad \Rightarrow \quad & \mathcal{H}^m(s) + 2\delta\frac{\omega_\xi}{\omega}\int_r^s \widehat{W}_{u\cdot v}(z)\left(\int_r^z (W_\xi(g) - W_\xi(r))\, dg\right) dz + \\ & +\delta^2\left(\frac{\omega_\xi}{\omega}\right)^2 \int_r^s \left(\int_r^z (W_\xi(g)-W_\xi(r))\, dg\right)^2 dz. \end{aligned} \tag{30}$$

This result shows that the magnitude of the additional terms depends, in addition to δ, upon 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.

Of course, the detectors have the discussed power properties under combinations of the changes considered separately in the above proposition. The local asymptotic power properties of the procedures are discussed in some detail in Supplementary Appendix B. A few additional useful observations about the detectors can be made: First, the detectors are consistent against the alternative of $\{u_t\}$ turning to a near-integrated process, as defined in Phillips (1987). Second, consistency prevails also against a change of $\{u_t\}$ from I(0) to being fractionally integrated. Third, the procedures can be used to detect bubbles, defined as explosive behavior of $\{u_t\}$, compare Astill et al. (2015), Homm and Breitung (2012) or Phillips et al. (2011). Considering, e.g., asset prices, the null hypothesis of random walk I(1) behavior is assessed against the alternative of explosive behavior from a certain point onwards. In this case the first difference of the residual series (after potential detrending) can be used, since under the null the first difference will be stationary, whereas under the alternative also the first difference will be explosive from the break-point onwards. Typically, bubbles are considered temporary phenomena and it is thus convenient that procedures like ours allow to detect *episodes* of I(1) or explosive behavior of $\{u_t\}$ as long as this episode (or these episodes) have asymptotically positive length (when scaling the time axis to [0,1] in the asymptotic analysis).

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):

$$y_t = \mu + \gamma t + x_{1t}\beta_1 + x_{2t}\beta_2 + u_t, \tag{31}$$

$$x_{it} = x_{i,t-1} + v_{it}, \quad x_{i0} = 0, \quad i = 1, 2, \tag{32}$$

where

$$u_t = \rho_1 u_{t-1} + \varepsilon_t + \rho_2(e_{1t} + e_{2t}), \quad u_0 = 0, \tag{33}$$

$$v_{it} = e_{it} + 0.5 e_{i,t-1}, \quad i = 1, 2, \tag{34}$$

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 $\mu = 3$, $\beta_1 = \beta_2 = \gamma = 1$.[9] 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 detectors 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

[9] Results for the intercept only case with $\gamma = 0$ are available upon request. A rough one sentence summary is that things work typically better in the intercept only case.

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.

We compare our monitoring procedures with the end-of-sample structural break test in cointegrating regressions of Andrews and Kim (2006). In particular we consider their test statistic P based on the full sample residuals using the same three modified least squares estimators.[10] In our notation, the three corresponding test statistics are given by $P = \sum_{[mT]+1}^{T} (\hat{u}_{t,1}^{+})^2$, using again the same notation for both the FM- and D-OLS residuals, and $P = \sum_{[mT]+1}^{T} (\Delta\hat{S}_{t,1}^{u})^2$ when using IM-OLS. Critical values are calculated from the empirical distribution of sub-sampled P statistics, using $2[mT] - T + 1$ consecutive sub-samples of length $T - [mT]$ in the calibration period $t = 1, \ldots, [mT]$. Note that this construction necessarily implies that $m > 0.5$.[11] The considered sample sizes are $T = 200, 500$ and for all procedures the number of replications is 10,000. Monitoring respectively testing is performed at the 5% nominal level.

We start by considering empirical null rejection probabilities for our detectors 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 1 and $T = 500$ in Figure 2. Several main patterns in line with expectations emerge: First, size distortions decrease with increasing m and increasing sample size T. For $T = 200$ and very small values of m the size distortion of D-OLS are substantially larger than for FM-OLS and IM-OLS. Second, for given T 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 smallest size distortions occur for IM-OLS. The larger ρ_1, ρ_2, the bigger is the performance advantage of IM-OLS over the other two estimation methods.

[10] The main part of the discussion in Andrews and Kim (2006) focuses on the OLS residuals since sub-sampling avoids the need for modification. However, it is of course possible (see also the discussion in their appendix) to use residuals based on consistent modified least squares estimators.

[11] This stems from the fact that Andrews and Kim (2006) are concerned with structural breaks at the end of the sample, i.e., in our notation they are concerned with very large values of m. To be sure, the letter m has a different meaning in Andrews and Kim (2006) than here.

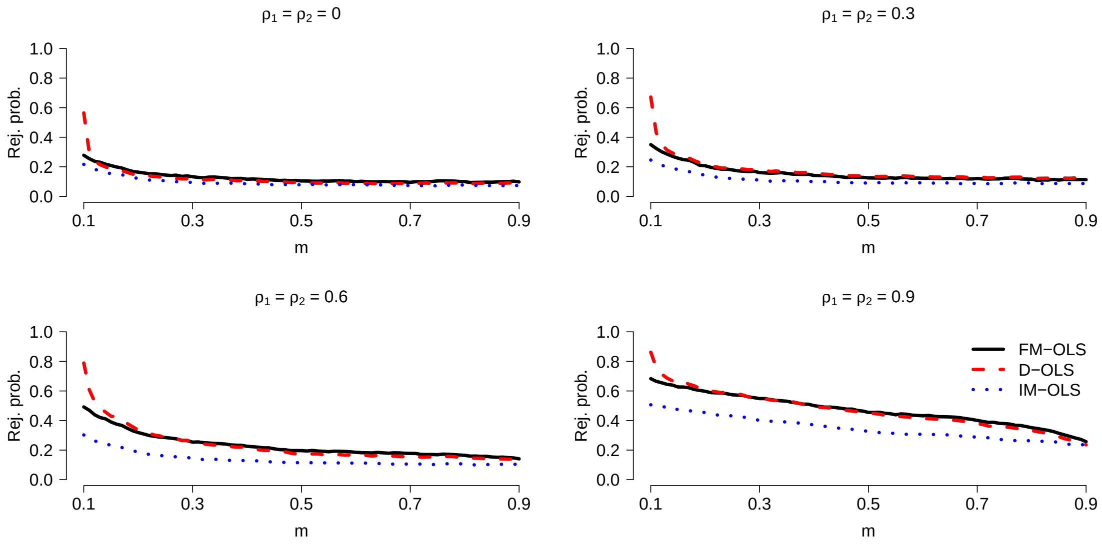

Figure 1: Empirical null rejection probabilities for a grid of values of m and $T = 200$.

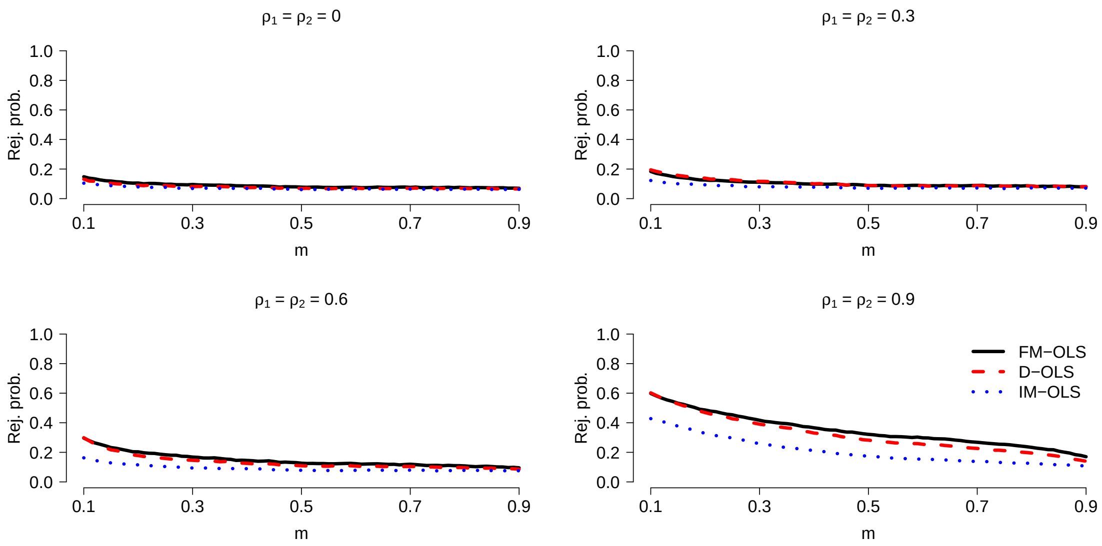

Figure 2: Empirical null rejection probabilities for a grid of values of m and $T = 500$.

In Figures 3 and 4 we consider the same setting for the P-test of Andrews and Kim (2006), now for a grid of mesh 0.01 for $m = 0.55, \ldots, 0.9$. The results are quite clear: For the smaller values of m the size distortions are huge, but this is of course not the situation for which this test is designed. Interestingly, size distortions are partly only reduced by a small extent when moving from $T = 200$ to $T = 500$. For large values of m the size distortions are small and partly smaller than for our detectors. Depending upon precise setting, the Andrews and Kim (2006) test leads to lower size distortions than our detectors for m starting around $m = 0.7$ for $\rho_1, \rho_2 = 0$ to around $m = 0.9$, i.e., the largest value displayed, for $\rho_1, \rho_2 = 0.9$. With respect to the three different estimators, the smallest size distortions occur mostly with IM-OLS and sometimes with D-OLS, whilst FM-OLS typically leads to the largest size distortions. More details are provided in Supplementary Appendix B.

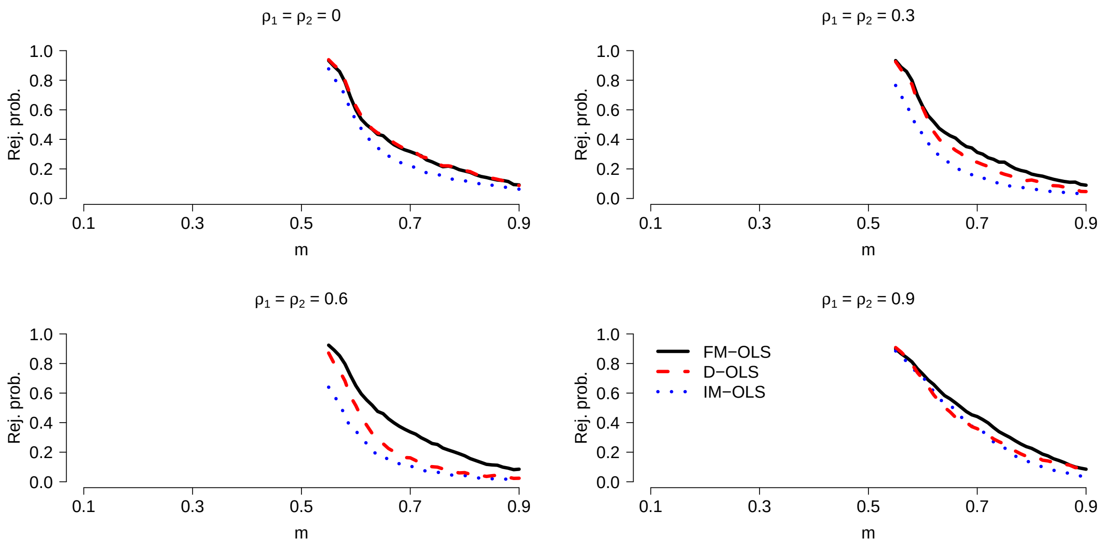

Figure 3: Andrews-Kim: Empirical null rejection probabilities for a grid of values of m and $T = 200$.

We next turn to *size-corrected* power, focusing here only on the case that $\{u_t\}$ changes its behavior from I(0) to I(1) at sample fraction $[rT]$. We refer to this case as I(1) breaks henceforth. The other cases of the alternative, i.e., trend and slope (coefficient) breaks, are considered in Supplementary Appendix B due to space constraints. We consider for both m and r all three values $0.25, 0.5$ and 0.75 and thus include cases where the break occurs in the calibration period ($r < m$) to assess the implications of a *too long* calibration period. It can be shown, most easily for stationarity monitoring, that in case $r < m$ the detectors are of order $O_p(T/b_T)$, with b_T denoting the bandwidth chosen for long-run variance estimation. Given that in case of I(1) errors, under the considered alternative, typically large bandwidths are chosen, we expect a low divergence rate of the detector in

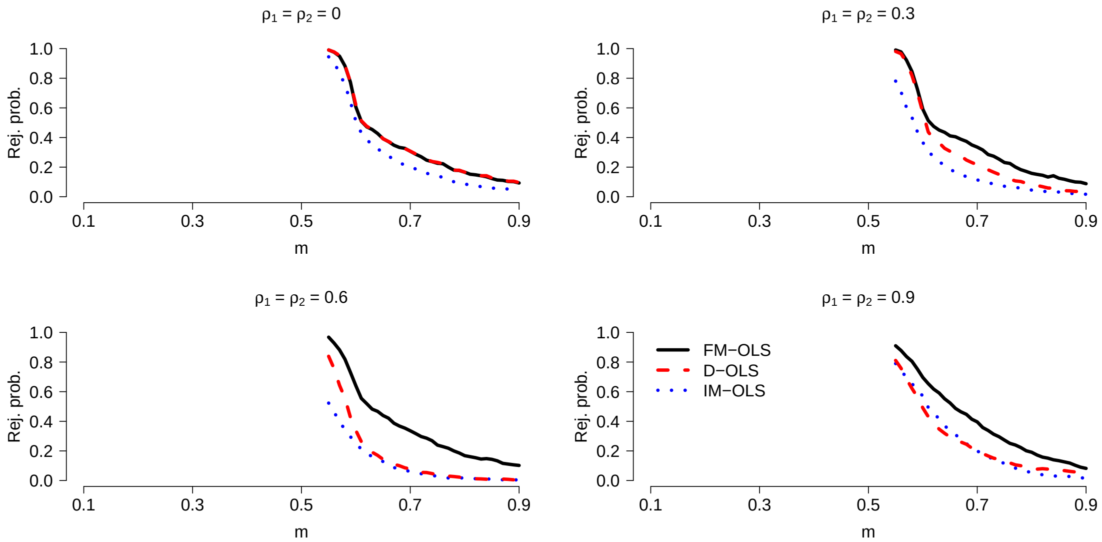

Figure 4: Andrews-Kim: Empirical null rejection probabilities for a grid of values of m and $T = 500$.

the $r < m$ cases and thus low power. This is exactly what happens.

The results are shown for $T = 200$ in Table 1 and for $T = 500$ in Table 2. In addition to the above observation also some additional expectations are confirmed: Power increases with the sample size T, decreases with increasing ρ_1, ρ_2 and depends upon both m and r. With the exception of $T = 200$, $m = 0.25$ and large values of ρ_1, ρ_2, power is largest for $m = r$. In case that $m \leq r$ power decreases with increasing r (later breaks) for fixed m and increases with m (longer calibration period) for fixed r. The differences in power across estimation methods are minor and no clear ranking emerges. The fact that the IM-OLS based detector has often smaller size distortions under the null, but does not suffer from systematically lower power, makes this our slightly preferred choice.[12]

Finally we investigate the estimated detection times of our detectors against I(1) breaks. We show in Figure 5 the detection times for $\rho_1, \rho_2 = 0.9$ and $T = 200$ and in Figure 6 the detection times for $\rho_1, \rho_2 = 0$ and $T = 500$. We display the detection times in the form of Box-Whiskers plots. The numbers below the abbreviated method names indicate the null

[12]We abstain from presenting power here since size-correction cannot be performed in the usual way – or does not have the same effect – for sub-sampled statistics. In Supplementary Appendix B we report and discuss for completeness both raw and size-corrected power for the Andrews-Kim test.

$\rho_1 = \rho_2$ \ r		$m = 0.25$			$m = 0.5$			$m = 0.75$		
		0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
	FM	0.61	0.51	0.43	0.21	0.87	0.55	0.06	0.35	0.90
0	D	0.63	0.53	0.45	0.23	0.88	0.58	0.06	0.37	0.90
	IM	0.52	0.46	0.45	0.16	0.80	0.56	0.06	0.26	0.84
	FM	0.46	0.46	0.39	0.14	0.79	0.50	0.06	0.24	0.83
0.3	D	0.44	0.44	0.38	0.13	0.77	0.50	0.05	0.22	0.82
	IM	0.39	0.43	0.42	0.11	0.71	0.51	0.05	0.17	0.76
	FM	0.22	0.32	0.30	0.07	0.57	0.39	0.05	0.11	0.66
0.6	D	0.22	0.35	0.32	0.08	0.60	0.42	0.05	0.12	0.69
	IM	0.22	0.35	0.37	0.08	0.54	0.45	0.05	0.09	0.62
	FM	0.06	0.07	0.09	0.05	0.09	0.10	0.05	0.05	0.15
0.9	D	0.06	0.08	0.10	0.05	0.10	0.11	0.05	0.05	0.16
	IM	0.06	0.09	0.14	0.05	0.10	0.14	0.05	0.05	0.13

Table 1: Size corrected power against I(1) breaks for $T = 200$.

$\rho_1 = \rho_2$ \ r		$m = 0.25$			$m = 0.5$			$m = 0.75$		
		0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
	FM	0.94	0.57	0.45	0.60	0.99	0.60	0.11	0.72	0.99
0	D	0.95	0.58	0.48	0.63	0.99	0.64	0.12	0.74	0.99
	IM	0.88	0.52	0.46	0.50	0.97	0.59	0.08	0.62	0.98
	FM	0.87	0.53	0.42	0.48	0.97	0.56	0.08	0.61	0.98
0.3	D	0.87	0.52	0.44	0.47	0.97	0.58	0.08	0.61	0.98
	IM	0.80	0.49	0.44	0.38	0.95	0.56	0.07	0.50	0.95
	FM	0.67	0.44	0.37	0.25	0.91	0.49	0.06	0.40	0.93
0.6	D	0.70	0.47	0.39	0.28	0.92	0.52	0.06	0.44	0.94
	IM	0.64	0.46	0.42	0.23	0.88	0.52	0.06	0.34	0.89
	FM	0.09	0.17	0.18	0.05	0.29	0.24	0.05	0.06	0.41
0.9	D	0.09	0.20	0.22	0.06	0.33	0.29	0.05	0.06	0.47
	IM	0.09	0.24	0.31	0.05	0.32	0.36	0.05	0.06	0.41

Table 2: Size corrected power against I(1) breaks for $T = 500$.

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.[13] The six graphs within the figures display the six combinations of m and r from the nine combinations considered in the power tables for which $m \leq r$, i.e., we consider only the cases where the break does

[13]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.

not occur within the calibration period.

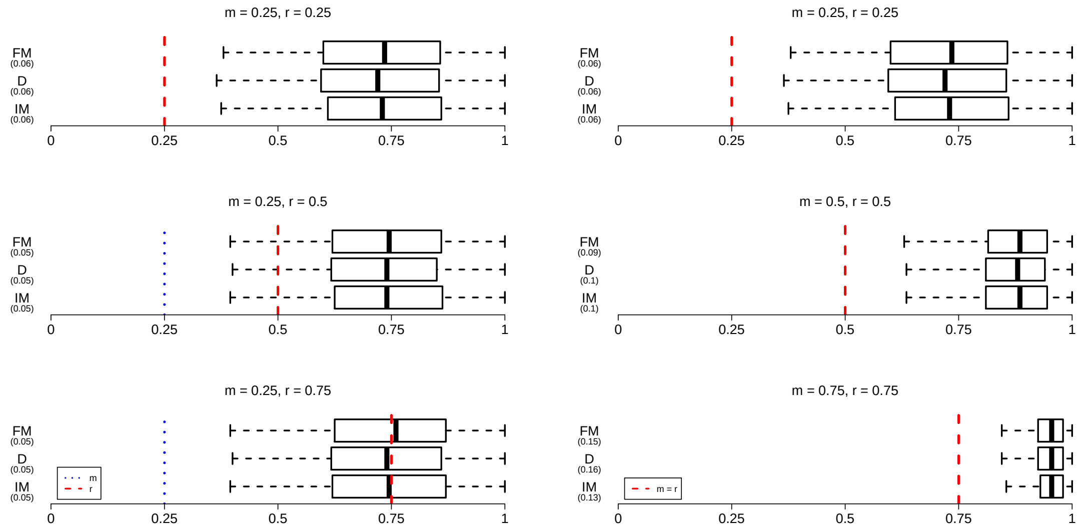

Figure 5: Detection times for I(1) breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.9$.

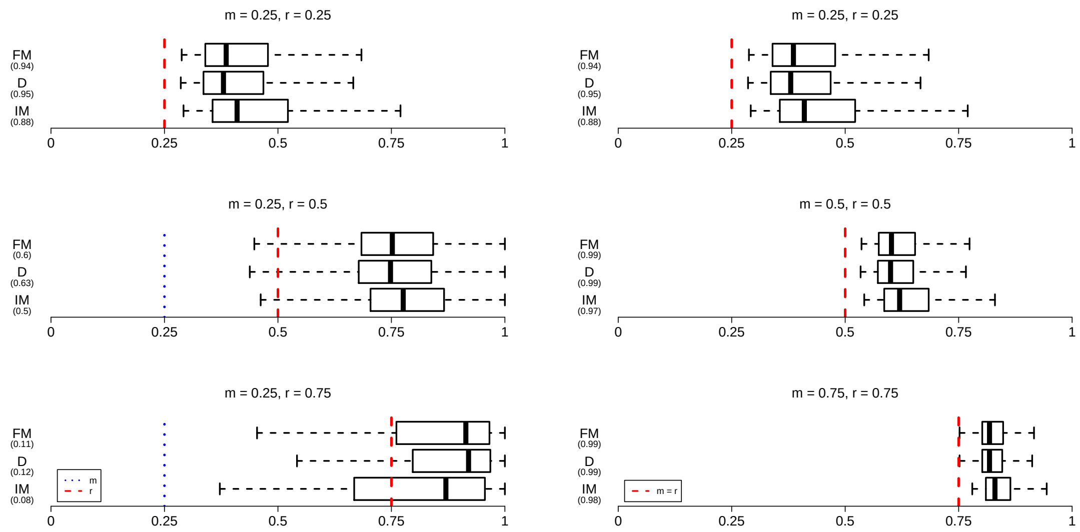

Figure 6: Detection times for I(1) breaks for $T = 500$ and $\rho_1 = \rho_2 = 0$.

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 throughout lead to smaller average delays. As expected (compare also some additional figures in Supplementary

Appendix B), increasing endogeneity and error serial correlation often lead to increasing detection times, i.e., bigger delays. For $m = 0.25$ and $r = 0.75$ detection occurs sometimes already *prior* to the structural change with IM-OLS. Increasing values of $m = r$ lead to smaller delays of detecting the structural change. The three methods typically lead to very similar delays. The delays are partly substantial, in particular for $\rho_1, \rho_2 = 0.9$ and $T = 200$. 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 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.[14]

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.[15] This leads to the following approximate equilibrium relationship (ignoring deterministic components and stochastic errors at this point):

$$p_t = \theta_r r_t + \theta_{UC} UC_t, \tag{35}$$

[14]Another illustrative application of our procedure is contained in the 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 stationarity of Euro area real exchange rates. Reynolds et al. (2017) use the developed procedure to assess deviations from triangular parity relationships.

[15]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.

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 mentioned above leads to the following approximate equilibrium relationship (again ignoring deterministic components and errors here):

$$p_t = \theta_y y_t + \theta_h h_t + \theta_{UC} UC_t, \tag{36}$$

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 (35) and (36), 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):[16]

$$p_t = \theta_c + \theta_t t + \theta_r r_t + \theta_{UC} UC_t + u_t \tag{37}$$

$$p_t = \theta_c + \theta_t t + \theta_y y_t + \theta_h h_t + \theta_{UC} UC_t + u_t, \tag{38}$$

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

[16] 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.

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 model are real housing prices p_t, real rents r_t and the real direct user cost UC_t. In the inverted demand 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 (37) and (38), 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

Method	Coefficient	Estimate	Std. Error	t-value	p-value
FM-OLS	θ_c	0.1856	0.0136	13.6398	0.0000
	θ_t	0.0008	0.0004	2.3189	0.0231
	θ_r	0.6855	0.1941	3.5314	0.0007
	θ_{UC}	-0.8564	0.2374	-3.6068	0.0006
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	9.7125	0.4972	19.5353	0.0000
	θ_t	-0.0066	0.0004	-15.9889	0.0000
	θ_y	0.6893	0.0473	14.5834	0.0000
	θ_h	1.6961	0.1466	11.5656	0.0000
	θ_{UC}	-0.3742	0.0674	-5.5533	0.0000
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.

included. 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 for the inverted demand model, compare Table 4.

The detected break-points are reported in Table 5. For the inverted demand model, the results show some variation in the detection times across methods, whereas the variation

	FM-OLS	D-OLS	IM-OLS
Price-to-rent model	2007:Q2	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.

is very small for the price-to-rent model. The rankings across methods are different for the two 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.[17] 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 based detectors, 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. Again we have to note that finding optimal weighting functions to minimize expected delay remains an important open but difficult issue.

Let us close this section by looking at the residuals in Figure 7 and the detectors in Figure 8. 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 7 are obtained from parameter estimation only over the calibration period that ranges until

[17]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 as early as 1996:Q3.
Earlier break-points are also detected when considering a calibration period only until 1990:Q2, i.e. prior to the onset of the recession in the US in the early 1990s. In this case the break-points range, depending upon model and estimator, from 1992:Q2 to 1999:Q3.

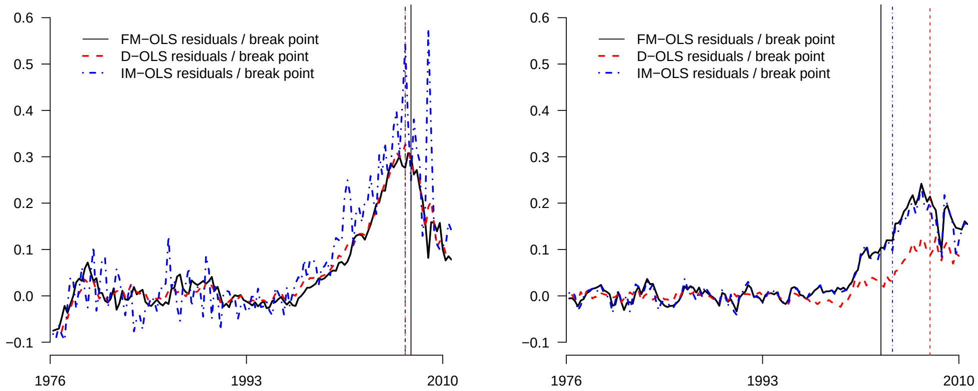

Figure 7: Residuals of the price-to-rent (left graph) and inverted demand (right graph) models. The vertical lines indicate the detected break points for the three used estimation methods. For IM-OLS we display the first differences of the residuals of the partial summed regression used in IM-OLS estimation.

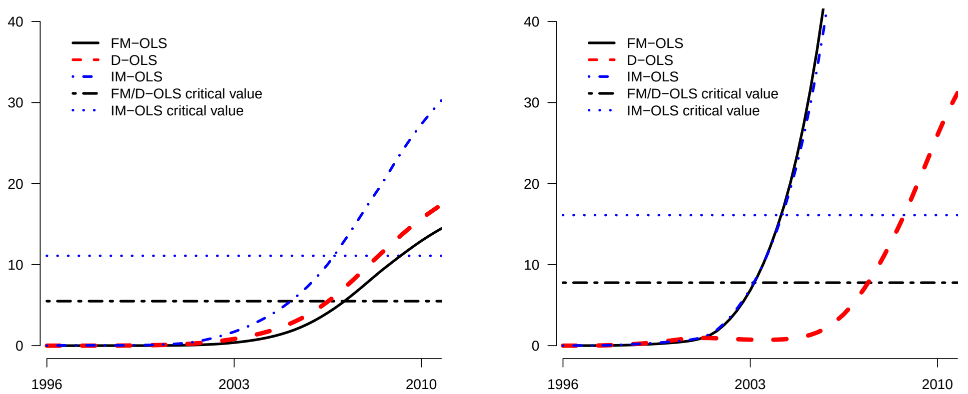

Figure 8: Detectors for the price-to-rent (left graph) and inverted demand (right graph) models. The two horizontal lines indicate the critical values for the three versions of the monitoring procedure.

1995:Q4. Thus, by construction until the end of 1995 the graphs display 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 7, 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 8 shows information also contained in the previous graph by plotting the detectors. The intersection of the detectors with the corresponding critical values occurs – by construction – exactly at the break-points displayed by means of vertical lines in Figure 7.

5. Summary and Conclusions

We have proposed a closed-end monitoring procedure for structural change in cointegrating regressions. The detector is inspired by Chu et al. (1996) in being based on parameter estimation over a calibration period and the cointegration test of Shin (1994) for the precise form of the statistic. To achieve nuisance parameter free limiting distributions despite error serial correlation and regressor endogeneity, parameter estimation rests upon modified least squares, in particular FM-OLS, D-OLS and IM-OLS. The asymptotic distributions coincide for FM-OLS and D-OLS, whilst IM-OLS leads to a different limit. The effects of sample size, regressor endogeneity and error serial correlation are as typically found in the cointegration literature. The finite sample performance of the three variants is often quite similar. In a variety of settings IM-OLS leads to smaller size distortions under the null. This fact in conjunction with essentially no differences in size-corrected power under the alternative makes IM-OLS our slightly preferred choice. Detection occurs by construction with delays, which are in some configurations substantial. In this respect, unsurprisingly, a longer calibration period, resulting in more precise parameter estimates, is beneficial. With respect to the estimated break-points there are no systematic differences across estimation methods. A comparison with the test of Andrews and Kim (2006) indicates that the latter performs very well – and potentially even better than our detectors – for the late break situation it has been designed for. For earlier breaks our procedure exhibits better performance, and additionally provides estimated break-points.

In our application using the data of Anundsen (2015) the estimated break-points range from 2003:Q2 to 2007:Q3 depending upon estimation method and relationship considered. For all cases, at least, the detected break-points are prior to the start of the fall of US housing prices. Here it has to be noted that a smaller calibration period leads to considerably earlier break-points as discussed in Footnote 17.

Several extensions of our work are conceivable. First, a better understanding of the impact of the weighting function on the performance of the monitoring scheme needs to be obtained. In extension of the experience collected so far, compare Footnote 4, a more systematic knowledge of the performance of variants of our detectors based on potentially other test statistics and other principles may be beneficial. Second, it may be relevant to consider monitoring schemes for multiple cointegrating relationships. Third, especially relevant for financial data, the impact of non-constant (conditional or unconditional) variances needs to be understood. A fourth extension, the opposite pair of null and alternative hypothesis is studied in Sakarya et al. (2015).

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. and M. Wagner (2016): *cointReg: Parameter Estimation and Inference in a Cointegrating Regression*, R package version 0.2.0.9000.

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.

——— (2016): *cointmonitoR: Consistent Monitoring of Stationarity and Cointegrating Relationships*, R package version 0.1.0.9000.

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.

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

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.

Homm, U. and J. Breitung (2012): "Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods," *Journal of Financial Econometrics*, 10, 198–231.

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

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.

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

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

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.

Reynolds, J., L. Sögner, M. Wagner, and D. Wied (2017): "The Triangular Parity and Market Dislocations," *Mimeo*.

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. 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.