

## Article

# Structural Break Tests Robust to Regression Misspecification

Alaa Abi Morshed <sup>1</sup>, Elena Andreou <sup>2</sup> and Otilia Boldea <sup>3,\*</sup>

<sup>1</sup> Department of Finance and Business Intelligence, AEGON Nederland, 2591 TV Den Haag, The Netherlands; alaa.abimorshed@aegon.nl

<sup>2</sup> Department of Economics and CEPR, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus; elena.andreou@ucy.ac.cy

<sup>3</sup> Department of Econometrics and Operation Research and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

\* Correspondence: o.boldea@uvt.nl

Received: 23 December 2017; Accepted: 14 May 2018; Published: 22 May 2018



**Abstract:** Structural break tests for regression models are sensitive to model misspecification. We show—analytically and through simulations—that the sup Wald test for breaks in the conditional mean and variance of a time series process exhibits severe size distortions when the conditional mean dynamics are misspecified. We also show that the sup Wald test for breaks in the unconditional mean and variance does not have the same size distortions, yet benefits from similar power to its conditional counterpart in correctly specified models. Hence, we propose using it as an alternative and complementary test for breaks. We apply the unconditional and conditional mean and variance tests to three US series: unemployment, industrial production growth and interest rates. Both the unconditional and the conditional mean tests detect a break in the mean of interest rates. However, for the other two series, the unconditional mean test does not detect a break, while the conditional mean tests based on dynamic regression models occasionally detect a break, with the implied break-point estimator varying across different dynamic specifications. For all series, the unconditional variance does not detect a break while most tests for the conditional variance do detect a break which also varies across specifications.

**Keywords:** structural change; sup Wald test; dynamic misspecification

**JEL Classification:** C01; C12

## 1. Introduction

There is vast literature on alternative structural break tests, as well as empirical evidence that many economic indicators went through periods of structural change. Most structural break tests are developed for the slope parameters of a regression model (see inter alia [Andrews 1993](#); [Andrews and Ploberger 1994](#); [Bai and Perron 1998](#); [Ploberger and Krämer 1992](#)).

Macroeconomic variables may often exhibit long-run mean shifts, that is, structural breaks in their unconditional mean. Mean shifts in unemployment rates, interest rates, GDP growth, inflation and other macroeconomic variables may signal permanent changes in the structure of the economy and are therefore themselves of interest to practitioners. Nevertheless, very few papers test for unconditional mean shifts; instead, most of the literature refers to “mean shifts” as breaks in the short-run conditional

mean and proceed with the usual regression based tests for break-points (see inter alia [Vogelsang 1997, 1998](#); [McKittrick and Vogelsang 2014](#); [Perron and Yabu 2009](#)).<sup>1</sup>

In this paper, we show, analytically and through simulations, that tests for conditional mean breaks are severely oversized when the functional form is misspecified, leading to detection of spurious breaks. Their unconditional counterparts are not plagued by the same issues and we propose using them instead, or in conjunction with, the conditional mean break tests.

The sensitivity of the conditional mean break tests to functional form misspecifications has been documented earlier. [Chong \(2003\)](#) focused on cases with iid, conditionally homoskedastic errors. [Bai et al. \(2008\)](#) focused on models with measurement error and proposed a new break-point test that corrects for measurement error. Another strand of literature focuses on trend breaks rather than mean breaks. It shows that dynamic misspecification of the conditional mean may result in severe size distortions and nonmonotonic power (see inter alia [Vogelsang 1997, 1998, 1999](#); [Kejriwal 2009](#); [McKittrick and Vogelsang 2014](#); [Perron and Yabu 2009](#)). The theoretical and simulation results in these papers are supported by the empirical studies of [Altansukh \(2013\)](#) and [Bataa et al. \(2013\)](#), who exposed the undesirable effects of misspecifying the conditional mean seasonalities, outliers, dynamics and heteroskedasticity in practice.

In this paper, we analyze the effects of static and dynamic misspecification on conditional mean break tests. We focus on the sup Wald test of [Andrews \(1993\)](#) because it is widely used in applied work. We prove that its asymptotic distribution is nonstandard and highly data-dependent when the number of lags is underspecified. Our analysis focuses on stationary weakly dependent and heteroskedastic processes, generalizing the iid homoskedastic results derived in [Chong \(2003\)](#).<sup>2</sup>

Most of the literature proposed correcting for dynamic misspecification by better lag selection procedures in possibly non-stationary series ([Perron and Yabu 2009](#); [Vogelsang and Perron 1998](#)) or by fixed bandwidth asymptotics ([Cho and Vogelsang 2014](#); [Sayginsoy and Vogelsang 2011](#); [Vogelsang 1998](#)). Moreover, [Perron \(1989\)](#), [Diebold and Inoue \(2001\)](#) and [Smallwood \(2016\)](#), among others, established that both unit root tests and long memory tests are size distorted in the presence of structural breaks. In our paper, we focus exclusively on testing for structural breaks for otherwise stationary processes, and using fixed bandwidth asymptotics is beyond the scope of our paper.

Concerning lag selection, AIC and BIC are routinely used for selecting the number of lags in dynamic models. In the presence of breaks, [Kurozumi and Tuvaandorj \(2011\)](#) proposed AIC- and BIC-type information criteria to jointly estimate the number of lags and breaks in a dynamic model. In our simulations, we show that, in a model with no breaks, AIC and BIC underselect the true number of lags even for large samples, especially when the true model has higher order lags with smaller coefficients than the first two lags, which is typically the case in empirical applications. Moreover, even if a data generating process has higher order lags with large coefficients, where our simulations show that the BIC performs well in large samples, it is unclear how the method in [Kurozumi and Tuvaandorj \(2011\)](#) would fare in selecting the true number of breaks, because it is derived under the assumption of shrinking (moderate) shifts, while some breaks typically encountered in macroeconomic series such as the Great Moderation break are typically treated as fixed (large) shifts. Therefore, we focus instead on how dynamic misspecification affects the break point inference when using conditional mean tests.

---

<sup>1</sup> There are a few notable exceptions. For example, [Rapach and Wohar \(2005\)](#) directly tested and found multiple unconditional mean shifts in international interest rates and inflation using the [Bai and Perron \(1998\)](#) tests. [Elliott and Müller \(2007\)](#) and [Eo and Morley \(2015\)](#) considered in their simulations a break in the unconditional mean (and variance) of a time series. Their focus was on constructing confidence sets with good coverage for small and large breaks, by inverting structural break tests. Recently, [Müller and Watson \(2017\)](#) proposed new methods for detecting low-frequency mean or trend changes. Our paper is different as it highlights the properties of the existing sup Wald test for a break in the unconditional and conditional mean and variance of a time series.

<sup>2</sup> Non-stationary processes with a trend break and unit root errors, whose first-differences exhibit mean shifts with stationary errors, have been analyzed in many papers. However, as [Vogelsang \(1998, 1999\)](#) showed, to recover monotonic power, testing the first-differenced series for a mean shift is better than testing the level series for a trend shift.

Additionally, we propose testing for breaks in the unconditional mean as an alternative to testing for breaks in the conditional mean. Breaks in the conditional mean are not equivalent, yet closely related, to breaks in the unconditional mean, as long as the conditional mean is correctly specified. [Aue and Horvath \(2012\)](#), among others, illustrated tests for both type of breaks in a recent comprehensive review on structural break tests. Our extensive simulation study shows that the unconditional mean break test, corrected for autocorrelation, yields close to correctly-sized tests, while, for most common static and dynamic misspecifications in the conditional mean, the conditional mean tests are severely oversized. Moreover, the power of both tests is very similar, especially as the sample size increases.<sup>3</sup> Similar results hold for the unconditional versus conditional break tests in variance. Therefore, the approach of testing first for a break in the unconditional mean and variance of the variable of interest is not only complementary to the regression approach, but is robust to alternative sources of misspecification.

There is a plethora of empirical evidence for breaks in the conditional mean and volatility of many US macroeconomic time series during the early and mid 1980s, associated with the Great Moderation (see, for example, [Bataa et al. 2013](#); [McConnell and Perez-Quiros 2000](#); [Sensier and van Dijk 2004](#); [Stock and Watson 2002](#)). Most studies employ dynamic regression models to detect such breaks. Focusing on unemployment, the industrial production and the real interest rate, we show that the unconditional mean and volatility tests mostly indicate no breaks in the unemployment or the industrial production growth. We further show that while the conditional mean tests occasionally detect breaks, the implied break-point estimates vary across the dynamic specifications employed. Therefore, it is plausible that the breaks found by the conditional tests are spurious because of size distortions, or that they do not result in long-run breaks in the mean of unemployment or industrial production growth. We also find that the evidence for a break in the variance of unemployment rates and real interest rates around the Great Moderation is not as strong as previously shown in [Stock and Watson \(2002\)](#).

This paper is organized as follows: Section 2 defines the unconditional break tests in mean and variance and derives their asymptotic properties in a unified framework. Section 3 defines the conditional structural break tests in mean and variance. It contains asymptotic results for the conditional break tests under correct specification and misspecification. Section 4 presents the simulation evidence comparing the size and power of the conditional and unconditional break tests. Section 5 illustrates the difference between these alternative structural break tests approaches with three empirical applications: the US civilian unemployment rate, the short-term real interest rate and the industrial production growth rate. A final section concludes. All the proofs are relegated to Appendix A.

## 2. Unconditional Mean and Variance Break Tests

In this section, we define the unconditional sup Wald test for an unknown break in the mean or variance of a dynamic univariate process.<sup>4</sup>

To our knowledge, a test for an unknown unconditional mean break, adjusted for autocorrelation, is rarely used in the literature.<sup>5</sup> Most papers test for a break in the conditional mean of a series; when they intend to test for an unconditional mean break, they routinely test for a trend break or an intercept break instead, after specifying a conditional mean (see, e.g., [Stock and Watson 2002](#)).

<sup>3</sup> The only case where our test has comparatively low power to the conditional mean test is in a correctly specified dynamic model with an intercept very close to zero. This case is further discussed in Section 3.

<sup>4</sup> Throughout the paper, we use the sup Wald test definition in [Andrews \(1993\)](#); alternative definitions of the sup Wald test are available, but they are not equivalent to the original sup Wald test in [Andrews \(1993\)](#) and should not be confused with it.

<sup>5</sup> Even though UM tests are not routinely used, they are a special case of the HAC-adjusted conditional break-point test in, e.g., [Bai and Perron \(1998\)](#), when the only regressor is an intercept. In addition, a CUSUM (cumulative sum) variant of this test for iid data is in [Pitarakis \(2004\)](#). As shown in Appendix A, proof of Theorem 1, for unconditional break tests, there is an explicit asymptotic relationship between the CUSUM test and the sup Wald test. However, as the Appendix shows, the conclusion of the two tests based on asymptotic critical values is in general different. Since there is strong evidence for the non-monotonic power of the CUSUM test (see, e.g., [Vogelsang 1999](#)), the paper focuses on the sup Wald test instead.

Such approaches have the disadvantage that they are highly dependent on the correct specification of the conditional mean. They also do not shed light on unconditional mean shifts, which may not be equivalent to conditional mean shifts. Therefore, in this paper, we propose using UM break tests complementarily to CM break tests, to uncover long-run mean shifts in the presence of potential static and dynamic misspecification.

We denote the unconditional mean by UM, and the unconditional variance by UV henceforth. In contrast to UM breaks, UV breaks are routinely tested in applications, for example to uncover the Great Moderation break. It is common to test for a break in the absolute value of the demeaned data, as a proxy for testing a variance break (see McConnell and Perez-Quiros 2000; Sensier and van Dijk 2004; Stock and Watson 2002). We call these tests UA (unconditional absolute deviation) break tests. One can also use the squared demeaned data to test for a variance break, as in Pitarakis (2004) and Qu and Perron (2007). We call these UV break tests, because they test directly for a variance break.<sup>6</sup>

Below, we state the null asymptotic distributions of UM, UA and UV break tests under fairly general assumptions on the data. These distributions are not dependent on regressor, functional form, or seasonality misspecifications, simply because a conditional mean is not specified. The only misspecification that affects the null asymptotic distribution of these tests are UM breaks for the UA and UV tests, or UV breaks for the UM test. Fortunately, this misspecification is easy to correct; we discuss this correction at the end of this section.

The true model takes the general form:

$$y_t = \mu_1 \mathbf{1}[t \leq T_{UM}] + \mu_2 \mathbf{1}[t > T_{UM}] + u_t, \quad (1)$$

where  $\mu_1, \mu_2$  are deterministic, the break  $T_{UM} = [T\lambda_{UM}]$  is an unknown, fixed fraction of the sample  $0 < \lambda_{UM} < 1$ , and  $u_t$  satisfies the assumption below, in which  $\mathbf{AVar} = \lim_{T \rightarrow \infty} \text{Var}$ .

### Assumption 1.

- (i)  $\mathbf{E}(u_t) = 0$  and  $\mathbf{AVar} \left( T^{-1/2} \sum_{t=1}^{[T\lambda]} u_t \right) = \lambda \mathbf{v}_u$ ;
- (ii) for some  $d > 4$ ,  $\sup_t \mathbf{E}|u_t|^d < \infty$  and  $\{u_t\}$  is  $\mathcal{L}_2$ -near epoch dependent of size  $c_m = O(m^{-1})$  on  $\{g_t\}$ , i.e.,  $\|u_t - \mathbf{E}[u_t | \mathcal{G}_{t-m}^{t+m}]\|_2 \leq c_m$  with  $c_m = O(m^{-1})$  where  $\mathcal{G}_{t-m}^{t+m} = \sigma(g_{t-m}, \dots, g_{t+m})$ , and  $\{g_t\}$  is either  $\phi$ -mixing of size  $m^{-d/(2(d-1))}$  or  $\alpha$ -mixing of size  $m^{-d/(d-2)}$ .<sup>7</sup>

With these assumptions,  $y_t$  can exhibit very general dependence—ARMA, GARCH, and nonlinear dependence—but it cannot have unit roots or UV breaks.<sup>8</sup>

For a UM break, the null and alternative hypotheses are:

$$H_0^{UM} : \mu_1 = \mu_2 \quad \text{vs.} \quad H_A^{UM} : \mu_1 \neq \mu_2.$$

For a UA break, let  $a_t = \mathbf{E}|y_t - \bar{y}|$ , and test:

$$H_0^{UA} : a_t = a_u \quad \text{vs.} \quad H_A^{UA} : a_t = a_{u,1} \mathbf{1}[t \leq T_{UA}] + a_{u,2} \mathbf{1}[t > T_{UA}], a_{u,1} \neq a_{u,2}.$$

For a UV break, let  $v_{ut} = \mathbf{E}(y_t - \bar{y})^2$ , and test:

$$H_0^{UV} : v_t = v_u \quad \text{vs.} \quad H_A^{UV} : v_t = v_{u,1} \mathbf{1}[t \leq T_{UV}] + v_{u,2} \mathbf{1}[t > T_{UV}], v_{u,1} \neq v_{u,2}.$$

<sup>6</sup> Note that a break in the expected absolute value of a demeaned series is not the same as a variance break only under certain conditions.

<sup>7</sup> Here,  $\|\cdot\|_2 = (\mathbf{E}\|\cdot\|^2)^{1/2}$  stands for the  $\mathcal{L}_2$ -norm, and  $|\cdot|$  stands for the Euclidean norm.

<sup>8</sup> For a proof that the most common GARCH model, GARCH(1,1), is near-epoch dependent and therefore fits our assumptions, see (Hansen 1991).

Under the alternative hypotheses, all breaks  $T_{UM} = [T\lambda_{UM}]$ ,  $T_{UV} = [T\lambda_{UV}]$ ,  $T_{UA} = [T\lambda_{UA}]$  are assumed to occur at unknown fixed fractions  $0 < \lambda_{UM}, \lambda_{UV}, \lambda_{UA} < 1$  of the sample.

The UM test is defined below. It is a special case of the Andrews (1993) sup Wald test when the only regressor is an intercept, and when the variance is estimated under the null of no break. Therefore, it is not new; nevertheless, to our knowledge, it is rarely used in the empirical literature in the form defined below:

$$UM_T^* = \sup_{\lambda \in [\epsilon, 1-\epsilon]} UM_T(\lambda), \quad UM_T(\lambda) = T(\bar{y}_{1\lambda} - \bar{y}_{2\lambda})^2 / \hat{\mathbf{v}}_{u\lambda}, \quad (2)$$

$\epsilon > 0$  is a small cut-off, typically  $\epsilon = 0.15$  in applications,  $\bar{y}_{i\lambda} = T_{i\lambda}^{-1} \sum_{t=1}^{T_{i\lambda}} y_t$  for  $i = 1, 2$ ,  $T_{1\lambda} = [T\lambda]$ ,  $T_{2\lambda} = T - [T\lambda]$ ,  $\sum_{1\lambda} = \sum_{t=1}^{T_{1\lambda}}$ ,  $\sum_{2\lambda} = \sum_{t=T_{1\lambda}+1}^T$  and  $\hat{\mathbf{v}}_{u\lambda}$  is a HAC consistent estimator of  $\mathbf{v}_{u\lambda} = \mathbf{AVar}(\sqrt{T}(\bar{y}_{1\lambda} - \bar{y}_{2\lambda})) = \mathbf{v}_u / [\lambda(1-\lambda)]$  under  $H_0^{UM}$ .

For the HAC estimator  $\hat{\mathbf{v}}_{u\lambda}$ , it is crucial to calculate it over the full sample, i.e., under the null  $H_0^{UM}$ . If we use sub-sample estimators in its computation—i.e., we estimate the variances  $T^{1/2}(\bar{y}_{1\lambda} - \mu)$  and  $T^{1/2}(\bar{y}_{2\lambda} - \mu)$  separately—we need a separate bandwidth for each. Since the bandwidth estimation is only accurate in large samples, for those  $\lambda$ 's that are close to  $\epsilon$  and  $1 - \epsilon$ , such an estimation would be highly inaccurate, resulting in high size distortions.<sup>9</sup>

Thus, we define:

$$\hat{\mathbf{v}}_{u\lambda} = \frac{1}{\lambda(1-\lambda)} \sum_{j=-b_T+1}^{b_T-1} k\left(\frac{j}{b_T}\right) \hat{\gamma}_j, \quad \hat{\gamma}_j = \begin{cases} \frac{1}{T-1} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y}), & j \geq 0 \\ \hat{\gamma}_{-j} & j < 0 \end{cases}$$

where  $\bar{y} = T^{-1} \sum_{t=1}^T y_t$ ,  $k(x) = (1 - |x|) \mathbf{1}[|x| \leq 1]$  is throughout the paper the Bartlett kernel, with the optimal data-dependent bandwidth in Newey and West (1994).<sup>10</sup> Specifically, we let  $b_T = \min[T, \hat{\eta} T^{\frac{1}{3}}]$ , where  $\hat{\eta} = 1.1447(\hat{f}^{(1)} / \hat{f}^{(0)})^{\frac{2}{3}}$ , and  $\hat{f}^{(1)} = 2 \sum_{j=1}^{\tau} j \hat{\gamma}_j$ ,  $\hat{f}^{(0)} = \hat{\gamma}_0 + 2 \sum_{j=1}^{\tau} \hat{\gamma}_j$ , with  $\tau = [(T/100)^{2/9}]$ . The lag truncation parameter  $\tau$  governs how many auto-covariances should be used in forming the nonparametric estimates  $\hat{f}^{(1)}$  and  $\hat{f}^{(0)}$ , which estimate the spectral density at frequency one and zero.<sup>11</sup> Therefore,  $\hat{f}^{(1)}$ ,  $\hat{f}^{(0)}$ , and  $\hat{\eta}$  are computed over the full sample.

The UA and UV tests are denoted by  $UA_T^*$  and  $UV_T^*$ . They are computed as  $UM_T^*$ , but with  $y_t$  replaced by  $\hat{a}_t = |y_t - \bar{y}|$  for UA,  $\hat{v}_t = (y_t - \bar{y})^2$  for UV, and  $\hat{\mathbf{v}}_{u\lambda}$  replaced by the HAC consistent estimator of the asymptotic variance of  $\hat{a}_t$  or  $\hat{v}_t$ .

Define the distribution:

$$\mathcal{G}_p = \sup_{\lambda \in [\epsilon, 1-\epsilon]} \frac{[B_p(\lambda) - \lambda B_p(1)]' [B_p(\lambda) - \lambda B_p(1)]}{\lambda(1-\lambda)},$$

where  $B_p(\cdot)$  is a  $p \times 1$  vector of independent standard Brownian motions, for some  $p \geq 1$ . As Theorem 1 shows,  $\mathcal{G}_1$  is the null asymptotic distributions of the UM, UA and UV break tests. Although the distribution of various break point tests under different (more restrictive) assumptions is available, an explicit proof for the UM, UV and UA tests under Assumption 1 is not available in a unified setting to our knowledge, and we provide it in Appendix A.

For the UA, respectively, UV tests, we need the following additional assumptions.

<sup>9</sup> Simulation evidence for this statement is available from the authors upon request.

<sup>10</sup> Additional simulations not reported here show that the fixed optimal bandwidth proposed in Andrews (1991) leads to worse performance of the UM break test.

<sup>11</sup> The weights mentioned in Newey and West (1994) are set equal to one as usual for scalar cases.

**Assumption 2.**

- (i)  $\mathbf{AVar} \left( T^{-1/2} \sum_{t=1}^{[T\lambda]} |y_t - \bar{y}| \right) = \lambda \mathbf{v}_a$ , for some  $\mathbf{v}_a > 0$ ; and
- (ii)  $\mathbf{AVar} \left( T^{-1/2} \sum_{t=1}^{[T\lambda]} (y_t - \bar{y})^2 \right) = \lambda \mathbf{v}_s$ , for some  $\mathbf{v}_s > 0$ .

**Theorem 1.** Let the model be as in (1), and let Assumption 1 hold. Then: (i) under  $H_0^{UM}$ ,  $UM_T^* \Rightarrow \mathcal{G}_1$ ; (ii) under  $H_0^{UM}$ ,  $H_0^{UA}$  and Assumption 2(i),  $UA_T^* \Rightarrow \mathcal{G}_1$ ; and (iii) under  $H_0^{UM}$ ,  $H_0^{UV}$  and Assumption 2(ii),  $UV_T^* \Rightarrow \mathcal{G}_1$ .<sup>12</sup>

Note that the distributions are non-standard, but critical values are available in, e.g., Andrews (1993) and Bai and Perron (1998).

Theorem 1 assumes no UV breaks for UM tests, via imposing Assumption 1(i), and no UM breaks for UV/UA tests, via imposing  $H_0^{UM}$ . If there is a UM break ( $H_A^{UM}$  holds instead of  $H_0^{UM}$ ), as shown in Pitarakis (2004), we can obtain  $\hat{v}_t$  and  $\hat{a}_t$  via subsample demeaning, and Theorem 1(ii)–(iii) will hold. That is, we let  $\hat{v}_t = (y_t - \bar{y}_t)^2$ ,  $\hat{a}_t = |y_t - \bar{y}_t|$ , and  $\bar{y}_t = \hat{T}_{UM}^{-1} \sum_{t=1}^{\hat{T}_{UM}} y_t \mathbf{1}[t \leq \hat{T}_{UM}] + (T - \hat{T}_{UM})^{-1} \sum_{t=\hat{T}_{UM}+1}^T y_t \mathbf{1}[t > \hat{T}_{UM}]$ , where  $\hat{T}_{UM}$  is the Bai and Perron (1998) OLS break-point estimator of  $T_{UM}$  in (1). If there is a UV break, the asymptotic distribution of the UM test is affected, but one can employ the fixed-regressor bootstrap in Hansen (2000) to correct for this. The correction for the UV tests via sub-sample demeaning is necessary and employed in our empirical analysis in Section 5.

**3. Conditional Mean and Variance Break Tests****3.1. Correct Specification**

Unlike unconditional break tests, regression-based break tests are pervasive in empirical work, despite their sensitivity to misspecification (this sensitivity is discussed in Section 3.2). The most common regression specification is of the linear form:

$$y_t = x_t' \theta_1 \mathbf{1}[t \leq T_{CM}] + x_t' \theta_2 \mathbf{1}[t > T_{CM}] + \epsilon_t, \quad (3)$$

where  $T_{CM} = [T\lambda_{CM}]$ ,  $0 < \lambda_{CM} < 1$ ,  $x_t$  is a  $p \times 1$  vector of regressors that includes an intercept and possibly lagged dependent variables, and  $\epsilon_t$  are scalar errors.

We denote by CM, CA and CV the conditional mean, conditional absolute deviation and the conditional variance, where the word “conditional” simply refers to specifying the conditional mean in (3). To derive the asymptotic distribution of the CM, CA and CV break tests, we need additional assumptions on the joint dependence of regressors and errors.

**Assumption 3.**

- (i)  $\mathbf{E}(x_t \epsilon_t) = 0$ ,  $\mathbf{AVar}(T^{-1/2} \sum_{t=1}^{[T\lambda]} x_t \epsilon_t) = \lambda \mathbf{V}$  and  $T^{-1} \sum_{t=1}^{[T\lambda]} x_t x_t' \xrightarrow{P} \lambda \mathbf{Q}$ , two positive definite (pd) matrices of constants; and
- (ii) for some  $d > 4$ ,  $\sup_t \|x_t \epsilon_t\|_d < \infty$  and  $\{x_t \epsilon_t\}$  is  $\mathcal{L}_2$ -near epoch dependent of size  $c_m = O(m^{-1})$  on  $\{h_t\}$ , and  $\{h_t\}$  is either  $\phi$ -mixing of size  $m^{-d/(2(d-1))}$  or  $\alpha$ -mixing of size  $m^{-d/(d-2)}$ .

Note that we need the assumption  $T^{-1} \sum_{t=1}^{[T\lambda]} x_t x_t' \xrightarrow{P} \lambda \mathbf{Q}$  for two reasons. First, as explained in Hansen (2000), if this assumption does not hold, then the asymptotic distribution of the test statistic is not pivotal. Second, this assumption does not allow for unit roots in  $x_t$ , but it allows for lagged dependent variables in  $x_t$ .

<sup>12</sup> Here, “ $\Rightarrow$ ” indicates weak convergence in the Skorohod metric.



The null and alternative hypotheses of the conditional tests are:

$$\begin{aligned} H_0^{CM} : \theta_1 = \theta_2 & \quad vs. \quad H_A^{CM} : \theta_1 \neq \theta_2, \\ H_0^{CA} : a_{\epsilon t} = a_{\epsilon} & \quad vs. \quad H_A^{CA} : a_{\epsilon t} = a_{\epsilon 1} \mathbf{1}[t \leq T_{CA}] + a_{\epsilon 2} \mathbf{1}[t > T_{CA}], \quad a_{\epsilon 1} \neq a_{\epsilon 2}, \\ H_0^{CV} : v_{\epsilon t} = v_{\epsilon} & \quad vs. \quad H_A^{CV} : v_{\epsilon t} = v_{\epsilon 1} \mathbf{1}[t \leq T_{CV}] + v_{\epsilon 2} \mathbf{1}[t > T_{CV}], \quad v_{\epsilon 1} \neq v_{\epsilon 2}, \end{aligned}$$

where  $a_{\epsilon t} = \mathbf{E}[\epsilon_t]$ ,  $v_{\epsilon t} = \mathbf{Var}(\epsilon_t)$ ,  $T_{CA} = [T\lambda_{CA}]$ ,  $T_{CV} = [T\lambda_{CV}]$ ,  $0 < \lambda_{CA}, \lambda_{CV} < 1$ .

The corresponding sup Wald test for a CM break is defined in, e.g., [Andrews \(1993\)](#):

$$CM_T^* = \sup_{\lambda \in [\epsilon, 1-\epsilon]} CM_T(\lambda), \quad CM_T(\lambda) = T(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda})' \hat{\mathbf{V}}_{\lambda}^{-1} (\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}),$$

where  $\hat{\theta}_{1\lambda}, \hat{\theta}_{2\lambda}$  are the OLS estimators of  $\theta$  in Equation (3) in subsamples  $\{1, \dots, T_{1\lambda}\}$  and  $\{T_{1\lambda} + 1, \dots, T\}$ , and  $\hat{\mathbf{V}}_{\lambda}$  is a consistent estimator of  $\mathbf{AVar}(T^{1/2}(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}))$  under  $H_0^{CM}$ .

For the conditional test, the asymptotic variance  $\mathbf{AVar}(T^{1/2}(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}))$  is routinely estimated over subsamples, i.e., separately for  $T^{1/2}(\hat{\theta}_{1\lambda} - \theta^0)$  and  $T^{1/2}(\hat{\theta}_{2\lambda} - \theta^0)$ , or under the alternative. If a HAC estimator under the alternative would be used, the same problems would arise as for the unconditional test: there would be size distortions due to inaccurate bandwidth estimation for  $\lambda$  close to the beginning or the end of the sample. However, in most studies, the conditional mean specification in (3) is assumed to be correct, in which case all lags of the dependent variable are included as regressors, and correcting for autocorrelation is no longer necessary. If this is the case, the variance can be estimated under the alternative without further size distortions. Thus, as in most empirical studies, we use variance estimators that are not autocorrelation-robust in all the simulations except those where the model is static. In a static model, the researcher might suspect that the errors are autocorrelated, and a HAC estimator is justified.

For the theory section, we consider two potential estimators for  $\mathbf{AVar}(T^{1/2}(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}))$ , under homoskedasticity or heteroskedasticity. The one under homoskedasticity is:

$$\begin{aligned} \hat{\mathbf{V}}_{\lambda} &= \hat{\mathbf{V}}_{1\lambda} + \hat{\mathbf{V}}_{2\lambda}, \quad \hat{\mathbf{V}}_{i\lambda} = \hat{\mathbf{v}}_{\epsilon, i\lambda} (T^{-1} \sum_{i\lambda} x_t x_t')^{-1}, \quad \hat{\mathbf{v}}_{\epsilon, i\lambda} = T_{i\lambda}^{-1} \sum_{i\lambda} \hat{\epsilon}_t^2, \quad (i = 1, 2), \\ \hat{\epsilon}_t &= y_t - x_t' \hat{\theta}_{1\lambda} \mathbf{1}\{t \leq T_{1\lambda}\} - x_t' \hat{\theta}_{2\lambda} \mathbf{1}\{t > T_{1\lambda}\}. \end{aligned} \quad (4)$$

The one under heteroskedasticity is:

$$\hat{\mathbf{V}}_{\lambda} = \hat{\mathbf{V}}_{1\lambda} + \hat{\mathbf{V}}_{2\lambda}, \quad \hat{\mathbf{V}}_{i\lambda} = (T^{-1} \sum_{i\lambda} x_t x_t')^{-1} (T^{-1} \sum_{i\lambda} \hat{\epsilon}_t^2 x_t x_t') (T^{-1} \sum_{i\lambda} x_t x_t')^{-1}, \quad (i = 1, 2). \quad (5)$$

We define the CA and CV tests as the UA and the UV tests, but with  $\hat{a}_t, \hat{v}_t$  replaced by  $\hat{a}_{\epsilon t} = |\hat{\epsilon}_t|, \hat{v}_{\epsilon t} = \hat{\epsilon}_t^2$ , and with  $\hat{\epsilon}_t = y_t - x_t' \hat{\theta}$  the residuals from estimating (3) under the null  $H_0^{CM}$ . We emphasize that the name “conditional” refers exclusively to pre-specifying the conditional mean in (3), and not the conditional variance of  $y_t$  or  $\epsilon_t$ . Therefore, the tests in this paper should not be confused with the conditional variance tests proposed by, e.g., [Andreou and Ghysels \(2002\)](#), who wrote down a model for the conditional variance of  $\epsilon_t$ .

Theorem 2 states the asymptotic distribution of the CM, CA and CV break tests.

**Theorem 2.** Let the model be as in (3), and let Assumption 3 hold. Then: (i) under  $H_0^{CM}$ ,  $CM_T^* \Rightarrow \mathcal{G}_p$ ; (ii) under  $H_0^{CM}$  and  $H_0^{CA}$ ,  $CA_T^* \Rightarrow \mathcal{G}_1$ ; and (iii) under  $H_0^{CM}$  and  $H_0^{CV}$ ,  $CV_T^* \Rightarrow \mathcal{G}_1$ .

Note that the distributions are similar to the unconditional break tests, but there are more degrees of freedom used up by the conditional break tests.

As for the UA and UV tests, the asymptotic distributions of the CA and CV tests are not valid if there is a CM break; in that case, as [Pitarakis \(2004\)](#) shows, the CM break at  $T_{CM}$  can be pre-estimated by  $\hat{T}_{CM}$  along with the slope parameters  $\hat{\theta}_1, \hat{\theta}_2$  before and after the break, via the methods in [Bai and](#)

Perron (1998). Then, we can redefine  $\hat{\epsilon}_t = y_t - x_t' \hat{\theta}_1 \mathbf{1}[t \leq \hat{T}_{CM}] - x_t' \hat{\theta}_2 \mathbf{1}[t > \hat{T}_{CM}]$  in the computation of  $\hat{a}_{\epsilon t}, \hat{v}_{\epsilon t}$ , obtaining the same asymptotic distributions as stated in Theorem 2. Under the alternative  $H_A^{CV}$ , the asymptotic null distribution of the CM test is not valid, but, as for the unconditional break tests, it can be bootstrapped via the fixed regressor bootstrap in Hansen (2000).

Note that UM and CM tests are in general testing equivalent null hypotheses under correct specification, with some exceptions. Consider the following general AR( $p$ ) model with additional covariates:

$$y_t = \left( \alpha_1 + \sum_{i=1}^p y_{t-i} \beta_{1i} + x_t' \gamma_1 \right) \mathbf{1}[t \leq T_{CM}] + \left( \alpha_2 + \sum_{i=1}^p y_{t-i} \beta_{2i} + x_t' \gamma_2 \right) \mathbf{1}[t > T_{CM}] + \epsilon_t$$

The CM statistic tests the null hypothesis  $H_0^{CM} : \theta_1 = \theta_2$ , where  $\theta_j = (\alpha_j, \beta_{j1}, \dots, \beta_{jp}, \gamma_j')'$  for  $j = 1, 2$ . The UM statistic tests the null hypothesis  $H_0^{UM} : \mu_1 = \mu_2$ , where  $\mu_j = (\alpha_j + [E(x_t)]' \theta_j) / (1 - \sum_{i=1}^p \beta_{ji})$ . In principle, a change in any of the elements of  $\theta_1$  results in a change in  $\mu_1$ . In addition, a small change in  $\beta_{1i}$ , a number usually between zero and one, typically results in a larger change in  $\mu_1$ , which means that the UM test will tend to have larger power than the CM test against these changes. However, it is useful to note that, if  $\alpha_j + [E(x_t)]' \theta_j = 0$  (for example,  $\alpha_j = \theta_j = 0$  for  $j = 1, 2$ ), then the unconditional mean is zero and the UM test will have no power against changes in the other parameters  $\beta_{ji}$ . Therefore, the UM test should be used only for series which do not have a zero unconditional mean over the whole sample, a condition that can be easily verified for any dataset before proceeding with the UM test.

### 3.2. Dynamic Misspecification

Unlike the unconditional break tests, all the conditional break tests are highly dependent on the correct specification of the functional form, including seasonality and dynamics. Bataa et al. (2013) and Altansukh (2013) empirically showed the effects of misspecifying the conditional mean seasonalities, outliers, dynamics and heteroskedasticity on the conditional break tests. Chong (2003) and Bai et al. (2008) theoretically showed that misspecification of the functional form leads to different null asymptotic distributions for the CM break tests. They focus on iid errors and static misspecifications, although some of their theoretical results apply to dynamic misspecification as well. The impact of dynamic misspecification of Equation (3) on conditional break tests was analyzed by Vogelsang and Perron (1998), Vogelsang (1999), Perron and Yabu (2009), inter alia. However, all these studies correct for omitted autocorrelation in the errors by either better selection of lags in the regression equation, or directly correcting the error variance via HAC estimators. The first correction is successful if the method used indeed selects the number of lags correctly. The second correction is not always valid if the regression model is already dynamic, as omitted autocorrelation in the errors often violates the exogeneity Assumption 3(i), so a HAC variance estimator does not fix the dynamic misspecification problems.

To our knowledge, the effect of misspecifying the regressors or number of lags on CM break tests has not been studied before under general dependence and conditionally heteroskedastic data as allowed for in Assumption 4. The result in Theorem 3 is a generalization of the result in Chong (2003). The assumptions in Chong (2003) allow for lagged dependent variables, however the authors assumed that the error term is iid, and constructed the sup Wald test imposing this assumption (i.e., imposing homoskedasticity). Our results are more general than Chong (2003) in two ways. First, we allow the error term in the true model to be a near epoch dependent process, thus also allowing for conditionally heteroskedastic series, which is empirically important for analyzing both macroeconomic and financial time series. Second, we construct the sup Wald test such that it corrects for heteroskedasticity. We prove below that the asymptotic distribution of the CM break test is data-dependent and different than that stated in Theorem 3. Therefore, in the presence of dynamic misspecification, the critical values of the



CM tests will be incorrect<sup>13</sup>, while the critical values for the UM break test are correct. Thus, the UM break test provides a valuable tool for assessing stability of the process  $y_t$  in the presence of dynamic misspecification.

To formalize the results under dynamic misspecification, let  $x_t = \text{vec}(x_{t(1)}, x_{t(2)})$  and  $\theta = \text{vec}(\theta_{(1)}, \theta_{(2)})$ , where  $x_{t(1)}, \theta_{(1)}$  are  $p_1 \times 1$ ,  $x_{t(2)}, \theta_{(2)}$  is  $p_2 \times 1$ , and  $p_1 + p_2 = p$ .<sup>14</sup> The true model is (3), but we mistakenly regress  $y_t$  only on  $x_{t(1)}$  (which we assume includes the intercept). Thus, we underspecify the number of regressors; in particular, we are interested in the effects of underspecifying the number of lags.<sup>15</sup> Partition  $Q = \begin{bmatrix} Q_{(1)} & Q_{(12)} \\ Q'_{(12)} & Q_{(2)} \end{bmatrix}$ , with  $Q_{(1)}, Q_{(2)}$  of dimensions  $p_1 \times p_1$  and  $p_2 \times p_2$ , respectively.

**Assumption 4.** Let  $k_t = x_{t(1)}\epsilon_t$ ,  $L_t = x_{t(1)}x'_{t(2)} - Q_{(12)}$ ,  $M_t = x_{t(1)}x'_{t(1)} - Q_{(1)}$ , and  $w_t = \text{vech}(k_t, L_t, M_t)$ , where  $\text{vech}(A, B)$  selects, in order, the unique elements and the first occurrence of the repeating elements in  $\text{vec}(A, B)$ .

- (i)  $E(w_t) = 0$ ,  $\text{AVar}(T^{-1/2} \sum_{t=1}^{[T\lambda]} w_t) = \lambda H$ , a pd matrix of constants;
- (ii) for some  $d > 4$ ,  $\sup_t \|w_t\|_d < \infty$  and  $\{w_t\}$  is  $\mathcal{L}_2$ -near epoch dependent of size  $d_m = O(m^{-1})$  on either an  $\phi$ -mixing process of size  $m^{-d/(2(d-1))}$  or an  $\alpha$ -mixing process of size  $m^{-d/(d-2)}$ ;
- (iii)  $Q_{(12)} \neq O_{p_1 \times p_2}$ , where  $O_{p_1 \times p_2}$  is the  $p_1 \times p_2$  null matrix; and
- (iv)  $T^{-1} \sum_{1 \leq \lambda} w_t w'_t \xrightarrow{P} \lambda \Omega$ , a pd matrix of constants.

Assumption 4(iii) states that the omitted regressors are correlated with the included regressors, as is the case when the number of lags is underspecified. The rest of the statements in Assumption 4 are standard. Let  $r = p_1(1 + p_2 + (p_1 + 1)/2)$ , the dimension of  $w_t$ . Then, under Assumption 4, the functional central limit theorem in (Wooldridge and White 1988, Theorem 2.11) can be applied to yield  $T^{-1/2} \sum_{1 \leq \lambda} w_t \Rightarrow H^{1/2} B_r(\lambda)$ . To state the asymptotic distribution of the CM break test under misspecification, let  $B_r(\lambda) = B_r(\lambda) - \lambda B_r(1)$ ,  $s = p_1(1 + p_1 + p_2)$ , and  $B_s^*(\lambda) = B_s^*(\lambda) - \lambda B_s^*(1)$ , where  $B_s^*(\lambda)$  is constructed from  $B_r(\lambda)$  by repeating its elements exactly in the positions where  $w_t^* = \text{vec}(k_t, L_t, M_t)$  repeats the elements of  $w_t = \text{vech}(k_t, L_t, M_t)$ . Similarly, let  $H^{*1/2}$  and  $\Omega^*$  be positive semidefinite matrices constructed from  $H^{1/2}$  and  $\Omega$ , which were defined in Assumption 4, so that  $\text{AVar}(T^{-1/2} \sum_{t=1}^{[T\lambda]} w_t^*) = \lambda H^{*1/2} H^{*1/2'}$ , and  $T^{-1} \sum_{1 \leq \lambda} w_t^* w_t^{*'} \xrightarrow{P} \lambda \Omega^*$ . With this notation, the asymptotic distribution of the CM test is stated in Theorem 3.

**Theorem 3.** Let Assumptions 3 and 4 and  $H_0^{CM}$  hold,  $\delta = Q_{(1)}^{-1} Q_{(12)} \theta_{(2)}$ ,  $\xi = \text{vec}(1, \theta_{(2)}, -\delta)$  and  $A = H^{*1/2'} [\xi \otimes Q_{(1)}^{-1}] \left\{ [\xi' \otimes Q_{(1)}^{-1}] \Omega^* [\xi \otimes Q_{(1)}^{-1}] \right\}^{-1} [\xi' \otimes Q_{(1)}^{-1}] H^{*1/2}$ .

- (i) If  $CM_T^*$  is constructed under heteroskedasticity,

$$CM_T^* \Rightarrow \sup_{\lambda} \frac{B_s^{*'}(\lambda) A B_s^*(\lambda)}{\lambda(1 - \lambda)}.$$

- (ii) Let  $v = \sigma_e^2 + \theta'_{(2)} [Q_{(2)} - Q'_{(12)} Q_{(1)}^{-1} Q_{(12)}] \theta_{(2)}$ . If  $CM_T^*$  is constructed under homoskedasticity, then the result in (i) holds, with  $A = v^{-1} H^{*1/2'} [(\xi \xi') \otimes Q_{(1)}^{-1}] H^{*1/2}$ .

<sup>13</sup> The simulation section shows that the CM tests are severely oversized with dynamic misspecification.

<sup>14</sup> We extend the  $\text{vec}(A, B)$  notation to denote stacking in a vector all columns of  $A$ , then all columns of  $B$ , one by one, in order, even when  $A, B$  do not have the same number of rows, and we let  $\text{vec}'(A, B) = [\text{vec}(A, B)]'$ .

<sup>15</sup> Overspecifying the number of lags or regressors is not a problem, as the coefficients on the additional regressors or lags will converge to zero.

*Comment 1.* The theorem above shows that the asymptotic distribution of the CM test is nonstandard and highly dependent on the data parameters and the unknown number of lags omitted. Our theorem is a generalization of Theorem 3 in Chong (2003), who proved the same result, but under iid and conditionally homoskedastic errors, with  $CM_T^*$  constructed only under homoskedasticity.

*Comment 2.* As we expect, in the presence of no misspecification ( $\theta_{(2)} = 0$ ), we can show that Theorem 3 reduces to Theorem 2. To see this, note that, when  $\theta_{(2)} = 0$ ,  $\zeta' = (1, 0, 0)$ . Therefore,  $[Q_{(1)}^{-1} \otimes \zeta'] \mathbf{H}^{*1/2} \mathcal{B}_s^*(\lambda) = Q_{(1)}^{-1} \Omega_{kk}^{*1/2} \mathcal{B}_{p_1}(\lambda)$ , where  $\Omega_{kk}^*$  was defined in the Appendix as  $\Omega_{kk}^* \equiv \mathbf{AVar}(T^{-1/2} \sum_{t=1}^T x_{t(1)} \epsilon_t)$ , and  $\mathcal{B}_{p_1}(\lambda)$  are the first  $p_1$  elements of  $\mathcal{B}_s(\lambda)$ . Moreover, note that  $\Omega_{kk}^* = \mathbf{AVar}(T^{-1/2} \sum_{t=1}^T x_t \epsilon_t)$  because in the true model,  $p_1 = p$ , so  $x_t = x_{t(1)}$ . Finally, note that  $\left\{ [\zeta' \otimes Q_{(1)}^{-1}] \Omega^* [\zeta \otimes Q_{(1)}^{-1}] \right\}^{-1} = \left\{ Q_{(1)}^{-1} \Omega_{kk}^* Q_{(1)}^{-1} \right\}^{-1} = Q_{(1)} \Omega_{kk}^{*-1} Q_{(1)}$ . Therefore, Theorem 3(i) reduces to

$$CM_T^* \Rightarrow \sup_{\lambda} \frac{\mathcal{B}_{p_1}'(\lambda) \Omega_{kk}^{*1/2'} Q_{(1)}^{-1} Q_{(1)} \Omega_{kk}^{*-1} Q_{(1)} Q_{(1)}^{-1} \Omega_{kk}^{*1/2} \mathcal{B}_{p_1}(\lambda)}{\lambda(1-\lambda)} = \sup_{\lambda} \frac{\mathcal{B}_{p_1}'(\lambda) \mathcal{B}_{p_1}(\lambda)}{\lambda(1-\lambda)} = \mathcal{G}_p(\lambda),$$

exactly the distribution in Theorem 2. By similar arguments, Theorem 3(ii) also reduces to Theorem 2, as it should under conditional homoskedastic errors.

*Comment 3.* The distributions in Theorem 3 are also the same as in Theorem 2 when the model is static, except that they have fewer degrees of freedom ( $p_1$  instead of  $p$ ). The reason for this is that the functional form of the model is still linear when misspecified, so it will be correctly specified for a modified version of the initial model.<sup>16</sup> To see the intuition for this result, note that, under the null hypothesis, the true model is  $y_t = x_{t(1)}\theta_{(1)} + x_{t(2)}\theta_{(2)} + \epsilon_t$ . The parameters will be consistently estimated by OLS if we regress  $y_t$  on  $x_t = (x_{t(1)}', x_{t(2)}')'$ . However, the initial model can be rewritten as  $y_t = x_{t(1)}\theta_{(1)}^* + u_t$ , where  $\theta_{(1)}^* = \theta_{(1)} + Q_{(1)}^{-1} Q_{(12)} \theta_{(2)} = \theta_{(1)} + \delta$  and  $u_t = x_{t(1)}'[\theta_{(1)} - \theta_{(1)}^*] + x_{t(2)}\theta_{(2)} + \epsilon_t = \epsilon_t + x_{t(2)}'\theta_{(2)} - x_{t(1)}'\delta$ . This new model satisfies  $E(x_{t(1)} u_t) = 0$  by construction and therefore it is also correctly specified in the sense that it will consistently estimate the new parameter  $\theta_{(1)}^*$  when regressing  $y_t$  on  $x_{t(1)}$  via OLS. This would suggest that the test statistic  $CM_T^*$  for breaks should have a similar distribution as for the correct specification but using only  $p_1$  degrees of freedom pertaining to  $x_{t(1)}$ . However, this intuition is only true for static models (by static models, we mean models where the long-run variance  $\mathbf{H}^* = \mathbf{AVar}(T^{-1/2} \sum_{t=1}^T w_t)$  is equal to the short-run variance  $\Omega^* = T^{-1} \sum_{t=1}^T w_t w_t'$ . In such models,  $A = \mathbf{H}^{*1/2'} [\zeta \otimes Q_{(1)}^{-1}] \left\{ [\zeta' \otimes Q_{(1)}^{-1}] \mathbf{H}^* [\zeta \otimes Q_{(1)}^{-1}] \right\}^{-1} [\zeta' \otimes Q_{(1)}^{-1}] \mathbf{H}^{*1/2}$  and it is therefore a projection matrix of rank  $p_1$ , acting as a matrix that selects only the first  $p_1$  elements of  $\mathcal{B}_s^*(\lambda)$ .<sup>17</sup> It follows that  $\mathcal{B}_s^{*'}(\lambda) A \mathcal{B}_s^*(\lambda) = \mathcal{B}_{p_1}'(\lambda) \mathcal{B}_{p_1}(\lambda)$ , therefore  $CM_T^* \Rightarrow \mathcal{G}_{p_1}(\lambda)$ , the same distribution as in Theorem 2, but with  $p_1$  degrees of freedom instead of  $p$ .

*Comment 4.* If the model is dynamic in the sense that the long-run and short-run variances differ ( $\mathbf{H}^* \neq \Omega^*$ ), then the distribution in Theorem 3 does not simplify, as the intuition outlined in *Comment 3* no longer holds. For example, if  $x_{t(1)}\epsilon_t$  are autocorrelated, then  $x_{t(1)}u_t$  are autocorrelated, but the  $CM_T^*$  test does not correct for autocorrelation, yielding a more complicated distribution. Even if a HAC correction would be employed, there is still the problem of dynamic misspecification: lags of  $x_{t(1)}\epsilon_t$  might be correlated with lags of  $x_{t(1)}x_{t(2)}$ , yielding  $\mathbf{H}^* \neq \Omega^*$  and therefore the more complicated distribution in Theorem 3.

*Comment 5.* Theorem 3 shows that, in the general case of dynamic misspecification (with  $Q_{(12)} \neq 0$ ), the usual critical values from Theorem 2 no longer apply. Allowing for conditional heteroskedasticity, Theorem 3 demonstrates that the size distortions of the CM test are dependent on several parameters

<sup>16</sup> This was also mentioned in (Chong 2003), in the comments after their Theorem 3.

<sup>17</sup> A formal proof of this statement can be found in (Hall et al. 2012, Supplemental Appendix, page 23.)

of the data generating process, and that correcting for heteroskedasticity does not help in overcoming this problem.

## 4. Simulation Results

### 4.1. Correct Specification and Various Misspecifications

The objective of the simulation analysis was to compare the size and power of unconditional moments, UM/UV, break tests to their conditional moments, CM/CV, counterparts, under correct regression model specification, and under static and dynamic misspecification. We evaluated the size and power of the tests for alternative model specifications, sample sizes, as well as structural break sources and sizes.<sup>18</sup> We considered sample sizes  $T = 100, 200, 500, 1000$  with a break in the middle of the sample,  $T_0 = [0.5T]$  and four data generating processes (DGPs). We also considered alternative break points and our results are robust to  $T_0 = [0.25T]$  and  $T_0 = [0.75T]$ . For all simulations, we used the critical values reported in (Andrews 2003). For DGPs with static errors, we calculated  $CM_T^*$  with  $\hat{V}_\lambda$  as described in (4). For a static DGP with i.i.d. errors, we calculated  $UM_T^*$  without the HAC adjustment, but with split sample variance estimators as for the  $CM_T^*$  to make the comparison fair. For the same purpose of fair comparison, for the DGP with static regressors and AR(1) errors, both the  $UM_T^*$  and the  $CM_T^*$  tests employed the Newey–West HAC estimator over the full sample as an estimator of the variances present in these test statistics.

The results are organized as follows: the sizes of all tests are reported in Tables and the (size-adjusted) powers in Figures.<sup>19,20</sup> Tables 1 and 2 and Figures 1–3 refer to mean tests, and Tables 3 and 4 and Figure 4 refer to variance tests. Tables 1 and 3 and Figures 1–4 are for correctly specified models, while Tables 2 and 4 are for misspecified models.

We consider four DGPs, some of which we analyze under both correct specification and misspecification. The first DGP is a simple AR(1) model with iid errors:

$$DGP1: y_t = \alpha_t + \beta_t y_{t-1} + \epsilon_t, \epsilon_t \sim iid \mathcal{N}(0, 1), (t = 1, \dots, T).$$

All simulations were performed in Matlab for 10,000 replications<sup>21</sup> and for the AR models we used zero as the starting value and 100 burn-in observations. Under the null,  $\alpha_t = \alpha = 1$ , and the persistence ranges  $\beta_t = \beta \in [0.1, 0.7]$ . Under the alternative, there is one break either in the intercept, with  $\alpha_t = 1 + \delta_\alpha \mathbf{1}_{t > T_0}$ , and  $\delta_\alpha \in (0, 2]$ , or in the slope, with  $\beta_t = 0.1 + \delta_\beta \mathbf{1}_{t > T_0}$ , and  $\delta_\beta \in (0, 0.6]$ .

For DGP1, we estimated only the correctly specified dynamic model. The size of the CM and UM break tests (under the null) are reported in the top panel of Table 1. Using the 5% critical values, we found that the UM test exhibits slightly better size for small sample sizes of  $T = 100$  relative to the CM test which yields size around 10%. For large sample sizes of  $T > 500$ , both tests approached the nominal level, as expected. Under the alternative, we plot the size-adjusted power functions in Figure 1. When the break occurs in the slope parameter, the UM and CM tests have similar power as

<sup>18</sup> The unconditional mean and variance sup Wald tests require a long-run variance estimator. We report the Newey and West (1994) HAC estimator with the data dependent bandwidth therein and the Bartlett kernel, as explained in detail in Section 2. The Andrews (1991) fixed bandwidth HAC estimator leads to slightly worse performance across all tests and designs; results are available upon request from the authors.

<sup>19</sup> For a static DGP with i.i.d. errors (DGP3, detailed below), we report the power of the tests based on asymptotic critical values rather than size-adjusted powers, because the size distortions are minor.

<sup>20</sup> The size-adjusted powers are computed as follows: for a DGP under the alternative of one break, we take the parameter values after the break, and use these parameter values for generating the DGP under the null, which will have the same sample size as the DGP under the alternative. We simulate this null DGP and take the 95% quantile of the empirical distribution of a test statistic as the empirical critical value to be used. We then simulate the corresponding DGP under the alternative, and calculate the empirical rejection frequency using the corresponding empirical critical values. Note that, by construction, all size-adjusted power plots start at 5%, which is the corresponding empirical rejection frequency for a DGP under the null of no break using its simulated empirical 95% critical values.

<sup>21</sup> Only for the static model in DGP3, we used 2000 simulations because they were sufficient to get accurate Monte Carlo results.

the sample size grows. The CM test performs only mildly better for moderate changes in the AR slope parameter (with maximum relative gains in power of 10% for  $T = 100$ ). On the other hand, when the break is in the intercept, the UM test has better power in small sample sizes (of  $T = 100, 200$ ), with up to 20% gains vis-a-vis the CM test.<sup>22</sup>

The second DGP is an AR(4) model with iid errors:

$$DGP2 : y_t = \alpha_t + \beta'_t \text{vec}(y_{t-1}, \dots, y_{t-4}) + \epsilon_t, \epsilon_t \sim iid \mathcal{N}(0, 1), (t = 1, \dots, T).$$

We set  $\beta'_t = (\beta_{t,1}, 0.2, 0.15, 0.075)$  to represent the memory decaying pattern encountered in many time series in economics. Under the null, we set  $\alpha_t = \alpha = 1$  and vary  $\beta_{t,1} = \beta \in [0.1, 0.3]$ .

We analyze the impact of dynamic misspecification: the true DGP is an AR(4) model, but we estimate an AR(1) model or an AR(2) model instead.<sup>23</sup>

The top two panels of Table 2 show that underestimating the number of lags causes severe size distortions of the CM test, of up to 60% even for small levels of forgone persistence. This effect does not die out even for large sample sizes of  $T = 1000$ . In contrast, the UM test is not so severely oversized, especially for large samples; the size distortions reach a maximum of 13% for large samples of  $T = 1000$ . Our simulation results indicate that, although the HAC estimator that corrects for dynamics in error term of the UM test may be less reliable in small samples, it results in much smaller size distortions than if we instead used a CM test for a misspecified model. In Section 4.2, we provide further evidence for this using DGP2; we consider information criteria to select the number of lags and show the performance of the CM test for various lag lengths.

The third DGP is a static model with iid errors:

$$DGP3 : y_t = \alpha_t + \beta_t X_t + \epsilon_t, \epsilon_t \sim iid \mathcal{N}(0, 1), X_t \sim iid \mathcal{N}(1, 1), X_t \perp \epsilon_s, (t, s = 1, \dots, T).$$

Under the null, we set  $\alpha_t = \alpha = 1$  and vary  $\beta_t = \beta \in [0.1, 0.9]$ . Under the alternative, there is one break either in the intercept, where  $\alpha_t = 1 + \delta_\alpha \mathbf{1}_{t > T_0}$ , and  $\delta_\alpha \in (0, 2]$ , or in the slope, where  $\beta_t = 0.1 + \delta_\beta \mathbf{1}_{t > T_0}$ , and  $\delta_\beta \in (0, 0.8]$ .

For DGP3, we analyzed both correctly specified and misspecified models. As expected, if we estimated the correctly specified static model in DGP3, the size of both the CM and UM break tests is close to the nominal size, as shown by the second panel in Table 1. The corresponding power curves in Figure 2 are again similar for the two tests, especially as  $T$  increases.

However, if instead we estimated an AR(1) model, the results in the third panel of Table 2 show that the UM test is undersized for small sample sizes and that its size improves for  $T > 500$ . In contrast, the CM test is oversized for small samples. For smaller samples, misspecifying the regressors compromises the power of the CM test which can be up to around 20% smaller than that of the UM test when  $T = 100$ .

The fourth DGP is a static model with AR(1) errors:

$$DGP4 : y_t = \alpha_t + \beta_t X_t + \epsilon_t, \epsilon_t = 0.6\epsilon_{t-1} + v_t, v_t \sim iid \mathcal{N}(0, 1), (t = 1, \dots, T).$$

For comparison purposes, in DGP4, the  $X_t$  and the null and alternatives are generated in the same way as in DGP3. For DGP4, we analyzed both correctly specified and misspecified models. Under correct specification, the last panel of Table 1 shows that the UM test is correctly sized for all sample sizes, whereas the CM test is oversized even for large sample sizes. The size of the CM tests can reach up to 10% even when  $T = 1000$  (and the nominal size is 5%).

<sup>22</sup> Note that, for DGP1, the unconditional mean of  $y_t$  is equal to  $\alpha_t/(1 - \beta_t)$ . If  $\alpha_t/(1 - \beta_t)$  is close to zero regardless of  $t$ , the UM test will, by design, have little power for a break in the slope  $\beta_t$ . Therefore, if a slope break is the only break of interest, it should be tested directly via the CM test for partial structural change in slopes.

<sup>23</sup> Note that we do not plot size-adjusted powers for this DGP, and we only do so in general for correctly specified models.

Furthermore, we consider a nonlinear misspecification by estimating the model with  $X_t^2$  instead of  $X_t$ , similar to (Chong 2003). The nonlinear misspecification yields oversized CM tests across all sample sizes. The last panel of Table 2 shows that, even for  $T = 1000$ , the traditional CM test yields size of around 13%.

We now turn to examine the size and power of tests for breaks in the variance of the residuals of the regression models by comparing the UV and CV tests.<sup>24</sup> We considered the same DGPs as before, but we set  $\alpha_t = 1$ ,  $\beta_t = 0.5$ . For DGP1–3, we let  $\epsilon_t \sim iid \mathcal{N}(0, \sigma_t)$ , and, for DGP4, we let  $\nu_t \sim iid \mathcal{N}(0, \sigma_t)$ . Under the null hypotheses, we fixed  $\sigma_t = \sigma \in [1, 2.6]$ . Under the alternative, we set  $\sigma_t = 1 + \delta_\sigma \mathbf{1}[t > T_0]$ , and let  $\delta_\sigma \in (0, 1.6]$ . As before, we estimated both correctly specified and misspecified models.

When the estimated model is correctly specified, as considered in DGP1 and DGP4, the size of both CV and UV tests are close to the nominal size for  $T \geq 200$ , as shown in Table 3. However, the powers of these two tests differ. Figure 4 shows that, for DGP1, the CV test has better power for small sample sizes across all break sizes, including small breaks, as  $T$  increases. For DGP4, Figure 4 shows that the power curves of the CV tests and UV tests are the same.

If instead, a misspecified model is estimated for DGP1–DGP4, the CV test appears to enjoy good size properties, as shown in Table 4. The exception is the oversizing reported in the top panel of Table 4, which is due to underestimating the lag order; in this case the size does not improve as the sample increases. Our analysis shows that misspecifying the dynamics of the conditional mean of the regression model yields an oversized CV test.

Summarizing, the simulation results show that under correct model specification, the UM/UV and CM/CV have similar size and power. In contrast, under static nonlinear and dynamic misspecifications, the CM/CV tests are severely oversized, having both finite and large sample distortions. While the UM/UV tests may also occasionally exhibit mild size distortions, they feature similar power properties as the CM/CV tests, especially in larger samples. Therefore, the UM/UV tests can be a valuable tool for detecting breaks, because, in applied work, misspecification is likely to occur and bias the CM/CV break test results.<sup>25</sup>

<sup>24</sup> The results are very similar for the UA and CA tests and they are available upon request.

<sup>25</sup> Other types of model misspecifications may also affect the size and power of the (CM and CV) structural break tests. Analyzing them is beyond the scope of this paper, but further results regarding these misspecifications can be found in (Chong 2003; Pitarakis 2004), among others.

**Table 1.** Size of the UM/CM tests in correctly specified models.

DGP	Trim	Model		$UM_T^*$					$CM_T^*$		
		$\alpha$	$\beta$	$T = 100$	200	500	1000	100	200	500	1000
DGP1—AR(1) model, iid errors	15%	1	0.1	0.034	0.040	0.047	0.052	0.067	0.056	0.050	0.050
			0.2	0.039	0.047	0.054	0.056	0.078	0.060	0.050	0.051
			0.3	0.040	0.048	0.053	0.057	0.082	0.063	0.056	0.052
			0.4	0.044	0.055	0.056	0.056	0.091	0.064	0.058	0.048
			0.5	0.052	0.052	0.062	0.060	0.107	0.068	0.063	0.050
			0.6	0.061	0.058	0.070	0.065	0.120	0.079	0.055	0.058
			0.7	0.082	0.072	0.075	0.069	0.141	0.092	0.062	0.058
DGP3—static model, iid errors	15%	1	0.1	0.067	0.052	0.060	0.050	0.075	0.063	0.059	0.041
			0.2	0.073	0.053	0.049	0.043	0.086	0.052	0.054	0.047
			0.3	0.067	0.056	0.054	0.052	0.083	0.062	0.049	0.049
			0.4	0.084	0.055	0.049	0.052	0.091	0.058	0.042	0.052
			0.5	0.061	0.053	0.058	0.051	0.070	0.056	0.061	0.051
			0.6	0.058	0.058	0.046	0.054	0.080	0.056	0.044	0.052
			0.7	0.073	0.056	0.056	0.055	0.081	0.059	0.054	0.048
			0.8	0.063	0.060	0.047	0.048	0.079	0.059	0.051	0.048
			0.9	0.069	0.062	0.055	0.050	0.092	0.065	0.050	0.052
DGP 4—static model, AR(1) errors	15%	1	0.1	0.063	0.055	0.067	0.064	0.115	0.107	0.090	0.098
			0.3	0.065	0.061	0.066	0.062	0.113	0.111	0.090	0.100
			0.5	0.063	0.060	0.063	0.059	0.122	0.109	0.091	0.106
			0.7	0.060	0.061	0.070	0.060	0.111	0.113	0.104	0.109
			0.9	0.064	0.061	0.065	0.065	0.109	0.106	0.086	0.099



**Table 2.** Size of the UM/CM tests in misspecified models.

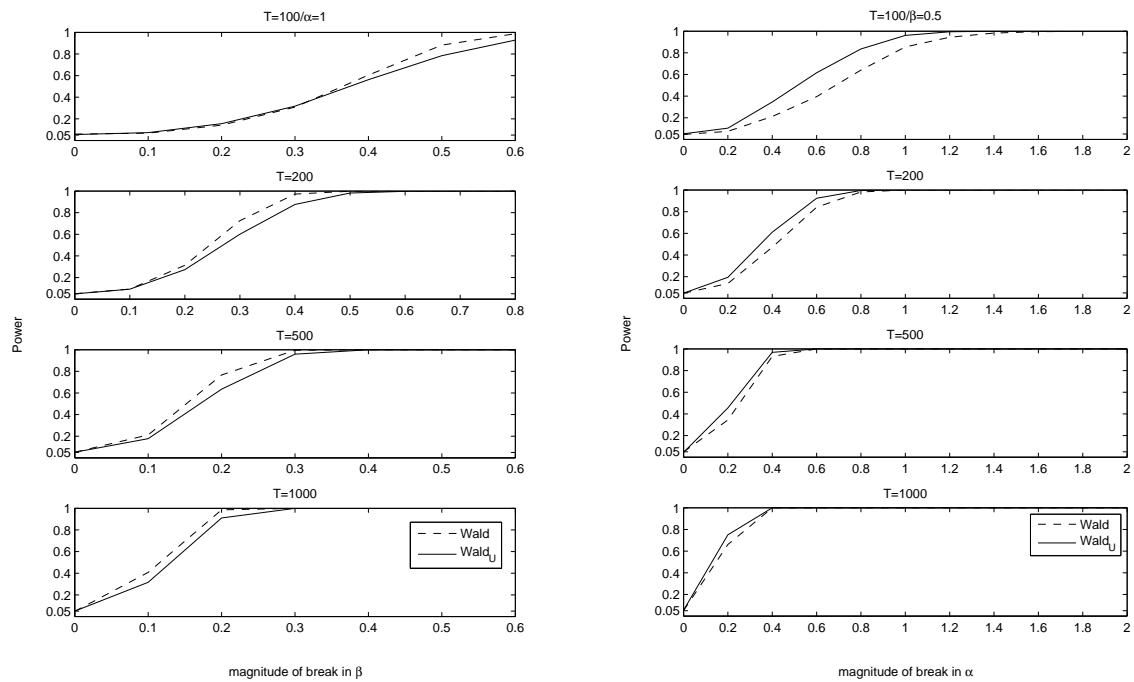
DGP	Estimated Model	Trim	Model		$UM_T^*$				$CM_T^*$			
			$\alpha$	$\beta$	$T = 100$	200	500	1000	100	200	500	1000
DGP2—AR(4), iid errors	AR(1)	15%	1	0.1	0.157	0.114	0.114	0.088	0.427	0.463	0.486	0.508
				0.2	0.191	0.131	0.132	0.102	0.493	0.509	0.531	0.560
				0.3	0.231	0.171	0.174	0.130	0.554	0.579	0.606	0.616
DGP2—AR(4), iid errors	AR(2)	15%	1	0.1	0.157	0.113	0.114	0.088	0.419	0.371	0.336	0.332
				0.2	0.190	0.131	0.131	0.103	0.455	0.378	0.344	0.336
				0.3	0.230	0.170	0.173	0.129	0.496	0.421	0.374	0.358
DGP3—static model, iid errors	AR(1)	15%	1	0.1	0.067	0.052	0.060	0.050	0.067	0.054	0.053	0.046
				0.2	0.073	0.053	0.049	0.043	0.067	0.054	0.047	0.049
				0.3	0.067	0.056	0.054	0.052	0.070	0.056	0.048	0.056
				0.4	0.084	0.055	0.049	0.052	0.070	0.053	0.048	0.052
				0.5	0.061	0.053	0.058	0.051	0.065	0.055	0.051	0.051
				0.6	0.058	0.058	0.046	0.054	0.074	0.053	0.049	0.051
				0.7	0.073	0.056	0.056	0.055	0.067	0.055	0.049	0.048
				0.8	0.063	0.060	0.047	0.048	0.073	0.054	0.050	0.051
				0.9	0.069	0.062	0.055	0.050	0.065	0.053	0.053	0.052
DGP4—static model, AR(1) errors	$X_t^2$ instead of $X_t$	15%	1	0.1	0.063	0.055	0.067	0.064	0.207	0.165	0.111	0.121
				0.3	0.065	0.061	0.066	0.062	0.204	0.158	0.120	0.125
				0.5	0.063	0.060	0.063	0.059	0.211	0.162	0.121	0.131
				0.7	0.060	0.061	0.070	0.060	0.209	0.160	0.127	0.128
				0.9	0.064	0.061	0.065	0.065	0.216	0.173	0.137	0.143

**Table 3.** Size of the UV/CV break tests in correctly specified models.

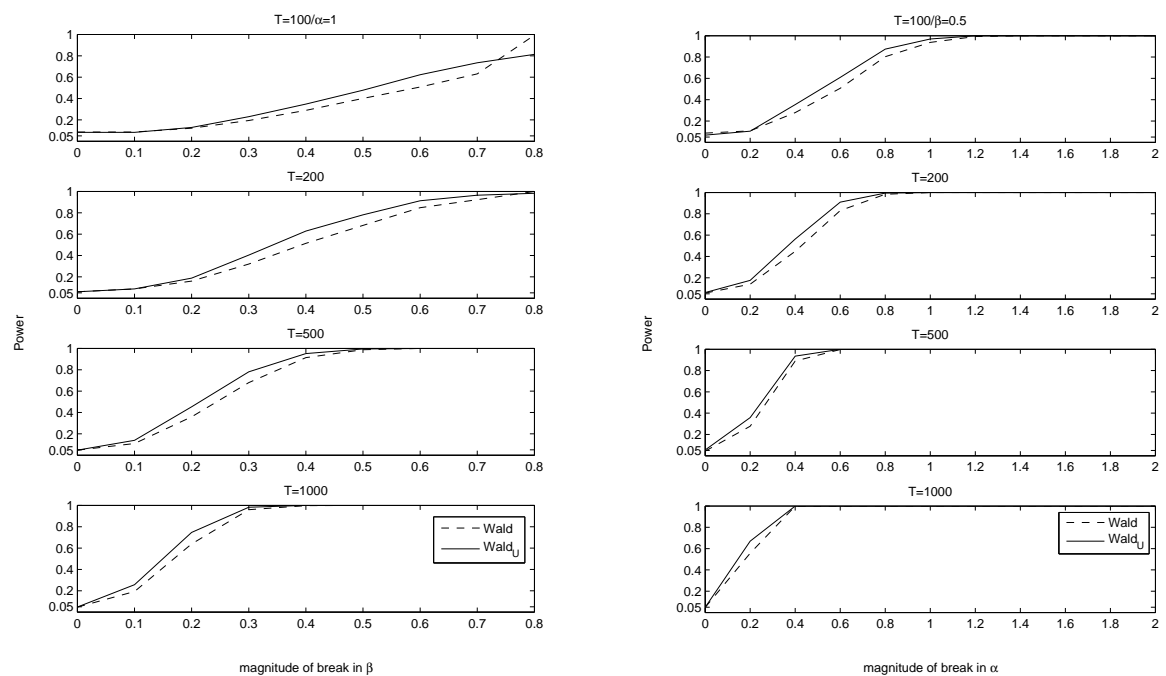
DGP	Trim	Model		$UV_T^*$					$CV_T^*$		
		$\alpha$	$\sigma$	$T = 100$	200	500	1000	100	200	500	1000
DGP1—AR(1) model, iid errors	15%	1	1	0.042	0.052	0.051	0.059	0.0753	0.0570	0.050	0.048
			1.2	0.042	0.047	0.051	0.050	0.0730	0.0533	0.050	0.048
			1.4	0.039	0.046	0.050	0.056	0.0754	0.0584	0.050	0.052
			1.6	0.040	0.043	0.052	0.054	0.0745	0.0517	0.053	0.047
			1.8	0.040	0.045	0.052	0.055	0.0735	0.0543	0.053	0.051
			2.0	0.040	0.047	0.052	0.054	0.0775	0.0572	0.053	0.052
			2.2	0.041	0.044	0.054	0.057	0.0768	0.0592	0.050	0.051
			2.4	0.043	0.049	0.054	0.055	0.0774	0.0595	0.051	0.048
			2.6	0.043	0.047	0.052	0.052	0.0738	0.0581	0.049	0.048
DGP4—static model, AR(1) errors	15%	1	1	0.041	0.051	0.052	0.057	0.076	0.067	0.059	0.060
			1.2	0.041	0.049	0.053	0.055	0.075	0.067	0.060	0.058
			1.4	0.042	0.048	0.053	0.058	0.076	0.066	0.060	0.060
			1.6	0.039	0.047	0.056	0.054	0.071	0.064	0.063	0.056
			1.8	0.041	0.048	0.056	0.053	0.077	0.066	0.063	0.057
			2.0	0.043	0.048	0.055	0.057	0.076	0.062	0.063	0.061
			2.2	0.039	0.049	0.058	0.052	0.073	0.064	0.063	0.056
			2.4	0.042	0.053	0.056	0.054	0.075	0.072	0.061	0.057
			2.6	0.045	0.048	0.053	0.057	0.072	0.061	0.058	0.060

**Table 4.** Size of the CV/UV tests in misspecified models.

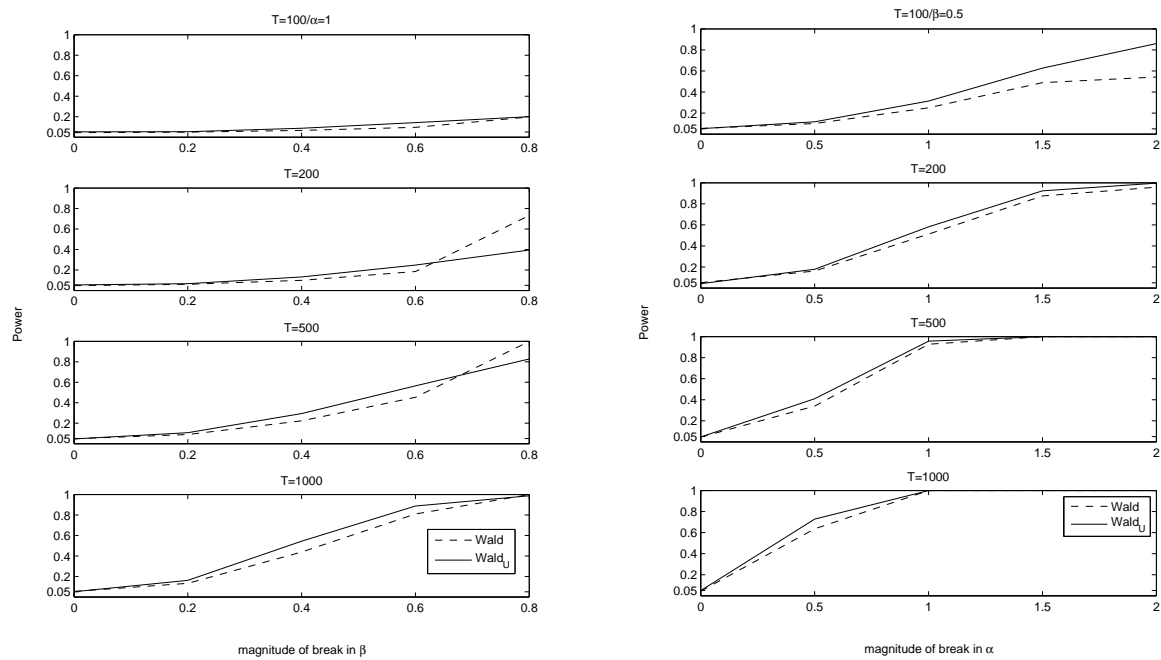
DGP	Estimated Model	Trim	Model		$UV_T^*$				$CV_T^*$			
			$\alpha$	$\sigma$	$T = 100$	200	500	1000	100	200	500	1000
DGP2—AR(4) model, iid errors	AR(1)	15%	1	1	0.057	0.064	0.075	0.063	0.110	0.107	0.124	0.121
				1.2	0.061	0.063	0.072	0.065	0.112	0.102	0.118	0.129
				1.4	0.059	0.068	0.072	0.065	0.116	0.112	0.119	0.129
				1.6	0.062	0.061	0.072	0.063	0.116	0.100	0.118	0.125
				1.8	0.055	0.056	0.073	0.067	0.113	0.104	0.120	0.125
				2	0.057	0.064	0.068	0.062	0.119	0.106	0.115	0.123
				2.2	0.060	0.063	0.077	0.064	0.118	0.099	0.119	0.122
				2.4	0.057	0.063	0.075	0.067	0.108	0.108	0.121	0.125
DGP1—AR(1) model, iid errors	AR(4)	15%	1	2.6	0.062	0.065	0.074	0.065	0.111	0.113	0.116	0.125
				1	0.042	0.052	0.051	0.059	0.075	0.054	0.047	0.047
				1.2	0.042	0.047	0.051	0.050	0.075	0.052	0.048	0.046
				1.4	0.039	0.046	0.050	0.056	0.070	0.054	0.049	0.050
				1.6	0.040	0.043	0.052	0.054	0.079	0.052	0.049	0.046
				1.8	0.040	0.045	0.052	0.055	0.076	0.051	0.049	0.050
				2	0.040	0.047	0.052	0.054	0.075	0.056	0.050	0.050
				2.2	0.041	0.044	0.054	0.057	0.073	0.051	0.047	0.049
DGP3—static model, iid errors	AR(1)	15%	1	2.4	0.043	0.049	0.054	0.055	0.074	0.053	0.048	0.046
				2.6	0.043	0.047	0.052	0.052	0.076	0.055	0.047	0.046
				1	0.030	0.033	0.041	0.047	0.077	0.057	0.051	0.051
				1.2	0.030	0.032	0.040	0.048	0.076	0.058	0.048	0.053
				1.4	0.029	0.032	0.043	0.047	0.075	0.055	0.055	0.051
				1.6	0.028	0.032	0.041	0.044	0.072	0.054	0.049	0.050
				1.8	0.028	0.033	0.039	0.049	0.071	0.054	0.047	0.053
				2	0.029	0.037	0.040	0.048	0.074	0.061	0.049	0.054
DGP4—static model, AR(1) errors	$X_t^2$ instead of $X_t$	15%	1	2.2	0.029	0.033	0.040	0.043	0.080	0.057	0.049	0.045
				2.4	0.026	0.036	0.039	0.044	0.069	0.058	0.048	0.049
				2.6	0.029	0.032	0.045	0.045	0.074	0.055	0.056	0.050
				1	0.041	0.051	0.052	0.057	0.075	0.064	0.059	0.058
				1.2	0.041	0.049	0.053	0.055	0.074	0.063	0.058	0.055
				1.4	0.042	0.048	0.053	0.058	0.073	0.064	0.058	0.058
				1.6	0.039	0.047	0.056	0.054	0.071	0.063	0.063	0.057
				1.8	0.041	0.048	0.056	0.053	0.070	0.064	0.060	0.059
				2	0.043	0.048	0.055	0.057	0.076	0.065	0.058	0.059
				2.2	0.039	0.049	0.058	0.052	0.069	0.064	0.063	0.054
				2.4	0.042	0.053	0.056	0.054	0.075	0.067	0.060	0.052
				2.6	0.045	0.048	0.053	0.057	0.075	0.060	0.057	0.058



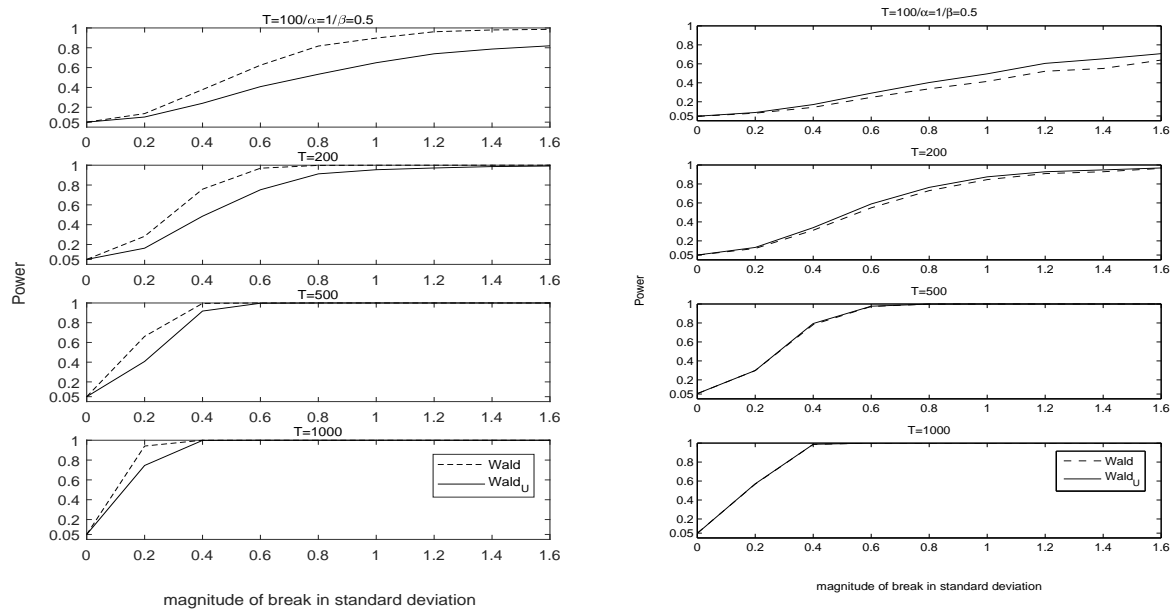
**Figure 1.** DGP1: Size-adjusted power for a correctly specified AR(1) model with iid errors. Note: *Wald* is the CM test, and *Wald<sub>U</sub>* is the UM test.



**Figure 2.** DGP3: Power for a correctly specified static model with iid errors. Note: *Wald* is the CM test, and *Wald<sub>U</sub>* is the UM test.



**Figure 3.** DGP4: Size-adjusted power for a correctly specified static model with AR(1) errors. Note: *Wald* is the CM test, and *Wald<sub>U</sub>* is the UM test.



**Figure 4.** DGP1 (left); and DGP4 (right) with size-adjusted power for a correctly specified model: with an AR(1) lag and iid errors (left); or with static regressors and AR(1) errors (right). Note: *Wald* is the CV test, and *Wald<sub>U</sub>* is the UV test.

#### 4.2. Dynamic Misspecification

In this section, we further analyze DGP2, given by the AR(4) model below, with i.i.d.  $\mathcal{N}(0, 1)$  errors, and 400 burn-in observations:

$$y_t = 1 + \beta y_{t-1} + \gamma_1 y_{t-2} + \gamma_2 y_{t-3} + \gamma_3 y_{t-4} + \epsilon_t.$$

We consider two variants of this model: DGP2-A and DGP2-B. DGP2-A is a typical DGP encountered in applied work, where the coefficients on the higher order lags are smaller than those of the first and second lag ( $\beta \in \{0.1, 0.2, 0.3\}$ ,  $\gamma_1 = 0.2$ ,  $\gamma_2 = 0.15$ ,  $\gamma_3 = 0.075$ ). Note that the three empirical applications in Section 5 show similar patterns of smaller coefficients on higher order lags. The second DGP, DGP2-B, is less realistic, allowing for the coefficient on the fourth lag to be larger than that on the other lags ( $\beta = 0.1$ ,  $\gamma_1 = 0$ ,  $\gamma_2 = 0$ ,  $\gamma_3 \in \{0.175, 0.275, 0.375\}$ ). Such DGP is plausible if the seasonality at lag four was not yet removed from the data, and it is encountered less in macroeconomic datasets, because they are typically seasonally adjusted. Nevertheless, we include it for completeness.

In most empirical applications, the number of lags are selected with AIC and BIC. Hence, we begin by showing the empirical distribution of the number of lags selected by AIC and BIC in 10,000 simulations for both DGP2-A and DGP2-B. Figures 5–7 show that, for DGP2-A, both the AIC and BIC tend to incorrectly select the number of lags even for sample sizes of  $T = 1000$ . In particular, the BIC tends to underestimate the number of lags about 60% of the time. What is perhaps more surprising is the even the AIC underestimates the number of lags about 20% of the times in large samples. This result implies that both the AIC and BIC are not reliable for typical datasets where the higher order lags enter with smaller coefficients; unfortunately, those are the same datasets typically encountered in applications, for which the problem of dynamic misspecification of the conditional mean seems unavoidable.

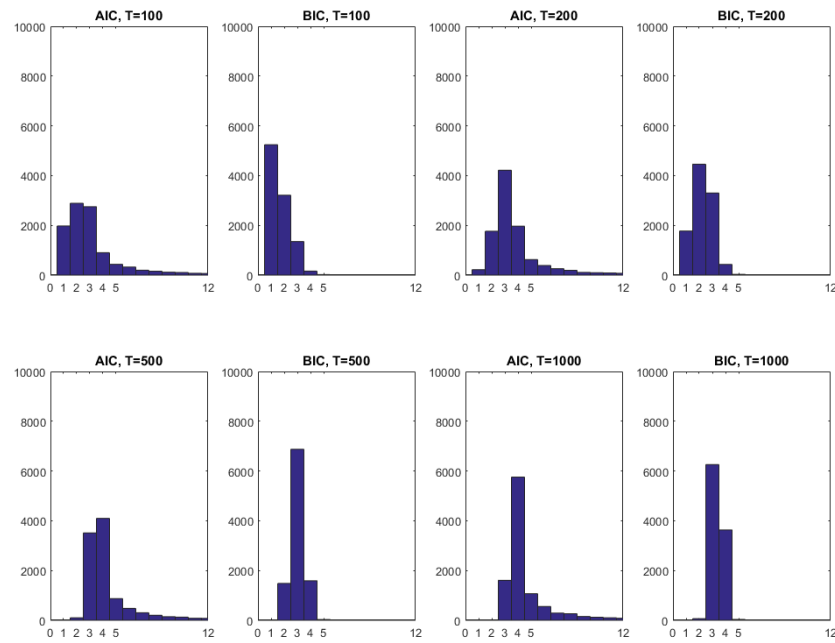
For completeness, we also show the empirical distribution of the number of lags estimated by AIC and BIC for DGP2-B, for which the fourth coefficient is larger than all the three others. Figures 8–10 show that in this case, BIC selects the true number of lags with probability approaching one as the sample size gets large. However, notice that, when  $\beta = 0.1$  and  $\gamma_3 = 0.175$ , it still selects only one lag about 10% of the time for large sample sizes  $T = 1000$ . This implies that  $\gamma_3$  has to be quite large, reminiscent of models with uncorrected seasonalities, for a correct selection of the number of lags. We also notice that AIC tends to select the true number of lags or a larger number, but only as the sample size rises to  $T = 500$ .

From these figures, we conclude that AIC and BIC can both incorrectly choose the number of lags, leading to dynamic misspecification in the conditional mean of the CM tests. In Tables 5 and 6, we show for DGP2-A and DGP2-B the empirical sizes of both the UM and CM tests for a nominal size of 5% and 10,000 simulations. For the CM test, we impose different lags, from  $k = 0$  to  $k = 8$ .

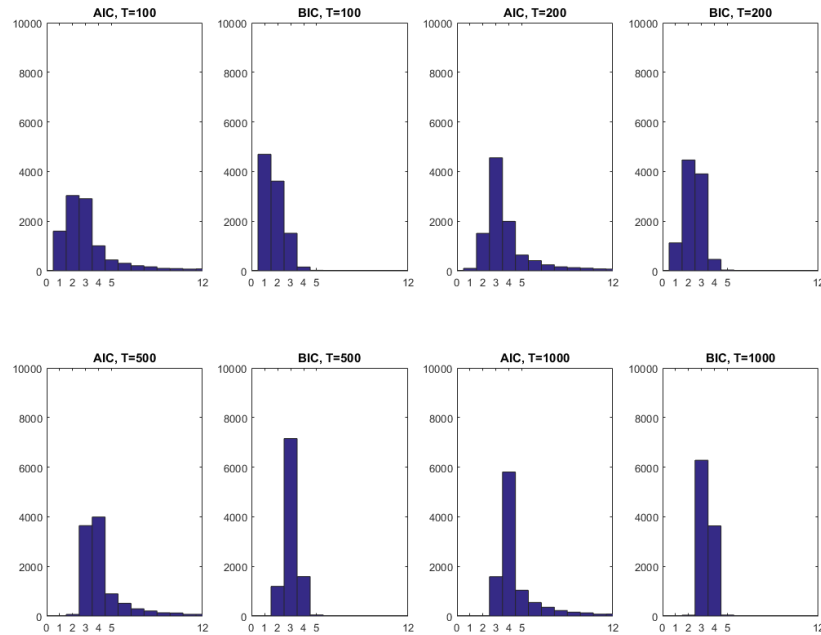
For DGP2-A presented in Table 5, we notice severe size distortions when underspecifying the number of lags, with the sizes of the CM test up to 61.4% for one lag and 21.7% for two lags and  $T = 1000$ . When the number of lags is overspecified, the size distortions of the CM tests remain severe for  $T = 100$ , with sizes going up to 69.9%, although they do decrease towards the nominal level as the sample size increases. The size distortions in the UM test are less severe for small sample sizes like  $T = 100$ , and can reach up to 22.1%, due to the HAC correction this test employs. As the sample size increases, the size of the UM test improves to around 10% for  $T = 1000$ .

For DGP2-B in Table 6, we notice the same patterns. Given these results and Figures 8–10, we conclude that, even for seasonally unadjusted models such as DGP2-B, performing the CM test with the number of lags selected by BIC is only reliable if the sample sizes are large; otherwise, the UM test would be a good alternative.

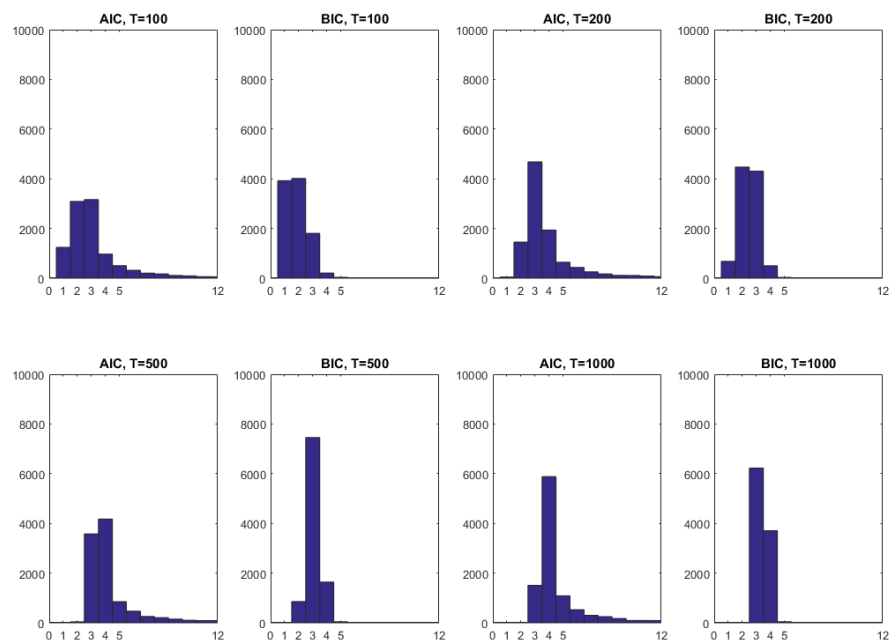




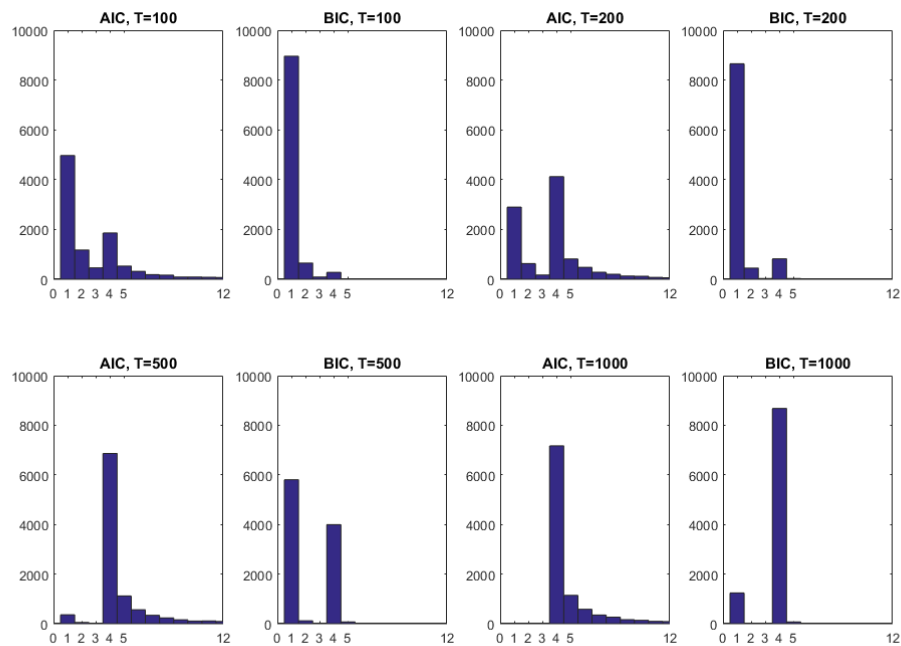
**Figure 5.** Empirical distribution of the number of lags selected by AIC and BIC. DGP2: AR(4) model:  $y_t = 1 + 0.1 y_{t-1} + 0.2 y_{t-2} + 0.15 y_{t-3} + 0.075 y_{t-4} + \epsilon_t$ .



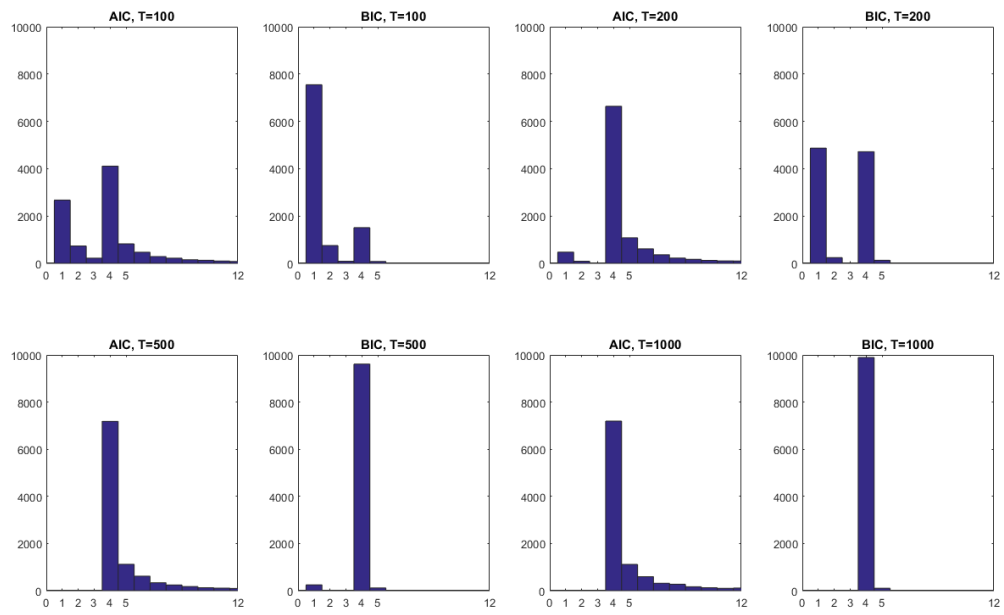
**Figure 6.** Empirical distribution of the number of lags selected by AIC and BIC. DGP2: AR(4) model:  $y_t = 1 + 0.2 y_{t-1} + 0.2 y_{t-2} + 0.15 y_{t-3} + 0.075 y_{t-4} + \epsilon_t$ .



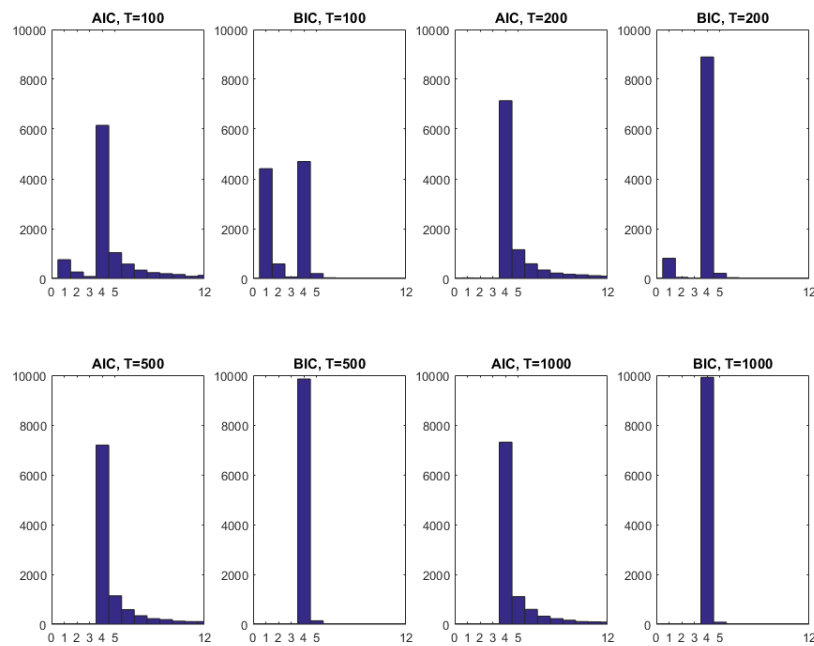
**Figure 7.** Empirical distribution of the number of lags selected by AIC and BIC. DGP2: AR(4) model:  $y_t = 1 + 0.3 y_{t-1} + 0.2 y_{t-2} + 0.15 y_{t-3} + 0.075 y_{t-4} + \epsilon_t$ .



**Figure 8.** Empirical distribution of the number of lags selected by AIC and BIC. DGP2: AR(4) model:  $y_t = 1 + 0.1 y_{t-1} + 0.175 y_{t-4} + \epsilon_t$ .



**Figure 9.** Empirical distribution of the number of lags selected by AIC and BIC. DGP2: AR(4) model:  $y_t = 1 + 0.1 y_{t-1} + 0.275 y_{t-4} + \epsilon_t$ .



**Figure 10.** Empirical distribution of the number of lags selected by AIC and BIC. DGP2: AR(4) model:  $y_t = 1 + 0.1 y_{t-1} + 0.375 y_{t-4} + \epsilon_t$ .

**Table 5.** Size of the UM/CM tests in possibly misspecified models.

<b>DGP 2: AR(4) Model: <math>y_t = 1 + \beta y_{t-1} + 0.2 y_{t-2} + 0.15 y_{t-3} + 0.075 y_{t-4} + \epsilon_t</math></b>											
		<i>k</i>	0	1	2	3	4	5	6	7	8
		$UM_T^*$	$CM_T^*$								
<i>T</i> = 100	$\beta = 0.1$	0.153	0.563	0.468	0.309	0.246	0.292	0.346	0.447	0.536	0.661
	$\beta = 0.2$	0.178	0.714	0.515	0.332	0.271	0.323	0.377	0.465	0.558	0.676
	$\beta = 0.3$	0.221	0.849	0.577	0.373	0.300	0.354	0.407	0.500	0.592	0.699
<i>T</i> = 200	$\beta = 0.1$	0.110	0.612	0.464	0.243	0.151	0.140	0.154	0.169	0.184	0.215
	$\beta = 0.2$	0.134	0.774	0.527	0.265	0.168	0.149	0.160	0.180	0.200	0.230
	$\beta = 0.3$	0.170	0.893	0.585	0.292	0.182	0.165	0.178	0.198	0.217	0.245
<i>T</i> = 500	$\beta = 0.1$	0.121	0.660	0.492	0.210	0.109	0.082	0.083	0.088	0.092	0.095
	$\beta = 0.2$	0.131	0.813	0.545	0.214	0.105	0.077	0.076	0.082	0.085	0.087
	$\beta = 0.3$	0.170	0.924	0.606	0.243	0.119	0.086	0.086	0.094	0.095	0.101
<i>T</i> = 1000	$\beta = 0.1$	0.087	0.682	0.501	0.190	0.097	0.064	0.061	0.067	0.067	0.066
	$\beta = 0.2$	0.099	0.838	0.569	0.203	0.097	0.065	0.065	0.069	0.069	0.070
	$\beta = 0.3$	0.119	0.944	0.614	0.217	0.096	0.066	0.063	0.064	0.065	0.065

Note: *k* is the number of lags imposed in the model when performing the  $CM_T^*$  test.

**Table 6.** Size of the UM/CM tests in possibly misspecified models.

<b>DGP 2: AR(4) Model: <math>y_t = 1 + 0.1 y_{t-1} + \gamma_3 y_{t-4} + \epsilon_t</math></b>											
		<i>k</i>	0	1	2	3	4	5	6	7	8
		$UM_T^*$	$CM_T^*$								
<i>T</i> = 100	$\gamma_3 = 0.175$	0.084	0.233	0.157	0.235	0.275	0.254	0.306	0.399	0.498	0.624
	$\gamma_3 = 0.275$	0.143	0.315	0.219	0.335	0.383	0.272	0.327	0.421	0.513	0.637
	$\gamma_3 = 0.375$	0.214	0.405	0.305	0.454	0.516	0.295	0.355	0.455	0.542	0.656
<i>T</i> = 200	$\gamma_3 = 0.175$	0.106	0.239	0.139	0.191	0.203	0.122	0.127	0.152	0.164	0.191
	$\gamma_3 = 0.275$	0.168	0.342	0.221	0.315	0.327	0.128	0.137	0.163	0.183	0.208
	$\gamma_3 = 0.375$	0.242	0.455	0.326	0.453	0.479	0.138	0.148	0.172	0.194	0.226
<i>T</i> = 500	$\gamma_3 = 0.175$	0.138	0.262	0.149	0.193	0.187	0.076	0.081	0.088	0.091	0.096
	$\gamma_3 = 0.275$	0.191	0.365	0.221	0.300	0.293	0.071	0.072	0.075	0.079	0.084
	$\gamma_3 = 0.375$	0.279	0.509	0.351	0.469	0.467	0.077	0.078	0.086	0.084	0.091
<i>T</i> = 1000	$\gamma_3 = 0.175$	0.072	0.266	0.138	0.171	0.164	0.060	0.063	0.062	0.062	0.062
	$\gamma_3 = 0.275$	0.077	0.385	0.224	0.312	0.303	0.063	0.062	0.063	0.065	0.069
	$\gamma_3 = 0.375$	0.092	0.539	0.357	0.475	0.460	0.063	0.063	0.063	0.066	0.068

Note: *k* is the number of lags imposed in the model when performing the  $CM_T^*$  test.

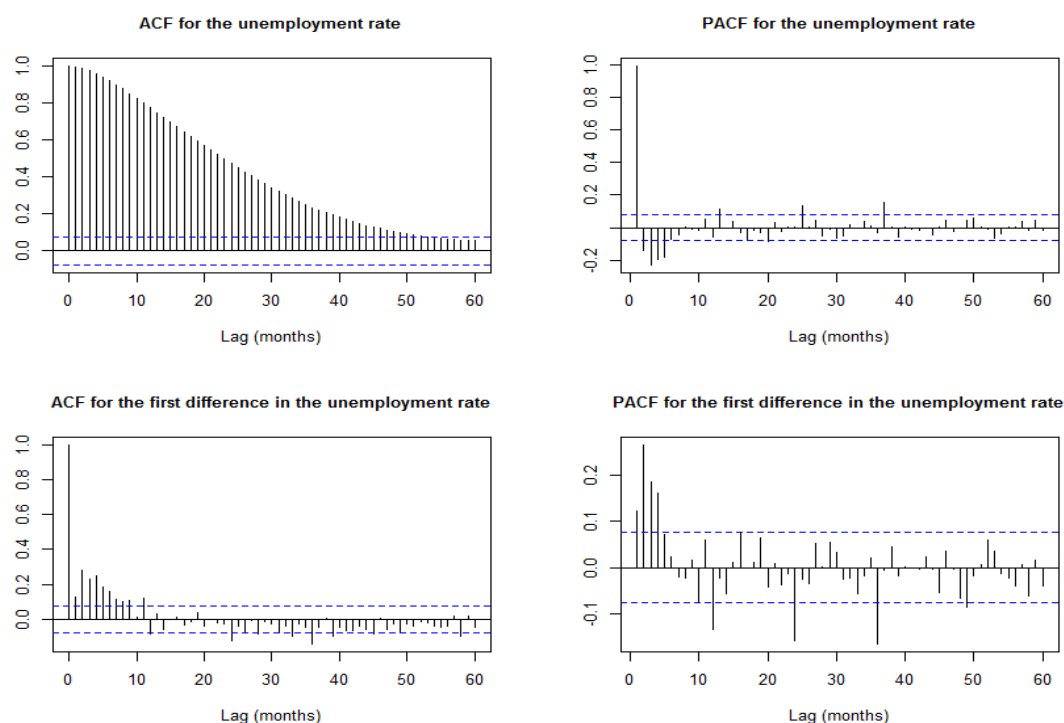
Overall, the simulation results show that, under both static and dynamic misspecifications and for data generating processes typically encountered in applications, the CM/CV tests are severely oversized even in large samples, while the UM/UV tests occasionally exhibit size distortions that are relatively mild. Especially because the size distortions due to dynamic misspecification cannot be fixed by choosing the number of lags with information criteria, while imposing too many lags also leads to size distortions in small samples as we showed above, the UM/UV tests can be a valuable complementary alternative to the CM/UM tests.

## 5. Empirical Illustrations

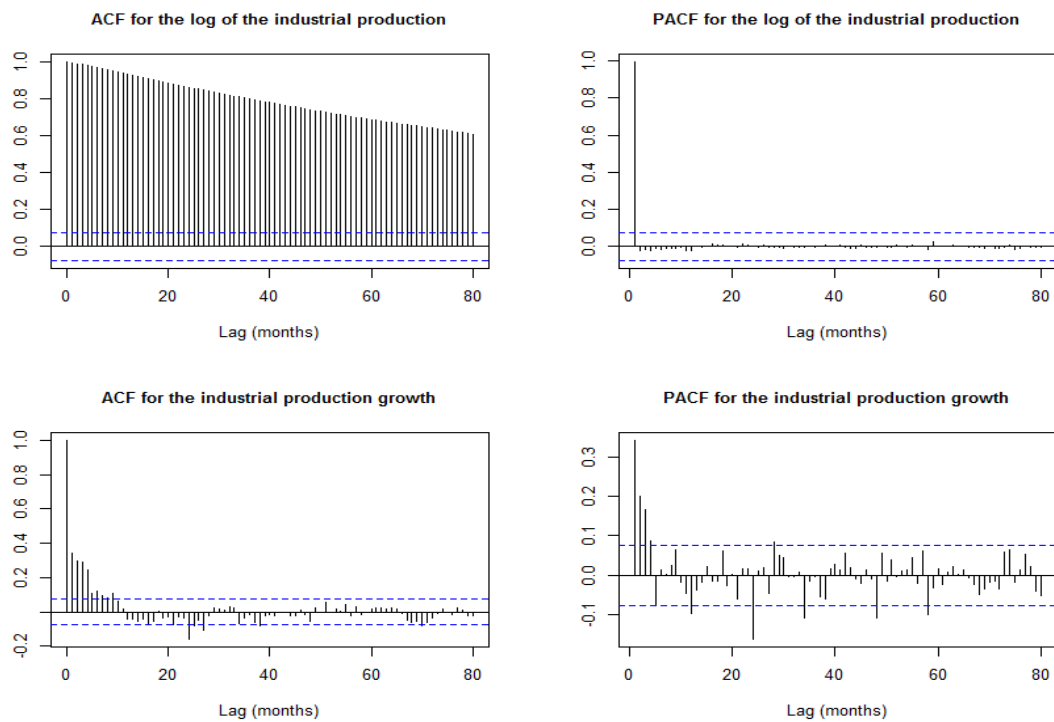
This section illustrates the use of UM/UV in conjunction with CM/CV breaks tests for three macroeconomic series: the US civilian unemployment rate, the industrial production growth and the short term real interest rates. These variables were also examined for breaks in the conditional mean and volatility by [Stock and Watson \(2002\)](#) and [Sensier and van Dijk \(2004\)](#), among others, at quarterly frequency. We employed the analysis on a monthly sample during January 1960–October 2014 with

$T = 658$  to benefit from larger samples (except for the interest rates for which data are available only from April 1960, therefore we used the sample April 1960–October 2014). Our data source is the FRED database at the Federal Reserve Bank of St. Louis. The unemployment rate is the civilian unemployment rate, and the real interest rates are computed as the annualized nominal three-month Treasury Bill minus the three-month CPI inflation rate over the sample period.

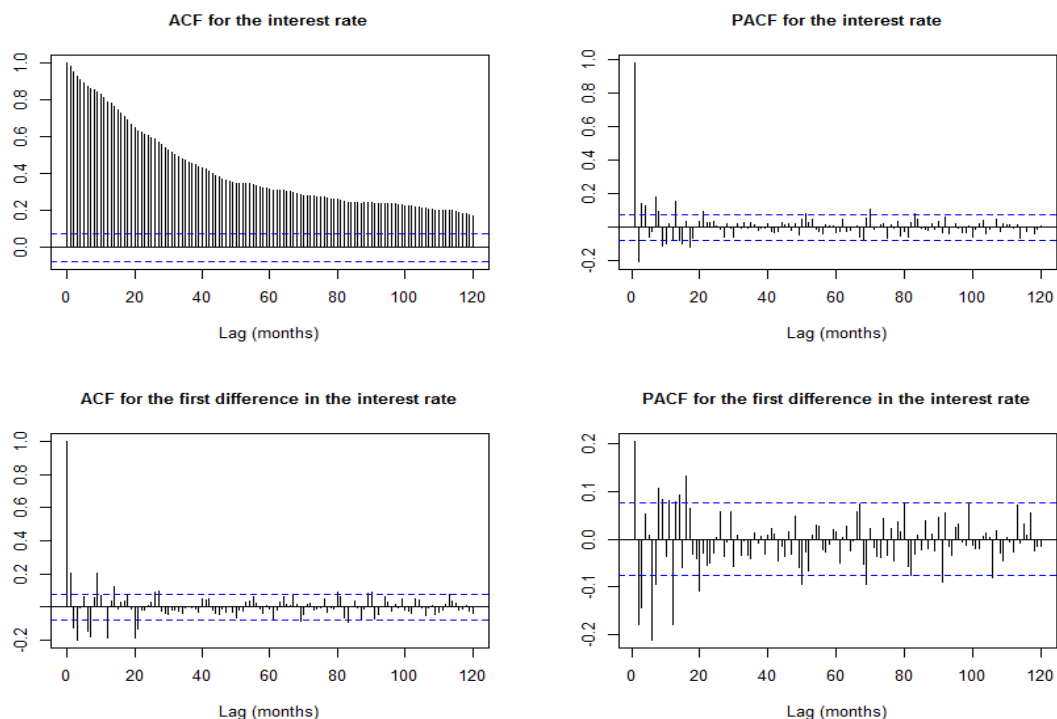
The unemployment rate and the real interest rates are analyzed in levels for two reasons: (1) because both series are measured in rates so they are bounded over the sample period; and (2) because first differencing removes most of the variation in these two series. The industrial production series is typically treated as a unit root process with a possible trend; therefore, as in (Stock and Watson 2002), we analyze the first differences of the logs of the series. Note that the unit root tests give mixed results, but they are also unreliable even if there were no breaks in these series, because the lag length selection may fail as shown in the simulation section. We therefore show the ACF and PACF plots for these series in Figures 11–13. The ACF of the unemployment rate is dying out reasonably fast for a monthly series, so if we believe that the mean of this series does not have breaks, as argued in Section 5.1, then indeed this series should be treated as stationary. The ACF for the interest rates is dying out much slower. However, if we believe that there are mean breaks in the interest rates, as argued in Section 5.2, then this plot is unreliable, and the interest rates should be treated as piece-wise stationary with breaks. The ACF is very persistent for the industrial production series, and if we believe that this series has no mean breaks, as argued in Section 5.3, then it is non-stationary in the sense of having a unit root and/or a trend, and it therefore needs to be first differenced. We refer to the first differences in the log of the industrial production series as the industrial production growth in the rest of the paper.



**Figure 11.** ACF and PACF for the unemployment rate.



**Figure 12.** ACF and PACF for the industrial production.



**Figure 13.** ACF and PACF for the interest rates.

We apply the UM/CM with 5% and 10% trimming to allow detection of potential breaks due to the recent economic crisis. However, the 5% trimming turns out to be problematic in many cases, as we argue in Sections 5.1–5.3. Therefore, we report the UV/CV tests only at the 10% level. For all tests,



we use the critical values in (Andrews 2003).<sup>26</sup> For all series, the CM and CV tests are employed for an AR( $p$ ) model with an intercept, where  $p \in \{1, 4, 12\}$  as these are typical choices in the literature, or  $p$  is selected by AIC and by BIC. When using the AIC and BIC in Tables 7–13, we impose a maximum of 12 lags because this is the typical maximum choice in applied work for monthly data. If we increase the maximum to 30 lags, then only the AIC selects more lags for some series, and this is discussed explicitly below for the series for which this occurs.

We also report for all series a distributed lag model with an intercept and  $p$  lags—DL( $p$ )—of each of two monthly factors: a macro factor, extracted from the mean of a large cross-section of economic and financial US series, and a macro uncertainty factor. The number of lags is selected with AIC and BIC. The two factors are taken from (Jurado et al. 2014) and their use is further motivated in (Benigno et al. 2015), inter alia. Note that, as shown in Table 8, the macro factor may have a variance break in July 2008. Since the CM tests are not robust to changes in the variance of regressors, as shown in (Hansen 2000), we employ a simple correction to these CM tests in Tables 7, 10 and 12, by interacting the lags of the macro factor with the break in variance in July 2008, as recommended in (Pitarakis 2004). Similarly, all variance tests in Tables 9, 11 and 13 are not valid if a mean break was detected in the unconditional or conditional mean of that particular model used. Therefore, we follow (Pitarakis 2004) and correct the UV/CV tests for a potential break as follows. If for a particular model, a mean break is detected by the UM or the CM test, then the implicit break is imposed in the unconditional or conditional mean of the model (in the conditional mean, all regressors and the intercept are interacted with this break). From this model, the residuals are calculated, and the CV test is testing for a mean break in the squared residuals of the corresponding model.

**Table 7.** Structural breaks in the mean of the unemployment rate.

Moments/Models	Trimming	Statistic Value	Critical Value	Break Fraction	Break Date
Unconditional Mean sup Wald tests:					
$E(y_t)$	10%	9.585 *	9.11	0.892	Oct-08
	5%	9.585	9.71	0.892	-
Conditional Mean sup Wald tests:					
AR(1)	10%	11.408	12.17	0.419	-
	5%	15.494 *	12.80	0.929	Nov-10
AR(4)	10%	14.736	18.86	0.139	-
	5%	38.485 *	19.57	0.911	Dec-09
AR(5)—BIC	10%	18.205	20.81	0.121	-
	5%	36.777 *	21.53	0.928	Nov-10
AR(6)—AIC	10%	25.214 *	22.62	0.880	Apr-08
	5%	31.279 *	23.41	0.928	Nov-10
AR(12)	10%	30.343	32.76	0.879	-
	5%	56.090 *	33.63	0.949	Jan-12
DL(1) with macro factors—BIC	10%	13.935	16.91	0.451	-
	5%	18.612 *	17.54	0.949	Mar-09
DL(7) with macro factors—AIC	10%	19.665	37.43	0.619	-
	5%	25.653	38.35	0.934	-
DL(12) with uncertainty factors—AIC/BIC	10%	150.259 *	32.76	0.173	Mar-70
	5%	1826 *	33.63	0.941	Jan-09

Note: (a) Superscript \* means that the test rejects the null of no breaks at the 5% level; (b) DL( $p$ ) refers to a distributed lag model with  $p$  lags of a given factor; and (c) the tests with the macro factors are corrected for a potential break in variance as explained in Section 5.

<sup>26</sup> For one test in Table 10, the critical values are not available because this test entails 26 parameters, while critical values are available, to our knowledge, only for maximum 20 parameters. However, from (Andrews 2003), it is evident that the critical values are strictly increasing in the number of parameters, so it is reasonable to assume that the critical values for 26 parameters should be above the critical values for 20 parameters.

**Table 8.** Structural breaks in the regressors.

Moments	Test	Trimming	Statistic Value	Critical Value	Break Fraction	Break Date
Macro Factor						
$E(y_t)$	$UM_T^*$	10%	2.646	9.11	0.281	-
		5%	2.646	9.71	0.281	-
Macro Uncertainty Factor						
$E(y_t)$	$UM_T^*$	10%	5.214	9.11	0.898	-
		5%	6.295	9.71	0.921	-
Macro Factor						
$Var(y_t)$	$UV_T^*$	10%	4.277	9.11	0.898	-
		5%	10.181 *	9.71	0.935	Jul-08
Macro Uncertainty Factor						
$Var(y_t)$	$UV_T^*$	10%	4.969	9.11	0.898	-
		5%	8.317	9.71	0.928	-

Note: The variance tests are corrected for a mean break, if needed, as in (Pitarakis 2004).

**Table 9.** Structural breaks in the variance of the unemployment rate.

Moments/Models	Statistic Value	Critical Value	Break Fraction	Break Date
Unconditional Variance sup Wald tests:				
$Var(y_t)$	5.011	9.11	0.447	-
Conditional Variance sup Wald tests:				
AR(1)	15.278 *	9.11	0.476	Feb-86
AR(4)	11.948 *	9.11	0.471	Feb-86
AR(5)—BIC	13.024 *	9.11	0.469	Feb-86
AR(6)—AIC	17.660 *	9.11	0.468	Feb-86
AR(12)	14.753 *	9.11	0.457	Feb-86
DL(1) with macro factors—BIC	15.098 *	9.11	0.275	May-74
DL(7) with macro factors—AIC	15.044 *	9.11	0.259	Apr-74
DL(12) with uncertainty factors—AIC/BIC	6.375	9.11	0.415	-

Note: (a) A 10% trimming is used for all tests. Superscript \* means that the test rejects the null of no breaks at the 5% level; and (b) the variance tests are corrected for a mean break as explained in Section 5.

It is worth noting that the unconditional mean over the full-sample of our three macroeconomic series is significantly different from zero at the 1% level<sup>27</sup>, which implies that it is non-zero at least for some subsamples of the data, and we therefore expect the UM test to have power against structural change for at least. In addition, for all three series, the full sample coefficient estimates on higher order lags in AR( $p$ ) models are typically larger than the coefficients on the first two lags, so, from the simulation evidence in Section 4.2, we expect the conditional mean to be misspecified when choosing the number of lags with AIC or BIC.

<sup>27</sup> The means are: 6.11 for the unemployment rate, 17.73 for the interest rate, and 0.002 for the industrial production growth. For the power of the UM test, the means themselves are not important as long as they are non-zero, and as long as the t-statistics for these means (also known as signal-noise ratios) reject the null hypothesis of a zero mean.

**Table 10.** Structural breaks in the mean of the industrial production growth.

Moments/Models	Trimming	Statistic Value	Critical Value	Break Fraction	Break Date
Unconditional Mean sup Wald tests:					
$E(y_t)$	10%	3.191	9.11	0.252	-
	5%	3.191	9.71	0.252	-
Conditional Mean sup Wald tests:					
AR(1)	10%	13.949 *	12.17	0.401	Jan-82
	5%	28.948 *	12.80	0.933	Feb-11
AR(3)—BIC	10%	19.147 *	16.91	0.399	Jan-82
	5%	22.813 *	17.54	0.922	Jul-10
AR(4)	10%	25.432 *	18.86	0.398	Jan-82
	5%	25.432 *	19.57	0.398	Jan-82
AR(5)—AIC	10%	28.486 *	20.81	0.397	Jan-82
	5%	28.486 *	21.53	0.397	Jan-82
AR(12)	10%	35.846 *	32.76	0.392	Feb-82
	5%	37.517 *	33.63	0.924	Sep-10
DL(12) with macro factors—AIC/BIC	10%	32.600	>43.47	0.102	-
	5%	36.215	>44.46	0.060	-
DL(1) with uncertainty factors—BIC	10%	11.632	12.17	0.539	-
	5%	26.729 *	12.80	0.948	May-09
DL(5) with uncertainty factors—AIC	10%	17.761	20.81	0.254	-
	5%	20.115	21.53	0.935	-

Note: (a) Superscript \* means that the test rejects the null of no breaks at the 5% level; (b) DL( $p$ ) refers to a distributed lag model with  $p$  lags of a given factor; and (c) the test with the macro factors are corrected for a potential break in variance as explained in Section 5. In addition, note that, for the CM test for the DL(12) model with macro factors, there are 12 regressors and the intercept plus the same regressors interacted with the potential break in variance, adding up to 26 parameters. Critical values for more than 20 parameters are not available to our knowledge. However, the critical values in (Andrews 1993) strictly increase with the number of parameters, so it is plausible that the critical values with 26 parameters are strictly larger than the critical values for 20 parameters.

**Table 11.** Structural breaks in the variance of the industrial production growth.

Moments/Models	Statistic Value	Critical Value	Break Fraction	Break Date
Unconditional Variance sup Wald tests:				
$Var(y_t)$	5.535	9.11	0.437	-
Conditional Variance sup Wald tests:				
AR(1)	10.164 *	9.11	0.437	Jan-84
AR(3)—BIC	12.535 *	9.11	0.433	Jan-84
AR(4)	13.302 *	9.11	0.413	Jan-83
AR(5)—AIC	12.168 *	9.11	0.411	Jan-83
AR(12)	10.495 *	9.11	0.398	Jan-83
DL(12) with macro factors—AIC/BIC	8.267	9.11	0.442	-
DL(1) with uncertainty factors—BIC	13.318 *	9.11	0.455	Jan-84
DL(5) with uncertainty factors—AIC	10.938 *	9.11	0.448	Jan-84

Note: (a) A 10% trimming is used for all tests. Superscript \* means that the test rejects the null of no breaks at the 5% level, (b) The variance tests are corrected for a mean break, as explained in Section 5.

**Table 12.** Structural breaks in the mean of real interest rates.

Moments/Models	Trimming	Statistic Value	Critical Value	Break Fraction	Break Date
Unconditional Mean sup Wald tests:					
$E(y_t)$	10%	15.797 *	9.11	0.752	Apr-01
	5%	15.797*	9.71	0.752	Apr-01
Conditional Mean sup Wald tests:					
AR(1)	10%	57.077 *	12.17	0.893	Dec-08
	5%	57.077 *	12.80	0.893	Dec-08
AR(4)	10%	54.459 *	18.86	0.893	Dec-08
	5%	54.459 *	19.57	0.893	Dec-08
AR(9)—BIC	10%	125.086 *	27.77	0.892	Dec-08
	5%	125.086 *	28.64	0.892	Dec-08
AR(12)—AIC	10%	129.647 *	32.76	0.891	Dec-08
	5%	129.647 *	33.63	0.891	Dec-08
DL(1) with macro factors—AIC/BIC	10%	8.081	16.91	0.409	-
	5%	25.995 *	17.54	0.934	Jun-08
DL(1) with uncertainty factors—AIC/BIC	10%	850.945 *	12.17	0.791	Apr-01
	5%	1840 *	12.80	0.942	Jan-09

Note: (a) Superscript \* means that the test rejects the null of no breaks at the 5% level; (b) DL( $p$ ) refers to a distributed lag model with  $p$  lags of a given factor; and (c) the test with the macro factors are corrected for a potential break in variance as explained in Section 5.

**Table 13.** Structural breaks in the variance of real interest rates.

Moments/Models	Statistic Value	Critical Value	Break Fraction	Break Date
Unconditional Variance sup Wald tests:				
$Var(y_t)$	4.852	9.11	0.409	-
Conditional Variance sup Wald tests:				
AR(1)	21.591 *	9.11	0.410	Sep-82
AR(4)	24.721 *	9.11	0.407	Oct-82
AR(9)—BIC	20.799 *	9.11	0.394	Aug-82
AR(12)—AIC	23.787 *	9.11	0.388	Aug-82
DL(1) with macro factors—AIC/BIC	12.637 *	9.11	0.433	Aug-82
DL(1) with uncertainty factors—AIC/BIC	9.678 *	9.11	0.428	Aug-82

Note: (a) A 10% trimming is used for all tests. Superscript \* means that the test rejects the null of no breaks at the 5% level. (b) The variance tests are corrected for a mean break as explained in Section 5.

### 5.1. Unemployment Rate

Stock and Watson (2002) estimated AR(4) models for the quarterly difference in the US unemployment rate and found no break in the conditional mean, over a shorter period (1959–2001), but that is likely because the series was first differenced so a lot of the variation in the series has been removed. When analyzing the unemployment rate in levels, Table 7 shows that the evidence for breaks in this series is inconclusive. The UM test indicates a break in the recent crises when using a 10% cut-off, but not when using a 5% cut-off. Some of the CM tests also give mixed evidence: no break in the conditional (or short-run) mean at the 10% level, and one break in the recent crisis at the 5% level. The PACF in Figure 11 indicates spikes at lags 13 and 25 (also at higher lags), lags that are to our knowledge rarely used in applied work. Nevertheless, when rerunning the model selection with AIC and BIC and a maximum of 30 lags, we observe that the BIC does not change, but the AIC selects 25 lags. The CM tests on an AR(25) model are 51.508 at the 10% level, and 148.311 at the 5% level, so they reject in both cases the null of no break because they are above the critical values of 43.47 and 44.46, respectively (these are critical values for 20 parameters, and as explained in Footnote 23, they only provide a lower bound for the true critical values which are in this case not available). Nevertheless, these tests achieve their maximum value at the boundaries of the cut-off: 0.9 and 0.05.

Because at a 5% cut-off we only have 33 observations relative to 26 parameters to estimate, there are strong reasons to believe that these tests are plagued by numerical inaccuracies driven by too many parameters relative to the cut-off. Note that this numerical inaccuracy may also plague the CM test for the DL(12) model with uncertainty factor at the 5% level. Moreover, the uncertainty factor seems to be very persistent; an AR(1) model of the uncertainty factor yields a coefficient of 0.98 on the first lag. Because typically this factor is used in predictions in levels rather than first differences, we just report the CM tests for the models with the uncertainty factor in levels. However, note that the large values of the CM test can be explained by large distortions in the presence of long memory regressors, and therefore these results are also not to be fully trusted.

To summarize, it could be that CM tests are oversized and spuriously reject the null because of heavy underspecification of the true number of lags, an explanation in line with our simulations in Section 4.2. However, this problem cannot be remedied by including more lags because of numerical issues. When examining the UM test, it can also be that the HAC correction in the UM test is inaccurate for a large number of lags; it could explain why this test rejects the null hypothesis when using a 10% cut-off.

Next, we test for breaks in the variance of unemployment via the UV and CV tests; these results are reported in Table 9. The UV test and the CV test with 12 lags of the uncertainty factor used in the conditional mean does not reject the null of no breaks. In contrast, the CV tests based on AR( $p$ ) models show a structural break in the conditional variance of the unemployment in the mid 1980s, associated with the Great Moderation period. In Section 4.2, we showed that dynamic misspecification, especially underestimation of the number of lags, yields severely oversized CV tests even for  $T = 1000$ , while their power is comparable to the UV tests. This might explain the difference in results between the two tests. Alternatively, the results in Table 9 can be taken to suggest that although there is no break in the unconditional (long-run) variance of unemployment, there is a structural change in the conditional (short-run) variance of the unemployment dynamics, possibly related to the Great Moderation. It is worth mentioning that the mixed empirical evidence on volatility breaks in unemployment provided by the two tests is also found in other studies using quarterly data. Namely, while [Stock and Watson \(2002\)](#) report no evidence of breaks using the UV test for quarterly unemployment, [Sensier and van Dijk \(2004\)](#) find support for a Great Moderation volatility break when using an AR(4) model.

## 5.2. Industrial Production Growth

Table 10 reports the UM and CM tests for industrial production growth. Here, the UM test and the CM tests for DL models with lags of the uncertainty factor and 10% cut-off do not show evidence of breaks. Neither do the CM tests with lags of the macro factor. The other tests for AR( $p$ ) models do find breaks, but notice that as the number of lags increases, the CM tests become closer to the critical values, so the evidence for breaks is getting weaker. Moreover, the PACF in Figure 12 indicates a sizable negative spike at lag 24; it seems that underspecification of the number of lags (for example with AIC and BIC) yields to incorrect rejection of the null of no breaks, as shown in the simulation Section 4. This may also explain why [Stock and Watson \(2002\)](#) find a break in the conditional mean of the quarterly industrial production growth around the Great Moderation. To investigate this further, we reran the AIC and BIC model selection of an AR( $p$ ) model using a maximum of 30 lags; the BIC does not change, but the AIC now selects 18 lags. The corresponding CM test statistic is 137.039 with both 10% and 5% cutoffs and rejects the null of no breaks at the 5% level, but the implied break-fraction estimate is in both cases at 0.89, so near the cut-off, indicating that this test cannot be fully trusted as evidence of a mean break in the recent period, given its inflated value near the cut-off due to estimating many parameters with few observations.

As for breaks in the variance of the industrial production growth, the UV test and the CV test for a model with lags of the macro factor do not reject the null of no breaks, but the other CV tests indicate a break around the Great Moderation. This result is also found in [\(Stock and Watson 2002\)](#) for the quarterly industrial production growth, where the UV test does no reject, but the CV tests

reject. Note however that the CV tests starting with the one for the AR(4) model are decreasing when including more lags and getting very close to the critical values at 12 lags, indicating that the evidence for a Great Moderation break seems to diminish as more lags are taken into account.

### 5.3. Interest Rates

Several papers find breaks in the conditional mean of the short-term interest rates (see, e.g., [Garcia and Perron 1996](#); [Sensier and van Dijk 2004](#); [Stock and Watson 2002](#)). For example, [Garcia and Perron \(1996\)](#) employed a Markov switching model with three possible regimes in mean and variance for ex-post real interest rates. They find that there are two mean shifts, one in 1973 associated with the sudden rise in oil prices and another in 1981 which is in line with a federal budget deficit. Furthermore, [Rapach and Wohar \(2005\)](#) examined structural breaks in the mean real interest rate for 13 industrialized countries over the period 1960 till 1998. For the US series, they find breaks in the late 1960s, early 1970s and early 1980s.

Table 12 confirms these findings: it shows strong evidence of breaks in both the long-run and short-run mean of the interest rates, some around the Great Moderation and some around 2001, perhaps tied to the 11 September 2001 attack, which sent the stock market plummeting after several days in which it was closed.<sup>28</sup>

The UV test in Table 13 does not detect a break whilst the CV tests indicate at least one break, whose estimate is again around the Great Moderation. In this case, also because the CV test statistics seem to still be large at 12 lags, it is more likely that there is a variance break, and the UV test does not have enough power to detect it, which can happen for some data generating processes as indicated in Figure 2. We conclude that there is strong evidence of breaks in the unconditional mean and in the variance of short-term interest rates.

## 6. Conclusions

In this paper, we propose an alternative and complementary approach to the sup Wald test for breaks in the conditional mean and variance. We show that the corresponding unconditional mean and variance break tests exhibit comparable size and power properties. The unconditional mean and variance do not employ a conditional mean specification, so they do not suffer from potential regression model misspecification. We show that under certain commonly encountered forms of regression model misspecification, especially dynamic misspecification, the traditional conditional mean break tests suffer from severe oversizing, even for large sample sizes. Moreover, both tests for a break in mean have similar size-adjusted power as the sample size grows. In a comprehensive empirical analysis, we applied these tests to show that there is no clear evidence of long-run breaks in the mean of the unemployment rate or the industrial production growth. Similarly, the evidence for breaks in the long-run variance of the unemployment rate and the industrial production growth is mixed. It is worth noting that we only focus on the sup Wald test, which is among the most popular break point tests in empirical work. It would be interesting to repeat the analysis in this paper for other break-point tests, including more powerful tests, and we leave this for future research.

**Author Contributions:** Alaa Abi Morshed has done most of the coding and the simulations in Section 4.1. For the rest of the paper, all three authors have equal contributions.

**Funding:** The second author Elena Andreou would like to acknowledge the support of the Horizon 2020 ERC-PoC 2014 grant and of the UCY internal grant 2017–2018.

**Acknowledgments:** We are grateful to two anonymous referees, the seminar participants at the Tilburg Econometrics Seminars in 2014 and 2015 and to the conference participants at the NESG 2014 and IAAE 2015 conferences for useful comments.

<sup>28</sup> Note that, in this case, the AIC and BIC with a maximum number of 30 lags picks the same number of lags as when a maximum of 12 lags is imposed.



**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs of Theorems 1–3

**Proof of Theorem 1.** Part (i). Aue and Horvath (2012) define a CUSUM test for  $H_0^{UM}$  versus  $H_A^{UM}$  as follows:

$$\mathcal{Z}_T^* = \sup_{\lambda \in [\epsilon, 1-\epsilon]} \mathcal{Z}_T(\lambda), \quad \mathcal{Z}_T(\lambda) = \frac{1}{\sqrt{T}} \left( \sum_{t=1}^{[T\lambda]} y_t - \frac{[T\lambda]}{T} \sum_{t=1}^T y_t \right) / \hat{\mathbf{v}}_u^{1/2},$$

where  $\hat{\mathbf{v}}_u$  is a HAC consistent estimator of  $\mathbf{v}_u = \mathbf{AVar} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T y_t \right)$  under  $H_0^{UM}$  and Assumption 1(i). They state that if a functional central limit theorem (FCLT) holds under  $H_0^{UM}$  for  $\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\lambda]} u_t$ , then  $\mathcal{Z}_T(\lambda) \Rightarrow [B_1(\lambda) - \lambda B_1(1)] = \mathcal{B}_1(\lambda)$ , and so by the continuous mapping theorem (CMT),

$$\mathcal{Z}_T^* \Rightarrow \sup_{\lambda \in [\epsilon, 1-\epsilon]} \mathcal{B}_1(\lambda), \quad (\text{A1})$$

where  $\mathcal{B}_1(\lambda) = B_1(\lambda) - \lambda B_1(1)$  is a scalar independent Brownian bridge. Below, we show that there is a clear connection between the CUSUM and the UM test, so the asymptotic distribution of the second follows from the first.

$$\begin{aligned} T(\overline{y_{1\lambda}} - \overline{y_{2\lambda}})^2 &= T \left( \frac{1}{T_{1\lambda}} \sum_{t=1}^{T_{1\lambda}} y_t - \frac{1}{T_{2\lambda}} \sum_{t=1}^{T_{2\lambda}} y_t \right)^2 = \frac{T^3}{T_{1\lambda}^2 T_{2\lambda}^2} \left( \frac{T_{2\lambda}}{T} \sum_{t=1}^{T_{1\lambda}} y_t - \frac{T_{1\lambda}}{T} \sum_{t=1}^{T_{2\lambda}} y_t \right)^2 \\ &= \frac{T^3}{T_{1\lambda}^2 T_{2\lambda}^2} \left( \sum_{t=1}^{T_{1\lambda}} y_t - \frac{T_{1\lambda}}{T} \sum_{t=1}^T y_t \right)^2 = \left[ \frac{1}{\lambda^2(1-\lambda)^2} + o(1) \right] \left[ \frac{1}{\sqrt{T}} \left( \sum_{t=1}^{T_{1\lambda}} y_t - \frac{T_{1\lambda}}{T} \sum_{t=1}^T y_t \right) \right]^2 \\ &\Rightarrow \mathbf{v}_u [B_1(\lambda) - \lambda B_1(1)]^2 / [\lambda^2(1-\lambda)^2] = \mathbf{v}_u \mathcal{B}_1^2(\lambda) / [\lambda^2(1-\lambda)^2]. \end{aligned}$$

Since  $\mathbf{v}_{u\lambda} = \mathbf{AVar}(\sqrt{T}(\overline{y_{1\lambda}} - \overline{y_{2\lambda}})) = \mathbf{v}_u / [\lambda(1-\lambda)]$ ,  $UM_T(\lambda) \Rightarrow \mathcal{B}_1^2(\lambda) / [\lambda(1-\lambda)]$ , so:

$$UM_T^* \Rightarrow \sup_{\lambda \in [\epsilon, 1-\epsilon]} \mathcal{B}_1^2(\lambda) / [\lambda(1-\lambda)]. \quad (\text{A2})$$

Comparing (A1) and (A2), the two limiting distributions attain their supremum at different  $\lambda$ s, and thus the size of these tests will in general be different.<sup>29</sup> However, underlying the asymptotic theory is the same assumption, that the FCLT holds for  $T^{-1/2} \sum_{t=1}^{[T\lambda]} u_t$ . Assumption 1 guarantees that the FCLT in (Wooldridge and White 1988), Theorem 2.11, can be applied for  $u_t$  (in fact, we only need  $d_m = O(m^{-1/2})$ ), completing the proof of (i).

Part (ii). Here, we just verify Assumption 1 for  $|y_t - \bar{y}| - a$  instead of  $u_t$ . The rest of the proof is as in part (i) of the proof. Assumption 1(i) holds by the null hypothesis and by Assumption 2(i), and we are left to verify Assumption 1(ii). Since  $u_t$  is  $\mathcal{L}_2$ -near epoch dependent of size  $m^{-1/2}$  on  $\{g_t\}$  with positive constants equal to 1 (these constants appear in the near epoch dependent definition in Davidson (1994) but since here they are fixed, they are absorbed into the definition for  $d_m$ ), it follows that so is  $y_t - \bar{y}$ , with constants  $2 \sup_t(1) = 2$ . In Theorem 17.12 in Davidson (1994), let  $\phi_t(\cdot) = |\cdot|$ , a uniform Lipschitz function, with the argument  $y_t - \bar{y}$ . Then,  $y_t - \bar{y}$  is  $\mathcal{L}_2$ -near epoch dependent of size  $m^{-1/2}$ .

Part (iii). Here, we just verify Assumption 1(ii) for  $(y_t - \bar{y})^2 - v_u$ , instead of  $u_t$ , because Assumption 1(i) holds by the null hypothesis and Assumption 2(ii). In Theorem 17.12 in Davidson (1994), under  $H_0^{UV}$  and  $H_0^{UM}$ , define the function  $\phi_t(y_t - \bar{y}) = (y_t - \bar{y})^2 - v_u$ . From part (ii) of the

<sup>29</sup> In addition, note that the test statistic  $\sup_{\lambda \in [\epsilon, 1-\epsilon]} \sqrt{UM}_T$  is known in statistics as a “weighted version” of the CUSUM test—see (Aue and Horvath 2012, p. 5).

proof,  $(y_t - \bar{y})$  is a  $\mathcal{L}_2$ -near epoch dependent process of size  $m^{-1}$  on  $\{g_t\}$  with constants equal to 2. Below, we show that under  $H_0^{UM}$ ,  $\phi_t$  is uniform Lipschitz almost surely:

$$\begin{aligned} |\phi_t(y_t - \bar{y}) - \phi_t(y_k - \bar{y})| &= |(y_t - \bar{y})^2 - (y_k - \bar{y})^2| \\ &\leq |y_t + y_k - 2\bar{y}| |(y_t - \bar{y}) - (y_k - \bar{y})| \leq |u_t + u_k - 2\bar{u}| |(y_t - \bar{y}) - (y_k - \bar{y})| \\ &\leq (4 \sup_t |u_t|) |(y_t - \bar{y}) - (y_k - \bar{y})| \leq \kappa |(y_t - \bar{y}) - (y_k - \bar{y})|, \text{ almost surely,} \end{aligned}$$

for some  $\kappa > 0$ , by Assumption 1 for  $u_t$ , where  $\bar{u} = T^{-1} \sum_{t=1}^T u_t$ . Hence, by Theorem 17.12 in Davidson (1994),  $(y_t - \bar{y})^2 - v_u$  is also  $\mathcal{L}_2$ -near epoch dependent of size  $m^{-1/2}$ .  $\square$

**Proof of Theorem 2.** (i). Since Assumption 3 is a special case of Assumption 8 in (Hall et al. 2012), the result follows directly from their Theorem 6, setting  $x_t = z_t$ . (ii)–(iii). Primitive assumptions for CV test can be found in, e.g., Qu and Perron (2007), and involve joint mixing assumptions on  $\{x_t \epsilon_t\}$  and  $\epsilon_t^2$ . They mention that these conditions can be replaced by sufficient conditions to yield a FCLT for  $\{x_t \epsilon_t\}$  and  $\epsilon_t^2 - v_\epsilon$  under the null. By similar reasoning, for the CA test, sufficient conditions to yield a joint FCLT for  $\{x_t \epsilon_t\}$  and  $|\epsilon_t| - E|\epsilon_t|$  suffice. Since  $x_t$  includes an intercept, these conditions can be verified as for the proof of Theorem 1(ii)–(iii). Note that they all require  $H_0^{CM}$ .  $\square$

**Proof of Theorem 3.** Denote, for  $i = 1, 2$ ,  $\hat{Q}_{i\lambda, (1)} = T^{-1} \sum_{i\lambda} x_{t(1)} x'_{t(1)}$ ,  $\hat{Q}_{i\lambda, (12)} = T^{-1} \sum_{i\lambda} x_{t(1)} x'_{t(2)}$ ,  $\hat{Q}_{i\lambda, (2)} = T^{-1} \sum_{i\lambda} x_{t(2)} x'_{t(2)}$ , where recall that  $\Sigma_{1\lambda} = \sum_{t=1}^{[T\lambda]}$ , and  $\Sigma_{2\lambda} = \sum_{t=[T\lambda]+1}^T$ . By Assumption 3,  $\hat{Q}_{i\lambda, (1)} \xrightarrow{P} \lambda_i Q_{(1)}$ ,  $\hat{Q}_{i\lambda, (2)} \xrightarrow{P} \lambda_i Q_{(2)}$  and  $\hat{Q}_{i\lambda, (12)} \xrightarrow{P} \lambda_i Q_{(12)}$ , where  $i = 1, 2$ ,  $\lambda_1 = \lambda$  and  $\lambda_2 = 1 - \lambda_1$ . Recall that we mistakenly regress  $y_t$  only on  $x_{t(1)}$ ; let  $\hat{\theta}_{1\lambda}$  and  $\hat{\theta}_{2\lambda}$  be the OLS estimators in  $\{1, \dots, [T\lambda]\}$ , respectively  $\{[T\lambda] + 1, \dots, T\}$ .

$$\begin{aligned} \hat{\theta}_{1\lambda} &= \hat{Q}_{1\lambda, (1)}^{-1} \Sigma_{1\lambda} x_{t(1)} y_t = \theta_{(1)} + \hat{Q}_{1\lambda, (1)}^{-1} \hat{Q}_{1\lambda, (12)} \theta_{(2)} + \hat{Q}_{1\lambda, (1)}^{-1} T^{-1} \Sigma_{1\lambda} x_{t(1)} \epsilon_t \\ \hat{\theta}_{2\lambda} &= \hat{Q}_{2\lambda, (1)}^{-1} \Sigma_{1\lambda} x_{t(2)} y_t = \theta_{(1)} + \hat{Q}_{2\lambda, (1)}^{-1} \hat{Q}_{2\lambda, (12)} \theta_{(2)} + \hat{Q}_{2\lambda, (1)}^{-1} T^{-1} \Sigma_{2\lambda} x_{t(1)} \epsilon_t \\ T^{1/2}(\hat{\theta}_{1\lambda} - \theta_{(1)}) &= \hat{Q}_{1\lambda, (1)}^{-1} T^{1/2} \hat{Q}_{1\lambda, (12)} \theta_{(2)} + \hat{Q}_{1\lambda, (1)}^{-1} T^{-1/2} \Sigma_{1\lambda} x_{t(1)} \epsilon_t \\ T^{1/2}(\hat{\theta}_{2\lambda} - \theta_{(1)}) &= \hat{Q}_{2\lambda, (1)}^{-1} T^{1/2} \hat{Q}_{2\lambda, (12)} \theta_{(2)} + \hat{Q}_{2\lambda, (1)}^{-1} T^{-1/2} \Sigma_{2\lambda} x_{t(1)} \epsilon_t \\ T^{1/2}(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}) &= \hat{Q}_{1\lambda, (1)}^{-1} T^{-1/2} \Sigma_{1\lambda} [x_{t(1)} \epsilon_t + (x_{t(1)} x'_{t(2)} - Q_{(12)}) \theta_{(2)}] \\ &\quad - \hat{Q}_{2\lambda, (1)}^{-1} T^{-1/2} \Sigma_{2\lambda} [x_{t(1)} \epsilon_t + (x_{t(1)} x'_{t(2)} - Q_{(12)}) \theta_{(2)}] \\ &\quad + (\lambda \hat{Q}_{1\lambda, (1)}^{-1} - (1 - \lambda) \hat{Q}_{2\lambda, (1)}^{-1}) T^{1/2} Q_{(12)} \theta_{(2)} + o_P(1) \\ &= \hat{Q}_{1\lambda, (1)}^{-1} T^{-1/2} \Sigma_{1\lambda} (k_t + L_t \theta_{(2)}) - \hat{Q}_{2\lambda, (1)}^{-1} T^{-1/2} \Sigma_{2\lambda} (k_t + L_t \theta_{(2)}) + o_P(1) \\ &\quad - \lambda(1 - \lambda) T^{1/2} \hat{Q}_{1\lambda, (1)}^{-1} [\hat{Q}_{1\lambda, (1)} / \lambda - \hat{Q}_{2\lambda, (1)} / (1 - \lambda)] \hat{Q}_{2\lambda, (1)}^{-1} Q_{(12)} \theta_{(2)} + o_P(1) \\ &= \hat{Q}_{1\lambda, (1)}^{-1} T^{-1/2} \Sigma_{1\lambda} (k_t + L_t \theta_{(2)}) - \hat{Q}_{2\lambda, (1)}^{-1} T^{-1/2} \Sigma_{2\lambda} (k_t + L_t \theta_{(2)}) + o_P(1) \\ &\quad - \hat{Q}_{1\lambda, (1)}^{-1} T^{-1/2} [\Sigma_{1\lambda} x_{t(1)} x'_{t(1)} - \lambda \sum_{t=1}^T x_{t(1)} x'_{t(1)}] \hat{Q}_{2\lambda, (1)}^{-1} Q_{(12)} \theta_{(2)} + o_P(1) \\ &\equiv I + II - III. \\ I + II &= Q_{(1)}^{-1} T^{-1/2} [\Sigma_{1\lambda} [k_t L_t] / \lambda - \Sigma_{2\lambda} [k_t L_t] / (1 - \lambda)] \text{vec}(1, \theta_{(2)}) + o_P(1) \\ &= \frac{1}{\lambda(1-\lambda)} Q_{(1)}^{-1} T^{-1/2} \left[ \Sigma_{1\lambda} [k_t L_t] - \lambda \sum_{t=1}^T [k_t L_t] \right] \text{vec}(1, \theta_{(2)}) + o_P(1). \end{aligned}$$

Let  $\delta = Q_{(1)}^{-1}Q_{(12)}\theta_{(2)}$ , and note:

$$\begin{aligned} III &= \frac{1}{\lambda(1-\lambda)}Q_{(1)}^{-1}T^{-1/2}[\sum_{1\lambda} x_{t(1)}x'_{t(1)} - \lambda \sum_{t=1}^T x_{t(1)}x'_{t(1)}]Q_{(1)}^{-1}Q_{(12)}\theta_{(2)} + o_P(1) \\ &= \frac{1}{\lambda(1-\lambda)}Q_{(1)}^{-1}T^{-1/2}\left(\sum_{1\lambda} M_t - \lambda \sum_{t=1}^T M_t\right)\delta + o_P(1) \\ &= \frac{1}{\lambda(1-\lambda)}Q_{(1)}^{-1}T^{-1/2}\left(\sum_{1\lambda} M_t - \lambda \sum_{t=1}^T M_t\right)\delta + o_P(1). \end{aligned}$$

With  $s_t = [k_t \ L_t \ M_t]$ , we have:

$$T^{1/2}(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}) = \frac{1}{\lambda(1-\lambda)}Q_{(1)}^{-1}\left(\sum_{1\lambda} s_t - \lambda \sum_{t=1}^T s_t\right)\mathbf{vec}(1, \theta_{(2)}, -\delta) + o_P(1).$$

By Assumption 4 and the FCLT,  $T^{-1/2}\sum_{1\lambda} \mathbf{vec}(s_t) \Rightarrow \mathbf{H}^{*1/2}B_s^*(\lambda)$ , and so  $T^{-1/2}\sum_{1\lambda} s_t \Rightarrow \mathbf{H}^{*1/2}B_{mat}^*(\lambda)$ , where we denoted

$$B_{mat}^*(\lambda) = (B_{1:p_1}^*(\lambda), B_{p_1+1:2p_1}^*(\lambda), \dots, B_{p_1p_2+1:p_1(p_2+1)}^*(\lambda), B_{p_1(p_2+1)+1:p_1(p_2+1)+p_1}^*(\lambda), \dots, B_{p_1p+1:p_1(p+1)}^*(\lambda)),$$

so that  $\mathbf{vec}(B_{mat}^*(\lambda)) = B_s^*(\lambda)$ . Letting  $\xi = \mathbf{vec}(1, \theta_{(2)}, \delta)$ , we obtain:

$$\begin{aligned} T^{1/2}(\hat{\theta}_{1\lambda} - \hat{\theta}_{2\lambda}) &\Rightarrow Q_{(1)}^{-1}\mathbf{H}^{*1/2}[B_{mat}^*(\lambda)/\lambda - (B_{mat}^*(1) - B_{mat}^*(\lambda))/(1-\lambda)]\mathbf{vec}(1, \theta_{(2)}, \delta) \\ &= Q_{(1)}^{-1}\mathbf{H}^{*1/2}[B_{mat}^*(\lambda) - \lambda B_{mat}^*(1)]\xi/[\lambda(1-\lambda)] \\ &= Q_{(1)}^{-1}\mathbf{H}^{*1/2}\left(B_{1:p_1}^*(\lambda) + \sum_{i=1}^{p_2} B_{p_1i+1:p_1(i+1)}^*\theta_{i(2)} + \sum_{i=1}^{p_1} B_{p_1(p_2+1)+p_1(i-1)+1:p_1(p_2+1)+p_1i}^*(\lambda)\delta_i\right) \\ &= [Q_{(1)}^{-1} \otimes \xi']\mathbf{H}^{*1/2}B_{p_1(p+1)}^*, \end{aligned}$$

where  $\theta_{i(2)}, \delta_i$  are the  $i$ th elements of  $\theta_{(2)}$ , respectively  $\delta$ .

Part (i). Recall that  $\hat{V}_{i\lambda} = (\sum_{i\lambda} x_{t(1)}x'_{t(1)})^{-1}\hat{\Omega}_{i\lambda}(\sum_{i\lambda} x_{t(1)}x'_{t(1)})^{-1}$  and  $\hat{\Omega}_{i\lambda} = T^{-1}\sum_{i\lambda} \hat{\epsilon}_t^2 x_{t(1)}x'_{t(1)}$ , for  $i = 1, 2$ . Since  $\hat{\epsilon}_t = \epsilon_t - x'_{t(1)}(\hat{\theta}_{1\lambda} - \theta_{(1)}) + x'_{t(2)}\theta_{(2)}$ ,  $\hat{\epsilon}_t^2 = \epsilon_t^2 + (\hat{\theta}_{1\lambda} - \theta_{(1)})'x_{t(1)}x'_{t(1)}(\hat{\theta}_{1\lambda} - \theta_{(1)}) + \theta'_{(2)}x_{t(2)}x'_{t(2)}\theta_{(2)} - 2(\hat{\theta}_{1\lambda} - \theta_{(1)})'\epsilon_t x_{t(1)} + 2\theta'_{(2)}\epsilon_t x_{t(2)} - 2(\hat{\theta}_{1\lambda} - \theta_{(1)})'x_{t(1)}x'_{t(2)}\theta_{(2)}$  and so:

$$\begin{aligned} \hat{\Omega}_{1\lambda} &= T^{-1}\sum_{1\lambda} \hat{\epsilon}_t^2 x_{t(1)}x'_{t(1)} = T^{-1}\sum_{1\lambda} \epsilon_t^2 x_{t(1)}x'_{t(1)} + T^{-1}\sum_{1\lambda} [x'_{t(1)}(\hat{\theta}_{1\lambda} - \theta_{(1)})]^2 x_{t(1)}x'_{t(1)} \\ &\quad + T^{-1}\sum_{1\lambda} [\theta'_{(2)}x_{t(2)}]^2 x_{t(1)}x'_{t(1)} - 2T^{-1}\sum_{1\lambda} (\hat{\theta}_{1\lambda} - \theta_{(1)})'\epsilon_t x_{t(1)}x_{t(1)}x'_{t(1)} \\ &\quad + 2T^{-1}\sum_{1\lambda} \epsilon_t x'_{t(2)}\theta_{(2)}x_{t(1)}x'_{t(1)} \\ &\quad - 2T^{-1}\sum_{1\lambda} (\hat{\theta}_{1\lambda} - \theta_{(1)})'x_{t(1)}x'_{t(2)}\theta_{(2)}x_{t(1)}x'_{t(1)} \\ &= IV + V + VI - VII + VIII - IX. \end{aligned}$$

Partition  $\Omega^* = \begin{bmatrix} \Omega_{kk}^* & \Omega_{k\ell}^* & \Omega_{km}^* \\ \Omega_{k\ell}^* & \Omega_{\ell\ell}^* & \Omega_{\ell m}^* \\ \Omega_{km}^* & \Omega_{\ell m}^* & \Omega_{mm}^* \end{bmatrix}$ , such that  $\Omega_{kk}^*, \Omega_{\ell\ell}^*, \Omega_{mm}^*$  are  $p_1 \times p_1, (p_1p_2) \times (p_1p_2)$ , and

$p_1^2 \times p_1^2$  respectively. First,  $IV = T^{-1}\sum_{1\lambda} \epsilon_t^2 x_{t(1)}x'_{t(1)} = T^{-1}\sum_{1\lambda} k_t k'_t = \lambda\Omega_{kk}^* + o_P(1)$ . In addition, from the above,  $\hat{\theta}_{1\lambda} - \theta_{(1)} = \delta + o_P(1)$ , where  $\delta = Q_{(1)}^{-1}Q_{(12)}\theta_{(2)}$ . Thus, because of existence of fourth order moments of  $x_t$  by Assumption 4, it can be shown that:

$$\begin{aligned}
V &= T^{-1} \sum_{1\lambda} [x'_{t(1)} (\hat{\theta}_{1\lambda} - \theta_{(1)})]^2 x_{t(1)} x'_{t(1)} = T^{-1} \sum_{1\lambda} (x_{t(1)} x'_{t(1)} \delta) (\delta' x_{t(1)} x'_{t(1)}) + o_P(1) \\
&= T^{-1} \sum_{1\lambda} (x_{t(1)} x'_{t(1)} - Q_{(1)}) \delta \delta' (x_{t(1)} x'_{t(1)} - Q_{(1)}) + o_P(1) + Q_{(1)} \delta \delta' T^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(1)} \\
&\quad + T^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(1)} \delta \delta' Q_{(1)} - \lambda Q_{(1)} \delta \delta' Q_{(1)} + o_P(1) \\
&= T^{-1} \sum_{1\lambda} (x_{t(1)} x'_{t(1)} - Q_{(1)}) \delta \delta' (x_{t(1)} x'_{t(1)} - Q_{(1)}) + \lambda Q_{(1)} \delta \delta' Q_{(1)} + o_P(1) \\
&= T^{-1} \sum_{1\lambda} M_t \delta \delta' M'_t + \lambda Q_{(12)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} + o_P(1).
\end{aligned}$$

We have  $T^{-1} \sum_{1\lambda} \text{vec}(M_t) \text{vec}'(M'_t) \xrightarrow{P} \lambda \Omega_{mm}^*$  by Assumption 4. With  $m_{t,ij}, \ell_{t,ij}$  the  $(i, j)$ th element of  $M_t, L_t$  and  $\delta_i$  the  $i$ th element of  $\delta$ , we have:

$$\begin{aligned}
\delta' M'_t &= \text{vec}[\sum_{n=1}^{p_1} \delta_n m_{t,n1}, \dots, \sum_{n=1}^{p_1} \delta_n m_{t,np_1}] \\
&= \text{vec}[\delta' \{\text{vec}[M_t]\}_{1:p_1}, \dots, \delta' \{\text{vec}[M_t]\}_{(p_1^2-p_1+1):p_1^2}] = \text{vec}'[M_t][I_{p_1} \otimes \delta], \\
M_t \delta &= [\sum_{n=1}^{p_1} \delta_n M_{t,1n}, \dots, \sum_{n=1}^{p_1} \delta_n M_{t,p_1 n}] = [I_{p_1} \otimes \delta'] \text{vec}[M'_t], \\
L_t \theta_{(2)} &= [I_{p_1} \otimes \theta'_{(2)}] \text{vec}[L'_t], \\
\theta'_{(2)} L'_t &= \text{vec}'[L_t][I_{p_1} \otimes \theta_{(2)}].
\end{aligned}$$

It follows that:

$$\begin{aligned}
M_t \delta \delta' M'_t &= [I_{p_2} \otimes \delta'] \text{vec}[M'_t] \text{vec}'[M_t][I_{p_1} \otimes \delta], \\
V &= T^{-1} \sum_{1\lambda} M_t \delta \delta' M'_t + \lambda Q_{(12)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} + o_P(1) \\
&= \lambda [I_{p_1} \otimes \delta'] \Omega_{mm}^* [I_{p_1} \otimes \delta] + \lambda Q_{(12)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} + o_P(1).
\end{aligned}$$

Similarly, it follows that:

$$\begin{aligned}
VI &= T^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(2)} \theta_{(2)} \theta'_{(2)} x_{t(2)} x'_{t(1)} = T^{-1} \sum_{1\lambda} (x_{t(1)} x'_{t(2)} - Q_{(12)}) \theta_{(2)} \theta'_{(2)} (x_{t(2)} x'_{t(1)} - Q'_{(12)}) \\
&\quad + Q_{(12)} T^{-1} \sum_{1\lambda} \theta_{(2)} \theta'_{(2)} x_{t(2)} x'_{t(1)} + T^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(2)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} - \lambda Q_{(12)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} \\
&= T^{-1} \sum_{1\lambda} L_t \theta_{(2)} \theta'_{(2)} L'_t + \lambda Q_{(12)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} + o_P(1) \\
&= \lambda [I_{p_1} \otimes \theta'_{(2)}] \Omega_{\ell\ell}^* [I_{p_1} \otimes \theta_{(2)}] + \lambda Q_{(12)} \theta_{(2)} \theta'_{(2)} Q'_{(12)} + o_P(1).
\end{aligned}$$

In addition,

$$\begin{aligned}
VII &= 2T^{-1} \sum_{1\lambda} (\hat{\theta}_{1\lambda} - \theta_{(1)})' \epsilon_t x_{t(1)} x'_{t(1)} = 2T^{-1} \sum_{1\lambda} (\delta' x_{t(1)} \epsilon_t) x_{t(1)} x'_{t(1)} + o_P(1) \\
&= 2T^{-1} \sum_{1\lambda} (\delta' x_{t(1)} \epsilon_t) (x_{t(1)} x'_{t(1)} - Q_{(1)}) + 2T^{-1} \sum_{1\lambda} (\delta' x_{t(1)} \epsilon_t) Q_{(1)} + o_P(1) \\
&= 2T^{-1} \sum_{1\lambda} x_{t(1)} \epsilon_t \delta' (x_{t(1)} x'_{t(1)} - Q_{(1)}) + o_P(1) = 2T^{-1} \sum_{1\lambda} k_t \delta' M'_t + o_P(1) \\
&= T^{-1} \sum_{1\lambda} k_t \delta' M'_t + T^{-1} \sum_{1\lambda} \delta' M'_t k_t + o_P(1) = \lambda \Omega_{km}^* [I_{p_1} \otimes \delta] + \lambda [I_{p_1} \otimes \delta'] \Omega_{km}^* + o_P(1),
\end{aligned}$$

$$\begin{aligned}
VIII &= 2T^{-1} \sum_{1\lambda} \theta'_{(2)} \epsilon_t x_{t(2)} x_{t(1)} x'_{t(1)} = 2T^{-1} \sum_{1\lambda} x_{t(1)} \epsilon_t \theta'_{(2)} x_{t(2)} x'_{t(1)} \\
&= 2T^{-1} \sum_{1\lambda} x_{t(1)} \epsilon_t \theta'_{(2)} (x_{t(2)} x'_{t(1)} - Q'_{(12)}) + o_P(1) = 2T^{-1} \sum_{1\lambda} k_t \theta'_{(2)} L'_t + o_P(1) = \\
&= \lambda \Omega_{k\ell}^* [I_{p_1} \otimes \theta_{(2)}] + \lambda [I_{p_1} \otimes \theta'_{(2)}] \Omega_{k\ell}^{*'} + o_P(1), \\
IX &= 2T^{-1} \sum_{1\lambda} (\hat{\theta}_{1\lambda} - \theta_{(1)})' x_{t(1)} x'_{t(2)} \theta_{(2)} x_{t(1)} x'_{t(1)} = 2T^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(1)} \delta \theta'_{(2)} x_{t(2)} x'_{t(1)} + o_P(1) \\
&= 2T^{-1} \sum_{1\lambda} M_t \delta \theta'_{(2)} L'_t + 2Q_{(1)} \delta \theta'_{(2)} T^{-1} \sum_{1\lambda} x_{t(2)} x'_{t(1)} + 2T^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(1)} \delta \theta'_{(2)} Q'_{(12)} \\
&\quad - 2\lambda Q_{(1)} \delta \theta'_{(2)} Q'_{(12)} + o_P(1) \\
&= \lambda [I_{p_1} \otimes \delta'] \Omega_{\ell m}^{*'} [I_{p_1} \otimes \theta_{(2)}] + \lambda [I_{p_1} \otimes \theta'_{(2)}] \Omega_{\ell m}^* [I_{p_1} \otimes \delta] + 2\lambda Q_{(1)} \delta \theta'_{(2)} Q'_{(12)} + o_P(1).
\end{aligned}$$

Putting all of the above together,

$$\begin{aligned}
\hat{\Omega}_{1\lambda}^* &= \lambda \left\{ \Omega_{kk}^* + [I_{p_1} \otimes \delta'] \Omega_{mm}^* [I_{p_1} \otimes \delta] + [I_{p_1} \otimes \theta'_{(2)}] \Omega_{\ell\ell}^* [I_{p_1} \otimes \theta_{(2)}] - \Omega_{km}^* [I_{p_1} \otimes \delta] - [I_{p_1} \otimes \delta'] \Omega_{km}^{*'} \right. \\
&\quad \left. + \Omega_{k\ell}^* [I_{p_1} \otimes \theta_{(2)}] + [I_{p_1} \otimes \theta'_{(2)}] \Omega_{k\ell}^{*'} - [I_{p_1} \otimes \delta'] \Omega_{\ell m}^{*'} [I_{p_1} \otimes \theta_{(2)}] - [I_{p_1} \otimes \theta'_{(2)}] \Omega_{\ell m}^* [I_{p_1} \otimes \delta] \right\} + o_P(1) \\
&= \lambda \left\{ [I_{p_1} \otimes \xi'] \Omega^* [I_{p_1} \otimes \xi] \right\} + o_P(1) \\
\hat{V}_{1\lambda} &= \frac{1}{\lambda} Q_{(1)}^{-1} \left\{ [I_{p_1} \otimes \xi'] \Omega^* [I_{p_1} \otimes \xi] \right\} Q_{(1)}^{-1} + o_P(1) = \frac{1}{\lambda} \left\{ [Q_{(1)}^{-1} \otimes \xi'] \Omega^* [Q_{(1)}^{-1} \otimes \xi] \right\} + o_P(1) \\
\hat{V}_{2\lambda} &= \frac{1}{1-\lambda} \left\{ [Q_{(1)}^{-1} \otimes \xi'] \Omega^* [Q_{(1)}^{-1} \otimes \xi] \right\} + o_P(1) \\
\hat{V}_\lambda &= \frac{1}{\lambda(1-\lambda)} \left\{ [Q_{(1)}^{-1} \otimes \xi'] \Omega^* [Q_{(1)}^{-1} \otimes \xi] \right\} + o_P(1).
\end{aligned}$$

Hence,

$$CM_T^* \Rightarrow \sup_\lambda \left\{ \frac{1}{\lambda(1-\lambda)} \mathcal{B}_{p_1(p+1)}^{*'}(\lambda) A \mathcal{B}_{p_1(p+1)}^*(\lambda) \right\},$$

with  $A = \mathbf{H}^{*1/2'} [Q_{(1)}^{-1} \otimes \xi] \left\{ [Q_{(1)}^{-1} \otimes \xi'] \Omega^* [Q_{(1)}^{-1} \otimes \xi] \right\}^{-1} [Q_{(1)}^{-1} \otimes \xi'] \mathbf{H}^{*1/2}$ .

Part (ii). In this case,

$$\begin{aligned}
\hat{v}_{\epsilon,1\lambda} &= T_{1\lambda}^{-1} \sum_{1\lambda} \epsilon_t^2 + (\hat{\theta}_{1\lambda} - \theta_{(1)})' T_{1\lambda}^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(1)} (\hat{\theta}_{1\lambda} - \theta_{(1)}) + \theta'_{(2)} T_{1\lambda}^{-1} \sum_{1\lambda} x_{t(2)} x'_{t(2)} \theta_{(2)} \\
&\quad - 2(\hat{\theta}_{1\lambda} - \theta_{(1)})' T_{1\lambda}^{-1} \sum_{1\lambda} \epsilon_t x_{t(1)} + 2\theta'_{(2)} T_{1\lambda}^{-1} \sum_{1\lambda} \epsilon_t x_{t(2)} - 2(\hat{\theta}_{1\lambda} - \theta_{(1)})' T_{1\lambda}^{-1} \sum_{1\lambda} x_{t(1)} x'_{t(2)} \theta_{(2)} \\
&= \sigma_\epsilon^2 + \delta' Q_{(1)} \delta + \theta'_{(2)} Q_{(2)} \theta_{(2)} - 2\delta' Q_{(12)} \theta_{(2)} + o_P(1) \\
&= \sigma_\epsilon^2 + \theta'_{(2)} Q'_{(12)} Q_{(1)}^{-1} Q_{(12)} \theta_{(2)} + \theta'_{(2)} Q_{(2)} \theta_{(2)} - 2\theta'_{(2)} Q'_{(12)} Q_{(1)}^{-1} Q_{(12)} \theta_{(2)} + o_P(1) \\
&= \sigma_\epsilon^2 - \theta'_{(2)} Q'_{(12)} Q_{(1)}^{-1} Q_{(12)} \theta_{(2)} + \theta'_{(2)} Q_{(2)} \theta_{(2)} + o_P(1) \\
&= \lambda \sigma_\epsilon^2 + \theta'_{(2)} [Q_{(2)} - Q'_{(12)} Q_{(1)}^{-1} Q_{(12)}] \theta_{(2)} + o_P(1) \\
\hat{V}_{1\lambda} &= \left\{ \sigma_\epsilon^2 + \theta'_{(2)} [Q_{(2)} - Q'_{(12)} Q_{(1)}^{-1} Q_{(12)}] \theta_{(2)} \right\} Q_{(1)}^{-1} / \lambda + o_P(1) \\
\hat{V}_{2\lambda} &= \left\{ \sigma_\epsilon^2 + \theta'_{(2)} [Q_{(2)} - Q'_{(12)} Q_{(1)}^{-1} Q_{(12)}] \theta_{(2)} \right\} Q_{(1)}^{-1} / (1-\lambda) + o_P(1) \\
\hat{V}_\lambda &= \frac{1}{\lambda(1-\lambda)} \left\{ \sigma_\epsilon^2 + \theta'_{(2)} [Q_{(2)} - Q'_{(12)} Q_{(1)}^{-1} Q_{(12)}] \theta_{(2)} \right\} Q_{(1)}^{-1} + o_P(1) = \frac{1}{\lambda(1-\lambda)} Q_{(1)}^{-1} + o_P(1).
\end{aligned}$$

Thus,  $CM_T^*$  weakly converges to:

$$= \sup_\lambda \left\{ \frac{1}{\nu\lambda(1-\lambda)} \mathcal{B}_{p_1(p+1)}^{*'}(\lambda) \mathbf{H}^{*1/2'} \{Q_{(1)}^{-1} \otimes (\xi \xi')\} \mathbf{H}^{*1/2} \mathcal{B}_{p_1(p+1)}^*(\lambda) \right\}.$$

□

## References

- Altansukh, Gantungalag. 2013. International Inflation Linkages and Forecasting in the Presence of Structural Breaks. PhD thesis. University of Manchester, Manchester, UK. Available online: [https://www.research.manchester.ac.uk/portal/files/54550450/FULL\\_TEXT.PDF](https://www.research.manchester.ac.uk/portal/files/54550450/FULL_TEXT.PDF) (accessed on 1 May 2015).
- Andreou, Elena, and Eric Ghysels. 2002. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17: 579–600. [CrossRef]
- Andrews, Donald W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–58. [CrossRef]
- Andrews, Donald W. K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61: 821–56. [CrossRef]
- Andrews, Donald W. K. 2003. Tests of parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* 71: 395–97. [CrossRef]
- Andrews, Donald W. K., and Werner Ploberger. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62: 1383–414. [CrossRef]
- Aue, Alexander, and Lajos Horvath. 2012. Structural breaks in time series. *Journal of Time Series Analysis* 34: 1–16. [CrossRef]
- Bai, Jushan, Haiqiang Chen, Terence Tai-Leung Chong, and Seraph Xin Wang. 2008. Generic consistency of the break-point estimators under specification errors in a multiple-break model. *Econometrics Journal* 11: 287–307. [CrossRef]
- Bai, Jushan, and Pierre Perron. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66: 47–78. [CrossRef]
- Bataa, Erdenebat, Denise Osborn, Marianne Sensier, and Dick van Dijk. 2013. Structural breaks in the international dynamics of inflation. *Review of Economics and Statistics* 95: 646–59. [CrossRef]
- Benigno, Pierpaolo, Luca Antonio Ricci, and Paolo Surico. 2015. Unemployment and productivity in the long run: The role of macroeconomic volatility. *The Review of Economics and Statistics* 97: 698–709. [CrossRef]
- Cho, Cheol-Keun, and Timothy J. Vogelsang. 2014. Fixed-b Inference for Testing Structural Change in a Time Series Regression. Working Paper. Available online: <https://msu.edu/~tjv/cho-vogelsang.pdf> (accessed on 1 May 2016).
- Chong, Terence Tai-Leung. 2003. Generic consistency of the break-point estimator under specification errors. *Econometrics Journal* 6: 167–92. [CrossRef]
- Davidson, James. 1994. *Stochastic Limit Theory*. Oxford: Oxford University Press.
- Diebold, Francis X., and Atsushi Inoue. 2001. Long memory and regime switching. *Journal of Econometrics* 105: 131–59. [CrossRef]
- Elliott, Graham, and Ulrich K. Müller. 2007. Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141: 1196–218. [CrossRef]
- Eo, Junjong, and James Morley. 2015. Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics* 6: 463–97. [CrossRef]
- Garcia, Rene, and Pierre Perron. 1996. An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics* 78: 111–25. [CrossRef]
- Hall, Alastair R., Sanggohn Han, and Otilia Boldea. 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics* 170: 281–302. [CrossRef] [PubMed]
- Hansen, Bruce E. 1991. GARCH(1,1) processes are near epoch dependent. *Economics Letters* 36: 181–86. [CrossRef]
- Hansen, Bruce E. 2000. Testing for structural change in conditional models. *Journal of Econometrics* 97: 93–115. [CrossRef]
- Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng. 2014. Measuring uncertainty. *American Economic Review* 105: 1177–216. [CrossRef]
- Kejriwal, Mohitosh. 2009. Tests for a mean shift with good size and monotonic power. *Economic Letters* 102: 78–82. [CrossRef]
- Kurozumi, Eiji, and Purevdorj Tuvaandorj. 2011. Model selection criteria in multivariate models with multiple structural changes. *Journal of Econometrics* 164: 218–38. [CrossRef]
- McConnell, Margaret M., and Gabriel Perez-Quiros. 2000. Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review* 90: 1464–76. [CrossRef]

- McKittrick, Ross R., and Timothy J. Vogelsang. 2014. Hac robust trend comparisons among climate series with possible level shifts. *Environmetrics* 25: 528–47. [\[CrossRef\]](#)
- Müller, Ulrich K., and Mark W. Watson. 2017. Low-frequency econometrics. In *Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society*. Edited by Bo Honoré, Ariel Pakes, Monika Piazzesi and Larry Samuelson. Cambridge: Cambridge University Press, Volume II: 53–94.
- Newey, Whitney K., and Kenneth D. West. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61: 631–53. [\[CrossRef\]](#)
- Perron, Pierre. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57: 1361–401. [\[CrossRef\]](#)
- Perron, Pierre, and Tomoyoshi Yabu. 2009. Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business and Economic Statistics* 27: 369–96. [\[CrossRef\]](#)
- Pitarakis, Jean-Yves. 2004. Least squares estimation and tests of breaks in mean and variance under misspecification. *Econometrics Journal* 7: 32–54. [\[CrossRef\]](#)
- Ploberger, Werner, and Walter Krämer. 1992. The CUSUM test with OLS residuals. *Econometrica* 60: 271–85. [\[CrossRef\]](#)
- Qu, Zhongjun, and Pierre Perron. 2007. Estimating and testing structural changes in multivariate regressions. *Econometrica* 75: 459–502. [\[CrossRef\]](#)
- Rapach, David E., and Mark E. Wohar. 2005. Regime changes in international real interest rates: Are they a monetary phenomenon? *Journal of Money, Credit and Banking* 37: 887–906. [\[CrossRef\]](#)
- Sayginsoy, Özgen, and Timothy J. Vogelsang. 2011. Testing for a shift in trend at an unknown date: A fixed-b analysis of heteroskedasticity autocorrelation robust OLS-based tests. *Econometric Theory* 27: 992–1025. [\[CrossRef\]](#)
- Sensier, Marianne, and Dick van Dijk. 2004. Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics* 86: 833–39. [\[CrossRef\]](#)
- Smallwood, Aaron D. 2016. A Monte Carlo investigation of unit root tests and long memory in detecting mean reversion in  $i(0)$  regime switching, structural break, and nonlinear data. *Econometric Reviews* 35: 986–1012. [\[CrossRef\]](#)
- Stock, James H., and Mark W. Watson. 2002. Has the business cycle changed and why? *NBER Macroeconomics Annual* 17: 159–218. [\[CrossRef\]](#)
- Vogelsang, Timothy J. 1997. Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory* 13: 818–48. [\[CrossRef\]](#)
- Vogelsang, Timothy J. 1998. Trend function hypothesis testing in the presence of serial correlation. *Econometrica* 66: 123–48. [\[CrossRef\]](#)
- Vogelsang, Timothy J. 1999. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics* 88: 283–99. [\[CrossRef\]](#)
- Vogelsang, Timothy J., and Pierre Perron. 1998. Additional tests for a unit root allowing for a break in the trend function at an unknown time. *International Economic Review* 39: 1073–1100. [\[CrossRef\]](#)
- Wooldridge, Jeffrey M., and Halbert White. 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4: 210–30. [\[CrossRef\]](#)

