

NBER WORKING PAPER SERIES

TESTING FOR PARAMETER INSTABILITY AND STRUCTURAL CHANGE IN  
PERSISTENT PREDICTIVE REGRESSIONS<sup>^</sup>

Torben G. Andersen  
Rasmus T. Varneskov

Working Paper 28570  
<http://www.nber.org/papers/w28570>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
March 2021

We thank Atsushi Inoue for helpful comments and suggestions. This manuscript subsumes a previous working paper entitled "On the Informational Efficiency of Option-Implied and Time Series Forecasts of Realized Volatility." Financial support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation, is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Torben G. Andersen and Rasmus T. Varneskov. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Testing for Parameter Instability and Structural Change in Persistent Predictive Regressions<sup>^</sup>  
Torben G. Andersen and Rasmus T. Varneskov  
NBER Working Paper No. 28570  
March 2021  
JEL No. G12,G17

### **ABSTRACT**

This paper develops parameter instability and structural change tests within predictive regressions for economic systems governed by persistent vector autoregressive dynamics. Specifically, in a setting where all – or a subset – of the variables may be fractionally integrated and the predictive relation may feature cointegration, we provide sup-Wald break tests that are constructed using the Local speCtrum (LCM) approach. The new tests cover both parameter variation and multiple structural changes with unknown break dates, and the number of breaks being known or unknown. We establish asymptotic limit theory for the tests, showing that it coincides with standard testing procedures. As a consequence, existing critical values for tied-down Bessel processes may be applied, without modification. We implement the new structural change tests to explore the stability of the fractionally cointegrating relation between implied- and realized volatility (IV and RV). Moreover, we assess the relative efficiency of IV forecasts against a challenging time-series benchmark constructed from high-frequency data. Unlike existing studies, we find evidence that the IV-RV cointegrating relation is unstable, and that carefully constructed time-series forecasts are more efficient than IV in capturing low-frequency movements in RV.

Torben G. Andersen  
Kellogg School of Management  
Northwestern University  
2001 Sheridan Road  
Evanston, IL 60208  
and NBER  
t-andersen@kellogg.northwestern.edu

Rasmus T. Varneskov  
Copenhagen Business School  
Department of Finance  
Solberg Plads 3  
2000 Frederiksberg  
Denmark  
rtv.fi@cbs.dk

# 1 Introduction

Predictive regressions are routinely invoked to assess the forecast prowess of candidate predictors for the future value of an economic quantity of interest. This approach provides great flexibility in accommodating multiple predictors, and it affords direct and transparent interpretability of the findings. Under standard regularity conditions, robust inference techniques may be applied to test for the significance of an individual regressor or the joint significance of (a subset of) the regressors.

The use of predictive regressions and least squares inference in time series contexts is, however, subject to controversy due to the inherent nature of economic and financial data, which has spurred a large literature dealing with potential pitfalls. One major issue is the strong persistence displayed by many economic time series. In particular, the spurious regression case, where one  $I(1)$  process is projected onto another independent  $I(1)$  process, generates large size distortions in standard least squares significance tests and artificially inflates model-fit measures such as the  $R^2$  statistic; see e.g., Granger & Newbold (1974) and Phillips (1986). Similar problems arise for nearly integrated systems, e.g., Ng & Perron (1997) and Valkanov (2003), or if the variables are fractionally integrated,  $I(d)$ , where  $d$  is possibly a non-integer, see Tsay & Chung (2000). Moreover, the regressions may be subject to sizable predictive biases if the system features endogenous correlation between the regression errors and innovations to the predictor variables, e.g., Stambaugh (1999) and Phillips & Lee (2013).

A second major complication is potential instability of the predictive regression relation. The evidence for the structural breaks in the data generating process is extensive within economics and finance; see, e.g., Stock & Watson (1996), Paye & Timmermann (2006), Peseran, Pettenuzzo & Timmermann (2006), Chen & Hong (2012), Dangl & Hailing (2012) and Farmer, Schmidt & Timmermann (2019). This issue is particularly salient when estimating predictive relations using data spanning multiple decades, and sometimes even centuries, or if using cointegration models to generate long-horizon forecasts, relying critically on accurate identification of the long-run equilibrium, see, e.g., Christoffersen & Diebold (1998). Since predictive regressions are reduced-form specifications, it is therefore not surprising that they evolve along with the economic environment over long time spans.

Several studies develop inference techniques for systems with persistent variables. For example, fractional (and standard) cointegration frameworks facilitate inference on linear predictive relations, whose error is purged of (at least some of) the persistence in the original variables; see, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006), and Johansen & Nielsen (2012) for parametric and semi-parametric procedures. Moreover, robust techniques have been proposed for the nearly integrated case by, among others, Cavanagh, Elliott & Stock (1995), Jansson & Moreira (2006), and Elliott, Müller & Watson (2015). These procedures are reviewed by Phillips & Lee (2013), who develop the IVX methodology, allowing for stationary, nearly integrated, unit root and locally explosive processes. Finally, Andersen & Varneskov (2020a) provides a robust procedure for systems where all variables may be fractionally integrated of potentially different orders, covering both (asymptotically) stationary and non-stationary processes, and where the predictive relation may feature fractional cointegration, with a priori unknown persistence and cointegration properties. Furthermore, their Local speCtrUM

(LCM) approach accommodates regressor endogeneity and achieves asymptotic Gaussian inference.

Issues related to parameter instability and structural change have similarly received considerable attention. Andrews (1993), Andrews & Ploberger (1994) and Diebold & Chen (1996) develop tests for a structural change assuming that the break point is unknown, and Bai & Perron (1998) provide a comprehensive treatment for multiple structural changes with the timing, and possibly also the number, of breaks allowed to be unknown. These studies, however, require the system to be weakly dependent. In contrast, Hansen (1992), Hansen (2003) and Kerjiwal & Perron (2010), among others, develop similar structural change tests in  $I(1)$ - $I(0)$  cointegration models that also allow for  $I(0)$  regressors. Moreover, Hidalgo & Robinson (1996) extend the sup-Wald test by Andrews (1993) to systems with stationary fractional integration, and Georgiev, Harvey, Leybourne & Taylor (2018) consider nearly integrated regressors. No prior study, however, has provided tests in the flexible persistence setting of Andersen & Varneskov (2020a), that is, when the system can feature a mixture between stationary and non-stationary fractionally integrated variables and (fractional) cointegration, which is not known prior to the analysis. In fact, as discussed in the reviews by Perron (2006) and Casini & Perron (2019), existing procedures cannot simultaneously accommodate  $I(0)$  and  $I(1)$  errors, because the latter induce spurious inference. In this paper, we tackle these problems.

Specifically, we provide new tests for parameter instability and structural change, applicable for economic systems in which the variables (or a subset thereof) may be fractionally integrated of potentially different orders, and the predictive relation may feature cointegration, thereby nesting standard weakly dependent as well as cointegration settings. Our tests connect the LCM approach with the sup-Wald testing framework by Andrews (1993) and Bai & Perron (1998). Hence, we provide tests for structural changes with unknown break dates, where the number of breaks may be either known or unknown ex ante. These tests are generally applicable as parameter instability tests, despite being designed for the detection of structural breaks. We establish an asymptotic limit theory for the new tests. Importantly, by leveraging the asymptotic Gaussian distribution theory obtained via the LCM approach, we show that our tests, despite allowing for diverse and highly persistent systems, attain identical asymptotic distributions as the corresponding tests in Andrews (1993) and Bai & Perron (1998), who consider weakly dependent settings without cointegration. Hence, their critical values for tied-down Bessel processes may readily be applied in our context. Interestingly, due to the robustness properties of the LCM procedure, we also accommodate regressor endogeneity in the testing framework. Consequently, the new tests may be used as “off-line” diagnostics for a wide array of predictive systems, when one wants to retrospectively examine whether breaks occur in a given sample.<sup>1</sup>

We illustrate the finite sample properties of our LCM-based sup-Wald testing procedure in settings that accommodate flexible, fractionally integrated persistence in the predictive relation, both in the absence and presence of cointegration. Despite the semi-parametric nature of our approach, the test

---

<sup>1</sup>In contrast, if one wants to construct forecasts in real time, two alternatives include using rolling windows to implement the predictive regressions, e.g., Pesaran & Timmermann (2007) and Clark & McCracken (2009), or the explicit modeling of time-variation in the parameters, e.g., Peseran et al. (2006) and Dangl & Hailing (2012). None of their results, however, pertain to our flexible setting with fractionally integrated variables and, possibly, cointegration.

has excellent finite sample size and power properties for sample sizes typically encountered in financial and macroeconomic applications. Moreover, it is robust towards tuning parameter selections.

We use our new sup-Wald testing procedure to study the stability of the predictive relation between option implied and realized volatility (IV and RV), which has received considerable attention in empirical finance and financial econometrics. Specifically, using predictive regressions, this literature has been concerned with examining the bias and efficiency of IV as a predictor of future RV. After correcting for various methodological issues, consensus is converging towards the conclusion that IV forecasts are biased, but efficient predictors of future volatility; see, e.g., Christensen & Prabhala (1998), Jiang & Tian (2005), Chernov (2007), Chraoenwong, Jenwittayaroje & Low (2009), Taylor, Yadav & Zhang (2010) and Cheng & Fung (2012). These studies, however, all rely on inference procedure for weakly dependent processes, despite comprehensive evidence that return RV exhibits long memory; e.g., Andersen & Bollerslev (1997), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001, 2003), Corsi (2009) and Varneskov & Perron (2018).

The evidence of long-memory in IV and RV has inspired a number of studies to assess whether their predictive relation is fractionally cointegrated and whether IV is “long-run unbiased” for the persistent variation in RV. In particular, Bandi & Perron (2006), Christensen & Nielsen (2006), Nielsen (2007), Kellard, Dunis & Sarantis (2010) and Nielsen & Frederiksen (2011) provide solid support for these hypotheses. In addition, Bollerslev, Sizova & Tauchen (2011) and Bollerslev, Osterrieder, Sizova & Tauchen (2013), Osterrieder, Ventosa-Santaularia & Vera-Valdes (2019), Li, Izzeldin & Yao (2020) confirm the findings of fractional cointegration and proceed to exploit the equilibrium relation between IV and RV to estimate (components of) the volatility risk premium and forecast asset returns.

We revisit the long-run relation between IV and RV by first examining the relative efficiency of the long-run information in IV for RV. We compare it to forecasts from a long-memory volatility model exploiting high-frequency data, thereby providing a more challenging benchmark than historical volatility, which typically is used in the literature. Second, we examine the stability of the fractional cointegrating relations using our new LCM-based sup-Wald tests. Unlike existing studies, we find the IV-RV cointegrating relation to be unstable, and that carefully constructed time-series forecasts are more efficient than IV in capturing low-frequency movements in RV. These results are interesting from a volatility forecasting perspective, indicating a change in the low-frequency characteristics of RV over time. In addition, they have implications for asset pricing and return prediction, because our evidence is at odds with the presumed stability of the cointegration models, which are often used to estimate the volatility risk premium component associated with a long-run equilibrium and to forecast returns. In fact, we provide reduced-form evidence that the significant forecasting power of the volatility risk premium for returns materializes during a period surrounding the global financial crises of 2008-2009, where the low-frequency relation between IV and RV breaks down, with the former being completely spanned by our long-memory time-series forecasts. In contrast, during periods where a “normal” long-run equilibrium between IV and RV prevails, the volatility risk premium has no significant predictive power for returns. These findings provide a challenge for existing asset pricing models that stipulate

a stationary relation between RV, IV, the volatility risk premium and future returns.

The paper proceeds as follows. Section 2 describes the modeling framework, the underlying assumptions, and the competing parameter (in)stability hypotheses. Section 3 introduces the LCM procedure and derives the requisite partial-sample limit theory. Section 4 provides the new LCM-based sup-Wald tests and establishes their asymptotic properties. The simulation study is contained in Section 5, and Section 6 explores the IV-RV relation empirically. Section 7 concludes. The proofs of the theoretical results are relegated to the Appendix together with additional simulation results.

Throughout the paper, “ $\sim$ ” and “ $\asymp$ ” signify, respectively, that the ratio of the left- and right-hand-side of the relevant equation tends to one and a non-zero constant in the limit, element-wise, while we use  $\|\cdot\|$  to denote the Frobenius matrix norm,  $\circ$  the Hadamard product,  $i = \sqrt{-1}$ , and  $\mathbf{I}_k$  the  $k$ -dimensional identity matrix. Moreover, the notation “ $\rightarrow$ ”, “ $\xrightarrow{\mathbb{D}}$ ” and “ $\xrightarrow{\mathbb{P}}$ ” describe the limit, convergence in distribution and convergence in probability. Finally, we denote by “ $\Rightarrow$ ” weak convergence under the uniform metric on the space  $D[0, 1]$ , as defined by Pollard (1984, Chapter V).

## 2 Framework and Competing Hypotheses

First, to set the stage, let  $(1 - L)^d$  denote a generic fractional filter,

$$(1 - L)^d = \sum_{i=0}^{\infty} \frac{\Gamma(i - d)}{\Gamma(i + 1)\Gamma(-d)} L^i, \quad (1)$$

with  $\Gamma(\cdot)$  being the gamma function. Next, we assume observations are available for a  $(k + 1) \times 1$  vector of economic variables  $\mathbf{z}_t = (y_t, \mathbf{x}'_{t-1})'$ , which obeys a Type II fractional model,

$$\mathbf{D}(L) (\mathbf{z}_t - \boldsymbol{\mu}) = \mathbf{v}_t \mathbf{1}_{\{t \geq 1\}} \quad (2)$$

where  $\boldsymbol{\mu}$  is a  $(k + 1) \times 1$  vector of nonrandom unknown finite numbers, either capturing the means or initial values of the variables,  $\mathbf{v}_t = (e_t, \mathbf{u}'_{t-1})'$  is a weakly dependent vector process, and,

$$\mathbf{D}(L) = \text{diag}[(1 - L)^{d_1}, \dots, (1 - L)^{d_{k+1}}].$$

In this setting, following Andersen & Varneskov (2020a), we define and study the predictive relation between  $y_t$  and  $\mathbf{x}_{t-1}$  through their weakly dependent components. Specifically, we assume,

$$e_t = \boldsymbol{\beta}'_t \mathbf{u}_{t-1} + \eta_t^{(b)}, \quad t = 1, \dots, n, \quad (3)$$

where  $\eta_t^{(b)} = (1 - L)^b \eta_t$  for some constant  $b \geq 0$ ,  $\eta_t \sim I(0)$ , and the  $k$ -dimensional parameter vector,  $\boldsymbol{\beta}_t$ , decomposes as  $\boldsymbol{\beta}_t \equiv (\boldsymbol{\beta}'_t, \boldsymbol{\delta}'_t)'$  with  $\boldsymbol{\beta}_t \in \mathbb{R}^{k_\beta}$ ,  $\boldsymbol{\delta}_t \in \mathbb{R}^{k_\delta}$  and  $k = k_\beta + k_\delta$ .

The main difference between the relation (3) and the corresponding assumption in Andersen & Varneskov (2020a) concerns the potential time variation of  $\boldsymbol{\beta}_t$ . They consider estimation and inference

for predictability under the standard invariant parameter vector assumption,

$$\mathcal{H}_0: \boldsymbol{\beta}_t = \boldsymbol{\beta}, \quad \text{for all } t = 1, \dots, n. \quad (4)$$

Importantly, even under the restriction,  $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ , the observation equation (2) and functional form (3) encompass most fractionally integrated systems within the long memory literature, and they translate into a well-defined, *balanced* relation between  $y_t$  and  $\mathbf{x}_{t-1}$ , despite the variables exhibiting distinct degrees of fractional integration. This is readily seen by combining the two equations, obtaining,

$$y_t = a + \mathbf{B}'\mathcal{Q}(L)\mathbf{x}_{t-1} + v_t, \quad t = 1, \dots, n, \quad (5)$$

where  $\mathcal{Q}(L) = \mathbf{D}_x(L)(1-L)^{-d_1}$ ,  $\mathbf{D}_x(L)$  is the  $k \times k$  lower-right submatrix of the fractional filter  $\mathbf{D}(L)$ ,  $a = \mu_y - \mathbf{B}'\mathcal{Q}(L)\boldsymbol{\mu}_x$  for  $\boldsymbol{\mu} = (\mu_y, \boldsymbol{\mu}_x)'$  and  $v_t = (1-L)^{-d_1}\eta_t^{(b)} = (1-L)^{b-d_1}\eta_t$ . The filtering matrix  $\mathcal{Q}(L)$  adjusts the persistence of the predictors so that regression balance is achieved in this implied relation between the observed variables. If the system is balanced a priori, then  $\mathcal{Q}(L) = \mathbf{I}_k$ , and the adjustment is trivial. Its presence, however, ensures that the implied dynamics of the dependent variable,  $y_t \sim I(d_1)$ , are equivalent, both in the presence and absence of predictive power,

$$\mathbf{B} = \mathbf{0} \quad \text{and} \quad \mathbf{B} \neq \mathbf{0}, \quad (6)$$

respectively, and apply for settings with and without fractional cointegration;

- (i)  $\mathbf{B} = \mathbf{0}$  and  $b = 0$ :  $\mathbf{x}_{t-1}$  contains no predictive power for  $y_t$ .
- (ii)  $\mathbf{B} \neq \mathbf{0}$  and  $b = 0$ :  $\mathbf{x}_{t-1}$  partially spans the persistent component(s) of  $y_t$ .
- (iii)  $\mathbf{B} \neq \mathbf{0}$  and  $b > 0$ :  $\mathbf{x}_{t-1}$  cointegrate with  $y_t$ , spanning its persistent component(s).

Andersen & Varneskov (2020a) show that most predictive long-memory systems, with and without cointegration, are nested in scenarios (i)-(iii), and they propose the Local SpectrumM (LCM) procedure (described below) to draw inference on  $\mathbf{B}$ , whose validity holds uniformly across these settings, without prior knowledge of the persistence of the system, or whether it features cointegration. Hence, the LCM procedure overcomes problems with spurious inference in persistent systems, as documented by, among others, Granger & Newbold (1974), Phillips (1986) and Tsay & Chung (2000).

Their analysis, however, does rest critically on the null hypothesis,  $\mathcal{H}_0$ . In contrast, in this paper, we are concerned with testing whether  $\mathcal{H}_0$  holds, allowing for either parameter instability or structural change in (a subset of)  $\boldsymbol{\beta}_t$  under the alternative. In line with previous studies offering instability tests in regression contexts, see, e.g., Andrews (1993) and Bai & Perron (1998) for one and multiple breaks, respectively, our framework nests pure and partial structural change models, when  $k_\delta = 0$  and  $0 < k_\delta < k$ . Our new structural change tests within this flexible long-memory setting substantively generalizes the dynamic specifications assumed in prior work, both under the null hypothesis, as conveyed by scenarios (i)-(iii), and under the alternative. The specific alternatives are detailed next.

## 2.1 Alternative Hypotheses

First, define the set  $\Phi \in (\epsilon, 1 - \epsilon)$ , for some  $\epsilon > 0$ , let  $\{\varphi_1 < \dots < \varphi_q\} \in \Phi$  denote the sample break fractions, and fix  $\varphi_0 = 0$  and  $\varphi_{q+1} = 1$ . In addition, we label the corresponding observation subsets  $\mathcal{T}_g = \{t = \lfloor \varphi_g n \rfloor + 1, \dots, \lfloor \varphi_{g+1} n \rfloor\}$  for the  $q + 1$  regimes  $g = 0, \dots, q$  and denote the number of observations in  $\mathcal{T}_g$  by  $n_g = n(\mathcal{T}_g)$ . We will entertain multiple alternative hypotheses concerning structural change of the form,

$$\mathcal{H}_{A1} : \boldsymbol{\beta}_t = \boldsymbol{\beta}_g, \quad \text{for } t \in \mathcal{T}_g, \quad g = 0, \dots, q, \quad (7)$$

where the break dates are unknown, while  $q$  may be known or unknown, a priori. This alternative allows the predictive power of  $\boldsymbol{x}_{t-1}$  to be characterized by distinct regimes of the form (i)-(iii). For example,  $\boldsymbol{x}_{t-1}$  may partially span the persistence of  $y_t$  in some subsets of the sample, but not in others, thereby switching between scenarios (i) and (ii). Moreover, we allow for predictive systems, where the cointegration properties for a subset of the regressors may break down in specific economic regimes, as long as the system remains in scenario (iii). To see this, suppose we have a cointegration model, where  $y_t \in I(1)$ ,  $\boldsymbol{x}_{t-1} = (x_{1,t-1}, x_{2,t-1})' \in I(1)$  and  $b = 1$ . Then, our setting may feature a partial structural change model, with  $k_\beta = k_\delta = 1$ , where cointegration arises through a time-invariant relation with  $x_{2,t-1}$ , but the predictive power of  $x_{1,t-1}$  pertains only to certain regimes.

As a final example, the alternative hypothesis accommodates fractionally cointegrated systems, where persistence is balanced, that is, with  $y_t \in I(d)$ ,  $\boldsymbol{x}_{t-1} \in I(d)$  and  $0 \leq b \leq d$ , and where the relative efficiency of (a subset of) the predictors may change over the course of the sample.

The flexibility of scenarios (i)-(iii), allowing all variables to have different fractional integration orders and the predictive relation to feature cointegration, in conjunction with the alternative hypothesis (7), facilitates testing for parameter changes in the vast majority of the predictive long-memory systems studied in the literature. Furthermore, it is robust to a feature not accommodated by prior structural break tests; namely, we remain agnostic as to whether the system involves cointegration. Specifically, Andrews (1993) and Bai & Perron (1998), among others, provide structural change tests with unknown break dates in weakly dependent settings, and Hansen (1992), Hansen (2003) and Kerjiwal & Perron (2010) consider cointegrated systems that may also feature  $I(0)$  regressors, but their respective procedures cannot simultaneously accommodate  $I(0)$  and  $I(1)$  errors, because the latter induce spurious inference; see, e.g., the discussions in Perron (2006) and Casini & Perron (2019). In our setting, the errors may be  $I(0)$  or  $I(1)$ , and we also accommodate fractional cointegration within the system. As will be explained in detail subsequently, this added generality stems from our use of the LCM procedure in the construction of the structural change tests.

Finally, we emphasize that, despite our tests being designed with alternatives such as the structural break specification (7) in mind, they will also have power against other alternatives of the form,

$$\mathcal{H}_{A2} : \boldsymbol{\beta}_t \neq \boldsymbol{\beta}_s, \quad \text{for some } t, s \geq 1,$$



as noted by Andrews (1993) in a weakly dependent setting without cointegration. Following the analysis in the latter, we show below that our new tests, indeed, have power against local alternatives of the form  $\beta_t = \beta + \eta(t/n)/\mathcal{S}_n(b)$  for some bounded function  $\eta(t/n)$  on  $[0, 1]$ , that is not almost everywhere constant on  $\Phi$ , and where  $\mathcal{S}_n(b)$  is an increasing deterministic function of the sample size  $n$ , which depends on the cointegration strength,  $b$ . Hence, our tests will capture more general forms of parameter instability under the predictive scenarios (i)-(iii).

## 2.2 Assumptions

The theoretical analysis rests on formal assumptions for the components of equations (2) and (3). Moreover, we assume an estimator of the fractional integration orders,  $\hat{d}_i$  for  $i = 1, \dots, k+1$ , satisfying mild consistency requirements, is available. The assumptions, stated in terms of  $\mathbf{q}_t = (\mathbf{u}'_{t-1}, \eta_t)'$ , mirror the corresponding ones in Andersen & Varneskov (2020a), and they are consistent with the regularity conditions underpinning the fractional cointegration analyses in, e.g., Robinson & Marinucci (2003), Christensen & Nielsen (2006) and Christensen & Varneskov (2017).

**Assumption D1.** *The vector process  $\mathbf{q}_t$ ,  $t = 1, \dots$ , is covariance stationary with spectral density matrix satisfying  $\mathbf{f}_{qq}(\lambda) \sim \mathbf{G}_{qq}$  as  $\lambda \rightarrow 0^+$ , where the upper left  $k \times k$  submatrix,  $\mathbf{G}_{uu}$ , has full rank, and the  $(k+1)$ th element of the diagonal,  $G_{\eta\eta}$ , is strictly greater than zero. Moreover, there exists a  $\varpi \in (0, 2]$  such that  $\|\mathbf{f}_{qq}(\lambda) - \mathbf{G}_{qq}\| = O(\lambda^\varpi)$  as  $\lambda \rightarrow 0^+$ . Finally, let  $\mathbf{G}_{qq}(i, k+1)$  be the  $(i, k+1)$ th element of  $\mathbf{G}_{qq}$ , which has  $\mathbf{G}_{qq}(i, k+1) = \mathbf{G}_{qq}(k+1, i) = 0$  for all  $i = 1, \dots, k$ .*

**Assumption D2.**  *$\mathbf{q}_t$  is a linear process,  $\mathbf{q}_t = \sum_{j=0}^{\infty} \mathbf{A}_j \boldsymbol{\epsilon}_{t-j}$ , with square summable coefficients  $\sum_{j=0}^{\infty} \|\mathbf{A}_j\|^2 < \infty$ , the innovations satisfy, almost surely,  $\mathbb{E}[\boldsymbol{\epsilon}_t | \mathcal{F}_{t-1}] = 0$  and  $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t | \mathcal{F}_{t-1}] = \mathbf{I}_{k+1}$ , and the matrices  $\mathbb{E}[\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t | \mathcal{F}_{t-1}]$  and  $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t \otimes \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t | \mathcal{F}_{t-1}]$  are nonstochastic, finite, and do not depend on  $t$ , with  $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon}_s, s \leq t)$ . There exists a random variable  $\zeta$  such that  $\mathbb{E}[\zeta^2] < \infty$  and, for all  $c$  and some  $C$ ,  $\mathbb{P}[\|\mathbf{q}_t\| > c] \leq C \mathbb{P}[\|\zeta\| > c]$ . For future reference, we denote the periodogram of  $\boldsymbol{\epsilon}_t$  by  $\mathbf{J}(\lambda)$ .*

**Assumption D3.** *For  $\mathbf{A}(\lambda, i)$ , the  $i$ -th row of  $\mathbf{A}(\lambda) = \sum_{j=0}^{\infty} \mathbf{A}_j e^{ij\lambda}$ , its partial derivative satisfies  $\|\partial \mathbf{A}(\lambda, i) / \partial \lambda\| = O(\lambda^{-1} \|\mathbf{A}(\lambda, i)\|)$  as  $\lambda \rightarrow 0^+$ , for  $i = 1, \dots, k+1$ .*

**Assumption M.** *Let  $0 \leq d_i < 2$  for all  $i = 1, \dots, k+1$  and define  $\underline{d} = \min_{i=1, \dots, k+1} d_i$ . Moreover, suppose that the cointegration strength parameter satisfies  $0 \leq b \leq \min(\underline{d}, 1)$ .*

**Assumption F.** *Let  $m_d \asymp n^\rho$  be a sequence of integers where  $0 < \rho \leq 1$ , then, for all  $i = 1, \dots, k+1$  elements of  $\mathbf{z}_t$ , we assume to have an estimator with the property,*

$$\hat{d}_i - d_i = O_p(1/\sqrt{m_d}), \quad \text{and we then let, } \hat{\mathbf{D}}(L) = \text{diag} \left[ (1-L)^{\hat{d}_1}, \dots, (1-L)^{\hat{d}_{k+1}} \right].$$

The regularity conditions on  $\mathbf{q}_t$  in Assumptions D1-D3 are standard in the literatures analyzing semi-parametric fractional cointegration as well as the estimation of multivariate fractional time series models, e.g., Shimotsu (2007) and Nielsen (2015). Specifically, conditions D1 and D3 impose a mild rate

of convergence for the spectral density  $\mathbf{f}_{qq}(\lambda)$  as  $\lambda \rightarrow 0^+$ , which depends on the smoothness parameter  $\varpi \in (0, 2]$ . Moreover, D1 requires full rank of  $\mathbf{u}_{t-1}$  and the vector being locally exogenous to  $\eta_t$  as  $\lambda \rightarrow 0^+$ , but not global exogeneity. That is,  $\mathbf{u}_{t-1}$  and  $\eta_t$  may have non-trivial covariance for frequencies  $\lambda \rightarrow c > 0$ . Finally, D2 specifies linearity, martingale and moment conditions for  $\mathbf{q}_t$ , allowing for general multivariate dependence, but rules out time-variation in the conditional covariance between the innovations as well as their third and fourth conditional moments. Assumption M restricts the fractional integration order of the variables to be below 2 as well as the cointegration range in relation to the memory of the system. The first part is innocuous, since it is satisfied by most macroeconomic and financial time series. Moreover, the cointegration restriction is rather intuitive. If the system is balanced  $d_i = d$ ,  $i = 1, \dots, k+1$ , then we accommodate both  $I(1)$ - $I(0)$  and fractional cointegration settings, without any additional restrictions. If the memory parameters differ, Assumption M implies that the cointegration relation must be balanced and, at least, weakly dependent. To see this, suppose  $y_t \in I(0.3)$  and  $\mathbf{x}_{t-1} \in I(0.6)$ . Then, the upper bound  $b = 0.3$ , in conjunction with the balanced regression requirement (5), implies a cointegrating relation with  $I(0)$  residuals. On the other hand, if  $y_t \in I(0.6)$  and  $\mathbf{x}_{t-1} \in I(0.3)$ , then the restriction generates  $I(0.3)$  residuals. Hence, the regressors cannot purge  $y_t$  of more persistence than their own degree of fractional integration.

These conditions allow for  $\mathbf{z}_t$  to constitute vector ARFIMA processes (for  $\varpi = 2$ ), thus nesting VAR and integrated VAR dynamics as special cases with  $d_i = 0$  and  $d_i = 1$ , respectively, for  $i = 1, \dots, k+1$ , and  $b = 0$ . Generally, however, the properties of the variables depend on their integration orders. Specifically, the  $i$ th variable have (asymptotically) stationary long memory, when  $0 < d_i < 1/2$ , it is non-stationary when  $d_i \geq 1/2$ , and it has a well-defined mean for  $d_i < 1$ . Hence, the assumptions allow for very flexible persistence of the variables as well as fractional cointegration amongst them.

Finally, the conditions on  $\hat{d}_i$  in Assumptions F are mild, essentially only requiring the existence of an estimator which, under the Assumptions D1-D3 and M, is consistent. Examples of estimators, that satisfy these requirements for the entire range,  $0 \leq d_i < 2$ , are the parametric ARFIMA estimators by Hualde & Robinson (2011) and Nielsen (2015), the mean-robust exact Local Whittle (ELW) estimator by Shimotsu (2010) and the trimmed ELW (TELW) estimator by Andersen & Varneskov (2020a).

**Remark 1.** *The LCM procedure, and thus our new tests below, further accommodates regressor endogeneity in the spirit of Stambaugh (1999) and Pastor & Stambaugh (2009) as well as regressors that have been pre-estimated as fractional cointegration residuals. These effects are formally analyzed in Andersen & Varneskov (2020a, Section 4), who show that robustness is achieved by slightly stricter trimming conditions than those required by Assumption T below. As these conditions are not binding in our empirical analysis, we refrain from formally introducing endogeneity and pre-estimated regressors here for simplicity of exposition, noting, however, that all our subsequent qualitative results will apply.*

### 3 Partial Sample LCM Inference

This section introduces a partial sample version of the two-step LCM estimator by Andersen & Varneskov (2020a). Furthermore, it provides a joint functional limit theory for  $q + 1$  partitions of the sample under the null hypothesis of parameter stability and develop associated feasible inference procedures. These estimators, and their resulting limit theory, are of independent interest but, importantly, also form the basic building blocks in the design of the structural change tests in Section 4.

#### 3.1 Partial Sample LCM Estimation and Limit Theory

The LCM estimation procedure is carried out in two steps.

**Step 1: Fractional filtering.** As in Assumption F, we use a consistent estimator of the fractional integration orders  $\hat{d}_i$ , for  $i = 1, \dots, k + 1$ , to compute  $\hat{\mathbf{D}}(L)$ , and then fractionally filter  $\mathbf{z}_t$  to obtain estimates of  $\mathbf{v}_t$ , the weakly dependent components of the variables in equation (2),

$$\hat{\mathbf{v}}_t \equiv (\hat{e}_t, \hat{\mathbf{u}}'_{t-1})' = \hat{\mathbf{D}}(L)\mathbf{z}_t, \quad (8)$$

without, however, accounting for the mean, or initial value, of  $\mathbf{z}_t$ . Rather than treating “de-meaning” of the series on a case-by-case basis, depending on the magnitude of  $d_i$ , we account for the residual impact of the mean component,  $\hat{\mathbf{D}}(L)\boldsymbol{\mu}$ , in a unified manner during the second stage estimation.

**Step 2: Medium Band Least Squares.** For a given partition  $g = 0, \dots, q$ , denote the Fourier frequencies for observation set  $\mathcal{T}_g$  by  $\lambda_j(g) = 2\pi j/n_g$ ,  $j = 1, \dots, n_g$ . Then, the partial sample discrete Fourier transform and cross-periodogram are defined as,

$$\mathbf{w}_h(\lambda_j, g) = \frac{1}{\sqrt{2\pi n_g}} \sum_{t \in \mathcal{T}_g} \mathbf{h}_t e^{it\lambda_j(g)}, \quad \mathbf{I}_{hk}(\lambda_j, g) = \mathbf{w}_h(\lambda_j, g) \bar{\mathbf{w}}_k(\lambda_j, g), \quad (9)$$

respectively, where  $\mathbf{h}_t$  and  $\mathbf{k}_t$  are generic (and compatible) vector time series. Moreover, let  $\mathbf{I}_{hk}(\lambda_j, g) = \Re(\mathbf{I}_{hk}(\lambda_j, g)) + i\Im(\mathbf{I}_{hk}(\lambda_j, g))$  be the real and imaginary decomposition of  $\mathbf{I}_{hk}(\lambda_j, g)$ . Finally, define the corresponding trimmed discretely averaged co-periodogram (TDAC) as,

$$\hat{\mathbf{F}}_{hk}(\ell, m, g) = \frac{2\pi}{n_g} \sum_{j=\ell_g}^{m_g} \Re(\mathbf{I}_{hk}(\lambda_j, g)), \quad 1 \leq \ell_g \leq m_g \leq n_g, \quad (10)$$

where  $\ell_g = \ell(n_g)$  and  $m_g = m(n_g)$  are partition-specific trimming and bandwidth functions. Then, we can write the TDAC of  $\hat{\mathbf{u}}_{t-1}$  as  $\hat{\mathbf{F}}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}(\ell, m, g)$  and, similarly, of  $\hat{\mathbf{u}}_{t-1}$  and  $\hat{e}_t$  as  $\hat{\mathbf{F}}_{\hat{\mathbf{u}}\hat{e}}(\ell, m, g)$ , and use these to define the partial sample medium band least squares (MBLS) estimator as,

$$\hat{\mathbf{B}}(\ell, m, g) = \hat{\mathbf{F}}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}(\ell, m, g)^{-1} \hat{\mathbf{F}}_{\hat{\mathbf{u}}\hat{e}}(\ell, m, g), \quad (11)$$

for which  $\ell_g, m_g \rightarrow \infty$  and  $\ell_g/m_g + m_g/n_g \rightarrow 0$ , as  $n_g \rightarrow \infty$ .<sup>2</sup> Moreover, by definition of the partial sample observation scheme, we have  $n_g/n \rightarrow \varphi_{g+1} - \varphi_g > 0$  as  $n \rightarrow \infty$ , that is, each subsample grows proportionally to  $n$ . Importantly, the MBLS estimator in equation (11) is equivalent to the second-step estimator in Andersen & Varneskov (2020a), although it applies only to a subsample rather than the full set of observations.<sup>3</sup> Hence, it inherits some useful properties. First, it is semiparametric about the dynamics of  $\mathbf{z}_t$ , requiring only some specific structure on the spectrum as  $\lambda \rightarrow 0^+$ . Second, the trimming and bandwidth functions aid in the recovery of the parameter vector within each subsample by asymptotically eliminating the first-stage estimation errors through filtering. Specifically, the trimming component eliminates any slippage from the means, or initial values,  $\widehat{\mathbf{D}}(L)\boldsymbol{\mu}$ , occurring at lower frequencies and, in conjunction with the bandwidth, the estimation errors given in Assumptions F.

Consistent with subsample estimation of the parameter vector  $\boldsymbol{\mathcal{B}}$ , we require additional conditions on the partition-specific bandwidth and trimming functions. These are partial sample analogs of the corresponding assumptions in Andersen & Varneskov (2020a).

**Assumption T.** *Let the bandwidth  $m_g \asymp n_g^\kappa$  and  $\ell_g \asymp n_g^\nu$  with  $0 < \nu < \kappa < \varrho \leq 1$  for sample partitions  $g = 0, \dots, q$ . Moreover, recall that the parameter  $\varpi \in (0, 2]$  measures the smoothness of the spectral density in Assumption D1. Then, the following cross-restrictions are imposed,*

$$\frac{m_g^{1+2\varpi}}{n_g^{2\varpi}} + \frac{\ell_g^{1+\varpi+b}}{n_g^\varpi m_g^{1/2+b}} + \frac{n_g^{1/2+b}}{m_d^{1/2} m_g^b \ell_g} + \frac{n_g^{1-2d+b}}{m_g^{1/2-2d+b} \ell_g^2} + \frac{n_g^b}{m_g^{1/2+b}} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

The conditions in Assumption T are mild. The first condition is standard for semiparametric estimation in the frequency domain, e.g., Robinson (1995) and Lobato (1999), while the last is needed, because we only require local exogeneity between  $\mathbf{u}_{t-1}$  and  $\eta_t$  as  $\lambda \rightarrow 0^+$ . For the empirically relevant vector ARFIMA process ( $\varpi = 2$ ), this condition implies  $\kappa < 4/5$ . The remaining conditions are LCM-specific upper and lower bounds in the trimming and bandwidth rates,  $\nu$  and  $\kappa$ , which are discussed at length in Andersen & Varneskov (2020a). They depend on  $\underline{d}$ ,  $b$  and  $\varrho$ , thus reflecting the persistence properties of the economic system, cointegration and the convergence rate of the first-step estimator of the fractional integration orders, respectively. For the vector ARFIMA process, the most important implications of conditions 2-4 of Assumption T amount to,

$$(3/5 - (2\underline{d} - b)/5)/2 \vee (1 - \varrho)/2 + b(1 - \kappa) < \nu < \kappa < 4/5, \quad (12)$$

<sup>2</sup>We have suppressed dependence on the (lagged) time  $t$  indicator in  $\widehat{\mathbf{F}}_{\widehat{a}\widehat{a}}(\ell, m, g)$  and  $\widehat{\mathbf{F}}_{\widehat{a}\widehat{c}}(\ell, m, g)$  to ease exposition. The dependence on time, however, is explicated when necessary, e.g., when establishing Theorem 1 below.

<sup>3</sup>The MBLS estimator is also related to the estimator in Christensen & Varneskov (2017), who, however, input observable variables rather than fractionally filtered ones. This has important implications; the framework in the latter only applies to (asymptotically) stationary processes and cannot handle situations where the integration orders differ across variables in the system. Moreover, the asymptotic analysis in Christensen & Varneskov (2017) assumes cointegration. However, since the latter shows robustness of their estimator (via trimming) towards outliers, structural breaks (in means), Markov switching means, certain deterministic trends, etc., which are known to contaminate co-periodograms at frequency ordinates close to the origin, see, e.g., Diebold & Inoue (2001) and Perron & Qu (2010), this suggests that our estimator will enjoy similar robustness properties. We do not formally analyze those effects here.

for the trimming rate, thereby generating a conservative bound  $3/10 < \nu$  for standard choices of the parameters  $\varrho$  and  $\kappa$ . The last condition generates a mild lower bound restriction on the bandwidth rate,  $b/(1/2 + b) \leq \min(\underline{d}, 1)/(1/2 + \min(\underline{d}, 1)) < \kappa$ . Naturally, the trimming and bandwidth rates can be made subsample specific, as long as they satisfy Assumption T in each partition. We refrain from doing so for simplicity of exposition and to avoid dealing with a large set of tuning parameters.

Before stating the joint functional limit theory for the subsample LCM estimators, we introduce some auxiliary notation. Specifically, we let,

$$\boldsymbol{\vartheta}(b) = \text{diag} \left[ (\varphi_1 - \varphi_0)^{2b(\kappa-1)-\kappa-1}, \dots, (\varphi_{q+1} - \varphi_q)^{2b(\kappa-1)-\kappa-1} \right] \quad (13)$$

be a  $(q+1)$ -dimensional diagonal matrix, depending on the sample partition, and,

$$\mathbf{C}(b) = \mathbf{G}_{uu}^{-1} \mathbf{V}(b) \mathbf{G}_{uu}^{-1}, \quad \text{where} \quad \mathbf{V}(b) = \mathbf{G}_{uu} \frac{G_{\eta\eta}}{2(1+2b)}, \quad (14)$$

denote the  $k \times k$  asymptotic covariance matrix of the LCM estimator. Moreover, write the full sample Fourier frequencies as  $\lambda_j = 2\pi j/n$ ,  $j = 1, \dots, n$ , and let  $\ell = \ell(n)$  and  $m = m(n)$  be the full sample trimming and bandwidth sequences, which also satisfy Assumption T. Finally, let  $\mathbf{Y}(\boldsymbol{\varphi})$ , for some  $\boldsymbol{\varphi} \in [0, 1]$ , denote a  $k \times 1$  vector of independent Brownian motions on  $[0, 1]$ , and we then define,

$$\mathbf{Y}(\boldsymbol{\varphi}) = \begin{pmatrix} \mathbf{Y}(\varphi_1) - \mathbf{Y}(\varphi_0) \\ \vdots \\ \mathbf{Y}(\varphi_{q+1}) - \mathbf{Y}(\varphi_q) \end{pmatrix}, \quad (15)$$

with dimension  $(q+1) \times 1$ , for the sample partitions  $\boldsymbol{\varphi}$  on the set  $\Phi$ . We are now ready to characterize the joint limit theory for subsample LCM estimators given any partition,  $g = 0, \dots, q$ .

**Theorem 1.** *Let Assumptions D1-D3, F, M and T hold. Moreover, suppose that the mutual consistency condition  $\max(0, (1 - 3\kappa/2)/(1 + \kappa/2)) < \varpi \leq 2$  is satisfied, then, for every sequence of sample partitions  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_q)'$ , whose elements are in the set  $\Phi$ , it follows,*

$$\sqrt{m} \lambda_m^{-b} \begin{pmatrix} \widehat{\mathbf{B}}(\ell, m, 0) - \mathbf{B} \\ \vdots \\ \widehat{\mathbf{B}}(\ell, m, q) - \mathbf{B} \end{pmatrix} \implies \left( \boldsymbol{\vartheta}(b)^{1/2} \circ (\mathbf{G}_{uu}^{-1} \mathbf{V}(b)^{1/2}) \right) \mathbf{Y}(\boldsymbol{\varphi}).$$

Theorem 1 extends Andersen & Varneskov (2020a, Theorem 1) to allow for non-overlapping partitions of the sample and demonstrates that subsample LCM estimators enjoy a joint Gaussian uniform limit. Importantly, the convergence rate depends on whether  $\mathbf{B} \neq \mathbf{0}$  and, if this is the case, whether the system features cointegration. Specifically, the rate is  $\sqrt{m}$  in its absence ( $b = 0$ ), consistent with semiparametric estimators in the frequency domain, e.g., Brillinger (1981, Chapters 7-8). Conversely, the rate is  $\sqrt{m} \lambda_m^{-b} \asymp \sqrt{m} (n/m)^b$  in its presence ( $b > 0$ ), in analogy with super consistency properties in

standard cointegration settings. Moreover, cointegration lowers the asymptotic variance, as reflected by the scale  $1/(2(1+2b))$ . The joint Gaussian uniform distribution theory is remarkable; holding in scenarios with and without cointegration, across (asymptotically) stationary and non-stationary variables in the observable persistent system (2), and for both weak ( $b < 1/2$ ) and strong ( $b \geq 1/2$ ) cointegration settings. This feature is exactly what allows our structural change tests to remain, a priori, agnostic about the persistence of the system and whether it features cointegration, thus considerably generalizing the scope relative to, e.g., Andrews (1993), Bai & Perron (1998) and Kerjiwal & Perron (2010), who either rely on standard  $\sqrt{n}$  asymptotic properties of OLS (or GMM) estimators in weakly dependent settings or rate- $n$  results in cointegration settings.

### 3.2 Feasible Inference

The design of our structural change tests below require estimates of the asymptotic covariance matrix for each sample partition. To this end, we must provide consistent subsample estimators of the long-run covariance matrix  $\mathbf{G}_{uu}$  and variance  $G_{\eta\eta}$ . The main challenge is that we observe  $\hat{\mathbf{v}}_t$ , not  $\mathbf{v}_t$  – raising issues similar to those that motivated us to apply the robust MBLs estimator in the second step of the LCM procedure. The residuals  $\eta_t$  are latent, and we estimate them as,

$$\hat{\eta}_t^{(b)}(g) = \hat{e}_t - \hat{\mathbf{B}}(\ell, m, g)' \hat{\mathbf{u}}_{t-1}, \quad \hat{\eta}_t(g) \equiv (1-L)^{-\hat{b}(g)} \hat{\eta}_t^{(b)}(g), \quad t \in \mathcal{T}_g, \quad (16)$$

for each subsample  $g = 0, \dots, q$ , where  $\hat{b}(g)$  denotes a consistent estimator of  $b$ . The intuition behind the estimator  $\hat{\eta}_t(g)$  is the following; we must undo the (over-)differencing, indicated by  $b \neq 0$ , to recover estimates of the residuals, that are then used to estimate  $G_{\eta\eta}$ . This is carried out on each subsample for our estimates to be valid under the null,  $\mathcal{H}_0$ , and the alternative hypothesis,  $\mathcal{H}_{A1}$ .

Next, we use  $\hat{\mathbf{v}}_t$  and  $\hat{\eta}_t(g)$ ,  $g = 0, \dots, q$ , to define a generic class of subsample estimators of the asymptotic covariance matrices,

$$\hat{\mathbf{G}}_{hh}(\ell_G, m_G, g) = \frac{1}{m_{G,g} - \ell_{G,g} + 1} \sum_{j=\ell_{G,g}}^{m_{G,g}} \Re(\mathbf{I}_{hh}(\lambda_j, g)), \quad (17)$$

for some arbitrary vector  $\mathbf{h}_t$ , where  $m_{G,g} = m_G(n_g)$  and  $\ell_{G,g} = \ell_G(n_g)$  are other, partition-specific, bandwidth and trimming functions, which, as for MBLs, are used to eliminate estimation errors from fractional filtering, low-frequency slippage from  $\hat{\mathbf{D}}(L)\boldsymbol{\mu}$ , as well as estimation errors from both MBLs and  $\hat{b}(g)$ . Now, using equation (17), the asymptotic covariance of  $\hat{\mathbf{B}}(\ell, m, g)$  is estimated as,

$$\widehat{\mathbf{AVAR}}(g, b) = \hat{\mathbf{G}}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}(\ell_G, m_G, g)^{-1} \frac{\hat{\mathbf{G}}_{\hat{\eta}\hat{\eta}}(\ell_G, m_G, g) \lambda_{m_g}^{2\hat{b}(g)}(g)}{2(1+2\hat{b}(g))m_g}, \quad (18)$$

which implicitly accounts for the scale  $(\varphi_{g+1} - \varphi_g)^{2b(\kappa-1)-\kappa}$  by applying  $\lambda_{m_g}(g)$  and  $m_g$  instead of the full sample versions  $\lambda_m$  and  $m$ . To retain flexibility in terms of implementing the inference and testing

procedures, in analogy to our treatment of the estimators  $\widehat{d}_i$ ,  $i = 1, \dots, k + 1$ , we assume an estimator  $\widehat{b}(g)$  is available, for  $g = 0, \dots, q$ , which satisfies mild consistency requirements. For example, these conditions are satisfied by the TELW estimator of Andersen & Varneskov (2020a). Before proceeding, we also need to impose restrictions on the new bandwidth and trimming parameters.

**Assumption B.** Let  $m_{b,g} \asymp n_g^\varepsilon$ , with  $g = 0, \dots, q$ , be a partition-specific sequence of integers where, for some  $0 < \varepsilon \leq 1$ , the subsample estimators satisfy  $\widehat{b}(g) - b = O_p(1/\sqrt{m_{b,g}})$ .

**Assumption T-G.** Let  $m_{G,g} \asymp n_g^{\kappa_G}$  and  $\ell_{G,g} \asymp n_g^{\nu_G}$  with  $0 < \nu_G < \kappa_G \leq \varrho \leq 1$ , for  $g = 0, \dots, q$ , and define the sequence of integers  $\underline{m}_g = m_d \wedge m_{b,g} \wedge m_g$ , then the following cross-restrictions are imposed on the trimming and bandwidth parameters:  $n/(m_{G,g} \ell_{G,g}^2) + n^2/(m_{G,g} \ell_{G,g}^2 \underline{m}_g) \rightarrow 0$ , as  $n \rightarrow \infty$ .

Finally, to illustrate the use of the subsample inference, we consider general linear hypotheses on the parameters  $\mathbf{B}$  as  $\mathcal{R}\mathbf{B} = \mathbf{r}$  for some  $h \times k$  selection matrix  $\mathcal{R}$  and  $h \times 1$  vector  $\mathbf{r}$ . The following theorem establishes validity of the feasible inference and partition-specific testing.

**Theorem 2.** Let the conditions of Theorem 1 and Assumptions B and T-G hold. Then, for every sequence of partitions  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_q)'$ , whose elements are on the set  $\Phi$ , it follows,

- (a)  $\|m_g \lambda_{m_g}^{-2b}(g) \widehat{\mathbf{AVAR}}(g, b) - \mathbf{C}(b)\| \xrightarrow{\mathbb{P}} 0$  for all  $g = 0, \dots, q$ , and,
- (b)  $\mathcal{LCM}(\ell, m, g) \equiv (\mathcal{R} \widehat{\mathbf{B}}(\ell, m, g) - \mathbf{r})' (\mathcal{R} \widehat{\mathbf{AVAR}}(g, b) \mathcal{R}')^{-1} (\mathcal{R} \widehat{\mathbf{B}}(\ell, m, g) - \mathbf{r}) \xrightarrow{\mathbb{D}} \chi_h^2$ .

Theorem 2 shows that Wald test statistics may be constructed for every sample partition and, importantly, have standard  $\chi^2$ -distribution theory. This feature is essential for the design and validity of our structural break tests, which are introduced next.

## 4 Testing for Parameter Instability and Structural Change

First, we provide a new test for general parameter instability as well as a single structural break with unknown break date, exploiting the Gaussian asymptotic limit theory for the LCM approach along with the sup-Wald testing procedure of Andrews (1993). Second, we extend this test to better accommodate multiple structural changes, occurring at unknown time points, where the number of regimes may be known or unknown, a priori, leveraging insights from Bai & Perron (1998).

### 4.1 Testing for Parameter Instability and a Single Structural Change

We exploit the standard asymptotic properties of the subsample LCM estimators in Theorems 1 and 2 to construct our Wald test statistic for a *single* structural break,  $g = 1$ ,

$$\mathcal{W}_n(\varphi_1) \equiv \left( \widehat{\boldsymbol{\beta}}(\ell, m, 0) - \widehat{\boldsymbol{\beta}}(\ell, m, 1) \right)' \left( \widehat{\mathbf{AVAR}}_{\boldsymbol{\beta}}(0, b) + \widehat{\mathbf{AVAR}}_{\boldsymbol{\beta}}(1, b) \right)^{-1} \left( \widehat{\boldsymbol{\beta}}(\ell, m, 0) - \widehat{\boldsymbol{\beta}}(\ell, m, 1) \right),$$

in line with the alternative hypothesis  $\mathcal{H}_{A1}$ , where  $\widehat{\beta}(\ell, m, g)$  denotes the first  $k_\beta$  elements of the sub-sample LCM estimate  $\widehat{\mathcal{B}}(\ell, m, g)$  and, similarly,  $\widehat{\mathbf{AVAR}}_\beta(g, b)$  denotes the upper-left  $k_\beta \times k_\beta$  submatrix of  $\widehat{\mathbf{AVAR}}(g, b)$ . Then, we introduce a new class of sup-Wald tests,

$$\mathcal{P}_n(1) \equiv \sup_{\varphi_1 \in \Phi} \mathcal{W}_n(\varphi_1), \quad (19)$$

with, again,  $\Phi \in (\epsilon, 1 - \epsilon)$  for some  $0 < \epsilon < 1$ . The asymptotic properties of the  $\mathcal{P}_n(1)$  class of parameter instability tests under the null hypothesis are formalized by the following theorem.

**Theorem 3.** *Suppose the conditions of Theorems 1-2 hold. Let  $\zeta = \kappa/2 + (1 - \kappa)b$ , and,*

$$\begin{aligned} Q_{k_\beta}(\varphi_1, \zeta) &= \frac{\widetilde{\mathbf{Y}}_{k_\beta}(\varphi_1, \zeta)' \widetilde{\mathbf{Y}}_{k_\beta}(\varphi_1, \zeta)}{\left( (1 - \varphi_1)^{2\zeta} + \varphi_1^{2\zeta} \right) (\varphi_1(1 - \varphi_1))}, \quad \text{where} \\ \widetilde{\mathbf{Y}}_{k_\beta}(\varphi_1, \zeta) &= \left( \mathbf{Y}_{k_\beta}(\varphi_1) (1 - \varphi_1)^{\zeta+1/2} - (\mathbf{Y}_{k_\beta}(1) - \mathbf{Y}_{k_\beta}(\varphi_1)) \varphi_1^{\zeta+1/2} \right), \end{aligned}$$

and with  $\mathbf{Y}_{k_\beta}(\varphi_1)$  containing the first  $k_\beta$  elements of the vector of independent Brownian motions,  $\mathbf{Y}(\varphi_1)$ . Then, the following convergence results hold for all  $\varphi_1 \in \Phi$ ,

$$\mathcal{W}_n(\varphi_1) \Rightarrow Q_{k_\beta}(\varphi_1, \zeta), \quad \mathcal{P}_n(1) \xrightarrow{\mathbb{D}} \sup_{\varphi_1 \in \Phi} Q_{k_\beta}(\varphi_1, \zeta).$$

Despite the challenging setting, where all variables may exhibit different degrees of fractional integration and the system may feature cointegration, as conveyed by scenarios in (i)-(iii), the LCM-based sup-Wald test,  $\mathcal{P}_n(1)$ , enjoys asymptotic properties under the null hypothesis, which are identical to those achieved by the corresponding test in Andrews (1993) under weak dependence. This readily follows as,

$$Q_{k_\beta}(\varphi_1, \zeta) \stackrel{\mathbb{D}}{=} \frac{(\mathbf{Y}_{k_\beta}(\varphi_1) - \mathbf{Y}_{k_\beta}(1) \varphi_1)' (\mathbf{Y}_{k_\beta}(\varphi_1) - \mathbf{Y}_{k_\beta}(1) \varphi_1)}{\varphi_1 (1 - \varphi_1)}, \quad (20)$$

which is the tied-down Bessel process describing the asymptotic distribution theory in Andrews (1993, Theo. 3). The standardization in the construction of  $\mathcal{W}_n(\varphi_1)$ , thus, eliminates any potential changes in the asymptotic distribution of the LCM-based sup-Wald test stemming from the, possibly, non-standard convergence rate of LCM, as signified by  $\zeta$ , where  $\zeta = 1/2$  corresponds to  $\sqrt{n}$ . Consequently, our test is readily implementable in practice, with critical values provided in Andrews (1993, Table 1). Moreover, one may apply the identical arguments, as for Theorems 4-5 and Corollary 2 of the latter, due to the LCM Gaussian limit theory, to show that  $\mathcal{P}_n(1)$  has local power against alternatives of the form  $\mathcal{H}_{A2}$ . To obtain equivalent results, we simply set  $\mathcal{S}_n(b) = \sqrt{m} \lambda_m^{-b}$  instead of  $\sqrt{n}$ , in line with the differences in convergence rates between LCM and GMM in their respective settings.

**Remark 2.** *Due to the standard asymptotic behavior of the Wald test statistics,  $\mathcal{W}_n(\varphi_1)$ , one may also consider exp-Wald and mean-Wald tests in the spirit of Andrews & Ploberger (1994), using the former as building blocks. However, Kim & Perron (2009) show that tests for structural change based*



on either sup-Wald or exp-Wald formulations compares favorably to mean-Wald tests as well as tests based on Langrange Multiplier (LM) and Likelihood Ratio (LR) measures, according to an approximate Bahadur measure of efficiency. Hence, we abstain from pursuing such extensions here.

## 4.2 Testing for Multiple Structural Changes with Unknown Break Points

The LCM-based sup-Wald test in equation (19) is asymptotically valid against alternative hypotheses involving multiple structural changes, as shown Andrews (1993, Theo. 4-5, Cor. 2). Its finite sample power may be lacking, however; see, e.g., Vogelsang (1997). To alleviate concerns about the latter, we amend our LCM approach with the tests for multiple structural changes in Bai & Perron (1998), designed in weakly dependent settings and absent cointegration, to explicitly account for  $(q+1)$  regimes with unknown break dates. To this end, let us define,

$$\mathcal{R}(\hat{\beta}, q) \equiv \left( \hat{\beta}(\ell, m, 0)' - \hat{\beta}(\ell, m, 1)', \dots, \hat{\beta}(\ell, m, q-1)' - \hat{\beta}(\ell, m, q)' \right)', \quad (21)$$

$$\mathcal{A}(\hat{\beta}, q) \equiv \text{diag} \left( \widehat{\mathbf{AVAR}}_{\beta}(0, b) + \widehat{\mathbf{AVAR}}_{\beta}(1, b), \dots, \widehat{\mathbf{AVAR}}_{\beta}(q-1, b) + \widehat{\mathbf{AVAR}}_{\beta}(q, b) \right), \quad (22)$$

of dimensions  $qk_{\beta} \times 1$  and  $qk_{\beta} \times qk_{\beta}$ , respectively. Then, we apply our subsample LCM framework and generalize the Wald statistic and sup-Wald test in equation (19) as,

$$\mathcal{P}_n(q) \equiv \sup_{\varphi \in \Phi} \mathcal{W}_n(\varphi), \quad \mathcal{W}_n(\varphi) = \mathcal{R}(\hat{\beta}, q)' \mathcal{A}(\hat{\beta}, q)^{-1} \mathcal{R}(\hat{\beta}, q) / (qk_{\beta}), \quad (23)$$

This class of tests requires specifying the number of structural changes under the alternative, which may not be desirable in some applications. Hence, as also considered by Bai & Perron (1998), we further introduce a “double maximum” version of our LCM-based sup-Wald test,

$$\mathcal{D}_n(q_{\max}) \equiv \max_{1 \leq q \leq q_{\max}} \mathcal{P}_n(q), \quad (24)$$

for some upper bound  $q_{\max}$ . The main advantage of “double maximum” tests, as thoroughly discussed by Bai & Perron (2006) in a weakly dependent setting, is that they alleviate problems with non-monotonic power, if the true number of breaks is larger than what was implemented in the testing procedure, e.g., if we implement  $\mathcal{P}_n(1)$  when  $q = 2$ . Moreover, the power of the former is almost as high as what can be achieved by  $\mathcal{P}_n(q)$  in cases where  $q$  is known.

Next, as for our sup-Wald test in equation (19), the properties of the  $\mathcal{P}_n(q)$  and  $\mathcal{D}_n(q_{\max})$  classes of structural change tests are characterized under the null hypothesis in the following theorem.

**Theorem 4.** *Suppose the conditions of Theorems 1-2 hold. Let  $\zeta = \kappa/2 + (1 - \kappa)b$  and*

$$Q_{k_{\beta}}(\varphi, \zeta, g) = \frac{\tilde{\mathbf{Y}}_{k_{\beta}}(\varphi, \zeta, g)' \tilde{\mathbf{Y}}_{k_{\beta}}(\varphi, \zeta, g)}{\left( (\varphi_{g+1} - \varphi_g)^{2\zeta} + (\varphi_{g+2} - \varphi_{g+1})^{2\zeta} \right) \left( (\varphi_{g+1} - \varphi_g)(\varphi_{g+2} - \varphi_{g+1}) \right)}, \quad \text{where}$$

$$\tilde{\mathbf{Y}}_{k_{\beta}}(\varphi, \zeta, g) = \left( \Delta \mathbf{Y}_{k_{\beta}}(\varphi_{g+1})(\varphi_{g+2} - \varphi_{g+1})^{\zeta+1/2} - \Delta \mathbf{Y}_{k_{\beta}}(\varphi_{g+2})(\varphi_{g+1} - \varphi_g)^{\zeta+1/2} \right),$$

for  $g = 0, \dots, q-1$ , and with  $\Delta \mathbf{Y}_{k_\beta}(\varphi_{g+1}) = \mathbf{Y}_{k_\beta}(\varphi_{g+1}) - \mathbf{Y}_{k_\beta}(\varphi_g)$ , with  $\mathbf{Y}_{k_\beta}(\varphi_g)$  containing the first  $k_\beta$  elements of the vector of independent Brownian motions,  $\mathbf{Y}(\varphi_g)$ . Moreover, define,

$$\mathcal{Q}_{k_\beta}(\varphi, \zeta, q) = \frac{1}{qk_\beta} \sum_{g=0}^{q-1} \mathcal{Q}_{k_\beta}(\varphi, \zeta, g).$$

Then, the following convergence results hold, for all  $\varphi \in \Phi$ ,

$$\mathcal{P}_n(q) \xrightarrow{\mathbb{D}} \sup_{\varphi \in \Phi} \mathcal{Q}_{k_\beta}(\varphi, \zeta, q), \quad \mathcal{D}_n(q_{max}) \xrightarrow{\mathbb{D}} \max_{1 \leq q \leq q_{max}} \sup_{\varphi \in \Phi} \mathcal{Q}_{k_\beta}(\varphi, \zeta, q).$$

Theorem 4 shows that the LCM-based sup-Wald tests for multiple structural breaks with unknown break dates deliver an asymptotic distribution theory, which is equivalent to the corresponding limits in Bai & Perron (1998, Proposition 6), despite our more challenging setting. The latter is readily seen by each partition satisfying distributional equivalence with a multi-break extension of representation (20). Hence, we may apply the critical values in Bai & Perron (1998, Table 1) to implement our LCM-based tests. Moreover, the power of the latter against alternatives of the form  $\mathcal{H}_{A2}$  follow by the inequalities  $\mathcal{P}_n(1) \leq 2\mathcal{P}_n(2) \leq q\mathcal{P}_n(q)$ , and consistency of our test in equation (19).

The results in Theorems 3 and 4 significantly broaden the applicability of sup-Wald tests for structural changes and parameter instability. Contrary to prior studies, such as Andrews (1993), Hansen (1992), Bai & Perron (1998), Hansen (2003), and Kerjiwal & Perron (2010), we accommodate the following dynamic features: (i) all variables of the system may be fractionally integrated of potentially different orders; (ii) the predictive relation may be fractionally cointegrated, covering  $I(1)$ - $I(0)$  cointegration as a special case; (iii) we remain agnostic about the persistence and cointegration properties, ex ante.<sup>4</sup> These empirically relevant extensions are rendered feasible by our utilization of LCM, and its Gaussian limit theory, in the design of the tests.

**Remark 3.** *An interesting alternative to applying the critical values in Andrews (1993) and Bai & Perron (1998) for structural change testing is to use bootstrap techniques such as those in Diebold & Chen (1996) and Hansen (2000), who document favorable size properties of their tests. Moreover, Georgiev et al. (2018) and Boldea, Cornea-Madeira & Hall (2019) show that bootstrap procedures are critical for valid testing of parameter instability in OLS regressions, if the predictors are persistent, modeled as local-to-unity processes, and endogenous, respectively. In contrast, we accommodate persistence through reliance on the flexible class of fractionally integrated processes, cointegration and endogenous regressors (cf., Remark 1). The primary distinction of our approach from theirs is that they correct biases in the asymptotic theory for OLS in such settings, whereas we eliminate corresponding biases via the LCM procedure, rendering our asymptotic theory Gaussian.*

<sup>4</sup>See also the recent review in Casini & Perron (2019) for a discussion of the state of the literature.

## 5 Simulation Study

This section complements the asymptotic analysis by exploring the finite sample properties of the LCM-based sup-Wald test for parameter instability,  $\mathcal{P}_n(1)$ , through realistic simulations. Specifically, Section 5.1 describes the simulation setup, which is designed to mimic quantitative features of our empirical application in Section 6. Whereas Section 5.2 and 5.3 provide the main numerical results, we have included additional tables in Appendix B to supplement the analysis

### 5.1 Simulation Setting

We consider testing for parameter instability in an setting reminiscent of the ones in Hong (1996), Shao (2009) and Andersen & Varneskov (2020a), albeit allowing for time-varying predictive power of the regressors. In particular, we restrict attention to the “pure” structural break model,  $\mathcal{B}_t = \beta_t$ , and consider the univariate regressor case, thus writing  $x_{t-1}$  and  $\beta_t$  for the predictor and regression coefficient, respectively. Then, we generate fractional ARMA(1, 0) processes as,

$$(1 - L)^{d_1}(y_t - \mu_y) = \beta_t u_{t-1} + (1 - L)^b \eta_t, \quad (1 - L)^{d_2}(x_{t-1} - \mu_x) = u_{t-1}, \quad (25)$$

where  $\eta_t$  and  $u_{t-1}$  are AR(1) processes with persistence parameters  $\phi_\eta = \phi_u = 0.2$  and mean-zero Gaussian errors with volatility  $\sigma_\eta = 0.5$  and  $\sigma_u = 1$ , respectively. Moreover, we fix the means  $\mu_y = \mu_x = 1/2$ , the integration orders  $d_1 = d_2 = d$  and consider selections  $d \in \{0.30, 0.60, 1.00\}$  as well as models with ( $b = d$ ) and without ( $b = 0$ ) fractional cointegration. These configurations nest long memory regressions with variables that are (asymptotically) stationary, non-stationary and it covers standard unit root regressions and cointegration models. For each specification, we consider sample sizes  $n \in \{350, 700\}$ , noting that the case  $d = b = 0.60$  and  $n = 350$  corresponds well with our empirical application to the IV-RV forecast relation in Section 6.

To analyze the size and power properties of  $\mathcal{P}_n(1)$ , we simulate the system (25) under the null hypothesis without breaks in the predictive relation  $\mathcal{H}_0 : \beta_t = \beta = 1$  as well as examine various alternatives. For specificity, the alternative hypotheses stipulate a single break in the regression coefficient, that is, we consider parameters of the form  $\beta_t(\varphi_1) = \beta_0 \mathbf{1}_{\{t \in \mathcal{T}_0(\varphi_1)\}} + \beta_1 \mathbf{1}_{\{t \in \mathcal{T}_1(\varphi_1)\}}$ , for different values of the break fraction,  $\varphi_1$ , and regime-dependent coefficients,  $\beta_0$  and  $\beta_1$ . Our *baseline* set of alternatives feature  $\beta_0 = 0.50$ ,  $\beta_1 = 1.50$  as well as  $\mathcal{H}_1 : \varphi_1 = 0.25$ ,  $\mathcal{H}_2 : \varphi_1 = 0.50$ , and  $\mathcal{H}_3 : \varphi_1 = 0.75$ , thus capturing sizable breaks in the beginning, middle and end of the sample. In addition, we consider an equivalent set of *small* break alternatives, where  $\beta_0 = 0.90$ ,  $\beta_1 = 1.10$  and  $\mathcal{H}_4 : \varphi_1 = 0.25$ ,  $\mathcal{H}_5 : \varphi_1 = 0.50$  and  $\mathcal{H}_6 : \varphi_1 = 0.75$ , thus providing a very challenging set of tests in finite samples.<sup>5</sup>

We need to select an estimator of the fractional integration order along with a few tuning parameters

<sup>5</sup>In addition to the single break models, we have considered three different alternatives. First,  $\beta_t = 0.50 \times \mathbf{1}_{\{t \leq \lfloor 0.25n \rfloor\}} + \mathbf{1}_{\{\lfloor 0.25n \rfloor < t \leq \lfloor 0.75n \rfloor\}} + 1.50 \times \mathbf{1}_{\{t > \lfloor 0.75n \rfloor\}}$  has two structural breaks. Second,  $\beta_t = 0.50(1 - t/n) + 1.50(t/n)$  captures a trending regression coefficient. Finally, we have examined a predictive relation with regression parameter exhibiting periodic fluctuation of the form  $\beta_t = 1 + \sin(t/(n/10))$ . The power results are indistinguishable from those reported for the baseline set of alternatives and, thus, omitted for brevity.

to implement the LCM-based sup-Wald test. First, we carry out the fractional filtering and estimation of the cointegration strength parameter using the TELW estimator of Andersen & Varneskov (2020a), as it applies to all of our testing scenarios (i.e., values of  $d$  and  $b$ ).<sup>6</sup> To this end, we follow the latter and use trimming  $\ell_d = \lfloor n^{0.20} \rfloor$  and a bandwidth  $m_d = \lfloor n^{0.75} \rfloor$ . In addition, we will initially use the tuning parameters  $\nu = \nu_G = 0.20$ ,  $\kappa = 0.70$  and  $\kappa_G = 0.90$  for all configurations of the predictive model. To verify that these selections, indeed, correspond well with the restrictions implied by Assumption T and equation (12), we have explicated the requisite bounds on the trimming and bandwidth rates for the different values of  $d$  and  $b$  in Table 1. Moreover, we examine to what extent the tuning parameter selections are robust in Section 5.3. Finally, we set  $\epsilon = 0.20$ , controlling the length of the search space for the sup-Wald test,  $\Phi$ , the nominal size of the test to 5% (again, see Andrews (1993, Table 1) for critical values) and consider simulations with 1000 replications.

## 5.2 Simulation Results

The size and power results for the *baseline* alternatives are reported in Table 2 and deserve a few comments. First, the size of the test in the absence of cointegration ( $b = 0$ ) is 6.9-8.5%, thus close to the nominal 5% level. Moreover, it improves for the cointegration case ( $b = d$ ) when  $d \leq 0.60$  due to a faster rate of convergence of the LCM estimator (cf. Theorem 1), whereas the test is slightly conservative for unit root persistence,  $d = 1.00$ . Second, the test is very powerful, as evidenced by the 100% rejection rates for all *baseline* alternatives. The corresponding rejection rates for the *small* break alternatives in Table 3 underscores its considerable finite sample power, which improves when the sample size is increased from  $n = 350$  to  $n = 700$ . These results further show that the test has higher power to detect breaks in the middle of the sample than in the beginning or the end, for which, however, the rejection rates remain high. Finally, the simulations show that the power is generally higher for cointegrating regressions, not surprisingly, as these induce a faster rate of convergence.

The supplementary tables in Appendix B show that the full sample LCM estimate converges to a weighted average of the regime-dependent coefficients under the break alternative and that breaks generates higher variance of the coefficient estimates. Moreover, they demonstrate that the LCM-based sup-Wald procedure precisely estimates the break fraction for the *baseline* alternatives, yet have a bias for the *small* break alternatives, consistent with the power properties of the test.

## 5.3 Tuning Parameter Robustness

To assess the robustness of the LCM-based sup-Wald test for parameter instability,  $\mathcal{P}_n(1)$ , with respect to its tuning parameters, we consider various combinations with  $\nu \in \{0.10, 0.20, 0.30\}$ ,  $\kappa \in \{0.50, 0.60, 0.70, 0.75\}$  and  $\kappa_G \in \{0.75, 0.80, 0.90\}$ , while fixing the remaining TELW configurations. Moreover, in order keep the exposition simple, we focus on predictive regressions where  $d = 0.60$  and  $n = 350$ , which, as mentioned above, is closely aligned with our empirical application below, and we

<sup>6</sup>Note that estimation of the integration order and fractional filtering are performed using the full sample of observations before implementing the sup-Wald testing procedure on different partitions of the observations.

consider settings with ( $b = d$ ) and without ( $b = 0$ ) cointegration. The results are provided in Tables 4 and 5 for the *baseline* and *small* break alternatives, respectively. Whereas the properties of the sup-Wald testing procedure is robust to changes in the trimming rate,  $\nu$ , as well as the bandwidth of the long-run covariance estimator,  $\kappa_G$ , we find the test properties to display some sensitivity to the MBLS bandwidth parameter,  $\kappa$ . Specifically, when  $\kappa = 0.50$ , we observe both size distortions and lower power of the procedure. The finite sample properties improve for  $\kappa = 0.60$ , and the test performs very well when  $\kappa \geq 0.70$ . This pattern is not particularly surprising given that the effective number of Fourier frequencies used to construct the MBLS estimator and sup-Wald test statistic over each subinterval is substantially smaller when  $\kappa = 0.50$  compared with  $\kappa = 0.70$ , as evidence by the corresponding “effective” full sample sizes of frequencies  $\lfloor 350^{0.50} \rfloor = 18$  and  $\lfloor 350^{0.70} \rfloor = 60$ , respectively.

Generally, we find the LCM-based sup-Wald test,  $\mathcal{P}_n(1)$ , to perform very well across realistic persistence configurations and sample sizes. Moreover, the procedure is robust to changes tuning parameters, except when selecting the bandwidth parameter,  $\kappa$ , too low.

## 6 Empirical Analysis: Stability of the IV-RV Relation

As detailed in the introduction, and going back to Andersen & Bollerslev (1997), many empirical studies have documented the presence of long memory persistence in RV. Moreover, there is an extensive literature finding: (i) IV is an unbiased predictor of future RV after correcting for various methodological issues, and it subsumes the information content in historical RV, e.g., Christensen & Prabhala (1998) and Jiang & Tian (2005); (ii) IV and RV are fractionally cointegrated and their long-run relation is unbiased, e.g., Christensen & Nielsen (2006) and Bandi & Perron (2006); (iii) the fractionally cointegrating equilibrium between IV and RV carry information about the variance risk premium and forecast asset returns, e.g., Bollerslev et al. (2013) and Li et al. (2020).

In this section, we revisit the long-run relation between IV and RV by first examining the relative efficiency of the long-run information in IV for RV, comparing it to forecasts constructed from a long-memory model exploiting high-frequency data, thereby providing a more challenging benchmark than historical volatility, which the literature typically has relied upon. Second, we examine the stability of the fractional cointegrating relations using our new LCM-based sup-Wald tests. Finally, we explore the impact of “instability” on the forecasting prowess of the variance risk premium for returns.

### 6.1 Data and Summary Statistics

Our main dataset consists of three monthly series; implied volatility (IV) and realized volatility (RV) on the S&P 500 as well as a long-memory (LM) forecast of RV. We construct the RV series from one-minute observations of S&P 500 futures over January 1988 - December 2019. On each trading day, we use the flat-top realized kernel estimator of Varneskov (2016, 2017) to compute the intra-day quadratic variation, and we add the squared close-to-open return from the preceding trading day to account for the overnight return variation. We construct the RV series by summing all daily quadratic variation

estimates across the trading days in a given month, then apply a square-root transform, and annualize. The IV is proxied by the VIX. Finally, the LM forecasts are obtained using an ARFIMA(1,  $d$ , 0) model. Specifically, we fit the model to daily log-transformed quadratic variation estimates and apply a log-normal correction to map the forecast back into levels. The latter is motivated by Andersen, Bollerslev, Diebold & Ebens (2001) and Andersen, Bollerslev, Diebold & Labys (2001, 2003), who show that RV is approximately log-normal and that long-memory models capture their dynamic properties well.<sup>7</sup> The ARFIMA model is estimated using a two-year rolling window.<sup>8</sup> The first two years of the sample are used to initialize the LM forecasts, so our final dataset spans January 1990 through December 2019, featuring  $n = 360$  monthly observations. The series and their autocorrelation functions are depicted in Figure 1. Clearly, they are all highly persistent and strongly comoving, consistent with a large literature stipulating that IV and RV are fractionally cointegrated.

We further assess the properties of the series in Table 6, providing standard unconditional summary statistics and estimates of their degree of fractional integration. For the latter, we exploit the TELW estimator of Andersen & Varneskov (2020a) and the local polynomial Whittle with noise (LPWN) estimator by Frederiksen, Nielsen & Nielsen (2012), which is more robust against short-memory dynamics and measurement errors. The TELW estimator is implemented using a trimming parameter  $\ell_d = \lfloor n^{0.30} \rfloor$  and a bandwidth  $m_d = \lfloor n^{0.75} \rfloor$ , whereas the LPWN estimator is implemented with a bandwidth  $m_d = \lfloor n^{0.85} \rfloor$  and short-memory polynomials of orders zero and one. From Table 6, we observe that the unconditional mean of IV is larger than for RV, consistent with the presence of a volatility risk premium; see, e.g., Carr & Wu (2009) and Bollerslev et al. (2011). Second, RV is more volatile and has fatter tails than LM and IV, consistent with the former being a realization and the latter capturing volatility expectations. Finally, we confirm that the series are fractionally integrated, with orders falling in the range 0.5-0.6, when assessed through their LPWN estimates.

## 6.2 Econometric Issues and Implementation

There is an extensive literature exploring the IV–RV relation and correcting for various methodological issues. Although our exposition focuses on VIX, our adopted proxy for the risk-neutral expectation of IV, the following decomposition summarizes the econometric issues more broadly. We define the volatility risk premium for the coming month to be the difference between the expectation of RV under the risk-neutral measure,  $\mathbb{Q}$  and the objective measure,  $\mathbb{P}$ , yielding,

$$\text{VIX}_t = \mathbb{E}_t^{\mathbb{Q}}[\text{RV}_{t+1}] + \mathcal{E}_t = \mathbb{E}_t^{\mathbb{P}}[\text{RV}_{t+1}] + \underbrace{\mathbb{E}_t^{\mathbb{Q}}[\text{RV}_{t+1}] - \mathbb{E}_t^{\mathbb{P}}[\text{RV}_{t+1}]}_{\text{volatility risk premium}} + \mathcal{E}_t, \quad (26)$$

where  $\mathcal{E}_t$  captures a Jensen’s inequality bias from the square-root transformation and measurement errors in VIX, generated by a (time-varying) truncation of the strike range in the construction of the

<sup>7</sup>To corroborate this, we report unconditional summary statistics for both RV and log-RV in Table 6 below, which show that the logarithmic transformation alleviates skewness and kurtosis, rendering the series much closer to Gaussian.

<sup>8</sup>We have tested several different window lengths. The results are robust and, thus, omitted for brevity.

index and observation errors in the underlying options; see, e.g., Jiang & Tian (2005), Andersen & Bondarenko (2007), Andersen, Bondarenko & Gonzalez-Perez (2015) and Andersen, Fusari, Todorov & Varneskov (2020).<sup>9</sup> Hence, we have two sources of errors; one arising from standard measurement errors and another from the volatility risk premium. The latter is known to be positive on average (per our definition above), correlated with the level of volatility and weakly dependent; see, e.g., Bandi & Perron (2006), Chernov (2007), Bollerslev, Tauchen & Zhou (2009), Carr & Wu (2009) and Bollerslev et al. (2011). Next, we complement equation (26) with a decomposition of the latent  $\mathbb{P}$ -measure forecast of RV,

$$\mathbb{E}_t^{\mathbb{P}}[\mathbf{RV}_{t+1}] \equiv \mathcal{V}_{t,1} + \mathcal{V}_{t,2}, \quad (27)$$

where  $\mathcal{V}_{t,1} \in I(d_{\text{VIX}})$  and  $\mathcal{V}_{t,2} \in I(d_{\text{RES}})$  with  $0 \leq d_{\text{RES}} < d_{\text{VIX}}$ . That is, the latent RV forecast embedded in VIX has a two-factor structure, and the most persistent component is fractionally integrated, as documented in Table 6.<sup>10</sup> The decompositions imply that VIX is fractionally integrated of order  $I(d_{\text{VIX}})$ , with the remaining components being of strictly smaller orders (often assumed to be  $I(0)$ ). This motivated Bandi & Perron (2006), Christensen & Nielsen (2006), Nielsen (2007), Kellard et al. (2010), and Nielsen & Frederiksen (2011) to assess whether IV is “long-run unbiased” for RV, i.e., whether the most persistent forecast component,  $\mathcal{V}_{t,1}$ , achieves a coefficient  $\mathcal{B}_{\text{IV}} = 1$ . These studies take advantage of the differing persistence among the  $\text{VIX}_t$  components and use NBLs-type estimators to identify the accuracy of the most persistent component. By the same logic, Bollerslev et al. (2011), Bollerslev et al. (2013), Osterrieder et al. (2019) and Li et al. (2020) specify fractional cointegration models between IV and RV to estimate the volatility risk premium component associated with this long-run equilibrium and use it to forecast asset returns.

While these studies, indeed, demonstrate that IV is long-run unbiased, they do not assess the efficiency of its predictions for the low-frequency movements in RV.<sup>11</sup> Our analysis goes one step further by examining the relative information embedded in the persistent IV component against a competing ARFIMA forecast constructed using high-frequency data, in addition to their associated biases. Moreover, we use our new testing framework to assess whether these long-run relations, exploring bias and efficiency, are invariant over time. To this end, we test for parameter (in)stability using our new LCM-based sup-Wald test,  $\mathcal{P}_n(1)$ . As explained above, the LCM procedure is robust to errors-in-variables (cf., Remark 1), as long as they are less persistent than the signal of interest. Hence, it facilitates robust inference and testing in this fractional cointegration setting.

We implement the LCM procedure using the LPWN estimator in the fractional filtering step with a bandwidth  $\lfloor n^{0.85} \rfloor$ , since it achieves a slightly faster convergence rate than TELW, and all the variables have  $d < 1$ . In addition, we follow existing semi-parametric studies of fractional cointegration and take the average integration order as a global  $d$  estimate. Table 6 verifies that this is innocuous, as

<sup>9</sup>The Jensen’s inequality bias arises, because  $\text{VIX}_t^2$  captures expectations about  $\mathbb{E}_t^{\mathbb{Q}}[\mathbf{RV}_{t+1}^2]$ .

<sup>10</sup>Naturally,  $\mathbb{E}_t^{\mathbb{P}}[\mathbf{RV}_{t+1}]$  may have more than two components, as long as  $\mathcal{V}_{t,1}$  remains the more persistent one.

<sup>11</sup>Among others, Christensen & Prabhala (1998), Jiang & Tian (2005) and Chernov (2007) find IV to subsume the information content in historical volatility for S&P 500, applying inference techniques resting on weakly dependent dynamics. Chraoenwong et al. (2009), Taylor et al. (2010) and Cheng & Fung (2012) extend these findings to other assets.

the individual estimates are close for LPWN. The MBLs estimator in the second step is configured with trimming and bandwidth rates  $(\nu, \kappa) \in \{(0.2, 0.5), (0.2, 0.6), (0.3, 0.6), (0.3, 0.7)\}$ , both for the full sample fractional cointegration analysis and for the LCM-based sup-Wald testing. Whereas the choice of  $\nu$  is well-aligned with the recommendation in Andersen & Varneskov (2020a), the bandwidth rate is slightly smaller to emphasize robustness rather than efficiency, when estimating the cointegrating relation, by alleviating the impact of high-frequency components. These choices are in line with the values considered in the simulation study as well as the bounds provided by Table 1.<sup>12</sup>

In computing the asymptotic variance, we further set  $\nu_G = \nu$  and  $\kappa_G = \kappa + 0.2$ , and use the TELW estimator to find  $b$  with corresponding trimming and bandwidth rates,  $\nu_b = 0.30$  and  $\kappa_b = 0.75$ . The  $\mathcal{P}_n(1)$  test is applied with  $\epsilon = 0.2$ , determining the search space  $\Phi$ , to initialize the LCM estimates in the beginning and the end of the sample. The test results are, however, robust to this selection.

### 6.3 Implied and Realized Volatility

First, we carry out a full sample fractional cointegration analysis of the predictive relation between RV and either IV or LM. The regression results, assessing whether the two respective forecasting procedures are long-run unbiased, are reported in Table 7. The IV results, in particular, confirm the findings from the literature, demonstrating that its relation to RV is fractionally cointegrated with strength  $0.4 \leq b \leq 0.5$ . Moreover, we cannot reject that IV is long-run unbiased,  $\mathcal{B}_{IV} = 1$ . Note that the bounds on  $b$  imply convergence rates of  $\mathcal{S}_n(b) \asymp n^{\kappa/2 + \frac{2}{5}(1-\kappa)}$  and  $\mathcal{S}_n(b) \asymp \sqrt{n}$ , respectively, as cointegration improves on the semi-parametric rate,  $\sqrt{m}$ , obtained for  $b = 0$ . The corresponding LCM results for the LM forecasts similarly demonstrate fractional cointegration with equivalent strength. However, these forecasts may be slightly upward biased at the lowest frequencies, as the Wald test rejects  $\mathcal{B}_{LM} = 1$  at a 5% significance level, when  $\kappa < 0.7$ . This bias disappears for  $\kappa = 0.7$ .

Second, we provide full sample multivariate predictive regression results in Table 8. These show the LM forecasts emerging as more efficient than IV for capturing the low-frequency movements in RV. In particular, we reject  $\mathcal{B}_{LM} = \mathcal{B}_{IV}$  at a 5% significance level for all combinations of tuning parameters and even fail to reject  $(\mathcal{B}_{LM}, \mathcal{B}_{IV}) = (1, 0)$  in two of the four configurations. While these results appear to contradict Christensen & Prabhala (1998), Jiang & Tian (2005) and Chernov (2007), who find IV to be informationally efficient when assessed against historical volatility, it is important to note the subtle difference in interpretation. Our analysis only speaks to the relative quality of the low-frequency signals in the variables. Hence, IV may have important high-frequency information, which is not captured by our cointegration analysis. Our empirical results, on the other hand, demonstrate that the most persistent component of IV may be well-approximated by an ARFIMA(1,  $d$ , 0) process, generating more efficient predictions of the low-frequency variation in RV relative to IV.

Third, we test the stability of the predictive relations using the LCM-based sup-Wald test, with

<sup>12</sup>Whereas the lower bandwidth provides unbiased, yet less efficient, full sample LCM estimates, as seen in Andersen & Varneskov (2020a), our simulations show that the sup-Wald test will exhibit poor size properties when  $\kappa = 0.50$ . However, as the sup-Wald test size and the LCM coefficient estimate are both accurate when  $\kappa = 0.70$ , the use of multiple bandwidth parameters will allow us to assess robustness across different frequency ordinates.



results reported in Table 9. Importantly, the test overwhelmingly rejects the null hypothesis (4) for both univariate and multivariate predictions.<sup>13</sup> These rejections imply that full sample coefficient estimates obtained from the LCM procedure in Tables 7-8 will reflect the average coefficients across regimes in the sample, as conveyed by Tables 11 and 15 in Appendix B (from the simulation study). Unfortunately, they will conceal any regime-specific information from the forecasting variables over the course of the sample. Hence, to explore the nature of the rejections in greater detail, we depict rolling window LCM coefficient estimates (with 95% confidence bands) in Figure 2. The window length is five years ( $n = 60$ ), and LCM is implemented with  $(\nu, \kappa) = (0.3, 0.6)$ . The figure illustrates parameter instability associated with the Asian crisis in 1997, the global financial crisis 2008-2009, and the low-volatility bull-market environment after 2013. In fact, from the rolling multivariate regressions we observe that both LM and IV carry important low-frequency information about RV during the periods from 1990 through 2004 and again from 2014 through 2019. However, the low-frequency information embedded in IV is completely subsumed in the period leading up to and including the financial crisis. These results have implications for volatility forecasting, indicating a shift in the low-frequency RV characteristics over the 30-year period, particularly during periods of financial turmoil.

Finally, these results have ramifications for asset pricing and return predictions. Specifically, most studies measure the volatility risk premium as  $VRP_t = IV_t - RV_t$ . Other studies estimate the corresponding cointegrating relation between  $IV_t$  and  $RV_t$  to more accurately recover the “long-run” component of the volatility risk premium; see, e.g., Bollerslev et al. (2011), Bollerslev et al. (2013), Osterrieder et al. (2019) and Li et al. (2020).<sup>14</sup> Our rejection of stability of the cointegration models for RV, thus, suggest that such measures may convey different information across predictive regimes. That is, since the equilibrium cointegrating relation and “long-run” risk premium may differ across the sample, the forecast variable  $VRP_t$  may contain different information across such regimes, being a mixture of long- and short-run components. As a simple gauge of this conjecture, we predict monthly returns for the S&P 500 using the  $VRP_t$  measure and standard OLS regressions.<sup>15</sup> Specifically, we consider full sample predictions as well as subsamples from 1990.1-2004.12, 2005.1-2013.12 and 2014.1-2019.12, where IV either provides incremental low-frequency information about RV (first and third interval) or is subsumed by LM (second interval). Summary statistics for the return and  $VRP_t$  series are provided in Table 10 together with the predictive regressions results. Interestingly, we find that all predictive power from  $VRP_t$  materializes during the subsample 2005.1-2013.12, where the low-frequency information from IV is subsumed by time series forecasts. This suggests that it is the information from IV associated with high-frequency movements in RV, and thus the short-run component of the volatility

<sup>13</sup>Hence, despite the size distortions of the sup-Wald test  $\mathcal{P}_n(1)$  for  $\kappa = 0.50$ , the results using the remaining tuning parameter configurations provide strongly significant evidence against stability of the cointegration parameter.

<sup>14</sup>Importantly, this does not change the definition of the volatility risk premium,  $\mathbb{E}_t^{\mathbb{Q}}[RV_{t+1}] - \mathbb{E}_t^{\mathbb{P}}[RV_{t+1}]$ . Indeed, the alternative ways of constructing the empirical forecast measures reflect an emphasis on identifying different components of the risk premium; namely, those associated with variation operating at different frequencies.

<sup>15</sup>Note that we implement the predictive return regressions assuming that both returns and the volatility risk premium are weakly dependent. A rigorous treatment of return regression using fractionally integrated forecast variables is given by Andersen & Varneskov (2020b). A related analysis of such issues is beyond the scope of this paper.

risk premium, which generates significant forecasting power for returns.

Hence, in addition to accommodating regime-dependent low-frequency dynamics in forecasting models for RV, our empirical results suggests that a potentially fruitful avenue for future research would be to extend existing asset pricing models with a volatility risk premium to allow for regime-dependent cointegration between IV and RV as well as corresponding return predictability. Moreover, since, among others, Bollerslev & Todorov (2011), Andersen, Fusari & Todorov (2015) and Kilic & Shaliastovich (2019) have shown that the volatility risk premium is mainly associated with downside risk and/or left-tail jumps, it would be interesting to examine potential parameter instabilities in such persistent forecast variables as well. Again, this is left for future research.

## 7 Conclusion

This paper develops parameter instability and structural change tests for predictive regressions in economic systems governed by persistent vector autoregressive dynamics. Specifically, in a setting where all – or a subset – of the variables may be fractionally integrated, and the predictive relation may feature cointegration (possibly unknown, *ex-ante*), we provide sup-Wald break tests using the LCM approach of Andersen & Varneskov (2020*a*). The new tests cover both parameter variation and multiple structural changes with unknown break dates, and allows for the number of breaks to be unknown. The adoption of LCM ensures the tests, asymptotically, avoid problems associated with spurious inference, regression balance and endogenous variation in the predictors, which typically arise in such settings. We establish limit theory for the tests, showing it coincides with those for standard break testing procedures in, e.g., Andrews (1993) and Bai & Perron (1998), implying that existing critical values may be applied, without modification. Hence, the tests are easy to implement.

We apply the new structural change tests to explore the stability of the fractionally cointegrating relation between option implied and realized volatility. Moreover, we assess the relative efficiency of the low-frequency information embedded in IV forecasts against a challenging benchmark, a time series long-memory model-based forecast constructed from high-frequency data. Unlike existing studies, we find evidence that the IV-RV cointegrating relation is unstable, rejecting the null hypothesis of parameter stability at all conventional significance levels, and that carefully constructed time-series forecasts are more efficient than IV in capturing low-frequency movements in RV. Finally, we show that such parameter instabilities manifest themselves as instabilities in the forecasting relation between asset returns and the volatility risk premium.

Tuning Parameter Bounds in Simulations						
	no cointegration: $b = 0$			cointegration: $b = d$		
	$d = 0.30$	$d = 0.60$	$d = 1.00$	$d = 0.30$	$d = 0.60$	$d = 1.00$
$\nu_{\min}$	0.240	0.180	0.125	0.125 + $0.30(1 - \kappa)$	0.125 + $0.60(1 - \kappa)$	0.125 + $(1 - \kappa)$
$\nu_{\max}$	$\kappa$	$\kappa$	$\kappa$	$\kappa$	$\kappa$	$\kappa$
$\kappa_{\min}$	$\nu$	$\nu$	$\nu$	$\nu \vee 0.375$	$\nu \vee 6/11$	$\nu \vee 2/3$
$\kappa_{\max}$	4/5	4/5	4/5	4/5	4/5	4/5

**Table 1: Tuning parameter bounds.** This table compiles the lower and upper bounds on the trimming and bandwidth rates required by Assumption T for the different configurations of  $d$  and  $b$  considered for the simulation study in Section 5. Specifically, the bounds are derived using  $\varrho = 0.75$  for the TELW estimator. Moreover, for simplicity of exposition, the lower trimming rate bound,  $\nu_{\min}$ , is derived under the assumption  $\kappa \geq 1/2$  to avoid the appearance of multiple terms, as in equation (12). The lower bound  $\nu_{\min}$  for  $d \leq 0.60$  and  $b = 0$  follows from the first restriction in equation (12), that is, from  $(3/5 - (2d - b)/5)/2 < \nu$ , and the remaining cases from  $(1 - \varrho)/2 + b(1 - \kappa) < \nu$ .

Size and Power of the LCM-based Sup-Wald Test: Baseline								
	no cointegration: $b = 0$							
	$\mathcal{H}_0, n =$		$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.0850	0.0690	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$d = 0.60$	0.0850	0.0690	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$d = 1.00$	0.0850	0.0690	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	cointegration: $b = d$							
	$\mathcal{H}_0, n =$		$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.0610	0.0640	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$d = 0.60$	0.0610	0.0470	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$d = 1.00$	0.0260	0.0200	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

**Table 2: Size and power: Baseline.** This table provides size and power results for the LCM-based sup-Wald test (19), which is robust against general forms of parameter instability. Specifically, the test is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ , for  $(\nu, \kappa, \kappa_G) = (0.2, 0.7, 0.9)'$ . Robustness towards the tuning parameters is studied in Table 4. As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability, i.e., it speaks to test size.  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  describes three *baseline* predictive models with breaks, thus illustrating the power of the test. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

Size and Power of the LCM-based Sup-Wald Test: Small Breaks								
no cointegration: $b = 0$								
	$\mathcal{H}_0, n =$		$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.0850	0.0690	0.4740	0.7010	0.5840	0.8110	0.4830	0.7160
$d = 0.60$	0.0850	0.0690	0.4740	0.7010	0.5840	0.8110	0.4830	0.7160
$d = 1.00$	0.0850	0.0690	0.4740	0.7010	0.5840	0.8110	0.4830	0.7160
cointegration: $b = d$								
	$\mathcal{H}_0, n =$		$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.0610	0.0640	0.4870	0.8120	0.6060	0.8770	0.4310	0.7420
$d = 0.60$	0.0610	0.0470	0.5470	0.8850	0.6320	0.9480	0.4310	0.7930
$d = 1.00$	0.0260	0.0200	0.4860	0.8960	0.5630	0.9500	0.3370	0.7440

**Table 3: Size and power: Small breaks.** This table provides size and power results for the LCM-based sup-Wald test (19), which is robust against general forms of parameter instability. Specifically, the test is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ , for  $(\nu, \kappa, \kappa_G) = (0.2, 0.7, 0.9)'$ . Robustness towards the tuning parameters is studied in Table 5. As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability, i.e., it speaks to test size.  $\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6$  describes three predictive models with *small* breaks, thus illustrating the power of the test. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

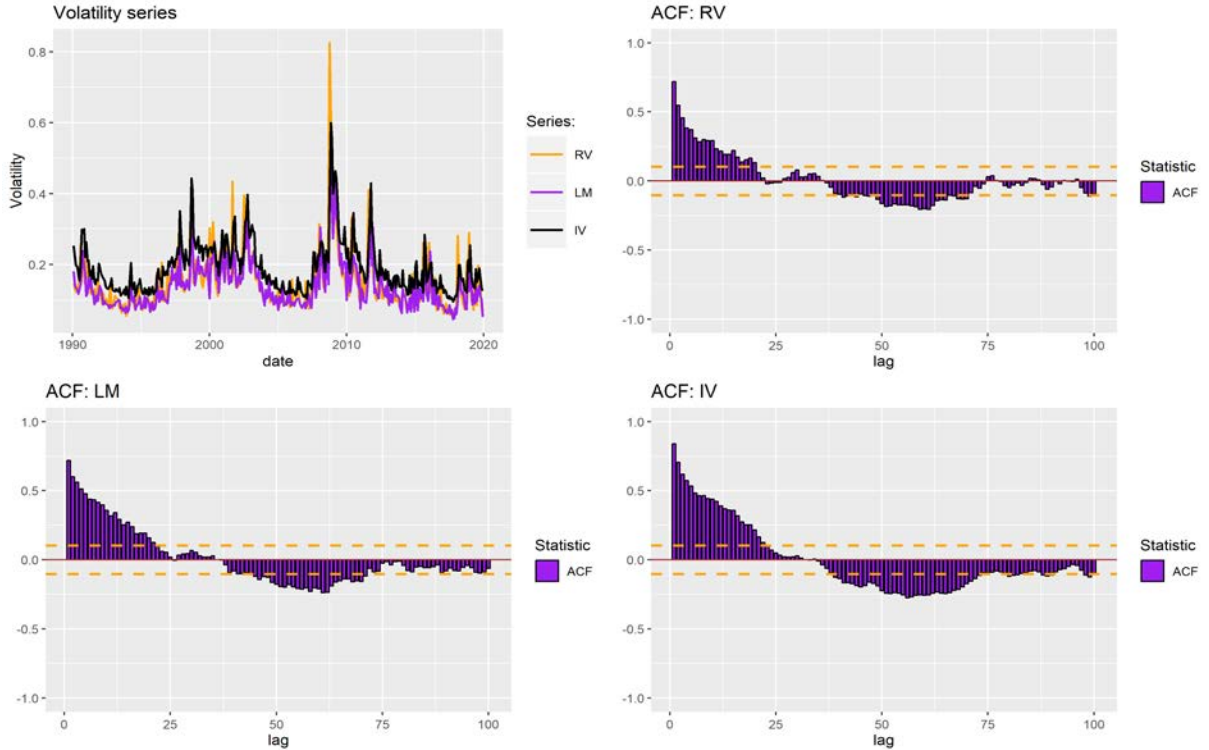
Size and Power of the LCM-based Sup-Wald Test: Tuning Parameters, Baseline								
	no cointegration: $b = 0$				cointegration: $b = d$			
$(\nu, \kappa, \kappa_G) =$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$
(0.10, 0.70, 0.90)	0.0710	1.0000	1.0000	1.0000	0.0530	1.0000	1.0000	1.0000
(0.20, 0.70, 0.90)	0.0850	1.0000	1.0000	1.0000	0.0610	1.0000	1.0000	1.0000
(0.30, 0.70, 0.90)	0.0890	1.0000	1.0000	1.0000	0.0750	1.0000	1.0000	1.0000
(0.20, 0.50, 0.90)	0.1580	1.0000	1.0000	0.9990	0.2020	1.0000	1.0000	1.0000
(0.20, 0.60, 0.90)	0.1090	1.0000	1.0000	1.0000	0.0970	1.0000	1.0000	1.0000
(0.20, 0.75, 0.90)	0.0800	1.0000	1.0000	1.0000	0.0450	1.0000	1.0000	1.0000
(0.20, 0.70, 0.75)	0.0800	1.0000	1.0000	1.0000	0.0710	1.0000	1.0000	1.0000
(0.20, 0.70, 0.80)	0.0750	1.0000	1.0000	1.0000	0.0580	1.0000	1.0000	1.0000

**Table 4: Size and power: Tuning parameters, baseline.** This table provides size and power results for the LCM-based sup-Wald test (19), which is robust against general forms of parameter instability. Specifically, the test is implemented using several combinations of the tuning parameters; trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability, i.e., it provides size results.  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  describes three *baseline* predictive models with breaks, thus speaking to the power of the test. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is  $n = 350$ , the integration order is  $d = 0.60$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

**Size and Power of the LCM-based Sup-Wald Test: Tuning Parameters, Small Breaks**

$(\nu, \kappa, \kappa_G) =$	no cointegration: $b = 0$				cointegration: $b = d$			
	$\mathcal{H}_0$	$\mathcal{H}_4$	$\mathcal{H}_5$	$\mathcal{H}_6$	$\mathcal{H}_0$	$\mathcal{H}_4$	$\mathcal{H}_5$	$\mathcal{H}_6$
(0.10, 0.70, 0.90)	0.0710	0.4710	0.5780	0.4720	0.0530	0.5360	0.6330	0.4210
(0.20, 0.70, 0.90)	0.0850	0.4740	0.5840	0.4830	0.0610	0.5470	0.6320	0.4310
(0.30, 0.70, 0.90)	0.0890	0.4890	0.5950	0.4970	0.0750	0.5450	0.6420	0.4470
(0.20, 0.50, 0.90)	0.1580	0.3350	0.4150	0.3610	0.2020	0.6530	0.7080	0.5060
(0.20, 0.60, 0.90)	0.1090	0.3750	0.4550	0.3780	0.0970	0.5800	0.6680	0.4620
(0.20, 0.75, 0.90)	0.0800	0.5650	0.6870	0.5670	0.0450	0.5560	0.6590	0.4330
(0.20, 0.70, 0.75)	0.0800	0.4810	0.5880	0.4840	0.0710	0.6280	0.7200	0.5190
(0.20, 0.70, 0.80)	0.0750	0.4770	0.5860	0.4720	0.0580	0.5850	0.6880	0.4800

**Table 5: Size and power: Tuning parameters, small breaks.** This table provides size and power results for the LCM-based sup-Wald test (19), which is robust against general forms of parameter instability. Specifically, the test is implemented using several combinations of the tuning parameters; trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability, i.e., it speaks to the test size.  $\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6$  describes three predictive models with *small* breaks, thus illustrating the power of the test. The predictive model is specified with ( $b = d$ ) and without ( $b = 0$ ) cointegration. The sample size is  $n = 350$ , the integration order is  $d = 0.60$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.



**Figure 1: Volatility Series and Autocorrelation Functions (ACFs.)** This figure depicts the series and the ACFs for the realized volatility (RV) series as well as the ARFIMA (LM) and model-free implied volatility (IV) forecasts. All variables are in square-root form. The sample spans January 1990 through December 2019 ( $n = 360$ ). The sample autocorrelation functions are computed for the first 100 lags for each variable in the full sample.

Summary Statistics for Volatility Measures							
	mean	StDev	Skewness	Kurtosis	TELW	LPWN0	LPWN1
RV	0.1509	0.0826	2.9141	18.0456	0.5566 (0.0552)	0.4763 (0.0809)	0.5098 (0.1217)
LM	0.1403	0.0667	1.9758	9.6354	0.6309 (0.0552)	0.6167 (0.0751)	0.6101 (0.1118)
IV	0.1919	0.0737	1.7536	7.7572	0.6957 (0.0552)	0.5904 (0.0720)	0.5423 (0.1181)
log-RV	-1.9978	0.4392	0.6457	3.6146	0.5297 (0.0552)	0.5478 (0.0766)	0.5731 (0.1150)

**Table 6: Summary statistics.** This table displays summary statistics for the realized volatility (RV) series as well as the ARFIMA (LM) and model-free implied volatility (IV) forecasts. Furthermore, we include corresponding summary statistics for log-RV, since the monthly LM forecast are generated from daily log-transformed RV estimates. In addition to the standard unconditional statistics, we present estimates of the fractional integration order for all variables using the trimmed exact local Whittle (TELW) estimator from Andersen & Varneskov (2020a) and the local polynomial Whittle with noise (LPWN) estimator from Frederiksen et al. (2012), with standard errors indicated in parentheses. The TELW estimator is implemented using trimming  $\ell_d = \lfloor n^{0.30} \rfloor$  and bandwidth  $m_d = \lfloor n^{0.75} \rfloor$ . The LPWN estimator is implemented with bandwidth  $m_d = \lfloor n^{0.85} \rfloor$ . The respective numbers for the LPWN signifies the order of the short-memory polynomial. All variables are in square-root form. The sample spans January 1990 through December 2019 ( $n = 360$ ).

Univariate LCM Cointegration Analysis								
	LM, $(\nu, \kappa) =$				IV, $(\nu, \kappa) =$			
	(0.2, 0.5)	(0.2, 0.6)	(0.3, 0.6)	(0.3, 0.7)	(0.2, 0.5)	(0.2, 0.6)	(0.3, 0.6)	(0.3, 0.7)
$\mathcal{B}$	1.1717 (15.0419)	1.1664 (14.7894)	1.1736 (14.8583)	1.0616 (12.5238)	0.9638 (11.8750)	0.9421 (11.2118)	0.9540 (11.3776)	0.8787 (9.6458)
$b$	0.4906 (8.8855)	0.4898 (8.8855)	0.4909 (8.8909)	0.4620 (8.3677)	0.4478 (8.1099)	0.4388 (7.9462)	0.4438 (8.0368)	0.4106 (7.4363)
WALD	4.8587	4.4523	4.8322	0.5288	0.1993	0.4750	0.3010	1.7738
$p$ -WALD	0.0275	0.0349	0.0279	0.4671	0.6552	0.4907	0.5832	0.1829

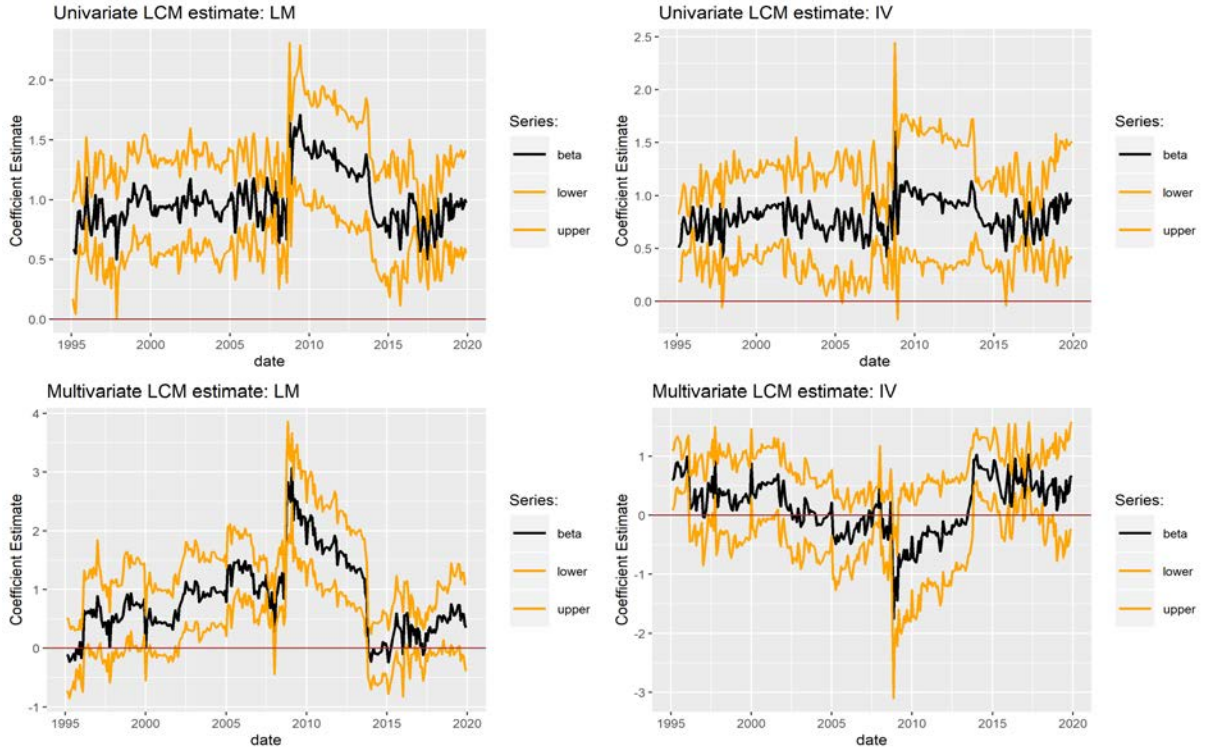
**Table 7: Univariate LCM analysis.** This table displays results from a univariate LCM analysis, where RV is predicted using either ARFIMA (LM) or model-free implied volatility (IV) forecasts. Specifically, the coefficient estimate and associated  $t$ -statistic (in parenthesis) is provided along with a TELW estimate of the fractional cointegration strength,  $b$ , and associated  $t$ -statistics (in parenthesis) and, finally, a Wald test of  $\mathcal{B} = 1$  together with its  $p$ -values. The results are provided for  $\hat{\mathcal{B}}(\ell, m)$ , where several combinations of trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$  parameters are considered. The TELW estimator is implemented as described in Table 6. All variables are in square-root form. The sample spans January 1990 through December 2019 ( $n = 360$ ).

Multivariate LCM Cointegration Analysis									
$(\nu, \kappa) =$	$\mathcal{B}_{LM}$	$\mathcal{B}_{IV}$	$b$	W1	$p$ -W1	W2	$p$ -W2	W3	$p$ -W3
(0.2, 0.5)	1.0627 (6.1677)	0.0969 (0.6130)	0.5077 (9.1946)	4.6694	0.0968	38.198	0.0000	11.783	0.0006
(0.2, 0.6)	0.8389 (6.4918)	0.2905 (2.2244)	0.5097 (9.2310)	5.8705	0.0531	42.219	0.0000	6.3382	0.0118
(0.3, 0.6)	0.8170 (6.3337)	0.3198 (2.4524)	0.5125 (9.2812)	6.9762	0.0306	40.132	0.0000	5.2258	0.0223
(0.3, 0.7)	0.9232 (8.1353)	0.1326 (0.9696)	0.4813 (8.7168)	0.9401	0.6250	67.009	0.0000	14.732	0.0001

**Table 8: Multivariate LCM analysis.** This table displays results from a multivariate LCM analysis, where RV is predicted using ARFIMA (LM) and model-free implied volatility (IV) forecasts. The coefficient estimates and  $t$ -statistics (in parentheses) are provided along with a TELW estimate of the fractional cointegration strength,  $b$ , and associated  $t$ -statistics (in parenthesis). Finally, the Wald test statistics and  $p$ -values from the three hypotheses  $(\mathcal{B}_{LM}, \mathcal{B}_{IV}) = (1, 0)$ ,  $(\mathcal{B}_{LM}, \mathcal{B}_{IV}) = (0, 1)$  and  $\mathcal{B}_{LM} = \mathcal{B}_{IV}$  are reported, labeled W1, W2 and W3 as well as  $p$ -W1,  $p$ -W2 and  $p$ -W3, respectively. The results are provided for  $\hat{\mathcal{B}}(\ell, m)$ , where several combinations of trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$  parameters are considered. The TELW estimator is implemented as in Table 6. All variables are in square-root form. The sample spans January 1990 through December 2019 ( $n = 360$ ).

Testing Stability of IV-RV Relations				
	Sup-Wald test, $(\nu, \kappa) =$			
	(0.2, 0.5)	(0.2, 0.6)	(0.3, 0.6)	(0.3, 0.7)
LM	16.5755***	19.3363***	19.6658***	15.8335***
IV	9.4495**	10.7750**	13.7734***	11.3375**
LM + IV	25.7693***	41.4712***	37.5787***	28.1861***

**Table 9: Stability testing.** This table displays results of the sup-Wald test for stability of univariate and multivariate LCM cointegrating regressions for RV. The latter is predicted using either ARFIMA (LM) and/or model-free implied volatility (IV) forecasts. Specifically, the sup-Wald test is implemented using  $\epsilon = 0.2$ . The critical values of the test may be found in Andrews (1993, Table 1) for both univariate and multivariate regressions, with degrees of freedoms equal to  $k = 1$  and  $k = 2$ , respectively. The results are provided for  $\hat{\mathcal{B}}(\ell, m)$ , where several combinations of trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$  parameters are considered. The TELW estimator is implemented as described in Table 6, and the variables are fractionally filtered using the full-sample global estimate of the integration order prior to sample splits and MBLS estimation. All variables are in square-root form. The sample spans January 1990 through December 2019 ( $n = 360$ ). Finally, \*\* and \*\*\* signify rejection of the null hypothesis,  $\mathcal{H}_0$ , of parameter stability at the 5% and 1% significance level, respectively.



**Figure 2: LCM parameter estimates.** This figure depicts rolling window LCM estimates and 95% confidence bands from regressions of realized volatility (RV) series on ARFIMA (LM) and model-free implied volatility (IV) forecasts. All variables are in square-root form. The sample spans January 1990 through December 2019 ( $n = 360$ ). The window length is 5 years ( $n = 60$ ), and LCM is implemented with  $(\nu, \kappa) = (0.3, 0.6)$ .

Return Predictions using VRP				
Panel A: Summary	mean	StDev	Skewness	Kurtosis
returns	0.0063	0.0412	-0.8109	4.7860
VRP	0.0407	0.0376	-1.5217	11.0565
Panel B: Regressions	$\mathcal{T}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$
constant	-0.0003 (-0.0945)	0.0051 (1.1293)	-0.0093 (-1.8588)	0.0083 (1.5307)
$\mathcal{B}_{\text{VRP}}$	0.1641 (2.5570)	0.0478 (0.6217)	0.3430 (4.9452)	-0.0191 (-0.1656)
adj. $R^2$	0.0170	-0.0098	0.1009	-0.0287

**Table 10: Return regressions.** This table displays summary statistics for monthly S&P 500 returns and volatility risk premium (VRP) estimates in Panel A along with standard predictive return regression results in Panel B. Specifically, the latter is based on OLS estimates and the heteroskedasticity and autocorrelation consistent (HAC) inference procedure by Andrews (1991) using the Parzel kernel and a bandwidth  $[4(n/100)^{2/9}]$ . The summary statistics and regressions results for “ $\mathcal{T}$ ” are based on the full sample from January 1990 through December 2019 ( $n = 360$ ). Regression results are also provided for three subsamples from 1990.1-2004.12, 2005.1-2013.12 and 2014.1-2019.12, labeled  $\mathcal{T}_1$ ,  $\mathcal{T}_2$  and  $\mathcal{T}_3$ , respectively. These subsamples correspond approximately to the periods in Figure 2, where the multivariate coefficient estimate on IV is positive, turns negative, and then back to positive.



## A Proofs

This section contains the proofs of the asymptotic results in the paper. Before proceeding, we introduce some notation. For a generic vector  $\mathbf{V}$ , let  $\mathbf{V}(i)$  index the  $i$ th element, and, similarly, for a matrix  $\mathbf{M}$ , let  $\mathbf{M}(i, q)$  denote its  $(i, q)$ th element. Moreover, denote by  $K \in (0, \infty)$  a generic constant, which may take different values from line to line or from (in)equality to (in)equality. Finally, we remark that sometimes the (stochastic) orders refer to scalars, sometimes to vectors and matrices.

### A.1 Proof of Theorem 1

The proof relies heavily on asymptotic results and stochastic bounds from the corresponding proof of the full sample LCM result in the Online Appendix to Andersen & Varneskov (2020a). We will henceforth refer to the latter as AVOA (2020). First, by applying the decomposition in AVOA (2020, equation (A.2)) in conjunction with AVOA (2020, Lemma A.2(b)), we may write,

$$\sqrt{m_g} \lambda_{m_g}^{-b}(g) \left\| \widehat{\mathbf{B}}(\ell, m, g) - \mathbf{B} - \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m, g)^{-1} \widehat{\mathbf{F}}_{u, \eta}^{(t, 1, b)}(1, m, g) \right\| \xrightarrow{\mathbb{P}} 0, \quad (\text{A.1})$$

for all partitions  $g = 0, \dots, q$ , where  $\widehat{\mathbf{F}}_{u, \eta}^{(t, 1, b)}(1, m, g)$  is the TDAC between  $\mathbf{u}_{t-1}$  and  $\eta_t^{(b)}$ . Moreover, by applying AVOA (2020, Lemma A.2(a)), we have,

$$\left\| \lambda_{m_g}^{-1}(g) \widehat{\mathbf{F}}_{\widehat{u}\widehat{u}}(\ell, m, g) - \mathbf{G}_{uu} \right\| \xrightarrow{\mathbb{P}} 0, \quad (\text{A.2})$$

similarly, for all  $g = 0, \dots, q$ , where  $\mathbf{G}_{uu}$  is positive definite by Assumption D1. Next, define

$$\widetilde{\mathbf{F}}_{u, \eta}^{(t, 1, b)}(1, m, g) \equiv \frac{2\pi}{n_g} \sum_{j=1}^{m_g} \Re \left( \widetilde{\mathbf{I}}_{u\eta}^{(t, 1, b)}(\lambda_j, g) \right), \quad \widetilde{\mathbf{I}}_{u\eta}^{(t, 1, b)}(\lambda_j, g) = \mathbf{w}_u(\lambda_j, g) \lambda_j^b(g) e^{(\pi/2)bi} \bar{w}_\eta(\lambda_j, g), \quad (\text{A.3})$$

where  $\mathbf{w}_u(\lambda_j, g)$  and  $w_\eta(\lambda_j, g)$  are the discrete Fourier transforms of  $\mathbf{u}_{t-1}$  and  $\eta_t^{(b)}$ , respectively. Moreover, define some arbitrary vector  $\boldsymbol{\psi} = (\psi_1, \dots, \psi_k)'$  as well as,

$$\mathcal{U}_4(g, b) \equiv \sum_{t \in \mathcal{T}_g} \boldsymbol{\epsilon}'_t \sum_{s=1}^{t-1} \mathbf{C}_{t-s}^n(g, b) \boldsymbol{\epsilon}_s, \quad \mathbf{C}_t^n(g, b) \equiv \frac{1}{2\pi n_g \sqrt{m_g}} \sum_{j=1}^{m_g} \lambda_j^b(g) \boldsymbol{\chi}_j(g, b), \quad (\text{A.4})$$

where the sequence of coefficients,  $\boldsymbol{\chi}_j(g, b)$ , is defined as  $\boldsymbol{\chi}_j(g, b) \equiv \sum_{i=1}^k \lambda_{m_g}^{-b}(g) \psi_i \mathcal{A}_j(g, b, i)$ , with

$$\begin{aligned} \mathcal{A}_j(g, b, i) \equiv & \Re \left( \left( \mathbf{A}(\lambda_j(g), i)' \bar{\mathbf{A}}(\lambda_j(g), k+1)' e^{-i(t-s)\lambda_j(g)} \right) e^{(\pi/2)bi} \right) \\ & + \Re \left( \left( \mathbf{A}(\lambda_j(g), k+1)' \bar{\mathbf{A}}(\lambda_j(g), i)' e^{i(t-s)\lambda_j(g)} \right) e^{(\pi/2)bi} \right). \end{aligned}$$

Then, by applying AVOA (2020, Lemma A.11), we have,

$$\sqrt{m_g} \lambda_{m_g}^{-b}(g) \left\| \widehat{\mathbf{F}}_{u,\eta}^{(t,1,b)}(1, m, g) - \widetilde{\mathbf{F}}_{u,\eta}^{(t,1,b)}(1, m, g) \right\| \xrightarrow{\mathbb{P}} 0, \quad (\text{A.5})$$

for all partitions  $g = 0, \dots, q$ . Moreover, by using the decomposition in AVOA (2020, Equation (A.16)) in conjunction with the same arguments for their terms  $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3$  and  $\mathcal{U}_4$ , it follows,

$$\left| \sqrt{m_g} \lambda_{m_g}^{-b}(g) \boldsymbol{\psi}' \widetilde{\mathbf{F}}_{u,\eta}^{(t,1,b)}(1, m, g) - \mathcal{U}_4(g, b) \right| \xrightarrow{\mathbb{P}} 0, \quad (\text{A.6})$$

similarly, for all  $g = 0, \dots, q$ . Hence, since  $\mathcal{M}_t(g, b) \equiv \boldsymbol{\epsilon}'_t \sum_{s=1}^{t-1} \mathbf{C}_{t-s}^n(g, b) \boldsymbol{\epsilon}_s$  in  $\mathcal{U}_4(g, b)$  is a martingale difference sequence with respect to the filtration  $\mathcal{F}_{t-1}$ , we use the same arguments as for AVOA (2020, Lemma A.3) to show,

$$\mathcal{U}_4(g, b) \xrightarrow{\mathbb{D}} N \left( 0, \boldsymbol{\psi}' \mathbf{G}_{uu} \boldsymbol{\psi} \frac{G_{\eta\eta}}{2(1+2b)} \right), \quad (\text{A.7})$$

which applies, again, for all  $g = 0, \dots, q$ . Next, as  $\mathcal{M}_t(g, b)$  is a martingale difference sequence and the sample partitions  $\mathcal{T}_g$ ,  $g = 0, \dots, q$ , are non-overlapping, it readily follows that the covariance between  $\mathcal{U}_4(g, b)$  and  $\mathcal{U}_4(h, b)$  is trivial, when  $g \neq h$ . Hence, we may apply (A.1)-(A.7) in conjunction with the continuous mapping theorem, Slutsky's theorem and the Cramér-Wold theorem to show that,

$$\sqrt{m} \lambda_m^{-b} \begin{pmatrix} \widehat{\mathbf{B}}(\ell, m, 0) - \mathbf{B} \\ \vdots \\ \widehat{\mathbf{B}}(\ell, m, q) - \mathbf{B} \end{pmatrix} \xrightarrow{\mathbb{D}} N \left( \mathbf{0}, \widetilde{\boldsymbol{\vartheta}}(b) \circ \mathbf{C}(b) \right), \quad \text{where} \quad (\text{A.8})$$

$$\widetilde{\boldsymbol{\vartheta}}(b) = \text{diag} \left[ (\varphi_1 - \varphi_0)^{2b(\kappa-1)-\kappa}, \dots, (\varphi_{q+1} - \varphi_q)^{2b(\kappa-1)-\kappa} \right], \quad (\text{A.9})$$

holds finite-dimensionally. Next, we establish *tightness* in the sample partition  $\boldsymbol{\varphi}$  of the functional limit sequence. To this end, let us define  $\widetilde{\mathcal{U}}_4(g, b) \equiv (\sqrt{m/m_g} \lambda_{m_g}^b \lambda_m^{-b}) \mathcal{U}_4(g, b)$ . Then, since it follows that  $\sqrt{m/m_g} \lambda_{m_g}^b \lambda_m^{-b} \rightarrow (\varphi_{g+1} - \varphi_g)^{b(\kappa-1)-\kappa/2} > 0$  for all partitions  $g = 0, \dots, q$  on the set  $\Phi$ , we have that the convergence results in (A.1), (A.2), (A.5) and (A.6) still apply when taking the supremum over the partition vector  $\boldsymbol{\varphi}$  on the set  $\Phi$ . Hence, we may study tightness of the joint limit theory via the re-scaled variable  $\widetilde{\mathcal{U}}_4(g, b)$ . By the law of iterated expectations and conditional independence of the martingale difference sequences over disjoint sets of observations, we have

$$\begin{aligned} \mathbb{E} \left[ \left( \widetilde{\mathcal{U}}_4(g, b) - \widetilde{\mathcal{U}}_4(h, b) \right)^2 \right] &\leq K \left( (\varphi_{g+1} - \varphi_g)^{b(\kappa-1)-\kappa/2} - (\varphi_{h+1} - \varphi_h)^{b(\kappa-1)-\kappa/2} \right)^2 \\ &\leq K \left( (\varphi_{g+1} - \varphi_g) - (\varphi_{h+1} - \varphi_h) \right)^{2b(\kappa-1)-\kappa}, \end{aligned} \quad (\text{A.10})$$

where the second inequality follows from the algebraic relation  $|a + b|^p - |a|^p \leq |b|^p$  for  $p \in (0, 1)$ , along with addition and subtraction. Hence, by applying (A.10) together with (A.8) and Ibragimov & Hasminskii (1981, Theorem 20), the sequence is tight and the joint uniform limit theory holds, whose representation in terms of independent Brownian motions follows by Wooldridge & White (1988,

Corollary 4.2), with their conditions D.1 and D.2 being satisfied by (A.8), tightness, and Assumptions D1-D3. Finally, the mutual consistency condition follows exactly as in AVOA (2020, p. 2).  $\square$

## A.2 Proof of Theorem 2

Part (a) follows by applying Assumption B and AVOA (2020, Lemma A.4) in conjunction with the continuous mapping theorem for all partition  $g = 0, \dots, q$  of every sequence of  $\boldsymbol{\varphi}$  on  $\Phi$ . (b) follows by applying (a), Theorem 1, the continuous mapping theorem and Slutsky's theorem.  $\square$

## A.3 Proof of Theorem 3

First, by applying Theorem 1, we may write,

$$\begin{aligned} \mathcal{S}_n(b) \left( \widehat{\boldsymbol{\beta}}(\ell, m, 0) - \widehat{\boldsymbol{\beta}}(\ell, m, 1) \right) &\Rightarrow \mathbf{C}_{k_\beta}(b)^{1/2} \left( \frac{\mathbf{Y}(\varphi_1)}{\varphi_1^{\zeta+1/2}} - \frac{\mathbf{Y}(1) - \mathbf{Y}(\varphi_1)}{(1 - \varphi_1)^{\zeta+1/2}} \right) \\ &= \mathbf{C}_{k_\beta}(b)^{1/2} \left( \frac{\mathbf{Y}(\varphi_1)(1 - \varphi_1)^{\zeta+1/2} - (\mathbf{Y}(1) - \mathbf{Y}(\varphi_1))\varphi_1^{\zeta+1/2}}{\varphi_1^{\zeta+1/2}(1 - \varphi_1)^{\zeta+1/2}} \right) \end{aligned} \quad (\text{A.11})$$

with, again,  $\mathcal{S}_n(b) = \sqrt{m}\lambda_m^{-b}$ , and where  $\mathbf{C}_{k_\beta}(b)$  is the upper-left  $k_\beta \times k_\beta$  submatrix of  $\mathbf{C}(b)$ . Next, by applying Theorem 2 and the continuous mapping theorem,

$$\begin{aligned} \mathcal{S}_n(b)^2 \left( \widehat{\mathbf{AVAR}}_\beta(0, b) + \widehat{\mathbf{AVAR}}_\beta(1, b) \right) \\ \xrightarrow{\mathbb{P}} \left( \frac{1}{\varphi_1^{2\zeta}} + \frac{1}{(1 - \varphi_1)^{2\zeta}} \right) \mathbf{C}_{k_\beta}(b) = \left( \frac{(1 - \varphi_1)^{2\zeta} + \varphi_1^{2\zeta}}{\varphi_1^{2\zeta}(1 - \varphi_1)^{2\zeta}} \right) \mathbf{C}_{k_\beta}(b). \end{aligned} \quad (\text{A.12})$$

Hence, the final convergence results follow by combining (A.11) and (A.12) with the continuous mapping theorem and Slutsky's theorem, for both statements.  $\square$

## A.4 Proof of Theorem 4

The result follows by applying the same arguments used to establish Theorem 3 for each partition  $g = 0, \dots, q$  on  $\Phi$  in conjunction with asymptotic independence between partitions, by Theorem 1.  $\square$

## B Additional Simulation Results

Tables 11-14 provide supplementary results for  $\mathcal{H}_0$  and the *baseline* alternatives  $\mathcal{H}_1$ ,  $\mathcal{H}_2$  and  $\mathcal{H}_3$ . Specifically, they display the mean and standard deviation of the “full sample” LCM coefficient estimates to gauge the impact of breaks under the alternative hypotheses as well as the bias and RMSE for the estimated break fraction,  $\varphi_1$ , using the LCM-based sup-Wald procedure. Tables 15-18 contain corresponding results for the *small* break alternatives.

Mean LCM Estimate: Baseline								
No Cointegration: $b = 0$								
	$\mathcal{H}_0, n =$		$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.9978	0.9983	1.2493	1.2477	0.9955	0.9966	0.7469	0.7463
$d = 0.60$	0.9978	0.9983	1.2493	1.2477	0.9955	0.9966	0.7469	0.7463
$d = 1.00$	0.9978	0.9983	1.2493	1.2477	0.9955	0.9966	0.7469	0.7463
Cointegration: $b = d$								
	$\mathcal{H}_0, n =$		$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.9984	0.9992	1.2498	1.2486	0.9960	0.9974	0.7473	0.7470
$d = 0.60$	0.9987	0.9997	1.2500	1.2491	0.9962	0.9979	0.7475	0.7475
$d = 1.00$	0.9990	1.0002	1.2503	1.2496	0.9965	0.9983	0.7477	0.7479

**Table 11: Mean estimate: Baseline.** This table reports the mean of the full-sample LCM estimates for different predictive models. Specifically, the estimator is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$ , for  $(\nu, \kappa) = (0.2, 0.7)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability.  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  describes three different *baseline* predictive models with breaks. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

Standard deviation of LCM Estimate: Baseline								
No Cointegration: $b = 0$								
	$\mathcal{H}_0, n =$		$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.0471	0.0366	0.0730	0.0562	0.0802	0.0603	0.0727	0.0545
$d = 0.60$	0.0471	0.0366	0.0730	0.0562	0.0802	0.0603	0.0727	0.0545
$d = 1.00$	0.0471	0.0366	0.0730	0.0562	0.0802	0.0603	0.0727	0.0545
Cointegration: $b = d$								
	$\mathcal{H}_0, n =$		$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700	350	700
$d = 0.30$	0.0388	0.0295	0.0678	0.0521	0.0760	0.0566	0.0673	0.0502
$d = 0.60$	0.0335	0.0241	0.0648	0.0494	0.0738	0.0545	0.0643	0.0476
$d = 1.00$	0.0291	0.0185	0.0625	0.0469	0.0722	0.0530	0.0622	0.0456

**Table 12: Standard deviation: Baseline.** This table provides the standard deviation of the full sample LCM estimates for different predictive models. Specifically, the estimator is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$ , for  $(\nu, \kappa) = (0.2, 0.7)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability.  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  describes three different *baseline* predictive models with breaks. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

Bias of LCM-based Break Fraction Estimate: Baseline						
No Cointegration: $b = 0$						
	$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700
$d = 0.30$	-0.0060	-0.0031	0.0002	-0.0001	0.0023	0.0031
$d = 0.60$	-0.0060	-0.0031	0.0002	-0.0001	0.0023	0.0031
$d = 1.00$	-0.0060	-0.0031	0.0002	-0.0001	0.0023	0.0031
Cointegration: $b = d$						
	$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700
$d = 0.30$	-0.0042	-0.0014	-0.0008	-0.0005	-0.0012	0.0003
$d = 0.60$	-0.0031	-0.0010	-0.0021	-0.0026	-0.0027	-0.0013
$d = 1.00$	-0.0033	-0.0016	-0.0032	-0.0042	-0.0046	-0.0016

**Table 13: Break fraction bias: Baseline.** This table reports the bias of the break fraction estimates from the LCM-based sup-Wald test in equation (19). Specifically, the test is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$ , and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ , for  $(\nu, \kappa, \kappa_G) = (0.2, 0.7, 0.9)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  describes three different *baseline* predictive models with breaks, whose break fractions are 0.25, 0.50 and 0.75, respectively. The predictive model is specified with ( $b = d$ ) and without ( $b = 0$ ) cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

RMSE of LCM-based Break Fraction Estimate: Baseline						
No Cointegration: $b = 0$						
	$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700
$d = 0.30$	0.0160	0.0107	0.0141	0.0088	0.0162	0.0113
$d = 0.60$	0.0160	0.0107	0.0141	0.0088	0.0162	0.0113
$d = 1.00$	0.0160	0.0107	0.0141	0.0088	0.0162	0.0113
Cointegration: $b = d$						
	$\mathcal{H}_1, n =$		$\mathcal{H}_2, n =$		$\mathcal{H}_3, n =$	
	350	700	350	700	350	700
$d = 0.30$	0.0161	0.0116	0.0159	0.0099	0.0186	0.0120
$d = 0.60$	0.0147	0.0092	0.0152	0.0099	0.0190	0.0120
$d = 1.00$	0.0133	0.0080	0.0155	0.0122	0.0216	0.0128

**Table 14: Break fraction RMSE: Baseline.** This table reports the bias of the break fraction estimates from the LCM-based sup-Wald test in equation (19). Specifically, the test is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ , for  $(\nu, \kappa, \kappa_G) = (0.2, 0.7, 0.9)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3$  describes three different *baseline* predictive models with breaks, whose break fractions are 0.25, 0.50 and 0.75, respectively. The predictive model is specified with ( $b = d$ ) and without ( $b = 0$ ) cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

Mean LCM Estimate: Small Breaks									
No Cointegration: $b = 0$									
	$\mathcal{H}_0, n =$		$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$		
	350	700	350	700	350	700	350	700	
$d = 0.30$	0.9978	0.9983	1.0482	1.0482	0.9976	0.9981	0.9479	0.9480	
$d = 0.60$	0.9978	0.9983	1.0482	1.0482	0.9976	0.9981	0.9479	0.9480	
$d = 1.00$	0.9978	0.9983	1.0482	1.0482	0.9976	0.9981	0.9479	0.9480	
Cointegration: $b = d$									
	$\mathcal{H}_0, n =$		$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$		
	350	700	350	700	350	700	350	700	
$d = 0.30$	0.9984	0.9992	1.0488	1.0491	0.9982	0.9990	0.9485	0.9489	
$d = 0.60$	0.9987	0.9997	1.0491	1.0497	0.9984	0.9995	0.9488	0.9495	
$d = 1.00$	0.9990	1.0002	1.0494	1.0502	0.9987	1.0000	0.9490	0.9500	

**Table 15: Mean estimate: Small breaks.** This table reports the mean of the full-sample LCM estimates for different predictive models. Specifically, the estimator is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$ , for  $(\nu, \kappa) = (0.2, 0.7)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability.  $\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6$  describes three predictive models with *small* breaks. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

Standard deviation of LCM Estimate: Small breaks									
No Cointegration: $b = 0$									
	$\mathcal{H}_0, n =$		$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$		
	350	700	350	700	350	700	350	700	
$d = 0.30$	0.0471	0.0366	0.0483	0.0376	0.0486	0.0375	0.0482	0.0372	
$d = 0.60$	0.0471	0.0366	0.0483	0.0376	0.0486	0.0375	0.0482	0.0372	
$d = 1.00$	0.0471	0.0366	0.0483	0.0376	0.0486	0.0375	0.0482	0.0372	
Cointegration: $b = d$									
	$\mathcal{H}_0, n =$		$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$		
	350	700	350	700	350	700	350	700	
$d = 0.30$	0.0388	0.0295	0.0403	0.0308	0.0408	0.0307	0.0401	0.0302	
$d = 0.60$	0.0335	0.0241	0.0352	0.0257	0.0359	0.0257	0.0350	0.0252	
$d = 1.00$	0.0291	0.0185	0.0309	0.0206	0.0319	0.0210	0.0308	0.0202	

**Table 16: Standard deviation: Small breaks.** This table reports the standard deviation of the full-sample LCM estimates for different predictive models. Specifically, the estimator is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$  and bandwidth  $m = \lfloor n^\kappa \rfloor$ , for  $(\nu, \kappa) = (0.2, 0.7)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_0$  describes the null hypothesis *without* parameter instability.  $\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6$  describes three predictive models with *small* breaks. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

Bias of LCM-based Break Fraction Estimate: Small Breaks						
No Cointegration: $b = 0$						
	$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$	
	350	700	350	700	350	700
$d = 0.30$	0.1026	0.0712	-0.0023	0.0055	-0.1062	-0.0591
$d = 0.60$	0.1026	0.0712	-0.0023	0.0055	-0.1062	-0.0591
$d = 1.00$	0.1026	0.0712	-0.0023	0.0055	-0.1062	-0.0591
Cointegration: $b = d$						
	$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$	
	350	700	350	700	350	700
$d = 0.30$	0.0772	0.0363	-0.0270	-0.0167	-0.1231	-0.0769
$d = 0.60$	0.0572	0.0242	-0.0324	-0.0228	-0.1218	-0.0733
$d = 1.00$	0.0500	0.0136	-0.0342	-0.0175	-0.1300	-0.0564

**Table 17: Break fraction bias: Small breaks.** This table reports the bias of the break fraction estimates from the LCM-based sup-Wald test in (19). Specifically, the test is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ , for  $(\nu, \kappa, \kappa_G) = (0.2, 0.7, 0.9)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6$  describes three predictive models with *small* breaks, whose break fractions are 0.25, 0.50 and 0.75, respectively. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

RMSE of LCM-based Break Fraction Estimate: Small Breaks						
No Cointegration: $b = 0$						
	$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$	
	350	700	350	700	350	700
$d = 0.30$	0.1992	0.1643	0.1532	0.1286	0.2061	0.1449
$d = 0.60$	0.1992	0.1643	0.1532	0.1286	0.2061	0.1449
$d = 1.00$	0.1992	0.1643	0.1532	0.1286	0.2061	0.1449
Cointegration: $b = d$						
	$\mathcal{H}_4, n =$		$\mathcal{H}_5, n =$		$\mathcal{H}_6, n =$	
	350	700	350	700	350	700
$d = 0.30$	0.1683	0.1051	0.1396	0.1056	0.2201	0.1641
$d = 0.60$	0.1373	0.0864	0.1307	0.0926	0.2146	0.1587
$d = 1.00$	0.1273	0.0604	0.1282	0.0783	0.2220	0.1373

**Table 18: Break fraction RMSE: Small breaks.** This table reports the bias of the break fraction estimates from the LCM-based sup-Wald test in (19). Specifically, the test is implemented using trimming  $\ell = \lfloor n^\nu \rfloor$ , bandwidth  $m = \lfloor n^\kappa \rfloor$  and a corresponding covariance estimator bandwidth  $m_G = \lfloor n^{\kappa_G} \rfloor$ , for  $(\nu, \kappa, \kappa_G) = (0.2, 0.7, 0.9)'$ . As described in the main text,  $\{y_t, x_{t-1}\} \in I(d)$ , with  $x_{t-1}$  predicting  $y_t$ .  $\mathcal{H}_4, \mathcal{H}_5, \mathcal{H}_6$  describes three different predictive models with *small* breaks, whose break fractions are 0.25, 0.50 and 0.75, respectively. The predictive model is specified with  $(b = d)$  and without  $(b = 0)$  cointegration. The sample size is either  $n = 350$  or  $n = 700$ , and the nominal test size is 5%. Simulations are implemented with 1000 replications.

## References

- Andersen, T. G. & Bollerslev, T. (1997), ‘Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns’, *Journal of Finance* **52**, 975–1005.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), ‘The distribution of realized stock return volatility’, *Journal of Financial Economics* **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), ‘The distribution of exchange rate volatility’, *Journal of the American Statistical Association* **96**, 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**, 579–625.
- Andersen, T. G. & Bondarenko, O. (2007), Construction and interpretation of model-free implied volatility, in I. Nelken, ed., ‘Volatility as an Asset Class’, London: Risk Books.
- Andersen, T. G., Bondarenko, O. & Gonzalez-Perez, M. (2015), ‘Exploring return dynamics via corridor implied volatility’, *The Review of Financial Studies* **28**(10), 2902–2945.
- Andersen, T. G., Fusari, N. & Todorov, V. (2015), ‘The risk premia embedded in index options’, *Journal of Financial Economics* **117**, 558–584.
- Andersen, T. G., Fusari, N., Todorov, V. & Varneskov, R. T. (2020), ‘Spatial dependence in option observation errors’, *Econometric Theory* **forthcoming**.
- Andersen, T. G. & Varneskov, R. T. (2020a), ‘Consistent inference for predictive regressions in persistent economic systems’, *Journal of Econometrics* **forthcoming**.
- Andersen, T. G. & Varneskov, R. T. (2020b), Consistent local spectrum (LCM) inference for predictive return regressions. Unpublished manuscript, Northwestern University and Copenhagen Business School.
- Andrews, D. W. (1991), ‘Heteroskedasticity and autocorrelation consistent covariance matrix estimation’, *Econometrica* **59**(3), 817–858.
- Andrews, D. W. K. (1993), ‘Tests for parameter instability and structural change with unknown change point’, *Econometrica* **61**, 821–856.
- Andrews, D. W. K. & Ploberger, W. (1994), ‘Optimal tests when a nuisance parameter is present only under the alternative’, *Econometrica* **62**, 1383–1314.
- Bai, J. & Perron, P. (1998), ‘Estimating and testing linear models with multiple structural change’, *Econometrica* **66**, 47–78.
- Bai, J. & Perron, P. (2006), Multiple structural change models: A simulation analysis, in D. Corbea, S. Durlauf & B. E. Hansen, eds, ‘Econometric Theory and Practice: Frontiers of Analysis and Applied Research’, Cambridge University Press, pp. 212–237.
- Bandi, F. M. & Perron, B. (2006), ‘Long memory and the relation between implied and realized volatility’, *Journal of Financial Econometrics* **4**(4), 636–670.
- Boldea, O., Cornea-Madeira, A. & Hall, A. R. (2019), ‘Bootstrapping structural change tests’, *Journal of Econometrics* **213**, 359–397.



- Bollerslev, T., Osterrieder, D., Sizova, N. & Tauchen, G. (2013), ‘Risk and return: Long-run relationships, fractional cointegration, and return predictability’, *Journal of Financial Economics* **108**, 409–424.
- Bollerslev, T., Sizova, N. & Tauchen, G. (2011), ‘Volatility in equilibrium: Asymmetries and dynamic dependencies’, *Review of Finance* **16**, 31–80.
- Bollerslev, T., Tauchen, G. & Zhou, H. (2009), ‘Expected stock returns and variance risk premia’, *Review of Financial Studies* **22**(11), 4463–4492.
- Bollerslev, T. & Todorov, V. (2011), ‘Tails, fears, and risk premia’, *Journal of Finance* **66**(6), 2165–2211.
- Brillinger, D. R. (1981), *Time Series. Data Analysis and Theory*, Siam: Classics in Applied Mathematics.
- Carr, P. & Wu, L. (2009), ‘Variance risk premiums’, *Review of Financial Studies* **22**(3), 1311–1341.
- Casini, A. & Perron, P. (2019), Structural breaks in time series, in ‘Oxford Research Encyclopedia of Economics and Finance’, Oxford University Press.
- Cavanagh, C., Elliott, G. & Stock, J. (1995), ‘Inference in models with nearly integrated regressors’, *Econometric Theory* **11**, 1131–1147.
- Chen, B. & Hong, Y. (2012), ‘Testing for smooth structural changes in time series models via nonparametric regression’, *Econometrica* **80**, 1157–1183.
- Cheng, X. & Fung, J. W. (2012), ‘The information content of model-free implied volatility’, *Journal of Futures Markets* **32**, 792–806.
- Chernov, M. (2007), ‘On the role of risk premia in volatility forecasting’, *Journal of Business and Economic Statistics* **25**(4), 411–426.
- Chraoenwong, C., Jenwittayaroje, N. & Low, B. (2009), ‘Who knows more about future currency volatility?’, *Journal of Futures Markets* **29**, 270–295.
- Christensen, B. J. & Nielsen, M. O. (2006), ‘Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting’, *Journal of Econometrics* **133**, 343–371.
- Christensen, B. J. & Prabhala, N. R. (1998), ‘The relation between implied and realized volatility’, *Journal of Financial Economics* **50**(2), 125–150.
- Christensen, B. J. & Varneskov, R. T. (2017), ‘Medium band least squares estimation of fractional cointegration in the presence of low-frequency contamination’, *Journal of Econometrics* **197**, 218–244.
- Christoffersen, P. & Diebold, F. X. (1998), ‘Cointegration and long-horizon forecasting’, *Journal of Business and Economic Statistics* **16**, 450–458.
- Clark, T. E. & McCracken, M. W. (2009), ‘Improving forecast accuracy by combining recursive and rolling forecasts’, *International Economic Review* **50**(2), 363 – 395.
- Corsi, F. (2009), ‘A simple approximate long-memory model of realized volatility’, *Journal of Financial Econometrics* **7**, 174–196.
- Dangl, T. & Hailing, M. (2012), ‘Predictive regressions with time-varying coefficients’, *Journal of Financial Economics* **106**, 157–181.

- Diebold, F. X. & Chen, C. (1996), ‘Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures’, *Journal of Econometrics* **70**, 221–241.
- Diebold, F. X. & Inoue, A. (2001), ‘Long memory and regime switching’, *Journal of Econometrics* **105**, 131–159.
- Elliott, G., Müller, U. & Watson, M. (2015), ‘Nearly optimal tests when a nuisance parameter is present under the null hypothesis’, *Econometrica* **83**, 771–811.
- Farmer, L., Schmidt, L. & Timmermann, A. (2019), Pockets of predictability. Unpublished manuscript, University of California, San Diego.
- Frederiksen, P. H., Nielsen, F. S. & Nielsen, M. O. (2012), ‘Local polynomial whittle estimation of perturbed fractional processes’, *Journal of Econometrics* **167**, 426–447.
- Georgiev, I., Harvey, D. I., Leybourne, S. J. & Taylor, A. M. R. (2018), ‘Testing for parameter instability in predictive regression models’, *Journal of Econometrics* **204**, 101–118.
- Granger, C. V. J. & Newbold, P. (1974), ‘Spurious regression in econometrics’, *Journal of Econometrics* **2**, 111–120.
- Hansen, B. E. (1992), ‘Tests for parameter instability in regressions with I(1) processes’, *Journal of Business and Economic Statistics* **10**, 321–335.
- Hansen, B. E. (2000), ‘Testing for structural change in conditional models’, *Journal of Econometrics* **97**, 93–115.
- Hansen, P. R. (2003), ‘Structural changes in the cointegrated vector autoregressive model’, *Journal of Econometrics* **114**, 261–295.
- Hidalgo, J. & Robinson, P. M. (1996), ‘Testing for structural change in a long-memory environment’, *Journal of Econometrics* **70**, 159–174.
- Hong, Y. (1996), ‘Testing for independence between two covariance stationary time series’, *Biometrika* **83**, 615–625.
- Hualde, J. & Robinson, P. M. (2011), ‘Gaussian pseudo-maximum likelihood estimation of fractional time series models’, *Annals of Statistics* **39**, 3152–3181.
- Ibragimov, I. & Hasminskii, R. (1981), *Statistical Estimation: Asymptotic Theory*, Springer, Berlin.
- Jansson, M. & Moreira, M. J. (2006), ‘Optimal inference in regression models with nearly integrated regressors’, *Econometrica* **74**, 681–714.
- Jiang, G. J. & Tian, Y. S. (2005), ‘Model-free implied volatility and its information content’, *Review of Financial Studies* **18**, 1305–1342.
- Johansen, S. & Nielsen, M. O. (2012), ‘Likelihood inference for a fractionally cointegrated vector autoregressive model’, *Econometrica* **80**, 2667–2732.
- Kellard, N., Dunis, C. & Sarantis, N. (2010), ‘Foreign exchange, fractional cointegration and the implied-realized volatility relation’, *Journal of Banking and Finance* **34**, 882–891.
- Kerjiwal, M. & Perron, P. (2010), ‘Testing for multiple structural changes in cointegrated regression models’, *Journal of Business and Economic Statistics* **28**, 503–522.

- Kilic, M. & Shaliastovich, I. (2019), ‘Good and bad variance risk premia and expected returns’, *Management Science* **65**, 2522–2544.
- Kim, D. & Perron, P. (2009), ‘Assessing the relative power of structural break tests using the framework based on the approximate Bahadur slope’, *Journal of Econometrics* **149**, 26–51.
- Li, Z., Izzeldin, M. & Yao, X. (2020), ‘Return predictability of variance differences: A fractionally cointegrated approach’, *Journal of Futures Markets* **40**, 1072–1089.
- Lobato, I. (1999), ‘A semiparametric two-step estimator in a multivariate long memory model’, *Journal of Econometrics* **90**, 129–155.
- Ng, S. & Perron, P. (1997), ‘Estimation and inference in nearly unbalanced and nearly cointegrated systems’, *Journal of Econometrics* **79**, 53–81.
- Nielsen, M. O. (2007), ‘Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation’, *Journal of Business and Economic Statistics* **25**, 427–446.
- Nielsen, M. O. (2015), ‘Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time series models’, *Journal of Time Series Analysis* **36**, 154–188.
- Nielsen, M. O. & Frederiksen, P. (2011), ‘Fully modified narrow-band least squares estimation of weak fractional cointegration’, *The Econometrics Journal* **14**(1), 77–120.
- Osterrieder, D., Ventosa-Santaularia, D. & Vera-Valdes, J. E. (2019), ‘The VIX, the variance premium, and expected returns’, *Journal of Financial Econometrics* **17**, 527–558.
- Pastor, L. & Stambaugh, R. F. (2009), ‘Predictive systems: Living with imperfect predictors’, *Journal of Finance* **64**, 1583–1628.
- Paye, B. & Timmermann, A. (2006), ‘Instability of return prediction models’, *Journal of Empirical Finance* **13**, 274–315.
- Perron, P. (2006), Dealing with structural breaks, in K. Patterson & T. C. Mills, eds, ‘Palgrave Handbook of Econometrics, Vol 1: Econometric Theory’, Palgrave Macmillan, pp. 278–352.
- Perron, P. & Qu, Z. (2010), ‘Long memory and level shifts in the volatility of stock market return indices’, *Journal of Business and Economic Statistics* **28**, 275–290.
- Pesaran, M. H. & Timmermann, M. H. (2007), ‘Selection of estimation window in the presence of breaks’, *Journal of Econometrics* **137**, 134–161.
- Peseran, M. H., Pettenuzzo, D. & Timmermann, A. (2006), ‘Forecasting time series subject to multiple structural breaks’, *Review of Economic Studies* **73**, 1057–1084.
- Phillips, P. C. B. (1986), ‘Understanding spurious regressions in econometrics’, *Journal of Econometrics* **33**, 311–340.
- Phillips, P. C. B. & Lee, J. H. (2013), ‘Predictive regression under various degrees of persistence and robust long-horizon regression’, *Journal of Econometrics* **177**, 250–264.
- Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.
- Robinson, P. M. (1995), ‘Gaussian semiparametric estimation of long range dependence’, *The Annals of Statistics* **23**, 1630–1661.

- Robinson, P. M. & Marinucci, D. (2003), ‘Semiparametric frequency domain analysis of fractional cointegration’. In: Robinson, P.M. (Ed.), *Time Series with Long Memory*. Oxford University Press, Oxford, pp. 334-373.
- Shao, X. (2009), ‘A generalized portmanteau test for independence between two stationary time series’, *Econometric Theory* **25**, 195–210.
- Shimotsu, K. (2007), ‘Gaussian semiparametric estimation of multivariate fractionally integrated processes’, *Journal of Econometrics* **137**, 277–310.
- Shimotsu, K. (2010), ‘Exact local whittle estimation of fractional integration with unknown mean and time trend’, *Econometric Theory* **26**, 501–540.
- Stambaugh, R. F. (1999), ‘Predictive regressions’, *Journal of Financial Economics* **54**, 783–820.
- Stock, J. H. & Watson, M. W. (1996), ‘Evidence of structural stability in macroeconomic time series’, *Journal of Business & Economic Statistics* **14**, 11–30.
- Taylor, S., Yadav, P. & Zhang, Y. (2010), ‘The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks’, *Journal of Banking and Finance* **34**, 871–881.
- Tsay, W.-J. & Chung, C.-F. (2000), ‘The spurious regression of fractionally integrated processes’, *Journal of Econometrics* **96**, 155–182.
- Valkanov, R. (2003), ‘Long-horizon regressions: Theoretical results and applications’, *Journal of Financial Economics* **68**, 201–232.
- Varneskov, R. T. (2016), ‘Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices’, *Journal of Business and Economic Statistics* **31**(1), 1–22.
- Varneskov, R. T. (2017), ‘Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels’, *Econometric Theory* **33**(6), 1457–1501.
- Varneskov, R. T. & Perron, P. (2018), ‘Combining long memory and level shifts in modeling and forecasting the volatility of asset returns’, *Quantitative Finance* **18**, 371–393.
- Vogelsang, T. J. (1997), ‘Wald-type tests for detecting breaks in the trend function of a dynamic time series’, *Econometric Theory* **13**, 818–849.
- Wooldridge, J. M. & White, H. (1988), ‘Some invariance principles and central limit theorems for dependent heterogeneous processes’, *Econometric Theory* **4**, 210–230.