

UC San Diego

UC San Diego Previously Published Works

Title

Pre and post break parameter inference

Permalink

<https://escholarship.org/uc/item/4j733246>

Journal

Journal of Econometrics, 180(2)

ISSN

0304-4076

Authors

Elliott, Graham
Müller, Ulrich K

Publication Date

2014-06-01

DOI

10.1016/j.jeconom.2014.03.007

Peer reviewed



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconomPre and post break parameter inference[☆]Graham Elliott^a, Ulrich K. Müller^{b,*}^a University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0508, United States^b Princeton University, Princeton, NJ, 08544-1021, United States

ARTICLE INFO

Article history:

Received 23 August 2012

Received in revised form

20 February 2014

Accepted 16 March 2014

Available online 31 March 2014

JEL classification:

C32

C12

Keywords:

Structural breaks

Time varying parameters

Convergence of experiments

Asymptotic efficiency of tests

ABSTRACT

Consider inference about the pre and post break value of a scalar parameter in a time series model with a single break at an unknown date. Unless the break is large, treating the break date estimated by least squares as the true break date leads to substantially oversized tests and confidence intervals. To develop a suitable alternative, we first establish convergence to a Gaussian process limit experiment. We then determine a nearly weighted average power maximizing test in this limit experiment, and show how to implement a small sample analogue in GMM time series models.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

This paper is concerned with testing hypotheses about the pre and post break value of a parameter in a time series model with a single break. By reversing the time ordering, inference about the post break value becomes inference about the pre break value, so that for simplicity, we will exclusively discuss the post break case. If the break date is known, inference is straightforward, as standard results apply after restricting attention to the stable post break data. The effect of an incorrectly chosen break date crucially depends on whether the chosen date is earlier or later than the true break date. On the one hand, if the chosen date is later, then a restriction to the presumed post break data still yields a stable model, and inference remains valid. There is a cost of efficiency, though, since more post break data could have been used. On the other hand, if the chosen date is earlier than the true break date, then the presumed post break data stems from an unstable model.

Parameter estimators in this unstable model tend to estimate the average parameter value, which is different from the true post break value, so standard inference yields tests and confidence intervals with distorted size.

These distortionary effects are small if the chosen break date is sufficiently close to the true break date. With the break date unknown, this requires a precise break date estimator. As formally shown by Bai (1994, 1997) and Bai and Perron (1998) for linear regressions, and Hall et al. (2012) and Perron and Yamamoto (2012) for two stage least squares, the least squares break date estimator is indeed sufficiently precise for these distortionary effects to become asymptotically negligible if the parameter shift is sufficiently pronounced. Formally, these papers study asymptotics in which the break magnitude, while possibly shrinking, is outside the local $T^{-1/2}$ neighborhood of zero. Relative to the sampling uncertainty of the parameter estimator in a stable model, the parameter shift thus becomes infinitely large. We consider the behavior of the 5% nominal level two-sided test based on the least squares break date estimator under local asymptotics in Section 2.3.1, where the break magnitude is measured in multiples of standard deviations of the full sample parameter estimator in a stable model. With the break date restricted to the middle 70% of the sample, the largest null rejection probability is almost 30%, a break magnitude of less than 5 standard deviations leads to effective size of more than 10% for all break dates, and a break of 11 standard deviations still yields size greater than 10% for some break dates. These distortions are further exacerbated by attempts to pre-test for parameter stability.

[☆] The authors thank the co-editor, Peter Robinson and three anonymous referees for useful comments and suggestions, and seminar and workshop participants at Chicago GSB, Columbia, Cowles Summer Conference, Harvard, MIT, Montreal, NBER Summer Institute, Pennsylvania State, St. Louis Federal Reserve Bank, U.C. Irvine, University of Pennsylvania and USC for helpful discussion. Jia Li provided excellent research assistance. Müller gratefully acknowledges the NSF for financial support via grant SES-0751056.

* Corresponding author.

E-mail address: umueller@princeton.edu (U.K. Müller).

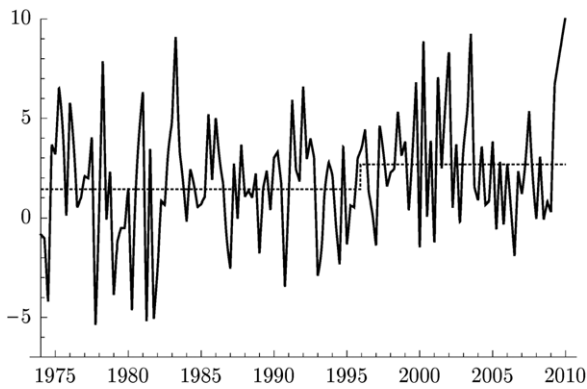


Fig. 1. Quarterly productivity growth in the non farm business sector and average growth pre and post 1995Q4 (Fred Database series OPHNFB).

In other words, for any sample size, there exists a break magnitude for which these standard methods yield incorrect inference, so they are not uniformly valid.

It is instructive to consider these issues in the context of an example. Productivity growth in most developed nations appears to have suffered shifts in its mean a number of times. Most researchers agree on a downward shift in productivity following the oil crisis of 1973, typically dated at the third quarter. In post 1973 data, there is a suggestion of a mid 1990's upward shift in average productivity growth and speculation about its cause. A recent summary appears in Jorgenson et al. (2008), who informally suggest a break in the fourth quarter of 1995. Gordon (2003) also dates the increase to 1995. Anderson and Klieson (2006) refer to the increase in productivity as the “defining economic event of the past decade”.

Fig. 1 plots quarterly observations of US non farm business productivity growth, along with the pre and post break means using the Jorgenson et al. (2008) break date of 1995 Q4. With this break date, quarterly productivity growth is estimated to have increased by 1.1% in 1995 Q4. The least-squares estimate of the break date is slightly later at 1997 Q1, although the parameter stability test of Elliott and Müller (2006) fails to reject the null hypothesis of stable mean growth at the 5% level. With a (long-run) standard deviation of productivity growth of approximately 2.6% and a sample of $T = 143$ observations, 5–11 standard deviations of the full sample estimator correspond to a 1.0%–2.3% shift in quarterly productivity growth. In this example, one would thus not want to assume that the true break magnitude is sufficiently large for least-squares break date based inference to be reliable. These concerns are corroborated in a small sample Monte Carlo study modeled around the productivity example in Section 4.3. This illustrates that lack of uniform asymptotic validity directly translates into poor small sample properties with practically relevant parameter values.

More generally, it seems hard to argue that break magnitudes of less than 11 standard deviations of the full sample estimator are not part of the empirically relevant parameter space in most applications. One way to see this is that shifts of, say, 8 standard deviations tend to induce highly significant rejections of the null hypothesis of parameter stability with high probability—see Elliott and Müller (2007). But in practice, rejections of the stable model are rarely overwhelming. For instance, there is continued debate about the stability of monetary policy with Orphanides (2004) arguing for rather stable relationships, while Cogley and Sargent (2005) find instabilities that they deem “economically important”, but of a magnitude that would be detected by a formal parameter stability test less than 25% of the time.

Simple adjustments to the standard procedure that ensure size control over all break magnitudes do not deliver reasonable tests.

Table 1
Inference about current US productivity growth.

Method	95% Confidence interval
Break date chosen at 1995Q4	[2.00, 3.36]
Least squares break date estimator	[2.02, 3.49]
Corrected LS break date estimator	[0.92, 4.59]
Using last 15% of data	[0.89, 2.99]
Test of this paper	[1.74, 3.43]

Note: Sample size is $T = 143$. All confidence intervals use Newey and West (1987) standard errors with 4 lags.

For instance, in the set-up mentioned above and described in detail in Section 2.3.1, the critical value for the 5% two-sided t -statistic would need to be increased from 1.96 to approximately 4.9 to ensure uniform size control, with obvious adverse effects on power.¹ A Bonferroni procedure based on uniformly valid confidence sets for the break date developed by Elliott and Müller (2007) performs well for large breaks, but has poor power for breaks of moderate magnitude. Also, given the results of Andrews and Guggenberger (2010a), it seems clear that standard bootstrap or subsampling techniques would not yield uniformly valid inference in this problem.

The main contribution of this paper is the construction of a powerful test about the post break parameter value in a general GMM time series model with unknown break date in the middle 70% of the sample that controls size uniformly over the break magnitude. This test follows a switching rule: if there is strong evidence for a large break, then the test essentially reduces to standard inference using post break data determined by a break date estimator, with a slight increase of the 5% critical value from 1.96 to 2.01. In the absence of strong evidence for a large break, the test switches to a likelihood ratio test. This likelihood ratio test is carefully constructed to ensure both overall size control and approximate efficiency in the sense that for a particular weighting function, (local asymptotic) weighted average power of the suggested test is demonstrably at most one percentage point smaller than the weighted average power of any test that controls size. Section 4.1 gives a fully self-contained explanation of how to construct the test statistic, and is the relevant section for readers wishing to apply the method to data.

Returning to the US productivity example, Table 1 contains nominal 95% confidence intervals for current US productivity growth, constructed using different methods. The first method assumes the Jorgenson et al. (2008) break date to be correct and reports the post 1995 Q4 sample mean ± 1.96 times the standard error. The second method estimates the break date by least squares and, conditional on this estimate, again applies standard critical values. The third method also relies on the least-squares break date, but instead uses a critical value of 4.9 instead of 1.96 to ensure uniform asymptotic coverage control, as explained above. The fourth method conducts standard inference using the last 15% of the data only. The final row reports the 95% confidence interval constructed by inverting the 5% level test developed in this paper.

As argued above, in this example it is not reasonable to assume that the true break magnitude is large enough for the (uncorrected) least-squares based method to yield reliable inference. In contrast, the last three methods are asymptotically uniformly valid. Among those valid methods, the inference developed here is by far the most informative. The resulting confidence interval is nevertheless wider than those with the break date presumed known, and its center is shifted towards smaller values. This makes intuitive sense: with the break date unknown, part of the low productivity

¹ This adjustment to the critical value is also the end result of the size corrected hybrid subsampling method advanced by Andrews and Guggenberger (2009, 2010b) when applied to this problem.

growth quarters of, say, the early 90s are potentially realizations from the post break period, which suggests a relatively lower post break population mean.

The efficiency claim of the derived test is obtained as follows. Consider first a parametric underlying model. We apply LeCam’s limit of experiments theory to establish that quite generally, time variation of parameters of the order of $T^{-1/2}$ in a nicely behaved parametric model leads to a corresponding limit experiment involving the observation of an appropriately defined Gaussian process on the unit interval. Specializing this result to a model with a single break then leads to a hypothesis testing problem in the limit experiment where both the null and alternative hypothesis are composite. In particular, under the null hypothesis about the post break parameter value, the nuisance parameters are the break date and break magnitude. This non-standard testing problem falls into the class of problems considered by Elliott et al. (2012) (EMW in the following), who derive a set of upper bounds on the weighted average power of any valid test, and suggest a numerical algorithm to determine both a low upper bound and a test with weighted average power close to the bound. We apply EMW’s approach to the limit experiment corresponding to post break parameter inference.² We further show how to construct an analogue to the nearly optimal test in the limit experiment for the original testing problem. The local asymptotic power of this test matches the power of the nearly optimal test in the limit experiment, which establishes our efficiency claim via standard limit of experiment arguments. For an underlying GMM model, we invoke the results of Müller (2011) to obtain the corresponding asymptotic efficiency claims in the class of tests that are “robust” in the sense of Müller (2011).

The remainder of the paper is organized as follows. Section 2 is dedicated to the analysis of parametric models. Section 2.1 establishes convergence to a limit experiment under quite general forms of (parameterized) time variation in parameters of the order $T^{-1/2}$. This is then specialized to the problem of post break parameter inference in Section 2.2. Section 2.3 analyzes the limit experiment for a scalar parameter, illustrates the failure of least-squares based inference and describes how we apply the ideas of EMW to obtain a powerful test. Section 2.4 translates this result for a scalar parameter to an efficiency claim for models with a vector parameter. Section 3 considers GMM models and contains details on the application of Müller’s (2011) approach to large sample efficiency. Section 4 defines the suggested test statistic in terms of partial sample GMM statistics (cf. Andrews (1993)), establishes its asymptotic properties, and analyzes its small sample behavior in three Monte Carlo experiments. Proofs are collected in an Appendix.

2. Asymptotic analysis for parametric models

2.1. Limit experiment for general forms of time variation

Let $X_T = (x_{T,1}, \dots, x_{T,T}) \in \mathbb{R}^{qT}$ be the available data in a sample of size T . The data is the realization of a random vector X_T with density $\prod_{t=1}^T f_{T,t}(\Gamma_t)$ relative to some σ -finite measure μ_T when the parameter $\Gamma \in \mathbb{R}^k$ takes on the value Γ_t at time t . This form of the likelihood arises naturally when $f_{T,t}(\Gamma_t)$ is the density of $x_{T,t}$ conditional on $\mathfrak{F}_{T,t-1}$, the σ -field generated by $\{x_{T,s}\}_{s=1}^{t-1}$. We will refer to the model with density $\prod_{t=1}^T f_{T,t}(\Gamma_0)$ as the “stable” model. Define $l_{T,t}(\Gamma) = \ln f_{T,t}(\Gamma)$, $s_{T,t}(\Gamma) = \partial l_{T,t}(\Gamma) / \partial \Gamma$ and $h_{T,t} = \partial s_{T,t}(\Gamma) / \partial \Gamma'$, and write $\lfloor x \rfloor$ for the largest integer smaller or equal to $x \in \mathbb{R}$. Assume that the stable model is nicely behaved, as described in the following regularity condition.

² The examples worked out in EMW all have a single nuisance parameter under the null hypothesis; the presence of two nuisance parameters renders this exercise relatively more involved from an “engineering” point of view.

Condition 1. In the stable model with parameter Γ_0

- (i) in some neighborhood \mathcal{B}_0 of Γ_0 , $l_{T,t}(\Gamma)$ is twice differentiable a.s. with respect to Γ for $t = 1, \dots, T$, $T \geq 1$;
- (ii) $\{s_{T,t}(\Gamma_0), \mathfrak{F}_{T,t}\}$ is a square-integrable martingale difference array with $T^{-1} \sum_{t=1}^{\lfloor sT \rfloor} E[s_{T,t}(\Gamma_0)s_{T,t}(\Gamma_0)' | \mathfrak{F}_{T,t-1}] \xrightarrow{p} sH$ for all $0 \leq s \leq 1$ and some $k \times k$ full rank matrix H , $T^{-1} \sup_{t \leq T} \|E[s_{T,t}(\Gamma_0)s_{T,t}(\Gamma_0)' | \mathfrak{F}_{T,t-1}]\| \xrightarrow{p} 0$ and there exists $\nu > 0$ such that $T^{-1} \sum_{t=1}^T E[\|s_{T,t}(\Gamma_0)\|^{2+\nu} | \mathfrak{F}_{T,t-1}] = O_p(1)$;
- (iii) $T^{-1} \sum_{t=1}^T \|h_{T,t}(\Gamma_0)\| = O_p(1)$, $T^{-1} \sup_{t \leq T} \|h_{T,t}(\Gamma_0)\| \xrightarrow{p} 0$ and for any decreasing neighborhoods $\mathcal{B}_T \subset \mathcal{B}_0$ of Γ_0 that converge to $\{\Gamma_0\}$, $T^{-1} \sum_{t=1}^T \sup_{\Gamma \in \mathcal{B}_T} \|h_{T,t}(\Gamma) - h_{T,t}(\Gamma_0)\| \xrightarrow{p} 0$;
- (iv) for all $0 \leq s \leq 1$, $T^{-1} \sum_{t=1}^{\lfloor sT \rfloor} h_{T,t}(\Gamma_0) \xrightarrow{p} -sH$.

These conditions are a special case of Condition 2 in Li and Müller (2009) and Condition 1 in Müller and Petalas (2010), who provide further discussion and references.

Now consider a version of this model with time varying parameters, that is for all $t \leq T$

$$\Gamma_{T,t} = \Gamma_0 + T^{-1/2}B(t/T; \theta). \tag{1}$$

For a given $\theta \in \Theta$, $B(\cdot; \theta)$ is a function on the unit interval that describes the shape of the parameter instability through time. This shape is parameterized by the finite dimensional parameter θ , so that the parameter space for the time varying version of the model is $(\Gamma_0, \theta) \in \mathbb{R}^k \times \Theta$; in the next subsection, we will specialize this set-up to functions B that correspond to a single shift in the parameter value. The scaling by $T^{-1/2}$ in (1) ensures that even in large samples, there is only a finite amount of information about B , and thus θ . Such asymptotics provide good approximations for small samples when the parameter variability is of the same order of magnitude as sampling uncertainty about Γ .

At this point, we only put some mild constraints on the shape of the parameter instability B (cf. Condition 2 (i) in Li and Müller (2009)).

Condition 2. For all $\theta \in \Theta$, the function $s \mapsto B(s; \theta)$ is a bounded and piece-wise continuous \mathbb{R}^k valued function with at most a finite number of discontinuities, and left and right limits everywhere.

Now by a second order Taylor expansion and Condition 1, one can show that the likelihood ratio statistic $LR_T(\theta)$ between the model with parameter evolution described by (1) and the stable model, with data generated from the stable model, satisfies

$$\begin{aligned} \ln LR_T(\theta) &= T^{-1/2} \sum_{t=1}^T B(t/T; \theta)' s_{T,t}(\Gamma_0) \\ &\quad + \frac{1}{2} T^{-1} \sum_{t=1}^T B(t/T; \theta)' h_{T,t}(\Gamma_0) B(t/T; \theta) + o_p(1) \\ &\Rightarrow \int_0^1 B(s; \theta)' H^{1/2} dW(s) - \frac{1}{2} \int_0^1 B(s; \theta)' HB(s; \theta) ds \end{aligned} \tag{2}$$

where $W(\lambda)$ is a k dimensional standard Wiener process.

Now consider the $k \times 1$ vector continuous time process $G(\cdot)$ on the unit interval

$$G(s) = W(s) + H^{1/2} \int_0^s B(\lambda; \theta) d\lambda. \tag{3}$$

By standard arguments (see, for instance, Kailath (1971)), the log Radon–Nikodym derivative of the measure of G in (3) with parameter $\theta \in \Theta$ relative to the measure ν of the standard Wiener process W , evaluated at W , is given by

$$\ln f_\theta(W) = \int_0^1 B(s; \theta)' H^{1/2} dW(s) - \frac{1}{2} \int_0^1 B(s; \theta)' HB(s; \theta) ds \tag{4}$$

under **Condition 2**. This is the same expression as the r.h.s. of (2), suggesting that the information about θ from observing X_T converges in large samples to the information obtained by observing G .

This intuition is made formally precise in the limit of experiments theory pioneered by LeCam; see, for instance, **van der Vaart (1998)** for an introduction. Formally, we have the following result. Let $C_{[0,1]}^k$ be the space of \mathbb{R}^k valued continuous functions on the unit interval.

Proposition 1. *Let \mathcal{E}_0 be the statistical experiment of observing $G \in C_{[0,1]}^k$ in (3) with parameter space $\theta \in \Theta$. Let \mathcal{E}_T be the statistical experiment of observing X_T with parameter space $\theta \in \Theta$ and known Γ_0 under **Conditions 1** and **2**. Then \mathcal{E}_T converges (weakly) to \mathcal{E}_0 .*

The asymptotic representation theorem (see, for instance, **van der Vaart (1991)**) now implies that for any statistic ξ_T that converges weakly in the time varying parameter model \mathcal{E}_T under all $\theta \in \Theta$, $\xi_T \Rightarrow \xi$, there exists a corresponding statistic ξ_0 in the limit experiment \mathcal{E}_0 with the same distribution $\xi_0 \sim \xi$ under all $\theta \in \Theta$. Thus, any weakly converging statistic in the original model has asymptotic properties that can be “matched” by some statistic in the limit experiment.

We will be especially interested in the implications of this general result for hypothesis tests

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \Theta_1.$$

A (possibly randomized) test in the parametric model is a sequence of measurable functions $\varphi_T : \mathbb{R}^{qT} \mapsto [0, 1]$, with rejection probability $E_\theta[\varphi_T(X_T)] = \int \varphi_T \prod_{t=1}^T f_{T,t}(\Gamma_{T,t}) d\mu_T$. The test is of asymptotic level α if $\sup_{\theta \in \Theta_0} \limsup_{T \rightarrow \infty} E_\theta[\varphi_T(X_T)] \leq \alpha$. The asymptotic representation theorem now implies that if $\varphi_T(X_T) \Rightarrow \xi$ in the time varying parameter model with parameter θ , then there exists a corresponding test $\varphi : C_{[0,1]}^k \mapsto [0, 1]$ with the same asymptotic local power $E_\theta[\varphi(G)] = \int \varphi f_\theta dv = E_\theta[\xi]$.

In many testing problems there does not exist a uniformly most powerful test, even in the limit experiment. If that is the case, then a useful criterion to choose among tests is their weighted average power for some given integrable weighting function F with support in the (closure of) Θ_1 , as in **Wald (1943)** and **Andrews and Ploberger (1994)**. In particular, the weighted average power of a test φ_T in the time varying parameter model of sample size T is $\int E_\theta[\varphi_T(X_T)] dF(\theta)$, and the weighted average power of a test φ in the limit experiment is $\int E_\theta[\varphi(G)] dF(\theta)$. The following proposition is a simple corollary of theorem 15.1 of **van der Vaart (1998)**.

Proposition 2. *Suppose **Conditions 1** and **2** hold, and let $\varphi_T : \mathbb{R}^{qT} \mapsto [0, 1]$ be a test such that $E_\theta[\varphi_T(X_T)] \rightarrow \pi(\theta)$ for all $\theta \in \Theta$, and that φ_T is of asymptotic level α .*

(i) *If $\varphi^* : C_{[0,1]}^k \mapsto [0, 1]$ maximizes weighted average power $\int E_\theta[\varphi(G)] dF(\theta)$ among all level α tests φ in \mathcal{E}_0 , then*

$$\limsup_{T \rightarrow \infty} \int E_\theta[\varphi_T(X_T)] dF(\theta) \leq \int E_\theta[\varphi^*(G)] dF(\theta).$$

(ii) *Suppose there does not exist a level α test $\varphi : C_{[0,1]}^k \mapsto [0, 1]$ in \mathcal{E}_0 with weighted average power $\int E_\theta[\varphi(G)] dF(\theta)$ larger than $\bar{\pi}$. Then*

$$\limsup_{T \rightarrow \infty} \int E_\theta[\varphi_T(X_T)] dF(\theta) \leq \bar{\pi}.$$

Proposition 2 formalizes the notion that the limit experiment provides bounds on the asymptotic performance of tests in the underlying parametric model. Part (ii) is of particular interest in this paper, as the next sections derive upper bounds on power in the limit experiment. Note that both parts of **Proposition 2** are not constructive, but only provide limits on how well tests in the underlying small sample problem may perform asymptotically. At the same time, it seems that for a given φ^* of part (i), one should be able to construct tests in \mathcal{E}_T with the same asymptotic power as φ^* .

We describe the translation of a good test in the limit experiment \mathcal{E}_0 to a test with good asymptotic properties in \mathcal{E}_T in Section 4.

2.2. Limiting problem for post break parameter inference

The set-up of the last subsection is quite general and can potentially accommodate many inference issues in time varying parameter models. In this paper we focus on the particular issue of tests about the post break parameter value. Thus, let B in (1) be of the following form

$$B(s; \theta) = H^{-1/2} \beta + H^{-1/2} \delta \mathbf{1}[s \leq \rho] \tag{5}$$

where $\theta = (\beta, \delta, \rho) \in \mathbb{R}^k \times \mathbb{R}^k \times [0.15; 0.85]$. Under (5), the parameter $\Gamma_{T,t}$ is equal to $\Gamma_0 + T^{-1/2} H^{-1/2} (\beta + \delta)$ for $t \leq \lfloor \rho T \rfloor$, and equal to $\Gamma_0 + T^{-1/2} H^{-1/2} \beta$ for $t > \lfloor \rho T \rfloor$. As in previous studies on break models, such as **Andrews (1993)** or **Bai and Perron (1998)**, we assume that the break date ρ , expressed in fractions of the sample size, is not too close to the boundaries, and restrict $\rho \in [0.15; 0.85]$. The normalization by $H^{-1/2}$ in (5) ensures that the magnitudes of β and δ are expressed in multiples of the limiting standard deviation of the full sample maximum likelihood estimator in the stable model.

Suppose the scalar parameter of interest is the first element of Γ . The null hypothesis that its post break value is equal to the first element of Γ_0 then corresponds to the hypothesis

$$H_0 : v' \beta = 0 \tag{6}$$

where v' is proportional to the first row of $H^{-1/2}$, normalized such that $v'v = 1$. Let R_v be a $k \times (k - 1)$ matrix such that (v, R_v) is full rank. The null hypothesis (6) has $2k$ nuisance parameters: $R_v' \beta \in \mathbb{R}^{k-1}$, $\delta \in \mathbb{R}^k$ and $\rho \in [0.15; 0.85]$.

The process (5) satisfies **Condition 2** for all $\theta \in \Theta$, so that the limit experiment is given by

$$G(s) = W(s) + \beta s + \delta \min(\rho, s), \quad s \in [0, 1] \tag{7}$$

with associated density (cf. (4))

$$f_\theta(G) = \exp \left[\delta' G(\rho) + \beta' G(1) - \frac{1}{2} (\beta' \beta + 2\rho \beta' \delta + \rho \delta' \delta) \right]. \tag{8}$$

To heuristically motivate the form of the limit experiment (7), consider observations $\{y_t\}_{t=1}^T$ from a scalar Gaussian model with a break in the mean at time τ

$$y_t = \mu + \Delta \mathbf{1}[t \leq \tau] + \varepsilon_t, \quad \varepsilon_t \sim iid \mathcal{N}(0, H^{-1}). \tag{9}$$

We then have the following equality in distribution

$$T^{-1/2} H^{1/2} \sum_{t=1}^{sT} y_t \sim G(s) \quad \text{for any } s \in \left\{ \frac{t}{T} \right\}_{t=1}^T$$

where $\rho = \tau/T$, $\beta = T^{1/2} H^{1/2} \mu$ and $\delta = T^{1/2} H^{1/2} \Delta$. Except for the scaling by $T^{-1/2}$ and the discretization $s \in \{ \frac{t}{T} \}_{t=1}^T$, the testing problem (6) involving the observation G is therefore identical to inference about the post-break mean in the Gaussian location model (9). It is also straightforward to show that model (9) satisfies **Condition 1**, so that observing G is indeed also the formal limit experiment via **Proposition 1**.

2.3. Scalar limit experiment

Initially, we consider the limit experiment associated with an underlying model that only contains a scalar parameter Γ , that is $k = 1$. We treat the vector case in Section 2.4.

In this section, we seek to derive good tests for the value of the post mean parameter in this scalar limit experiment. In the notation developed in the last subsection, this corresponds to the hypothesis test

$$H_0 : \theta = (\beta, \delta, \rho) \in \{0\} \times \mathbb{R} \times [0.15; 0.85] = \Theta_0$$

$$\text{vs } H_1 : \theta = (\beta, \delta, \rho) \in \mathbb{R}\{0\} \times \mathbb{R} \times [0.15; 0.85] = \Theta_1. \tag{10}$$

Geometrically, the null hypothesis asserts that the slope of the deterministic component $\beta s + \delta \min(\rho, s)$ of $G(s)$ is equal to zero after the (potential) kink at $s = \rho$.

We consider two approaches. First, we analyze the properties of tests that estimate the break date ρ by least squares, and then take that estimate as the true value. This approach corresponds to the suggestion by Bai (1994, 1997) and Bai and Perron (1998) for a linear regression setting, such as the simple mean shift example (9). We find that this way of conducting inference does very poorly under small and moderate break magnitudes ($|\delta| < 12$, say). Second, we apply the general approach of EMW to derive nearly weighted average power maximizing tests to this specific testing problem.

2.3.1. Tests based on least-squares break date estimator

As discussed in the introduction, Bai (1994, 1997) and Bai and Perron (1998) suggest and analyze the following procedure for conducting inference about the post break value of a coefficient in a linear regression that is subject to a break: estimate the break date by least-squares, construct a post break dummy using this estimate, and perform the usual t -test on that dummy. These papers show that this procedure results in asymptotically correct inference for break magnitudes that diverge when multiplied by the square root of the sample size. This corresponds to $|\delta| \rightarrow \infty$ in the notation developed here. Another procedure, arguably most prevalent in applied work, first performs a pretest for a break on some conventional significance level, followed by standard full sample inference about the post break value if no significant break is found, and performs least-squares based post break inference as above if the pretest rejects.

In the scalar limit experiment (7), the nominally 5% level least squares and pretest procedure (based on the supF statistic) correspond to the test

$$\mathbf{1}[\text{supF} > \text{cv}_{pre}(\alpha_{pre})] \mathbf{1}[|\hat{t}| > 1.96] + \mathbf{1}[\text{supF} < \text{cv}_{pre}(\alpha_{pre})] \mathbf{1}[|G(1)| > 1.96] \tag{11}$$

where $\text{cv}_{pre}(\alpha_{pre})$ is the critical value of the supF statistic of level α_{pre} as tabulated in Andrews (1993), (and $\text{cv}_{pre}(1) = 0$ corresponds to always using the least squares break date estimator, with no pretest), and

$$\hat{\rho} = \underset{r \in [0.15; 0.85]}{\text{argmax}} \frac{(G(r) - rG(1))^2}{r(1-r)}, \tag{12}$$

$$\text{supF} = \frac{(G(\hat{\rho}) - \hat{\rho}G(1))^2}{\hat{\rho}(1-\hat{\rho})}, \quad \hat{t} = \frac{G(1) - G(\hat{\rho})}{\sqrt{1-\hat{\rho}}}$$

Under standard assumptions, the large sample properties of the small sample pretest and least square based tests in a linear regression with a single coefficient converge in distribution to those in (11) for breaks of magnitude $T^{-1/2}$ —this follows, for instance, from Proposition 1 of Elliott and Müller (2007).

Fig. 2 displays the null rejection probability of (11) for $\alpha_{pre} = 1$ (no pretest), $\alpha_{pre} = 0.10$, $\alpha_{pre} = 0.05$ and $\alpha_{pre} = 0.01$ as a function of δ for selected values of ρ , based on 50,000 Monte Carlo draws and 1,000 step random walk approximations to Wiener processes. None of these tests comes close to controlling size uniformly over δ . The approximately largest null rejection probability of the pure least-squares break date based test is approximately 29% at $(\rho, \delta) = (0.85, 2.6)$. Pre-testing for a break seems to both substantially exacerbate and shift the size control problem to larger values of δ .³

In fact, to obtain a 5% level test in the no pretest case, one must employ a critical value of approximately 4.9 instead of 1.96 for the t -statistic, with size still approximately equal to the nominal 5% level at the point $(\rho, \delta) = (0.85, 6.8)$. Denote this size corrected

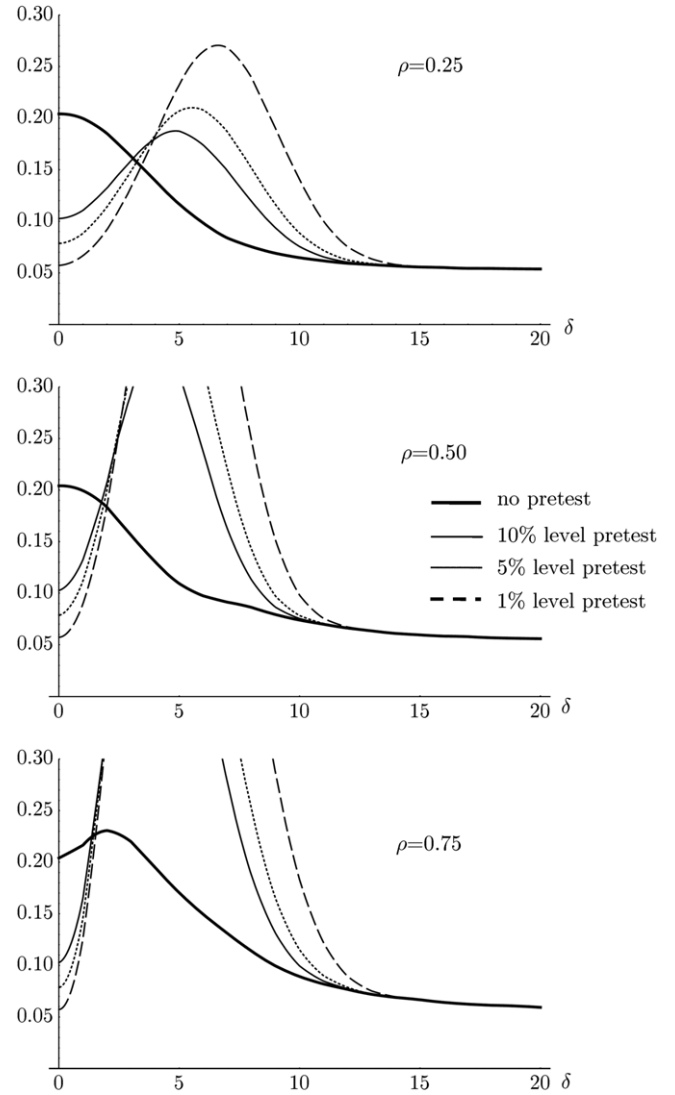


Fig. 2. Asymptotic null rejection probabilities of nominal 5% level tests for the post break parameter value based on the least squares break date estimator.

test φ_{LS} . Because the worst size distortion occurs for a strictly positive break magnitude, the size corrected hybrid subsampling method of Andrews and Guggenberger (2009, 2010b) reduces to the size corrected fixed critical value test φ_{LS} . Alternatively, one could always use the test $\varphi_{0.85}(G) = \mathbf{1}[|G(1) - G(0.85)|/\sqrt{0.15} > 1.96]$ which corresponds to a usual 5% level t -test based on the last 15% of the data. As one would expect, though, the power of both φ_{LS} and $\varphi_{0.85}$ is quite poor (we provide power computations in Fig. 4), motivating the construction of an alternative test.

2.3.2. Approximately weighted average power maximizing test

The testing problem (10) has exactly the structure of testing problems analyzed in EMW: it is a non-standard but parametric testing problem with composite null and alternative hypotheses. We thus follow their general approach to determine a test that comes close to maximizing a weighted average power criterion.

As discussed in EMW, in problems of this kind the weighted average power maximizing test rejects for large values of a likelihood ratio type statistic. The numerator in this ratio is the likelihood f_θ integrated over the weighting function F , which can be chosen freely by the researcher. The intuition here is that the weighted average power criterion $\int E_\theta[\varphi(G)]dF(\theta) = \int \int \varphi_\theta d\nu dF(\theta)$ is, by a change of the order of integration, equivalent to $E_h[\varphi(G)] =$

³ Unreported results show that these asymptotic results provide very good approximations for the small sample Gaussian location problem (9) with $T = 100$.

$\int \varphi h d\nu$, where $h = \int f_{\theta} dF(\theta)$. Thus, maximizing weighted average power amounts to maximizing power against the single mixture density h .

The denominator in the likelihood ratio statistic is the likelihood integrated over the so-called “least favorable distribution” Λ with support in Θ_0 , so that the best test rejects for large values of $h / \int f_{\theta} d\Lambda(\theta)$. The distribution Λ is endogenous to the problem and depends on the weighting function F and the level of the test. Its name may be understood by thinking of a game between the econometrician and nature: nature selects the parameter value under the null hypothesis to make it as hard as possible for the econometrician to devise a powerful level α test between the null hypothesis and h . In general, nature’s best strategy is randomized, and described by the least favorable distribution.

Now in many non-standard problems, such as the one under consideration here, it is very difficult to analytically derive the form of the least favorable distribution. For this reason, EMW devise an algorithm to numerically determine an approximately least favorable distribution Λ^* as a mixture of a finite set of base distributions with support in Θ_0 . The distribution Λ^* is useful for two purposes: on the one hand, it implies an upper bound on weighted average power of any level α test. The reasoning is that the econometrician’s best response to any given randomized strategy of nature provides an upper bound for the power against nature’s best randomized strategy. On the other hand, by EMW’s definition of an approximately least favorable distribution Λ^* , the level α test that rejects for large values of the likelihood ratio statistic $h / \int f_{\theta} d\Lambda^*(\theta)$ has power only slightly below this upper bound. For practical purposes, this feasible test thus comes numerically close to solving the problem of determining the weighted average power maximizing test.

In order to implement EMW’s algorithm, we thus need to choose a weighting function F and select a set of base distributions that discretizes the null parameter space Θ_0 .

We choose F such that ρ is uniformly distributed on $[0.15, 0.85]$ and (β, δ) is bivariate normal independent of ρ , with $\beta \sim \mathcal{N}(0, \sigma_{\beta}^2)$, $\delta \sim \mathcal{N}(0, \sigma_{\delta}^2)$ and β independent of $\beta + \delta$, where $\sigma_{\beta}^2 = 22$ and $\sigma_{\delta}^2 = 400$. The Gaussianity of the marginal distribution of β and δ in F is chosen for analytical convenience, and the independence of the pre and post break parameter values β and $\beta + \delta$ mirrors the independence of the pre and post break sample information $G(\rho)$ and $G(1) - G(\rho)$. The value of $\sigma_{\beta}^2 = 22$ is motivated by King’s (1988) discussion of overall reasonable point-optimal tests, since it turns out that for $\sigma_{\beta}^2 = 22$, the best 5% level test has approximately 50% weighted average power. The uniform weighting over ρ accords to the choice in Andrews and Ploberger (1994) and is intended to generate reasonable power for all break dates $\rho \in [0.15, 0.85]$. Finally, the value $\sigma_{\delta}^2 = 400$ is motivated as follows: for large δ , (say, $|\delta| > 12$ or so) there is a lot of information about the break date ρ , and good tests can come close to performing as well as if ρ was known. By putting enough weight on rather large values of δ , the choice $\sigma_{\delta}^2 = 400$ ensures that tests that have large weighted average power share this desirable feature. At the same time, the distribution $\mathcal{N}(0, 400)$ concentrates about half of its mass on $|\delta| < 12$, so tests with large weighted average power also perform reasonably well when there is only limited sample information about the break date ρ .

We choose the base distributions to be of the form “ ρ is uniformly distributed on $[\max(0.15, a/100 - \frac{1}{2}), \min(0.85, b/100 + \frac{1}{2})]$ and δ is an equal probability mixture of $\delta \sim \mathcal{N}(\mu_{\delta}, \sigma_{\delta}^2)$ and $\delta \sim \mathcal{N}(-\mu_{\delta}, \sigma_{\delta}^2)$ ” for some integers $a, b \in \{15, \dots, 85\}$, $b \geq a$, and $\mu_{\delta} \in \mathbb{R}$, $\sigma_{\delta}^2 > 0$.

The algorithm now proceeds as follows. Set the candidate Λ^* equal to some base distribution, and construct the optimal test against the weighted average alternative distribution h . Slightly

adjust the critical value of this test upwards so that it has null rejection probability slightly below α . Check whether the adjusted test controls size for data drawn from the other base null distributions. If it does not, determine a new Λ^* that is a mixture of the previous Λ^* and the violating base distribution. Iterate until the adjusted test controls size uniformly over all base distributions.

The result of this process is a test that has power that is close to the power bound, and that controls size under all mixtures of base distributions. It may not, however, control size for all possible values of the nuisance parameters (δ, ρ) . So the final step is to check size control. If size control is not obtained, then additional null base distributions with smaller σ_{δ}^2 and/or smaller values of $b - a$ are added to the set of base distributions, and the iterative process goes through another round. Table 6 in the Appendix describes the sets of base distributions in this process in detail.

In the application of this paper, we determine a level α test φ^* in the limit experiment with weighted average power close to the upper bound of a slightly simplified form. The computation of rejection probabilities requires the simulation of (pseudo) random continuous time processes G . Discrete time Gaussian random walk approximation based on, say, 1000 steps are computationally expensive and of uncertain accuracy. For this reason, we approximate all integrals over ρ by sums with $\rho \in \{1/100, 2/100, \dots, 1\}$. The rejection probability of the (approximate) likelihood ratio type test thus becomes a function of

$$\left\{ G\left(\frac{l}{100}\right) \right\}_{l=1}^{100}. \tag{13}$$

The 100 random variables (13) can be simulated without any approximation error, as they are exactly jointly Gaussian. Also, looking ahead, implementation of this test only requires evaluation of an appropriate partial sum process at the 71 values 0.15, 0.16, ..., 0.85, which numerically simplifies applications in potentially computationally demanding GMM problems.

In addition, we modify the behavior of the test when the break is obviously large. The motivation is as follows: by construction, the level α test that rejects for large values of $h / \int f_{\theta} d\Lambda^*(\theta)$ comes close to maximizing weighted average power over F . But F puts very little mass on very large breaks (say, $|\delta| > 40$). Thus, the test might perform quite poorly there. At the same time, for a large break $|\delta|$, there is ample information about the break date ρ . With ρ known, inference about the post break value is a standard problem—simply reject for the analogue of a t -test based on the last $1 - \rho$ fraction of the data, that is $\varphi_{\rho}(G) = \mathbf{1}[|t_{\rho}| > 1.96]$, where $t_{\rho} = (G(1) - G(\rho)) / \sqrt{1 - \rho}$. The suggestion of EMW for problems of this kind is to switch between the standard test if there is overwhelming evidence that $|\delta|$ is large, and the likelihood ratio type test $h / \int f_{\theta} d\Lambda^*(\theta)$ otherwise. Numerical calculations suggest that once the supF statistic (12) takes on very large values, inference based on $\mathbf{1}[|\hat{t}| > 1.96]$ is nearly correct. Thus, whenever $\text{supF} > 90$, we switch to using the test statistic \hat{t} in (12), but with critical value of 2.01 rather than 1.96 on the 5% level. The adjustment of the critical value is to ensure overall size control, including the switching.

These modifications are only applied to the nearly weighted average power maximizing test φ^* , however. The computation of the power bound does not involve any switching and requires evaluation of and integration over the actual densities (8), which we approximate by Riemann sums with $8 \cdot 10^5$ steps.

Fig. 3 plots the approximate least favorable distribution obtained by the algorithm. It is a mixture of 18 base distributions; see Table 2 for the corresponding weights and values of $(a, b, \mu_{\delta}, \sigma_{\delta}^2)$.

Proposition 3. In model (7) with $k = 1$,
 (i) any 5% level test φ of (10) has $\int E_{\theta}[\varphi(G)]dF(\theta) \leq \bar{\pi} \approx 50.0\%$;
 (ii) the (nonrandomized) test φ^* defined in the Appendix is of level $\alpha^* \approx 5\%$, and $\int E_{\theta}[\varphi^*(G)]dF(\theta) = \pi^* \approx 49.0\%$.

The numbers for $\bar{\pi}$, α^* and π^* in Proposition 3 are estimated based on 50,000 independent Monte Carlo draws.

Table 2
Approximate least favorable distribution.

j	1	2	3	4	5	6	7	8	9
p_j	0.588	0.123	0.067	0.057	0.038	0.032	0.026	0.02	0.009
a_j	15	85	85	20	75	20	20	75	45
b_j	85	85	85	74	85	74	74	82	59
$\sigma_{\delta,j}^2$	100	10	4	300	200	10	3	10	10
μ_j	20	5	3	16	28	9	6	7	11
j	10	11	12	13	14	15	16	17	18
p_j	0.009	0.008	0.006	0.005	0.004	0.004	0.002	0.001	0.001
a_j	70	15	15	60	80	60	83	85	75
b_j	74	19	24	69	82	69	84	85	82
$\sigma_{\delta,j}^2$	10	10	200	10	10	3	10	3	3
μ_j	9	5	28	12	11	8	13	15.5	13

Notes: The least favorable distribution is a mixture of $j = 1, \dots, 18$ base distributions. Under each base distribution, ρ is uniform on $[\max(0.15, a/100 - \frac{1}{2}), \min(0.85, b/100 + \frac{1}{2})]$ and δ is an equal probability mixture of $\delta \sim N(\mu_\delta, \sigma_\delta^2)$ and $\delta \sim N(-\mu_\delta, \sigma_\delta^2)$. The values in this table also enter the definition of the nearly weighted average power maximizing test $\hat{\varphi}^*$ of Section 4.1.

Fig. 4 plots, for selected values of ρ and δ , the power of the suggested test $\hat{\varphi}^*$, of the size corrected least squares break date based test φ_{LS} , of the test using the last 15% of the sample $\varphi_{0.85}$ only, and the infeasible test using actual post break data $\varphi_\rho = \mathbf{1}[|t_\rho| > 1.96]$ introduced above. All tests are invariant under the transformation $G \rightarrow -G$, so there is no need to plot the power for negative δ . As one might expect given the large critical value of 4.9, φ_{LS} has very low power. In contrast, $\hat{\varphi}^*$ is almost always more powerful than $\varphi_{0.85}$, often substantially so, and comes close to the power of the infeasible benchmark φ_ρ for large $|\delta|$.

2.4. Multivariate limit experiment

We now return to the vector limit experiment (6) with $k > 1$. In principle, one could again try to apply EMW’s algorithm directly, but due to the high-dimensional nuisance parameter space with $2k$ parameters, this is computationally difficult.

Instead, note that one could always choose to ignore the information contained in G beyond the scalar process $v'G$. But the testing problem treating $v'G$ as the only observation is identical to the one treated in the last subsection. In particular, simply applying $\hat{\varphi}^*$ of Proposition 3 to $v'G$ obviously yields a valid 5% level test, with a rejection probability that only depends on $\rho \in [0.15; 0.85]$, $v'\beta \in \mathbb{R}$ and $v'\delta \in \mathbb{R}$.

Now it seems that ignoring the information in G beyond $v'G$ is very inefficient. But the following proposition shows that this approach is nearly admissible.

Proposition 4. In model (7) with $k > 1$, let F_v be the probability distribution for θ such that $(v'\beta, v'\delta, \rho)$ has distribution F of Section 2.3.2, and $R'_v\beta = R'_v\delta = 0$ almost surely under F_v . Then

- (i) any 5% level test $\hat{\varphi}$ of (6) has $\int E_\theta[\varphi(G)]dF_v(\theta) \leq \bar{\pi} \simeq 50.0\%$;
- (ii) the test $\hat{\varphi}^*(v'G)$ of (6), with $\hat{\varphi}^*$ defined as in Proposition 3, is of level $\alpha^* \simeq 5\%$, and $\int E_\theta[\hat{\varphi}^*(v'G)]dF_v(\theta) = \pi^* \simeq 49.0\%$.

Proposition 4 shows that for a particular weighting function F_v that is closed related to F on Section 2.3.2, no 5% level test can have weighted average power larger than $\bar{\pi}$. At the same time, $\hat{\varphi}^*$ applied to $v'G$ has weighted average power with respect to F_v equal to π^* , which is only slightly smaller than $\bar{\pi}$. Thus $\hat{\varphi}^*(v'G)$ is nearly admissible in the multivariate problem: no level α test can exist that has substantially higher power uniformly over $\theta \in \Theta \subset \mathbb{R}^{2k+1}$. What is more, in a more general limit experiment where the last $k - 1$ elements of Γ undergo breaks at a different time, or are subject to multiple breaks, the test $\hat{\varphi}^*(v'G)$ still controls size, as its null rejection properties are entirely unaffected by the form of parameter instabilities in parameters other than the parameter of interest.

We consider the combination of near admissibility and this robustness as quite attractive and will thus implement $\hat{\varphi}^*(v'G)$ in the multivariate problem.

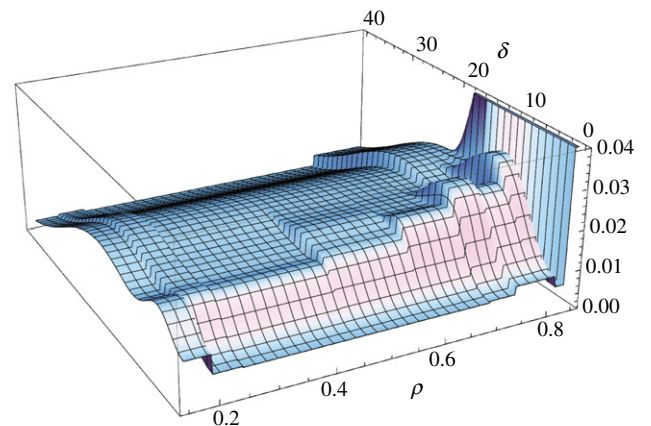


Fig. 3. Approximate least favorable distribution A^* .

3. GMM models

The last section discussed the problem of deriving asymptotically powerful inference about the post break parameter value using LeCam’s limits of experiments theory. The theory derived there applies to any well-behaved parametric model, such as, say, the post-break parameters of a Gaussian ARMA(p, q) model (see Section 4.3 for an illustration), or post break parameter values of a logit model with time series data.

The results do not, however, apply to general semiparametric models, such as those defined through moment conditions. It is in principle possible to apply limits of experiments ideas also to semiparametric models (see, for instance, Choi et al. (1996) for an application to semiparametric LAN models), but the non-standard nature of the testing problem here makes this an involved enterprise beyond the scope of the paper. As a more straightforward alternative, we now apply the asymptotic efficiency concept of tests introduced by Müller (2011) to post break parameter inference in GMM models.

Using data $X_T = (x_{1,T}, \dots, x_{T,T}) \in \mathbb{R}^{qT}$, let the moment condition be the \mathbb{R}^m -valued function $g(\cdot, \cdot)$, so that $E[g(x_{T,t}, \Gamma_t)] = 0$ when the true parameter at date t is given by $\Gamma = \Gamma_t$. Consider first the stable version of the GMM model with true parameter value Γ_0 . Then full sample efficient GMM estimation is based on

$$\hat{\Gamma} = \arg \min_{\Gamma} \left(\sum_{t=1}^T g(x_{T,t}, \Gamma) \right)' \hat{V} \left(\sum_{t=1}^T g(x_{T,t}, \Gamma) \right)$$

where \hat{V} is possibly data dependent $m \times m$ positive definite weighting matrix that converges to the inverse of the asymptotic variance

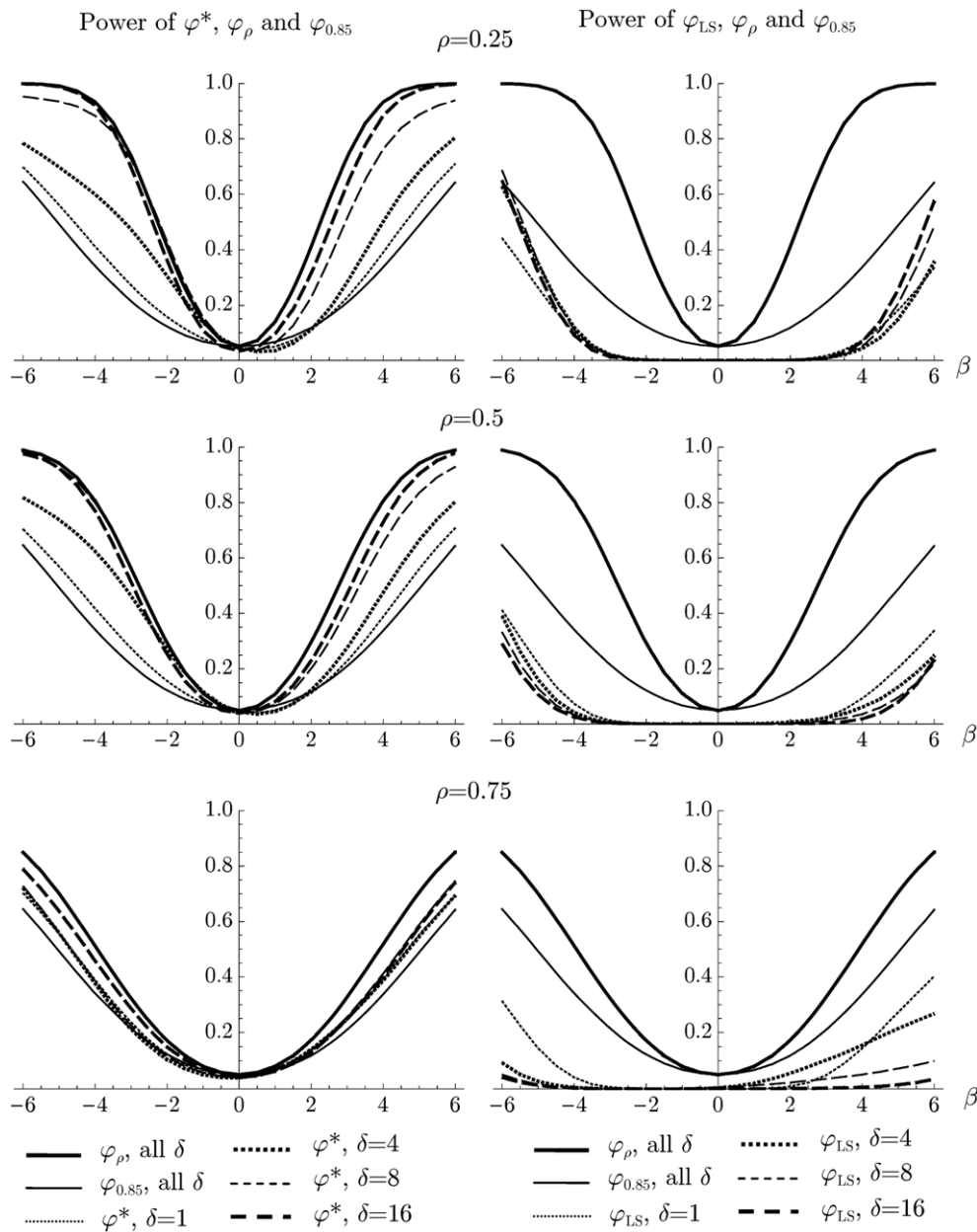


Fig. 4. Asymptotic power of the approximately optimal test φ^* and of the size corrected least-squares date estimator based test φ_{LS} , compared to the infeasible test φ_ρ with break date known, and the test using only the last 15% of the observations $\varphi_{0.85}$.

V^{-1} of $T^{-1/2} \sum_{t=1}^T g(x_{T,t}, \Gamma_0)$, and

$$\sqrt{T}(\hat{\Gamma} - \Gamma_0) \Rightarrow \mathcal{N}(0, H^{-1})$$

where $H^{-1} = (\tilde{\gamma}'V\tilde{\gamma})^{-1}$, $T^{-1} \sum_{t=1}^T \gamma(x_{T,t}, \Gamma_0) \xrightarrow{p} \tilde{\gamma}$ and the $\mathbb{R}^{k \times m}$ valued function $\gamma(\cdot, \Gamma)$ is the partial derivative of g with respect to Γ . Further, let $\hat{\Omega}$ be a consistent estimator of the variance of $\hat{\Gamma}$, in the sense that $T\hat{\Omega} \xrightarrow{p} H^{-1}$, and $\hat{\Omega}^{-1/2}(\hat{\Gamma} - \Gamma_0) \Rightarrow \mathcal{N}(0, I_k)$.

Now consider a time varying parameter version of this GMM model, where the parameter evolves according to (1) and (5),

$$\Gamma_{T,t} = \Gamma_0 + T^{-1/2}H^{-1/2}(\beta + \delta \mathbf{1}[t/T \leq \rho]) \tag{14}$$

so that the (local) parameter instability is parameterized by $\theta = (\beta, \delta, \rho) \in \Theta$, just as in the previous section. Conditional on $\tau = \lfloor \rho T \rfloor$, the natural estimators for the pre and post break value of Γ are given by the efficient GMM estimator using pre and post

break data,

$$\hat{\Gamma}_{pre}(\tau) = \arg \min_{\Gamma} \left(\sum_{t=1}^{\tau} g(x_{T,t}, \Gamma) \right)' \hat{V}_{pre}(\tau) \left(\sum_{t=1}^{\tau} g(x_{T,t}, \Gamma) \right)$$

$$\hat{\Gamma}_{post}(\tau) = \arg \min_{\Gamma} \left(\sum_{t=\tau+1}^T g(x_{T,t}, \Gamma) \right)' \hat{V}_{post}(\tau) \left(\sum_{t=\tau+1}^T g(x_{T,t}, \Gamma) \right)$$

with $\hat{V}_{pre}(\tau)$ and $\hat{V}_{post}(\tau)$ the appropriate $m \times m$ positive definite weighting matrices, and associated covariance matrix estimators $\hat{\Omega}_{pre}(\tau)$ and $\hat{\Omega}_{post}(\tau)$ of $\hat{\Gamma}_{pre}(\tau)$ and $\hat{\Gamma}_{post}(\tau)$, respectively. A natural starting point for post break parameter value inference in the absence of knowledge about $\rho \in [0.15; 0.85]$ is arguably the information contained in $\{\hat{\Omega}_{pre}(\tau), \hat{\Omega}_{post}(\tau), \hat{\Gamma}_{pre}(\tau), \hat{\Gamma}_{post}(\tau)\}$ for $[0.15T] \leq \tau \leq [0.85T]$. In other words, it seems natural to try to learn about the post parameter value from the relevant set of

“partial sample” parameter GMM estimates (along with their standard error estimates) in the jargon developed in Andrews (1993).

In large samples, the set of estimators $\{\hat{\Gamma}_{pre}(\lfloor \cdot T \rfloor), \hat{\Gamma}_{post}(\lfloor \cdot T \rfloor), \hat{\Omega}_{pre}(\lfloor \cdot T \rfloor), \hat{\Omega}_{post}(\lfloor \cdot T \rfloor)\}$ typically satisfies the following weak convergence

$$\begin{aligned}
 Y_T(s) &= \begin{pmatrix} T^{1/2}(\hat{\Gamma}_{pre}(\lfloor sT \rfloor) - \Gamma_0) \\ T^{1/2}(\hat{\Gamma}_{post}(\lfloor sT \rfloor) - \Gamma_0) \\ T\hat{\Omega}_{pre}(\lfloor sT \rfloor) \\ T\hat{\Omega}_{post}(\lfloor sT \rfloor) \end{pmatrix} \\
 \Rightarrow Y(s) &= \begin{pmatrix} H^{-1/2}G(s) \\ H^{-1/2}(G(1) - G(s)) \\ sH^{-1} \\ (1-s)H^{-1} \end{pmatrix}
 \end{aligned}
 \tag{15}$$

where G is distributed as in (7), and the convergence is in the space of $2k + 2k^2$ valued cadlag functions on the interval $[0.15; 0.85]$. We discuss relevant primitive conditions below Proposition 5 in Section 4.

Now consider tests of the null hypothesis that the post break value of the first element of Γ is equal to the first element of Γ_0 . Under (14), just as in Section 2.2, this corresponds to the composite null hypothesis

$$H_0 : v'\beta = 0
 \tag{16}$$

where v' is the normalized first row of $H^{-1/2}$.

Now one might demand that a test $\varphi_T(X_T)$ does not overreject, at least asymptotically, whenever the GMM model is well enough behaved for (15) to hold. Formally, Müller (2011) introduces the following definition: a test $\varphi_T : \mathbb{R}^{qT} \mapsto [0, 1]$ of (16) is considered *robust* if $\limsup_{T \rightarrow \infty} E_\theta[\varphi_T(X_T)] \leq 0.05$ for all $\theta \in \Theta$ with $v'\beta = 0$, for any sequence of distributions of X_T that induces the weak convergence (15). Theorem 1 in Müller (2011) shows that under this constraint, the asymptotically weighted average power maximizing test based on X_T is simply the corresponding weighted average power maximizing test in the limiting problem of observing $Y(\cdot)$ in (15), evaluated at the sample analogue $Y_T(\cdot)$. In contrast to the more traditional LeCam-type arguments of Section 4.1, no additional regularity assumptions on the likelihood or other properties of X_T are required for this efficiency claim. At the same time, the robustness requirement is arguably very strong, as there are many ways in which (15) may hold, correspondingly weakening the strength of the claim. See Müller (2011) for further discussion on the merits of this notion of efficiency.

The limiting problem of observing $Y(\cdot)$ in (15) is recognized as (essentially) identical to the limit experiment considered in Sections 2.3 and 2.4 (it is potentially less informative, as observing $Y(\cdot)$ on $[0.15; 0.85]$ is equivalent to observing $G(\cdot)$ on $[0.15; 0.85]$, rather than observing $G(\cdot)$ on the entire unit interval). Thus, the bound $\bar{\pi} \simeq 50.0\%$ on weighted average power derived in Propositions 3 and 4 also applies to tests of (16) based on X_T under Müller’s (2011) robustness constraint relative to (15). In Section 4, we discuss in detail the construction of the small sample analogue $\hat{\varphi}^*$, and show in Proposition 5 that its asymptotic rejection profile is equal to the rejection profile of φ^* whenever (15) holds, so its asymptotic weighted average power is equal to $\pi^* \simeq 49.0\%$. Thus, $\hat{\varphi}^*$ is the nearly asymptotic weighted average power maximizing test among all tests that are robust in the sense of Müller (2011) relative to (15).

4. Test statistic and properties

This section gives the precise form of the test statistic and provides the asymptotic justification for the test as well as some evidence on its small sample performance. Section 4.1 details the

construction of the statistic, and is the relevant subsection for those wishing to apply the method to data. The suggested statistic applies to inference on a single element of a $k \times 1$ parameter vector in a GMM problem. Hence the method applies to linear and nonlinear regression, maximum likelihood models with differentiable likelihoods, and structural and reduced form models. The method takes as its input the output of standard GMM estimation over various subsamples, as detailed below. In Section 4.2 we examine the large sample properties of the statistic, establish limit results and show that the test is consistent under local and non-local break magnitudes. The conditions on the data and models are similar to the set of assumptions in the breaks literature (for example Andrews (1993)), allowing for time heterogeneous data and a wide range of models, but ruling out unit (or near unit) roots and time trends. We examine small sample size and power properties in Monte Carlo experiments in Section 4.3 in a mean estimation problem, for inference about a parameter estimated by maximum likelihood, and in an instrumental variables regression problem.

4.1. Definition of test statistic

We define the statistic in a general method of moments framework with the k dimensional parameter $\Gamma = (\gamma, \tilde{\gamma})'$, with $\gamma \in \mathbb{R}$ and $\tilde{\gamma} \in \mathbb{R}^{k-1}$. The data in a sample of size T is given by $X_T = (x_1, \dots, x_T)$, and the population moment condition is $E[g(x_t, \Gamma_0)] = 0$ when the true parameter at date t is given by $\Gamma = \Gamma_0$ for some known, \mathbb{R}^m valued function $g(\cdot, \cdot)$. For example in a linear IV/2SLS problem we might have $g(x_t, \Gamma_0) = (x_{1t} - \Gamma_0'x_{2t})x_{3t}$ where the scalar variable x_{1t} is regressand on the vector x_{2t} , using x_{3t} as instruments. The weight matrix may be chosen in the usual way and, for the validity of our test, does not have to be the efficient one.

The parameter Γ changes its value from $\Gamma_{pre} = (\gamma_{pre}, \tilde{\gamma}'_{pre})'$ to $\Gamma_{post} = (\gamma_{post}, \tilde{\gamma}'_{post})'$ at some unknown date $\tau \in \mathbb{N}$ in the middle 70% of the sample, $\tau \in [0.15T, 0.85T]$ (the methods apply for any size of the change so are applicable where subsets or even all of the parameters in Γ do not change). The hypotheses of interest concerns the post break value of the first element γ of Γ ,

$$H_0 : \gamma_{post} = \gamma_{post,0} \quad \text{against} \quad H_1 : \gamma_{post} \neq \gamma_{post,0}.
 \tag{17}$$

Denote by $\hat{\Gamma}_{pre}(t)$ and $\hat{\Gamma}_{post}(t)$ standard GMM estimators of Γ using data $\{x_s\}_{s=1}^t$ and $\{x_s\}_{s=t+1}^T$, and denote by $\hat{\Omega}_{pre}(t)$ and $\hat{\Omega}_{post}(t)$ the estimators of the covariance matrix of $\hat{\Gamma}_{pre}(t)$ and $\hat{\Gamma}_{post}(t)$, respectively. These covariance matrix estimators, as well as the weighting matrix for efficient GMM estimation in the overidentified case, are based on data $\{x_s\}_{s=1}^t$ and $\{x_s\}_{s=t+1}^T$ for the pre and post break estimators. If necessary, the estimators $\hat{\Omega}_{pre}(t)$ and $\hat{\Omega}_{post}(t)$ account for serial correlation by employing a correction as in Newey and West (1987) or Andrews (1991), for instance, so that in an overall stable model with parameter Γ_0 , approximately, $\hat{\Omega}_{pre}(t)^{-1/2}\hat{\Gamma}_{pre}(t) \sim \mathcal{N}(\Gamma_0, I_k)$ and $\hat{\Omega}_{post}(t)^{-1/2}\hat{\Gamma}_{post}(t) \sim \mathcal{N}(\Gamma_0, I_k)$.

Our test statistic of (17) only requires evaluation of $\{\hat{\Gamma}_{pre}(t), \hat{\Gamma}_{post}(t), \hat{\Omega}_{pre}(t), \hat{\Omega}_{post}(t)\}$ at the 71 numbers $t = \lfloor lT/100 \rfloor$ for $l = 15, 16, \dots, 85$ where $\lfloor x \rfloor$ denotes the largest integer smaller or equal to $x \in \mathbb{R}$. Let $\hat{\gamma}_{pre}(l)$ and $\hat{\gamma}_{post}(l)$ be the first element of $\hat{\Gamma}_{pre}(\lfloor lT/100 \rfloor)$ and $\hat{\Gamma}_{post}(\lfloor lT/100 \rfloor)$, and denote by $\hat{\omega}_{pre}^2(l)$ and $\hat{\omega}_{post}^2(l)$ the 1, 1 element of $\hat{\Omega}_{pre}(\lfloor lT/100 \rfloor)$ and $\hat{\Omega}_{post}(\lfloor lT/100 \rfloor)$, respectively. With the definitions in Box 1,⁴ the 5% level test $\hat{\varphi}^*$ of

⁴ One obtains the test φ^* of Proposition 3 and defined in the Appendix by setting $\hat{\gamma}_{pre}(l) = 100G(l/100)/l$, $\hat{\gamma}_{post}(l) = (G(1) - G(l/100))/(1 - l/100)$, $\hat{\omega}_{pre}^2(l) = 100/l$ and $\hat{\omega}_{post}^2(l) = 100/(100 - l)$.

$$\begin{aligned}
 \widehat{\text{supF}} &= \max_{16 \leq l \leq 85} \frac{(\hat{\gamma}_{\text{post}}(l) - \hat{\gamma}_{\text{pre}}(l-1))^2}{\hat{\omega}_{\text{post}}^2(l) + \hat{\omega}_{\text{pre}}^2(l-1)} \\
 \hat{\Delta}_{\text{pre}}(l) &= l\hat{\gamma}_{\text{pre}}(l) - (l-1)\hat{\gamma}_{\text{pre}}(l-1) \\
 \hat{\Delta}_{\text{post}}(l) &= (101-l)\hat{\gamma}_{\text{post}}(l-1) - (100-l)\hat{\gamma}_{\text{post}}(l) \\
 \hat{l} &= \arg \min_{16 \leq l \leq 85} \left(\sum_{j=16}^{l-1} \hat{\Delta}_{\text{pre}}(j)^2 - (l-1)\hat{\gamma}_{\text{pre}}(l-1)^2 + \sum_{j=l+1}^{85} \hat{\Delta}_{\text{post}}(j)^2 - (100-l)\hat{\gamma}_{\text{post}}(l)^2 \right) \\
 \hat{t}_{\text{post}} &= \frac{\hat{\gamma}_{\text{post}}(\min(\hat{l} + 1, 85)) - \gamma_{\text{post},0}}{\sqrt{\hat{\omega}_{\text{post}}^2(\min(\hat{l} + 1, 85))}} \tag{18} \\
 \hat{\omega}^2 &= \frac{(\hat{l}-1)^2}{9900} \hat{\omega}_{\text{pre}}^2(\hat{l}-1) + \frac{(100-\hat{l})^2}{9900} \hat{\omega}_{\text{post}}^2(\hat{l}) \\
 \widehat{\text{LR}} &= \frac{\sum_{l=15}^{85} \exp \left[\frac{1}{2} \frac{\sigma_{\text{pre}}^2 (\hat{\gamma}_{\text{pre}}(l) - \gamma_{\text{post},0})^2 l^2}{100^2 \hat{\omega}^2 v(l, \sigma_{\text{pre}}^2)} + \frac{1}{2} \frac{\sigma_{\text{post}}^2 (\hat{\gamma}_{\text{post}}(l) - \gamma_{\text{post},0})^2 (100-l)^2}{100^2 \hat{\omega}^2 v(100-l, \sigma_{\text{post}}^2)} \right]}{71 \sqrt{v(l, \sigma_{\text{pre}}^2) v(100-l, \sigma_{\text{post}}^2)}} \\
 &\quad \frac{\sum_{j=1}^{18} \sum_{l=a_j}^{b_j} p_j \exp \left[-\frac{1}{2} \frac{\mu_{\delta,j}^2 l}{100 v(l, \sigma_{\delta,j}^2)} + \frac{1}{2} \frac{\sigma_{\delta,j}^2 (\hat{\gamma}_{\text{pre}}(l) - \gamma_{\text{post},0})^2 l^2}{100^2 \hat{\omega}^2 v(l, \sigma_{\delta,j}^2)} \right] \cosh \left[\frac{(\hat{\gamma}_{\text{pre}}(l) - \gamma_{\text{post},0}) \mu_j l}{100 v(l, \sigma_{\delta,j}^2) \hat{\omega}} \right]}{(b_j - a_j + 1) \sqrt{v(l, \sigma_{\delta,j}^2)}}
 \end{aligned}$$

where $v(l, \sigma^2) = 1 + \sigma^2 l / 100$, $\sigma_{\text{pre}}^2 = 378$, $\sigma_{\text{post}}^2 = 22$, and a_j, b_j, p_j, μ_j and $\sigma_{\delta,j}^2$ are defined in Table 2.

Box I.

(17) rejects H_0 if and only if either (i) $\widehat{\text{supF}} > 90$ and $|\hat{t}_{\text{post}}| > 2.01$, or (ii) $\widehat{\text{supF}} \leq 90$ and $\widehat{\text{LR}} > 2.41$.⁵

The intuition for these computations is roughly as follows. The statistic $\widehat{\text{supF}}$ is the largest F-statistic of the null hypothesis that the value of γ in the first $(l-1)\%$ is equal to the value of γ in the last $(100-l)\%$, maximized over $16 \leq l \leq 85$. By leaving out the middle 1% of observations, it is ensured that for any true break fraction within the middle 70% of the sample, one of these F-statistics (often, the largest) only involves estimators $(\hat{\gamma}_{\text{pre}}(l), \hat{\omega}_{\text{pre}}^2(l))$ and $(\hat{\gamma}_{\text{post}}(l), \hat{\omega}_{\text{post}}^2(l))$ from stable periods. When $\widehat{\text{supF}} > 90$, that is, when there is very strong evidence for the occurrence of a break, the test $\hat{\varphi}^*$ rejects if the usual t -statistic \hat{t}_{post} is larger than 2.01 (rather than the usual 1.96) in absolute value, where \hat{t}_{post} uses \hat{l} to determine the appropriate post-break data.

In (approximately) linear and stationary models, $\hat{\gamma}_{\text{pre}}(l)$ is centered at the average parameter value of the first $l\%$ of the sample, $[\text{IT}/100]^{-1} \sum_{t=1}^{\lfloor \text{IT}/100 \rfloor} (\gamma_{\text{pre}} + (\gamma_{\text{post}} - \gamma_{\text{pre}}) \mathbf{1}[t > \tau])$, and similarly, $\hat{\gamma}_{\text{post}}(l)$ is centered at the average parameter value of the last $(100-l)\%$ of the sample. The statistics $\hat{\Delta}_{\text{pre}}(l)$ and $\hat{\Delta}_{\text{post}}(l)$ thus approximately estimate the value of the parameter in the l th percent of the sample. If the true break τ is in the l_0 th percent of the sample, then $\hat{\Delta}_{\text{pre}}(l)$ should be approximately equal to γ_{pre} for all $l < l_0$, and similarly for $\hat{\Delta}_{\text{post}}(l)$. The break fraction estimator \hat{l} is usefully thought of as a least squares estimator based on these discretized estimators leaving out the middle one percent; note that for a given potential break date $l \in \{15, \dots, 85\}$ measured in percentage points

$$\begin{aligned}
 &\min_{\gamma_{\text{pre}}, \gamma_{\text{post}}} \left[\sum_{j=1}^{l-1} (\hat{\Delta}_{\text{pre}}(j) - \gamma_{\text{pre}})^2 + \sum_{j=l+1}^{100} (\hat{\Delta}_{\text{post}}(j) - \gamma_{\text{post}})^2 \right] \\
 &\approx \sum_{j=1}^{l-1} \hat{\Delta}_{\text{pre}}(j)^2 - (l-1)\hat{\gamma}_{\text{pre}}(l-1)^2 \\
 &\quad + \sum_{j=l+1}^{100} \hat{\Delta}_{\text{post}}(j)^2 - (100-l)\hat{\gamma}_{\text{post}}(l)^2
 \end{aligned}$$

⁵ Replace (2.01, 2.41) by (2.36, 10.6) for a 1% level test.

and the omission of $\sum_{j=1}^{15} \hat{\Delta}_{\text{pre}}(j)^2$ and $\sum_{j=86}^{100} \hat{\Delta}_{\text{post}}(j)^2$ in the definition of \hat{l} is immaterial, as neither term depends on l . By adding one to \hat{l} , it is ensured that with high probability, $\hat{l} + 1$ is at least as large as the true break fraction in percentage points, so that \hat{t}_{post} is based on estimates from a stable post-break period. The advantage of the least squares criterion (18) based on $(\hat{\Delta}_{\text{pre}}(l), \hat{\Delta}_{\text{post}}(l))$ over, say, an analogue to the F-statistics that underlie $\widehat{\text{supF}}$, is that \hat{l} in (18) has appealing properties under relatively weak conditions also when the break magnitude is not small—cf. Proposition 5.

In the absence of strong evidence for a break, that is $\widehat{\text{supF}} \leq 90$, the test $\hat{\varphi}^*$ switches to deciding the null hypothesis based on a likelihood ratio statistic. The numerator of this statistic is the result of the weighting over alternative values for γ_{post} , break dates and break magnitudes with the weighting function F defined in Section 2.3.2. The denominator is the result of a weighting over break dates and break magnitudes that make detection of this alternative as difficult as possible, the numerically determined approximate least favorable distribution Λ^* discussed in Section 2.3.2, ensuring size control of $\hat{\varphi}^*$ even when information about the true break date is scarce. In this likelihood ratio statistic, the scale of $\hat{\gamma}_{\text{pre}}(l)$ and $\hat{\gamma}_{\text{post}}(l)$ is normalized by $\hat{\omega}$. If the break date τ satisfies $(\hat{l}-1)/100 \leq \tau/T \leq \hat{l}/100$, then the estimator $\hat{\omega}^2$ is based on variance estimators of stable models, which improves the small sample properties of $\widehat{\text{LR}}$.

4.2. Asymptotic properties

The following proposition establishes formally that under suitable conditions, the suggested test $\hat{\varphi}^*$ is consistent for non-local alternatives for both local and non-local break magnitudes; it has local power close to that of the infeasible test that uses only actual post break data when the break magnitude is large; it has local asymptotic power equal to the power of φ^* of Proposition 3 under a local break and local alternative; and it controls size uniformly. In particular, this implies that the weighted average power in Proposition 1, $\pi^* \simeq 49.0\%$, and the power in Fig. 4 of Section 2.3.2 is also the (weighted average) asymptotic local power of $\hat{\varphi}^*$ under local breaks.

Proposition 5. Define $\delta_T = T^{1/2}(\gamma_{pre} - \gamma_{post})/\omega$, $\beta_T = T^{1/2}(\gamma_{post} - \gamma_{post,0})/\omega$, and $\rho_T = \tau/T$ for some $\omega > 0$.

(i) Suppose $\delta_T = \delta \in \mathbb{R}$, $\rho_T = \rho \in [0.15, 0.85]$ and

$$\begin{aligned} & \{T^{1/2}(\hat{\gamma}_{pre}(l) - \gamma_{pre}), T^{1/2}(\hat{\gamma}_{post}(l) - \gamma_{post}), \\ & T\hat{\omega}_{pre}^2(l), T\hat{\omega}_{post}^2(l)\}_{l=15}^{85} \\ & \Rightarrow \left\{ \omega \frac{G_0(l/100)}{l/100}, \omega \frac{G_0(1) - G_0(l/100)}{1 - l/100}, \frac{\omega^2}{l/100}, \right. \\ & \left. \frac{\omega^2}{1 - l/100} \right\}_{l=15}^{85} \end{aligned} \tag{19}$$

where $G_0(s) = W(s) + \delta \min(\rho, s)$. (i.a) If $\beta_T = \beta \in \mathbb{R}$, then $\hat{\varphi}^*(X_T) \Rightarrow \varphi^*(G)$. (i.b) If $\beta_T \rightarrow \pm\infty$, then $\hat{\varphi}^*(X_T) \xrightarrow{p} 1$.

(ii) Suppose $\delta_T \rightarrow \pm\infty$, $\rho_T = \rho \in [0.15, 0.85]$, and, for some $\omega_{pre} > 0$ not necessarily equal to ω ,

$$\begin{aligned} & \{T^{1/2}(\hat{\gamma}_{pre}(l) - \gamma_{pre}), T\hat{\omega}_{pre}^2(l)\}_{l=15}^{[100\rho]} \\ & \Rightarrow \left\{ \omega_{pre} \frac{W(l/100)}{l/100}, \frac{\omega_{pre}^2}{l/100} \right\}_{l=15}^{[100\rho]} \\ & \{T^{1/2}(\hat{\gamma}_{post}(l) - \gamma_{post}), T\hat{\omega}_{post}^2(l)\}_{l=[100\rho]}^{85} \\ & \Rightarrow \left\{ \omega \frac{W(1) - W(l/100)}{1 - l/100}, \frac{\omega^2}{1 - l/100} \right\}_{l=[100\rho]}^{85} \end{aligned} \tag{20}$$

$$P(T^{1/2}|\hat{\gamma}_{pre}([100\rho] + 1) - \gamma_{pre}| > M) \rightarrow 1$$

$$P(T^{1/2}|\hat{\gamma}_{post}([100\rho] - 1) - \gamma_{post}| > M) \rightarrow 1$$

for any $M \in \mathbb{R}$, where $\lceil x \rceil = -\lfloor -x \rfloor$, and the weak convergences do not have to hold jointly. (ii.a) If $\beta_T = 0$, then $\limsup_{T \rightarrow \infty} E[\hat{\varphi}^*(X_T)] \leq 5\%$. (ii.b) If $\beta_T = \beta \in \mathbb{R}$ and $\rho \notin R_{grid} = \{0.15, 0.16, \dots, 0.84\}$, then $\hat{\varphi}^*(X_T) \Rightarrow \mathbf{1}[|Z + \beta\sqrt{0.98 - [100\rho]/100}| > 2.01]$, where $Z \sim \mathcal{N}(0, 1)$. (ii.c) If $\beta_T \rightarrow \pm\infty$, then $\hat{\varphi}^*(X_T) \xrightarrow{p} 1$.

(iii) Suppose $\beta_T = 0$; (19) holds for all sequences $(\rho_T, \delta_T) \rightarrow (\rho, \delta) \in [0.15, 0.85] \times \mathbb{R}$; (20) holds for all sequences $(\rho_T, \delta_T) \rightarrow (\rho, \delta) \in ([0.15, 0.85] \setminus R_{grid}) \times \{-\infty, +\infty\}$; and for all sequences $(\rho_T, \delta_T) \rightarrow (\rho, \delta) \in R_{grid} \times \{-\infty, +\infty\}$, there exists a subsequence $(\rho_{T'}, \delta_{T'}) \rightarrow (\rho, \delta)$ such that along this subsequence, (20) holds except for

$$\begin{aligned} & (T^{1/2}(\hat{\gamma}_{pre}(l_0) - \gamma_{pre}), T\hat{\omega}_{pre}^2(l_0)) \\ & \Rightarrow \left(\omega_{pre} \frac{W(l_0/100)}{l_0/100} + v_{pre}(l_0), \frac{\omega_{pre}^2}{l_0/100} \right) \\ & (T^{1/2}(\hat{\gamma}_{post}(l_0) - \gamma_{post}), T\hat{\omega}_{post}^2(l_0)) \\ & \Rightarrow \left(\omega \frac{W(1) - W(l_0/100)}{1 - l_0/100} + v_{post}(l_0), \frac{\omega^2}{1 - l_0/100} \right) \end{aligned} \tag{21}$$

where $l_0 = 100\rho$ and $v_{pre}(l_0), v_{post}(l_0) \in \mathbb{R} \cup \{-\infty, +\infty\}$ are non-random constants, of which at most one is nonzero. Then $\limsup_{T \rightarrow \infty} \sup_{\theta \in \Theta_0} E_\theta[\hat{\varphi}^*(X_T)] \leq \alpha^* \simeq 5\%$.

Part (i) of Proposition 5 considers the case where the pre and post break value of γ only differ of the order $T^{-1/2}$, i.e. when the break size is small in a statistical sense. In that neighborhood, the GMM estimators do not contain enough information to pin down the true break fraction exactly, even asymptotically. This asymptotic embedding mirrors the small sample problem with substantial uncertainty about the true break date. The proposition establishes that for local alternatives, where correspondingly, the true value γ_{post} differs by the order $T^{-1/2}$ from the hypothesized value $\gamma_{post,0}$, the asymptotic properties of $\hat{\varphi}^*$ are just like those of φ^* in the limiting problem discussed in Section 2.3.2. So in particular, by Proposition 3, the test $\hat{\varphi}^*$ has asymptotic level $\alpha^* \simeq 5\%$,

and it has asymptotic weighted average power of $\pi^* \simeq 49\%$. Furthermore, $\hat{\varphi}^*$ is consistent in this scenario against any alternative where γ_{post} is outside the $T^{-1/2}$ neighborhood of the hypothesized value.

These results rely on the high level condition (19). The sequence of statistics $\{\hat{\gamma}_{pre}(l), \hat{\gamma}_{post}(l), \hat{\omega}_{pre}(l), \hat{\omega}(l)\}$ are a special case of partial sample GMM estimators analyzed by Andrews (1993). In particular, his primitive Assumption 1 with an appropriate modification to account for parameter time variation of order $T^{-1/2}$ imply (19)—see the Appendix for details. Alternatively, the approach of Li and Müller (2009) could be amended to yield (19) under a different set of assumptions. Also, for the special case of maximum likelihood estimation, Condition 1 and the identification Condition 3 in the Appendix can be shown to imply (19) for both the average Hessian and outer product of scores covariance estimators $\hat{\Omega}_{pre}$ and $\hat{\Omega}_{post}$. The conclusions of part (i) thus hold for a wide range of models.

Part (ii) of Proposition 5 establishes the asymptotic properties of $\hat{\varphi}^*$ when the break is (very) large relative to the sample information about γ . This is the situation that the majority of the literature on breaks has considered, where breaks are large enough that their date can effectively be treated as known for purposes of asymptotic inference about β . The test again controls size and is consistent against any non-local alternative. Under local alternatives $\beta_T = \beta \neq 0$, it asymptotically corresponds to a two-sided t -test about the mean of a unit variance Gaussian variate with non-centrality parameter $\beta\sqrt{0.98 - [100\rho]/100}$ and critical value 2.01, at least as long $\rho \neq R_{grid}$.⁶ In comparison, the two-sided 5% level t -test based on post break data (assuming the break date was known) converges to $\varphi_\rho(G) = \mathbf{1}[|\frac{G(1) - G(\rho)}{\sqrt{1 - \rho}}| > 1.96]$, and thus has non-centrality parameter $\beta\sqrt{1 - \rho}$. With $\beta \sim \mathcal{N}(0, 22)$ as in the weighting function F of Section 2.3.2, the average asymptotic power loss of $\hat{\varphi}^*$ relative to φ_ρ over $\rho \in [0.15, 0.85]$ is approximately 1.7 percentage points, with a largest difference of 3.6 percentage points occurring at ρ near 0.83.

The results in part (ii) of Proposition 5 require two sets of assumptions. On the one hand, the statistics $\{\hat{\gamma}_{pre}(l), \hat{\omega}_{pre}^2(l)\}$ and $\{\hat{\gamma}_{post}(l), \hat{\omega}_{post}^2(l)\}$ have to behave in the usual way over the stable pre and post break periods, respectively, and be asymptotically independent. Note that the limiting variance may change at the parameter break date; this accommodates, say, changes in the variance of the AR(1) coefficient estimator that are induced by a non-local break in the AR(1) coefficient. One can again invoke the primitive conditions of Andrews (1993) to justify these convergences. On the other hand, the estimators $\hat{\gamma}_{pre}(l)$ and $\hat{\gamma}_{post}(l)$ must diverge from the $T^{-1/2}$ neighborhood of the pre and post parameter values γ_{pre} and γ_{post} for two values of l that involve a positive fraction of post and pre break data, respectively.⁷ For a non-local but shrinking break magnitude $|\gamma_{pre} - \gamma_{post}| \rightarrow 0$ and $T^{1/2}(\gamma_{pre} - \gamma_{post}) \rightarrow \pm\infty$, $\hat{\gamma}_{pre}(l)$ typically estimates the average parameter value in the first $l\%$ of the sample, that is

$$\frac{\hat{\gamma}_{pre}(l) - \gamma_{pre}}{\gamma_{post} - \gamma_{pre}} \xrightarrow{p} \mathbf{1}[l > [100\rho]] \frac{l - [100\rho]}{l} \neq 0, \tag{22}$$

and similarly for $\hat{\gamma}_{post}(l)$, which is clearly sufficient for the purposes of Proposition 5 (ii). The convergence (22) can be shown to hold, for instance, under the high-level Condition 1 of Li and Müller (2009)

⁶ This qualification is necessary because if $\rho \in R_{grid}$, \hat{l} potentially takes on the two values $[100\rho] + 1$ and $[100\rho] + 2$ with positive probability, even asymptotically, despite the diverging break magnitude $T^{1/2}(\gamma_{pre} - \gamma_{post})$.

⁷ Note, however, that nothing is assumed about the behavior of the variance estimators $\{\hat{\omega}_{pre}(l), \hat{\omega}_{post}(l)\}$ that involve subsets of unstable periods.

Table 3
Small sample size and power in location model.

ρ	Size ($\beta = 0$)				Power ($\beta = 4$)				Power ($\beta = -4$)			
	δ				δ				δ			
	1	4	8	16	1	4	8	16	1	4	8	16
$\hat{\varphi}_\rho$, t -test with break date known												
0.25	0.06	0.06	0.06	0.06	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
0.50	0.07	0.07	0.07	0.07	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
0.75	0.09	0.09	0.09	0.09	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
t -test with least-squares break date, standard critical value												
0.25	0.22	0.19	0.14	0.10	0.66	0.73	0.78	0.78	0.62	0.58	0.56	0.57
0.50	0.22	0.17	0.10	0.07	0.63	0.66	0.74	0.80	0.66	0.77	0.85	0.85
0.75	0.22	0.23	0.16	0.11	0.63	0.56	0.54	0.56	0.65	0.67	0.65	0.61
t -test with least-squares break date, asymptotic uniform critical value												
0.25	0.00	0.00	0.00	0.00	0.10	0.09	0.08	0.08	0.11	0.12	0.11	0.08
0.50	0.00	0.00	0.00	0.00	0.10	0.07	0.06	0.06	0.11	0.11	0.09	0.07
0.75	0.00	0.00	0.01	0.00	0.13	0.15	0.08	0.04	0.08	0.05	0.04	0.03
$\hat{\varphi}_{0.85}$, t -test with last 15% of data												
0.25	0.11	0.11	0.11	0.11	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
0.50	0.11	0.11	0.11	0.11	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
0.75	0.11	0.11	0.11	0.11	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45
$\hat{\varphi}^*$, test of this paper												
0.25	0.06	0.06	0.05	0.04	0.46	0.57	0.78	0.90	0.48	0.64	0.90	0.93
0.50	0.06	0.06	0.06	0.05	0.45	0.53	0.69	0.76	0.48	0.67	0.82	0.79
0.75	0.06	0.05	0.06	0.07	0.45	0.44	0.44	0.45	0.46	0.43	0.41	0.50

Notes: The variance of sample means is estimated with a Newey and West (1987) estimator with four lags. The tests are the same as those in Table 1, except that the oracle test $\hat{\varphi}_\rho$ is based on the actual post break data. Power is not size-corrected. Based on 10,000 replications.

by proceeding as in their Theorem 1 as long as no break in the nuisance parameter $\tilde{\gamma}$ is an order of magnitude larger than the break in the parameter of interest γ . For a fixed break magnitude $\gamma_{post} - \gamma_{pre}$, the distortionary effects of the break becomes even stronger, of course, and barring pathological cancellations, one would expect the condition in Proposition 5(ii) to be satisfied. We refrain from a detailed discussion of sufficient primitive conditions for the sake of brevity.

Finally, part (iii) of Proposition 5 shows that the test $\hat{\varphi}^*$ controls size uniformly in large samples—for T large enough, no choice of $\theta \in \Theta_0$ leads to a rejection probability substantially above 5%. This result is established under the assumption that for local breaks $\delta_T = O(1)$, the convergences in part (i) hold for all converging sequences $(\rho_T, \delta_T) \rightarrow (\rho, \delta)$. We show in the Appendix that the primitive assumptions of Andrews (1993) are again sufficient. For non-local breaks, care must be taken for the case $\rho_T \rightarrow \rho \in R_{grid}$. If $\rho_T \rightarrow \rho = l_0/100 \in R_{grid}$ with $\rho_T < \rho$, then $\hat{\gamma}_{pre}(l_0)$ is an estimator from an unstable model, with a shrinking fraction of the data stemming from the post break model. Depending on the rate at which $\delta_T \rightarrow \pm\infty$, this contamination shifts the center of the asymptotic distribution of $\hat{\gamma}_{pre}(l_0)$ by $v_{pre}(l_0)$ —typically, given the arguments above (22), one would expect $v_{pre}(l_0) \in \mathbb{R}$ if $\delta_T(\rho_T - \rho) \rightarrow \xi \in \mathbb{R}$, and it is sufficient for the last assumption in Proposition 5(iii) to assume that when $(\rho_T, \delta_T) \rightarrow (\rho, \delta) \in R_{grid} \times \{-\infty + \infty\}$, (20) holds except for (21) for all sequences $\delta_T(\rho_T - \rho) \rightarrow \xi \in \mathbb{R} \cup \{-\infty, +\infty\}$.⁸

⁸ Inspection of the proof of Proposition 5 shows that these assumptions are only needed to ensure that $(\{\hat{\gamma}_{pre}(l)\}_{l=1}^{l_0}, \hat{\Delta}_{post}(l_0 + 1))$ are asymptotically independent of $(\{\hat{\gamma}_{post}(l), \hat{\omega}_{post}^2(l)\}_{l=l_0+1}^{l_0+2})$, as this independence is enough to establish that the mixture of the two \hat{t}_{post} statistics with $\hat{l} = l_0 + 1$ and $\hat{l} = l_0 + 2$ controls size. The condition could dispensed with entirely by increasing the critical value of \hat{t}_{post} to 2.07, since $P(|Z_1| < 2.07 \text{ and } |Z_2| < 2.07) > 0.95$ for all bivariate normal Z_1, Z_2 with $Z_1 \sim Z_2 \sim \mathcal{N}(0, 1)$ and correlation of at least $\sqrt{15/16}$, so that no mixture of the two \hat{t}_{post} -statistics can induce overrejections.

4.3. Small sample properties

We now turn to the small sample properties of the test $\hat{\varphi}^*$ suggested here, and compare it to the infeasible test $\hat{\varphi}_\rho$ that corresponds to standard inference using actual post break data only. We consider three Monte Carlo designs: inference about the post break mean of an otherwise stationary scalar time series, inference about the post break value of the moving average parameter of a scalar time series, and inference on a linear regression estimator in a problem where there is endogeneity but instruments are available. The first and third examples are GMM problems, the second is a maximum likelihood problem.

For inference about the mean, we design the data generating process to mimic the productivity growth example of the introduction. The data does not exhibit significant autocorrelation, and the sample kurtosis is 3.24. Accordingly, we set $T = 143$, and draw the data $\{x_t\}_{t=1}^T$ from

$$x_t = \beta T^{-1/2} + \delta T^{-1/2} \mathbf{1}[t \leq \rho T] + u_t$$

where u_t is i.i.d. student- t with 29 degrees of freedom, and scaled to be of unit variance. Table 3 reports size and power of various 5% nominal tests of the null hypothesis that $\beta = 0$ (this choice is without loss of generality, as the testing problem is translation and scale invariant). The results correspond closely to the asymptotic properties of the corresponding tests discussed in Section 2.3, and similar comments apply. The test developed in this paper, $\hat{\varphi}^*$, has close to nominal rejection probability for all values of the break magnitude, in contrast to the (usual) t -test on post break data with a break date determined by minimizing the sum of squared residuals. The mild overrejections of $\hat{\varphi}^*$ are caused by the variance estimation, and are smaller than those of the oracle test $\hat{\varphi}_\rho$ with known break date.

In the MA(1) design, we generate data from the model

$$x_t = \mu + \varepsilon_t + \mathbf{1}[t \leq \rho T] \psi_{pre} \varepsilon_{t-1} + \mathbf{1}[t > \rho T] \psi_{post} \varepsilon_{t-1}$$

with $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$ and $T = 480$ (think of 40 years of monthly data). We test the hypothesis $H_0 : \psi_{post} = 0$, which corresponds to the null hypothesis of no current predictability of x_t . This

Table 4
Small sample size and power of suggested test in MA(1) model.

ρ	Size ($\psi_{post} = 0$)					Power ($\psi_{post} = 0.2$)				
	$\hat{\varphi}^*, \psi_{pre}$				$\hat{\varphi}_\rho$	$\hat{\varphi}^*, \psi_{pre}$				$\hat{\varphi}_\rho$
	0.1	0.3	0.5	0.7		-0.1	0.1	0.3	0.5	
0.25	5.7	5.6	5.0	5.2	5.2	60.3	85.5	96.9	96.9	97.5
0.50	6.0	7.1	8.4	8.3	5.9	61.3	85.4	88.6	89.1	89.7
0.75	5.7	8.9	12.0	13.3	7.0	53.6	51.0	56.7	66.6	64.4

Notes: All entries are based on asymptotic critical values. The variance of the unconditional maximum likelihood estimators $\hat{\psi}$ is estimated by $(1 - \hat{\psi}^2)/[rT]$ in a sample of size $[rT]$. Based on 10,000 replications.

Table 5
Small sample size and power of suggested test in an instrumental variable model.

δ	Size ($H_0 : \beta = 0$)			Power ($\beta = 3$)			Power ($\beta = -3$)		
	ρ			ρ			ρ		
	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
1	4.3	4.4	4.6	30.1	30.3	31.8	26.3	26.4	22.3
4	4.5	4.9	5.9	38.3	37.6	33.5	46.9	44.4	20.2
8	4.6	5.3	7.5	57.2	48.2	34.8	69.8	52.4	20.3
16	3.9	4.1	6.2	65.6	50.6	31.2	68.6	48.4	25.0
32	5.2	5.1	6.0	72.2	57.6	38.5	70.9	50.7	23.7
64	5.2	5.1	5.8	72.5	57.6	38.5	71.4	50.8	23.7

Notes: All entries are based on asymptotic critical values. The variance of the 2SLS estimator is estimated by the usual (heteroskedasticity robust) GMM formula. Based on the based on 10,000 replications.

design is outside the linear regression framework for which least-squares methods have been developed, so with unknown break date, $\hat{\varphi}^*$ seems to be the only known way of performing this inference. Table 4 reports the null and alternative rejection probabilities of $\hat{\varphi}^*$ and the oracle test $\hat{\varphi}_\rho$, where the parameters are estimated by maximum likelihood. Small sample size distortions are mostly larger for $\hat{\varphi}^*$ compared to $\hat{\varphi}_\rho$, but remain modest for most considered parameter values. We experimented with smaller sample sizes and found worse size control, sometimes substantially so. Good properties of $\hat{\varphi}^*$ rely on reasonable accuracy of the usual large sample approximations over all partial sample estimators. But with small sample sizes T , the estimators over, say, the first and last 15% are based on very few observations, which leads to well known problems with the MA(1) maximum likelihood estimator.

For the instrumental variables example we follow the design of Hall et al. (2012), although we allow the position and the size of the break to vary. The data is generated by

$$x_{1t} = \mathbf{1}[t \leq \rho T](x_{2t}\gamma_{pre} + \tilde{\gamma}_{pre}) + \mathbf{1}[t > \rho T](x_{2t}\gamma_{post} + \tilde{\gamma}_{post}) + u_t \tag{23}$$

where $x_{2t} = 1 + 0.5 \sum_{i=1}^4 x_{3t,i} + v_t$, $x_{3t} \sim$ i.i.d. $\mathcal{N}(0, I_4)$ and (u_t, v_t) is i.i.d. normal with mean zero, variances equal to one and covariance equal to one half. The 4×1 vector x_{3t} is observed and serves as an instrument. A calculation shows that the full sample GMM (=2SLS) estimator of $\Gamma = (\gamma, \tilde{\gamma})'$ from the stable model (23) has asymptotic variance I_2 . We consider tests of $H_0 : \gamma_{post} = 0$. The true values of the parameters are $\gamma_{pre} = \tilde{\gamma}_{pre} = T^{-1/2}(\beta + \delta)$ and $\gamma_{post} = \tilde{\gamma}_{post} = T^{-1/2}\beta$, so both parameters break by the common value δ . This model satisfies the assumptions of Proposition 5, as we detail in the Appendix. Table 5 reports small sample rejection probabilities for $T = 480$. Size is generally well controlled for both small and large breaks, although some moderate size distortions arise when the break is near the end of the sample.

5. Conclusion

Models with discrete breaks in the parameters at an unknown or uncertain date have become popular in empirical work. This paper shows that inference about pre and post break parameters

using the standard approach of using the estimated break date as the true break date leads to substantially oversized tests and confidence intervals as long as the break magnitude is not very large relative to the sampling uncertainty about the parameters. For the important special case of a single break at an unknown date and a single parameter of interest, we derive an alternative test with uniform asymptotic size control that demonstrably comes close to maximizing a weighted average power criterion.

The method applies to a very wide range of econometric models—those that can be written as a GMM problem. This includes breaks in means and variances, in coefficients in linear and nonlinear regressions, and maximum likelihood models with differentiable likelihood (where the first order condition becomes the GMM problem). The required regularity conditions are similar to those in Andrews (1993). And while the derivation of the test statistic is somewhat complicated, its application is fairly straightforward: it only involves the estimation of four standard GMM objects (parameter estimates and their covariance matrices) over 71 specific subsamples, irrespective of the sample size.

Appendix

Proof of Proposition 1. Let $\Theta^\ell = (\theta_1, \dots, \theta_\ell) \subset \Theta$ be an arbitrary set of $\ell < \infty$ values of θ . By Definition 9.1 in van der Vaart (1998), we need to show that $\{\ln LR_T(\theta)\}_{\theta \in \Theta^\ell} \Rightarrow \{\ln f_\theta(G)\}_{\theta \in \Theta^\ell}$, where X_T and G are generated under any $\theta_0 \in \Theta$.

We first show the convergence $\{\ln LR_T(\theta)\}_{\theta \in \Theta^\ell} \Rightarrow \{\ln f_\theta(W)\}_{\theta \in \Theta^\ell}$ where X_T is generated from the stable model. As in the proof of Lemma 1 in Li and Müller (2009), note that by an exact Taylor expansion, for large enough T to ensure that $\Gamma_{T,t} \in \mathcal{B}_0$ for $\theta \in \Theta^\ell$,

$$\ln LR_T(\theta) = T^{-1/2} \sum_{t=1}^T B(t/T; \theta)' s_{T,t}(\Gamma_0) + \frac{1}{2} T^{-1} \sum_{t=1}^T B(t/T; \theta)' h_{T,t}(\tilde{\Gamma}_{T,t}) B(t/T; \theta)$$

where $\tilde{\Gamma}_{T,t}$ is on the line segment between Γ_0 and $\Gamma_0 + T^{-1/2} B(t/T; \theta)$. To show the desired convergence, it therefore suffices to show that (i) $T^{-1} \sum_{t=1}^T B(t/T; \theta)' h_{T,t}(\tilde{\Gamma}_{T,t}) B(t/T; \theta) \xrightarrow{p} - \int_0^1 B(\lambda; \theta)' HB(\lambda; \theta) d\lambda$ for all $\theta \in \Theta^\ell$ and (ii) $\left\{ T^{-1/2} \sum_{t=1}^T B(t/T; \theta)' s_{T,t}(\Gamma_0) \right\}_{\theta \in \Theta^\ell} \Rightarrow \left\{ \int_0^1 B(\lambda; \theta)' H^{1/2} dW(\lambda) \right\}_{\theta \in \Theta^\ell}$. From Condition 1(iii),

$$T^{-1} \sum_{t=1}^T \|h_{T,t}(\tilde{\Gamma}_{T,t}) - h_{T,t}(\Gamma_0)\| \leq T^{-1} \sum_{t=1}^T \sup_{\Gamma \in \mathcal{B}_T} \|h_t(\Gamma) - h_t(\Gamma_0)\| \xrightarrow{p} 0.$$

Table 6
Base distributions for approximate least favorable distribution.

i	$a, b + 1$	μ_δ	σ_δ^2
1	15, 20, 75, 83, 85, 86	0, 2, ..., 20	300
2	15, 20, 75, 83, 85, 86	0, 2, ..., 16	50
3	15, 20, 75, 83, 85, 86	0, 2, ..., 22	10
4	15, 20, 75, 83, 85, 86	0, 2, ..., 22	3
5	15, 20, 30, 45, 60, 70, 75, 80, 83, 85, 86	0, 2, ..., 22	10
6	83, 84, 85, 86	0, $\frac{1}{2}$, ..., 18	1
7	15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 83, 85, 86	0, 1, ..., 22	10

Note: The algorithm described in Section 2.3.2 is initialized with $(a, b, \mu_\delta, \sigma_\delta^2) \in \{(15, 24, 28, 200), (75, 85, 28, 200), (15, 85, 20, 4), (85, 85, 3, 4), (15, 85, 0, 300)\}$ and goes through 7 rounds in total, indexed by i . Each round increases the set of baseline distributions by the indicated values for $(a, b, \mu_\delta, \sigma_\delta^2)$, with $(a, b + 1)$ equal to all consecutive pairs of values in the second column.

Thus, in combination with Condition 2, for any fixed $\theta \in \Theta^\ell$, the assumptions of Lemma 4 of Li and Müller (2009) are satisfied, implying the following convergence in probability

$$\begin{aligned} & T^{-1} \sum_{t=1}^T B(t/T; \theta)' h_{T,t}(\tilde{\Gamma}_{T,t}) B(t/T; \theta) \\ &= T^{-1} \text{tr} \sum_{t=1}^T h_{T,t}(\tilde{\Gamma}_{T,t}) B(t/T; \theta) B(t/T; \theta)' \\ &\xrightarrow{p} -\text{tr} \int_0^1 HB(\lambda; \theta)' B(\lambda; \theta) d\lambda = - \int_0^1 B(\lambda; \theta)' HB(\lambda; \theta) d\lambda. \end{aligned}$$

For part (ii), note that $\left\{ \int_0^1 B(\lambda; \theta)' H^{1/2} dW(\lambda) \right\}_{\theta \in \Theta^\ell}$ is a ℓ dimensional normal vector. Thus, by the Cramer–Wold device, it suffices to show that for all $v \in \mathbb{R}^\ell$ with $v'v = 1$, $T^{-1/2} \sum_{t=1}^T (\sum_{i=1}^\ell v_i B(t/T; \theta_i)' s_{T,t}(\Gamma_0)) \Rightarrow \int_0^1 B_v(\lambda)' H^{1/2} dW(\lambda)$, where $B_v(\cdot) = \sum_{i=1}^\ell v_i B(\cdot; \theta_i)$. Let $q_{T,t} = B_v(t/T)' s_t(\beta_0)$. Then $\{q_{T,t}, \tilde{\mathfrak{F}}_{T,t}\}$ is a martingale difference array, and

$$\begin{aligned} & T^{-1} \sum_{t=1}^T E[|q_{T,t}|^{2+v} | \tilde{\mathfrak{F}}_{T,t-1}] \\ &\leq \sup_{0 \leq \lambda \leq 1} \|B_v(\lambda)\|^{2+v} T^{-1} \sum_{t=1}^T E[\|s_{T,t}(\beta_0)\|^{2+v} | \tilde{\mathfrak{F}}_{T,t-1}] \end{aligned} \tag{24}$$

which is $O_p(1)$ by Condition 1(ii) and Condition 2. Also

$$\begin{aligned} & T^{-1} \sum_{t=1}^T E[q_{T,t}^2 | \tilde{\mathfrak{F}}_{T,t-1}] \\ &= T^{-1} \sum_{t=1}^T B_v(t/T)' E[s_{T,t}(\beta_0) s_{T,t}(\beta_0)' | \tilde{\mathfrak{F}}_{T,t-1}] B_v(t/T) \\ &= \text{tr} T^{-1} \sum_{t=1}^T E[s_{T,t}(\beta_0) s_{T,t}(\beta_0)' | \tilde{\mathfrak{F}}_{T,t-1}] B_v(t/T) B_v(t/T)' \\ &\xrightarrow{p} \text{tr} \int_0^1 HB_v(\lambda) B_v(\lambda)' d\lambda = \int_0^1 B_v(\lambda)' HB_v(\lambda) d\lambda \end{aligned} \tag{25}$$

where the convergence in probability follows from a column-wise application of Lemma 4 in Li and Müller (2009). But (24) and (25) are sufficient for the central limit theorem Corollary 3.1 in Hall and Heyde (1980), so that

$$\begin{aligned} & T^{-1/2} \sum_{t=1}^T q_{T,t} \\ &\Rightarrow \mathcal{N} \left(0, \int_0^1 B_v(\lambda)' HB_v(\lambda) d\lambda \right) \sim \int_0^1 B_v(\lambda)' H^{1/2} dW(\lambda) \end{aligned}$$

concluding the argument for $\{\ln LR_T(\theta)\}_{\theta \in \Theta^\ell} \Rightarrow \{\ln f_\theta(W)\}_{\theta \in \Theta^\ell}$ under the stable model.

For the convergence under $\theta \in \Theta$, by the very argument just given, we also have the joint convergence $\{\ln LR_T(\theta_0), \{\ln LR_T(\theta)\}_{\theta \in \Theta^\ell}\} \Rightarrow \{\ln f_{\theta_0}(W), \{\ln f_\theta(W)\}_{\theta \in \Theta^\ell}\}$ in the stable model. Furthermore, since $E[f_{\theta_0}(W)] = 1$, by LeCam’s First Lemma (cf. Lemma 6.4 in van der Vaart (1998)) the model with parameter θ_0 is contiguous to the stable model. Thus, by a general version of LeCam’s Third Lemma (cf. Lemma 3.1 in van der Vaart (2002)), $\{\ln LR_T(\theta)\}_{\theta \in \Theta^\ell}$ also converges weakly in the model with $\theta = \theta_0$, and the limiting distribution is given by $\{\ln f_\theta(G)\}_{\theta \in \Theta^\ell}$, with G generated under $\theta = \theta_0$.

Proof of Proposition 2. By Proposition 1 and Theorem 15.1 of van der Vaart (1998), there exists a test φ_0 in \mathcal{E}_0 such that $E_\theta[\varphi_0(G)] = \pi(\theta)$. Since φ_T is of asymptotic level α , $\pi(\theta) \leq \alpha$ for all $\theta \in \Theta_0$. The claims now follow from the assumptions about φ^* and $\bar{\pi}$.

Definition of φ^ :*
See Box II.

Proof of Proposition 4. (i) Note that if $(b, d, \rho) \in \mathbb{R} \times \mathbb{R} \times [0.15, 0.85]$ are distributed F , then $(bv, dv, \rho) \in \mathbb{R}^k \times \mathbb{R}^k \times [0.15, 0.85]$ are distributed F_v . Thus,

$$\begin{aligned} h(G) &= \int f_\theta(G) dF_v(\theta) \\ &= \int \exp[\delta'G(\rho) + \beta'G(1) - \frac{1}{2}(\beta'\beta + 2\rho\beta'\delta + \rho\delta'\delta)] dF_v(\theta) \\ &= \int \exp[dv'G(\rho) + bv'G(1) - \frac{1}{2}(b^2 + 2\rho bd + \rho d^2)] dF(b, d, \rho). \end{aligned}$$

Further, let Λ_v^* be the distribution of (bv, dv, ρ) when (b, d, ρ) are distributed Λ^* . Then similarly

$$\int f_\theta(G) d\Lambda_v^*(\theta) = \int \exp[dv'G(\rho) - \frac{1}{2}\rho d^2] d\Lambda^*(b, d, \rho).$$

Thus, the best test to discriminate between H_1 : “ G has density h ” against H_0 : “ G has density $\int f_\theta(G) d\Lambda_v^*(\theta)$ ” rejects for large values of

$$\frac{h(G)}{\int f_\theta(G) d\Lambda_v^*(\theta)} = \frac{\int \exp[dv'G(\rho) + bv'G(1) - \frac{1}{2}(b^2 + 2\rho bd + \rho d^2)] dF(b, d, \rho)}{\int \exp[dv'G(\rho) - \frac{1}{2}\rho d^2] d\Lambda^*(b, d, \rho)}. \tag{26}$$

Note that conditional on (b, d, ρ) , $v'G(s) = W_1(s) + bs + d \min(\rho, s)$, where $W_1 = v'W$ is a scalar standard Wiener process. Thus, the distribution of (26) is just as in the scalar case underlying Proposition 3, and the same power bound applies.

(ii) The claim about power follows from the same argument as in part (i). For the claim about size control, note that $v'G(s) = v'W(s) + v'\beta s + v'\delta \min(\rho, s)$, which under the null hypothesis of $v'\beta = 0$ is a simple reparameterization of the scalar problem.

$\varphi^*(G) = \mathbf{1}[\text{supF}_0 > 90]\mathbf{1}[\hat{t}_0 > 2.01] + \mathbf{1}[\text{supF}_0 \leq 90]\mathbf{1}[\text{LR}_0 > 2.41]$, where

$$\text{supF}_0 = \max_{16 \leq l \leq 85} \frac{100 \left((100 - l)G\left(\frac{l-1}{100}\right) - (l-1) \left(G(1) - G\left(\frac{l}{100}\right) \right) \right)^2}{(100 - l)(l - 1)}$$

$$S_0(l) = \frac{100 \left(100G\left(\frac{l}{100}\right) - lG(1) \right)^2}{l(100 - l)} + 100^2 \frac{\left(lG\left(\frac{l-1}{100}\right) - (l-1)G\left(\frac{l}{100}\right) \right)^2}{l(l - 1)}$$

$$\hat{l}_0 = \arg \max_{16 \leq l \leq 85} S_0(l)$$

$$\hat{t}_0 = \frac{G(1) - G(\min(\hat{l}_0 + 1, 85)/100)}{\sqrt{1 - \min(\hat{l}_0 + 1, 85)/100}}$$

$$\text{LR}_0 = \frac{\sum_{l=15}^{85} v(l, \sigma_{pre}^2)^{-1/2} v(100 - l, \sigma_{\beta}^2)^{-1/2} \exp \left[\frac{1}{2} \frac{\sigma_{pre}^2 G(l/100)^2}{v(l, \sigma_{pre}^2)} + \frac{1}{2} \frac{\sigma_{\beta}^2 (G(1) - G(l/100))^2}{v(100 - l, \sigma_{\beta}^2)} \right]}{\sum_{j=1}^{18} \sum_{l=a_j}^{b_j} \frac{p_j}{b_j - a_j + 1} v(l, \sigma_{\delta, j}^2)^{-1/2} \exp \left[-\frac{1}{2} \frac{\mu_{\delta, j}^2 l}{100 v(l, \sigma_{\delta, j}^2)} + \frac{1}{2} \frac{\sigma_{\delta, j}^2 G(l/100)^2}{v(l, \sigma_{\delta, j}^2)} \right] \cosh \left[\frac{\mu_j G(l/100)}{v(l, \sigma_{\delta, j}^2)} \right]}$$

and $v(l, \sigma^2) = 1 + \sigma^2 l/100$, $\sigma_{pre}^2 = 378$, $\sigma_{\beta}^2 = 22$, and $p_j, a_j, b_j, \sigma_{\delta, j}^2$ and μ_j are defined in Table 2.

Box II.

Proof of Proposition 5. (i) The CMT, and, in the case of \hat{l} , some tedious algebra, imply that $\hat{l} \Rightarrow \hat{l}_0$ and $\widehat{\text{supF}} \Rightarrow \text{supF}_0$ with \hat{l}_0 and supF_0 as defined the definition of φ^* above (note that replacing G by G_0 leaves \hat{l}_0 and supF_0 unchanged).

(i.a) Since $G(s) = G_0(s) + \beta s$, the result follows from the CMT.

(i.b) We have $\hat{t}_{post}/\beta_T \Rightarrow 1/\sqrt{1 - \min(\hat{l}_0 + 1, 85)/100}$ and a calculation yields $\beta_T^{-2} \ln \widehat{\text{LR}} \xrightarrow{P} c > 0$, so that $P(\hat{t}_{post} > 2.01) \rightarrow 1$ and $P(\widehat{\text{LR}} > 2.41) \rightarrow 1$.

(ii) Define $l_0 = \lceil 100\rho \rceil$. Note that $\delta_T^{-2} T(\hat{\gamma}_{post}(l_0) - \hat{\gamma}_{pre}(l_0 - 1))^2 \xrightarrow{P} \omega^2 > 0$, $T\hat{\omega}_{post}^2(l_0) \xrightarrow{P} \omega^2/(1 - l_0/100) > 0$, and $T\hat{\omega}_{pre}^2(l_0 - 1) \xrightarrow{P} 100\omega_{pre}^2/(l_0 - 1) > 0$ so that $P(\widehat{\text{supF}} > 90) \rightarrow 1$. Let

$$\hat{S}(l) = \sum_{j=1}^{l-1} (\hat{\Delta}_{pre}(j) - \hat{\gamma}_{pre}(l-1))^2 + \sum_{j=l+1}^{100} (\hat{\Delta}_{post}(j) - \hat{\gamma}_{post}(l))^2$$

where $\hat{\Delta}_{pre}(j) = \hat{\gamma}_{pre}(15)$ for $j \leq 15$ and $\hat{\Delta}_{post}(j) = \hat{\gamma}_{post}(85)$ for $j > 85$. Note that $\hat{l} = \arg \min_{16 \leq l \leq 85} \hat{S}(l)$, because $\sum_{j=1}^l \hat{\Delta}_{pre}(j) = l\hat{\gamma}_{pre}(l)$ and $\sum_{j=l+1}^{100} \hat{\Delta}_{post}(j) = (100 - l)\hat{\gamma}_{post}(l)$. If $\rho \notin R_{grid}$, then $\hat{S}(l_0) = O_p(T^{-1})$, and for all $l \neq l_0$, $P(T\hat{S}(l) > M) \rightarrow 1$ for any $M \in \mathbb{R}$, so that $\hat{l} \xrightarrow{P} l_0$. Part (ii.b), and part (ii.a) for $\rho \notin R_{grid}$, now follow from the CMT.

Consider thus the proof of part (ii.a) if $\rho \in R_{grid}$. By the same argument, $P(\hat{l} \notin \{l_0, l_0 + 1\}) \rightarrow 0$. Define $t_{post}(l_0 + 1) = (W(1) - W(\frac{l_0+1}{100}))/\sqrt{1 - \frac{l_0+1}{100}} \sim \mathcal{N}(0, 1)$ and $t_{post}(l_0 + 2) = (W(1) - W(\frac{l_0+2}{100}))/\sqrt{1 - \frac{l_0+2}{100}} \sim \mathcal{N}(0, 1)$, and consider

$$\begin{aligned} \hat{S}(l_0) - \hat{S}(l_0 + 1) &= A_{pre} + \hat{\Delta}_{post}(l_0 + 1)^2 - (100 - l_0)\hat{\gamma}_{post}(l_0)^2 \\ &\quad + (99 - l_0)\hat{\gamma}_{post}(l_0 + 1)^2 \\ &\Rightarrow \Delta S(l_0) = a \cdot t_{post}(l_0 + 1)^2 \\ &\quad + t_{post}(l_0 + 1)B(l_0) + C(l_0) \end{aligned}$$

where A_{pre} is a function of $\{\hat{\gamma}_{pre}(l)\}_{l=15}^{l_0}$, the convergence follows from the CMT, a is a positive constant, and $B(l_0)$ and $C(l_0)$ are random variables that are independent of $t_{post}(l_0 + 1)$ and $t_{post}(l_0 + 2)$. Thus, by a further application of the CMT, we obtain

$$\begin{aligned} \hat{t}_{post} &\Rightarrow \mathbf{1}[\Delta S(l_0) \geq 0]\mathbf{1}[|t_{post}(l_0 + 2)| > 2.01] \\ &\quad + \mathbf{1}[\Delta S(l_0) < 0]\mathbf{1}[|t_{post}(l_0 + 1)| > 2.01]. \end{aligned} \tag{27}$$

Consider now the rejection probability of the r.h.s. of (27) conditional on $B(l_0) = b$ and $C(l_0) = c$. Noting that $\Delta S(l_0)$ is a quadratic function of $t_{post}(l_0 + 1)$ with positive coefficient on the square term, one obtains that the rejecting probability is bounded from above by the replacement of $\mathbf{1}[\Delta S(l_0) \geq 0]$ by $\mathbf{1}[t_{post}(l_0 + 1) < -2.01]$, for any value of b, c . A numerical calculation now shows that the conditional rejection probability remains below 5% even in that case for all $l_0 = 15, \dots, 83$, and the result follows.

(ii.b) Immediate from $P(\hat{l} \notin \{l_0, l_0 + 1\}) \rightarrow 0$ and $\beta_T^{-1}(\hat{\gamma}_{post}(l) - \gamma_{post,0})/\hat{\omega}_{post}(l) \xrightarrow{P} \sqrt{1 - l/100}/\omega > 0$ for all $l \geq l_0$.

(iii) Let θ_T be a parameter sequence such that $\lim \sup_{T \rightarrow \infty} E_{\theta_T}[\hat{\varphi}^*(X_T)] = \lim \sup_{T \rightarrow \infty} \sup_{\theta \in \Theta_0} E_{\theta}[\hat{\varphi}^*(X_T)]$. Pick a subsequence T' of T such that $\lim_{T' \rightarrow \infty} E_{\theta_{T'}}[\hat{\varphi}^*(X_{T'})] = \lim \sup_{T \rightarrow \infty} E_{\theta_T}[\hat{\varphi}^*(X_T)]$. Since $\bar{\Theta}_0 = \Theta_0 \cup \{(0, \rho, \delta) : \rho \in [0.15, 0.85], \delta \in \{+\infty, -\infty\}\}$ is compact under an appropriate metric, there exists a further subsequence $\theta_{T''}$ of $\theta_{T'}$ such that $\theta_{T''} \rightarrow \tilde{\theta} = (0, \tilde{\rho}, \tilde{\delta}) \in \bar{\Theta}_0$. If $\tilde{\rho} \notin R_{grid}$, then the result follows as in parts (i) and (ii). If $\tilde{\rho} \in R_{grid}$, then by assumption, there is a further subsequence such that (20) holds, except for (21). Proceeding as in the proof of part (ii.a) now again yields the result (and for the case $v_{pre}(l_0) \in \{+\infty, -\infty\}$ or $v_{post}(l_0) \in \{+\infty, -\infty\}$, $P(|\hat{S}(l_0) - \hat{S}(l_0 + 1)| < M) \rightarrow 0$ for all M , so that $\mathbf{1}[\Delta S(l_0) \geq 0]$ in (27) is replaced by either zero or one).

Condition 3. In the stable model with parameter Γ_0 , for all $s \in [0.15, 0.85]$ and $\epsilon > 0$ there exists $K(s, \epsilon) > 0$ such that $P(\sup_{\|r - r_0\| \geq \epsilon} T^{-1} \sum_{t=1}^{\lfloor sT \rfloor} (l_{T,t}(\Gamma) - l_{T,t}(\Gamma_0)) < -K(s, \epsilon)) \rightarrow 1$ and $P(\sup_{\|r - r_0\| \geq \epsilon} T^{-1} \sum_{t=\lfloor sT \rfloor + 1}^T (l_{T,t}(\Gamma) - l_{T,t}(\Gamma_0)) < -K(s, \epsilon)) \rightarrow 1$.

Justification of (19) via Andrews (1993):

The true parameters are given by $\Gamma_{T,t} = \Gamma_{post} + T^{-1/2} \mathbf{1}[\tau \leq 100\rho_T] \Delta_T$, where $\Delta_T = -(\omega \delta_T, \omega \tilde{\delta}_T)' \rightarrow \Delta_0 \in \mathbb{R}^k$ and $\rho_T \rightarrow \rho_0 \in [0.15, 0.85]$, so that $E[g(X_{T,t}, \Gamma_{T,t})] = 0$. Write $g_t(\Gamma)$ for $g(X_{T,t}, \Gamma)$, and define $\gamma_t(\Gamma) = \partial g_t(\Gamma)/\partial \Gamma'$. Our set-up corresponds to what Andrews (1993) refers to as 'pure' structural change. We now show that under Assumption 1 of Andrews (1993), with the assumption in part (b) $E[g_t(\Gamma_0)] = 0$ replaced by $E[g_t(\Gamma_{T,t})] = 0$, (19) holds. These computations closely correspond to what is derived in Section 5.4 of Andrews (1993), with two differences: (i) we do not assume that $\max_{t \leq T} \sup_{\|r - r_{post}\| \leq T^{-1/2}K} \|E[\gamma_t(\Gamma)] - \tilde{\gamma}\| \rightarrow 0$ for some $K > 0$ (cf. first line on page 832), where in his Assumption 1(f), $\tilde{\gamma}$ is defined as the unique limit $T^{-1} \sum_{t=1}^{\lfloor T \rfloor} E \gamma_t$

$(\Gamma_{post}) \xrightarrow{p} \cdot \tilde{\gamma}$; (ii) we consider sequences for Δ_T and ρ_T , so that $(\Delta_T, \rho_T) \rightarrow (\Delta_0, \rho_0)$. (We note that there is a typo in Assumption 1-LP of Andrews (1993); it should read $\sup_{\pi \in \Pi} \|\sqrt{T}E\tilde{m}_T(\theta_0, \pi) - \mu(\pi)\| \rightarrow 0$ in his notation.)

Define $\eta_T : [0, 1] \mapsto \mathbb{R}^k$ as $\eta_T(x) = \mathbf{1}[x \leq \rho_T]\Delta_T$ for $T = 0, 1, \dots$. By the argument in Andrews' proof of his Theorem 4(a), it suffices to show that $\sup_{0 \leq s \leq 1} \|T^{-1/2}E[\sum_{t=1}^{\lfloor sT \rfloor} g_t(\Gamma_{post}) - \tilde{\gamma} \int_0^s \eta_0(l)dl]\| \rightarrow 0$. Now clearly $\sup_{0 \leq s \leq 1} \|\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^{\lfloor sT \rfloor} \eta_T(t/T) - \int_0^s \eta_0(l)dl\| \rightarrow 0$. Also,

$$\begin{aligned} & T^{-1/2}E[\sum_{t=1}^{\lfloor sT \rfloor} g_t(\Gamma_{post}) - T^{-1/2}\tilde{\gamma}\eta_T(t/T)] \\ &= T^{-1/2}E[\sum_{t=1}^{\lfloor \min(s, \rho_T)T \rfloor} g_t(\Gamma_{post}) - T^{-1/2}\tilde{\gamma}\eta_T(t/T)] \\ &= (E[\tilde{\gamma}_T(\min(s, \rho_T))]) - \tilde{\gamma})\Delta_T \end{aligned}$$

where $\tilde{\gamma}_T(r)$ is equal to $T^{-1} \sum_{t=1}^{\lfloor rT \rfloor} \int_0^1 \gamma_t(\Gamma_{post} + \lambda T^{-1/2}\Delta_T)d\lambda$. Furthermore,

$$\begin{aligned} & \sup_{0 \leq s \leq 1} \|E[\tilde{\gamma}_T(\min(s, \rho_T)) - \tilde{\gamma}_T(\min(s, \rho_0))]\| \\ & \leq T^{-1}E[\sum_{t=\lfloor \min(\rho_0, \rho_T)T \rfloor}^{\lfloor \max(\rho_0, \rho_T)T \rfloor} \|\int_0^1 \gamma_t(\Gamma_{post} + \lambda T^{-1/2}\Delta_T)d\lambda\|] \\ & \leq (\lfloor \max(\rho_0, \rho_T)T \rfloor - \lfloor \min(\rho_0, \rho_T)T \rfloor) \sup_{t \leq T} E[\sup_{\Gamma \in \mathcal{G}_0} \|\gamma_t(\Gamma)\|] \\ & \rightarrow 0 \end{aligned}$$

where the convergence follows, since by Andrews' Assumption 1(f), $\sup_{t \leq T} E[\sup_{\Gamma \in \mathcal{G}_0} \|\gamma_t(\Gamma)\|] = O(1)$ for some neighborhood \mathcal{G}_0 of Γ_{post} . Finally, $\sup_{0 \leq s \leq 1} \|E[\tilde{\gamma}_T(\min(s, \rho_0)) - \tilde{\gamma}]\| \rightarrow 0$ follows from the same reasoning as below Andrews' equation (A.11).

Instrumental variables example of Section 4.3

We show that the instrumental variables example satisfies the assumptions of Proposition 5, establishing that the test controls size asymptotically and is consistent for this example. First, define $\tilde{x}_{2t} = (x_{2t}, 1)'$ and $\tilde{x}_{3t} = (x'_{3t}, 1)'$, $\Sigma_{33} = E[\tilde{x}_{3t}\tilde{x}'_{3t}]$ and $\Sigma_{23} = E[\tilde{x}_{2t}\tilde{x}'_{3t}]$. Note that for each $l \in \{15, 16, \dots, 100\}$ ($\hat{\Sigma}_{23}(l)$, $\hat{\Sigma}_{33}(l)$) $= (T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{2t}\tilde{x}'_{3t}, T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{3t}) \xrightarrow{p} (\frac{1}{100}\Sigma_{23}, \frac{1}{100}\Sigma_{33})$ by a standard LLN for i.i.d. data. Similarly, $(T^{-1} \sum_{t=\lfloor lT/100 \rfloor+1}^T \tilde{x}_{2t}\tilde{x}'_{3t}, T^{-1} \sum_{t=\lfloor lT/100 \rfloor+1}^T \tilde{x}_{3t}\tilde{x}'_{3t}) = (\hat{\Sigma}_{23}(100) - \hat{\Sigma}_{23}(l), \hat{\Sigma}_{33}(100) - \hat{\Sigma}_{33}(l)) \xrightarrow{p} ((1 - \frac{l}{100})\Sigma_{23}, (1 - \frac{l}{100})\Sigma_{33})$ for $l \in \{15, 16, \dots, 85\}$. With the 2SLS weighting matrices $\hat{V}_{pre}(\lfloor lT/100 \rfloor) = \hat{\Sigma}_{33}(l/100)^{-1}$, and $\hat{K}(l) = \hat{\Sigma}_{23}(l) \hat{\Sigma}_{33}(l)^{-1} \hat{\Sigma}_{23}(l)'$, we obtain for the GMM estimator

$$\hat{\Gamma}_{pre}(\lfloor lT/100 \rfloor) = \hat{K}(l)^{-1} \hat{\Sigma}_{23}(l) \hat{\Sigma}_{33}(l)^{-1} T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}x_{1t} \quad (28)$$

with associated GMM covariance matrix estimator

$$T\hat{\Omega}_{pre}(\lfloor lT/100 \rfloor) = \hat{K}(l)^{-1} \hat{\Sigma}_{23}(l) \hat{\Sigma}_{33}(l)^{-1} \hat{P}_{pre}(l) \hat{\Sigma}_{33}(l)^{-1} \hat{\Sigma}_{23}(l)' \hat{K}(l)^{-1} \quad (29)$$

where we define $\hat{P}_{pre}(l) = T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{3t}\hat{u}_{t,pre}^2(l)$ and $\hat{u}_{t,pre}(l) = x_{1t} - \tilde{x}'_{2t}\hat{\Gamma}_{pre}(\lfloor lT/100 \rfloor)$. The form of the post estimators follows similarly, with $\hat{P}_{post}(l) = T^{-1} \sum_{t=\lfloor lT/100 \rfloor+1}^T \tilde{x}_{3t}\tilde{x}'_{3t}\hat{u}_{t,post}^2(l)$ and $\hat{u}_{t,post}(l) = x_{1t} - \tilde{x}'_{2t}\hat{\Gamma}_{post}(\lfloor lT/100 \rfloor)$ in the place of $\hat{P}_{pre}(l)$. It follows from the limit results above that $\hat{\Sigma}(l) = \hat{K}(l)^{-1} \hat{\Sigma}_{23}(l) \hat{\Sigma}_{33}(l)^{-1} \xrightarrow{p} \frac{1}{l/100} (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{23}')^{-1} \Sigma_{23}\Sigma_{33}^{-1} = \frac{1}{l/100}\mathcal{E}$.

First, consider (19), where $T^{1/2}(\Gamma_{pre}(l) - \Gamma_{post}) = \Delta = \delta e_2$ with $e_2 = (1, 1)'$ and $\delta \in \mathbb{R}$. From (28), we find

$$\begin{aligned} T^{1/2}(\hat{\Gamma}_{pre}(l) - \Gamma_{post}) &= \hat{\Sigma}(l) \left(T^{-1/2} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}u_t \right) \\ &+ \hat{\Sigma}(l) \hat{\Sigma}'_{23} T^{-1} (\min(\lfloor lT/100 \rfloor, \lfloor \rho T \rfloor)) \Delta \\ &\Rightarrow \frac{1}{l/100} \mathcal{E} \Sigma_{33}^{1/2} W_2 \left(\frac{l}{100} \right) \\ &+ \frac{1}{l/100} \min \left(\frac{l}{100}, \rho \right) \Delta \quad (30) \end{aligned}$$

from the continuous mapping theorem jointly in $l \in \{15, \dots, 85\}$, since $\mathcal{E} \Sigma_{23}' = I_2$, and $T^{-1/2} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}u_t \Rightarrow \Sigma_{33}^{1/2} W_2(\cdot)$ with $W_2(\cdot)$ a standard 2×1 vector Brownian motion. Noting that the 1–1 element of $\mathcal{E} \Sigma_{33} \mathcal{E}'$ is equal to one, we find that the first element on the r.h.s. of (30) is distributed $\frac{1}{l/100} W(\frac{l}{100}) + \frac{1}{l/100} \min(\frac{l}{100}, \rho) \delta$, which establishes $T^{1/2}(\hat{\gamma}_{pre}(l) - \gamma_{post}) \Rightarrow \omega \frac{G_0(l/100)}{l/100}$ of (19) with $\omega = 1$. The convergence for $T^{1/2}(\hat{\gamma}_{post}(l) - \gamma_{post})$ follows entirely analogously.

We now show that $\hat{P}_{pre}(l) \xrightarrow{p} \frac{1}{100} \Sigma_{33}$, so that $T\hat{\Omega}_{pre}(\lfloor lT/100 \rfloor) \xrightarrow{p} \frac{1}{l/100} (\Sigma_{23}\Sigma_{33}^{-1}\Sigma_{23}')^{-1} = \frac{1}{l/100} I_2$, implying the required convergence for $T\hat{\omega}_{pre}^2(l)$. Let $\tau_T = \min(\lfloor lT/100 \rfloor, \lfloor \rho T \rfloor)$. We have

$$\begin{aligned} \hat{P}_{pre}(l) &= T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{3t}\hat{u}_{t,pre}^2(l) \\ &= T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{3t}(u_t - (\hat{\Gamma}_{pre}(l) - \Gamma_{post})' \tilde{x}_{2t})^2 \\ &+ T^{-1/2} \Delta' \tilde{x}_{2t} \mathbf{1}[t \leq \lfloor \rho T \rfloor]^2 \\ &= T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{3t}u_t^2 + T^{-2} \sum_{t=1}^{\tau_T} \tilde{x}_{3t}\tilde{x}'_{2t} \Delta' \Delta \tilde{x}_{2t}\tilde{x}'_{3t} \\ &+ T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{2t} (\hat{\Gamma}_{pre}(l) - \Gamma_{post}) (\hat{\Gamma}_{pre}(l) - \Gamma_{post})' \tilde{x}_{2t}\tilde{x}'_{3t} \\ &+ 2T^{-3/2} \sum_{t=1}^{\tau_T} \tilde{x}_{3t} \Delta' \tilde{x}'_{2t} (\hat{\Gamma}_{pre}(l) - \Gamma_{post})' \tilde{x}_{2t}\tilde{x}'_{3t} \\ &+ 2T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{2t} (\hat{\Gamma}_{pre}(l) - \Gamma_{post}) u_t \tilde{x}'_{3t} \\ &+ 2T^{-3/2} \sum_{t=1}^{\tau_T} \tilde{x}_{3t}\tilde{x}'_{2t} \Delta u_t \tilde{x}'_{3t}. \end{aligned}$$

It follows directly from a standard LLN that $T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{3t}u_t^2 \xrightarrow{p} E[\tilde{x}_{3t}\tilde{x}'_{3t}u_t^2] = \frac{1}{100} \Sigma_{33}$, so that it suffices to show that the remaining terms converge to zero. For the second term after the third equals sign in the above expression, we have that $\text{vec}(T^{-2} \sum_{t=1}^{\tau_T} \tilde{x}_{3t}\tilde{x}'_{2t} \Delta \Delta' \tilde{x}_{2t}\tilde{x}'_{3t}) = T^{-2} \sum_{t=1}^{\tau_T} [\tilde{x}_{3t}\tilde{x}'_{2t} \otimes \tilde{x}_{3t}\tilde{x}'_{2t}] \text{vec}(\Delta \Delta')$ which converges to zero as $E[\tilde{x}_{3t}\tilde{x}'_{2t} \otimes \tilde{x}_{3t}\tilde{x}'_{2t}]$ is finite and $\text{vec}(\Delta \Delta')$ are constants. Similarly for the third term we have $\text{vec}(T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{2t} (\hat{\Gamma}_{pre}(l) - \Gamma_{post}) (\hat{\Gamma}_{pre}(l) - \Gamma_{post})' \tilde{x}_{2t}\tilde{x}'_{3t}) = (T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} [\tilde{x}_{3t}\tilde{x}'_{2t} \otimes \tilde{x}_{3t}\tilde{x}'_{2t}]) \text{vec}((\hat{\Gamma}_{pre}(l) - \Gamma_{post}) (\hat{\Gamma}_{pre}(l) - \Gamma_{post})')$ which converges to zero as $\hat{\Gamma}_{pre}(l) \xrightarrow{p} \Gamma_{post}$ from above. A similar calculation gives the same result for the fourth term. For the second last term, we have $\text{vec}(T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} \tilde{x}_{3t}\tilde{x}'_{2t} (\hat{\Gamma}_{pre}(l) - \Gamma_{post}) u_t \tilde{x}'_{3t}) = (T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} [\tilde{x}_{3t}u_t \otimes \tilde{x}_{3t}\tilde{x}'_{2t}]) \text{vec}(\hat{\Gamma}_{pre}(l) - \Gamma_{post})$. Again since $E[\tilde{x}_{3t}u_t \otimes \tilde{x}_{3t}\tilde{x}'_{2t}]$ is finite, $T^{-1} \sum_{t=1}^{\lfloor lT/100 \rfloor} [\tilde{x}_{3t}u_t \otimes \tilde{x}_{3t}\tilde{x}'_{2t}]$ can be at most $O_p(1)$, so this term converges to zero via $\hat{\Gamma}_{pre}(l) \xrightarrow{p} \Gamma_{post}$. For the last term we

have $\text{vec}(T^{-1} \sum_{t=1}^{T^*} \tilde{x}_{3t} \tilde{x}'_{2t} \Delta u_t \tilde{x}'_{3t}) = T^{-1} \sum_{t=1}^{T^*} [\tilde{x}_{3t} u_t \otimes \tilde{x}_{3t} \tilde{x}'_{2t}] \Delta$ where again this is at most $O_p(1)$, hence the division by $T^{-1/2}$ ensures this term is $o_p(1)$.

For the assumption of Proposition 5 part (ii), the first two sets of conditions in (20) follow from the same set of arguments just employed. For the last condition, since $\Delta = \Delta_T = T^{-1/2} e_2 \delta_T$ for some $\delta_T \rightarrow \infty$, $\delta_T^{-1} T^{1/2} \Delta_T \rightarrow e_2$. Let $I^* = \lfloor 100\rho \rfloor + 1$. We have

$$T^{1/2} (\hat{\Gamma}_{pre}(I^*) - \Gamma_{pre}) = \hat{\Sigma}(I^*) T^{-1/2} \sum_{t=1}^{\lfloor IT/100 \rfloor} \tilde{x}_{3t} u_t + \hat{\Sigma}(I^*) T^{-1} \times \sum_{t=\lfloor (I^*-1)T/100 \rfloor + 1}^{\lfloor I^*T/100 \rfloor} \tilde{x}_{3t} \tilde{x}'_{2t} e_2 \delta_T.$$

From results above we have that the first term is $O_p(1)$, and $\hat{\Sigma}(I^*) T^{-1} \sum_{t=\lfloor (I^*-1)T/100 \rfloor + 1}^{\lfloor I^*T/100 \rfloor} \tilde{x}_{3t} \tilde{x}'_{2t} \xrightarrow{P} \frac{1}{I^*} I_2$, so that $\delta_T^{-1} T^{1/2} (\hat{\Gamma}_{pre}(I^*) - \Gamma_{pre}) = \frac{1}{I^*} e_2 + o_p(1)$. Thus $P(T^{1/2} |\hat{\gamma}_{pre} - \gamma_{pre}| > M) = P(|\delta_T^{-1} T^{1/2} (\hat{\gamma}_{pre}(I^*) - \gamma_{pre})| > M \delta_T^{-1}) \rightarrow 1$. A similar argument applies to the post break estimator.

References

- Anderson, R.J., Klieson, K.L., 2006. The 1990's Acceleration in Labor Productivity: Causes and Measurement. Federal Reserve Bank of St. Louis Review 88, 181–202.
- Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 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., Guggenberger, P., 2009. Hybrid and size-corrected subsampling methods. *Econometrica* 77, 721–762.
- Andrews, D.W.K., Guggenberger, G., 2010a. Asymptotic size and a problem with subsampling and with the M out of N bootstrap. *Econometric Theory* 26, 426–468.
- Andrews, D.W.K., Guggenberger, P., 2010b. Applications of subsampling, hybrid, and size-correction methods. *J. Econometrics* 158, 285–305.
- Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.
- Bai, J., 1994. Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.* 15, 453–470.
- Bai, J., 1997. Estimation of a change point in multiple regressions. *Rev. Econ. Stat.* 79, 551–563.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.
- Choi, A., Hall, W.J., Schick, A., 1996. Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Stat.* 24, 841–861.
- Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Rev. Econ. Dyn.* 8, 262–302.
- Elliott, G., Müller, U.K., 2006. Efficient tests for general persistent time variation in regression coefficients. *Rev. Econom. Stud.* 73, 907–940.
- Elliott, G., Müller, U.K., 2007. Confidence sets for the date of a single break in linear time series regressions. *J. Econometrics* 141, 1196–1218.
- Elliott, G., Müller, U.K., Watson, M.W., 2012. Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis, Working Paper, Princeton University.
- Gordon, R.J., 2003. Exploding productivity growth: context, causes, and implications. *Brookings Pap. Econ. Activity* 2, 207–279.
- Hall, A., Han, S., Boldea, O., 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. *J. Econometrics* 170, 281–302.
- Hall, P., Heyde, C.C., 1980. *Martingale Limit Theory and its Applications*. Academic Press, New York.
- Jorgenson, D., Ho, M.S., Stiroh, K.J., 2008. A retrospective look at the US productivity growth resurgence. *J. Econ. Perspect.* 22, 3–24.
- Kailath, T., 1971. The structure of Radon–Nikodym derivatives with respect to Wiener and related measures. *Ann. Math. Statist.* 42, 1054–1067.
- King, M.L., 1988. Towards a theory of point optimal testing. *Econom. Rev.* 6, 169–218.
- Li, H., Müller, U.K., 2009. Valid inference in partially unstable general method of moment models. *Rev. Econom. Stud.* 76, 343–365.
- Müller, U.K., 2011. Efficient tests under a weak convergence assumption. *Econometrica* 79, 395–435.
- Müller, U.K., Petalas, P., 2010. Efficient estimation of the parameter path in unstable time series models. *Rev. Econom. Stud.* 77, 1508–1539.
- Newey, W.K., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Orphanides, A., 2004. Monetary policy rules, macroeconomic stability, and inflation: a view from the trenches. *J. Money, Credit, Bank.* 36, 151–175.
- Perron, P., Yamamoto, Y., 2012. A Note on Estimating and Testing for Multiple Structural Changes in Models with Endogenous Regressors via 2SLS, Working paper, Boston University.
- van der Vaart, A.W., 1991. An asymptotic representation theorem. *Internat. Statist. Rev.* 259, 97–121.
- van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.
- van der Vaart, A.W., 2002. The statistical work of Lucien Le Cam. *Ann. Statist.* 30, 631–682.
- Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. American Mathematical Society* 54, 426–482.