Tulane Economics Working Paper Series

Identification Using Stability Restrictions

Leandro M. Magnusson
Department of Economics
Tulane University
New Orleans, LA
lmagnuss@tulane.edu

Sophocles Mavroeidis
Brown University
Providence, Rhode Island

Working Paper 1116
June 2010

Abstract

Structural change, typically induced by policy regime shifts, is a common feature of dynamic economic models. We show that structural change can be used constructively to improve the identification of structural parameters that are stable over time. A leading example is models that are immune to the well-known Lucas (1976) critique. This insight is used to develop novel econometric methods that extend the widely used generalized method of moments (GMM). The proposed methods yield improved inference in a leading macroeconomic policy model.

Keywords: GMM, identification, structural stability, Lucas critique, new Keynesian models
JEL: C22, E31

Identification using stability restrictions*

Leandro M. Magnusson
Tulane University

Sophocles Mavroeidis
Brown University

June 16, 2010

Abstract

Structural change, typically induced by policy regime shifts, is a common feature of dynamic economic models. We show that structural change can be used constructively to improve the identification of structural parameters that are stable over time. A leading example is models that are immune to the well-known Lucas (1976) critique. This insight is used to develop novel econometric methods that extend the widely used generalized method of moments (GMM). The proposed methods yield improved inference in a leading macroeconomic policy model.

Keywords: GMM, identification, structural stability, Lucas critique, new Keynesian models.

JEL: C22, E31

*We would like to thank Don Andrews, Victor Chernozhukov, Bruce Hansen, Frank Kleibergen, Blaise Melly, Ulrich Mueller, Whitney Newey, Serena Ng, Argia Sbordone and Jim Stock for helpful comments and discussion. We also benefited from comments by seminar participants at Columbia, MIT, the NBER summer institute, the Greater New York Metropolitan Area workshop at Princeton and the North American summer meeting of the Econometric Society at Boston University.

1 Introduction

Structural change, typically induced by policy regime shifts, is a common feature of dynamic economic models. We show that structural change can be used constructively to improve the identification of structural parameters that are stable over time. This insight is used to develop novel econometric methods that extend the widely used generalized method of moments (GMM). The proposed methods yield improved inference in models that are used for the analysis of macroeconomic policy, so they have the potential to be widely used in practice.

The contribution of this paper is twofold. First, it makes a formal case for using stability restrictions (e.g., immunity to the well-known Lucas (1976) critique) as a source of identification of the stable structural parameters in economic models, or put differently, for using structural change to identify stable dynamic causal effects. The key insight is that changes in the distribution of the data induced by, for example, policy regime shifts, provide additional exogenous variation that can be usefully exploited for inference. This information is ignored by the usual GMM approach that relies only on full-sample exclusion or cross-equation restrictions to identify the structural parameters of the model. The current practice can be justified if there are no breaks in the data generating process, but we argue that this assumption is too strong in many contexts. For example, there is considerable evidence of parameter instability in macroeconomic models, see Stock and Watson (1996), Clarida, Galí, and Gertler (2000) and Sims and Zha (2006). Therefore, we expect *a priori* that the information contained in stability restrictions will be nontrivial, and our application confirms this empirically.

The second contribution is to develop new econometric methods for structural inference that exploit the information in stability restrictions and require only mild assumptions about the nature of instability in the distribution of the data. Specifically, our methods do not require any prior knowledge about the incidence, number and timing of breaks. Our main assumption, which is used in the literature on structural breaks, see Perron (2005), is that partial-sample moments satisfy a functional central limit theorem. Because no assumptions about identification are required, the main regularity conditions are strictly weaker than those used to justify the stability tests that are widely used in applied work, e.g., Andrews

(1993), Andrews and Ploberger (1994), Elliott and Mueller (2006). Therefore, the scope of the proposed methods is very wide.

We examine the empirical relevance of the proposed methods by applying them to a widely used macroeconomic model, the new Keynesian Phillips curve (NKPC). This model is known to suffer from problems of weak identification, see Kleibergen and Mavroeidis (2009) and the references therein. Identification robust confidence intervals that use only full-sample information are very wide, in some cases containing the entire parameter space. However, when using methods that exploit the stability restrictions, confidence sets on the parameters become drastically smaller.

The message of this paper is that structural change can be used constructively to improve the identification of structural parameters that are assumed to be stable over time. This differs markedly from the existing approaches that ignore the implications of stability restrictions for identification and use them only for post-estimation model evaluation. Two related papers by Li (2008) and Li and Mueller (2006) provide formal justification for this approach under certain conditions. They show that standard Wald tests on stable parameters remain valid in the face of time-variation in other parameters, provided that (i) the instability is small in the sense that it is not detectable with probability one asymptotically, and (ii) the structural parameters are well-identified by the available full-sample moment conditions. Instead, we do not impose any identification assumption, since it has been shown to be unrealistic in many cases, see, e.g., Stock, Wright, and Yogo (2002), nor do we rule out large instabilities in the data generating process.

This paper relates to the literature on identification via heteroskedasticity, see Lewbel (2003), Rigobon (2003) and Klein and Vella (2010). These papers obtain identification by exploiting a certain heterogeneity in the data generating process. In the case of Rigobon, this heterogeneity is changes in the volatility of the shocks in structural vector autoregressions. By specifying the moment conditions appropriately, our framework nests Rigobon's method, and is more general, in that it does not require any rank condition for identification, nor knowledge about the timing of breaks. Moreover, like the aforementioned three papers, identification may be achieved even in models that would be under-identified by conventional GMM, i.e., even when there are more parameters than exclusion restrictions, because partial-sample moment conditions provide the necessary additional identifying restrictions.

The paper also relates to Rossi (2005) who proposed GMM-based methods for testing parametric restrictions jointly with the hypothesis of stability of the parameters. Rossi did not consider the implications of stability restrictions for the identification of structural parameters but focused instead on the implication of stability restrictions for non-nested model comparisons, see also Giacomini and Rossi (2009). Our proposed methods also differ from theirs in that they are robust to identification failure. Finally, the scope of this paper also relates to the notion of 'super-exogeneity', see Engle, Hendry, and Richard (1983), as well as to the notion of 'co-breaking', see Hendry and Massmann (2007).

The outline of the paper is as follows. Section 2 presents our assumptions and motivating examples and describes the proposed methods. The following section provides the underlying asymptotic theory. Section 4 reports asymptotic power comparisons of the tests and results on their size in finite-sample. Section 5 presents an empirical application, and 6 concludes.

2 Assumptions and tests

Consider a p-dimensional vector of structural parameters θ whose parameter region Θ is a compact subset of $\mathbb{R}^p$, and suppose that we observe a sample of size T given by a triangular array of random variables $\{Y_{T,t} : t \leq T, T \geq 1\}$. The triangular array construction is used to account for instabilities in the data generating process. For notational convenience, we will drop the dependence of random variables in the sample on T where no confusion arises.

We assume that economic theory gives rise to a set of moment conditions that can be represented in terms of a k-dimensional function of data and parameters $f(\theta, Y_{T,t})$, abbreviated as $f_t(\theta)$ dropping the dependence on T for convenience, whose expectation vanishes at the true value of θ, i.e.,

$$E[f_t(\theta)] = 0 \quad \text{for all } t \leq T,\ T \geq 1. \tag{1}$$

For example, a typical Euler equation model with G equations gives rise to a set of conditional moment restrictions of the form $E[h_t(\theta)|\mathcal{I}_t] = 0$, where $h_t(\theta)$ is a G-dimensional function of data and parameters, e.g., a vector of residuals or structural errors, and $\mathcal{I}_t$ is the information set at time t. Given any set of instrumental variables $Z_t \in \Re^{G\times k}$ in $\mathcal{I}_t$, the conditional moment restrictions can be converted to unconditional restrictions in (1) by defining

$$f_t(\theta) = Z_t' h_t(\theta). \tag{2}$$

The single-equation linear instrumental variable (IV) model as well as the simultaneous equations model are special cases where $h_t(\theta)$ is linear.

Our interest lies in testing the null and alternative hypotheses

$$H_0 : \theta = \theta_0 \quad \text{and} \quad H_1 : \theta \neq \theta_0, \tag{3}$$

using tests with significance level α. The robustness requirement is that α-level tests should not reject H_0 more often than the nominal level asymptotically for a wide range of data generating processes (DGPs), satisfying a multivariate invariance principle for the sample moments, see Mueller (2008) for a motivation. There is a large class of tests that meet this requirement, and we shall therefore also address the question of efficiency by means of weighted average power (WAP) criteria.

Let the partial sums of the moment function $f_t(\theta)$ be denoted by

$$F_{sT}(\theta) = \sum_{t=1}^{[sT]} f_t(\theta) \tag{4}$$

where $[x]$ denotes the integer part of x and $s \in [0,1]$. The moment conditions (1) are equivalent to $E[F_{sT}(\theta)] = 0$ for all $s \in [0,1]$. We will refer to $F_{sT}(\theta)$ as partial-sample moments, and $F_T(\theta) \equiv F_{1T}(\theta)$ as full-sample moments.

The moment conditions (1) together with the null hypothesis $H_0 : \theta = \theta_0$ give rise to kT identifying restrictions. These restrictions can be written equivalently as the k restriction that $E[f_t(\theta_0)]$ is *zero on average*, i.e., $E(F_T(\theta_0)) = 0$, and the restriction that $E[f_t(\theta_0)]$ is *stable* over t. The usual approach to inference on the hypothesis (3) utilizes only the first k restrictions on the average value of $E[f_t(\theta_0)]$. We show in the next section that this approach wastes information unless $E[f_t(\theta_0)]$ is constant over t.

We now turn to inference procedures that exploit the information in the stability restrictions. Since our objective is to do inference using weak assumptions, we consider first asymptotically efficient tests based on the weak assumption that the partial-sample moments $F_{sT}(\theta_0)$ satisfy a multivariate invariance principle. We show in the next section that the resulting test statistics can be expressed as generalizations of the Anderson and Rubin (1949) statistic for GMM.

2.1 Generalized Anderson-Rubin tests

Let $X_T(s) = T^{-1/2}F_{sT}(\theta_0)$ denote the partial-sample moments at θ_0. We make the following high-level assumption about the large sample behavior of X_T under both H_0 and H_1.

Assumption 1 *The process $X_T(s) = T^{-1/2}F_{sT}(\theta_0)$ satisfies: (i) $X_T(\cdot) - E[X_T(\cdot)] \Rightarrow V_{ff}^{1/2}W(\cdot)$, where W is a standard $k \times 1$ Wiener process, V_{ff} is a positive definite $k \times k$ matrix, and $V_{ff}^{1/2}$ denotes its symmetric square root.*
(ii) $\lim_{T\to\infty} var[X_T(s)] = sV_{ff}$ uniformly in s.
(iii) There exists a consistent estimator of V_{ff}, denoted $\hat{V}_{ff}(\theta_0)$.

Primitive conditions for the high-level Assumption 1 can be found in various papers in the stability literature, e.g., Andrews (1993) and Sowell (1996). For instance, when the moment functions are given by equation (2), Assumption 1 will be satisfied when $h_t(\theta_0)$ is strong mixing with finite moments of order greater than 2, and Z_t is asymptotically mse stationary, see Hansen (2000). Asymptotic mse stationarity is weaker than strict stationarity and allows for non-permanent changes in the marginal distribution of Z_t.

Assumption 1 strengthens Stock and Wright (2000, Assumption A), which corresponds to the special case of $s = 1$, above. In the context of a linear model, this assumption places no restrictions upon the so-called 'first-stage' (see examples below). This assumption is sufficient to provide useful tests with robustness and asymptotic efficiency properties. We consider also a stronger condition below which enables us to obtain score and quasi Likelihood ratio tests, see Assumption 3.

It is important to acknowledge that Assumption 1 excludes permanent changes in the variance of the moment conditions. This assumption is shared by all tests of structural change proposed in the literature, so it does not limit the applicability of our results any more than for any other stability test. Technically, this assumption is necessary for the proposed tests to control size asymptotically, see Hansen (2000). However, it does not preclude all changes in the variance of the moment conditions. For example, it is sufficient to assume an asymptotic sequence such that the magnitude of any changes in the variance of the sample moments converges to zero as the sample increases. This is similar to the approach used in Bai and Perron (1998) to obtain pivotal statistics for inference on break dates, see Bai and Perron (1998, Assumption A6). Therefore, Assumption 1 does not preclude changes in the variance that can be detected with very high probability. In addition, the results in Hansen (2000) indicate that the size distortions induced by such departures from assumption 1 are modest, and our numerical results below confirm that.

An important requirement for robustness is that tests should control size in cases when θ may be arbitrarily weakly identified. To make this precise, we adopt the local-to-zero asymptotic nesting of Stock and Wright (2000).

Assumption 2 $E[X_T(s)] \rightarrow \int_0^s m(\theta, r)\, dr$ *uniformly in* s*, where the function* $m(\theta, \cdot)$ *is bounded, it belongs to* $D^k_{[0,1]}$*, the space of functions on* $[0,1]$ *that are right-continuous with finite left limits (also known as cadlag), and* $m(\theta_0, s) = 0$ *for all* $s \in [0,1]$.

At $s = 1$, Assumption 2 corresponds to the weak identification assumption in Stock and Wright (2000).[1] This assumption makes precise the notion that the moment conditions (1) are nearly satisfied even when the true value θ is far from the hypothesized value θ_0. The distribution of the data under H_1 is contiguous to the distribution under H_0, meaning that no consistent test exists under these asymptotics. However, the assumption implies that efficient tests will have non-trivial power, except in the degenerate case $m(\theta; s) = 0$ for all s and θ.

The key addition to Stock and Wright's framework is that Assumption 2 allows us to characterize the behavior of the moment conditions also over subsamples, and thus model time variation in $E[f_t(\theta_0)]$ under H_1. The special case in which $E[f_t(\theta_0)]$ is approximately constant to order $T^{-1/2}$ in large samples corresponds to $m(\theta, s)$ being constant in terms of s. Assumption 2 implies that any time variation in $E[f_t(\theta_0)]$ under H_1 is of the same order of magnitude as the full-sample moment conditions $E[F_T(\theta_0)]$. This ensures that the informational content of stability restrictions is comparable to that of the full-sample moment conditions.

[1] Because we do not seek to characterize the behavior of estimators of θ, we do not need uniform convergence and differentiability of $m(\theta; s)$ with respect to θ.

The function $m(\theta, \cdot)$ in Assumption 2 can accommodate most types of instability that have been used in the literature on structural change. Specifically, $m(\theta, \cdot)$ can be a step function with a finite number of discontinuities, corresponding to a fixed number of structural 'breaks' or distinct 'regimes', as in Andrews (1993), Sowell (1996) or Bai and Perron (1998). It can also be a realization of a continuous stochastic process, such as a martingale process, as in Stock and Watson (1996), or the general persistent time variation process studied in Elliott and Mueller (2006), representing slow continuous time variation. It could also be a smooth deterministic function of time, such as a spline, representing a smooth transition between different regimes.

Next, we give a couple of motivating examples.

Example 1: Identification through policy regime shifts Consider the structural model

$$y_t = \beta E_t y_{t+1} + \gamma x_t + u_t, \tag{5}$$

where E_t denotes expectations conditional on information available at time t, and u_t is an unobserved shock, which is assumed to be uncorrelated with lags of the observables, y and x. The above equation can be thought of as a (possibly linearized version of) an Euler equation that determines the optimal choice of y_t by an economic agent given their objective function. The parameters β and γ will then be directly related to some 'deep' structural parameters that characterize the objective function (e.g., discount factors, elasticities, etc).

We want to do inference on β and γ using the identifying assumption (2) with $\theta = (\beta, \gamma)'$, $h_t(\theta) = y_t - \beta y_{t+1} - \gamma x_t$, and Z_t a $1 \times k$ vector containing lags of y_t and x_t. Identification depends on the distribution of x_t. Suppose that x_t is a policy variable determined according to an inertial feedback rule of the form

$$x_t = \rho x_{t-1} + (1 - \rho) \varphi y_t + \varepsilon_t, \tag{6}$$

where ε_t is an unobserved 'policy' shock. Then, it can be shown that, in a determinate rational expectations equilibrium, the only relevant instrument is x_{t-1}. Consequently, the parameters β and γ are under-identified, because there are two endogenous regressors y_{t+1}

and x_t.[2]

Now, suppose that policy changes over time, e.g., φ becomes φ_t, but the parameters in equation (5) remain stable, i.e., immune to the Lucas critique. Then, a single change in the policy parameters at date t_b, say, suffices to induce identification: interacting x_{t-1} with the indicator $1_{\{t>t_b\}}$, generates an additional relevant instrument.

The objective of this paper is to exploit the information in such changes that leave the structural parameters of interest unaltered, without making any *a priori* assumptions about the incidence, nature and timing of these changes. For example, it may be that φ changes 'very little' or not at all, or that it changes a fixed number of times at unknown dates, or that it drifts 'randomly' over time. Our proposed methods accommodate these alternatives. Assumption 2 (which is not necessary for the validity of the methods we propose later), will be satisfied if $\varphi_t = \varphi_0 + O\left(T^{-1/2}\right)$. This assumption also ensures that the instruments Z_t, which are lags of x_t, y_t, are asymptotically mse stationary, which was mentioned above as a primitive condition for Assumption 1.[3]

□

Example 2: Identification through changes in variance Consider the bivariate simultaneous equations model

$$\begin{pmatrix} 1 & -\beta \\ -\gamma & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix}$$

with mutually and serially uncorrelated errors u_t, ε_t, interpreted as structural shocks. The assumption $E(u_t \varepsilon_t) = 0$ implies a single identifying restriction of the form (1), with $\theta = (\beta, \gamma)'$ and $f_t(\theta) = y_t x_t (1 + \beta\gamma) - \beta x_t^2 - \gamma y_t^2$. Therefore, the parameters θ are under-identified. Now, suppose that the volatilities of the shocks are time-varying, so that $E(f_t(\theta_0)) = (\beta\gamma - \beta_0\gamma_0) E(y_t x_t) - (\beta - \beta_0) E\left(x_t^2\right) - (\gamma - \gamma_0) E\left(y_t^2\right)$ is not constant under H_1. A single break in the volatility of at least one of the shocks in this example suffices to induce identification. This idea was used before in Rigobon (2003) and Klein and Vella (2009), who developed it in the context of structural vector autoregressions.

□

[2] This follows because the dynamics are given by a restricted first order vector autoregression of the form $y_t = \alpha_1 x_{t-1} + v_t^y$, $x_t = \rho_1 x_{t-1} + v_t^x$, where v_t^y, v_t^x are innovations.

[3] The weaker condition $\varphi_t - \varphi_0 = o(1)$ is actually sufficient for asymptotic mse stationarity.

A suitable approach to asymptotic efficiency is the one proposed by Mueller (2008). Mueller shows that the robustness requirement that tests should control size asymptotically for all data generating processes that satisfy Assumption 1 means that we must restrict attention to statistics that are functionals of $F_{sT}(\theta_0)$. Moreover, asymptotically efficient tests can be obtained by evaluating efficient tests in the limiting problem at their sample analogue.

In the next section, we show that asymptotically efficient tests based on Assumption 1 can be expressed as joint tests of the validity of the full-sample moment restrictions $E[F_T(\theta_0)] = 0$, and the restriction that $E[f_t(\theta_0)]$ is stable. The test statistics we derive can be written in the form

$$gen\text{-}AR_T(\theta_0) = \frac{\bar{c}}{1+\bar{c}} GMM\text{-}AR_T(\theta_0) + stab\text{-}AR_T^{\tilde{c}}(\theta_0) \tag{7}$$

where $GMM\text{-}AR_T(\theta_0)$ tests the validity of the full-sample moment restrictions, $stab\text{-}AR_T^{\tilde{c}}(\theta_0)$ tests the stability restrictions under H_0, and $\bar{c}$ and $\tilde{c}$ are non-negative scalars that determine the weight the investigator attaches to violations of the full-sample moment restrictions and stability restrictions, respectively, under H_1. The statistic $gen\text{-}AR_T(\theta_0)$ is asymptotically pivotal, and a test that rejects for large values of $gen\text{-}AR_T(\theta_0)$ is asymptotically efficient as described in Section 3. Asymptotic critical values are nonstandard, but can be computed by simulation.

The first component of $gen\text{-}AR_T(\theta_0)$ is given by

$$GMM\text{-}AR_T(\theta_0) = \frac{1}{T} F_T(\theta_0)' \hat{V}_{ff}(\theta_0)^{-1} F_T(\theta_0) \tag{8}$$

where $\hat{V}_{ff}(\theta_0)$ is a consistent estimator of the variance of $T^{-1/2} F_T(\theta_0)$. This is the statistic proposed by Stock and Wright (2000), and it can be seen as a special case of the $gen\text{-}AR_T$ statistic when the investigator puts zero weight ($\tilde{c} = 0$) on instability under H_1. The specification of $stab\text{-}AR_T$ statistic, the second component of $gen\text{-}AR_T$, depends on the assumptions about the nature of instability under H_1. We consider the two leading cases: a single break at unknown date τ, and martingale time variation.

In the case of a single break at unknown date, the $stab\text{-}AR_T$ statistic can be obtained in the following steps:

1. Specify $[t_l, t_u]$, the range of candidate break dates, typically $t_l = [0.15T]$ and $t_u =$

$[0.85T]$.

2. Split the sample at each candidate break date $t_b \in [t_l, t_u]$, compute the moment functions in each subsample $F_T^1\left(\theta_0, \frac{t_b}{T}\right) = F_{t_b}(\theta_0)$ and $F_T^2\left(\theta_0, \frac{t_b}{T}\right) = F_T(\theta_0) - F_{t_b}(\theta_0)$, obtain estimates of their variance $\hat{V}_{ff}^i\left(\theta_0, \frac{t_b}{T}\right)$ $i = 1, 2$, and evaluate the partial-sample (continuously updated) GMM objective function $S_T\left(\theta_0; \frac{t_b}{T}\right) = \sum_{i=1}^{2} T_i^{-1} F_T^i\left(\theta_0, \frac{t_b}{T}\right)'$ $\hat{V}_{ff}^i\left(\theta_0, \frac{t_b}{T}\right)^{-1} F_T^i\left(\theta_0, \frac{t_b}{T}\right)$, with $T_1 = t_b$ and $T_2 = T - T_1$. [For $\hat{V}_{ff}^i\left(\theta_0, \frac{t_b}{T}\right)$, use either partial-sample estimators, e.g., Newey-West (1987) computed using data in subsample i, or the full-sample estimator $\hat{V}_{ff}(\theta_0)$].

3. Compute $GMM\text{-}AR_T(\theta_0)$ using equation (8), and obtain $S_T^B\left(\theta_0; \frac{t_b}{T}\right) = S_T\left(\theta_0; \frac{t_b}{T}\right) - GMM\text{-}AR_T(\theta_0)$ for each t_b.

4. Finally, compute $2\log\left[\frac{1}{\mathrm{T}(t_u,t_l)}\sum_{t_b=t_l}^{t_u}\exp\left\{\frac{1}{2}\frac{\tilde{c}}{1+\tilde{c}}S_T^B\left(\theta_0, \frac{t_b}{T}\right)\right\}\right]$ where $\mathrm{T}(t_u, t_l) = t_u - t_l + 1$.

Since the $stab\text{-}AR_T$ statistic depends on $\tilde{c}$, we select particular versions of it following the approach in the stability literature, see Andrews and Ploberger (1994). Specifically, letting $\tilde{c} \to 0$ or $\tilde{c} \to \infty$ we obtain, respectively, the special cases $ave\text{-}AR_T^B(\theta_0) = \frac{1}{\mathrm{T}(t_u,t_l)}\sum_{t_b=t_l}^{t_u} S_T^B\left(\theta_0, \frac{t_b}{T}\right)$ and $\exp\text{-}AR_T^B(\theta_0) = 2\log\left[\frac{1}{\mathrm{T}(t_u,t_l)}\sum_{t_b=t_l}^{t_u}\exp\left\{S_T^B\left(\theta_0, \frac{t_b}{T}\right)\right\}\right]$. Letting $\bar{c} = \tilde{c} = c$, the resulting test statistics for $c \to 0$ and $c \to \infty$, are $ave\text{-}AR_T(\theta_0) = \frac{1}{\mathrm{T}(t_u,t_l)}\sum_{t_b=t_l}^{t_u} S_T\left(\theta_0, \frac{t_b}{T}\right)$ and $\exp\text{-}AR_T(\theta_0) = 2\log\left[\frac{1}{\mathrm{T}(t_u,t_l)}\sum_{t_b=t_l}^{t_u}\exp\left\{S_T\left(\theta_0, \frac{t_b}{T}\right)\right\}\right]$, respectively.

In the case of martingale time variation, the statistic $stab\text{-}AR_T^{\tilde{c}}(\theta_0)$ can be obtained in the following steps.

1. Compute $v_t = \hat{V}_{ff}(\theta_0)^{-1/2} f_t(\theta_0)\ (k \times 1)$, and denote the ith element by $v_{t,i}$, $i = 1, ..., k$.

2. For each $\{v_{t,i}\}$, compute the new series $w_{1,i} = v_{1,i}$ and $w_{t,i} = \tilde{r}w_{t-1,i} + \Delta v_{t,i}$, for $t = 2, ..., T$, with $\tilde{r} = 1 - \frac{\tilde{c}}{T}$.

3. Regress $\{w_{t,i}\}$ on $\left\{\tilde{r}^t\right\}$ and obtain the squared residuals, sum over all $i = 1, ..., k$ and multiply by $\tilde{r}$.

4. Compute $\sum_{i=1}^{k}\sum_{t=1}^{T}(v_{i,t} - \overline{v}_i)^2$, where $\overline{v}_i = T^{-1}\sum_{t=1}^{T} v_{i,t}$ and subtract the quantity in step 3 from it.

This statistic is very similar to (the negative of) the qLL statistic proposed by Elliott and Mueller (2006) to test against persistent time variation in regression coefficients. (qLL stands for quasi local level). Following Elliott and Mueller (2006), we set $\tilde{c} = 10$, and denote the resulting stability statistic as $qLL\text{-}AR_T^B$. For the joint test statistic in Equation (7), we set $\bar{c} = 10$ in order to give equal weights to the two alternatives, and denote the resulting test as $qLL\text{-}AR_T$.

2.2 Split-sample tests

When the moment functions $f_t(\theta)$ are linear, e.g., $f_t(\theta) = f_t(\theta_0) + q_t(\theta - \theta_0)$, information from stability restrictions arises from time-variation in the expectation of their Jacobian $E(q_t)$. Consider the leading case of a one-time change in the expected Jacobian at some date t_b. If t_b were known, an obvious approach to inference would be to split the sample at t_b and proceed with *GMM* estimation using the additional k moment conditions generated by the break. The resulting 'split-sample' continuously-updated *GMM* criterion function would be a 'split-sample' $GMM\text{-}AR_T$ statistic, or simply $split\text{-}AR_T$, which is a special case of the $ave\text{-}AR_T$ and $exp\text{-}AR_T$ statistics when the break date is known.[4] Moreover, under Assumption 1, asymptotic critical values for the $split\text{-}AR_T$ test can be obtained from a $\chi^2(2k)$ distribution.

In addition to the split-sample $GMM\text{-}AR_T$ test, under an additional mild assumption on the Jacobian, as in Kleibergen (2005, Assumption 1), we could also obtain identification robust 'split-sample' conditional score (*split-KLM*) and 'split-sample' quasi-likelihood ratio (*split-MLR*) tests, the latter being a generalization of the conditional *LR* test of Moreira (2003). The motivation for considering such tests is that they are asymptotically more powerful than the $GMM\text{-}AR_T$ test under strong identification, see Andrews and Stock (2005).

Since we typically do not know the break date, we can obtain feasible versions of the aforementioned tests by evaluating them at an estimated break date. For this purpose, Assumption 1 is insufficient, because the break date is not identified under the null from the distribution of the sample moments alone. Therefore, we need an assumption about the joint distribution of the sample moments and their Jacobian.

[4]Think of the analogy between the Chow test and the Quandt likelihood ratio test.

Assumption 3 *Let* $q_t(\theta_0) = vec[\partial f_t(\theta_0)/\partial\theta']$, $Q_{sT}(\theta_0) = \sum_{t=1}^{[sT]} q_t(\theta_0)$. *We assume that*
(i)

$$T^{-1/2}\left\{\begin{pmatrix} F_{\cdot T}(\theta_0) \\ Q_{\cdot T}(\theta_0)\end{pmatrix} - E\left[\begin{pmatrix} F_{\cdot T}(\theta_0) \\ Q_{\cdot T}(\theta_0)\end{pmatrix}\right]\right\} \Rightarrow \begin{pmatrix} V_{ff} & V_{fq} \\ V_{qf} & V_{qq}\end{pmatrix}^{1/2}\begin{pmatrix} W_f(\cdot) \\ W_q(\cdot)\end{pmatrix}$$

where W_f *and* W_q *are independent standard* $k \times 1$ *and* $kp \times 1$ *Wiener processes.*
(ii) $\lim_{T\to\infty} var\left[F_{sT}(\theta_0) \vdots Q_{sT}(\theta_0)\right] = sV$ *uniformly in* s, *where* $V = \begin{pmatrix} V_{ff} & V_{fq} \\ V_{qf} & V_{qq}\end{pmatrix}$.
(iii) There exists a consistent estimator $\hat{V}_T$ *of* V.

Assumption 3 is a stronger version of Assumptions 1 and 2 in Kleibergen (2005). The latter correspond to the special case of $s = 1$ in the former. Assumption 3 avoids placing any restrictions on the (infinitely dimensional) nuisance parameter $\lim_{T\to\infty} T^{-1}E[Q_{sT}(\theta_0)]$, which are difficult to verify. In particular, we avoid making any assumptions about identification or the incidence and magnitude of breaks.

Example 3: Linear IV regression with time-varying first stage The model consists of a structural and a reduced-form equation. The structural equation is

$$y_{1,t} = Y_{2,t}\theta + u_t, \quad t = 1, ..., T \tag{9}$$

where $\{y_{1,t}, Y_{2,t}\}_{t=1}^T$ is a sequence of $1 \times (1+p)$ random vectors, $\{u_t\}_{t=1}^T$ is a (structural) error, and $\theta \in \Re^p$ is the unknown structural parameter. The reduced-form equation (also known as first-stage regression) is given by

$$Y_{2,t} = Z_t\Pi_t + V_{2,t}, \quad t = 1, ..., T \tag{10}$$

where $Z_t \in \Re^{1\times k}$, $t = 1, ..., T$ is a sequence of observed instrumental variables that are fixed, $V_{2,t} \in \Re^{1\times p}$, $t = 1, ..., T$ is a (reduced form) error vector, and $\Pi_t \in \Re^{k\times p}$, $t = 1, ..., T$ is a sequence of unknown parameters. The errors u_t and $v_{2,t}$ are iid and have a mean zero bivariate normal distribution. The identifying restrictions in this model are $E(Z_t u_t) = 0$ for all t, and the moment function $f_t(\theta)$ in equation (1) is $f_t(\theta) = Z_t(y_{1,t} - Y_{2,t}\theta)$.

Assumption 3 is satisfied if $\frac{1}{T}\sum_{t=1}^{[sT]} Z_t' Z_t \to sQ_{ZZ}$ uniformly in s, where Q_{ZZ} is nonsingular. Assumption 2 is satisfied when $\Pi_t = O\left(T^{-1/2}\right)$, as in Staiger and Stock (1997), with $m(\theta, s) = Q_{ZZ} \lim_{T\to\infty} T^{1/2}\Pi_{[sT]}(\theta - \theta_0)$.

□

Under Assumption 3, we can evaluate standard identification robust *GMM* tests based on the partial-sample *GMM* objective function, where the break date has been estimated. We refer to such tests as split-sample tests. The robustness objective is that the split-sample tests should control size irrespective of whether there has been a break or not, or whether the break date is consistently estimable, or even when the nature of instability has been misspecified. This is achieved by the following procedure.

1. Estimate the break date. Specify a range of break dates $[t_l, t_u]$, typically $t_l = [0.15T]$ and $t_u = [0.85T]$. For each $t_b \in [t_l, t_u]$ compute the two subsample moments $F_T^1\left(\theta_0, \frac{t_b}{T}\right) = F_{t_b}(\theta_0)$ and $F_T^2\left(\theta_0, \frac{t_b}{T}\right) = F_T(\theta_0) - F_{t_b}(\theta_0)$, and their Jacobians $Q_T^1\left(\theta_0, \frac{t_b}{T}\right) = Q_{t_b}(\theta_0)$ and $Q_T^2\left(\theta_0, \frac{t_b}{T}\right) = Q_T(\theta_0) - Q_{t_b}(\theta_0)$, and estimate their $k(p+1) \times k(p+1)$ variance matrix $\hat{V}_T^i\left(\theta_0, \frac{t_b}{T}\right)$, $i = 1, 2$, using either partial-sample estimators or a full-sample estimator. Compute the $k \times p$ matrices $D_T^i\left(\theta_0, \frac{t_b}{T}\right)$, $i = 1, 2$, from

$$vec\left[D_T^i(\theta_0, \tau)\right] = vec\left[Q_T^i(\theta_0, \tau)\right] - \hat{V}_{qf}^i(\theta_0, \tau)\, \hat{V}_{ff}^i(\theta_0, \tau)^{-1} F_T^i(\theta_0, \tau), \tag{11}$$

and estimate their variance by

$$\hat{V}_{qq.f}^i\left(\theta_0, \frac{t_b}{T}\right) = \hat{V}_{qq}^i\left(\theta_0, \frac{t_b}{T}\right) - \hat{V}_{qf}^i\left(\theta_0, \frac{t_b}{T}\right) \hat{V}_{ff}^i\left(\theta_0, \frac{t_b}{T}\right)^{-1} \hat{V}_{qf}^i\left(\theta_0, \frac{t_b}{T}\right)'. \tag{12}$$

Obtain the restricted estimator of the break date by

$$\widetilde{t}_b(\theta_0) = \arg\max_{t_b} \sum_{i=1}^{2} T_i^{-1} vec\left[D_T^i\left(\theta_0, \frac{t_b}{T}\right)\right]' \hat{V}_{qq.f}^i\left(\theta_0, \frac{t_b}{T}\right)^{-1} vec\left[D_T^i\left(\theta_0, \frac{t_b}{T}\right)\right],$$

with $T_1 = t_b$ and $T_2 = T - T_1$. This can be expressed equivalently as $\widetilde{\tau}(\theta_0) = \widetilde{t}_b(\theta_0)/T$.

2. Split-sample test statistics. The identification robust split-sample statistics are given by

$$
\begin{aligned}
&split\text{-}AR_T(\theta_0,\tau) = \textstyle\sum_{i=1}^{2} T_i^{-1} F_T^i(\theta_0,\tau)' \hat{V}_{ff}^i(\theta_0,\tau)^{-1} F_T^i(\theta_0,\tau) \\
&split\text{-}KLM_T(\theta_0,\tau) = \textstyle\sum_{i=1}^{2} T_i^{-1} F_T^i(\theta_0,\tau)' \hat{V}_{ff}^i(\theta_0,\tau)^{-1/2} P_{\hat{V}_{ff}^i(\theta_0,\tau)^{-1/2} D_T^i(\theta_0,\tau)} \\
&\qquad\qquad \hat{V}_{ff}^i(\theta_0,\tau)^{-1/2} F_T^i(\theta_0,\tau) \\
&split\text{-}JKLM_T(\theta_0,\tau) = split\text{-}AR_T(\theta_0,\tau) - split\text{-}KLM_T(\theta_0,\tau)
\end{aligned}
$$

and

$$
\begin{aligned}
split\text{-}MLR_T(\theta_0,\tau) = \tfrac{1}{2}\Big[& split\text{-}AR_T(\theta_0,\tau) - rk_\theta(\theta_0,\tau) + \\
& \sqrt{(split\text{-}AR_T(\theta_0,\tau) + rk_\theta(\theta_0,\tau))^2 - 4 split\text{-}JKLM_T(\theta_0,\tau)\, rk_\theta(\theta_0,\tau)}\Big]
\end{aligned}
\tag{13}
$$

with

$$
\begin{aligned}
rk_\theta(\theta,\tau) = \min_{\varphi\in\mathbb{R}^{p-1}} & \tbinom{1}{\varphi}' \textstyle\sum_{i=1}^{2} \frac{1}{T_i} D_T^i(\theta,\tau)' \\
& \left[\left(\tbinom{1}{\varphi}\otimes I_k\right)' \hat{V}_{qq.f}^i(\theta) \left(\tbinom{1}{\varphi}\otimes I_k\right)\right]^{-1} D_T^i(\theta,\tau)\tbinom{1}{\varphi}.
\end{aligned}
\tag{14}
$$

Evaluate the above statistics at $\tau = \widetilde{\tau}(\theta_0)$. We can also define a split-sample version of the KLM-JKLM combination test of Kleibergen (2005).

3. Critical values. Conditional critical values for the split-sample tests evaluated at $\widetilde{\tau}(\theta_0)$ are given by the asymptotic distributions that arise as if $\widetilde{\tau}(\theta_0)$ were nonrandom. This is because, under Assumption 3 and $H_0 : \theta = \theta_0$, $\widetilde{\tau}(\theta_0)$ is asymptotically independent of $T^{-1/2} F_{\cdot T}(\theta_0)$. In the case of the *split-AR*$_T$, *split-KLM*$_T$ and *split-JKLM*$_T$ statistics, the asymptotic distributions are $\chi^2(2k)$, $\chi^2(p)$ and $\chi^2(2k-p)$, respectively. For the *split-MLR*$_T$ test, critical values conditional on $rk_\theta(\theta_0, \widetilde{\tau}(\theta_0))$ can be computed by simulation.

3 Asymptotic theory

Assumption 1 and the continuous mapping theorem imply that, under H_0, $G(X_T(s)) \Rightarrow G(W(s))$ for any continuous functional $G(\cdot)$ on $D_{[0,1]}^k$. Thus, there exists a large class of asymptotically pivotal statistics that can be used to test the null hypothesis $H_0 : \theta = \theta_0$.

Define the random element X, such that $X_T \Rightarrow X$, and let ν_0, ν_1 denote the probability measures of X under H_0 and H_1 respectively. We shall obtain efficient tests for the limiting problem of testing ν_0 against ν_1, and then we will evaluate these tests at their sample analogue using X_T and an estimator of the long-run variance V_X. Since no uniformly most powerful test exists, we will make use of weighted average power (WAP) criteria over different alternatives.

Under Assumptions 1 and 2, ν_1 is determined by the stochastic differential equation $dX(s) = m(\theta, s)\, ds + V_X^{1/2} dW(s)$, and ν_0 is determined by $dX(s) = V_X^{1/2} dW(s)$. Therefore, ν_1 is absolutely continuous with respect to ν_0, and the Radon Nikodym derivative of ν_1 with respect to ν_0 conditional on the entire path of $m(\theta, \cdot)$ is given by

$$\xi(m) = \exp\left\{\int_0^1 m(\theta, s)' V_X^{-1} dX(s) - \frac{1}{2}\int_0^1 m(\theta, s)' V_X^{-1} m(\theta, s)\, ds\right\}. \tag{15}$$

Under the maintained assumptions, the process $X(s)$ can be decomposed orthogonally into $\overline{X} \equiv X(1)$ and $\tilde{X}(s) \equiv X(s) - sX(1)$. Define $\overline{m}(\theta) = \int_0^1 m(\theta, s)\, ds$ and $\widetilde{m}(\theta, s) = m(\theta, s) - \overline{m}(\theta)$, and note that the random function $\xi(m)$ in equation (15) factors into the product of

$$\overline{\xi}(\overline{m}) = \exp\left\{\overline{m}(\theta)' V_X^{-1} \overline{X} - \frac{1}{2}\overline{m}(\theta)' V_X^{-1} \overline{m}(\theta)\right\}, \tag{16}$$

and

$$\widetilde{\xi}(\widetilde{m}) = \exp\left\{\int_0^1 \widetilde{m}(\theta, s)' V_X^{-1} d\tilde{X}(s) - \frac{1}{2}\int_0^1 \widetilde{m}(\theta, s)' V_X^{-1} \widetilde{m}(\theta, s)\, ds\right\}. \tag{17}$$

It follows that the statistic $\tilde{X}(s)$ is ancillary for θ if and only if $m(\theta, \cdot)$ is constant. The finite-sample analogue of the statistic $\tilde{X}(s)$ is $X_T(s) - sX_T(1)$ or $T^{-1/2}[F_{sT}(\theta_0) - sF_T(\theta_0)]$, and captures subsample variations in the moment functions $f_t(\theta_0)$ that is asymptotically independent of the full-sample moments $T^{-1/2} F_T(\theta_0)$. Therefore, we have established the following result.

Proposition 1 *Under Assumptions 1 and 2, the statistic $F_{sT}(\theta_0) - sF_T(\theta_0)$ is asymptotically ancillary for θ if and only if $T^{1/2} E[f_t(\theta_0)]$ is approximately constant.*

Proposition 1 shows that ignoring the stability restrictions implicit in the moment conditions $E[f_t(\theta_0)] = 0$ for all $t \leq T$, cannot be optimal, except in the special case when $E[f_t(\theta_0)]$ is constant, i.e., when the stability restrictions are vacuous. Therefore, a test that ignores subsample variations in the moment functions $f_t(\theta_0)$ cannot be efficient, in general.

Example 3 (continued): Linear IV The constant-parameter IV regression model is a special case of the model given by equations (9) and (10) with $\Pi_t = \Pi$ for all t. The assumption of normality of the errors implies the availability of the low-dimensional sufficient statistic, $\sum_{t=1}^{T} Z_t' Y_t \in \Re^{k\times(1+p)}$, where $Y_t = (y_{1,t}, Y_{2,t})$, see Andrews, Moreira, and Stock (2006). When $\Pi_t \neq \Pi$, the aforementioned statistic is no longer sufficient. From the factorization theorem, a sufficient statistic is given by the sequence $\{Z_t' Y_t\}_{t=1}^{T}$, so the statistic $\sum_{t=1}^{T} Z_t' Y_t - s \sum_{t=1}^{sT} Z_t' Y_t$ is not ancillary, in line with Proposition 1.

□

The limiting problem of testing $H_0 : \theta = \theta_0$ against the composite alternative $H_1 : \theta \neq \theta_0$ is equivalent to testing $H_0 : m(\theta, s) = 0$ for all s against $H_1 : m(\theta, s) \neq 0$ for some s. An asymptotically point optimal test in the limiting problem is given by the statistic $\xi(m)$ in (15). Let ν_m denote a probability measure for the process $m(\theta, s)$ under H_1. A WAP maximizing test in the limiting problem is then given by the likelihood ratio (LR) statistic $\int \xi(m)\, d\nu_m$. The finite sample counterpart of the (LR) statistic is obtained by substituting $T^{-1/2} F_{sT}(\theta_0)$ and $V_T(\theta_0)$ for $X(s)$ and V_X, respectively, in equation (15). For simplicity, we omit the dependence of $m(\theta, s)$ on θ in the remainder of this section.

Alternative WAP tests differ in the specification of ν_m. Since m is a linear function of $\overline{m}$ and $\widetilde{m}$, corresponding to violation of the full-sample moment restrictions and the stability restrictions, respectively, we can equivalently specify ν_m as a joint measure over $\overline{m}$ and $\widetilde{m}$. Because $\overline{m}$ and $\widetilde{m}$ correspond to the independent statistics $\overline{X}$ and $\widetilde{X}$, it is reasonable to specify independent distributions of weights over $\overline{m}$ and $\widetilde{m}$, so the joint measure is given by the product of $\nu_{\overline{m}}$ and $\nu_{\widetilde{m}}$.

For $\nu_{\overline{m}}$, we will use the conventional weight distribution $\overline{m} \sim N(0, \bar{c}V_X)$, which puts equal weights over alternatives that are 'equally hard to detect'. The scalar parameter $\bar{c}$ measures the magnitude of the violation of the full-sample moment conditions. This is the distribution used to motivate the standard Wald (1943) test.

For $\nu_{\widetilde{m}}$, which corresponds to the stability restrictions, we will consider the two leading alternatives in the stability literature: (i) a fixed number of breaks at unknown break dates, as in Andrews (1993), Andrews and Ploberger (1994), and Sowell (1996); and (ii) persistent time variation, as in Stock and Watson (1998) and Elliott and Mueller (2006). In both cases, we will index $\nu_{\widetilde{m}}$ by a scalar parameter $\tilde{c}$ that measures the magnitude of the instability under H_1. Power can be directed towards stability restrictions versus full-sample moment restrictions by varying $\tilde{c}$ relative to $\bar{c}$.

3.1 Single break at unknown date

We focus on the leading case of a single break at an unknown date $\tau \in \varsigma \subset (0,1)$, defining two regimes in $m(s)$.

Assumption 4 *$m(s) = m_1 1_{\{s<\tau\}} + m_2 1_{\{s\geq\tau\}}$, for some $\tau \in \varsigma \subset (0,1)$.*

The analysis of multiple breaks is straightforward, but the resulting tests are much more computationally intensive. In addition, efficient tests against a single break typically have high power even against multiple breaks, and they have better finite-sample size properties than tests that are optimal against multiple breaks. So, for practical purposes, restricting attention to the case of a single break seems reasonable.

Assumption 4 implies $\overline{m} = \tau m_1 + (1-\tau) m_2$ and $\widetilde{m}(s) = \left(1_{\{s<\tau\}} - \tau\right)(m_1 - m_2)$. Let $\delta = \tau(1-\tau)(m_1 - m_2)$ denote the mean of $\widetilde{X}(\tau)$ under H_1, i.e., $\widetilde{X}(\tau) \overset{H_1}{\sim} N(\delta, \tau(1-\tau)V_X)$. Then, $\widetilde{\xi}(\widetilde{m})$ in equation (17) can be written in terms of δ and τ as:

$$\widetilde{\xi}(\delta,\tau) = \exp\left\{\delta'\left[\tau(1-\tau)V_X\right]^{-1}\widetilde{X}(\tau) - \frac{1}{2}\delta'\left[\tau(1-\tau)V_X\right]^{-1}\delta\right\}.$$

To obtain WAP maximizing tests we need a probability measure for δ and τ, which we specify as $\nu_{\delta|\tau} \times \nu_\tau$, where $\nu_{\delta|\tau}$ has density $N(0, c\tau(1-\tau)V_X)$. For ν_τ we will choose a uniform distribution, following the convention in the stability literature, see Andrews and Ploberger (1994). For any ν_τ we have

$$\widetilde{LR}^c = \int_\varsigma \int_{-\infty}^{+\infty} LR(\delta,\tau)\, d\nu_{\delta|\tau} d\nu_\tau = (1+c)^{-k/2}\int_\varsigma \exp\left\{\frac{1}{2}\frac{c}{1+c}\widetilde{X}(\tau)'\left[\tau(1-\tau)V_X\right]^{-1}\widetilde{X}(\tau)\right\} d\nu_\tau.$$

So, $LR^{\bar{c},\tilde{c}} = \overline{LR}^{\bar{c}}\widetilde{LR}^{\tilde{c}}$, where:

$$\overline{LR}^c = (1+c)^{-k/2}\exp\left\{\frac{1}{2}\frac{c}{1+c}\overline{X}'V_X^{-1}\overline{X}\right\}. \tag{18}$$

A test that rejects for large values of $LR^{\bar{c},\tilde{c}}$ is a point-optimal test for testing H_0 in the limiting problem against the point alternative given by the probability measures $\nu_{\overline{m}}, \nu_{\delta|\tau}$, indexed by $\bar{c}$ and $\tilde{c}$, respectively, and ν_τ.

The parameters $\bar{c}, \tilde{c}$ measure the importance of the full-sample versus the stability restrictions. If we put zero weight on instability, $\tilde{c} = 0$, the resulting test $LR^{\bar{c},0}$ is equivalent to a test that rejects for large values of $\overline{X}' V_X^{-1} \overline{X}$. The finite-sample analogue of this statistic is the *GMM-AR*$_T$ statistic. Therefore, the *GMM-AR*$_T$ test is asymptotically efficient under Assumption 1 only when there is no instability under the alternative, in accordance with Proposition 1.

For $\tilde{c} > 0$, the optimal test generally depends on $\bar{c}$ and $\tilde{c}$. By setting $\bar{c} = \tilde{c} = c$ we put equal weights on the two alternatives, and the finite-sample analogue of the $LR^{c,c}$ statistic can be written as

$$LR_T^{c,c}(\theta_0) = (1+c)^{-k} \int_\varsigma \exp\left\{\frac{1}{2}\frac{c}{1+c} S_T(\theta_0,\tau)\right\} d\nu_\tau$$

where

$$S_T(\theta,\tau) = \frac{1}{T}\begin{pmatrix} F_{\tau T}(\theta) \\ F_T(\theta) - F_{\tau T}(\theta)\end{pmatrix}' \begin{pmatrix} \frac{\hat{V}^1_{ff}(\theta;\tau)^{-1}}{\tau} & 0 \\ 0 & \frac{\hat{V}^2_{ff}(\theta;\tau)^{-1}}{1-\tau}\end{pmatrix} \begin{pmatrix} F_{\tau T}(\theta) \\ F_T(\theta) - F_{\tau T}(\theta)\end{pmatrix}$$

is the continuously updated version of the "partial-sample" GMM objective function of Andrews (1993). For the estimators $\hat{V}^1_{ff}(\theta;\tau)$, $\hat{V}^2_{ff}(\theta;\tau)$ we can use either respective partial-sample estimators, or a full-sample estimator $\hat{V}_{ff}(\theta)$, see Andrews (1993). $S_T(\theta_0;\tau)$ can be thought of as a 'split-sample' *GMM-AR*$_T$ statistic that arises when we split the sample at date $[\tau T]$ and use the resulting $2k$ moment conditions $\left[F_{\tau T}(\theta_0)', F_T(\theta_0)' - F_{\tau T}(\theta_0)'\right]'$.

The split-sample statistic $S_T(\theta_0,\tau)$ can be decomposed orthogonally into the full-sample *GMM-AR*$_T$ statistic and the statistic

$$S_T^B(\theta_0,\tau) = S_T(\theta_0) - GMM\text{-}AR_T(\theta_0)$$

that depends primarily on $F_{sT}(\theta_0) - sF_T(\theta_0)$, and therefore, has power only against instability. When $\bar{c} > 0$, the joint LR test can be based equivalently on the statistic

$$\textit{break-}AR_T^{\bar{c},\tilde{c}}(\theta_0) = \frac{\bar{c}}{1+\bar{c}} GMM\textit{-}AR_T(\theta_0) + 2\log \int_{\varsigma} \exp\left\{ \frac{1}{2} \frac{\tilde{c}}{1+\tilde{c}} S_T^B(\theta_0, \tau) \right\} d\nu_\tau. \qquad (19)$$

Setting $\bar{c} = 0$, we obtain tests of the stability restrictions.

Asymptotic critical values for the above statistics can be obtained from the following Lemma, which follows from the continuous mapping theorem.

Lemma 1 *Under Assumption 1 and* $H_0 : \theta = \theta_0$,

$$\textit{break-}AR_T^{\bar{c},\tilde{c}}(\theta_0) \Rightarrow \frac{\bar{c}}{1+\bar{c}} \overline{\psi}_k + 2\log \int_{\varsigma} \exp\left\{ \frac{1}{2} \frac{\tilde{c}}{1+\tilde{c}} \widetilde{\psi}_k(\tau) \right\} d\nu_\tau$$

where $\widetilde{\psi}_k(\tau) = \frac{\widetilde{W}(\tau)'\widetilde{W}(\tau)}{\tau(1-\tau)}$, *with* $\widetilde{W}(\tau)$ *a standard* $k \times 1$ *Brownian Bridge process, and* $\overline{\psi}_k$ *a* $\chi^2(k)$ *distributed random variable independent of* $\widetilde{W}(\tau)$.

3.2 Persistent time variation

Suppose the probability measure over $m(s)$, ν_m is given by $m(s) - \overline{m} = \Omega^{1/2} W_{\widetilde{m}}(s)$, where $W_{\widetilde{m}}(\cdot)$ is a standard $k \times 1$ Wiener process, which is independent of $W(\cdot)$ in Assumption 1. This assumption incorporates the drifting parameter approach to modeling instability that was followed by Stock and Watson (1996, 1998) and Elliott and Mueller (2006).

Derivation of the optimal test in this problem is facilitated by looking at a particular member of the class of data generating processes that satisfy Assumptions 1 and 2, for which the WAP-maximizing test can be derived analytically. For this purpose, we use the Gaussian multivariate local level model, following Elliott and Mueller (2006). The theory in Mueller (2008) can then be invoked to show that the resulting test will be asymptotically efficient in a wider sense.

Specifically, consider the model $y_t = \mu_t + u_t$, for $t = 1, ..., T$, where $y_t, \mu_t, u_t \in \Re^k$. Assume $u_t \sim iidN(0, \Sigma)$ for some positive definite matrix Σ, such that $u \sim N(0, I_T \otimes \Sigma)$, where $u = (u_1', ..., u_T')' \in \Re^{Tk}$. The density of $y = (y_1', ..., y_T')'$ conditional on $\mu = (\mu_1', ..., \mu_T')'$

is given by

$$f\left(y|\mu\right)=\left(2\pi\right)^{-Tk/2}\left|\Sigma\right|^{-T/2}\exp\left\{-\frac{1}{2}\left(y-\mu\right)'\left(I_T\otimes\Sigma^{-1}\right)\left(y-\mu\right)\right\}.$$

We want to test the hypothesis $H_0:\mu=0$, against the alternative of persistent time variation $H_1:\mu_t=\mu_{t-1}+\Delta\mu_t$, where $T\Delta\mu_t\sim iidN\left(0,c\Sigma\right)$. H_0 can be decomposed into $H_0^1:\bar{\mu}=0$, where $\bar{\mu}=T^{-1}\sum_{t=1}^{T}\mu_t$, and $H_0^2:\mu_t-\bar{\mu}=0$ for all t, or, equivalently, $H_0^2:\widetilde{\mu}=0$, where $\widetilde{\mu}=\left(B_e'\otimes I_k\right)\mu$, B_e is a $T\times(T-1)$ matrix such that $B_e'e=0$, $B_e'B_e=I_{T-1}$, $B_eB_e'=M_e$, $M_e=I_T-e\left(e'e\right)^{-1}e'$ and e is the $T\times 1$ vector of ones.

Conditional on μ, the ratio $f\left(y|\mu\right)/f\left(y|0\right)$ is given by

$$\xi\left(\mu\right)=\exp\left\{y'\left(I_T\otimes\Sigma^{-1}\right)\mu-\frac{1}{2}\mu'\left(I_T\otimes\Sigma^{-1}\right)\mu\right\},$$

and the likelihood ratio is given by $\int\xi\left(\mu\right)d\nu_\mu$, where $d\nu_\mu$ is the density of μ. As in the case of a single break, we specify independent weights over $\bar{\mu}$ and $\tilde{\mu}$, with densities given by $\sqrt{T}\bar{\mu}\sim N\left(0,\bar{c}\Sigma\right)$ and $T\tilde{\mu}\sim N\left(0,\left[B_e'FF'B_e\otimes\tilde{c}\Sigma\right]\right)$, where $F=\left[\mathrm{f}_{i,j}\right]$ is a $T\times T$ lower triangular matrix of ones, i.e, $\mathrm{f}_{i,j}=1$ for all $i\leq j$ and 0 otherwise. The resulting likelihood ratio statistic depends on $\bar{c}$ and $\widetilde{c}$, and can be written as the product of the statistics

$$\overline{LR}_T^{\bar{c}}=\left(1+\bar{c}\right)^{-k/2}\exp\left\{\frac{1}{2}\frac{\bar{c}}{1+\bar{c}}T\bar{y}'\Sigma^{-1}\bar{y}\right\}$$

and

$$\widetilde{LR}_T^{\widetilde{c}}=\left(\frac{1-r_{\widetilde{c}}^{2T}}{T\left(1-r_{\widetilde{c}}^2\right)r_{\widetilde{c}}^{T-1}}\right)^{-k/2}\exp\left\{\frac{1}{2}\sum_{i=1}^{k}v_i'\left(M_e-G_{\widetilde{c}}\right)v_i\right\}\tag{20}$$

where $\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_t$, $v_i=\left(I_T\otimes\iota_{k,i}\Sigma^{-1/2}\right)y$, $i=1,...,k$, $i_{k,i}$ is the $k\times 1$ unit vector with one at position i, and $G_c=H_c^{-1}-H_c^{-1}e\left(e'H_c^{-1}e\right)^{-1}e'H_c^{-1}$, with $H_c=r_c^{-1}FA_cA_c'F'$, A_c is a $T\times T$ matrix with ones in the main diagonal, $-r_c$ in its subdiagonal, and zeros otherwise, i.e., $A_c=\left[\mathrm{a}_{i,j}\right]$ where $\mathrm{a}_{i,j}=1$ if $i=j$, $-r_c$ if $i=j+1$, and 0 otherwise, and $r_c=\frac{1}{2}\left(2+c^2T^{-2}-T^{-1}\sqrt{4c^2+c^4T^{-2}}\right)=1-cT^{-1}+o\left(T^{-1}\right)$. The derivation of $\widetilde{LR}_T^{\widetilde{c}}$ follows the same calculations as in the proof of Elliott and Mueller (2006, Lemma 1).

Taking logs, multiplying by 2 and dropping the constants, the joint log-likelihood ratio statistic can be written as

$$LR_T^{\bar{c},\tilde{c}} = \frac{\bar{c}}{1+\bar{c}} T\overline{y}'\Sigma^{-1}\overline{y} + \sum_{i=1}^{k} v_i' \left(M_e - G_{\tilde{c}}\right) v_i.$$

The parameters $\bar{c}$ and $\tilde{c}$ govern the weight given to deviations from H_0 in the direction $\bar{\mu} \neq 0$ and $\tilde{\mu} \neq 0$, respectively. $LR_T^{c,0}$ coincides with the usual Wald statistic for $\mu = 0$, which is independent of c, while $LR_T^{0,c}$ is a pure stability test, conditional on $\bar{\mu} = 0$, (apart from the sign, the main difference from the qLL statistic of Elliott and Mueller (2006) is that the latter uses demeaned y). $LR_T^{c,c}$ gives equal weight to the two types of departure from H_0.

The asymptotic distribution of $LR_T^{\bar{c},\tilde{c}}$ is given by the following Lemma.

Lemma 2 *(i)* $\sum_{i=1}^{k} v_i' \left(M_e - G_c\right) v_i \Rightarrow \psi_c$, *where*

$$\psi_c = \sum_{i=1}^{k} \left[cJ_i(1)^2 + c^2 \int J_i^2 + \frac{2c}{1-e^{-2c}} \left[e^{-c} J_i(1) + c \int e^{-cs} J_i \right]^2 - \left[J_i(1) + c \int J_i \right]^2 \right]$$

and $J_i(s)$ *is the* i*th element of the* k*-dimensional Ornstein-Uhlenbeck process* $J(s) = W(s) - c\int e^{-c(s-r)} W(r)\, dr$, *and* W *is a* $k \times 1$ *standard Wiener process.*
(ii) $LR_T^{\bar{c},\tilde{c}} \Rightarrow \frac{\bar{c}}{1+\bar{c}} \overline{\psi}_k + \psi_{\tilde{c}}$, *where* $\overline{\psi}_k \sim \chi^2(k)$ *independent of* ψ_c.

Part (i) of this result follows from Lemma 6 of Elliott and Mueller (2006). Part (ii) follows from the asymptotic independence between $\sqrt{T}\overline{y}$ and $\left[I_k \otimes \left(M_e - G_c\right)\right] v$, where $v = \left(v_1', \ldots, v_k'\right)'$, and the continuous mapping theorem.

The resulting test of $H_0 : E\left[f_t(\theta_0)\right] = 0$ against a persistent time-varying alternative is obtained by replacing y_t by $f_t(\theta_0)$, and Σ by $\hat{V}_{ff}(\theta_0)$. We shall denote the resulting statistic the qLL-AR_T statistic, and index it by the weights $\bar{c}, \tilde{c}$.

$$qLL\text{-}AR_T^{\bar{c},\tilde{c}}(\theta_0) = \frac{\bar{c}}{1+\bar{c}} GMM\text{-}AR_T(\theta_0) + \sum_{i=1}^{k} v_i' \left(M_e - G_{\tilde{c}}\right) v_i, \tag{21}$$
$$v_i' = \left[f_1(\theta_0)' V_T(\theta_0)^{-1/2} \iota_{k,i}, \ldots, f_T(\theta_0)' V_T(\theta_0)^{-1/2} \iota_{k,i} \right]$$

Theorem 1 *Under Assumption 1 and* $H_0 : \theta = \theta_0$, qLL-$AR_T^{\bar{c},\tilde{c}}(\theta_0) \Rightarrow \frac{\bar{c}}{1+\bar{c}} \overline{\psi}_k + \psi_{\tilde{c}}$, *where* $\overline{\psi}_k$ *and* $\psi_{\tilde{c}}$ *are given in Lemma 2. Moreover, a test that rejects for large values of the qLL-*$AR_T^{\bar{c},\tilde{c}}(\theta_0)$ *statistic is asymptotically efficient against persistent time-variation of the moment* *conditions* $E[f_t(\theta_0)]$.

3.3 Tests based on estimating break dates

We now examine the alternative procedures that are based on estimates of the break date. It is instructive to consider first the case of the linear IV model with time-varying first-stage, see Example 3 in Section 2 above.

3.3.1 Finite-sample analysis for a special case

The model can be written compactly as

$$\begin{aligned} Y_t &= Z_t \Pi_t A' + V_t, \quad t = 1, ..., T \quad \text{where} \\ Y_t &= [y_{1,t} \vdots Y_{2,t}], \quad A' = \left(\theta \vdots I_p\right) \text{ and} \\ V_t &= [v_{1,t} \vdots V_{2,t}] \sim N(0, \Omega). \end{aligned} \tag{22}$$

This is a generalization of the canonical constant-parameter linear IV model studied by Andrews, Moreira, and Stock (2006).

Consider the assumption of a single break in Π_t occurring at time $[\tau T]$ (the analysis generalizes easily to multiple breaks). Define $Z(\tau) \in \Re^{T \times 2k}$ by

$$Z(\tau) = \begin{pmatrix} \{Z_t\}_{t=1}^{[\tau T]} & 0 \\ 0 & \{Z_t\}_{t=[\tau T]+1}^{T} \end{pmatrix},$$

stack Y_t, V_t into $Y, V \in \Re^{T \times (1+p)}$, respectively, and define $\Pi \in \Re^{2k \times p}$ by $\Pi = [\Pi_1' \vdots \Pi_2']'$. Then, the model (22) can be written as

$$Y = Z(\tau) \Pi A' + V. \tag{23}$$

When Ω is known, the log-likelihood function is given by (up to a constant)

$$L(\theta,\Pi,\tau) = -\frac{1}{2}tr\left[\Omega^{-1}\left(Y - Z(\tau)\Pi A'\right)'\left(Y - Z(\tau)\Pi A'\right)\right]$$
$$= tr\left[\Omega^{-1}A\Pi' Z(\tau)' Y\right] - \frac{1}{2}tr\left[\Omega^{-1}A\Pi' Z(\tau)' Z(\tau)\Pi A'\right]. \quad (24)$$

Since $\{Z_t\}$ is non-random, the likelihood depends on the data only through the process $Z(s)' Y$, $s \in [0,1]$, which is the sufficient statistic. This $2k \times (p+1)$ process can be decomposed orthogonally into the processes:

$$\overline{F}(s) = \left[Z(s)' Z(s)\right]^{-1/2} Z(s)' Y b_0 \left(b_0' \Omega b_0\right)^{-1/2},$$
$$\overline{D}(s) = \left[Z(s)' Z(s)\right]^{-1/2} Z(s)' Y \Omega^{-1} A_0 \left(A_0' \Omega^{-1} A_0\right)^{-1/2},$$
$$\text{where } b_0' = \left(1, -\theta_0'\right), \quad A_0' = \left(\theta_0 \vdots I_p\right),$$

Also, define the following parameter:

$$\mu_{\Pi,\tau}(s) = \left[Z(s)' Z(s)\right]^{-1/2} Z(s)' Z(\tau)\Pi \in \Re^{2k\times p}$$

and

$$c_\theta = (\theta - \theta_0)\left(b_0' \Omega b_0\right)^{-1/2} \in \Re$$
$$d_\theta = A' \Omega^{-1} A_0 \left(A_0' \Omega^{-1} A_0\right)^{-1/2} \in \Re^p.$$

Then, the following result arises as a straightforward extension of Andrews, Moreira, and Stock (2006, Lemma 2).

Lemma 3 *For the model given by equation (23):*

1. $\overline{F}(s)$ *is a Gaussian process with mean* $\mu_{\Pi,\tau}(s)\, c_\theta$, *and covariance kernel* $K(s_1, s_2) = \left[Z(s_1)' Z(s_1)\right]^{-1/2} Z(s_1)' Z(s_2)\left[Z(s_2)' Z(s_2)\right]^{-1/2}$

2. $\overline{D}(s)$*is a Gaussian process with mean* $\mu_{\pi,\tau}(s)\, d_\theta$, *and covariance kernel* $K(s_1, s_2)$, *same as for* $\overline{F}(s)$

3. $\overline{F}(s_1)$ *and* $\overline{D}(s_2)$ *are independent for all* s_1, s_2.

We can now write the log-likelihood function (24) in terms of the statistics $\overline{F}(s)$ and $\overline{D}(s)$

$$L(\theta,\Pi,\tau) = \overline{F}(\tau)' \mu_{\Pi,\tau}(\tau) c_\theta - \frac{1}{2} tr\left[c_\theta^2 \mu_{\Pi,\tau}(\tau)' \mu_{\Pi,\tau}(\tau)\right] + tr\left[\overline{D}(\tau)' \mu_{\Pi,\tau}(\tau) d_\theta\right] - \frac{1}{2} tr\left[d_\theta' \mu_{\Pi,\tau}(\tau)' \mu_{\Pi,\tau}(\tau) d_\theta\right].$$

Under H_0, $c_\theta = 0$, and

$$L(\theta_0,\Pi,\tau) = tr\left[\overline{D}(\tau)' \mu_{\Pi,\tau}(\tau) d_{\theta_0}\right] - \frac{1}{2} tr\left[d_{\theta_0}' \mu_{\Pi,\tau}(\tau)' \mu_{\Pi,\tau}(\tau) d_{\theta_0}\right]. \tag{25}$$

In other words, the process $\overline{D}(\cdot)$ is sufficient for Π and τ (or $\overline{F}(\cdot)$ is specific ancillary for Π, τ under H_0). Thus, the restricted maximum likelihood estimator (MLE) of Π, τ given $\theta = \theta_0$ can be obtained by minimizing (25) wrt Π and τ. Concentrating (25) with respect to Π, we obtain $L^c(\theta_0,\tau) = \frac{1}{2} tr\left[\overline{D}(\tau)' \overline{D}(\tau)\right]$ and therefore, the MLE for τ is

$$\widetilde{\tau} = \arg\max_\tau tr\left[\overline{D}(\tau)' \overline{D}(\tau)\right]. \tag{26}$$

We can obtain similar tests of $H_0 : \theta = \theta_0$ either based on pivotal statistics, or based on non-pivotal statistics by conditioning. The split-sample AR, KLM and $JKLM$ statistics are given by

$$AR(s) = \overline{F}(s)' \overline{F}(s), \quad LM(s) = \overline{F}(s)' \overline{D}(s) \left[\overline{D}(s)' \overline{D}(s)\right]^{-1} \overline{D}(s)' \overline{F}(s) \quad \text{and} \quad J(s) = AR(s) - LM(s)$$

and the LR test is analytically available only in the case $p = 1$, see Moreira (2003), in which case it can be written as

$$MLR(s) = \frac{1}{2}\left[AR(s) - rk(s) + \sqrt{\left[AR(s) + rk(s)\right]^2 - 4J(s)\, rk(s)}\right] \tag{27}$$

where $rk(s) = \overline{D}(s)' \overline{D}(s)$. For $p > 1$, we will use the generalization of the LR statistic derived by Kleibergen (2005), which is given by equation (27) with $rk(s)$ being a statistic that tests that the rank of the matrix Π is $p-1$ under H_0, and which is only a function of $\overline{D}(s)$.

Since $\overline{F}(\cdot)$ is orthogonal to $\overline{D}(\cdot)$, and $\widetilde{\tau}$ is only a function of $\overline{D}(\cdot)$, we have the following result.

Theorem 2 *Let $\widetilde{\tau} = \arg\max_s \overline{D}(s)' \overline{D}(s)$. Then,*

1. *$LM(\widetilde{\tau})$ is distributed as $\chi^2(p)$ under H_0.*

2. *$J(\widetilde{\tau}) = AR(\widetilde{\tau}) - LM(\widetilde{\tau})$ is distributed as $\chi^2(2k-p)$ under H_0.*

3. *$LM(\widetilde{\tau})$ and $J(\widetilde{\tau})$ are independent under H_0.*

4. *The distribution of $LR(\widetilde{\tau})$ conditional on $rk(\widetilde{\tau})$ is the same as the conditional distribution of $\frac{1}{2}\left[\psi_p + \psi_{2k-p} - rk(\widetilde{\tau}) + \sqrt{\left[\psi_p + \psi_{2k-p} - rk(\widetilde{\tau})\right]^2 + 4\psi_{2k-p} rk(\widetilde{\tau})}\right]$, where ψ_p, ψ_{2k-p} are independent random variables distributed as $\chi^2(p)$ and $\chi^2(2k-p)$, respectively.*

3.3.2 Asymptotic analysis for the general case

The previous analysis extends to any data generating process that satisfies Assumption 3. First, notice that under Assumption 3 and $H_0 : \theta = \theta_0$, the entire partial sample moments $F_{\cdot T}(\theta_0)$ are asymptotically ancillary for τ, since their asymptotic distribution does not depend on it.

Next, define the following estimator of the Jacobian of the split-sample moments

$$\overline{D}_T(\theta_0, \tau) = \begin{pmatrix} D_T^1(\theta_0, \tau) & 0 \\ 0 & D_T^2(\theta_0, \tau) \end{pmatrix}$$

and its variance

$$\hat{V}_{\overline{D}}(\theta_0, \tau) = \begin{pmatrix} \tau \hat{V}_{qq.f}^1(\theta_0, \tau) & 0 \\ 0 & (1-\tau)\hat{V}_{qq.f}^1(\theta_0, \tau) \end{pmatrix}$$

where $D_T^i(\theta_0, \tau)$ and $\hat{V}_{qq.f}^1(\theta_0, \tau)$ are defined in Equations (11) and (12) above. The matrix $\overline{D}_T(\theta_0, \tau)$ is the split-sample analogue of the matrix $D_T(\theta_0)$ defined in Kleibergen (2005, Equation 16).

Consider the following estimator of $\widetilde{\tau}$

$$\widetilde{\tau}\left(\theta_0\right) = \arg\max_{\tau \in \varsigma} vec\left[\overline{D}_T\left(\theta_0, \tau\right)\right]' \hat{V}_{\overline{D}}\left(\theta_0, \tau\right)^{-1} vec\left[\overline{D}_T\left(\theta_0, \tau\right)\right].$$

This is a generalization of the estimator given by Equation (26) for the linear IV model. Under Assumption 3 and $H_0 : \theta = \theta_0$, $\overline{D}_T\left(\theta_0, \cdot\right)$ is asymptotically independent of $F_{\cdot T}\left(\theta_0\right)$, and hence, so is $\widetilde{\tau}\left(\theta_0\right)$. Therefore, we obtain the following result.

Theorem 3 *When Assumptions 3 and $H_0 : \theta = \theta_0$ hold, the limiting distributions of the split-sample GMM-AR, KLM, JKLM and MLR statistics are given by*

$$\begin{aligned}
split\text{-}AR_T\left(\theta_0, \widetilde{\tau}(\theta_0)\right) &\overset{d}{\to} \psi_p + \psi_{2k-p} \\
split\text{-}KLM_T\left(\theta_0, \widetilde{\tau}(\theta_0)\right) &\overset{d}{\to} \psi_p \\
split\text{-}JKLM_T\left(\theta_0, \widetilde{\tau}(\theta_0)\right) &\overset{d}{\to} \psi_{2k-p} \\
split\text{-}MLR_T\left(\theta_0, \widetilde{\tau}(\theta_0)\right) | rk_\theta\left(\theta_0\right), \widetilde{\tau}(\theta_0) &\overset{d}{\to} \tfrac{1}{2}\left[\psi_p + \psi_{2k-p} - rk_\theta\left(\theta_0, \widetilde{\tau}(\theta_0)\right) + \right. \\
&\left. \sqrt{\left(\psi_p + \psi_{2k-p} + rk_\theta\left(\theta_0, \widetilde{\tau}(\theta_0)\right)\right)^2 - 4\psi_{2k-p} rk_\theta\left(\theta_0, \widetilde{\tau}(\theta_0)\right)}\right],
\end{aligned}$$

where ψ_p and ψ_{2k-p} are independently distributed $\chi^2(p)$ and $\chi^2(2k-p)$ random variables.

4 Numerical results

4.1 Asymptotic power comparisons

We compare the methods derived in the previous section in terms of asymptotic power. We first compare power curves derived from the point-optimal tests in equations (19) and (21), with the power curves of generalized Anderson-Rubin tests derived in Section 3. Next we consider the linear instrumental variable with one break in the reduced form equation. Following closely the numerical analysis in Andrews, Moreira, and Stock (2006), the linear IV model provides useful benchmarks, e.g., the power of the 'oracle' test when break dates are known. In both experiments we set the sample size to 2000 observations, and the number of Monte Carlo simulations to 20,000.

4.1.1 Point-Optimal Test

The power curves are based on testing the null hypothesis $H_0 : \theta - \theta_0 = 0$ against the alternative $H_1 : \theta - \theta_0 \neq 0$, where θ is value of the structural parameter of interest, and θ_0 is the value of θ which is being tested. We set the parameters $\bar{c} = (\theta - \theta_0)^2\omega$ and $\tilde{c} = (\theta - \theta_0)^2(1 - \omega)$ where $\omega \in [0, 1]$ is a weighting parameter controlling if the information comes from the stability restriction ($\omega = 0$) or from the full-sample restriction ($\omega = 1$). We use significance level of 5%, and the number of instruments $k = 1$.

The left and right columns of Figure 1 illustrate, respectively, the cases of a single break at unknown point and of the persistent time variation (PTV). When the information is coming only from the stability restriction ($\omega = 0$), as expected, the *exp-*AR_T and the *qLL-*AR_T power curves are the closest to the point optimal test curves in the single break and in the PTV cases, respectively. As the information switches from the stability to the full sample restrictions, the *ave-*AR_T power dominates the *exp-*AR_T and *qLL-*AR_T. We also notice that the generalized AR tests power dominates the *GMM-*AR_T when there is instability, while the opposite happens when there is no instability ($\omega = 1$). However, in the latter case, the loss of power of the proposed tests relatively to the optimal tests is small. Analogous results (not reported here) are obtained when the number of instruments increases.

4.1.2 Linear Instrumental Variable Model

We use the linear IV model, given by equation (22) above, with a single endogenous regressor, $p = 1$, and $\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

It is well-known that in the constant-parameter IV regression model, the amount of information in the data about the structural parameters (or the quality of instruments) can be characterized using a unitless measure known as the 'concentration parameter', which is $\lambda = \sum \Pi' Z_t' Z_t \Pi$. We can think of the contribution of each observation to the identification of θ as being equal to $\Pi' Z_t' Z_t \Pi$, but when Π_t is time-varying, the incremental information is $\Pi_t' Z_t' Z_t \Pi_t$, and so the total amount of information is $\lambda = \sum_{t=1}^{T} \Pi_t' Z_t' Z_t \Pi_t$.

We consider the case of a single break in Π_t. Because all of the statistics we propose are invariant to the class of transformation $Z_t \to Z_t G$ for any nonsingular matrix G, we

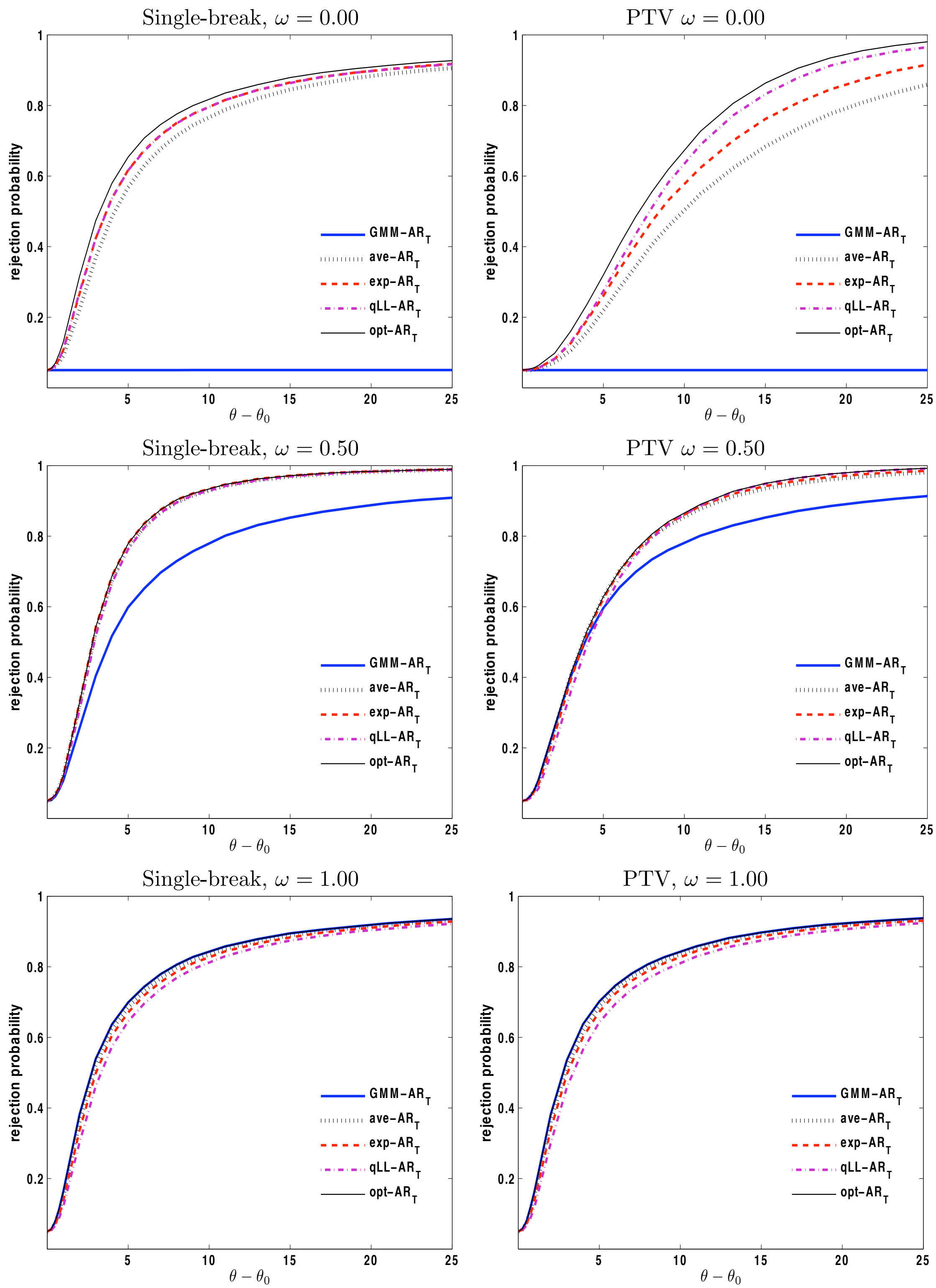

Figure 1: Asymptotic power curves of point-optimal and generalized AR - tests. $k = 1$, 20,000 Monte Carlo replications.

can set all but the first entry of Π_t to zero, without loss of generality. Let the first entry of Π_t and Z_t be π_t and z_t, respectively, with $\pi_t = \pi_1$ for $t < [\tau T]$ and π_2 for $t \geq [\tau T]$. Then, $\lambda = \sum_{t=1}^{T} \pi_t^2 z_t^2 = \pi_1^2 \sum_{t=1}^{[\tau T]} z_t^2 + \pi_2^2 \sum_{t=[\tau T]+1}^{T} z_t^2 \approx \left[\pi_1^2 \tau + \pi_1^2 (1-\tau)\right] Q_{zz}$, where $Q_{zz} = \sum_{t=1}^{T} z_t^2$. We can measure the information in the full-sample and the stability restrictions by $\lambda_F = \left[\pi_1 \tau + \pi_2 (1-\tau)\right]^2 Q_{zz}$ and $\lambda_S = (\pi_1 - \pi_2)^2 \tau (1-\tau) Q_{zz}$, respectively. In all the experiments, we fix $\lambda = 5$, to match the results of Andrews, Moreira, and Stock (2006).

Figure 2 presents the asymptotic power curves of the *GMM-MLR*$_T$, *ave-AR*$_T$ and *ave-AR*$_T^B$ tests when the distribution of information between full-sample and stability restrictions, λ_F and λ_S, respectively, is varied. We set the number of instruments k equal to 2. The appendix contains the case when the number of instruments is 5. Since the *GMM-MLR*$_T$ test is not invariant to ρ, we report results for two cases $\rho = 0.20$ and 0.95.

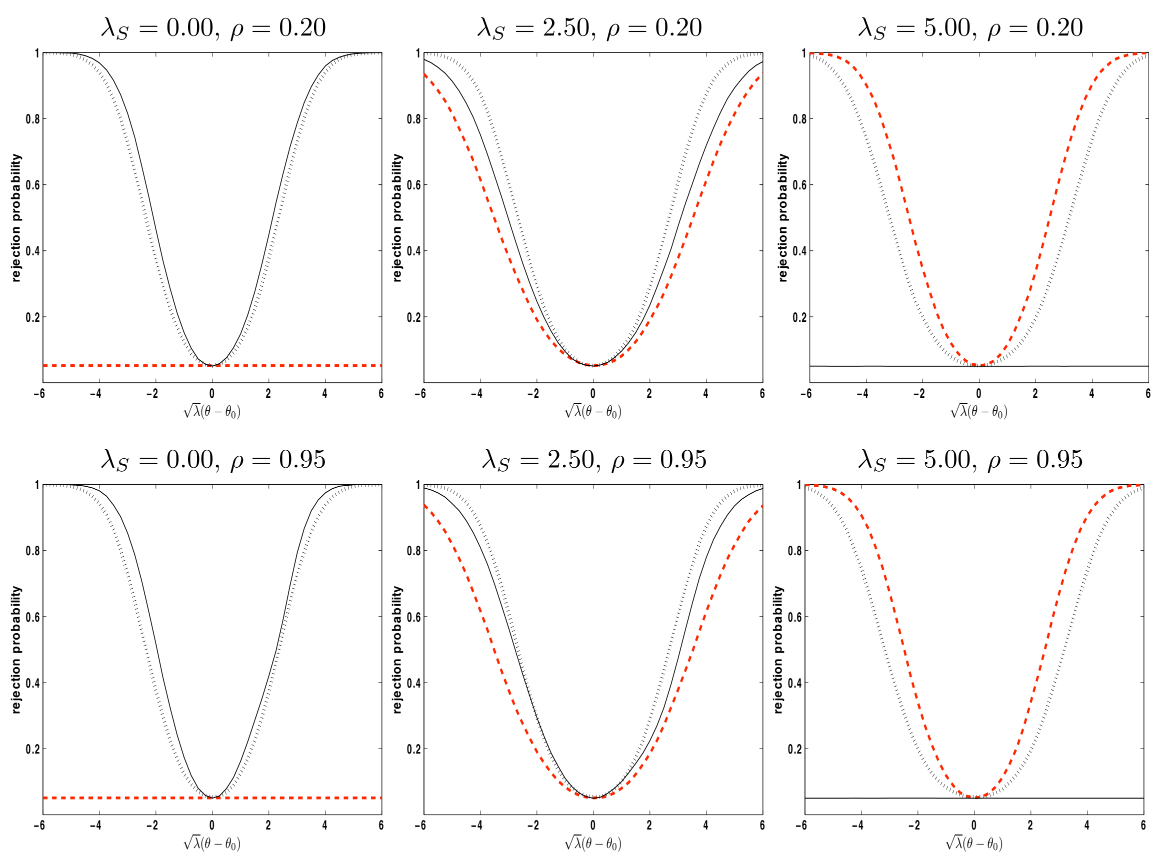

Figure 2: Power function of tests of the null hypothesis $H_0 : \theta = \theta_0$ using 5% significance level computed using 20,000 Monte Carlo replications; one deterministic break at $\tau = 0.5$; $k = 2$. The curves correspond to: *GMM-MLR*$_T$ (thin solid line); *ave-AR*$_T^B$ (dashed line); *ave-AR*$_T$ (dotted line).

The $GMM\text{-}MLR_T$ test dominates the other tests when there is no instability, $\lambda_S = 0.00$, since it is optimal, see Andrews, Moreira, and Stock (2006). In the case of $\lambda_S = 5$, the best test is the pure stability test $ave\text{-}AR_T^B$. We observe that the $ave\text{-}AR_T$, has good power in all cases, and its power is relatively insensitive to the source of information.

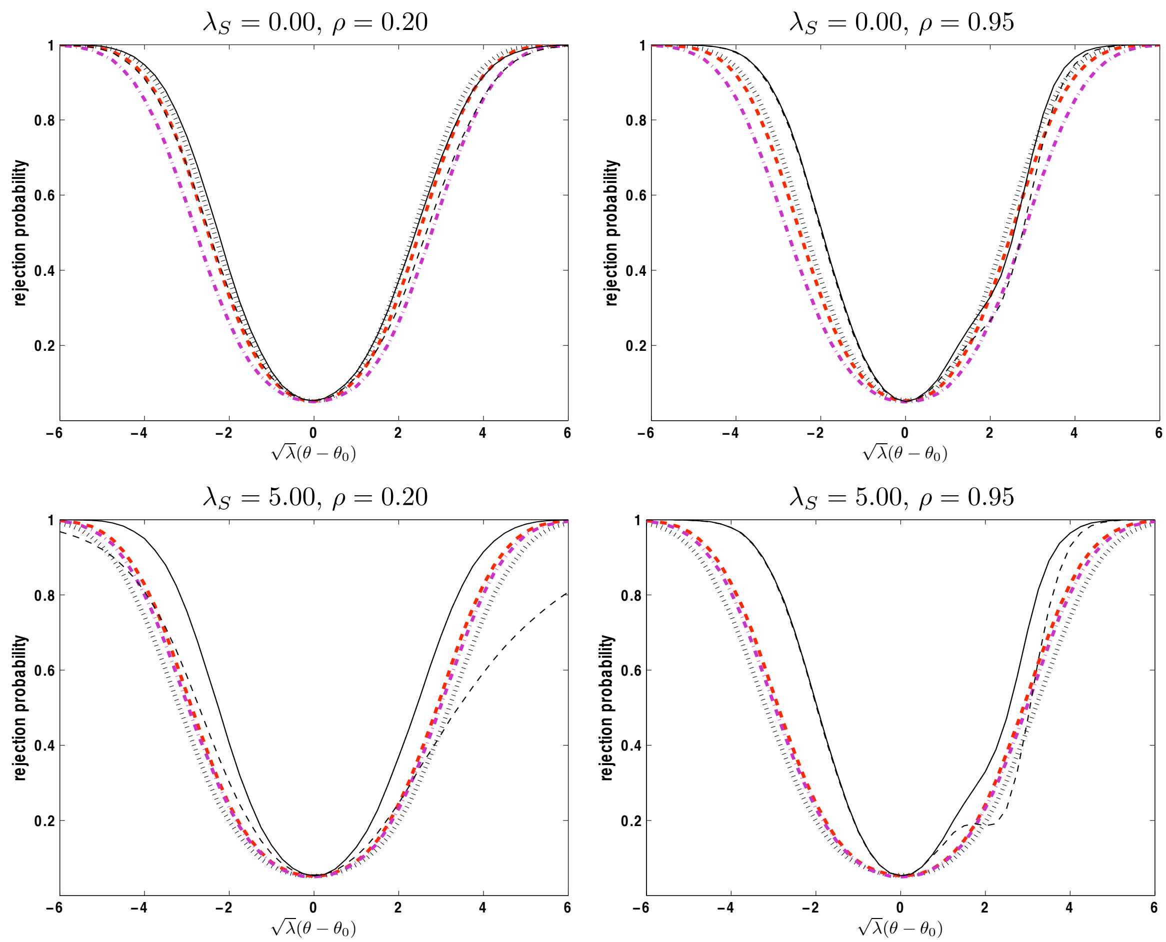

Figure 3: Power function of tests of the null hypothesis $H_0 : \theta = \theta_0$ using 5% significance level computed using 20,000 Monte Carlo replications; one deterministic break at $\tau = 0.5$; $k = 2$. The curves correspond to: MLR_T-"oracle" (thin solid line); $split\text{-}MLR_T$ (thin dashed line); $exp\text{-}AR_T$ (dashed line); $ave\text{-}AR_T$ (dotted line); and $qLL\text{-}AR_T$ (dash-dot line).

Next, we compare the power of the generalized AR tests with the split-sample tests based on estimating the break date. Figure 3 compares the power curves of the *exp*-, *ave*-, and $qLL\text{-}AR_T$ tests with the $split\text{-}MLR_T$ and the MLR_T tests in the two polar cases $\lambda_S = 0$ and $\lambda_S = 5$. The MLR_T test is computed assuming that that the break date is known. We see that no test dominates the others in terms of power. This is also true for the split-sample

AR and KLM statistics (see results reported in the appendix). Interestingly, we notice that the *ave*-AR_T and *exp*-AR_T dominate the remaining generalized AR tests when $\lambda_S = 0.00$ and $\lambda_S = 5.00$, respectively.

4.2 Size in finite samples

We study the finite-sample rejection frequencies of the proposed tests using a simulation experiment based on the structural model given by equation (5) in Example 1 in Section 2 above. This is a prototypical example of forward-looking model that is commonly used in macroeconomics and finance. We calibrate our simulations to a leading macroeconomic application, the new Keynesian Phillips curve, where y_t denotes inflation (π_t), and x_t is the labor share, see Galí and Gertler (1999). We assume that x_t follows $x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t$. We assume the shocks v_t and ε_t are jointly Normal with zero mean, variances σ_ε^2, σ_v^2 and covariance $\sigma_{\varepsilon v}$.

In this simple version of new Keynesian model, the parameter β is a discount factor, while, $\lambda = \frac{(1-\alpha)(1-\beta\alpha)}{\alpha}$, where α represents the price rigidity in the economy. The parameters are set to $\beta = 1$ and $\alpha = \frac{2}{3}$, ($\lambda = \frac{1}{6}$), while the remaining nuisance parameters are calibrated to post-1960 quarterly US data. We find $\rho_1 = 0.9$, $\rho_2 = 0.063$, $\sigma_\varepsilon^2 = .3$, $\sigma_v^2 = .011$ and $\sigma_{\varepsilon v} = .007$. Several authors have argued that there was a structural change in the US economy around 1984. Estimating the reduced form parameters over the two subsamples, we find that the first-order autocorrelation $\rho = \rho_1 / (1 + \rho_2)$ is constant, but ρ_2 goes from $-.09$ to .21, with a standard error of 0.15. We therefore set $\rho_{2,t} = 0.063 + \kappa\,(-1)^{1_{\{t<1984q1\}}}\,0.15$ and $\rho_{1,t} = \rho\,(1 + \rho_{2,t})$, with $\rho = 0.9$. The parameter κ is used to vary the magnitude of the change in the coefficients in terms of standard errors from zero, with $\kappa = 2$ corresponding to the subsamples estimates of ρ_2.

There is also evidence of a break in the variance of the shocks over that period, e.g., McConnell and Perez-Quiros (2000), a phenomenon known as the 'great moderation'. Indeed, we find that σ_ε^2 falls significantly after 1984, although σ_v^2 is constant over the two periods. It is important to check the implications of a change in the variance since permanent changes in the variance are not covered by Assumptions 1 and 3. Thus, large changes in the variance may lead to size distortion of our tests in finite samples. To examine this issue, we vary the change in the variance by ϕ multiples of its standard error and report the size of the tests for various values of ϕ.

Table 1 reports the null rejection frequencies of the proposed tests of $H_0 : \alpha = \frac{2}{3}$, in the model $\pi_t = E_t(\pi_{t+1}) + \frac{(1-\alpha)^2}{\alpha} x_t + u_t$ for a sample of $T = 180$, computed using 20,000 replications. The instruments used are x_{t-1} and x_{t-2}, and the variance estimator used is Newey and West (1987) with prewhitening. We consider six cases: for the magnitude of the parameter break we consider $\kappa = 0, 2$ and 4, and for magnitude of the change in the variance we consider $\phi = 0$ and 4.

	$\kappa = 0$				$\kappa = 2$				$\kappa = 4$			
	$\phi = 0$		$\phi = 4$		$\phi = 0$		$\phi = 4$		$\phi = 0$		$\phi = 4$	
Nom.level	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
GMM-AR$_T$	7.46	3.87	8.46	4.33	8.03	4.20	9.38	4.90	8.92	4.90	10.39	5.65
ave-AR$_T$	6.52	3.42	8.52	4.41	7.20	3.86	9.38	5.10	8.48	4.52	10.78	5.90
exp-AR$_T$	7.12	3.94	10.41	5.83	7.96	4.46	11.37	6.71	9.28	5.23	12.75	7.70
qLL-AR$_T$	6.84	3.26	9.88	5.44	7.64	3.84	10.66	6.06	8.82	4.76	11.47	6.78
split-AR$_T$	7.43	3.77	9.62	5.36	8.16	4.29	10.63	5.88	9.37	5.10	11.83	6.59
split-KJ$_T$	6.07	3.02	9.88	5.28	6.79	3.48	10.97	6.28	8.29	4.26	11.96	7.22
split-MLR$_T$	5.62	2.45	9.89	5.08	6.37	2.90	10.79	5.85	7.76	3.74	11.76	6.69

Table 1: Null rejection frequencies of tests of $H_0 : \alpha = \frac{2}{3}$ in the NKPC model. The instruments are x_{t-1}, x_{t-2}, the sample size is 180, and the number of Monte Carlo simulations is 20,000.

The rejection frequencies for all of the proposed the tests are close to their nominal level. Some tests appear to be undersized, the most severe being the *split-MLR*$_T$ test, whose rejection frequency is little above half the nominal level when there is no break in the variance. The important message is that the size is almost unaffected by the changes in the coefficients, and especially changes in the variance.

Figure 4 reports further evidence on the size of the tests as the magnitude of the break in the coefficients ρ_1 and ρ_2 changes. We keep the variance of the shocks fixed and consider changes of up to 7 standard errors. We find that the size is affected very little by the magnitude of the instability.

The top part of Figure 5 considers the implications of changes in the variance for the size of the tests. The magnitude of the change in the variance is measured in ϕ standard error

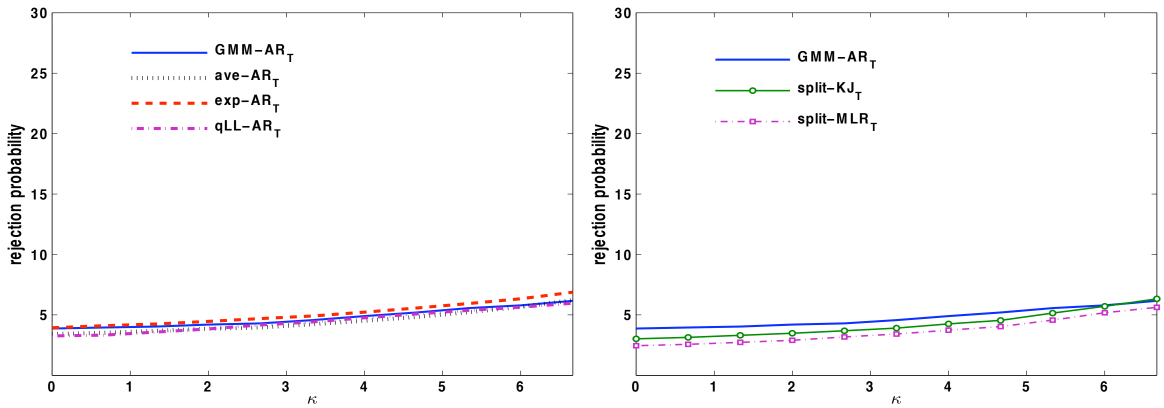

Figure 4: Rejection probability of 5% significance level tests of $H_0 : \alpha = \frac{2}{3}$ in the NKPC model with T=180 observations computed using 20,000 Monte Carlo replications when there is a change in the reduced form coefficients in the middle of the sample.

units. Even though the size of the tests increases with ϕ, the increase is very modest. Even with $\phi = 5$, the biggest size distortion does not exceed 8% at a nominal 5% level. These results are not particularly surprising, in view of the evidence reported by Hansen (2000), who studied this issue in a related context.

Finally, the bottom part of Figure 5 shows that the conclusions drawn from the top part Figure 5 remain unaltered when there is a change in the parameters, as well.

5 Empirical application

The new Keynesian Phillips curve is a forward-looking model of inflation dynamics that plays a central role in modern macroeconomic policy analysis. We consider the version of the model studied in Sbordone (2002):

$$\pi_t = \beta E_t \pi_{t+1} + \gamma \pi_{t-1} + \lambda x_t + \varepsilon_t, \tag{28}$$

where π_t is inflation, $\beta = \delta / (1 + \delta\varrho)$, δ is a discount factor, ϱ is the fraction of prices that are indexed to past inflation when they cannot be optimally reset, α is the probability that a price will be fixed in a given period, $\gamma = \varrho / (1 + \delta\varrho)$, and $\lambda = (1 - \alpha)(1 - \delta\alpha) / (\alpha (1 + \delta\varrho))$. The variable x_t is a measure of economic slack, and we shall proxy it using the labor share,

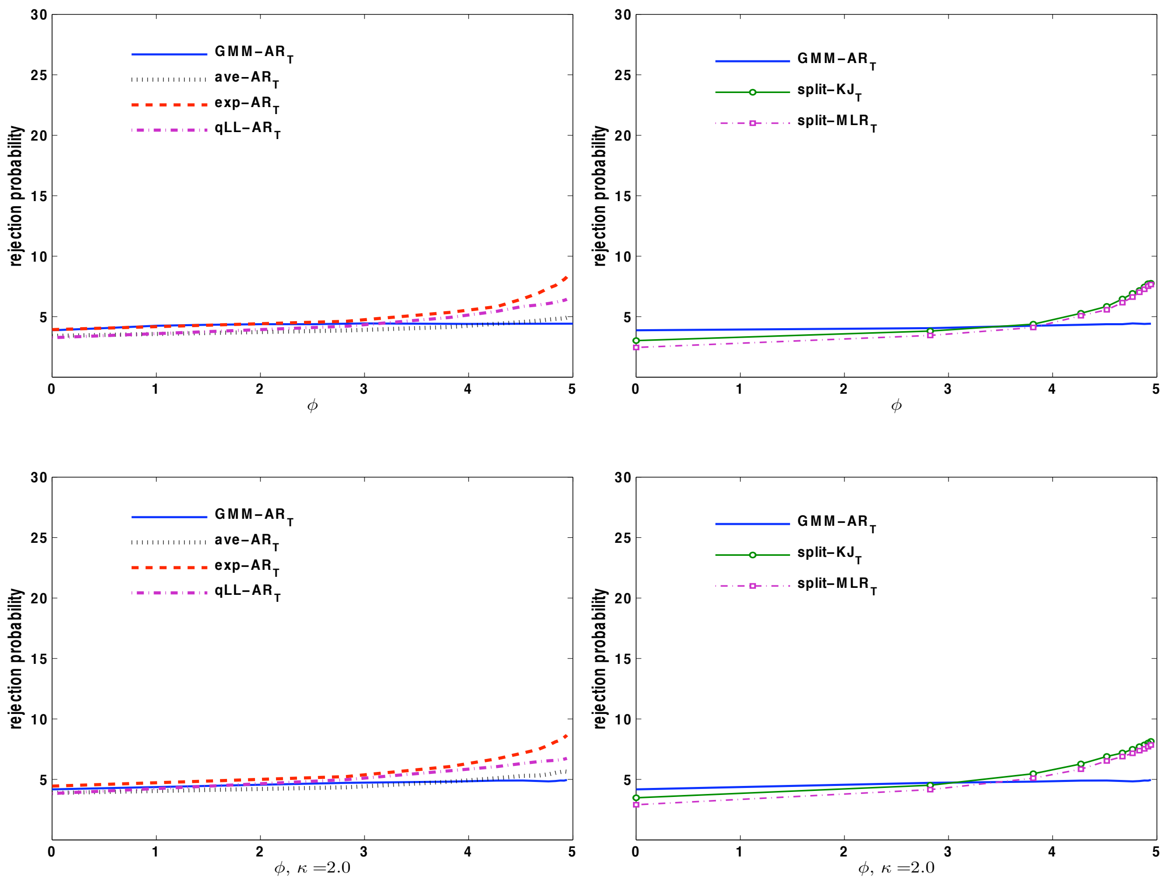

Figure 5: Rejection probability of 5% significance level tests of $H_0 : \alpha = \frac{2}{3}$ in the NKPC model with T=180 observations computed using 20,000 Monte Carlo replications. The top part consider the case of a change in the reduced form variance in the middle of the sample. The bottom part considers a change in the reduced form coefficients and variance in the middle of the sample.

following Galí and Gertler (1999) and Sbordone (2002).[5]

In accordance with the literature, see Kleibergen and Mavroeidis (2009), we impose the restriction $\delta = 1$, which allows us to write the model as

$$\Delta\pi_t = \beta E_t\left(\pi_{t+1} - \pi_{t-1}\right) + \lambda x_t + \varepsilon_t, \tag{29}$$

and we use obtain confidence sets two lags of $\Delta\pi$ and x as instruments. We allow for an unrestricted constant in equation (29) as well as in the set of instruments. The sample period is 1966q1-2008q4. Confidence sets at the 90% and 95% level for the parameters (α, ϱ) are constructed by inverting the various identification robust tests, and they.are plotted in Figures 6, 7 and 8. The following conclusions can be drawn from Figures 6 and 7.

First, the full-sample 95%-level confidence intervals for the coefficient ϱ based on both the GMM-AR and MLR tests cover the entire parameter space, so this parameter is completely unidentified by information over the full sample. This conclusion remains robust to changes in the sample size and number of instruments (not reported).

Second, the confidence regions based on inverting the generalized AR tests are a fraction of their full-sample counterparts. The confidence sets based on the *qLL-*AR_T test are the smallest followed by the *exp-*AR_T and *ave-*AR_T. This is consistent with the view of a slow persistent time variation in the parameters, see Stock and Watson (1996). Based on the *qLL-*AR_T test, we estimate that the indexation parameter is less than .4, which allows us to rule out the case of full indexation $\varrho = 1$ that was used in Christiano, Eichenbaum, and Evans (2005). This interval is less than half the size of the interval obtained from the *GMM-*AR_T test. The confidence interval for the price rigidity parameter α is also smaller, though the difference is not as dramatic as in the case of ϱ (about 25% smaller). The average duration of prices, computed from $\frac{1}{1-\alpha}$, is unbounded, which is consistent with the finding in the literature, see Kleibergen and Mavroeidis (2009).

Finally, it is interesting to look at the confidence sets that are obtained by inverting

[5] Inflation is measured by $\pi_t = 100 \times \left[\ln\left(\frac{GDP_t}{GDP_{t-1}}\right)\right]$, where GPD is the quarterly implicit gross domestic product price deflator. The labor share is $x_t = 100 \times .1226 \times \ln\left(\frac{w_t}{h_t}\right)$, where w_t is the ratio of compensation per hour and implicit price deflator in the nonfarm business sector, and h_t is the seasonally adjusted output per hour in the nonfarm business section over 100.

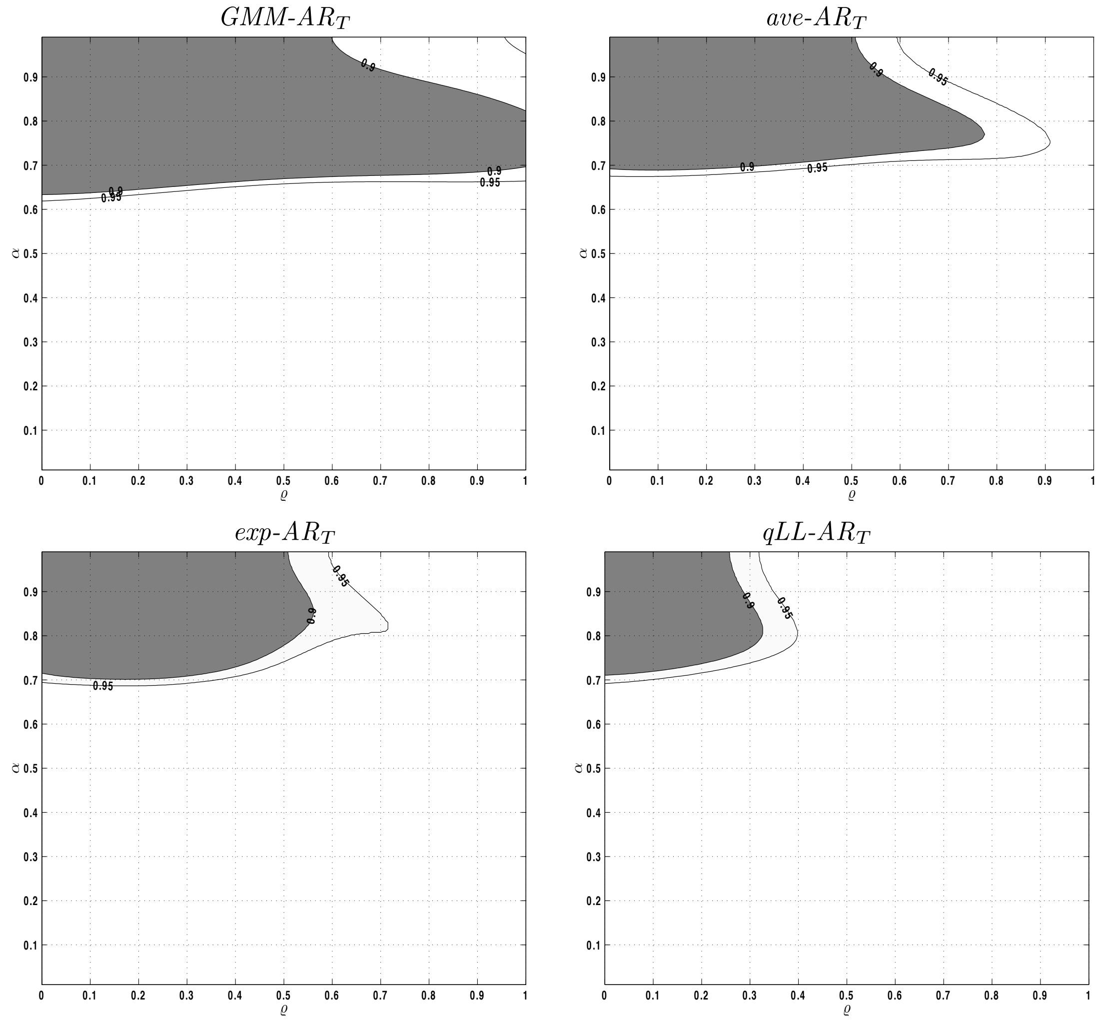

Figure 6: GMM-AR and generalized AR confidence sets for α and ϱ in the NKPC. Period: 1966q1 2008q4. The forcing variable is the labor share. Instruments include two lags of $\Delta\pi$ and the labor share.

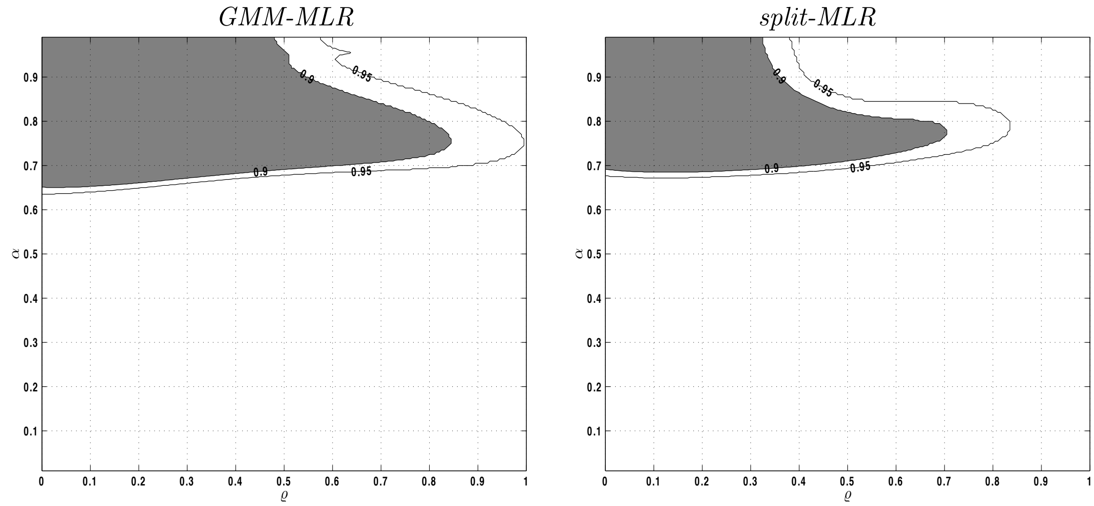

Figure 7: *GMM-MLR* and *split-MLR* confidence sets for α and ϱ in the NKPC. Period: 1966q1 2008q4. Instruments include two lags of $\Delta\pi$ and the labor share.

the pure stability tests *ave*-AR_T^B, *exp*-AR_T^B and *qLL*-AR_T^B. These are reported in Figure 8, where the GMM-AR confidence sets are also included for comparison. The confidence sets thus obtained are also considerably smaller than the confidence sets based on the full-sample statistics. This lending further support to the view that stability restrictions are an important source of identification in this application.

6 Conclusions

The contribution of this paper was twofold. First, we showed that structural change is useful for inference on structural parameters that are stable, a leading example being models that are immune to the Lucas critique. We demonstrated this both in theory as well as in practice. Secondly, we proposed methods for exploiting the information in the stability restrictions using only mild assumptions about the nature of instability. We considered two alternative approaches: (i) jointly testing the validity of the moment conditions on average and the stability restrictions implied by the model – this leads to generalized Anderson-Rubin tests; and (ii) estimating the break dates and using split-sample versions of existing identification robust GMM tests. None of the approaches dominates in terms of power, while generalized AR tests are more robust since they are based on weaker assumptions.

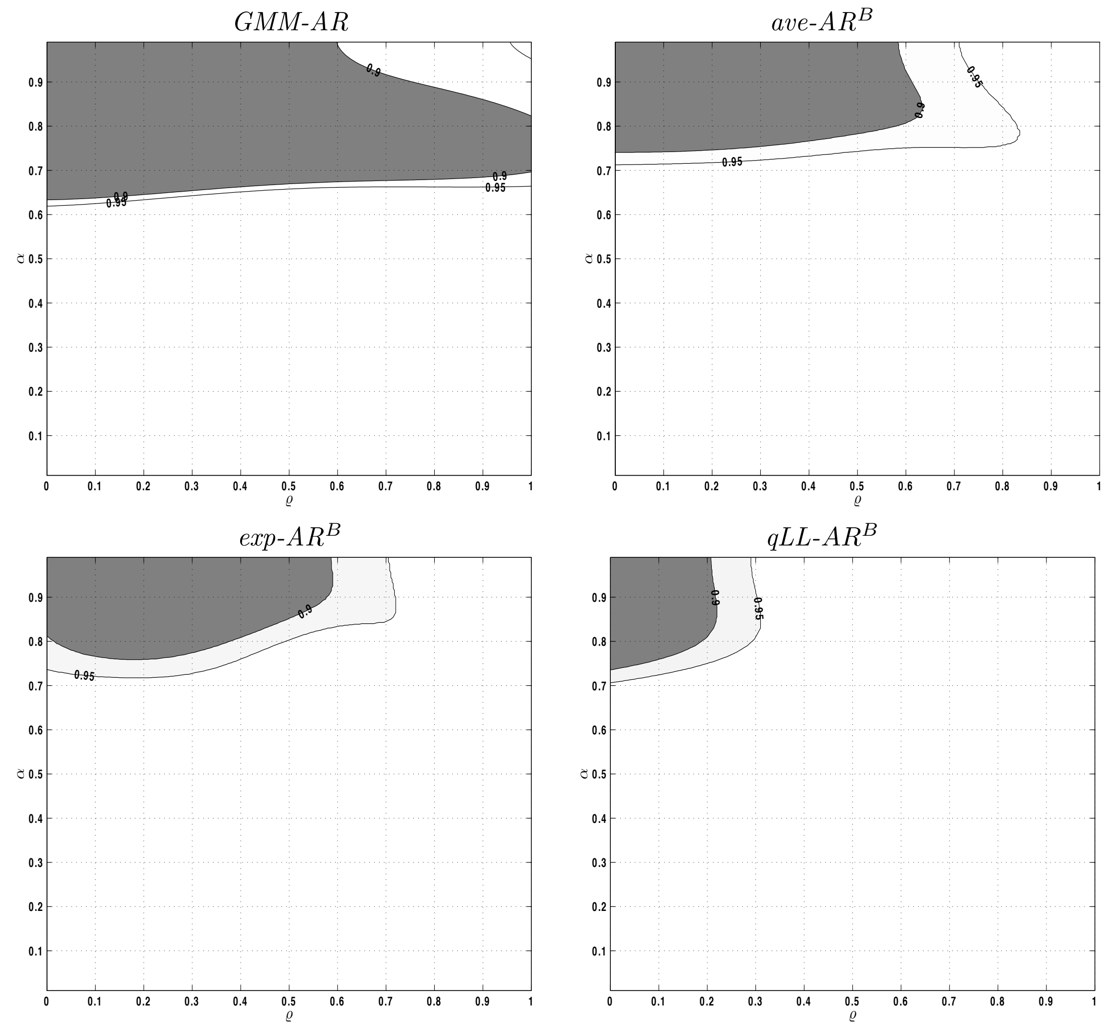

Figure 8: GMM-AR and stability-only confidence sets for α and ϱ in the NKPC. Period: 1966q1 2008q4. The forcing variable is the labor share. Instruments include two lags of $\Delta\pi$ and the labor share.

An interesting feature of our proposed method of inference is that it allows for identification of the parameters even when the usual GMM order condition for identification fails, i.e., when the number of instruments is smaller than the number of parameters. This may be useful in situations where alternative exclusion restrictions may be controversial.

References

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statistics 20*, 46–63.

Andrews, D. W. and J. H. Stock (2005). Inference with Weak Instruments. NBER Technical Working Papers 0313, National Bureau of Economic Research, Inc.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–56.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica 74*(3), 715–752.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Christiano, L. J., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *J. Political Economy 113*, 1–45.

Clarida, R., J. Galí, and M. Gertler (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics 115*, 147–180.

Elliott, G. and U. K. Mueller (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies 73*(4), 907–940.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*(2), 277–304.

Galí, J. and M. Gertler (1999). Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics 44*, 195–222.

Giacomini, R. and B. Rossi (2009). Model comparisons in unstable environments. ERID Working Paper 30, Duke.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*(1), 93–115.

Hendry, D. F. and M. Massmann (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business & Economic Statistics 25*, 33–51.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica 73*(4), 1103–1123.

Kleibergen, F. and S. Mavroeidis (2009). Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve. *Journal of Business and Economic Statistics 27*(3), 293–311.

Klein, R. and F. Vella (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics 154*(2), 154–164.

Lewbel, A. (2003). Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. Boston College Working Papers in Economics 587, Boston College Department of Economics.

Li, H. (2008). Estimation and testing of euler equation models with time-varying reduced-form coefficients. *Journal of Econometrics 127*(1), 425–448.

Li, H. and U. Mueller (2006). Valid inference in partially unstable gmm models. mimeo, Princeton University.

Lucas, R. E. J. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. Meltzer (Eds.), *The Philips Curve and Labor Markets.*, Carnegie-Rochester Conference Series on Public Policy. Amsterdam: North-Holland.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the United States: What has changed since the early 1980's? *The American Economic Review 90*(5), 1464–1476.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Mueller, U. (2008). An alternative sense of asymptotic efficiency. mimeo.

Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Perron, P. (2005). Dealing with structural breaks. Boston University - Department of Economics - Working Papers Series WP2005-017, Boston University - Department of Economics.

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics 85*(4), 777–792.

Rossi, B. (2005). Optimal tests for nested model selection with underlying parameter instability. *Econometric Theory 21*(05), 962–990.

Sbordone, A. M. (2002). Prices and unit labor costs: a new test of price stickiness. *Journal of Monetary Economics 49*, 265–292.

Sims, C. A. and T. Zha (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review 96*(1), 54–81.

Sowell, F. (1996). Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica 64*(5), 1085–1107.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *J. Amer. Statist. Assoc. 93*(441), 349–358.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H., J. H. Wright, and M. Yogo (2002). GMM, weak instruments, and weak identification. *Journal of Business and Economic Statistics 20*, 518–530.