Identification using stability restrictions

Leandro M. Magnusson† Sophocles Mavroeidis‡
University of Western Australia University of Oxford

May 31, 2012

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 on the parameters of a leading macroeconomic policy model, the new Keynesian Phillips curve.

Keywords: GMM, identification, structural stability, Lucas critique.

JEL: C22, E31

†Department of Economics, University of Western Australia, Email: leandro.magnusson@uwa.edu.au

‡Corresponding author: Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom. Email: sophocles.mavroeidis@economics.ox.ac.uk.
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 macroeconomic models used for policy analysis, 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 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 2009) 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 paper also relates to the notion of ‘super exogeneity’, see (Engle, Hendry, and Richard 1983) and (Hendry and Santos 2010), 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 followed by a brief conclusion. Proofs are given in the appendix.

2 Assumptions and tests

Consider a $p$-dimensional vector of structural parameters $\theta$ whose parameter region $\Theta$ 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 $\theta$, i.e.,

$$E[f_t(\theta)] = 0 \quad \text{for all } t \leq T, T \geq 1.$$  \hspace{1cm} (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 $Z_t \in \mathbb{R}^{G \times k}$ in $\mathcal{I}_t$, the conditional moment restrictions can be converted to unconditional restrictions in (1) by defining

$$f_t(\theta) = Z_t' h_t(\theta).$$  \hspace{1cm} (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

\[
H_0 : \theta = \theta_0 \quad \text{and} \quad H_1 : \theta \neq \theta_0,
\]

using tests with significance level \( \alpha \). The robustness requirement is that \( \alpha \)-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

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

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 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 restrictions \( E (F_T (\theta_0)) = 0 \). 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 $\theta_0$, and $\Rightarrow$ denote weak convergence of the underlying probability measures. We make the following high-level assumption about the large sample behavior of $X_T$ under both $H_0$ and $H_1$.

**Assumption 1** (i) The process $X_T(s) = T^{-1/2}F_{sT}(\theta_0)$ satisfies: $X_T(\cdot) - E[X_T(\cdot)] \Rightarrow V_{ff}^{1/2}W(\cdot)$, 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) 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 one of the other stability tests. 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 that the magnitude of any changes in the variance of the sample moments converges to zero as the sample increases, following the approach used by (Bai and Perron 1998) to obtain pivotal statistics for inference on break dates. Therefore, Assumption 1 does not preclude changes in the variance that can be detected with possibly high probability. In addition, the results in (Hansen
An important requirement for robustness is that tests should control size in cases when $\theta$ may be arbitrarily weakly identified. This covers situations when identification remains weak using both the full-sample and the stability restrictions. For the full-sample restrictions, weak identification is characterized using the local-to-zero asymptotic nesting of (Stock and Wright 2000). For the stability restrictions, weak identification corresponds to small instability/breaks under the alternative. To make these precise, define the expected value of the moment function under $H_0$, $f_t(\theta_0)$, taken with respect to the distribution of the data at the true value of the parameters $\theta$ as $m_T(\theta, \frac{t}{T}) \equiv E[f_t(\theta_0)]$, where $m_T(\theta, \cdot)$ is a step function such that $m_T(\theta, r) = m_T(\theta, \frac{t}{T})$ for $t/T \leq r < (t + 1)/T$. Then, $E[T^{-1}F_{sT}(\theta_0)] = \int_0^s m_T(\theta, r) dr$, $E[T^{-1}F_T(\theta_0)] = \int_0^1 m_T(\theta, r) dr = \bar{m}_T(\theta)$ and $E[T^{-1}\bar{F}_{sT}(\theta_0)] = \int_0^s \tilde{m}_T(\theta, r) dr$ where $\tilde{m}_T(\theta, r) = m_T(\theta, r) - \bar{m}_T(\theta)$. We distinguish the following cases. Strong full-sample identification: $\lim_{T \to \infty} \bar{m}_T(\theta) = \bar{m}(\theta)$, and $\bar{m}(\theta) = 0$ if and only if $\theta = \theta_0$. Weak full-sample identification: $\lim_{T \to \infty} \sqrt{T}\bar{m}_T(\theta) = \bar{m}(\theta)$, with $\bar{m}(\theta_0) = 0$ – this is the nesting of (Stock and Wright 2000). Large breaks (strong identification via stability restrictions): $\lim_{T \to \infty} \tilde{m}_T(\theta, r) = \tilde{m}(\theta, r)$ and $\bar{m}(\theta, r) = 0$ for all $r \in [0, 1]$ if and only if $\theta = \theta_0$. Finally, small breaks (weak identification via stability restrictions): $\lim_{T \to \infty} \sqrt{T}\tilde{m}_T(\theta, r) = \tilde{m}(\theta, r)$, with $\tilde{m}(\theta_0, r) = 0$ for all $r \in [0, 1]$. Thus, there are four cases to consider in total. We will focus on the case of weak full-sample identification/small breaks, which is the most appropriate one for robustness, and the most relevant empirically. This is described in the following assumption. For completeness, the other three cases are discussed in a supplementary appendix.

**Assumption 2** $E[X_T(s)] \to \int_0^s m(\theta, r) dr$ 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 $s \in [0, 1]$, and $m(\theta, 0) = 0$ for all $\theta$.

At $s = 1$, Assumption 2 corresponds to the weak identification assumption in Stock and Wright (2000). This assumption makes precise the notion that the moment conditions [1] are nearly satisfied even when the true value $\theta$ is far from the hypothesized value $\theta_0$.

---

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

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.

The following examples illustrate the assumptions.

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

\[
y_t = \beta E_t y_{t+1} + \gamma x_t + u_t,
\]

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 \( \beta \) and \( \gamma \) 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 \( \beta \) and \( \gamma \) 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

\[ x_t = \rho x_{t-1} + (1 - \rho) \varphi y_t + \varepsilon_t, \]  

(6)

where \( \varepsilon_t \) is an unobserved ‘policy’ shock. Then, in a determinate rational expectations equilibrium, the dynamics of \( y_t \) and \( x_t \) are given by a restricted first order vector autoregression of the form \( y_t = \alpha_1 x_{t-1} + v^y_t, \) \( x_t = \rho_1 x_{t-1} + v^x_t, \) where \( v^y_t, v^x_t \) are innovations. So, the only relevant instrument is \( x_{t-1}. \) Consequently, the parameters \( \beta \) and \( \gamma \) are under-identified, because there are two endogenous regressors \( y_{t+1} \) and \( x_t. \)

Now, suppose that policy changes over time, e.g., \( \varphi \) becomes \( \varphi_t, \) 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 \textit{a priori} assumptions about the incidence, nature and timing of these changes. For example, it may be that \( \varphi \) changes ‘very little’ or not at all, or that it changes a fixed number of times at unknown dates, or that it drifts slowly 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 \( \varphi_t = \varphi_0 + O(1). \) 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 \( 1. \)

\[ \square \]

\textbf{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 \( u_t, \varepsilon_t, \) 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 \( \theta \) are under-identified. Now, suppose that the volatilities of the shocks are time-varying, then the weaker condition \( \varphi_t - \varphi_0 = o(1) \) is actually sufficient for asymptotic mse-stationarity.
so that $E[f_t(\theta_0)] = \beta_0\gamma_0 E(y_t x_t) - \beta_0 E(x_t^2) - \gamma_0 E(y_t^2)$ 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. The identifying implication of changes in variance was exploited by [Rigobon 2003] and [Klein and Vella 2010], who developed it in the context of structural vector autoregressions under more restrictive assumptions.

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

$$gen-AR_T(\theta_0) = stab-AR_T^c(\theta_0) + \frac{\bar{c}}{1 + \bar{c}} GMM-AR_T(\theta_0),$$

(7)

where, under $H_0$, $GMM-AR_T(\theta_0)$ tests the validity of the full-sample moment restrictions, and $stab-AR_T^c(\theta_0)$ tests the stability restrictions. $\bar{c}$ and $\bar{c}$ 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 $gen-AR_T(\theta_0)$ is asymptotically pivotal, and a test that rejects for large values of $gen-AR_T(\theta_0)$ is asymptotically efficient as described in Section 3. Asymptotic critical values are nonstandard, but can be computed by simulation.

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

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

(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 $gen-AR_T$ statistic when the investigator puts zero weight ($\bar{c} = 0$) on instability under $H_1$. The specification of $stab-AR_T$ statistic, the second component of $gen-AR_T$,
depends on the assumptions about the nature of instability under $H_1$. We consider the two leading cases: a single break at unknown date $\tau$, and martingale time variation.

In the case of a single break at unknown date, the stab-$AR_T$ 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 $F_T^1 (\theta_0; \frac{t_b}{T}) = F_{t_b} (\theta_0)$ and $F_T^2 (\theta_0; \frac{t_b}{T}) = F_T (\theta_0) - F_{t_b} (\theta_0)$, obtain estimates of their variance $\hat{V}_{ff}^i (\theta_0; \frac{t_b}{T}) \ i = 1, 2$, and evaluate the partial-sample (continuously updated) GMM objective function $S_T (\theta_0; \frac{t_b}{T}) = \sum_{i=1}^{2} T^{-1} F_T^i (\theta_0; \frac{t_b}{T})' \hat{V}_{ff}^i (\theta_0; \frac{t_b}{T})^{-1} F_T^i (\theta_0; \frac{t_b}{T})$, with $T_1 = t_b$ and $T_2 = T - T_1$. For $\hat{V}_{ff}^i (\theta_0; \frac{t_b}{T})$, 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 $GMM-AR_T (\theta_0)$ using equation [8], and obtain $\tilde{S}_T (\theta_0; \frac{t_b}{T}) = S_T (\theta_0; \frac{t_b}{T}) - GMM-AR_T (\theta_0)$ for each $t_b$.

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

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

In the case of martingale time variation, the statistic $stab-AR_T^c (\theta_0)$ 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 $w_{t,i} = \tilde{\sigma} w_{t-1,i} + \Delta v_{t,i}$, for $t = 2, ..., T$, with $\tilde{\sigma} = 1 - \frac{1}{T}$. 

11
3. Regress \{w_{t,i}\} on \{\tilde{r}^t\} and obtain the squared residuals, sum over all \(i = 1, \ldots, k\) and multiply by \(\tilde{r}\).

4. Compute \(\sum_{i=1}^{k} \sum_{t=1}^{T} (v_{i,t} - \bar{v}_i)^2\), where \(\bar{v}_i = T^{-1} \sum_{t=1}^{T} v_{i,t}\) 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 \(\hat{c} = 10\), and denote the resulting stability statistic as \(qLL-AR_T^B\). 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 \(qLL-AR_T\).

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’ \(GMM-AR_T\) statistic, or simply \(split-AR_T\), which is a special case of the \(ave-AR_T\) and \(exp-AR_T\) statistics when the break date is known.\(^3\) Moreover, under Assumption 1, asymptotic critical values for the \(split-AR_T\) test can be obtained from a \(\chi^2(2k)\) distribution.

In addition to the split-sample \(GMM-AR_T\) 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 (\(split-KLM_T\)) and ‘split-sample’ quasi-likelihood ratio (\(split-MLR_T\)) tests, the latter being a generalization of the conditional likelihood ratio test of (Moreira 2003). The motivation for considering such tests is that they are asymptotically more powerful than the \(GMM-AR_T\) 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,

\(^3\)Think of the analogy between the Chow test and the Quandt likelihood ratio test.
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.

**Assumption 3** Let \( q_t(\theta_0) = \text{vec} \left[ \frac{\partial f_t(\theta_0)}{\partial \theta'} \right] \), \( Q_sT(\theta_0) = \sum_{t=1}^{[sT]} q_t(\theta_0) \). We assume that (i)

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

where \( W_f \) and \( W_q \) are independent standard \( k \times 1 \) and \( kp \times 1 \) Wiener processes; and (ii) There exists a consistent estimator \( \hat{V}_T(\theta_0) \) of

\[
V = \begin{pmatrix} V_{ff} & V_{fq} \\ V_{qf} & V_{qq} \end{pmatrix}.
\]

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

\[
y_{1,t} = Y_{2,t} \theta + u_t, \quad t = 1, \ldots, T
\]

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 \( \theta \in \mathbb{R}^p \) is the unknown structural parameter vector. The reduced-form equation (also known as first-stage regression) is given by

\[
Y_{2,t} = Z_t \Pi_t + V_{2,t}, \quad t = 1, \ldots, T
\]

where \( Z_t \in \mathbb{R}^{1 \times k}, t = 1, \ldots, T \) is a sequence of observed instrumental variables that are fixed, \( V_{2,t} \in \mathbb{R}^{1 \times p}, t = 1, \ldots, T \) is a (reduced form) error vector, and \( \Pi_t \in \mathbb{R}^{k \times p}, t = 1, \ldots, T \) is a sequence of unknown parameters. The errors \( u_t \) and \( V_{2,t} \) are iid and have a mean zero joint 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 $f_t(\theta) = Z_t (y_{t,1} - Y_{2,t} \theta)$.

Assumption 3 is satisfied if $\frac{1}{T} \sum_{t=1}^{[sT]} Z_t^T Z_t \rightarrow s Q_{ZZ}$ uniformly in $s$, where $Q_{ZZ}$ is nonsingular. Assumption 2 is satisfied when $\Pi_t = O(T^{-1/2})$, 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 $F_T^1(\theta_0, \frac{t_b}{T}) = F_{tb}(\theta_0)$ and $F_T^2(\theta_0, \frac{t_b}{T}) = F_T(\theta_0) - F_{tb}(\theta_0)$, and their Jacobians $Q_T^1(\theta_0, \frac{t_b}{T}) = Q_{tb}(\theta_0)$ and $Q_T^2(\theta_0, \frac{t_b}{T}) = Q_T(\theta_0) - Q_{tb}(\theta_0)$, and estimate their $k(p+1) \times k(p+1)$ variance matrix $\hat{V}_T^i(\theta_0, \frac{t_b}{T})$, $i = 1, 2$, using either partial-sample estimators or a full-sample estimator. Compute the $k \times p$ matrices $D_T^i(\theta_0, \frac{t_b}{T})$, $i = 1, 2$, from

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

and estimate their variance by

$$\hat{V}_{qf}^i(\theta_0, \tau) = \hat{V}_{qf}^i(\theta_0, \frac{t_b}{T}) - \hat{V}_{qf}^i(\theta_0, \frac{t_b}{T}) \hat{V}_{ff}^i(\theta_0, \frac{t_b}{T})^{-1} \hat{V}_{qf}^i(\theta_0, \frac{t_b}{T})'. $$

Obtain the restricted estimator of the break date by

$$\tilde{t}_b(\theta_0) = \arg\max_{t_b} \sum_{i=1}^{2} T_i^{-1} vec[D_T^i(\theta_0, \frac{t_b}{T})] \hat{V}_{qf}^i(\theta_0, \frac{t_b}{T})^{-1} vec[D_T^i(\theta_0, \frac{t_b}{T})],$$

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

14
2. Split-sample test statistics. The identification robust split-sample statistics are given by

\[ \text{split-AR}_T (\theta_0, \tau) = \sum_{i=1}^{2} T_i^{-1} F_i (\theta_0, \tau)' \hat{V}_i^f (\theta_0, \tau)^{-1} F_i (\theta_0, \tau) \]

\[ \text{split-KLM}_T (\theta_0, \tau) = \sum_{i=1}^{2} T_i^{-1} F_i (\theta_0, \tau)' \hat{V}_i^f (\theta_0, \tau)^{-1/2} P_{\hat{V}_i^f (\theta_0, \tau)^{-1/2} D_i (\theta_0, \tau)} \]

\[ \text{split-JKLM}_T (\theta_0, \tau) = \text{split-AR}_T (\theta_0, \tau) - \text{split-KLM}_T (\theta_0, \tau) \]

and

\[ \text{split-MLR}_T (\theta_0, \tau) = \frac{1}{2} [\text{split-AR}_T (\theta_0, \tau) - rk_\theta (\theta_0, \tau) + \sqrt{(\text{split-AR}_T (\theta_0, \tau) + rk_\theta (\theta_0, \tau))^2 - 4 \text{split-JKLM}_T (\theta_0, \tau) rk_\theta (\theta_0, \tau)}] \]

with \( rk_\theta (\theta, \tau) \) being a statistic that tests for a lower rank value of the Jacobian of the moment conditions, e.g.,

\[ rk_\theta (\theta, \tau) = \min_{\phi \in \mathbb{R}^{p-1}} \left( \frac{1}{\phi} \sum_{i=1}^{2} \frac{1}{T_i} D_i (\theta, \tau)' \left( \left( \frac{1}{\phi} \right) \otimes I_k \right)' \hat{\nu}_i^f (\theta) \left( \left( \frac{1}{\phi} \right) \otimes I_k \right)^{-1} D_i (\theta, \tau) \left( \frac{1}{\phi} \right) \right) \]

proposed by (Kleibergen and Mavroeidis 2009). Evaluate the above statistics at \( \tau = \tilde{\tau} (\theta_0) \). 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 \( \tilde{\tau} (\theta_0) \) are given by the asymptotic distributions that arise as if \( \tilde{\tau} (\theta_0) \) were nonrandom. This is because, under Assumption 3 and \( H_0 : \theta = \theta_0, \tilde{\tau} (\theta_0) \) is asymptotically independent of \( T^{-1/2} F_T (\theta_0) \). 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 \( rk_\theta (\theta_0, \tilde{\tau} (\theta_0)) \) can be computed by simulation.

3 Asymptotic theory

Assumption 1 and the continuous mapping theorem ((White 2001, Theorem 7.20)) imply that, under \( H_0, G (X_T (s)) \Rightarrow G (W (s)) \) for any continuous functional \( G (\cdot) \) on \( D_k [0, 1] \). 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 \( \nu_0, \nu_1 \) denote the probability measures of \( X \) under \( H_0 \) and \( H_1 \) respectively. We shall obtain efficient tests for the limiting problem of testing \( \nu_0 \) against \( \nu_1 \), 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, \( \nu_1 \) is determined by the stochastic differential equation \( dX (s) = m(\theta, s) \, ds + V_X^{1/2} dW (s) \), and \( \nu_0 \) is determined by \( dX (s) = V_X^{1/2} dW (s) \).

Therefore, \( \nu_1 \) is absolutely continuous with respect to \( \nu_0 \), and the Radon Nikodym derivative of \( \nu_1 \) with respect to \( \nu_0 \) conditional on the entire path of \( m(\theta, \cdot) \) is given by

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

(16)

Under the maintained assumptions, the process \( X (s) \) can be decomposed orthogonally into \( X \equiv X (1) \) and \( \tilde{X} (s) \equiv X (s) - sX (1) \). Define \( \overline{m} (\theta) = \int_0^1 m(\theta, s) \, ds \) and \( \tilde{m} (\theta, s) = m(\theta, s) - \overline{m} (\theta) \), and note that the random function \( \xi (m) \) in equation (16) factors into the product of

\[
\tilde{\xi} (\overline{m}) = \exp \left\{ \overline{m}(\theta)' V_X^{-1} \overline{X} - \frac{1}{2} \overline{m}(\theta)' V_X^{-1} \overline{m}(\theta) \right\},
\]

(17)

and

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

(18)

It follows that the statistic \( \tilde{X} (s) \) is ancillary for \( \theta \) if and only if \( m(\theta, \cdot) \) is constant, so that \( \tilde{m} (\theta, \cdot) = 0 \) for every \( \theta \). 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 \( F_{sT} (\theta_0) - sF_T (\theta_0) \), \( s \in [0, 1] \), is asymptotically ancillary for \( \theta \) 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
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 (16). Let $\nu_m$ 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 $\hat{V}_{ff} (\theta_0)$ for $X(s)$ and $V_{fX}$, respectively, in equation (16).

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

For $\nu_{\overline{m}}$, we will use the conventional weight distribution $\overline{m}(\theta) \sim N(0, \bar{c}V_{fX})$, 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 $\nu_{\tilde{m}}$, 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 $\nu_m$ 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 $m(\theta, s) = m_1(\theta) 1_{\{s < \tau\}} + m_2(\theta) 1_{\{s \geq \tau\}}$, 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 will have good power even against multiple breaks, see (Elliott and Mueller 2006). So, for practical purposes, restricting attention to the case of a single break seems reasonable.4

Assumption 4 implies $m(\theta) = \tau m_1(\theta) + (1 - \tau) m_2(\theta)$ and $\tilde{m}(\theta, s) = (1_{\{s < \tau\}} - \tau) (m_1(\theta) - m_2(\theta))$. Let $\delta(\theta) = \tau (1 - \tau) (m_1(\theta) - m_2(\theta))$ denote the mean of $\tilde{X}(\tau)$ under $H_1$, i.e., $\tilde{X}(\tau) \overset{H_1}{\sim} N(\delta(\theta), \tau (1 - \tau) V_X)$. Then, $\tilde{\xi}(\tilde{m})$ in equation (18) can be written in terms of $\delta(\theta)$ and $\tau$ as:

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

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

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

So, $LR^{\tilde{c}, \tilde{\delta}} = \tilde{LR}^\delta \times \tilde{LR}^\tilde{c}$, where:

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

Tests against multiple breaks may also have worse finite-sample properties, akin to the problem of ‘many instruments’, see (Andrews and Stock 2007).
By the Neyman-Pearson lemma, see also Sowell (1996, theorem 2), a test that rejects for large values of $LR^{\tilde{c},\tilde{c}}$ is an optimal test for testing $H_0$ in the limiting problem against the point alternative given by the probability measures $\nu_{\bar{c}}, \nu_{\tilde{c}}$, indexed by $\bar{c}$ and $\tilde{c}$, respectively, and $\nu_\tau$.

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 $X'V_X^{-1}X$. 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

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

where

$$S_T(\theta, \tau) = \frac{1}{T} \left( \begin{array}{c} F_T(\theta) \\ F_T(\theta) - F_{T_T}(\theta) \end{array} \right) \left( \begin{array}{cc} V_{ij}(\theta, \tau)^{-1} & 0 \\ 0 & V_{ij}(\theta, \tau)^{-1} \end{array} \right) \left( \begin{array}{c} F_T(\theta) \\ F_T(\theta) - F_{T_T}(\theta) \end{array} \right)$$

is the continuously updated version of the “partial-sample” GMM objective function of Andrews (1993). For the estimators $\hat{V}_{ij}^1(\theta; \tau)$, $\hat{V}_{ij}^2(\theta; \tau)$ we can use either respective partial-sample estimators, or a full-sample estimator $\hat{V}_{ij}(\theta)$, 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 $[F_{T_T}(\theta_0)', F_T(\theta_0)' - F_{T_T}(\theta_0)]'$.

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

$$\tilde{S}_T(\theta_0, \tau) = S_T(\theta_0, \tau) - GMM-AR_T(\theta_0)$$
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

$$break-AR_{T}^\bar{c} \cdot (\theta_0) = 2 \log \int_\varsigma \exp \left\{ \frac{1}{2} \frac{\bar{c}}{1 + \bar{c}} \tilde{S}_T(\theta_0, \tau) \right\} d\nu_{\tau} + \frac{\bar{c}}{1 + \bar{c}} GMM-AR_T(\theta_0). \quad (20)$$

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

The following result provides an asymptotically efficient test of $H_0 : \theta = \theta_0$ against alternatives that satisfy assumption 4.

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

$$break-AR_{T}^\bar{c} \cdot (\theta_0) \Rightarrow 2 \log \int_\varsigma \exp \left\{ \frac{1}{2} \frac{\bar{c}}{1 + \bar{c}} \tilde{\psi}_k(\tau) \right\} d\nu_{\tau} + \frac{\bar{c}}{1 + \bar{c}} \tilde{\psi}_k$$

where $\tilde{\psi}_k(\tau) = \frac{\tilde{W}(\tau)'\tilde{W}(\tau)}{\tau(1 - \tau)}$, with $\tilde{W}(\tau)$ a standard $k \times 1$ Brownian Bridge process, and $\tilde{\psi}_k$ is a \chi^2(k) distributed random variable independent of $\tilde{W}(\tau)$. Moreover, a test that rejects $H_0$ for large values of break-AR_{T}^\bar{c} \cdot (\theta_0)$ is asymptotically efficient against the alternative given by assumption 4.

### 3.2 Persistent time variation

Next, we consider the drifting parameter approach to modeling instability that was followed by Stock and Watson (1996, 1998) and Elliott and Mueller (2006).

**Assumption 5** $m(\theta, s) = \Omega^{1/2} W_m(s)$, where $W_m(\cdot)$ is a realization of a standard $k \times 1$ Wiener process, which is independent of $W(\cdot)$ in Assumption 1, and $\Omega$ is a positive definite $k \times k$ matrix, which is proportional to the long run variance of the moment conditions $V_{ff}$.

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 $y_t, \mu_t, u_t \in \mathbb{R}^k$. Assume $u_t \sim iidN(0, \Sigma)$ for some positive definite matrix $\Sigma$, such that $u \sim$
and 0 otherwise, and a subdiagonal, and zeros otherwise, i.e.,
\[ A_{i,j} = \begin{cases} 1 & \text{if } i = j, j+1, \text{ or } j+2, \\ 0 & \text{otherwise}. \end{cases} \]

The density of \( y = (y_1, \ldots, y_T)' \) conditional on \( \mu = (\mu_1', \ldots, \mu_T')' \) is given by
\[
f(y|\mu) = (2\pi)^{-Tk/2} |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} (y - \mu)' (I_T \otimes \Sigma^{-1}) (y - \mu) \right\}.
\]

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 \( T\Delta \mu_t \sim iidN(0, c\Sigma) \). \( H_0 \) can be decomposed into \( H_0^1 : \mu = 0, \) 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, \( H_0^3 : \tilde{\mu} = 0, \) where \( \tilde{\mu} = (B_e' \otimes I_k) \mu, \) \( B_e \) is a \( T \times (T-1) \) matrix such that \( B'e = 0, \) \( B'eB_e = I_{T-1}, B_eB_e' = M_e, \) \( M_e = I_T - e(e')^{-1}e' \) and \( e \) is the \( T \times 1 \) vector of ones.

Conditional on \( \mu \), the ratio \( f(y|\mu) / f(y|0) \) is given by
\[
\xi(\mu) = \exp \left\{ y' (I_T \otimes \Sigma^{-1}) \mu - \frac{1}{2} \mu' (I_T \otimes \Sigma^{-1}) \mu \right\},
\]
and the likelihood ratio is given by \( \int \xi(\mu) d\nu_\mu, \) where \( d\nu_\mu \) is the density of \( \mu \). As in the case of a single break, we specify independent weights over \( \bar{\mu} \) and \( \tilde{\mu} \), with densities given by \( \sqrt{T} \bar{\mu} \sim N(0, c\Sigma) \) and \( T\tilde{\mu} \sim N(0, [B'eF'F'B_e \otimes c\Sigma]), \) where \( F = [f_{i,j}] \) is a \( T \times T \) lower triangular matrix of ones, i.e., \( f_{i,j} = 1 \) for all \( i \leq j \) and 0 otherwise. The resulting likelihood ratio statistic depends on \( \bar{c} \) and \( \tilde{c} \), and can be written as the product of the statistics
\[
\tilde{LR}_T^c = (1 + \tilde{c})^{-k/2} \exp \left\{ \frac{1}{2} \tilde{c} T \bar{y}' \Sigma^{-1} \bar{y} \right\}
\]
and
\[
\tilde{LR}_T^\tilde{c} = \left( \frac{1 - r_c^2}{T (1 - r_c^2)^{T-1}} \right)^{-k/2} \exp \left\{ \frac{1}{2} \sum_{i=1}^{k} v_i' (M_e - G_c) v_i \right\}
\]
where \( \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t, v_i = (I_T \otimes l_i' \Sigma^{-1/2}) y, i = 1, \ldots, k, \) \( l_i \) is a \( k \times 1 \) vector with one at position \( i, \) and 0 otherwise, and \( G_c = H^{-1} - H^{-1} c (e'H^{-1} e)^{-1} e'H^{-1}, \) with \( H = r_c^{-1} F A_c A_c' F', A_c \) is a \( T \times T \) matrix with ones in the main diagonal, \( -r_c \) in its subdiagonal, and zeros otherwise, i.e., \( A_c = [a_{i,j}] \) where \( a_{i,j} = 1 \) if \( i = j, j+1, \) and 0 otherwise, and \( r_c = \frac{1}{2} (2 + c^2 T^{-2} - T^{-1} \sqrt{4c^2 + c^4 T^{-2}}) = 1 - c T^{-1} + o(T^{-1}). \) The derivation of \( \tilde{LR}_T^\tilde{c} \) 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

\[ LR_{T}^{\bar{c}, \tilde{c}} = \sum_{i=1}^{k} \hat{v}'_{i} (M_{e} - G_{\tilde{c}}) \hat{v}_{i} + \frac{\bar{c}}{1 + \bar{c}} T \bar{y}' \Sigma^{-1} \bar{y}. \]

The parameters \( \bar{c} \) and \( \tilde{c} \) govern the weights given to deviations from \( H_0 \) in the direction \( \mu \neq 0 \) and \( \tilde{\mu} \neq 0 \), respectively. \( LR_{T}^{\bar{c}, 0} \) coincides with the usual Wald statistic for \( \mu = 0 \), which is independent of \( c \), while \( LR_{T}^{0, \tilde{c}} \) is a pure stability test, conditional on \( \tilde{\mu} = 0 \). \(^5\)

\( LR_{T}^{c, \tilde{c}} \) gives equal weight to the two types of departure from \( H_0 \).

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 \( \Sigma \) by \( \hat{V}_{ff}(\theta_0) \). We shall denote the resulting statistic the \( qLL-AR_{T} \) statistic, and index it by the weights \( \bar{c}, \tilde{c} \).

\[ qLL-AR_{T}^{\bar{c}, \tilde{c}} (\theta_0) = \sum_{i=1}^{k} \hat{v}'_{i} (M_{e} - G_{\tilde{c}}) \hat{v}_{i} + \frac{\bar{c}}{1 + \bar{c}} GMM-AR_{T} (\theta_0), \tag{22} \]

\[ \hat{v}'_{i} = \left[ f_1 (\theta_0)' \hat{V}_{ff}(\theta_0)^{-1/2} \iota_{k,i}, \ldots, f_T (\theta_0)' \hat{V}_{ff}(\theta_0)^{-1/2} \iota_{k,i} \right] \]

The large sample properties of \( qLL-AR_{T}^{\bar{c}, \tilde{c}} \) are given by the following result.

**Theorem 2** Under Assumption \( \mathbb{I} \) and \( H_0 : \theta = \theta_0 \),

\[ qLL - AR_{T}^{\bar{c}, \tilde{c}} (\theta_0) \Rightarrow \psi_{\bar{c}} + \frac{\bar{c}}{1 + \bar{c}} \psi_{k}, \tag{23} \]

where \( \psi_{k} \sim \chi^2 (k) \) independent of \( \psi_{c} \), and

\[ \psi_{c} = \sum_{i=1}^{k} \left[ c J_i (1)^2 + c^2 \int_{0}^{1} J_i (s)^2 ds + \frac{2c}{1 - e^{-2c}} \left( e^{-c} J_i (1) + c \int_{0}^{1} e^{-cs} J_i (s) ds \right)^2 \right. \]
\[ \left. - \left( J_i (1) + c \int_{0}^{1} J_i (s) ds \right)^2 \right], \]

\( J_i (s) \) is the \( i \)th element of the \( k \)-dimensional Ornstein-Uhlenbeck process \( J (s) = W (s) - c \int_{0}^{s} e^{-c(s-r)} W (r) dr \), and \( W \) is a \( k \times 1 \) standard Wiener process. Moreover, a test that rejects for large values of the \( qLL-AR_{T}^{\bar{c}, \tilde{c}} (\theta_0) \) statistic is asymptotically efficient against

\(^5\)In this case, apart from the sign, the main difference of \( LR_{T}^{0, \tilde{c}} \) from the \( qLL \) statistic of Elliott and Mueller (2006) is that the latter uses demeaned \( y \).

22
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

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

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 $\Pi_t$ occurring at time $[\tau T]$ (the analysis generalizes easily to multiple breaks). Define $Z (\tau) \in \mathbb{R}^{T \times 2k}$ 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 $Y, V \in \mathbb{R}^{T \times (1+p)}$, respectively, and define $\Pi \in \mathbb{R}^{2k \times p}$ by $\Pi = [\Pi_1' : \Pi_2']'$. Then, the model (24) can be written as

$$Y = Z (\tau) \Pi A' + V. \quad (25)$$

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

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

$$= tr \left[ \Omega^{-1} A \Pi' Z (\tau)' Y \right] - \frac{1}{2} tr \left[ \Omega^{-1} A \Pi' Z (\tau)' Z (\tau) \Pi A' \right]. \quad (26)$$

Since $\{Z_t\}$ is non-random, the likelihood depends on the data only through the process.
where $b_0' = (1, -\theta_0')$, $A_0' = \begin{pmatrix} \theta_0 & \cdots & I_p \end{pmatrix}$.

Also, define the following quantities:

$$c_\theta = (\theta - \theta_0)(b_0'\Omega b_0)^{-1/2} \in \mathbb{R}, \quad d_\theta = A'\Omega^{-1}A_0 (A_0'\Omega^{-1}A_0)^{-1/2} \in \mathbb{R}^p.$$ 

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

**Lemma 1** For the model given by equation (25):

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

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

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

We can now write the log-likelihood function (26) in terms of the statistics $\mathcal{F}(s)$ and $\mathcal{D}(s)$

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

$$tr \left[ \mathcal{D}(\tau)' \mu_{\Pi,\tau}(\tau) d_\theta \right] - \frac{1}{2} tr \left[ d_\theta' \mu_{\Pi,\tau}(\tau)' \mu_{\Pi,\tau}(\tau) d_\theta \right].$$

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

$$L(\theta_0, \Pi, \tau) = tr \left[ \mathcal{D}(\tau)' \mu_{\Pi,\tau}(\tau) d_{\theta_0} \right] - \frac{1}{2} tr \left[ d_{\theta_0}' \mu_{\Pi,\tau}(\tau)' \mu_{\Pi,\tau}(\tau) d_{\theta_0} \right].$$ (27)
In other words, the process \( \overline{D}(\cdot) \) is sufficient for \( \Pi \) and \( \tau \) (or \( \overline{F}(\cdot) \) is specific ancillary for \( \Pi, \tau \) under \( H_0 \)). Thus, the restricted maximum likelihood estimator (MLE) of \( \Pi, \tau \) given \( \theta = \theta_0 \) can be obtained by minimizing (27) wrt \( \Pi \) and \( \tau \). Concentrating (27) with respect to \( \Pi \), we obtain

\[
L_c(\theta_0, \tau) = \frac{1}{2} \text{tr} \left[ \overline{D}(\tau)' \overline{D}(\tau) \right]
\]

and therefore, the MLE for \( \tau \) is

\[
\tilde{\tau} = \arg \max_{\tau} \text{tr} \left[ \overline{D}(\tau)' \overline{D}(\tau) \right].
\] (28)

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 JKM statistics are given by

\[
AR(s) = \overline{F}(s)' \overline{F}(s), \quad KLM(s) = \overline{F}(s)' \overline{D}(s) \left[ \overline{D}(s)' \overline{D}(s) \right]^{-1} \overline{D}(s)' \overline{F}(s)
\]

and

\[
JKLM(s) = AR(s) - KLM(s),
\]

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

\[
MLR(s) = \frac{1}{2} \left[ AR(s) - rk(s) + \sqrt{[AR(s) + rk(s)]^2 - 4J(s) rk(s)} \right],
\] (29)

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 (29) with \( rk(s) \) being a statistic that tests that the rank of the matrix \( \Pi \) 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 \( \tilde{\tau} \) is only a function of \( \overline{D}(\cdot) \), we have the following result.

**Theorem 3** Let \( \tilde{\tau} = \arg \max_{\tau} \overline{D}(s)' \overline{D}(s) \). Then, under \( H_0 \),

1. \( AR(\tilde{\tau}) \) is distributed as \( \chi^2(2k) \).
2. \( KLM(\tilde{\tau}) \) is distributed as \( \chi^2(p) \).
3. \( JKM(\tilde{\tau}) = AR(\tilde{\tau}) - LM(\tilde{\tau}) \) is distributed as \( \chi^2(2k-p) \).
4. \( KLM(\tilde{\tau}) \) and \( J(\tilde{\tau}) \) are independent.
5. The distribution of \( MLR(\tilde{\tau}) \) conditional on \( rk(\tilde{\tau}) \) is the same as the distribution of \( \frac{1}{2} \left[ \psi_p + \psi_{2k-p} - rk(\tilde{\tau}) + \sqrt{[\psi_p + \psi_{2k-p} - rk(\tilde{\tau})]^2 + 4\psi_{2k-p} rk(\tilde{\tau})} \right] \) conditional
on \( \text{rk}(\tilde{\tau}) \), where \( \psi_p, \psi_{2k-p} \) 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 \( F_T(\theta_0) \) are asymptotically ancillary for \( \tau \), since their asymptotic distribution does not depend on it.

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

\[
\bar{D}_T(\theta_0, \tau) = \begin{pmatrix}
D^1_T(\theta_0, \tau) & 0 \\
0 & D^2_T(\theta_0, \tau)
\end{pmatrix}
\]

and its variance

\[
\hat{V}_{\bar{D}}(\theta_0, \tau) = \begin{pmatrix}
\tau \hat{V}^{1, q}(\theta_0, \tau) & 0 \\
0 & (1 - \tau) \hat{V}^{1, q}(\theta_0, \tau)
\end{pmatrix}
\]

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

Consider the following estimator of \( \tilde{\tau} \)

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

This is a generalization of the estimator given by Equation (28) for the linear IV model. Under Assumption 3 and \( H_0 : \theta = \theta_0 \), \( \bar{D}_T(\theta_0, \cdot) \) is asymptotically independent of \( F_T(\theta_0) \), and hence, so is \( \tilde{\tau}(\theta_0) \). Therefore, we obtain the following result.

**Theorem 4** When Assumptions 3 and \( H_0 : \theta = \theta_0 \) hold, the limiting distributions of the split-sample GMM-AR, KLM, JKLM and MLR statistics, defined in equations (13)
and $\{14\}$, are given by

$$
\begin{align*}
\text{split-AR}_T (\theta_0, \tilde{\tau}(\theta_0)) & \xrightarrow{d} \psi_p + \psi_{2k-p} \\
\text{split-KLM}_T (\theta_0, \tilde{\tau}(\theta_0)) & \xrightarrow{d} \psi_p \\
\text{split-JKLM}_T (\theta_0, \tilde{\tau}(\theta_0)) & \xrightarrow{d} \psi_{2k-p} \\
\text{split-MLR}_T (\theta_0, \tilde{\tau}(\theta_0)) & \xrightarrow{d} \frac{1}{2} \left[ \psi_p + \psi_{2k-p} - r\kappa_0 (\theta_0, \tilde{\tau}(\theta_0)) + \sqrt{\left( \psi_p + \psi_{2k-p} + r\kappa_0 (\theta_0, \tilde{\tau}(\theta_0)) \right)^2 - 4\psi_{2k-p}r\kappa_0 (\theta_0, \tilde{\tau}(\theta_0))} \right],
\end{align*}
$$

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

### 3.4 Consistently estimable nuisance parameters

Often, the parameter of interest $\theta$ is a subset (or lower-dimensional function) of the parameters of the model, i.e., the model contains an additional ‘nuisance’ parameter $\zeta$ of dimension $p_\zeta \times 1$. With slight abuse of notation, assume that the moment conditions are given by $E[f_t (\theta, \zeta)] = 0$ for all $t \leq T$, where $f_t (\theta, \zeta)$ is a $k$-dimensional function of data and parameters and define the restricted estimator of $\zeta$

$$
\hat{\zeta}_0 = \arg \min_{\zeta} \sum_{t=1}^{T} f_t (\theta_0, \zeta) \psi_T \sum_{t=1}^{T} f_t (\theta_0, \zeta)
$$

where $\psi_T$ is an efficient weight matrix, as in two-step or continuously updated GMM, i.e., $\psi_T \xrightarrow{p} V_{ff}^{-1}$, where $V_{ff} = \lim \text{var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t (\theta, \zeta) \right]$. We make the following high-level assumption.

**Assumption 6** Under $H_0 : \tau = \tau_0$, the following conditions hold: (i) the process $X_T (s) = T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} f_t (\theta_0, \zeta_0)$ satisfies $X_T (\cdot) \Rightarrow V_{ff}^{1/2}W (\cdot)$, 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) $\psi_T \xrightarrow{p} V_{ff}^{-1}$; and (iii) the process $\hat{X}_T (s) = T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} f_t (\theta_0, \hat{\zeta}_0)$ satisfies $V_{ff}^{-1/2} \hat{X}_T (\cdot) \Rightarrow W (\cdot) - sPW (1)$, where $P = V_{ff}^{-1/2} \Gamma (\Gamma^\prime V_{ff}^{-1} \Gamma)^{-1} \Gamma V_{ff}^{-1/2}$ for some full rank $k \times p_\zeta$ matrix $\Gamma$.

Parts (i) and (ii) of Assumption 6 correspond to Assumption 1 above. Part (iii) is a special case of (Sowell 1996, Theorem 1) and (Li and Mueller 2009, Theorem 1 (iii)), where sufficient conditions for it can be found, for example $\hat{\zeta}_0 \xrightarrow{p} \zeta_0$, $f_t (\theta_0, \zeta)$ is differentiable wrt $\zeta$ in some neighborhood of $\zeta_0$ a.s. for all $t \leq T$ and $T > 0$, and
where

\[ T^{-1} \sum_{t=1}^{[sT]} \partial f_t(\theta_0, \zeta_0) / \partial \zeta' - s\Gamma \xrightarrow{p} 0. \]

These conditions are standard in the stability literature, see (Andrews 1993), (Sowell 1996) and (Elliott and Mueller 2006). Note that Assumption 6 rules out instabilities in \( \zeta \). This can be relaxed to allow for local instabilities in \( \zeta \) by appropriate modification of the stab-AR part in the gen-AR statistic, following the approach of (Li and Mueller 2009, Section 2.3), though the resulting tests need not have any asymptotic optimality properties.

Replace \( f_t(\theta_0) \) by \( f_t(\theta_0, \hat{\zeta}_0) \) in the definition of the generalized AR statistics described after equation (7) above, and denote the resulting statistics by

\[ \text{gen-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) = \frac{c}{1 + \bar{c}} \text{GMM-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) + \text{stab-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) \]

The following result shows that the asymptotic distribution of the resulting exp/ave/qLL-AR tests evaluated at \( \hat{\zeta}_0 \) is the same as in the case when there are no estimated nuisance parameters, given in Theorems 1 and 2 except for a degree of freedom correction in the limiting \( \chi^2 \) distribution of the GMM-AR statistic.

**Theorem 5** When Assumption 6 and \( H_0 : \theta = \theta_0 \) hold, \( \text{GMM-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) \) and \( \text{stab-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) \) are asymptotically independent, \( \text{GMM-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) \xrightarrow{d} \chi^2_{k-p_{\zeta}} \) and the asymptotic distribution of \( \text{stab-AR}_{T} \left( \theta_0, \hat{\zeta}_0 \right) \) is the same as in Theorems 1 and 2.

Split-sample GMM test statistics can be defined accordingly by replacing \( f_t(\theta_0) \) by \( f_t(\theta_0, \hat{\zeta}_0) \) in the formulae (13). A suitable extension of Assumption 3 enables us to obtain their limiting distribution.

**Assumption 7** Let \( F_{sT}(\theta, \zeta) = \sum_{t=1}^{[sT]} f_t(\theta, \zeta), q_t(\theta, \zeta) = \text{vec}[\partial f_t(\theta, \zeta) / \partial \theta'] \) and \( Q_{sT}(\theta, \zeta) = \sum_{t=1}^{[sT]} q_t(\theta, \zeta) \). We assume that (i)

\[ T^{-1/2} \left\{ \begin{bmatrix} F_{sT}(\theta_0, \hat{\zeta}_0) \\ Q_{sT}(\theta_0, \hat{\zeta}_0) \end{bmatrix} - E \left[ \begin{bmatrix} F_{sT}(\theta_0, \zeta_0) \\ Q_{sT}(\theta_0, \zeta_0) \end{bmatrix} \right] \right\} \xrightarrow{} \begin{pmatrix} V_{ff} & V_{fq} \\ V_{qf} & V_{qq} \end{pmatrix}^{1/2} \begin{pmatrix} W_f(s) - sPW_f(s) \\ W_q(s) \end{pmatrix} \]

where \( W_f \) and \( W_q \) are independent standard \( k \times 1 \) and \( kp \times 1 \) Wiener processes and \( P = V_{ff}^{-1/2} \Gamma (\Gamma' V_{ff}^{-1} \Gamma)^{-1} \Gamma' V_{ff}^{-1/2} \) for some full rank \( k \times p_{\zeta} \) matrix \( \Gamma \); and (ii) \( W_T \xrightarrow{p} V_{ff}^{-1} \) and there exist consistent estimators \( \hat{V}_{fq}(\theta_0) \) and \( \hat{V}_{qq}(\theta_0) \) of \( V_{fq} \) and \( V_{qq} \).

The part of Assumption 7 that refers to \( F_{sT}(\theta_0, \hat{\zeta}_0) \) is identical to Assumption 6. The part referring to the Jacobian \( Q_{sT}(\theta_0, \hat{\zeta}_0) \) has no counterpart in the stability
literature. Relative to Assumption 3, which restricts the behavior of the Jacobian at $\zeta_0$, this assumption can be verified with the additional requirement $\sup_{s \in [0,1]} \|Q_sT(\theta_0, \hat{\zeta}_0) - Q_sT(\theta_0, \zeta_0)\| \xrightarrow{p} 0$. This holds trivially in models that are linear in $\zeta$ (e.g., when $\zeta$ is an intercept), since $Q_sT(\theta_0, \hat{\zeta}_0) = Q_sT(\theta_0, \zeta_0)$.

**Theorem 6** When Assumption 7 and $H_0 : \theta = \theta_0$ hold, the limiting distributions of the split-sample GMM-AR, KLM, JKLM and MLR statistics evaluated at $\hat{\zeta}_0$, are given by

\[ \text{split-AR}_T \left( \theta_0, \hat{\zeta}_0, \bar{v}(\theta_0) \right) \xrightarrow{d} \psi_p + \psi_{2k-p-p_\zeta}, \]

\[ \text{split-KLM}_T \left( \theta_0, \hat{\zeta}_0, \bar{v}(\theta_0) \right) \xrightarrow{d} \psi_p, \]

\[ \text{split-JKLM}_T \left( \theta_0, \hat{\zeta}_0, \bar{v}(\theta_0) \right) \xrightarrow{d} \psi_{2k-p-p_\zeta}, \]

\[ \text{split-MLR}_T \left( \theta_0, \hat{\zeta}_0, \bar{v}(\theta_0) \right) \mid r_{k_\theta}(\theta_0), \bar{v}(\theta_0) \xrightarrow{d} \frac{1}{2} \left[ \psi_p + \psi_{2k-p-p_\zeta} - r_{k_\theta}(\theta_0, \bar{v}(\theta_0)) + \sqrt{\left( \psi_p + \psi_{2k-p-p_\zeta} + r_{k_\theta}(\theta_0, \bar{v}(\theta_0)) \right)^2 - 4\psi_{2k-p-p_\zeta} r_{k_\theta}(\theta_0, \bar{v}(\theta_0))} \right], \]

where $\psi_p$ and $\psi_{2k-p-p_\zeta}$ are independently distributed $\chi^2(p)$ and $\chi^2(2k-p-p_\zeta)$ 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 the power of the generalized Anderson-Rubin tests derived in Section 3 to the power envelope in the case of single-break and persistent time variation (PTV) alternatives, equations (20) and (22), respectively. Next, we compare the power of generalized Anderson-Rubin tests to the split sample tests in the linear IV model given by Example 3 above. The numerical analysis in [Andrews, Moreira, and Stock 2006] provides useful benchmarks, e.g., the power of the ‘oracle’ test when break dates are known. In all experiments, asymptotic results are approximated using a sample of 2000 observations, and the number of Monte Carlo simulations is 20,000.
4.1.1 Power under different alternatives

The power curves are based on testing the null hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta \neq \theta_0$, with moment conditions $E [X(s)] = 0$, where $X(s) = W(s) + \int_0^s m(\theta, r) \, dr$, and $m(\theta, s)$ is specified according to assumption 4 or 5 for the one-break and PTV alternatives, respectively. We set the parameters $\bar{c}$ and $\tilde{c}$ which scale the variance of $\bar{m}$ and $\tilde{m}$, respectively, as $\bar{c} = (\theta - \theta_0)^2 \omega$ and $\tilde{c} = (\theta - \theta_0)^2 (1 - \omega)$ where $\omega \in [0, 1]$ is a parameter that puts weights on the full-sample restriction relative to the stability restriction. Without loss of generality, we set $\theta_0 = 0$ in the computations, and due to symmetry, we only report the power curves for $\theta \geq \theta_0$. When $\omega = 0$, the full-sample statistic $X(1)$ is ancillary and when $\omega = 1$, the subsample statistics $X(s) - sX(1)$ are ancillary. We use significance level of 5% and we set $k = 1$ (this is the dimension of the data $X$).

The left and right hand columns of Figure 1 plot the power curves of the $GMM-AR_T$, $ave-AR_T$, $exp-AR_T$ and $qLL-AR_T$ tests for the cases of a single break at unknown point and of the persistent time variation, respectively. We also plot the power envelope constructed using a sequence of point-optimal tests for each alternative. As expected, when the information is coming only from the stability restriction ($\omega = 0$), the $exp-AR_T$ and the $qLL-AR_T$ power curves are the closest to the power envelope in the single break and in the PTV cases, respectively. The power of the $ave-AR$ test is lower, while the $GMM-AR_T$ test has trivial power. It is interesting to note that the $qLL-AR_T$ has very similar power to the $exp-AR_T$ in the case of a single break, and it dominates the latter in case of PTV. In the case when the information is split evenly across the two statistics, all the generalized Anderson-Rubin tests have power very close to the envelope, as expected, since they are designed with $\bar{c} = \tilde{c}$. In the other polar case when there is no instability ($\omega = 1$), the full-sample the $ave-AR_T$ power dominates the $exp-AR_T$ and $qLL-AR_T$, and the $GMM-AR_T$ test is optimal here, since $k = 1$. However, even in the case $\omega = 1$, the loss of power of the generalized AR tests relatively to the power envelope is noticeably small. Analogous results (not reported here) are obtained when the number of instruments (i.e., the dimension of $X$) is higher.

4.1.2 Generalized Anderson-Rubin versus split-sample tests

We compare the power of the generalized Anderson-Rubin tests to the split-sample GMM-AR, KLM and MLR tests in the linear IV regression model with time-varying
Figure 1: Asymptotic power curves of point-optimal and generalized AR - tests. $k = 1$, 20,000 Monte Carlo replications.
first stage. The data generating process is given by equation (24) 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 \( \lambda = \sum \Pi'Z_tZ_t\Pi \). We can think of the contribution of each observation to the identification of \( \theta \) as being equal to \( \Pi'Z_tZ_t\Pi \), but when \( \Pi_t \) is time-varying, the incremental information is \( \Pi'_tZ_tZ_t\Pi_t \), and so the total amount of information is \( \lambda = \sum_{t=1}^{T} \Pi'_tZ_tZ_t\Pi_t \).

We consider the case of a single break in \( \Pi_t \). Because all of the statistics we consider are invariant to the class of transformation \( Z_t \rightarrow Z_tG \), for any nonsingular matrix \( G \), we can set all but the first entry of \( \Pi_t \) to zero, without loss of generality. Let the first entry of \( \Pi_t \) and \( Z_t \) be \( \pi_t \) and \( z_t \), respectively, with \( \pi_t = \pi_1 \) for \( t < [\tau T] \) and \( \pi_2 \) for \( t \geq [\tau T] \). Then, \( \lambda = \sum_{t=1}^{T} \pi_t^2z_t^2 = \pi_1^2 \sum_{t=1}^{[\tau T]} z_t^2 + \pi_2^2 \sum_{t=[\tau T]+1}^{T} z_t^2 \approx \left[ \pi_1^2 \tau + \pi_2^2 (1 - \tau) \right] Q_{zz} \),

where \( Q_{zz} = \sum_{t=1}^{T} z_t^2 \). We can measure the information in the full-sample and the stability restrictions by \( \lambda_F = [\pi_1 \tau + \pi_2 (1 - \tau)]^2 Q_{zz} \) 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).

We draw the asymptotic power curves of the GMM-MLR_{T}, ave-AR_{T} and ave-AR_{B}^T tests when the distribution of information between full-sample and stability restrictions, \( \lambda_F \) and \( \lambda_S \), respectively, is varied. Figures 2 and 3 present the power curves with the number of instruments \( k \) equal to 2 and 5, respectively. Since the GMM-MLR_{T} test is not invariant to \( \rho \), we report results for two cases \( \rho = 0.20 \) and 0.95.

The GMM-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-AR_{B}^T. We observe that the ave-AR_{T}, has good power in all cases, and its power is relatively insensitive to the source of information.

Next, we compare the power of the generalized AR tests with the split-sample tests based on estimating the break date. Figures 4 and 5 compare the power curves of the exp-, ave-, and qLL-AR_{T} tests with the split-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
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).
\( \lambda_S = 0.00, \rho = 0.20 \)

\( \lambda_S = 2.50, \rho = 0.20 \)

\( \lambda_S = 5.00, \rho = 0.20 \)

\( \lambda_S = 0.00, \rho = 0.95 \)

\( \lambda_S = 2.50, \rho = 0.95 \)

\( \lambda_S = 5.00, \rho = 0.95 \)

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 \( \tau = 0.5; k = 5 \). The curves correspond to: GMM-MLR<sub>T</sub> (thin solid line); \( ave-AR_T^B \) (dashed line); \( ave-AR_T \) (dotted line).
\( \lambda_S = 0.00, \rho = 0.20 \)

\( \lambda_S = 0.00, \rho = 0.95 \)

\( \lambda_S = 5.00, \rho = 0.20 \)

\( \lambda_S = 5.00, \rho = 0.95 \)

Figure 4: 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); \( \text{split-MLR}_T \) (thin dashed line); \( \text{exp-AR}_T \) (dashed line); \( \text{ave-AR}_T \) (dotted line); and \( qLL-AR_T \) (dash-dot line).
Figure 5: 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 $\tau = 0.5$; $k = 5$. The curves correspond to: MLR$_T$—"oracle" (thin solid line); split-MLR$_T$ (thin dashed line); exp-AR$_T$ (dashed line); ave-AR$_T$ (dotted line); and qLL-AR$_T$ (dash-dot line).
that the break date is known. We see that no test dominates the others in terms of power. Unreported results, available on request, show that this is also true for the split-sample $AR$ and $KLM$ statistics. 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.

4.1.3 Power loss when there is no instability

One may worry about the possible loss of power of the proposed tests relative to standard GMM tests when the underlying data generating process is stable, since the stability restrictions will not be informative in this case. This issue is entirely analogous to the use of irrelevant instruments in standard GMM testing. To gain some insight into this problem, we compare the power loss of each of the proposed tests relative to the optimal test in the case of no instability.

We use the linear IV model described in the previous subsection, with concentration parameter $\lambda_F = 5$ and no instability. For each of the proposed tests, we calculate the average and maximum difference of the power curves at the 5% level of significance from the power of the optimal test. For comparison, we also report the corresponding results for the Anderson-Rubin and conditional LR (MLR) test with twice as many instruments. The intuition is that our proposed tests can be thought of as using twice as many instruments, which are all irrelevant in the case of no instability, so we want to compare the power loss of this compared to the conventional loss of power due to irrelevant instruments.

Table 1 reports the results for the case of one instrument ($k = 1$), and for two cases of endogeneity: $\rho = 0.2$ and 0.95. Not surprisingly, the power loss is the smallest for the $ave-AR_T$ test, since this optimal against small instability. It is on average 0.018 and at most 0.05. Moreover, this power loss compares favorably with the loss associated with the $AR_T$ and $MLR_T$ tests that use 2 instruments one of which is irrelevant. The $exp-AR_T$ and $qLL-AR_T$ suffer somewhat larger power losses, as do the $split-AR_T$, $-K_T$ and $-MLR_T$ tests.

Table 2 reports the results for $k = 2$. In this case, the relevant benchmark test is the conditional LR test, which is close to the relevant power envelope, see (Andrews, Moreira, and Stock 2006). The results show that the power loss is increasing in the number of instruments. The ordering of the tests remains the same as in the just-

\footnote{The Anderson-Rubin-type tests are invariant wrt $\rho$.}
identified case, except for the split-MLR$_T$ test, which outperforms the ave-AR$_T$ test when $\rho = 0.95$. The ave-AR$_T$ and exp-AR$_T$ tests suffer smaller power loss than the Anderson Rubin test with $2k$ instruments, though they seem to be doing worse than the corresponding MLR test with $k = 4$.

All in all, these results indicate that the loss of power associated with the proposed tests when there is no instability is similar to the power loss caused by using extra (irrelevant) instruments in standard tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>$\rho = 0.20$</th>
<th>$\rho = 0.95$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>average</td>
<td>maximum</td>
</tr>
<tr>
<td>ave-AR$_T$</td>
<td>0.018</td>
<td>0.050</td>
</tr>
<tr>
<td>exp-AR$_T$</td>
<td>0.037</td>
<td>0.103</td>
</tr>
<tr>
<td>qLL-AR$_T$</td>
<td>0.071</td>
<td>0.185</td>
</tr>
<tr>
<td>split-AR$_T$</td>
<td>0.040</td>
<td>0.107</td>
</tr>
<tr>
<td>split-KLM$_T$</td>
<td>0.173</td>
<td>0.392</td>
</tr>
<tr>
<td>split-MLR$_T$</td>
<td>0.060</td>
<td>0.172</td>
</tr>
<tr>
<td>MLR$_T$ $k = 2$</td>
<td>0.026</td>
<td>0.080</td>
</tr>
<tr>
<td>AR$_T$ $k = 2$</td>
<td>0.040</td>
<td>0.106</td>
</tr>
</tbody>
</table>

Table 1: The table lists the average and maximum differences in the rejection probabilities between each test and the Anderson-Rubin test in a linear IV model with a single endogenous regressor and one instrument. The significance level is 5%, and the concentration parameter is 5.

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, 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 $\varepsilon_t$ are jointly Normal.
with zero mean, variances $\sigma^2_\varepsilon$, $\sigma^2_v$ and covariance $\sigma_{\varepsilon v}$.

In this simple version of new Keynesian model, the parameter $\beta$ is a discount factor, while, $\gamma = \frac{(1-\alpha)(1-\beta\alpha)}{\alpha}$, where $\alpha$ represents the price rigidity in the economy. The parameters are set to $\beta = 1$ and $\alpha = \frac{2}{3}$, ($\gamma = \frac{1}{6}$), while the remaining nuisance parameters are calibrated to quarterly post-1960 US data. We find $\rho_1 = 0.9$, $\rho_2 = 0.063$, $\sigma^2_\varepsilon = .3$, $\sigma^2_v = .011$ and $\sigma_{\varepsilon v} = -.012$. 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 $\rho_2$ goes from $-0.09$ to $0.21$, with a standard error of $0.15$. We therefore set $\rho_{2,t} = 0.063 + 0.075 \times \kappa(-1)^{1\left(t<1984q1\right)}$ and $\rho_{1,t} = \rho(1 + \rho_{2,t})$, with $\rho = 0.9$. The parameter $\kappa$ 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 $\rho_2$.

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 $\sigma^2_\varepsilon$ falls significantly after 1984, although $\sigma^2_v$ and $\sigma_{\varepsilon v}$ remain constant over the two periods. Specifically, $\sigma^2_\varepsilon$ drop from 0.5 before 1984q1 to 0.1 thereafter, with a standard error of 0.068. It is important to check the implications of a change in

<table>
<thead>
<tr>
<th>Test</th>
<th>$\rho = 0.20$</th>
<th></th>
<th>$\rho = 0.95$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>average</td>
<td>maximum</td>
<td>average</td>
</tr>
<tr>
<td>ave-$AR_T$</td>
<td>0.035</td>
<td>0.107</td>
<td>0.046</td>
</tr>
<tr>
<td>exp-$AR_T$</td>
<td>0.059</td>
<td>0.167</td>
<td>0.070</td>
</tr>
<tr>
<td>qLL-$AR_T$</td>
<td>0.098</td>
<td>0.250</td>
<td>0.109</td>
</tr>
<tr>
<td>split-$AR_T$</td>
<td>0.062</td>
<td>0.174</td>
<td>0.073</td>
</tr>
<tr>
<td>split-$KLM_T$</td>
<td>0.208</td>
<td>0.469</td>
<td>0.064</td>
</tr>
<tr>
<td>split-$MLR_T$</td>
<td>0.069</td>
<td>0.180</td>
<td>0.032</td>
</tr>
<tr>
<td>MLR$_T$ $k = 4$</td>
<td>0.033</td>
<td>0.097</td>
<td>0.016</td>
</tr>
<tr>
<td>AR$_T$ $k = 4$</td>
<td>0.060</td>
<td>0.168</td>
<td>0.070</td>
</tr>
</tbody>
</table>

Table 2: The table lists the average and maximum differences in the rejection probabilities between each test and the CLR test in a linear IV model with a single endogenous regressor and two instruments. The significance level is 5%, and the concentration parameter is 5.
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 set $\sigma^2_{\varepsilon,t} = 0.3 + 0.034 \times \phi (-1)^{1_{t<1984t}1}$, where the scalar $\phi$ is used to vary the magnitude of the change in the variance in terms of standard errors from zero, with $\phi = 6$ corresponding to the subsamples estimates.\(^7\)

Table 3 reports the null rejection frequencies of the proposed tests of $H_0 : \alpha = \frac{2}{3}$, in the model (5) for a sample of $T = 180$, computed using 20,000 Monte Carlo replications. The instruments used are $x_{t-1}$ and $x_{t-2}$, and the variance estimator used is (Newey and West 1987) with prewhitening.\(^8\) 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 6.

<table>
<thead>
<tr>
<th></th>
<th>$\kappa = 0$</th>
<th>$\kappa = 2$</th>
<th>$\kappa = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi = 0$</td>
<td>$\phi = 6$</td>
<td>$\phi = 0$</td>
</tr>
<tr>
<td>Nom.level</td>
<td>10% 5%</td>
<td>10% 5%</td>
<td>10% 5%</td>
</tr>
<tr>
<td>$GMM-AR_T$</td>
<td>7.53 3.58</td>
<td>7.17 3.68</td>
<td>8.07 4.12</td>
</tr>
<tr>
<td>$exp-AR_T$</td>
<td>7.13 3.91</td>
<td>9.16 5.46</td>
<td>8.12 4.50</td>
</tr>
<tr>
<td>$qLL-AR_T$</td>
<td>5.83 2.83</td>
<td>7.80 4.25</td>
<td>6.88 3.35</td>
</tr>
<tr>
<td>$split-AR_T$</td>
<td>6.85 3.58</td>
<td>7.94 4.13</td>
<td>7.75 4.08</td>
</tr>
<tr>
<td>$split-KJ_T$</td>
<td>5.62 2.61</td>
<td>7.25 3.96</td>
<td>6.34 3.08</td>
</tr>
<tr>
<td>$split-MLR_T$</td>
<td>5.03 2.16</td>
<td>7.09 3.48</td>
<td>5.76 2.48</td>
</tr>
</tbody>
</table>

Table 3: 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 tests are close to their nominal level. Some tests appear to be undersized, the most severe being the $split-MLR_T$ test. Only the $exp-AR_T$ test appears somewhat oversized when there are large changes both in variance in the coefficients ($\kappa = 4$, $\phi = 6$), but the size distortion is modest: 12.26% at the 10% nominal level, and 7.5% at nominal 5% level. The important message is that the size is almost unaffected by the changes in the coefficients or in the variance.

\(^7\)Note that the estimates of $\sigma^2_{\varepsilon}$ from the data are conditional on the assumed values of the structural parameters $\beta, \alpha$, since $\sigma^2_{\varepsilon}$ is not independently identified.

\(^8\)For the exp- and ave-AR tests we use subsample estimates of the variance.
Figure 6 reports further evidence on the size of the tests as the magnitude of the break in the coefficients $\rho_1$ and $\rho_2$ varies. The figure plots the rejection frequencies of 5% nominal tests as a function of the magnitude of the break in the coefficients in standard error units $\kappa$. The $GMM-AR_T$, $exp-AR_T$, $ave-AR_T$ and $qLL-AR_T$ are shown on the left panel and the $split-MLR_T$ and $split-KLM_T$ tests are shown on the right panel. In this calculation, the variance is kept fixed ($\phi = 0$) and we consider changes of up to 6 standard errors in the coefficients. The results indicate that the size is affected very little by the magnitude of the instability, never exceeding 7% at the 5% nominal level.

In Figure 7, we report the corresponding rejection frequencies as functions of changes in the variance $\sigma^2_{\varepsilon}$, in $\phi$ standard error units. The top panels show the case when only the variance of the shock changes, i.e., the coefficients are constant, and the bottom panels report results when the coefficients also change by two standard errors (as in the data). Even though the size of the tests increases with $\phi$, the increase is modest, and becomes noticeable only when the break is more than 6 standard errors from zero. Even in those cases, the $exp-AR_T$ test is the only test that exhibits noticeable over-rejection, while the $qLL-AR_T$ and $ave-AR_T$ tests appear little affected. These results are not particularly surprising, in view of the evidence reported by [Hansen 2000], who
studied this issue in a related context.

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):

\[ \pi_t = \beta E_t \pi_{t+1} + \varphi \pi_{t-1} + \gamma x_t + \varepsilon_t, \]  

(30)

where \( \pi_t \) is inflation, \( \beta = \delta / (1 + \delta \varrho) \), \( \delta \) is a discount factor, \( \varrho \) is the fraction of prices that are indexed to past inflation when they cannot be optimally reset, \( \alpha \) is the probability that a price will be fixed in a given period, \( \varphi = \varrho / (1 + \delta \varrho) \), and \( \gamma = (1 - \alpha) (1 - \delta \alpha) / (\alpha (1 + \delta \varrho)) \). The variable \( x_t \) is a measure of economic slack, and we shall proxy it using the labor share, following (Galí and Gertler 1999) and (Sbordone 2002).

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

\[ \Delta \pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + \gamma x_t + \varepsilon_t, \]  

(31)

and we obtain confidence sets for \( \alpha, \varrho \) using two lags of \( \Delta \pi_t \) and \( x_t \) as instruments. We allow for an unrestricted constant in equation (31) 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 \((\alpha, \varrho)\) are constructed by inverting the various identification robust tests, and they are plotted in Figures 8, 9 and 10. The following conclusions can be drawn from Figures 8 and 9.

First, the full-sample 95%-level confidence intervals for the coefficient \( \varrho \) 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

---

9Inflation is measured by \( \pi_t = 100 \times \ln \left( \frac{GDP}{GDP_{t-1}} \right) \), where GDP 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 7: Rejection probability of 5% significance level tests of $H_0: \gamma = \frac{1}{6}$ in the NKPC model $y_t = \beta E_t(y_{t+1}) + \gamma x_t + \varepsilon_t$, where $x_t = \rho_{1,t}x_{t-1} + \rho_{2,t}x_{t-2} + v_t$, the sample size is $T=180$ and the variance of the structural error $\varepsilon_t$ and the reduced form coefficients $\rho_{1,t}$ and $\rho_{2,t}$ change by $\phi$ and $\kappa$ s.e.’s, respectively, in the middle of the sample. The top panels are for the case $\kappa = 0$ and the bottom panels for the case $\kappa = 2$. Computed using 20,000 Monte Carlo replications.
robust to changes in the sample size and number of instruments (not reported).

Figure 8: $GMM-AR_T$ and generalized AR confidence sets for $\alpha$ and $\varrho$ in the NKPC. Period: 1966q1 2010q4 $\Delta \pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + (1 - \alpha)^2 \alpha (1 + \varrho) x_t + \varepsilon_t$. The forcing variable is the labor share. Instruments include two lags of $\Delta \pi$ and three lags of the labor share.

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
Figure 9: GMM-$MLR_T$ and split-$MLR_T$ confidence sets for $\alpha$ and $\varrho$ in the NKPC \[ \Delta \pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + \frac{(1-\alpha)^2}{\alpha(1+\varrho)} x_t + \epsilon_t. \] Period: 1966q1 2010q4. Instruments include two lags of $\Delta \pi$ and three lags of the labor share.

(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 $\alpha$ is also smaller, though the difference is not as dramatic as in the case of $\varrho$ (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 the pure stability tests ave-$AR^B_T$, exp-$AR^B_T$ and qLL-$AR^B_T$. These are reported in Figure 10 where the GMM-$AR_T$ 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.

5.1 Robustness checks

We undertake a number of robustness checks to examine the robustness of the above results to possible violations of the underlying assumptions. In particular, we examine the impact of the great moderation (sharp drop in volatility of macroeconomic shocks after 1984), the possibility of time-varying trend inflation (Cogley and Sbordone 2008),
Figure 10: GMM-AR$_T$ and stability-only confidence sets for $\alpha$ and $\rho$ in the NKPC
$\Delta \pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + \frac{(1-\alpha)^2}{\alpha(1+\rho)} x_t + \varepsilon_t$. Period: 1966q1 2010q4. The forcing variable is the labor share. Instruments include two lags of $\Delta \pi$ and three lags of the labor share.
and the possibility of autocorrelated markup shocks, which would render the previous moment conditions invalid.

5.1.1 Results in the post-1984 sample

Figures 11 and 12 report full-sample AR, generalized AR, full sample MLR and split-sample MLR confidence sets on the parameters \((\alpha, \varrho)\) in the NKPC (31). As in the results of reported in Figures 8 and 9 the confidence sets that exploit information in the stability restrictions are smaller than their GMM counterparts. Moreover, the results are fairly similar to the results obtained over the 1966-2010 sample. Therefore, our conclusions appear to be robust to possible break in the variance in the early 1980s. This finding is consistent with monte Carlo evidence given above that even large changes in the variance will have modest impact on the size of the tests.

5.1.2 Trend inflation

Cogley and Sbordone (2008) show that if there is time-varying trend inflation, \(\bar{\pi}_t\) say, the coefficients in the NKPC are potentially time-varying. Even though the gen-AR tests have power against time-variation of the parameters, the resulting confidence sets may not be empty if the time-variation is not sufficiently strong. In other words, there is a concern that the generalized confidence sets may be misleadingly tight. We investigate this problem by considering an adaptation of the model to allow for local time-variation in \(\bar{\pi}_t\), i.e., \(\bar{\pi}_t = \bar{\pi} + O\left(\frac{T^{-1/2}}{2}\right)\).\(^{10}\) Under this assumption, we can rewrite the model in terms of three stable structural parameters, \(\alpha, \varrho\) and \(\mu\), the steady state value of the markup, and we can consider moment vectors that are orthogonal to the time-varying terms that involve \(\bar{\pi}_t\), details of this are given in the supplementary appendix.

Figures 13 and 14 give the respective 3D confidence of \((\alpha, \varrho, \mu)\). The confidence sets are noticeably wider than the 2D confidence sets reported earlier. However, it is still the case that the confidence sets that use the stability restrictions are tighter than the corresponding GMM sets. To emphasize that point further, Table 4 reports the fraction of the total volume in the cube defined by the grid in these pictures that is

\(^{10}\)Cogley and Sbordone (2008) model time variation using a time-varying parameter VAR, which also has this property.
Figure 11: GMM-AR$_T$ and generalized AR confidence sets for $\alpha$ and $\varrho$ in the NKPC. Period: 1984q1 2010q4 $\Delta\pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + \frac{(1-\alpha)^2}{\alpha(1+\varrho)} x_t + \varepsilon_t$. The forcing variable is the labor share. Instruments include two lags of $\Delta\pi$ and three lags of the labor share.
Figure 12: GMM-MLR<sub>T</sub> and split-MLR<sub>T</sub> confidence sets for α and ϱ in the NKPC
Δπ<sub>t</sub> = βE<sub>t</sub> (π<sub>t+1</sub> − π<sub>t−1</sub>) + \frac{(1−α)^2}{α(1+φ)} x<sub>t</sub> + ε<sub>t</sub>. Period: 1984q1 2010q4. Instruments include
two lags of Δπ and three lags of the labor share.

covered by the confidence sets.

5.1.3 Autocorrelated markup shocks
With autocorrelated markup shocks, the moment conditions need to be modified. We consider here the case of first-order autocorrelation in the error ε<sub>t</sub> in (31), i.e., ε<sub>t</sub> = φε<sub>t−1</sub> + ε<sub>t</sub>, so that the correct moment conditions are given by E (Z<sub>t</sub>ε<sub>t</sub>) = 0. 3D confidence sets on (α, ϱ, φ) can be obtained as before. In the interest of space, we only report the volume of the various confidence sets in table 5. The confidence sets that use stability restrictions are considerably smaller than their GMM counterparts.

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
Figure 13: GMM-AR$_T$ and generalized AR confidence sets for $\alpha$, $\varrho$ and $\mu$ in the NKPC with time-varying trend inflation. Period: 1966q1 2010q4 $\Delta \pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + (1-\alpha)^2 \frac{\alpha}{\alpha(1+\varrho)} x_t + \varepsilon_t$. The forcing variable is the labor share. Instruments include two lags of $\Delta \pi$ and three lags of the labor share.
Figure 14: GMM-MLR$_T$ and split-MLR$_T$ confidence sets for $\alpha$, $\varphi$ and $\mu$ in the NKPC with time-varying trending inflation. $\Delta \pi_t = \beta E_t (\pi_{t+1} - \pi_{t-1}) + \frac{(1-\alpha)^2}{\alpha(1+\varphi)} x_t + \varepsilon_t$. Period: 1966q1 2010q4. Instruments include two lags of $\Delta \pi$ and three lags of the labor share.

<table>
<thead>
<tr>
<th>(Volume of Confidence Region)/(Volume of the Cube)</th>
<th>(in percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tests</td>
<td>95%</td>
</tr>
<tr>
<td>GMM-AR</td>
<td>33.02</td>
</tr>
<tr>
<td>GMM-K</td>
<td>30.90</td>
</tr>
<tr>
<td>GMM-MLR</td>
<td>31.93</td>
</tr>
<tr>
<td>ave-AR</td>
<td>24.97</td>
</tr>
<tr>
<td>qLL-AR</td>
<td>23.76</td>
</tr>
<tr>
<td>stab-ave</td>
<td>28.80</td>
</tr>
<tr>
<td>stab-exp</td>
<td>17.21</td>
</tr>
<tr>
<td>stab-qLL</td>
<td>26.20</td>
</tr>
<tr>
<td>split-AR</td>
<td>24.73</td>
</tr>
<tr>
<td>split-K</td>
<td>20.21</td>
</tr>
<tr>
<td>split-MLR</td>
<td>21.02</td>
</tr>
</tbody>
</table>

Table 4: Volume of various 3D confidence sets relative to the parameter space in the NKPC model with time-varying trend inflation.
(Volume of Confidence Region)/(Volume of the Cube) (in percentage)

tests | 95%  | 90%  | 85%  | 80%
---|---|---|---|---
GMM-AR | 51.13 | 44.74 | 40.34 | 33.86|
GMM-K | 61.45 | 50.24 | 41.04 | 34.59|
GMM-MLR | 53.88 | 43.83 | 38.62 | 32.03|
ave-AR | 19.51 | 17.55 | 16.06 | 14.99|
exp-AR | 13.92 | 11.70 | 10.44 | 9.58|
qLL-AR | 18.42 | 14.98 | 13.61 | 12.77|
stab-ave | 18.52 | 16.83 | 15.65 | 14.84|
stab-exp | 13.15 | 11.08 | 9.64 | 8.27|
stab-qLL | 17.34 | 14.79 | 13.47 | 12.62|
split-AR | 25.39 | 18.40 | 15.19 | 12.24|
split-K | 58.81 | 47.09 | 38.82 | 33.52|
split-MLR | 26.97 | 17.87 | 13.66 | 11.13|

Table 5: Volume of various 3D confidence sets relative to the parameter space in the NKPC model with autocorrelated markup shocks.

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.

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


A Appendix: proofs

Proof of Theorem 1 By Assumption 1 under $H_0$, $V_X^{-1/2}X_T(1) \Rightarrow W(1)$ and $GMM-AR_T(\theta_0) = X_T(1)'V_X^{-1}X_T(1) + o_p(1) \Rightarrow \psi_k$. Also,

$$
\tilde{S}_T(\theta_0, \tau) = X_T(\tau)'X_T(\tau) + [X_T(1) - X_T(\tau)]'V_X^{-1}[X_T(1) - X_T(\tau)]
- X_T(1)'V_X^{-1}X_T(1) + o_p(1)
\frac{[X_T(\tau) - \tau X_T(1)]'V_X^{-1}[X_T(\tau) - \tau X_T(1)]}{\tau(1 - \tau)} + o_p(1)
$$

and $V_X^{-1/2}[X_T(\tau) - \tau X_T(1)] \Rightarrow \tilde{W}(\tau)$, which is independent of $W(1)$ and $\tilde{S}_T(\theta_0, \tau) \Rightarrow \tilde{\psi}_k(\tau)$. By the Neyman Pearson lemma, the test function

$$
1 \left\{ \frac{\bar{c}}{1 + \bar{c}} W(1)'W(1) + 2 \log \int \exp \left[ \frac{1}{2} \frac{\bar{c}}{1 + \bar{c}} \frac{\tilde{W}(\tau)'\tilde{W}(\tau)}{\tau(1 - \tau)} \right] d\nu_\tau > cv \right\}
$$

maximizes WAP in the limiting problem $H_0 : dX(s) = V_X^{1/2}dW(s)$ against $H_1 : dX(s) = m(\theta, s) + V^{1/2}dW(s)$ and it is continuous at almost all realizations of $W$. Asymptotic efficiency then follows from (Mueller 2008, Theorem 1).

Proof of Theorem 2 $\sum_{i=1}^k \hat{v}_i(M_e - G_c) \hat{v}_i \Rightarrow \psi_c$ follows from the consistency of $\hat{V}_{ff}(\theta_0)$ and the FCLT on $F_T(\theta_0)$ by Lemma 6 of Elliott and Mueller (2006). Independence of $\tilde{\psi}_k$ and $\psi_c$ follows from the asymptotic independence between $\sqrt{T}\hat{V}_{ff}(\theta_0)^{-1/2}$ $F_T(\theta_0)$ and $[I_k \otimes (M_e - G_c)] \hat{v}$, where $\hat{v} = (\hat{v}_1', \ldots, \hat{v}_k')'$, which is a direct consequence of Assumption 1 under $H_0$. Result (23) then follows from the continuous mapping theorem. Finally, asymptotic efficiency follows from (Mueller 2008, Theorem 1).

Proof of Lemma 1 The proof is analogous to (Andrews, Moreira, and Stock 2006, Lemma 2). Parts 1 and 2 follow from the fact that $Z(s)$ is nonstochastic and $V$ is Gaussian. For part 3 note that, for every $s_1, s_2$, $Z(s_1)'Yb_0$ and $Z(s_2)'Y\Omega^{-1}a_0$ are jointly normal, and their covariance is $cov(\hat{Z}(s_1)'Yb_0, \hat{Z}(s_2)'Y\Omega^{-1}a_0) = \sum_{t=1}^T Z_t(s_1)'Z_t(s_2)\sigma(Yb_0, Y\Omega^{-1}a_0) = \sum_{t=1}^T Z_t(s_1)'Z_t(s_2)b_0'\Omega^{-1}a_0 = 0$. 

56
Proof of Theorem 3  Since the random functions \( \overline{F}(\cdot) \) and \( \overline{D}(\cdot) \) are independent of each other, by Lemma 1 and \( \overline{\tau} \) only depends on \( \overline{D}(\cdot) \) by equation (28), it follows that \( \overline{\tau} \) is also independent of \( F(\cdot) \). Therefore, under \( H_0 \), Lemma \( \square \) part 1 implies that the (conditional) distribution of \( F(\overline{\tau}) \) is Gaussian with zero mean and variance matrix \( I_k \). Part 1 follows immediately. For part 2, note that conditional on \( D(\overline{\tau}) \) and \( \overline{\tau} \), \( D(\overline{\tau})' F(\overline{\tau}) \) is Gaussian with mean zero and variance \( D(\overline{\tau})' D(\overline{\tau}) \), and the matrix \( D(\overline{\tau})' D(\overline{\tau}) \) is invertible with probability 1. Parts 3 and 4 follow from the fact that \( J(\overline{\tau}) = F(\overline{\tau})' D(\overline{\tau}) \left[ D(\overline{\tau})' D(\overline{\tau}) \right]^{-1} D(\overline{\tau})' F(\overline{\tau}) \), where \( D(\overline{\tau}) \) is a \( k \times (k-p) \) matrix which is the orthogonal complement of \( D(\overline{\tau}) \), i.e., \( D(\overline{\tau})' D(\overline{\tau}) = 0 \), so \( D(\overline{\tau})' F(\overline{\tau}) \) and \( D(\overline{\tau})' F(\overline{\tau}) \) are independent conditionally on \( D(\overline{\tau}) \) and \( \overline{\tau} \). Part 5 now follows by combining the above results using the continuous mapping theorem.

Proof of Theorem 4  Assumption \( \square \) yields the asymptotic counterpart of Lemma \( \square \) for the linear model with fixed instruments and known variance. The asymptotic independence of \( D_T(\theta_0, \cdot) \) and \( F_T(\theta_0) \) implies \( \overline{\tau} \) will be asymptotically independent of \( F_T(\theta_0) \) as well, and so, conditional on \( \overline{\tau} \), \( F_T^1(\theta_0, \overline{\tau}) \) and \( F_T^2(\theta_0, \overline{\tau}) \) are jointly Gaussian and independent with zero mean and variance \( V_{ff} \), respectively. This establishes that \( \text{split-AR}_T(\theta_0, \overline{\tau}(\theta_0)) \overset{d}{\rightarrow} \chi^2(2k) \). The remaining results follow by direct analogy with the proof of Theorem \( \square \).

Proof of Theorem 5  Let \( \hat{X}_T^*(s) = T^{-1/2} W_T^{1/2} \sum_{t=1}^{[sT]} f_t(\theta_0, \zeta_0) \) and \( X_T^*(s) = T^{-1/2} W_T^{1/2} \sum_{t=1}^{[sT]} f_t(\theta_0, \zeta_0) \). Assumption \( \square \) (i) implies \( X_T^*(s) \Rightarrow W(s) \), while assumptions \( \square \) (ii) and (iii) imply \( \hat{X}_T^*(1) \Rightarrow MW(1) \) and \( \hat{X}_T^*(s) \Rightarrow W(s) - sPW(1) \), where \( M = I_k - P \). Hence, \( \hat{X}_T^*(s) - s\hat{X}_T^*(1) \Rightarrow W(s) - sW(1) = W(s) \). This is the same as the distribution of the statistic \( X_T^*(s) - sX_T^*(1) \) that does not involve estimation of any nuisance parameters \( \zeta \). Hence, the distribution of the stability component of the \( \text{gen-AR} \) statistics, which only involves \( X_T^*(s) - sX_T^*(1) \), is the same as in Theorems \( \square \) and \( \square \). On the other hand, \( \text{GMM-AR}_T(\theta_0, \hat{\zeta}_0) = \hat{X}_T^*(1)' \hat{X}_T^*(1) \Rightarrow W(1)' MW(1) \sim \chi^2_{k-p} \). Moreover, \( \hat{X}_T^*(s) - s\hat{X}_T^*(1) \) converges to a Brownian bridge \( W(s) \), which is independent of \( W(1) \), showing that \( \text{GMM-AR}_T(\theta_0, \hat{\zeta}_0) \) and \( \text{stab-AR}_T(\theta_0, \hat{\zeta}_0) \) are asymptotically independent.

Proof of Theorem 6  In the derivation of the split-sample statistics, \( D_T(\theta_0, \cdot) \) and \( F_T(\theta_0) \) are replaced with their counterparts that use \( f_t(\theta_0, \hat{\zeta}_0) \) instead of \( f_t(\theta_0) \) in
their definition. Denote these by $D_T \left( \theta_0, \hat{\zeta}_0, \cdot \right)$ and $F_T \left( \theta_0, \hat{\zeta}_0, \cdot \right)$. Under assumption 7, $D_T \left( \theta_0, \hat{\zeta}_0, \cdot \right)$ and $F_T \left( \theta_0, \hat{\zeta}_0, \cdot \right)$ are asymptotically independent, and since $\bar{\tau}$ only depends on $D_T \left( \theta_0, \hat{\zeta}_0, \cdot \right)$, it will be asymptotically independent of $F_T \left( \theta_0, \hat{\zeta}_0, \cdot \right)$, so we can condition on $\bar{\tau}$ to obtain the distribution of $\text{split-AR}_T \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) = \sum_{i=1}^{2} T_i^{-1} F_i^1 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) \hat{V}_i \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right)^{-1} F_i^1 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right)$.

Now, $F_1 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) = F_{\bar{\tau}} \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right)$, and $\hat{V}_1 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) \overset{p}{\to} V_{ff}$, so $\hat{V}_1 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right)^{-1/2} F_1 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) \overset{d}{\to} W (\bar{\tau}) - \bar{\tau} PW (1) = \tilde{W} (\bar{\tau}) + \bar{\tau} MW (1)$, where $\tilde{W} (\cdot)$ is a Brownian bridge. Similarly, since $F_2 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) = F_T \left( \theta_0, \hat{\zeta}_0, \cdot \right) - F_{\bar{\tau}} \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right)$, $\hat{V}_2 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right)^{-1/2} F_2^2 \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) \overset{d}{\to} MW (1) - [W (\bar{\tau}) - \bar{\tau} PW (1)] = (1 - \bar{\tau}) MW (1) - \tilde{W} (\bar{\tau})$. So,

$$\text{split-AR}_T \left( \theta_0, \hat{\zeta}_0, \bar{\tau} \right) \overset{d}{\to} \frac{\tilde{W} (\bar{\tau}) + \bar{\tau} MW (1)}{\bar{\tau}} + \frac{[1 - (1 - \bar{\tau}) MW (1) - \tilde{W} (\bar{\tau})]}{1 - \bar{\tau}} \cdot MW (1) \overset{d}{\to} \tilde{W} (\bar{\tau}) + MW (1)$$

and the first result follows by noting that $W (1)$ and $\tilde{W} (\bar{\tau})$ are independent standard normals with covariance matrices $I_k$ and $M$ is idempotent with rank $k - p_\zeta$. The remaining results can be established analogously.