Boundedly Pivotal Structural Change Tests in Continuous Updating GMM with Strong, Weak Identification and Completely Unidentified Cases

Mehmet Caner
University of Pittsburgh *

April 2005

Abstract

This paper develops structural change tests in the continuous updating GMM framework that are robust to weak identification. We propose likelihood ratio-like, Anderson-Rubin (1949), and Kleibergen (2005) type of tests. Since the limits of test statistics are not nuisance parameter free, bounds for the limit of the test statistics are derived. The bounds are nuisance parameter free and robust to identification problems. The bound for likelihood ratio-like test and Anderson-Rubin (1949) type of test is the same and depends on the number of orthogonality restrictions. The bound for Kleibergen (2005) type of test statistic is different and depends on the number of parameter restrictions. We also derive the test statistics limits' under standard strong identification in Continuous Updating Estimator. To obtain that result we derive a new weak convergence theorem for sequential empirical process with time series data. In order to understand how conservative the bound is in the worst case, we compare the limit in the bound with the limits derived under standard CUE. Simulations show that Anderson-Rubin (1949) and Kleibergen (2005) type of tests are slightly conservative, and have very good small sample properties. In the case of weak instruments, sup LM test of Andrews (1993a), which is widely used in the structural change literature, rejects the true null of parameter stability more than the nominal level, and also has low power in the weak instrument setup.

Keywords: Sequential Empirical Process, Constrained CUE, Weak Instruments, Stability.

JEL Classification: C13, C30.

---

*Department of Economics, WW Posvar Hall 4S01, Pittsburgh, PA 15260. email: caner@pitt.edu. I also thank James Stock, Bruce Hansen, Benedikt Pötscher, Fallaw Sowell, Yuichi Kitamura, and Atsushi Inoue for helpful discussions and suggestions. Comments by the participants of NBER/NSF Time Series Workshop 2001-Raleigh, Econometric Society Meetings at Atlanta 2002, Cornell, Penn State, Rochester, and University of Pennsylvania Econometric Seminars were greatly appreciated. Technical discussions with Erdem Basci, Tarik Kara, In Uck Park, and Jonathan Wright were helpful.

1 Introduction

In econometrics literature, one possible source of poor approximation of asymptotic normality to finite sample behavior of GMM estimates and test statistics is low correlation of the instruments with the first order conditions. Recently, in a seminal paper, Stock and Wright (2000) develop the weak instrument asymptotics for GMM estimators and show that the new limits provide better approximation for finite sample behavior of estimators. They also provide a new test statistic, which is called S-test. S-test is the continuous updating GMM objective function evaluated at the restriction imposed by the null. Kleibergen (2005) provides an LM-like test statistic. This test has more power than the test statistic of Stock and Wright (2000).

Many researchers in economics and finance are interested in detecting regime changes, and analyzing the stability of Euler equations. Andrews (1993a), in a path-breaking work, considers structural change tests with an unknown change point in the standard two-step GMM setup. Andrews (1993a) develops the limit theory for the structural change test based on LM-like test and likelihood ratio-like test. However, these tests are derived under the assumption that all of the parameters are identified. We want to learn whether the large sample distributions of the structural change tests can provide good finite sample approximations when there are weak instruments in the system. We also want to derive structural change tests that are robust to identification problems.

In this paper, we suggest three test statistics to test the null of no regime change against an unknown change point. These are likelihood ratio-like, Anderson-Rubin (1949), and Kleibergen (2005) type of tests. Under certain cases in weakly identified case, and in completely unidentified case, Caner (2003) shows that the test statistics are not asymptotically pivotal. However, in this paper we show that the limits are shown to be bounded with a distribution. This asymptotic bound is robust to identification problems and nuisance parameter free. In GMM with weak instruments, tests for structural change amounts to sub vector testing. In those cases we benefit from asymptotically conservative methods.

We also derive the limits for test statistics under only strong identification (standard GMM). To do that, we benefit from a result that we derive for sequential empirical processes. This result extends the iid sequential empirical process limit theorem in van der Vaart and Wellner (1996) to time series case. Next, we compare the limit bound in robust case to the limits that are derived in strong identification case. This can help us in assessing the magnitude of the mistake we may be encountering in the case of standard identification.

We conduct some simulations to understand the small sample behavior of the test statistics when the instruments are weakly correlated with the first order conditions. We realize that Anderson-Rubin (1949) and Kleibergen (2005) type of tests are only slightly conservative, and have very good small sample properties. We also show that sup LM test of Andrews (1993a), which is widely used in the literature, rejects the true null of stability much more often than the prescribed nominal level, and has also low power.

Section 2 introduces the model, main assumptions, and an important result for sequential empirical process. Section 3 introduces test statistics, and derives the distribution for the asymptotic bound. Section 4 conducts simulations. Section 5 concludes. The appendix covers the main proofs and the limit theory for the restricted partial sample estimator and the limits of test statistics proposed. Technical Appendix contains some of the basic but tedious proofs. Let Θ denote a compact parameter space, and Π denote a set whose closure is in (0,1). Let $\pi \in \Pi$, and $\Longrightarrow$ denote weak convergence of random functions with respect to the sup norm. $[T\pi]$ denotes the largest integer that is less than or equal to $T\pi$. The limits of the test statistics are also derived for two-step GMM case. These and some of the tedious proofs can be found in Caner (2003). This is available at www.pitt.edu/˜caner.

2 Partial Sample GMM Estimator with Weak Instruments

In this section we analyze the partial sample continuous updating GMM estimators. Before that we discuss the hypothesis of interest. The null hypothesis is parameter stability:

$$H_0 : \theta_t = \theta_0 \quad \text{for all} \quad t \geq 1 \quad \text{for some} \quad \theta_0 \in \Theta \subset R^n.$$

The alternative hypothesis is a one time structural change with unknown change-point $\pi \in (0, 1)$. T is the number of observations and $[T\pi]$ is the time of change, and for simplicity π is the change point,

$$H_{1T}(\pi) : \theta_t = \begin{cases} \theta_1 & \text{for} \quad t = 1, \cdots, [T\pi] \\ \theta_2 & \text{for} \quad t = [T\pi] + 1, \cdots, T. \end{cases}$$

for some constants $\theta_1, \theta_2 \in \Theta \subset R^n$.

2.1 The Estimators and Assumptions

We now consider the partial sample GMM estimator. We benefit from some of the arguments and definitions in section 3.2 of Andrews (1993a). The parameter vector θ takes two different values in the whole sample. In the first part of the sample, $t = 1, \cdots, [T\pi]$, it takes the value θ_1, and when $t = [T\pi] + 1, \cdots, T$, it takes the value θ_2. Define $\theta_1 = (\alpha_1', \beta_1')' \in \Theta = A \times B \subset R^{n_1} \times R^{n_2}$, $\theta_2 = (\alpha_2', \beta_2')' \in \Theta = A \times B \subset R^{n_1} \times R^{n_2}$. Set $\underline{\theta} = (\theta_1', \theta_2')' \in A \times B \times A \times B \subset R^{n_1} \times R^{n_2} \times R^{n_1} \times R^{n_2}$, and $n_1 + n_2 = n$. The parameters are in the interior of the compact set $\Theta \times \Theta = \underline{\Theta}$. The population values of θ_1 and θ_2 are $\theta_{10} = (\alpha_{10}', \beta_{10}')', \theta_{20} = (\alpha_{20}', \beta_{20}')'$ respectively. The population orthogonality conditions are, each with G equations:

$$E[h(Y_t, \theta_{10})|F_t] = 0, \quad \text{for all} \quad t = 1, \cdots, [T\pi],$$

$$E[h(Y_t, \theta_{20})|F_t] = 0, \quad \text{for all} \quad t = [T\pi] + 1, \cdots, T,$$

where F_t is the information set at time t, and $h(.,.)$ is a specified R^G valued function. Let Z_t be a K-dimensional vector of instruments contained in F_t. The observed sample is $\{(Y_t, Z_t), t = 1, \cdots, T\}$, and defined on a probability space $(\Delta, \mathcal{F}, \mathcal{P})$ where Δ is the sample space.

For each potential change point $\pi \in \Pi \subset (0,1)$, as in Andrews (1993a), we define the estimator that is based on the sample analogues of these orthogonality conditions. We define the unrestricted GMM estimators as $\hat{\underline{\theta}}(\pi) = (\hat{\theta}_1(\pi)', \hat{\theta}_2(\pi)')'$.

Definition 1. *The partial sample continuous updating GMM estimators* $\{(\hat{\underline{\theta}}(\pi) : \pi \in \Pi) : T \geq 1\}$ *are any sequence of estimators that minimize the objective function* $S_T(\underline{\theta}, \pi)$ *over* $\underline{\theta} \in A \times B \times A \times B \subset R^{n_1} \times R^{n_2} \times R^{n_1} \times R^{n_2}$ *for all* $\pi \in \Pi$ *where*

$$S_T(\underline{\theta}, \pi) = \bar{m}_T(\underline{\theta}, \pi)' W_T(\underline{\theta}, \pi) \bar{m}_T(\underline{\theta}, \pi),$$

and

$$\bar{m}_T(\underline{\theta}, \pi) = \begin{bmatrix} T^{-1/2} \sum_{t=1}^{[T\pi]} \psi_t(\theta_1) \\ T^{-1/2} \sum_{t=[T\pi]+1}^{T} \psi_t(\theta_2) \end{bmatrix}, \tag{1}$$

$\psi_t(\theta) = h(Y_t, \theta) \bigotimes Z_t$. $W_T(\underline{\theta}, \pi)$ *is an* $O_p(1)$ *positive definite, symmetric weighting matrix with dimensions of* $2GK \times 2GK$. $\hat{\underline{\theta}}(.)$ *is a random element.* $\bar{m}_T(.)$ *is a* R^{2GK} *valued function.*

We now introduce the assumptions.

Assumption 1.

$$T^{-1/2} \sum_{t=1}^{[T\pi]} \psi_t(\theta_0) \Longrightarrow B_{GK}(\pi),$$

where $B_{GK}(\pi)$ *is a Brownian Motion with,* $B_{GK}(\pi) \equiv \Omega_{\theta_0,\theta_0}^{1/2} W_{GK}(\pi)$, *where* $W_{GK}(\pi)$ *is a GK dimensional standard Brownian motion. The covariance matrix* $\Omega_{\theta_0,\theta_0} = \lim_{T\to\infty} E\Psi_T(\theta_0)\Psi_T(\theta_0)'$, *where* $\Psi_T(\theta_0) = T^{-1/2} \sum_{t=1}^{T} \psi_t(\theta_0)$. *Note that* $\Omega_{\theta_0,\theta_0}$ *is positive definite.*

It can be seen that primitives for Assumption 1 can be found in time series literature for both triangular array and strictly stationary random variables such as Andrews (1993a).

Before presenting the next assumption, we introduce the matrices that will be used in building an efficient heteroskedasticity robust weight matrix:

$$V_{T1}(\theta_1, \pi) = \frac{1}{[T\pi]} \sum_{t=1}^{[T\pi]} [\psi_t(\theta_1) - \bar{\psi}(\theta_1)][\psi_t(\theta_1) - \bar{\psi}(\theta_1)]',$$

$$V_{T2}(\theta_2, \pi) = \frac{1}{T - [T\pi]} \sum_{t=[T\pi]+1}^{T} [\psi_t(\theta_2) - \bar{\psi}(\theta_2)][\psi_t(\theta_2) - \bar{\psi}(\theta_2)]',$$

where $\bar{\psi}(\theta_1), \bar{\psi}(\theta_2)$ represent the sample means obtained from the first and second parts of the full sample respectively. Heteroskedasticity and autocorrelation consistent variance covariance matrix can also be formed by using the arguments in Andrews (1993a) or section 2 of Caner (2003). The efficient weight matrix is:

$$W_T(\underline{\theta}, \pi) = \begin{bmatrix} \frac{V_{T1}(\theta_1,\pi)^{-1}}{\pi} & 0 \\ 0 & \frac{V_{T2}(\theta_2,\pi)^{-1}}{1-\pi} \end{bmatrix}.$$

We need uniform consistency of weight matrices for the limit theory under the null and at the true values.

Assumption 2. *Under the null hypothesis of no structural change and at the true value of the parameters ($\theta_{10} = \theta_{20} = \theta_0$), and uniformly in π, for $j = 1, 2$,*

$$V_{Tj}(\theta_0, \pi) \xrightarrow{p} \Omega_{\theta_0,\theta_0}.$$

We want to introduce a Kleibergen (2005) type of test statistic for our problem. This may have better small sample properties. But before providing that we need the following Assumptions.

Assumption 3.*The $GK \times 1$ dimensional derivative of $\psi_t(\theta_0)$ with respect to θ_i, $i = 1, 2, \cdots, n$*

$$q_{it}(\theta_0) = \frac{\partial \psi_t(\theta_0)}{\partial \theta_i'},$$

is such that joint limiting behavior of the series $\bar{\psi}_t(\theta_0) = \psi_t(\theta_0) - E\psi_t(\theta_0)$ and $\bar{q}_t(\theta_0) = (\bar{q}_{1t}(\theta_0)', \cdots, \bar{q}_{nt}(\theta_0)')'$ with $\bar{q}_{it}(\theta_0) = q_{it}(\theta_0) - Eq_{it}(\theta_0)$ follows the functional central limit theorems

$$T^{-1/2} \sum_{t=1}^{[T\pi]} \begin{bmatrix} \bar{\psi}_t(\theta_0) \\ \bar{q}_t(\theta_0) \end{bmatrix} \Longrightarrow \begin{bmatrix} B_{GK}(\pi) \\ B_{nGK}(\pi) \end{bmatrix} \equiv B(\pi),$$

where $B_{GK}(\pi)$ is $GK \times 1$ Brownian Motion and $B_{nGK}(\pi)$ is $nGK \times 1$ Brownian Motion. They are jointly distributed $B(\pi) \equiv \Omega^{1/2} W(\pi)$, $W(\pi) = (W_{GK}(\pi)', W_{nGK}(\pi)')'$ and $W_{GK}(\pi), W_{nGK}(\pi)$ are GK, nGK dimensional standard Brownian Motions respectively. Ω is positive semi-definite, symmetric $(GK + nGK) \times (GK + nGK)$ matrix.

$$\Omega = \begin{bmatrix} \Omega_{\theta_0,\theta_0} & \Omega_{\theta_0,q_0} \\ \Omega_{q_0,\theta_0} & \Omega_{q_0,q_0} \end{bmatrix},$$

where $\Omega_{\theta_0,\theta_0}$ is $GK \times GK$, $\Omega_{q_0,\theta_0} = \Omega'_{\theta_0,q_0} = (\Omega'_{\theta_0,q_0,1}, \cdots, \Omega'_{\theta_0,q_0,n})$, and $\Omega_{q_0,\theta_0,i}$ is $GK \times GK$ matrices for $i = 1, \cdots, n$.

Furthermore

$$\Omega = \lim_{T\to\infty} var \begin{pmatrix} T^{-1/2} \sum_{t=1}^{T} \psi_t(\theta_0) \\ T^{-1/2} \sum_{t=1}^{T} q_t(\theta_0) \end{pmatrix}.$$

This Assumption is the partial sample counterpart of Assumption 1 in Kleibergen (2005). The functional central limit theorem can be obtained as described in Assumption 1 or by Andrews (1993a). These are standard in structural change literature.

Assumption 4. *We assume the following consistency results for variance-covariance matrix estimators, under the null of no structural change*

$$\hat{\Omega}^{j}_{\theta_0,\theta_0} \xrightarrow{p} \Omega_{\theta_0,\theta_0},$$

$$\hat{\Omega}^{j}_{q_0,\theta_0,i} \xrightarrow{p} \Omega_{q_0,\theta_0,i},$$

$$\hat{\Omega}^{j}_{q_0,q_0} \xrightarrow{p} \Omega_{q_0,q_0},$$

where $j = 1, 2$ represents the first part of the sample ($t = 1, \cdots, [T\pi]$), and second part of the sample ($t = [T\pi] + 1, \cdots, T$) respectively.

Similar assumptions are used in Kleibergen (2005), and Andrews (1993a). Specifically, this is the partial sample counterpart of Assumption 2 in Kleibergen (2005).

3 Tests for Structural Change

In this section we consider two structural change tests. First, we analyze an LR-like statistic. For fixed π we define the statistic as follows:

$$LR_T(\pi) = S_T(\tilde{\theta}(\pi), \pi) - S_T(\hat{\underline{\theta}}(\pi), \pi). \tag{2}$$

For the fixed change point π, this represents the difference between the restricted partial sample GMM objective function and the unrestricted version. Restricted partial sample GMM estimate, $\tilde{\theta}(\pi)$, is defined in (16) and (17).

We use the same arguments in Theorem 1 of Stock and Wright (2000) to have the unrestricted estimates. Denote the partial sample parameter vectors by $\underline{\alpha} = (\alpha_1', \alpha_2')'$, $\underline{\beta} = (\beta_1', \beta_2')'$. These vectors have dimensions of $2n_1 \times 1$ and $2n_2 \times 1$, respectively. Let $\hat{\underline{\beta}}(\underline{\alpha}, \pi)$ solve $\arg\min_{\underline{\beta} \in B \times B} S_T(\underline{\alpha}, \underline{\beta}, \pi)$ for all $\pi \in \Pi$. Let $\hat{\underline{\alpha}}(\pi) = \arg\min_{\underline{\alpha} \in A \times A} S_T(\underline{\alpha}, \hat{\underline{\beta}}(\underline{\alpha}, \pi), \pi)$ for all $\pi \in \Pi$. Then let $\hat{\underline{\beta}}(\pi) = \hat{\underline{\beta}}(\hat{\underline{\alpha}}(\pi), \pi)$ for all $\pi \in \Pi$, and $\underline{\theta} = (\theta_1', \theta_2')'$.

We are interested in the following version of an LR-like test statistic: $\sup_{\pi \in \Pi} LR_T(\pi)$. The null hypothesis, $H_0 : \theta_1 = \theta_2$, is rejected for large values of this statistic. The LR-like test is not asymptotically pivotal in weak instrument asymptotics setup. This limit is derived under the assumption that $\{Y_t, Z_t\}$ is governed by β (strongly identified parameter), not by α (weakly identified parameter) as seen in Caner (2003). The limit of completely unidentified case and the linear case are not nuisance parameter free and nonstandard (Caner, 2003).

The likelihood function is flat on a nonidentification subset of parameter space in simultaneous equations models. So the likelihood ratio test may have a better chance of being boundedly pivotal than the Wald type of tests. Dufour (1997) observes that LR statistic for testing hypothesis on unidentified parameters are boundedly pivotal. This may be helpful in our case since we have unidentified nuisance parameters under the null as well as the weak instruments problem.

Unlike Likelihood ratio, Lagrange multiplier and Wald tests, Anderson-Rubin (1949) type test is asymptotically pivotal in GMM with weak instruments. This is shown in Theorem 2 of Stock and Wright (2000). We decide to use this test for testing regime change as well. This test is based on partial sample continuous updating objective function. So we define the test statistic based on S-sets as follows [1]:

$$\sup_{\pi\in\Pi} S_T(\pi) = \sup_{\pi\in\Pi} S_T(\tilde{\theta}(\pi), \pi).$$

Another reason to use this test is: it may be boundedly pivotal in large samples since it is a minimized objective function. A similar argument is made in section 6 of Dufour (1997) in the case of the Anderson-Rubin (1949) test. We can rewrite sup S test as follows:

$$\sup_{\pi\in\Pi} S_T(\pi) = \sup_{\pi\in\Pi} LR_T(\pi) + \sup_{\pi\in\Pi} J_T(\pi),$$

where $\sup J_T(\pi) = \sup S_T(\underline{\hat{\theta}}(\pi), \pi)$. J test is the overidentifying restrictions test for the partial sample GMM in section 2.1. Even though sup S test has good properties in weak instruments case; it may reject the true null of stability of parameters when the overidentifying restrictions are invalid. The limit for S-based test statistic is not nuisance parameter free in weak instrument asymptotics. This can be seen in Caner (2003). The test statistics are not asymptotically pivotal since α cannot be consistently estimated. Therefore, we try to find an asymptotic bound which is free of nuisance parameters. The proof of Theorem 1 can be found in the appendix. Note that the following bound works under minimal assumptions. This bound is valid for strong, weakly identified cases, and completely unidentified case. These are explained in detail after the proof of Theorem 1 in Appendix.

Theorem 1. *Under Assumptions 1-2, and the null hypothesis of no structural change, both* $supLR_T(\pi)$ *and* $supS_T(\pi)$ *are asymptotically boundedly pivotal. The asymptotically pivotal bound distribution for both test statistics are the same, and given by:*

$$\sup_{\pi\in\Pi}\left[\frac{W_{GK}(\pi)'W_{GK}(\pi)}{\pi} + \frac{(W_{GK}(1) - W_{GK}(\pi))'(W_{GK}(1) - W_{GK}(\pi))}{1-\pi}\right],$$

where $W_{GK}(\pi)$ is a GK dimensional standard Brownian motion.

The critical values for the distribution in Theorem 1 are calculated in Table 1. The bound for sup S test comes from the fact that this is a minimized objective function for each π. We can have a bound for this function by using the version evaluated at the population values of the parameters. Then this test statistic converges to a limit which is free of nuisance parameters. This

[1] I owe special thanks to James Stock for this suggestion.

test statistic is also robust to all the identification problems. For LR-like test, we can see that it is the difference between sup S test and the unrestricted partial sample GMM objective function. Since the unrestricted objective function is either zero or positive with probability one, the bound for sup S test applies to LR test as well. The bound for LR test is conservative compared to sup S test.

Note that by simple algebra we can rewrite the bound in Theorem 1 into the two independent terms:

$$\sup_{\pi\in\Pi}\left[\frac{(W_{GK}(\pi)-\pi W_{GK}(1))'(W_{GK}(\pi)-\pi W_{GK}(1))}{\pi(1-\pi)}\right]+\chi^2_{GK}, \tag{3}$$

where χ^2_{GK} represents chi-square distribution with GK degrees of freedom. GK is the number of orthogonality conditions. The form in (3) can be used to compare the bound with the existing limit laws for sup LM, LR tests of Andrews (1993a) in the case of standard asymptotics of two-step GMM. We analyze this issue after Lemma 2.

Now we provide a Kleibergen (2005) type of test statistic [2]. This is based on the partial derivative of the continuous updating objective function. But before providing it, we describe notation. Let $\hat{\Omega}^j_{\theta\theta}, \hat{\Omega}^j_{q\theta,i}$ be the covariance matrix estimators described in Assumption 4, but evaluated at θ. Note that superscript "j" describes which part of the full sample is used. For example $j=1$ represents the estimator which uses observations $t=1,\cdots,[T\pi]$, and $j=2$ represents the estimator which uses observations $t=[T\pi]+1,\cdots,T$.

First we define the restricted K-estimator

$$\tilde{\theta}_K(\pi)=\arg\min_{\theta\in\Theta}K_T(\theta,\pi), \tag{4}$$

for each $\pi\in\Pi$. Upon following (12)-(16) of Kleibergen (2005) and taking into account the partial sample nature of our problem,

$$\begin{aligned} K_T(\theta,\pi) &= \frac{1}{\pi}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)]'(\hat{\Omega}^1_{\theta,\theta})^{-1/2}P_{(\hat{\Omega}^1_{\theta,\theta})^{-1/2}\tilde{D}^1_T(\theta,\pi)}(\hat{\Omega}^1_{\theta,\theta})^{-1/2}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)] \\ &+ \frac{1}{1-\pi}[T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta)]'(\hat{\Omega}^2_{\theta,\theta})^{-1/2}P_{(\hat{\Omega}^2_{\theta,\theta})^{-1/2}\tilde{D}^2_T(\theta,\pi)}(\hat{\Omega}^2_{\theta,\theta})^{-1/2}[T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta)]. \end{aligned} \tag{5}$$

Terms $\tilde{D}^1_T(\theta,\pi), \tilde{D}^2_T(\theta,\pi)$ are derived from the partial derivative of CUE objective function in section 2 here, and

$$\tilde{D}^1_T(\theta,\pi)=[q_{1T}(\theta,\pi)-\hat{\Omega}^1_{q\theta,1}(\hat{\Omega}^1_{\theta\theta})^{-1}(T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)),\cdots,q_{nT}(\theta,\pi)-\hat{\Omega}^1_{q\theta,n}(\hat{\Omega}^1_{\theta\theta})^{-1}(T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta))], \tag{6}$$

[2] I owe special thanks to an anonymous referee who suggested this test statistic

where

$$q_{iT}(\theta,\pi) = T^{-1/2}\sum_{t=1}^{[T\pi]} q_{it}(\theta), \tag{7}$$

$i = 1, \cdots, n$.

$$\begin{aligned}\tilde{D}_T^2(\theta,\pi) &= [(q_{1T}(\theta) - q_{1T}(\theta,\pi)) - \hat{\Omega}_{q\theta,1}^2(\hat{\Omega}_{\theta\theta}^2)^{-1}(T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta)), \cdots, \\ &\quad (q_{nT}(\theta) - q_{nT}(\theta,\pi)) - \hat{\Omega}_{q\theta,n}^2(\hat{\Omega}_{\theta\theta}^2)^{-1}(T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta))],\end{aligned} \tag{8}$$

where $q_{iT}(\theta) = T^{-1/2}\sum_{t=1}^{T} q_{it}(\theta)$, $i = 1, \cdots, n$.

The following test-statistic is suggested for testing the null of no structural change $H_0 : \theta_1 = \theta_2$:

$$\sup_{\pi\in\Pi} K_T(\tilde{\theta}_K(\pi), \pi). \tag{9}$$

Theorem 2.*Under Assumptions 3-4, and the null hypotheses of no structural change, the test statistic in (9) is asymptotically boundedly pivotal. The bound distribution is*

$$\sup_{\pi} \frac{[W_n(\pi) - \pi W_n(1)]'[W_n(\pi) - \pi W_n(1)]}{\pi(1-\pi)} + \chi_n^2,$$

where $W_n(\pi)$ is $n \times 1$ dimensional standard Brownian Motion, and χ_n^2 is chi-square distribution with n degrees of freedom.

Remark. This is very similar to the bound for sup S and LR tests in equation (3). However, in Theorem 2 the bound depends on number of parameter restrictions, rather than the number of orthogonality restrictions. Since the form of two limits in Theorems 1 (equation (3)) and 2 are the same, so the critical values can be found also from Table 1 replacing GK with n.

3.1 Standard Identified Case and Comparison

In this subsection we derive the limits of the tests under standard identification assumptions. Then we compare the limits with the bounds in Theorem 1 and Theorem 2 to assess whether the bounds are conservative or not. We put Kleibergen (2005) test at the end of this subsection since its analysis and assumptions are different from the other two test statistics.

Assumption S1.

(i) $\{Y_t, Z_t\}$ is strictly stationary, and β mixing with mixing coefficients satisfy $\beta(s) \leq Ds^{-A}$, where $D > 0$, and $A > 2 + 4/\delta$, for some $\delta > 0$. Note that $\{Y_t, Z_t\}$ is strictly stationary under the null hypothesis;

(ii)

$$\sup_{\theta\in\Theta} E|\psi_t(\theta)|^{2+\delta} < \infty, \quad \text{for some } \delta > 0;$$

(iii)

$$|\psi_t(\theta_1) - \psi_t(\theta_2)| \leq B_t|\theta_1 - \theta_2|,$$

where $E(B_t)^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption S2.

$$ET^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\theta) = \pi m_1(\theta),$$

where $m_1(\theta_0) = 0$, $m_1(\theta) \neq 0$, for $\theta \neq \theta_0$. $R(\theta)$ is continuous and $R(\theta_0)$ has full column rank. $R(\theta) = \partial m_1(\theta)/\partial\theta'$ is $GK \times n$.

Assumption S3. Under the null hypothesis of no structural change ($\theta_1 = \theta_2 = \theta$), and uniformly in $\theta \times \pi$, for $j = 1, 2$,

$$V_{Tj}(\theta, \pi) \xrightarrow{p} \Omega_{\theta,\theta},$$

where $V(.)$ is defined before Assumption 2, and $\Omega_{\theta,\theta} = \lim_{T\to\infty} E\Psi_T(\theta)\Psi_T(\theta)'$ and $\Psi_T(\theta) = T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta)$. This is also positive definite.

Remark. Assumption S1 is used in Lemma 1 below and discussed in the proof of that Lemma. This assumption is similar to the one used in Andrews (1993a) with one difference: we can not allow near epoch dependent processes. Lemma 1 only works for strictly stationary β mixing random variables. Assumption S2 is standard GMM identification condition. Assumption S3 is partial sample counterpart of the uniform consistency of the weight matrices assumption (Assumption D") of Stock and Wright (2000). Primitives for these assumptions can be found in Andrews (1993a). Note that Assumptions S1 and S3 are sufficient for Assumptions 1 and 2.

We have the following lemma which is helpful in establishing the limit law for the partial sample GMM estimator in the strongly identified case as well as the completely unidentified case and in certain cases of weakly identified GMM, (Caner, 2003). This is one of the main theoretical results of this article. That is the reason we use strong Assumptions S1-S3 instead of Assumptions 1 and 2. Before Lemma 1, define the sequential empirical process as in van der Vaart and Wellner (1996):

$$\Psi_T(\theta, \pi) = T^{-1/2}\sum_{t=1}^{[T\pi]}[\psi_t(\theta) - E\psi_t(\theta)].$$

Lemma 1 generalizes the weak convergence of sequential empirical process with iid data, Theorem 2.12.1 in van der Vaart and Wellner (1996) to time series context. This result also may be useful in other problems that involve partial sample results with time series data. The proof is in the appendix.

Lemma 1. *Under the null hypothesis of no structural change, and Assumption S1, uniformly in $\theta \times \pi \in \Theta \times \Pi$,*

$$\Psi_T(\theta, \pi) \Longrightarrow \Psi(\theta, \pi),$$

where $\Psi(\theta, \pi)$ is a multivariate, $GK \times 1$, Kiefer process, which is Gaussian zero mean with a variance covariance function $E\Psi(\theta_s, \pi_1)\Psi(\theta_k, \pi_2)' = (\pi_1 \wedge \pi_2)\Omega_{\theta_s,\theta_k}$, and $\Omega_{\theta_s,\theta_k} = \lim_{T\to\infty} E\Psi_T(\theta_s)\Psi_T(\theta_k)'$.

For the subsequent Theorems we need to define $\Psi(\theta)$ which is a Gaussian stochastic process on Θ with mean zero and covariance function $E\Psi(\theta_1)\Psi(\theta_2) = \Omega_{\theta_1,\theta_2}$. More explanations of this process are given after (26).

We now supply the limits of S and LR tests under the standard identification assumption for continuous updating GMM.

Theorem 3. *Under the null hypothesis of no structural change, and Assumptions S1-S3,*

(a)

$$\sup_{\pi\in\Pi} LR_T(\pi) \xrightarrow{d} \sup_{\pi\in\Pi}\left[\frac{(W_n(\pi)-\pi W_n(1))'(W_n(\pi)-\pi W_n(1))}{\pi(1-\pi)}\right].$$

(b)

$$\sup_{\pi\in\Pi} S_T(\pi) \xrightarrow{d} \sup_{\pi\in\Pi}\left[\frac{(W_{GK}(\pi)-\pi W_{GK}(1))'(W_{GK}(\pi)-\pi W_{GK}(1))}{\pi(1-\pi)}\right] + \chi^2_{GK-n},$$

where χ^2_{GK-n} is chi-square distribution with $GK-n$ degrees of freedom, n is the number of parameters, and the limit in Theorem 3b consists of two independent terms.

The proof of Theorem 3a can easily be deduced from two step GMM case of Andrews (1993a). The proof of Theorem 3b is in Appendix. $W_{GK}(\pi), W_n(\pi)$ are GK and n dimensional standard Brownian motions respectively. When $GK = n$, the limit in Theorem 3b consists of the first term only.

Remarks. For $\sup LR_T$ test, the distribution is exactly the same as in Andrews (1993a) for the case of the two-step GMM, which is a standardized tied-down Bessel process of order n. We can see that when the number of orthogonality conditions (GK) are large compared to the number of parameters (n), the bound becomes conservative as can be seen from the limits in (3), Theorem 3a. We also observe that critical values get large with the dimension of the standardized tied-down Bessel process, and the χ^2 distribution. (Andrews, 1993a).

The reason that sup LR test gets more conservative can be explained in the following way. The bound in (3) is

$$\sup_{\pi}\left[\frac{(W_{GK}(\pi)-\pi W_{GK}(1))'(W_{GK}(\pi)-\pi W_{GK}(1))}{\pi(1-\pi)}\right] + \chi^2_{GK}. \tag{10}$$

First term in (10) is a tied-down Bessel process of order GK. The limit in standard identified case in Theorem 3a is:

$$\sup_{\pi}\left[\frac{(W_n(\pi)-\pi W_n(1))'(W_n(\pi)-\pi W_n(1))}{\pi(1-\pi)}\right].$$

This last term is a tied-down Bessel process of order "n". First of all, when GK gets larger with respect to n, the critical values get larger when we compare these two tied-down Bessel processes (Andrews, 1993a). Furthermore, we have an additional χ^2_{GK} term in (10) which makes the bound more conservative when GK is large.

For sup S test, the reverse is true. If we compare the bound (10) with the limit for standard identified case in Theorem 3b, first terms are the same, but in the case of the bound we have additional χ^2_{GK} term, whereas in the case of standard identified case that term is χ^2_{GK-n}. So when GK is very large compared to n, the effect of n is minimal, and we see that bound in sup S test gets less conservative.

By comparing (3) and Theorem 3, we see that if n is large with respect to GK, we expect the bound to be conservative. To quantify this, we conduct the following exercise. First, we generate distributions in Theorem 3, the standard identified case for $n = 1, \cdots, 10$. Then, we substitute the 80^{th}, 90^{th} percentiles of the distribution of the bound (Theorem 1 or (3)) into the distribution of Theorem 3, for $n = 1, 2 \cdots, 10$. Then we find the corresponding percentiles of these $80^{th}, 90^{th}$ percentiles in Theorem 1 in the distribution of Theorem 3. The results are in Table 2. So in this sense LR, and sup S tests' limits' in strongly identified case are compared with the bound in Theorem 3. The first column represents the percentile in the distribution of the bound (Theorem 1 or (3)). The other columns show the corresponding percentiles in the distribution of Theorem 3. We see from Table 2 that with $GK = 10$ at $\pi_0 = 0.15$, when n gets larger than 6, the 90^{th} percentile of the bound (90^{th} percentile is 36.7 from Table 1) corresponds to 99-99.8 percentiles in the standard identified case of Theorem 3b. For S-test, we see that the bound is conservative when n is large. We also see LR test is very conservative from Tables 2a-c.

To test for $H_0 : \alpha_1 = \alpha_2$, we can use $\sup S_T(\tilde{\alpha}(\pi), \tilde{\beta}(\tilde{\alpha}(\pi), \pi), \pi)$ as a test statistic. The limit distribution can be found using Caner (2003). However, the test statistic is not asymptotically pivotal. A nuisance parameter free asymptotic bound can be derived using the technique in the proof of Theorem 1. An important issue is to do subtests of either weakly or strongly identified parameters. This is rather difficult and beyond the scope of this paper, however, one possible approach is to use Dufour and Taamouti (2001).

Now we analyze Kleibergen (2005) type of test statistic under standard identification which is introduced in (9). First we provide the assumptions needed for the standard identification case for Kleibergen (2005) type of structural change test.

Assumption S4.

(i).

$$\sup_{\theta \in \Theta} E|q_{it}(\theta)|^{2+\delta} < \infty, \quad \textit{for some}\, \delta > 0,$$

for $i = 1, \cdots, n$, *where* $q_{it}(\theta) = \partial \psi_t(\theta) / \partial \theta_i$.

(ii).$q_{it}(\theta)$ *is continuously differentiable in* θ, $i = 1, 2, \cdots, n$.

Assumption S5.$\lim_{T\to\infty} ET^{-1} \sum_{t=1}^{[T\pi]} q_{it}(\theta)$ *exists uniformly over* (θ, π) *and equals* $\pi q_i(\theta)$ *where* $q_i(\theta)$ *is finite and nonrandom.*

Assumption S6. *We assume the following consistency results for variance-covariance matrix estimators, under the null of no structural change* *(i).*

$$\hat{\Omega}^j_{\theta,\theta} \xrightarrow{p} \Omega_{\theta,\theta},$$

$$\hat{\Omega}^j_{q,\theta,i} \xrightarrow{p} \Omega_{q,\theta,i},$$

$$\hat{\Omega}^j_{q,q} \xrightarrow{p} \Omega_{q,q},$$

where $j = 1, 2$ represents the first part of the sample ($t = 1, \cdots, [T\pi]$), and second part of the sample ($t = [T\pi] + 1, \cdots, T$) respectively. The results are uniform over $\theta \times \pi$. The limit matrices are described in Assumption 3 but evaluated at θ_0 there.

(ii). We also assume $P_{\Omega_{\theta\theta}^{-1/2} D(\theta)} \Omega_{\theta\theta}^{-1/2} m_1(\theta) \neq 0 for \theta \neq \theta_0$. $m_1(\theta)$ is explained in Assumption S2 and $D(\theta)$ is explained in (77).

Assumptions S4 and S5 are helpful to provide a functional central limit theorem for the partial derivative $q_{it}(\theta)$ like Lemma 1 for the function $\psi_t(\theta)$. Assumption S6 is strengthened form of Assumption 4. Similar Assumptions are used in Andrews (1993a). Assumption S6ii, regarding the weight matrix is needed for consistency of $\tilde{\theta}_K(\pi)$. Since the estimator minimizes a Kleibergen type of test statistic in equation (4), this condition is essential in deriving consistency. This is similar in nature to positive semidefinite weight matrix used in GMM literature.

Now we provide the limit result for Kleibergen (2005) type of test for structural change in the case of standard identification.

Theorem 4. *Under the null hypothesis of no structural change, and Assumptions S1-S2, S4-S6*

$$\sup_{\pi} K_T(\tilde{\theta}_K(\pi), \pi) \Longrightarrow \sup_{\pi} \frac{[W_n(\pi) - \pi W_n(1)]'[W_n(\pi) - \pi W_n(1)]}{\pi(1-\pi)}.$$

Remark. Note that this is the limit result found in Andrews (1993a) for his LM test as well as the limit for the LR test.

We also compare the limit in Theorem 4 with the bound in Theorem 2 to see how conservative the bound is. We see that the bound and the limit does not depend on number of orthogonality conditions (GK) but only depends on number of restrictions (n). The bound should not be affected when the system has large number of moment conditions. The bound gets conservative when we test large number of restrictions. To quantify this we generate the distribution of Theorem 4. Then substitute $80^{th}, 90^{th}$ percentiles of the distribution of the bound (Theorem 2) into the distribution of Theorem 4 for $n = 1, 2, \cdots, 10$. The corresponding percentiles give us an idea how conservative the bound is. The results are in Table 2d. We also see that the bound here is conservative.

TABLE 1: Critical Values for the Asymptotic Bound

	$\pi_0 = 0.05$				$\pi_0 = 0.10$				$\pi_0 = 0.15$			
	20%	10%	5%	1%	20%	10%	5%	1%	20%	10%	5%	1%
$GK = 1$	7.8	9.6	11.3	15.1	7.3	9.1	10.9	14.4	6.9	8.7	10.4	14.0
$GK = 2$	11.8	13.8	15.7	19.8	11.2	13.4	15.4	19.9	10.7	12.9	14.8	18.5
$GK = 3$	15.1	17.5	19.6	24.1	14.5	16.7	18.8	23.3	14.0	16.4	18.5	23.2
$GK = 4$	18.2	20.7	23.1	28.0	17.6	20.1	22.5	27.5	17.0	19.7	22.0	27.1
$GK = 5$	21.4	24.2	26.5	31.6	20.6	23.4	25.9	30.9	20.0	22.8	25.4	30.7
$GK = 6$	24.2	27.2	29.7	34.6	23.5	26.4	29.0	35.5	22.8	25.9	28.4	33.7
$GK = 7$	27.1	30.1	32.8	38.7	26.2	29.2	31.7	37.7	25.6	28.7	31.3	37.0
$GK = 8$	29.9	33.1	35.9	41.6	29.0	32.1	35.2	41.0	28.4	31.6	34.5	40.5
$GK = 9$	32.6	35.9	39.0	45.3	31.8	34.9	37.9	44.2	31.1	34.6	37.6	44.0
$GK = 10$	35.2	38.6	41.4	47.9	34.4	37.9	41.0	47.5	33.4	36.7	40.2	46.4

Note: GK represents the number of population orthogonality equations (G) multiplied by the number of instruments (K). We use 10000 repetitions. $\Pi = [\pi_0, 1-\pi_0]$ and $GK = 1, 2, \cdots, 10$. π_0 takes the values of 0.05, 0.10, 0.15, which are used in practice. We approximate the distribution by a grid that is defined by $\Pi(N) = [\pi_0, 1-\pi_0] \cap \{\pi = j/N : j = 0, 1, \cdots, N\}$, $N = 3600$ as in Andrews (1993a).

Table 2a: Comparison of Identified GMM and The Bound Distributions

Test	Percentile	$n = 1$	2	3	4	5	6	7	8	9	10
LR	80	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.5	99.4
LR	90	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.8	99.8
S	80	83.1	86.3	89.3	91.7	94.2	96	97.5	97.9	98.9	99.3
S	90	91.8	93.7	95.3	96.7	97.8	98.7	99	99.5	99.6	99.8

Note: We conduct 10000 trials and generate the distributions in Theorem 3, for $GK = 10, \pi_0 = 0.15$, for $n = 1, 2 \cdots, 10$. Then we substitute $80^{th}, 90^{th}$ percentiles of the distribution of the bound in (3) (33.4, 36.7 respectively from Table 1) into the distributions in Theorem 3. Then we report the corresponding percentiles in the distribution for Theorem 3. The first column represents two tests analyzed. The second column represents the percentile in the distribution of the bound (Theorem 1 or (3)). The other columns show the corresponding percentiles in the distributions of Theorem 3.

Table 2b: Comparison of Identified GMM and The Bound Distributions

Test	Percentile	$n = 1$	2	3	4	5
LR	80	99.9	99.8	99.5	98.6	97.5
LR	90	100.0	99.9	99.8	99.5	99.2
S	80	85.6	89.7	92.5	95.2	97.4
S	90	93.3	95.7	96.8	98.1	99.1

Note: We conduct 10000 trials and generate the distributions in Theorem 3, for $GK = 5, \pi_0 = 0.15$, for $n = 1, 2 \cdots, 10$. Then we substitute $80^{th}, 90^{th}$ percentiles of the distribution of the bound in (3) (20.0, 22.8 respectively from Table 1) into the distributions in Theorem 3. Then we report the corresponding percentiles in the distribution for Theorem 3.

Table 2c: Comparison of Identified GMM and The Bound Distributions

Test	Percentile	$n=1$	2
LR	80	97.9	92.6
LR	90	99.3	96.8
S	80	87.3	92.1
S	90	94.5	96.7

Note: We conduct 10000 trials and generate the distributions in Theorem 3, for $GK = 2, \pi_0 = 0.15$, for $n = 1, 2 \cdots, 10$. Then we substitute $80^{th}, 90^{th}$ percentiles of the distribution of the bound in (3) (10.7, 12.9 respectively from Table 1) into the distributions in Theorem 3. Then we report the corresponding percentiles in the distribution for Theorem 3.

Table 2d: Comparison of Identified GMM and The Bound Distribution

Test	Percentile	$n=1$	2	3	4	5	6	7	8	9	10
K	80	89.2	92.5	94.5	96.0	97.5	97.7	98.6	98.8	99.3	99.3
K	90	95.0	97.0	97.8	98.5	99.1	99.3	99.5	99.6	99.8	99.8

Note: We conduct 10000 trials and generate the distribution in Theorem 4, $\pi = 0.15$, for $n = 1, 2 \cdots, 10$. Then we substitute $80^{th}, 90^{th}$ percentiles of the distribution of the bound in Theorem 2 into the distributions in Theorem 4. Then we report the corresponding percentiles in the distribution for Theorem 4.

4 Small Sample Properties

Our Monte Carlo setups use the standard consumption-based asset pricing model, the representative agent intertemporally separable CCAPM with CRRA preferences. In the simulation setup, we try to answer the question of how the existing structural change tests, sup LM and likelihood ratio-like (QLR) test of Andrews (1993a), and the tests proposed here, $\sup LR$, $\sup S$ perform in small samples when there are weak instruments. Andrews' (1993a) LR-like and sup LM tests are as follows:

$$QLR = \sup_{\pi \in \Pi} [\bar{m}_T(\tilde{\theta}_{TS}, \pi)' W_{T,GMM}(\underline{\hat{\theta}}_F, \pi) \bar{m}_T(\tilde{\theta}_{TS}, \pi) - \bar{m}_T(\underline{\hat{\theta}_{TS}}, \pi)' W_{T,GMM}(\underline{\hat{\theta}}_F, \pi) \bar{m}_T(\underline{\hat{\theta}_{TS}}, \pi)],$$

$$supLM = \sup_{\pi \in \Pi} [\frac{T}{\pi(1-\pi)} \bar{l}_{1T}(\tilde{\theta}_{TS}, \pi)' \tilde{V}^{-1} \tilde{D} (\tilde{D}' \tilde{V}^{-1} \tilde{D})^{-1} \tilde{D}' \tilde{V}^{-1} \bar{l}_{1,T}(\tilde{\theta}_{TS}, \pi)],$$

where the weight matrix in the QLR test, $W_{T,GMM}(\underline{\hat{\theta}}_F, \pi)$, has the same form as the weight matrix in Assumption S3. However, $W_{T,GMM}(\underline{\hat{\theta}}_F, \pi)$ uses first step estimates. We use only the unrestricted efficient weight matrix in QLR. Let $\tilde{\theta}_{TS}$ be the full sample standard two-step GMM estimators used in the standard GMM literature. In the QLR test, the second term in the brackets is the unrestricted partial sample two-step GMM objective function. This can be obtained by using Definition 1, replacing the continuous updating GMM with two-step GMM estimates. The other terms in the test statistics are defined as: $\bar{l}_{1,T}(\tilde{\theta}_{TS},\pi) = \frac{1}{T}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_{TS})$,

$$\tilde{V} = \frac{1}{T}\sum_{t=1}^{T}[\psi_t(\tilde{\theta}_{TS}) - 1/T\sum_{j=1}^{T}\psi_j(\tilde{\theta}_{TS})][\psi_t(\tilde{\theta}_{TS}) - 1/T\sum_{j=1}^{T}\psi_j(\tilde{\theta}_{TS})]',$$

and $\tilde{D} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\tilde{\theta}_{TS})}{\partial\theta'}$.

We closely follow the setup in Stock and Wright (2000) for analyzing the small sample properties of the various test statistics described in the above paragraph. The G Euler equations are

$$h(Y_t,\theta) = \delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{t+1} - \iota_G, \tag{11}$$

where δ is the discount factor, γ is the parameter for risk aversion, C_t is the consumption, R_t is a $G \times 1$ vector of asset returns, and ι_G is a G vector of ones. Then

$$\psi_t(\theta) = [\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{t+1} - \iota_G]\bigotimes Z_t, \tag{12}$$

where Z_t is a set of K instruments. Let $\theta = (\gamma, \delta)$, and assume parameters are in the interior of compact subset of R^2. As in Stock and Wright (2000), γ is deemed to be unidentified in large samples and δ identified. The errors are martingale difference sequences at true values so correction for autocorrelation is not used. There are no overlapping data as well. However, for $\sup K_T$, and sup LM tests we try also heteroskedasticity-autocorrelation correction for variance-covariance matrix estimator but this did not make any difference, so we use the less computationally intensive heteroskedasticity corrected variance covariance matrix estimators. The designs in the Monte Carlo are due to Tauchen (1986), Kocherlakota (1990), and Hansen, Heaton, and Yaron (1996). These designs are consistent with the Euler equations. Four designs are described in Table 3. Their method fits a VAR (1) to approximate consumption and dividend growth. Let c_t be the log growth rate of US per capita real annual consumption and d_t the log growth rate of real annual dividends on the S&P 500. This is given by

$$\begin{pmatrix} c_t \\ d_t \end{pmatrix} = \begin{pmatrix} 0.021 \\ 0.04 \end{pmatrix} + \begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}\begin{pmatrix} c_{t-1} \\ d_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{ct} \\ \epsilon_{dt} \end{pmatrix}, \tag{13}$$

where $var\epsilon_{dt} = 0.014, var\epsilon_{ct} = 0.0012, cov(\epsilon_{ct},\epsilon_{dt}) = 0.00177$. This is the setup for Tables 3 and 4.

In Table 4, we consider the rejection rates of Sup LR and $\sup S$ tests under the null of no structural change for the models in Table 3. The critical values are in Table 1. We also analyze the sup LM and LR-like tests of Andrews (1993a) in Table 4. The limits of the sup LM and LR-like tests of Andrews (1993a), and the critical values are tabulated in Table 1 of Andrews (1993a).

The rejection frequencies of sup LR and $\sup S$ tests are slightly above the nominal level of 10%. These improve when the sample size increases from 100 to 200. However, the LR-like and sup LM

Table 3: Monte Carlo Design

Design	γ_0	δ_0	Assets	Instruments
1	1.3	0.97	r_t^s	$1, r_{t-1}^s, c_{t-1}$
2	13.7	1.139	r_t^s	$1, r_{t-1}^s, c_{t-1}$
3	1.3	0.97	r_t^s, r_t^f	$1, c_{t-1}$
4	1.3	0.97	r_t^s, r_t^f	$1, c_{t-1}, r_{t-1}^s, r_{t-1}^f$

Note: $c_t = ln(C_t/C_{t-1}), r_t^f, r_t^s$ *represent consumption growth, the risk free rate, and the stock returns respectively. The program for Monte Carlo study can be obtained from the author on request.*

tests of Andrews (1993a) have large rejection rates of the true null of no structural change. In Table 4, we do not see declines in these rejection rates when the sample size increases for the sup LM test. In Design 4 when $GK = 8$, the size is very large for sup S, sup K, and QLR tests at $T = 100$, however when the sample size increases to 200, sup K test performs the best among the tests with 8% size.

The sup LM test of Andrews (1993a) is used extensively in the literature. Under weak instruments, sup LM test statistic may lead to finding structural change even when the parameters are stable. This has strong implications for applied work, especially in the case of asset pricing models and regarding the stability of Euler equations in those models.

We also run size exercises to see how tests fare under standard identification. These are given by two designs used in Kleibergen (2005). So we analyze Design 1 in Table 3 but with (13) coefficient matrix on $(c_{t-1}, d_{t-1})'$ multiplied by two. This is strong identification setup. This is called Design 5. Another possibility is using the same setup as Design 5 but changing the covariance between c_t, d_t in (13) to 0.0039. This is called Design 6. The results are in Table 5. In Table 5, we see that all tests have good size. Only sup K test is oversized at 18.9% for $T = 100$ in Design 5. However, clearly sup LR test is conservative. Then we compare the results in Table 5 with Design 1 in Table 4. This is comparison of a strongly identified systems with weakly identified one. We basically see a slight decline in size in all tests when there is strong identification. But we see that in small samples, the tests are not very conservative except from sup LR.

To see the degree of the conservativeness of the bounds we replicate Table 4 size exercise with $T = 400$ for Designs 1-3. Design 4 is very computationally intensive at this large sample size, so we are not able to analyze that. We see declines in size with this large sample size. The results are in Table 6. For example in Design 2 we see that at 10% nominal level, actual size is 6% at $T = 400$ for $\sup S, \sup K$ tests. We also see that QLR test has better size compared to Table 4 when the sample size is large in the case of weak identification. But the problem with sup LM test persists in the case of Design 2.

We analyze the power of the tests in Tables 7-8. The power is adjusted by using the finite

Table 4: Rejection Frequency of True Null at 10% level, Weak Instruments

	$T = 100$				$T = 200$			
Tests	Design 1	Design 2	Design 3	Design 4	Design 1	Design 2	Design 3	Design 4
$\sup S_T$	11.1	12.7	18.8	57.8	3.9	7.0	7.8	19.3
$\sup LR_T$	1.5	2.8	1.8	4.3	0.0	0.1	0.1	0.1
$\sup K_T$	5.7	8.6	13.4	39.5	3.6	7.0	4.8	8.0
$\sup LM$	17.9	28.1	16.9	9.3	15.2	30.5	19.8	14.4
QLR	16.9	17.2	1.6	49.0	13.3	16.0	1.2	26.8

Note: D1-4 represents the designs for Monte Carlo described in Table 3. The critical value for the test statistics sup S and sup LR is 16.4 for D1-2 for $GK = 3$ *in Table 1, and 19.7 for D3,* $GK = 4$*, for* $GK = 8$*, the critical value is 31.6. The critical value for sup K is 12.9 for* $n = 2$*. The critical values of QLR (the LR-like test of Andrews) and the sup LM of Andrews (1993a) is 10.01 corresponding to* $p = 2$*, Table 1 of Andrews (1993a), at the 10% level. 1000 repetitions are used to generate the table.*

Table 5: Rejection Frequency of True Null at 10% level, Standard Identification

	$T = 100$		$T = 200$	
Tests	Design 5	Design 6	Design 5	Design 6
$\sup S_T$	7.9	10.3	5.5	5.8
$\sup LR_T$	0.7	1.9	0.1	0.3
$\sup K_T$	18.9	7.0	8.5	6.1
$\sup LM$	12.2	3.5	7.7	4.2
QLR	14.6	0.1	10.1	0.0

Note: D5-D6 are described in section 4. The critical value for the test statistics sup S and sup LR is 16.4 for $GK = 3$ *in Table 1. The critical value for sup K is 12.9 for* $n = 2$*. The critical values of QLR (the LR-like test of Andrews) and the sup LM of Andrews (1993a) is 10.01 corresponding to* $p = 2$*, Table 1 of Andrews (1993a), at the 10% level. 1000 repetitions are used to generate the table.*

Table 6: Rejection Frequency of True Null at 10% level, $T = 400$

Tests	Design 1	Design 2	Design 3
$\sup S_T$	3.3	6.0	4.4
$\sup LR_T$	0.2	0.4	0.0
$\sup K_T$	2.2	5.4	3.4
$\sup LM$	10.0	29.0	22.6
QLR	0.8	8.1	2.7

Note: The critical values are described in Table 4.

sample critical values in Designs 1-3 under the null. Our design is as follows:

$$E[h(Y_t, \theta_{10})|F_t] = 0, \quad \text{for t} = 1, 2, \cdots [T/2], \tag{14}$$

$$E[h(Y_t, \theta_{20})|F_t] = 0, \quad \text{for t} = [T/2] + 1, \cdots, T, \tag{15}$$

where $\theta_{10} = (\gamma_{10}, \delta_{10}) = (1.3, 0.97)$ and $\theta_{20} = (\gamma_{20}, \delta_{20}) = (13.7, 1.139)$. The first half of the data is generated from Design 1 in Table 3, and the second half is generated from Design 2. In setup 1, the null is Design 1 and the finite sample critical values are used from Design 1 in the power exercise. In setup 2, the finite sample critical values are obtained from the null in Design 2 in Table 3. The finite sample critical values for each setup and test statistics can be obtained from the author on request. We also calculate size unadjusted power. This is more practical compared to size adjusted power. The results are in Table 7 too. First the rejection frequencies at 10% level calculated when the data is generated according to (13)-(15). The critical values are 16.4 (corresponding to $GK = 3, \pi_0 = 0.15$ in Table 1) for $\sup S, \sup LR$ tests, and 12.9 (corresponding to $n = 2$ in Theorem 2) for $\sup K$ test. For $\sup LM, QLR$ tests of Andrews (1993a) the critical value for $p = 2$ in his Table 1 is 10.01.

From Table 7, we see that the $\sup S$ test has the best power among the four. At $T = 100$, the $\sup S$ has moderate power whereas the other tests have low power. When $T = 200$, the $\sup S_T$ test has a 91-94% rejection rate whereas the other tests reject at a 41-59% rate. In the unadjusted power case, we see dramatic increases in power when the sample size increases from 100 to 200. Also we observe that $\sup S$ test dominates the others with 88% power, the next best is $\sup K$ test at 79% rejection rate. We also try the power exercise with different weights attached to Design 1 and Design 2 rather than equal weights of 50% each, but again the results stay the same: the $\sup S$ test has better power than the others.

Next, in Table 8 the first half of the data is generated as in Design 5, and the next half is generated according to Design 2. This involves standard identification in Design 5, and weak identification in Design 2. So the DGP follows (14)-(15), and the parameters are described immediately below (15). However, in Design 5 we use two times the coefficient matrix in front of $(c_{t-1}, d_{t-1})'$ in (13). Design 2 only uses (13) for generating instruments. In Setup 3, the null design is Design 2 and the finite sample critical values are used from Design 2 in size adjusted power and critical value for the unadjusted power is the same as in Table 7. Setup 4 uses the finite sample critical values from the null of Design 5 in size adjusted power, the critical values used for unadjusted power in Table 8 are the same as in Table 7. We see from Table 8 that at $T = 200$ $\sup S$ test has the best power among the tests for both size adjusted and unadjusted cases. For example, $\sup S$ test has 86.4% power whereas $\sup K$ test has 65.3%, and Andrews (1993a) sup LM test has 77.6% power in the unadjusted case for $T = 200$. Furthermore, at Setup 3, sup S test has 86.4% power whereas sup K has 77.3% power and sup LM test has 20.9% power.

Table 7: Power at 10% level, Weak Instruments

	Size Adjusted Power				Unadjusted Power	
	$T = 100$		$T = 200$		$T = 100$	$T = 200$
Tests	Setup 1	Setup 2	Setup 1	Setup 2		
$\sup S_T$	53.5	40.2	94.0	90.7	50.2	88.0
$\sup LR_T$	27.3	23.3	41.5	40.0	7.3	14.0
$\sup K_T$	49.0	28.5	82.8	71.1	27.6	79.4
$\sup LM$	29.8	19.2	57.3	23.2	44.4	70.2
QLR	23.3	23.1	58.9	29.2	41.1	63.5

Note: The numbers represent rejection frequency of the false null against the alternative (14)-(15). Setup 1 shows the power when the finite sample critical values are generated according to Design 1 in Table 3. Setup 2 shows the power when the finite sample critical values are generated from the distribution according to Design 2 in Table 3. 1000 replications are used.

Table 8: Power at 10% level

	Size Adjusted Power				Unadjusted Power	
	$T = 100$		$T = 200$		$T = 100$	$T = 200$
Tests	Setup 3	Setup 4	Setup 3	Setup 4		
$\sup S_T$	6.5	54.1	86.4	92.7	48.2	86.4
$\sup LR_T$	0.05	13.3	8.4	21.7	0.11	0.5
$\sup K_T$	31.8	12.6	73.3	68.9	29.3	65.3
$\sup LM$	9.0	34.2	20.9	81.1	38.4	77.6
QLR	12.5	30.6	42.8	69.9	37.7	69.9

Note: The numbers represent rejection frequency of the false null against the alternative (14)-(15). Setup 3 shows the power when the finite sample critical values are generated according to Design 2 in Table 3. Setup 4 shows the power when the finite sample critical values are generated from the distribution according to Design 5 in section 4. 1000 replications are used.

Table 9: Results from the Empirical Exercise

Model	Assets	GK	Instruments	$\sup S_T$	$\sup LR_T$	$\sup K_T$	sup LM	QLR
CRRA1	SR	3	SR,CG	11.6	3.2	14.2*	3.1	12.6*
CRRA2	SR	3	DY,CG	19.1*	7.1*	14.3*	7.9	12.9*
CRRA3	BR	3	BR,CG	22.2*	17.5*	8.4	15.4*	16.7*
CRRA4	BR	3	SP,CG	22.5*	17.1*	10.6	18.0*	12.6*

*Note: * indicates rejection at 10% level. The critical values for* $\sup LR_T, \sup S_T$ *tests are 16.4 for* $GK = 3$, *for* $\sup K_T$ *test the critical value is 12.9 for* $n = 2$, *from Table 1.* ($\pi = .15$). *GK represents the number of Euler equations multiplied by number of instruments.* n *represents the number of parameters tested. The critical value for QLR and sup LM is 10.01 from Table 1 of Andrews (1993a). This corresponds to* $n = 2$ *(two parameters) and* $\pi = .15$. *SR = Stock returns. CG =Consumption Growth. BR =Bond Returns, DY =Dividend Yield, SP=Spread. All the instruments are lagged once.*

5 Empirical Results

In this section we consider CRRA model that is used for simulation in section (i.e. equations (11)-(12)). The data set is the same one used in Stock and Wright (2000). We take γ to be the weakly identified parameter and δ to be the strongly identified parameter as given in Stock and Wright (2000). We conduct four different test statistics for structural change : $\sup S, \sup LR, \sup K$ which are proposed in this article and LR like (QLR) and $\sup LM$ tests of Andrews (1993a).

The data set is the updated version of Campbell and Shiller's (1987) annual data, covering the period 1889-1991[3]. It consists of U.S. stock returns, bond returns, consumption, spread, and dividend yield. Consumption is the real consumption of nondurable and services per capita. The bond returns are calculated using the nominal interest rate for prime 4-6 month commercial paper. The stock returns are obtained by using the Cowles Commission index and by following the annual average price of S&P monthly composite index. The returns are real (i.e. the producer price index is used). Spread is the difference between the yield on long term US treasury bonds and the commercial paper rate. The details are in Campbell and Shiller (1987), Shiller (1982).

We see that in all of the cases, sup LM test is in total contradiction with $\sup K$ test. $\sup K$ test has very good size and power and less conservative compared to other tests. These results are due to bad small sample properties of sup LM test as can be observed in simulations. $\sup S$ test is different from sup LM in one case, namely CRRA2. For calculation of the heteroskedasticity and autocorrelation consistent variance covariance matrix we use Bartlett Kernel with bandwidth size of 4 which is picked according to Andrews (1991). We also experiment with other bandwidth choices , but the results do not change.

[3] We thank J.Wright for pinpointing the typo regarding the period of the dates of this data set in their article

6 Conclusion

This paper develops weak instrument asymptotics for structural change tests: the likelihood ratio and S-based tests. Even though the test statistics are not asymptotically pivotal, they are asymptotically boundedly pivotal. In simulations, we realize that sup S and sup K tests have good sample properties. Another interesting topic to search may be structural change tests in the framework of many weak moment conditions.

Appendix

The main proofs are collected here.

Proof of Theorem 1.

Consider $S_T(\tilde{\theta}(\pi),\pi)$ which represents objective function at restricted partial sample continuous updating estimator where

$$\tilde{\theta}(\pi) = \arg\min_{\theta\in\Theta} S_T(\theta,\pi), \tag{16}$$

for all $\pi \in \Pi$, and

$$S_T(\theta,\pi) = \begin{bmatrix} T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta) \\ T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\theta)\end{bmatrix}' \begin{bmatrix} \frac{V_{T1}(\theta,\pi)^{-1}}{\pi} & 0 \\ 0 & \frac{V_{T2}(\theta,\pi)^{-1}}{1-\pi}\end{bmatrix} \begin{bmatrix} T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta) \\ T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\theta)\end{bmatrix}. \tag{17}$$

So by (16) for all $\pi \in \Pi$,

$$S_T(\tilde{\theta}(\pi),\pi) \leq S_T(\theta_0,\pi). \tag{18}$$

By (18), and $LR_T(\pi)$ definition in (2)

$$\sup_{\pi\in\Pi} LR_T(\pi) \leq \sup_{\pi\in\Pi}[S_T(\theta_0,\pi) - S_T(\underline{\hat{\theta}}(\pi),\pi)].$$

Note that the second term on the right hand side of the inequality is positive with probability one in the case of overidentification, and zero when there is exact identification. So the limit for $supLR_T(\pi)$ is bounded above by the limit for the first term on the right hand side of the inequality. In order to find the limit of $\sup S_T(\theta_0,\pi)$ see that by (17)

$$\begin{aligned}\sup_{\pi\in\Pi} S_T(\theta_0,\pi) &= \sup_{\pi}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)]'\left[\frac{V_{T1}(\theta_0,\pi)^{-1}}{\pi}\right][T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)] \\ &+ \sup_{\pi}[T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\theta_0)]'\left[\frac{V_{T2}(\theta_0,\pi)^{-1}}{1-\pi}\right][T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\theta_0)].\end{aligned}$$

Then by Assumptions 1 and 2,

$$\begin{aligned}\sup_{\pi\in\Pi} S_T(\theta_0,\pi) &\overset{d}{\to} \sup_{\pi\in\Pi} B_{GK}(\pi)'\frac{\Omega_{\theta_0,\theta_0}^{-1}}{\pi}B_{GK}(\pi) + \sup_{\pi\in\Pi}\frac{[B_{GK}(1)-B_{GK}(\pi)]'\Omega_{\theta_0,\theta_0}^{-1}[B_{GK}(1)-B_{GK}(\pi)]}{1-\pi} \\ &\equiv \sup_{\pi\in\Pi}\frac{W_{GK}(\pi)'W_{GK}(\pi)}{\pi} + \sup_{\pi\in\Pi}\frac{[W_{GK}(1)-W_{GK}(\pi)]'[W_{GK}(1)-W_{GK}(\pi)]}{1-\pi}.\end{aligned} \tag{19}$$

The results for $\sup S_T$ test statistic follows from (18), (19) . □

We now discuss why the limit in Theorem 1 is robust to identification problems. This is clear from the proof of Theorem 1, but we want to discuss this in detail. First let us take the case of weak identification, using the partial sample version of the identity in Stock and Wright (2000) we have

$$ET^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\theta)=\tilde{m}_T(\alpha,\beta,\pi)=\tilde{m}_T(\alpha_0,\beta_0,\pi)+\tilde{m}_{1T}(\alpha,\beta,\pi)+\tilde{m}_2(\beta,\pi),$$

where $\tilde{m}_{1T}(\alpha,\beta,\pi)=\tilde{m}_T(\alpha,\beta,\pi)-\tilde{m}_T(\alpha_0,\beta,\pi)$ and $\tilde{m}_{2T}(\beta,\pi)=\tilde{m}_T(\alpha_0,\beta,\pi)-\tilde{m}_T(\alpha_0,\beta_0,\pi)$ for all $\pi\in\Pi$. Note that for all π , $\tilde{m}_T(\alpha_0,\beta_0,\pi)=0$. Then set $\tilde{m}_{2T}(\beta,\pi)=\pi m_2(\beta)$ to simplify the notation. As in full sample case in Stock and Wright (2000) set

$$\tilde{m}_{1T}(\theta,\pi)=\frac{m_1(\theta,\pi)}{T^{1/2}}.$$

Note that to find the limit (19) in the weakly identified case instead of $T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)$ we could have used

$$T^{-1/2}\sum_{t=1}^{[T\pi]}(\psi_t(\theta_0)-E\psi_t(\theta_0))+ET^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0). \tag{20}$$

See that in the weakly identified case

$$ET^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)=\frac{m_{1T}(\alpha_0,\beta_0)}{T^{1/2}}+\pi m_2(\beta_0),$$

where at the true value of the parameters we know that

$$ET^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)=0.$$

So there is no point in adding and subtracting $ET^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)$ as used in weak identification literature. In other words, weak identification assumption as in equations before and after (20) are irrelevant.

In the unidentified case for α

$$ET^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)=\pi m_2(\beta_0),$$

but again we know that

$$ET^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)=0.$$

So there is no point in adding and subtracting the expectation term to $T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)$. The limit is definitely robust to identification and works under weak assumptions. Clearly assumptions 1 and 2 are enough to obtain the limit of the bound.

Now we provide the proof of Theorem 2. This is similar to the proof of Theorem 1.

Proof of Theorem 2. First of all, by the definition of $\tilde{\theta}_K(\pi)$ in (4) it is clear that

$$\sup_{\pi} K_T(\tilde{\theta}_K(\pi),\pi) \leq \sup_{\pi} K_T(\theta_0,\pi),$$

where we can set up $K_T(\theta_0,\pi)$ as

$$K_T(\theta_0,\pi) = K_T^1(\theta_0,\pi) + K_T^2(\theta_0,\pi),$$

where

$$K_T^1(\theta_0,\pi) = \frac{1}{\pi}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)]'(\hat{\Omega}^1_{\theta_0,\theta_0})^{-1/2}P_{(\hat{\Omega}^1_{\theta_0,\theta_0})^{-1/2}\tilde{D}_T^1(\theta_0,\pi)}(\hat{\Omega}^1_{\theta_0,\theta_0})^{-1/2}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)],$$

$$K_T^2(\theta_0,\pi) = \frac{1}{1-\pi}[T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta_0)]'(\hat{\Omega}^2_{\theta_0,\theta_0})^{-1/2}P_{(\hat{\Omega}^2_{\theta_0,\theta_0})^{-1/2}\tilde{D}_T^2(\theta_0,\pi)}(\hat{\Omega}^2_{\theta_0,\theta_0})^{-1/2}[T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta_0)],$$

These are basically taken from equation (5)-(8) and evaluated at θ_0. We consider the limit of $K_T^1(\theta_0,\pi)$ first. The analysis for $K_T^2(\theta_0,\pi)$ is the same. First, note that by (6)

$$\begin{aligned}\tilde{D}_T^1(\theta_0,\pi) &= [q_{1T}(\theta_0,\pi) - \hat{\Omega}^1_{q_0\theta_0,1}(\hat{\Omega}^1_{\theta_0\theta_0})^{-1}(T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)),\cdots,q_{nT}(\theta_0,\pi)\\ &- \hat{\Omega}^1_{q_0\theta_0,n}(\hat{\Omega}^1_{\theta_0\theta_0})^{-1}(T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0))],\end{aligned}$$

where $q_{iT}(\theta_0,\pi)$ is defined in equation (7) at θ_0. Define $J^1(\theta_0,\pi) = \lim_{T\to\infty} E[T^{-1}\sum_{t=1}^{[T\pi]} q_t(\theta_0)]$ and $q_t(\theta_0) = \partial\psi_t(\theta_0)/\partial\theta'$ $(GK\times n)$ is the partial derivative. Note that adapting Kleibergen (2005)'s ideas for our partial sample case $J^1(\theta_0,\pi)$ can have a fixed full rank value, or a weak value in which

$$J^1(\theta_0,\pi) = \frac{C_1}{T^{1/2}},$$

where $C_1: GK\times n$ matrix. The other possibility is that $J^1(\theta_0,\pi)=0$, the case of no identification.

We want to analyze Kleibergen (2005) type of test that is used in Theorem 2 under these three conditions, and show that the bound is the same regardless of identification issues. See that

$$\begin{aligned}\begin{bmatrix} I_{GK} & 0 \\ -\hat{\Omega}^1_{q_0,\theta_0}(\hat{\Omega}^1_{\theta_0,\theta_0})^{-1} & I_{nGK}\end{bmatrix} &\times \begin{bmatrix} T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0) \\ T^{-1/2}\sum_{t=1}^{[T\pi]}\bar{q}_t(\theta_0)\end{bmatrix}\\ &= \begin{bmatrix} T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0) \\ \tilde{D}_T^1(\theta_0,\pi) - J^1(\theta_0,\pi)\end{bmatrix}\\ &\Longrightarrow \begin{bmatrix} B_{GK}(\pi) \\ B_{2.1}(\pi)\end{bmatrix},\end{aligned} \tag{21}$$

by Assumptions 3 and 4. Since this is true for any value for $J^1(.)$ this is robust to identification issues. Note that $B_{2.1}(\pi)$ is a Brownian Motion and

$$B_{2.1}(\pi) = B_{nGK}(\pi) - \Omega_{q_0,\theta_0}(\Omega_{\theta_0,\theta_0})^{-1}B_{GK}(\pi),$$

and $B_{2.1}(\pi) = \Omega^{1/2}_{q_0 q_0 \theta_0} W_{nGK}(\pi)$ and $\Omega_{q_0 q_0 \theta_0} = \Omega_{q_0 q_0} - \Omega_{q_0 \theta_0}\Omega^{-1}_{\theta_0\theta_0}\Omega_{\theta_0 q_0}$. Clearly $B_{2.1}(\pi)$ is independent of $B_{GK}(\pi)$. This result (21) is the partial sample counterpart of Lemma 1 in Kleibergen (2005). Then via (21), since $\tilde{D}^1_T(\theta_0,\pi)$ is asymptotically independent from $T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)$, following the proof of Theorem 1 in Kleibergen (2005) for the partial sample case we derive

$$[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta_0)]'(\hat{\Omega}^1_{\theta_0\theta_0})^{-1}\tilde{D}^1_T(\theta_0,\pi)[\tilde{D}^1_T(\theta_0,\pi)'(\hat{\Omega}^1_{\theta_0\theta_0})^{-1}\tilde{D}^1_T(\theta_0,\pi)]^{-1/2} \Longrightarrow W_n(\pi), \tag{22}$$

where $W_n(\pi)$ is n dimensional standard Brownian Motion.

Use (22) in $K^1_T(\theta_0,\pi)$ to have

$$K^1_T(\theta_0,\pi) \Longrightarrow \frac{W_n(\pi)'W_n(\pi)}{\pi}. \tag{23}$$

In the same manner

$$K^2_T(\theta_0,\pi) \Longrightarrow \frac{(W_n(1)-W_n(\pi))'(W_n(1)-W_n(\pi))}{(1-\pi)}. \tag{24}$$

Then (23)-(24) provides

$$\sup_\pi K_T(\theta_0,\pi) \Longrightarrow \sup_\pi \left[\frac{W_n(\pi)'W_n(\pi)}{\pi} + \frac{(W_n(1)-W_n(\pi))'(W_n(1)-W_n(\pi))}{(1-\pi)}\right]. \tag{25}$$

Then simple algebra after (25) provides the result. □

Before proving Lemma 1, we need Ottaviani's inequality for α mixing variables in Theorem 8 of Doukhan (1994, p.43). Since α mixing implies β mixing, we can use the inequality for the β mixing variables. Lemma A.1 holds only under the null hypothesis of no structural change, since the random variables are strictly stationary. Lemma A.1 holds under Assumption 1.

Lemma A.1. *(Ottaviani's inequality for β mixing random variables) Let X_t be a strictly stationary random sequence centered at expectation such that $E|X_t|^{2+\delta} < \infty$ for some $\delta > 0$ and $S_T = \sum_{t=1}^T X_t$, then if the sequence is β mixing with $\beta(s) = O(s^{-A})$ for some $A > 2 + 4/\delta$, there exists positive C, c with*

$$\left[\frac{1}{2} - \max_{1\le k\le T} P\left(|S_T - S_k| > y\sqrt{T}\right)\right] P\left(\max_{1\le k\le T}|S_k| > x\right) \le P\left(|S_T| > x - 2y\sqrt{T}\right) + CT^{-d}$$

for $x > 0$, $y > c$, and $0 < d < \frac{A(2+\delta)}{2(2+\delta+A)} - 1$.

Note that there are two typographical errors in α mixing coefficients in Doukhan (1994). First, the correct coefficient should be $A > 2 + 4/\delta$ since we need $\frac{A(2+\delta)}{2(2+\delta+A)} > 1$. Another typographical error in Doukhan's inequality involves sum S_T. It should not be divided by $T^{1/2}$ in the statement of the inequality.

Next, we need the following result before the proof of Lemma 1. Under Assumptions S1i-iii, uniformly in θ, setting $\Psi_T(\theta) = T^{-1/2}\sum_{t=1}^T[\psi_t(\theta) - E\psi_t(\theta)]$,

$$\Psi_T(\theta) \Longrightarrow \Psi(\theta), \tag{26}$$

where $\Psi(\theta)$ is a Gaussian stochastic process on Θ with mean zero and covariance function $E\Psi(\theta_1)\Psi(\theta_2)' = \Omega_{\theta_1,\theta_2}$.

The m-dependent version of this is given as Assumption B of Stock and Wright (2000). Equation (26) is already shown in Doukhan, Massart and Rio (1995, application 2 of Theorem 1) and the same result is given in Andrews (1993b, p.200). The mixing rate assumption here is stronger than the one in Andrews (1993b). This is due to the fact that a new rate is needed to ensure the Ottaviani's inequality (Lemma A.1) is holding.

Proof of Lemma 1. The proof is very similar to Theorem 2.12.1 of van der Vaart and Wellner (1996) for iid data. We replace Ottaviani's inequality for independent data with our Lemma A.1. We also use (26) in the proof rather than weak convergence of iid empirical processes. Here we provide the proof. The pseudometric that is used is $|\pi_1 - \pi_2| + \rho(\theta_1, \theta_2)$ (on $[0,1] \times \Theta$) where $\rho(\theta_1, \theta_2) = [E(\psi_t(\theta_1) - \psi_t(\theta_2))^p]^{1/p}$, $p > 2$. Note that the proof is for real valued $\psi_t(.)$. The extension to the multivariate case is at the end of the proof of Lemma 1.

As in Theorem 2.12.1 of van der Vaart and Wellner (1996), we have to show that, given Assumptions S1i-iii, (26) implies Lemma 1. We have to show the asymptotic equicontinuity of $\Psi_T(\theta, \pi)$. Denote the set of functions $\psi_t(\theta) \in \mathcal{F}$ satisfying Assumptions S1i-iii. Then set, as in van der Vaart and Wellner (1996), $\mathcal{F}_\tau = \{\psi_t(\theta_1) - \psi_t(\theta_2) : \psi_t(.) \in \mathcal{F}, \rho(\theta_1, \theta_2) < \tau\}$. Denote $\|.\|_{\mathcal{F}_\tau}, \|.\|_{\mathcal{F}}$ as the uniform norm for maps from $\mathcal{F}_\tau$, $\mathcal{F}$ to R respectively. By the triangle inequality

$$\begin{aligned} \sup_{|\pi_1-\pi_2|+\rho(\theta_1,\theta_2)<\tau} |\Psi_T(\theta_1, \pi_1) - \Psi_T(\theta_2, \pi_2)| &\leq \sup_{|\pi_1-\pi_2|<\tau} \|\Psi_T(\theta_1, \pi_1) - \Psi_T(\theta_1, \pi_2)\|_{\mathcal{F}} \\ &+ \sup_{0\leq\pi_2\leq 1} \|\Psi_T(\theta_1, \pi_2)\|_{\mathcal{F}_\tau}. \end{aligned} \tag{27}$$

For the second term on the right-hand side of (27), we discretize π_2 such that it takes the value k/T with $k = 1, 2, \cdots, T$. We can rewrite the sequential empirical process as

$$\Psi_T(\theta, \pi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\pi]} [\psi_t(\theta) - E\psi_t(\theta)] = \sqrt{\frac{[T\pi]}{T}} \Psi_{[T\pi]}(\theta), \tag{28}$$

where $\Psi_{[T\pi]}(\theta) = \frac{1}{\sqrt{[T\pi]}} \sum_{t=1}^{[T\pi]} [\psi_t(\theta) - E\psi_t(\theta)]$. Use (28) and discretization of π_2 to rewrite the second term on the right hand side of (27) as

$$\sup_{0\leq\pi_2\leq 1} \|\Psi_T(\theta_1, \pi_2)\|_{\mathcal{F}_\tau} = \max_{k\leq T} \|\sqrt{\frac{k}{T}} \Psi_k(\theta)\|_{\mathcal{F}_\tau}. \tag{29}$$

Next analyze (29) via Lemma A.1, for all $\epsilon > 0$ and d in Lemma A.1,

$$\left[\frac{1}{2} - \max_{k\leq T} P(\sqrt{\frac{k}{T}} \|\Psi_k(\theta)\|_{\mathcal{F}_\tau} > \epsilon)\right] \left[P\left(\max_{k\leq T} \sqrt{\frac{k}{T}} \|\Psi_k(\theta)\|_{\mathcal{F}_\tau} > 2\epsilon\right)\right] \leq P(\|\Psi_T(\theta)\|_{\mathcal{F}_\tau} > \epsilon) + CT^{-d}. \tag{30}$$

By (26), $\Psi_T(\theta)$ is asymptotically equicontinuous, so the right hand side of the inequality in (30) converges to zero as $T \to \infty$ followed by $\tau \downarrow 0$. We analyze the first term in square brackets on the left hand side of (30). When $k \leq T_0$, $\sqrt{k}\|\Psi_k(\theta)\|_{\mathcal{F}_\tau}$ is bounded. This can be seen in van der Vaart and Wellner (1996, p.227) regardless of the time series nature of the data. For large T_0, when $k > T_0$, the first term with square brackets on the left hand side of (30) is bounded away from zero. This is obtained by asymptotic equicontinuity of $\Psi_T(\theta)$ via (26). Therefore, the second term on the right hand side of (27) converges to zero in probability as $T \to \infty$, followed by $\tau \downarrow 0$.

We need to prove the same result for the first term on the right hand side of (27), so we have to show that the following converges to zero:

$$P(\max_{0 \leq j\tau \leq 1} \sup_{j\tau \leq \pi_1 \leq (j+1)\tau} \|\Psi_T(\theta_1, \pi_1) - \Psi_T(\theta_1, j\tau)\|_{\mathcal{F}} > 2\epsilon), \tag{31}$$

where $j\tau$ helps discretize $\pi_2 - \tau$ and $\pi_2 + \tau$ in (27) as in van der Vaart and Wellner (1996, p.227). By stationarity of increments of $\Psi_T(\theta, \pi)$ in π (by Assumption S1), probability in (31) can be bounded by

$$\left[\frac{1}{\tau}\right] P(\sup_{0 \leq \pi_1 \leq \tau} \|\Psi_T(\theta_1, \pi_1)\|_{\mathcal{F}} > 2\epsilon).$$

Again discretize π_1, and use Lemma A.1 and (28) to have

$$\left[1/2 - \max_{k \leq [T\tau]} \left\{ P\left(\sqrt{\frac{k}{T}}\|\Psi_k(\theta)\|_{\mathcal{F}} > \epsilon\right)\right\}\right] \times \left[\frac{1}{\tau}\right]\left[P\left(\max_{k \leq T\tau} \sqrt{\frac{k}{T}}\|\Psi_k(\theta)\|_{\mathcal{F}} > 2\epsilon\right)\right] \leq$$
$$\left[\frac{1}{\tau}\right] P\left(\sqrt{\frac{[T\pi]}{T}}\|\Psi_{[T\pi]}(\theta)\|_{\mathcal{F}} > \epsilon\right) + CT^{-d}. \tag{32}$$

Benefiting from (8), and using the portmanteau theorem in van der Vaart and Wellner (1996, Theorem 1.3.4iii) as $T \to \infty$, the limit superior of probability of the right-hand side term in (32) is bounded by $P(\|\Psi(\theta)\|_{\mathcal{F}} \geq \epsilon/\tau^{1/2})$. Since norm of the Gaussian process $\Psi(\theta)$ has moments of all orders by Proposition A.2.3 of van der Vaart and Wellner (1996), limit for the probability of the right-hand side term converges to zero faster than any power of τ as $\tau \downarrow 0$. So the right hand side term in (32) converges to zero as $T \to \infty$ followed by $\tau \downarrow 0$. By a similar argument as before, the first term with square brackets on the left hand side of (32) converges to 1/2. This proves that probability in (31) converges to zero. We show that both of the terms on the right-hand side of (27) converge to zero in probability. So the stochastic equicontinuity of the sequential empirical process is proved. So we have

$$\Psi_T(\theta, \pi) \Longrightarrow \Psi(\theta, \pi). \tag{33}$$

However, in Lemma 1 $\psi(.)$ is of $GK > 1$ dimension. Therefore, we have to extend the result (33) to the multivariate case. In order to do that, we have to show $\Psi_T(\theta, \pi)$ is stochastically equicontinuous in the multivariate case. Then we have to show the convergence of finite dimensional distributions of $\Psi_T(\theta, \pi)$ in the multivariate case.

First, we analyze stochastic equicontinuity. Note that vector $\Psi_T(\theta, \pi)$ is asymptotically tight if, and only if, the components of vector are asymptotically tight by Lemma 1.4.3 of van der Vaart and Wellner (1996). So we have to show the component random elements of the vector $\Psi_T(\theta,\pi)$ are asymptotically tight. Note that by (33) we have

$$\gamma'\Psi_T(\theta,\pi) \Longrightarrow \gamma'\Psi(\theta,\pi),$$

for all $\gamma \in R^{GK}$. Let UB(S) represent uniformly bounded functions on S where S is an arbitrary set. Then UB $(\mathcal{F}\times[0,1])^{GK}$ is separable and complete by metric $d(f_1,f_2)=\sup_{\theta,\pi}\|f_1(\theta,\pi)-f_2(\theta,\pi)\|$ where $\|.\|$ is the Euclidean norm and $f\in UB(.)$ (van der Vaart and Wellner 1996, p.29-30). Now we can use the converse of Prohorov's Theorem (van der Vaart and Wellner 1996, Problem 1.12.4, p.74) to obtain asymptotic tightness of component random element of $\Psi_T(\theta,\pi)$. This provides the stochastic equicontinuity of the vector $\Psi_T(\theta,\pi)$.

Now we show finite dimensional convergence. For $\theta_1,\theta_2\in\Theta$, $0\le\pi_1<\pi_2\le 1$, we want convergence of $[\Psi_T(\theta_1,\pi_1)',\Psi_T(\theta_2,\pi_2)']'$. This is possible when $[\Psi_T(\theta_1,\pi_1)',\Psi_T(\theta_2,\pi_1)',\Psi_T(\theta_2,\pi_2)'-\Psi_T(\theta_2,\pi_1)']'$ converges. But this can happen if we have asymptotically independent increments in π given θ and weak convergence of $\Psi_T(\theta_2,\pi_2)-\Psi_T(\theta_2,\pi_1)$, for all $0\le\pi_1<\pi_2\le 1$.

First by (33) we have, for all $\gamma\in R^{GK}$,

$$\gamma'\Psi_T(\theta,\pi) \Longrightarrow \gamma'\Psi(\theta,\pi).$$

This implies weak convergence of $\gamma'\Psi_T(\theta_2,\pi_2)-\gamma'\Psi_T(\theta_2,\pi_1)$ to $\gamma'\Psi(\theta_2,\pi_2)-\gamma'\Psi(\theta_2,\pi_1)$. This in turn implies weak convergence of $\Psi_T(\theta_2,\pi_2)-\Psi_T(\theta_2,\pi_1)$ to $\Psi(\theta_2,\pi_2)-\Psi(\theta_2,\pi_1)$ by Cramer-Wold.

Now we show asymptotically independent increments in π given θ. Set $0\le\pi_0<\pi_1\le\pi_2\le 1$. Note that for $\theta_1,\theta_2\in\Theta$,

$$\gamma_1'\Psi_T(\theta_2,\pi_2)-\gamma_1'\Psi_T(\theta_2,\pi_1)+\gamma_2'\Psi_T(\theta_1,\pi_0)\xrightarrow{d} N(0,(\pi_2-\pi_1)\gamma_1'\Omega_{\theta_2,\theta_2}\gamma_1+\pi_0\gamma_2'\Omega_{\theta_1,\theta_1}\gamma_2),$$

by (33) β mixing property of $\psi_t(\theta)$, and the nature of the variance covariance matrix of the limit in Lemma 1. Then we have

$$\begin{pmatrix}\Psi_T(\theta_2,\pi_2)-\Psi_T(\theta_2,\pi_1)\\ \Psi_T(\theta_1,\pi_0)\end{pmatrix}\xrightarrow{d} N\left(0,\begin{bmatrix}(\pi_2-\pi_1)\Omega_{\theta_2,\theta_2} & 0\\ 0 & \pi_0\Omega_{\theta_1,\theta_1}\end{bmatrix}\right),$$

which proves asymptotically independent increments in π given θ. Note that, instead of $\Psi_T(\theta_1,\pi_0)$ we could have used $\Psi_T(\theta_2,\pi_0)$ in the result above; this does not change the result of asymptotically independent increments. □

Proof of Theorem 3b. The proof for Theorem 3b basically consists of three main steps. In the first step, we need the limit for restricted CUE estimator. This is shown in Technical Lemma A.2. Then we substitute this result into $\sup S_T$ test statistic and derive the limit. In the third step, the limit expression is put into a useful form for comparison by benefiting from the properties of Brownian Motion.

We can immediately benefit from the limit of

$$T^{1/2}(\tilde{\theta}(\pi)-\theta_0) \Longrightarrow u^*,$$

where

$$u^* \equiv -[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}[R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0}W_{GK}(1)].$$

where $W_{GK}(\pi)$ is a GK dimensional standard Brownian motion, whereas $W_{GK}(1)$ is a GK dimensional standard normal vector. This is established in Technical Lemma A.2.

Using the test statistics

$$\sup_{\pi} S_T(\tilde{\theta}(\pi),\pi) \Longrightarrow \sup_{\pi} \bar{S}(u^*,\pi). \tag{34}$$

Using (17)

$$\begin{aligned} S_T(\tilde{\theta}(\pi),\pi) &= [\Psi_T(\tilde{\theta}(\pi),\pi)+T^{1/2}\pi m_1(\tilde{\theta}(\pi))]'\frac{V_{T1}(\tilde{\theta}(\pi),\pi)^{-1}}{\pi}[\Psi_T(\tilde{\theta}(\pi),\pi)+T^{1/2}\pi m_1(\tilde{\theta}(\pi))] \\ &+ [\Psi_T(\tilde{\theta}(\pi))-\Psi_T(\tilde{\theta}(\pi),\pi)+T^{1/2}(1-\pi)m_1(\tilde{\theta}(\pi))]'\frac{V_{T2}(\tilde{\theta}(\pi),\pi)^{-1}}{1-\pi} \\ &\times [\Psi_T(\tilde{\theta}(\pi))-\Psi_T(\tilde{\theta}(\pi),\pi)+T^{1/2}(1-\pi)m_1(\tilde{\theta}(\pi))]. \end{aligned}$$

Then use Lemma 1 Assumptions S1ii, S2 and Technical Lemmata A1,A2 (or (40)(41)) to have

$$\begin{aligned} \sup_{\pi} \bar{S}(u^*,\pi) &\equiv \sup_{\pi}[B_{GK}(\pi)+\pi R(\theta_0)u^*]'\frac{\Omega^{-1}_{\theta_0,\theta_0}}{\pi}[B_{GK}(\pi)+\pi R(\theta_0)u^*] \\ &+ \sup_{\pi}[B_{GK}(1)-B_{GK}(\pi)+(1-\pi)R(\theta_0)u^*]'\frac{\Omega^{-1}_{\theta_0,\theta_0}}{1-\pi} \\ &\times [B_{GK}(1)-B_{GK}(\pi)+(1-\pi)R(\theta_0)u^*]. \end{aligned} \tag{35}$$

The limit in (35) can be simplified benefiting from $B_{GK}(\pi)\equiv\Omega^{1/2}_{\theta_0,\theta_0}W_{GK}(\pi)$ and u^* expression

$$\begin{aligned} \sup_{\pi} \bar{S}(u^*,\pi) &= \sup_{\pi}\{\frac{1}{\pi}[W_{GK}(\pi)-\pi\Omega^{-1/2}_{\theta_0,\theta_0}R(\theta_0)[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0}W_{GK}(1)]' \\ &\times [W_{GK}(\pi)-\pi\Omega^{-1/2}_{\theta_0,\theta_0}R(\theta_0)[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0}W_{GK}(1)]\} \\ &+ \sup_{\pi}\{\frac{1}{1-\pi}[W_{GK}(1)-W_{GK}(\pi)-(1-\pi)\Omega^{-1/2}_{\theta_0,\theta_0}R(\theta_0)[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0}W_{GK}(1)] \\ &\times [W_{GK}(1)-W_{GK}(\pi)-(1-\pi)\Omega^{-1/2}_{\theta_0,\theta_0}R(\theta_0)[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0}W_{GK}(1)]\}. \end{aligned} \tag{36}$$

Now let

$$C=[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0},$$

where C is $n\times GK$ matrix and $(CC')^{-1/2}C[(CC')^{-1/2}C]'=I_n$. We set $A_C=[(CC')^{-1/2}C]$ where A_C is $n\times GK$ matrix. See that $A_CW_{GK}(\pi)\equiv W_n(\pi)$, $W_n(\pi)$ is an "n" vector Standard Brownian

Motion. Using these, limit in (36) further simplified as

$$\begin{aligned}\sup_{\pi} \bar{S}(u^*,\pi) &= \sup_{\pi} \frac{1}{\pi}[W_{GK}(\pi)'W_{GK}(\pi) - 2\pi W_{GK}(\pi)'A_C'A_C W_{GK}(1) + \pi^2 W_{GK}(1)'A_C'A_C W_{GK}(1)] \\ &+ \sup_{\pi} \frac{1}{1-\pi}[(W_{GK}(1) - W_{GK}(\pi))'(W_{GK}(1) - W_{GK}(\pi)) \\ &- 2(1-\pi)(W_{GK}(1) - W_{GK}(\pi))'A_C'A_C W_{GK}(1) \\ &+ (1-\pi)^2 W_{GK}(1)'A_C'A_C W_{GK}(1)]. \end{aligned} \tag{37}$$

This last limit also can be expressed in a more convenient way

$$\begin{aligned}\sup_{\pi} \bar{S}(u^*,\pi) &= \sup_{\pi}\left[\frac{W_{GK}(\pi)'W_{GK}(\pi)}{\pi} + \frac{[W_{GK}(1) - W_{GK}(\pi)]'[W_{GK}(1) - W_{GK}(\pi)]}{1-\pi}\right] \\ &- W_{GK}(1)'A_C'A_C W_{GK}(1). \end{aligned} \tag{38}$$

After Theorem 3 we simplify the term with square brackets in (38) with (3). Then, the limit expression in (38) can be rewritten as:

$$\sup_{\pi\in\Pi}\left[\frac{(W_{GK}(\pi) - \pi W_{GK}(1))'(W_{GK}(\pi) - \pi W_{GK}(1))}{\pi(1-\pi)}\right] + [W_{GK}(1)'W_{GK}(1) - W_{GK}(1)'A_C'A_C W_{GK}(1)]. \tag{39}$$

Analyzing the following terms in (39)

$$W_{GK}(1)'W_{GK}(1) - W_{GK}(1)'A_C'A_C W_{GK}(1) = W_{GK}(1)'[I_{GK} - A_C'A_C]W_{GK}(1).$$

However, using C and A_C definitions

$$I_{GK} - A_C'A_C = I_{GK} - \Omega_{\theta_0,\theta_0}^{-1/2}R(\theta_0)[R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}R(\theta_0)]^{-1}R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1/2}.$$

Note that $I_{GK} - A_C'A_C$ is idempotent with rank "GK -n". So

$$W_{GK}(1)'[I_{GK} - A_C'A_C]W_{GK}(1) \equiv \chi^2_{GK-n}. \tag{40}$$

It is easy to see (40) is independent from the first term in the square brackets in (39) so we have

$$\sup_{\pi\in\Pi} S_T(\pi) \xrightarrow{d} \sup_{\pi\in\Pi}\left[\frac{(W_{GK}(\pi) - \pi W_{GK}(1))'(W_{GK}(\pi) - \pi W_{GK}(1))}{\pi(1-\pi)}\right] + \chi^2_{GK-n},$$

where the limit consists of two independent terms. □

Proof of Theorem 4. Before the analysis of the limit for the restricted Kleibergen type of test statistic we consider

$$T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) = [T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) - E\psi_t(\tilde{\theta}_K(\pi))] + [T^{-1/2}\sum_{t=1}^{[T\pi]}E\psi_t(\tilde{\theta}_K(\pi))].$$

Under Assumption S1, following proof of Lemma 1, and using consistency of $\tilde{\theta}_K(\pi)$

$$T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde\theta_K(\pi)) - E\psi_t(\tilde\theta_K(\pi)) \Longrightarrow B_{GK}(\pi) \equiv \Omega_{\theta_0\theta_0}^{1/2}W_{GK}(\pi). \tag{41}$$

Then by Assumption S2, mean value expansion, and by Technical Lemma A.4

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[T\pi]}E\psi_t(\tilde\theta_K(\pi)) &= T^{-1/2}\sum_{t=1}^{[T\pi]}E\psi_t(\theta_0) + \frac{[T\pi]}{T}R(\theta^*)T^{1/2}(\tilde\theta_K - \theta_0) \\ &\Longrightarrow -\pi R(\theta_0)[R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0)]^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1). \end{aligned} \tag{42}$$

So by (41)(42)

$$T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde\theta_K(\pi)) \Longrightarrow \Omega_{\theta_0\theta_0}^{1/2}W_{GK}(\pi) - \pi R(\theta_0)[R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0)]^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1). \tag{43}$$

In the same way

$$\begin{aligned} T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde\theta_K(\pi)) &\Longrightarrow \Omega_{\theta_0\theta_0}^{1/2}(W_{GK}(1) - W_{GK}(\pi)) \\ &- (1-\pi)R(\theta_0)[R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0)]^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1). \end{aligned} \tag{44}$$

Now benefit from equations (71)-(73) such that

$$K_T(\tilde\theta_K(\pi),\pi) = K_T^1(\tilde\theta_K(\pi),\pi) + K_T^2(\tilde\theta_K(\pi),\pi),$$

where

$$\begin{aligned} K_T^1(\tilde\theta_K(\pi),\pi) &= \frac{1}{\pi}[\frac{1}{T^{1/2}}\sum_{t=1}^{[T\pi]}\psi_t(\tilde\theta_K(\pi))]'(\hat\Omega^1_{\tilde\theta_K(\pi)\tilde\theta_K(\pi)})^{-1/2} \\ &\times P_{(\hat\Omega^1_{\tilde\theta_K(\pi)\tilde\theta_K(\pi)})^{-1/2}\tilde D_T^1(\tilde\theta_K(\pi),,\pi)}(\hat\Omega^1_{\tilde\theta_K(\pi)\tilde\theta_K(\pi)})^{-1/2}[\frac{1}{T^{1/2}}\sum_{t=1}^{[T\pi]}\psi_t(\tilde\theta_K(\pi))], \end{aligned} \tag{45}$$

and

$$\begin{aligned} K_T^2(\tilde\theta_K(\pi),\pi) &= \frac{1}{1-\pi}[\frac{1}{T^{1/2}}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde\theta_K(\pi))]'(\hat\Omega^2_{\tilde\theta_K(\pi)\tilde\theta_K(\pi)})^{-1/2} \\ &\times P_{(\hat\Omega^2_{\tilde\theta_K(\pi)\tilde\theta_K(\pi)})^{-1/2}\tilde D_T^2(\tilde\theta_K(\pi),,\pi)}(\hat\Omega^2_{\tilde\theta_K(\pi)\tilde\theta_K(\pi)})^{-1/2}[\frac{1}{T^{1/2}}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde\theta_K(\pi))]. \end{aligned} \tag{46}$$

Now use (43)(44)(90) in (45)(46) to have

$$\begin{aligned} K_T^1(\tilde\theta_K(\pi),\pi) &\Longrightarrow \frac{1}{\pi}[\Omega_{\theta_0\theta_0}^{1/2}W_{GK}(\pi) - \pi R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1)]' \\ &\times \Omega_{\theta_0\theta_0}^{-1/2}P_{\Omega_{\theta_0\theta_0}^{-1/2}R(\theta_0)}\Omega_{\theta_0\theta_0}^{-1/2} \\ &\times [\Omega_{\theta_0\theta_0}^{1/2}W_{GK}(\pi) - \pi R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1)], \end{aligned} \tag{47}$$

and

$$\begin{aligned} K_T^2(\tilde{\theta}_K(\pi),\pi) &\Longrightarrow \frac{1}{1-\pi}[\Omega_{\theta_0\theta_0}^{1/2}(W_{GK}(1)-W_{GK}(\pi))-(1-\pi)R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1)]' \\ &\times \Omega_{\theta_0\theta_0}^{-1/2}P_{\Omega_{\theta_0\theta_0}^{-1/2}R(\theta_0)}\Omega_{\theta_0\theta_0}^{-1/2} \\ &\times [\Omega_{\theta_0\theta_0}^{1/2}(W_{GK}(1)-W_{GK}(\pi))-(1-\pi)R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1)]. \quad (48) \end{aligned}$$

Before analyzing the last two terms more carefully, we need the following for simplification from the proof of Theorem 3b, regarding the matrix A_C

$$A_C = [R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0)]^{-1/2}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}, \quad (49)$$

$$A_C W_{GK}(\pi) \equiv W_n(\pi), \quad (50)$$

$$A_C'A_C = \Omega_{\theta_0\theta_0}^{-1/2}R(\theta_0)[R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0)]^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}. \quad (51)$$

Now rewrite (47) as

$$\begin{aligned} K_T^1(\tilde{\theta}_K(\pi),\pi) &\Longrightarrow \frac{1}{\pi}[W_{GK}(\pi)'\Omega_{\theta_0\theta_0}^{-1/2}R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(\pi)] \\ &- \frac{1}{\pi}[2\pi W_{GK}(1)'\Omega_{\theta_0\theta_0}^{-1/2}R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(\pi)] \\ &+ \frac{1}{\pi}[\pi^2 W_{GK}(1)'\Omega_{\theta_0\theta_0}^{-1/2}R(\theta_0)(R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1}R(\theta_0))^{-1}R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2}W_{GK}(1)]. \quad (52) \end{aligned}$$

Then use (50)(51) in (47) to have

$$\begin{aligned} K_T^1(\tilde{\theta}_K(\pi),\pi) &\Longrightarrow \frac{1}{\pi}[W_{GK}(\pi)'A_C'A_CW_{GK}(\pi)-2\pi W_{GK}(\pi)'A_C'A_CW_{GK}(1)+\pi^2W_{GK}(1)'A_C'A_CW_{GK}(1)] \\ &\equiv \frac{1}{\pi}[W_n(\pi)'W_n(\pi)-2\pi W_n(\pi)'W_n(1)+\pi^2W_n(1)'W_n(1)]. \quad (53) \end{aligned}$$

In the same way we have for (48)

$$\begin{aligned} K_T^2(\tilde{\theta}_K(\pi),\pi) &\Longrightarrow \frac{1}{1-\pi}[(W_n(1)-W_n(\pi))'(W_n(1)-W_n(\pi)) \\ &- 2(1-\pi)(W_n(1)-W_n(\pi))'W_n(1)+(1-\pi)^2W_n(1)'W_n(1)]. \quad (54) \end{aligned}$$

Then multiply (53) by $(1-\pi)$ and (54) by π, after some simple algebra and adding these two limit terms we obtain the desired result. □

REFERENCES

Anderson, T. W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations, Annals of Mathematical Statistics 20, 46-63.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

Andrews, D. W. K., 1993a. Tests for parameter instability and structural change with unknown change point, Econometrica 61, 821-856.

Andrews, D. W. K., 1993b. An introduction to econometric applications of empirical process theory for dependent random variables, Econometric Reviews 12, 183-216.

Campbell, J.Y., Shiller, R.J., 1987. Cointegration and tests of present value models, Journal of Political Economy 95, 1062-1088.

Caner, M. 2003. Structural change tests in GMM with weak identification, Manuscript. University of Pittsburgh, Department of Economics.

Doukhan, P. 1994. Mixing: Properties and Examples. (New York, Springer).

Doukhan, P., Massart, P., Rio, E., 1995. Invariance principles for absolutely regular empirical processes, Annales de l'Institut Henri Poincare 31, 393-427.

Dufour, J. M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models, Econometrica 65, 1365-1387.

Dufour, J. M., Taamouti, M., 2001. Projection based statistical inference in linear structural models with possibly weak instruments, Manuscript, University of Montreal. Department of Economics.

Hansen, L. P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators, Journal of Business and Economic Statistics 14, 262-280.

Jennrich, R.I., 1969. Asymptotic properties of nonlinear least squares estimators, Annals of Mathematical Statistics 40, 633-643.

Kim, J., Pollard, D., 1990. Cube root asymptotics, The Annals of Statistics 18, 191-219.

Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified, Manuscript, Brown University.

Kocherlakota, N., 1990. On tests of representative consumer asset pricing models, Journal of Monetary Economics 26, 285-304.

Shiller, R.J., 1982. Consumption, asset markets and macroeconomic fluctuations, Carnegie Rochester Series on Public Policy 17, 203-238.

Sohrab, H.H. 2003. Basic Real Analysis (Boston, Birkhauser).

Stock, J. H., Wright, J., 2000. GMM with weak identification, Econometrica 68, 1055-1096.

Tauchen, G., 1986. Statistical properties of Generalized Method of Moments Estimators of structural parameters obtained from financial market data, Journal of Business and Economic Statistics 4, 397-425.

Van Der Vaart, A. W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. (New York,Springer).

Technical Appendix

We develop limit theory for restricted partial sample continuous updating estimator in standard identified framework, using empirical process theory. This estimator is defined in the proof of Theorem 1 , (17). We need the following :

Technical Lemma A.1. *Under Assumptions S1-S3,*

$$T^{1/2}(\tilde{\theta}(\pi)-\theta_0)=O_p(1).$$

Proof of Technical Lemma A.1. This proof is very similar to rate of convergence proof in Stock and Wright (2000). They deal with weakly identified case in full sample, whereas we show the standard case in partial samples. We first prove the consistency. Rewrite the objective function

$$\begin{aligned} S_T(\theta,\pi;\theta,\pi) &= [T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)]'\frac{V_{T1}(\theta,\pi)^{-1}}{\pi}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)] \\ &+ [T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\theta)]'\frac{V_{T2}(\theta,\pi)^{-1}}{1-\pi}[T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\theta)] \end{aligned}$$

Then add and subtract to have

$$T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)=\Psi_T(\theta,\pi)+ET^{-1/2}\sum_{t=1}^{[T\Pi]}\psi_t(\theta). \tag{55}$$

By Lemma 1

$$T^{-1/2}\Psi_T(\theta,\pi)\xrightarrow{p}0. \tag{56}$$

By Assumption S2, S3, and (55), (56)

$$\begin{aligned} T^{-1}S_T(\theta,\pi;\theta,\pi) &\xrightarrow{p} [\pi m_1(\theta)]'\frac{\Omega_{\theta,\theta}^{-1}}{\pi}[\pi m_1(\theta)] \\ &+ [(1-\pi)m_1(\theta)]'\frac{\Omega_{\theta,\theta}^{-1}}{1-\pi}[(1-\pi)m_1(\theta)], \end{aligned}$$

where $m_1(\theta)=0$ iff $\theta=\theta_0$, then we have a unique minimum and the limit is continuous, by Theorem 2.7 of Kim and Pollard (1991) or Andrews (1993a) we have

$$\tilde{\theta}(\pi)-\theta_0\xrightarrow{p}0.$$

The consistency result helps in deriving the rate of convergence.

For the rate of convergence part since $\tilde{\theta}(\pi)$ minimizes the objective function $S_T(\theta,\pi;\theta,\pi)$ by definition

$$\begin{aligned} S_T(\tilde{\theta}(\pi),\pi;\tilde{\theta}(\pi),\pi) &- S_T(\theta_0,\pi;\theta_0,\pi) \\ &= [\bar{\Psi}_T(\tilde{\theta}(\pi),\pi)+T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)]'W_T(\tilde{\theta}(\pi),\pi)[\bar{\Psi}_T(\tilde{\theta}(\pi),\pi)+T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)] \\ &- [\bar{\Psi}_T(\theta_0,\pi)+T^{1/2}\bar{m}_1(\theta_0,\pi)]'W_T(\theta_0,\pi)[\bar{\Psi}_T(\theta_0,\pi)+T^{1/2}\bar{m}_1(\theta_0,\pi)] \\ &\leq 0. \end{aligned} \tag{57}$$

where

$$W_T(\theta,\pi) = \begin{bmatrix} \frac{V_{T1}(\theta,\pi)^{-1}}{\pi} & 0 \\ 0 & \frac{V_{T2}(\theta,\pi)^{-1}}{1-\pi} \end{bmatrix}, \tag{58}$$

$$\bar{\Psi}_T(\theta,\pi) = \begin{bmatrix} T^{-1/2}\sum_{t=1}^{[T\pi]}(\psi_t(\theta) - E\psi_t(\theta)) \\ T^{-1/2}\sum_{t=[T\pi]+1}^{T}(\psi_t(\theta) - E\psi_t(\theta)) \end{bmatrix}. \tag{59}$$

and

$$\bar{m}_1(\theta,\pi) = \begin{bmatrix} \pi m_1(\theta) \\ (1-\pi)m_1(\theta) \end{bmatrix}. \tag{60}$$

Then rewrite (57) in the following way

$$\begin{aligned} T\bar{m}_1(\tilde{\theta}(\pi),\pi)'W_T(\tilde{\theta}(\pi),\pi)\bar{m}_1(\tilde{\theta}(\pi),\pi) \quad &+ \quad 2T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)W_T(\tilde{\theta}(\pi),\pi)\bar{\Psi}_T(\tilde{\theta}(\pi),\pi) \\ &+ \quad d_{1T}(\tilde{\theta}(\pi),\pi) \le 0, \end{aligned} \tag{61}$$

where

$$d_{1T}(\tilde{\theta}(\pi),\pi) = \bar{\Psi}_T(\tilde{\theta}(\pi),\pi)'W_T(\tilde{\theta}(\pi),\pi)\bar{\Psi}_T(\tilde{\theta}(\pi),\pi) - [\bar{\Psi}_T(\theta_0,\pi)'W_T(\theta_0,\pi)\bar{\Psi}_T(\theta_0,\pi)]. \tag{62}$$

Then since W_T is symmetric

$$T\bar{m}_1(\tilde{\theta}(\pi),\pi)'W_T(\tilde{\theta}(\pi),\pi)\bar{m}_1(\tilde{\theta}(\pi),\pi) \ge \|T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)\|^2 meval(W_T(\tilde{\theta}(\pi),\pi)), \tag{63}$$

where meval A denotes the minimum eigenvalue of the matrix. Next,

$$T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)'W_T(\tilde{\theta}(\pi),\pi)\bar{\Psi}_T(\tilde{\theta}(\pi),\pi) \ge -\|T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)\|W_T(\tilde{\theta}(\pi),\pi)\bar{\Psi}_T(\tilde{\theta}(\pi),\pi). \tag{64}$$

Use (64)(63) in (61) and divide (61) by meval $W_T(\tilde{\theta}(\pi),\pi)$ (this is positive with probability one by Assumption S3)

$$\|T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)\|^2 - 2d_{2T}\|T^{1/2}\bar{m}_1(\tilde{\theta}(\pi),\pi)\| + d_{3T} \le 0, \tag{65}$$

where

$$d_{2T} = \|W_T(\tilde{\theta}(\pi),\pi)\bar{\Psi}_T(\tilde{\theta}(\pi),\pi)\|/mevalW_T(\tilde{\theta}(\pi),\pi),$$

$$d_{3T} = d_{1T}(\tilde{\theta}(\pi),\pi)/meval(W_T(\tilde{\theta}(\pi),\pi).$$

By Assumption S2

$$T^{1/2}m_1(\tilde{\theta}(\pi),\pi) = R(\bar{\theta}(\pi),\pi)T^{1/2}(\tilde{\theta}(\pi) - \theta_0),$$

where $\bar{\theta}(\pi) \in (\theta_0, \tilde{\theta}(\pi))$. (See also Jennrich 1969, Lemma 3 for mean value theorem for random valued functions).

Taking the roots of (65) and for (65) to hold

$$\|\bar{R}_T(\tilde{\theta}(\pi), \pi)\| \leq d_{2T} + (d_{2T}^2 - d_{3T})^{1/2},$$

where

$$\bar{R}_T(\tilde{\theta}(\pi), \pi) = \begin{bmatrix} \pi R(\bar{\theta}(\pi)) T^{1/2}(\tilde{\theta}(\pi) - \theta_0) & 0 \\ 0 & (1-\pi) R(\bar{\theta}(\pi)) T^{1/2}(\tilde{\theta}(\pi) - \theta_0) \end{bmatrix}.$$

By Assumption S2 and consistency

$$R(\bar{\theta}(\pi)) \xrightarrow{p} R(\theta_0).$$

The desired rate of convergence follows since by Lemma 1 and Assumption S3, $d_{2T} = O_p(1), d_{3T} = O_p(1)$. □

Technical Lemma A.2. *Under Assumptions S1-S3*

$$T^{1/2}(\tilde{\theta}(\pi) - \theta_0) \Longrightarrow -[R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1} R(\theta_0)]^{-1}[R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1/2} W_{GK}(1),]$$

where $W_{GK}(1)$ is a GK dimensional standard normal vector.

Proof of Technical Lemma A.2.

First set up the objective function

$$S_T(\tilde{\theta}(\pi), \pi; \theta_0, \pi) = [\bar{\Psi}_T(\tilde{\theta}(\pi), \pi) + T^{1/2}\bar{m}_1(\tilde{\theta}(\pi), \pi)]' W_T(\theta_0, \pi) [\bar{\Psi}_T(\tilde{\theta}(\pi), \pi) + T^{1/2}\bar{m}_1(\tilde{\theta}(\pi), \pi)]'. \tag{66}$$

By Lemma 1, consistency and Assumption S1 with (59)

$$\bar{\Psi}_T(\tilde{\theta}(\pi), \pi) \Longrightarrow \begin{bmatrix} B(\pi) \\ B(1) - B(\pi) \end{bmatrix}, \tag{67}$$

where we also use $\Psi(\theta_0, \pi) \equiv B(\pi)$.

By (60), and Assumption S2, using the rate of convergence result

$$T^{1/2}\bar{m}_1(\tilde{\theta}(\pi), \pi) \rightarrow \begin{bmatrix} \pi R(\theta_0) u \\ (1-\pi) R(\theta_0) u \end{bmatrix}. \tag{68}$$

The limit is uniform in $u \in \Delta$ which is a compact subset of R^n. Using (67) (68) in (66) we obtain the limit

$$S_T(\tilde{\theta}(\pi), \pi; \theta_0, \pi) \Longrightarrow \bar{S}(u, \pi), \tag{69}$$

where

$$\bar{S}(u,\pi) = \begin{bmatrix} B_{GK}(\pi)+\pi R(\theta_0)u \\ B_{GK}(1)-B_{GK}(\pi)+(1-\pi)R(\theta_0)u \end{bmatrix}' \times \begin{bmatrix} \frac{\Omega^{-1}_{\theta_0,\theta_0}}{\pi} & 0 \\ 0 & \frac{\Omega^{-1}_{\theta_0,\theta_0}}{1-\pi} \end{bmatrix} \begin{bmatrix} B_{GK}(\pi)+\pi R(\theta_0)u \\ B_{GK}(1)-B_{GK}(\pi)+(1-\pi)R(\theta_0)u \end{bmatrix}. \tag{70}$$

Since the limit is unique, by Lemma 3.2.1 of van der Vaart and Wellner (1996) we have uniformly over $\Delta \times \Pi$

$$T^{1/2}(\tilde{\theta}(\pi)-\theta_0) \Longrightarrow u^*(\pi),$$

where

$$u^*(\pi) = \arg\min_u \bar{S}(u,\pi).$$

Then differentiate this limit with respect to u we find

$$u^*(\pi) = -[R(\theta_0)'\Omega^{-1}_{\theta_0,\theta_0}R(\theta_0)]^{-1}[R(\theta_0)'\Omega^{-1/2}_{\theta_0,\theta_0}W_{GK}(1)].$$

$\square$.

But since $u^*(\pi)$ does not depend on π on the equation above we can also denote that as u^*.

The following Lemmata provide the limit for restricted Kleibergen (2005) type of estimator defined in equation (4). First we provide consistency.

Technical Lemma A.3. *Under Assumptions S1-S2, S4-S6*

$$(\tilde{\theta}_K(\pi)-\theta_0) \xrightarrow{p} 0.$$

Proof of Technical Lemma A.3. We follow the standard consistency proof. First,

$$K_T(\theta,\pi) = K_T^1(\theta,\pi) + K_T^2(\theta,\pi), \tag{71}$$

where

$$K_T^1(\theta,\pi) = \frac{1}{\pi}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)]'(\hat{\Omega}^1_{\theta\theta})^{-1/2}P_{(\hat{\Omega}^1_{\theta\theta})^{-1/2}\tilde{D}^1_T(\theta,\pi)}(\hat{\Omega}^1_{\theta\theta})^{-1/2}[T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\theta)], \tag{72}$$

and

$$K_T^2(\theta,\pi) = \frac{1}{1-\pi}[T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta)]'(\hat{\Omega}^2_{\theta\theta})^{-1/2}P_{(\hat{\Omega}^2_{\theta\theta})^{-1/2}\tilde{D}^2_T(\theta,\pi)}(\hat{\Omega}^2_{\theta\theta})^{-1/2}[T^{-1/2}\sum_{[T\pi]+1}^{T}\psi_t(\theta)], \tag{73}$$

where $\tilde{D}^1_T(\theta,\pi), \tilde{D}^2_T(\theta,\pi)$ are defined in equations (6)-(8).

Following the proof of Lemma 1 for $q_{it}(\theta)$ instead of $\psi_t(\theta)$, Assumptions S1,S4 provide the following

$$T^{-1/2}\sum_{t=1}^{[T\pi]} q_{it}(\theta) - Eq_{it}(\theta) = O_p(1). \tag{74}$$

Use (74) and Assumption S5 to have

$$T^{-1}\sum_{t=1}^{[T\pi]} q_{it}(\theta) \xrightarrow{p} \pi q_i(\theta). \quad (75)$$

So by Assumption S6, (75) in combination with Assumption S2 and (56), uniformly over $\theta \times \pi$

$$T^{-1/2}\tilde{D}_T^1(\theta) \xrightarrow{p} \pi D(\theta), \quad (76)$$

where

$$D(\theta) = [q_1(\theta) - \Omega_{q\theta,1}(\Omega_{\theta\theta})^{-1}m_1(\theta), \cdots, q_n(\theta) - \Omega_{q\theta,n}(\Omega_{\theta\theta})^{-1}m_1(\theta)]. \quad (77)$$

Same analysis shows that uniformly over $\theta \times \pi$

$$T^{-1/2}\tilde{D}_T^2(\theta) \xrightarrow{p} (1-\pi)D(\theta). \quad (78)$$

By (55),(56) Assumption S2, and (75)-(78), uniformly over $\theta \times \pi$

$$\begin{aligned} T^{-1}K_T(\theta,\pi) &\xrightarrow{p} \frac{1}{\pi}[\pi m_1(\theta)]'(\Omega_{\theta,\theta})^{-1/2}P_{(\Omega_{\theta,\theta})^{-1/2}D(\theta)}(\Omega_{\theta,\theta})^{-1/2}[\pi m_1(\theta)] \\ &+ \frac{1}{1-\pi}[1-\pi m_1(\theta)]'(\Omega_{\theta,\theta})^{-1/2}P_{(\Omega_{\theta,\theta})^{-1/2}D(\theta)}(\Omega_{\theta,\theta})^{-1/2}[1-\pi m_1(\theta)] \\ &= m_1(\theta)'(\Omega_{\theta,\theta})^{-1/2}P_{(\Omega_{\theta,\theta})^{-1/2}D(\theta)}(\Omega_{\theta,\theta})^{-1/2}m_1(\theta). \end{aligned}$$

Then we know that $m_1(\theta) = 0$ if $\theta = \theta_0$, and by Assumption S6ii ($P_{\Omega_{\theta\theta}^{-1/2}D(\theta)}\Omega_{\theta\theta}^{-1/2}m_1(\theta) \neq 0, \theta \neq \theta_0$) unique minimum exists and the limit is continuous by Assumption S1, S2, S4 so by Theorem 2.7 of Kim and Pollard (1990) we have

$$\tilde{\theta}_K(\pi) - \theta_0 \xrightarrow{p} 0.$$

□

Technical Lemma A.4. *Under Assumptions S1,S2,S4-S6,*

$$T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0) \xrightarrow{d} N(0, (R(\theta_0)'\Omega_{\theta_0,\theta_0}^{-1}R(\theta_0))^{-1}).$$

Proof of Technical Lemma A.4.

First take the partial derivative of $K_T(\theta,\pi)$ with respect to θ and divide the partial derivative equation's each side by $2T^{1/2}$, and note that $q_t(\theta) = \partial\psi_t(\theta)/\partial\theta$,

$$
\begin{aligned}
\frac{1}{\pi}\left[T^{-1}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi))\right]' \quad & \times \quad W_{T1}(\tilde{\theta}_K(\pi))\left[\frac{\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi))}{T^{1/2}}\right] \\
& + \quad \frac{1}{1-\pi}\left[T^{-1}\sum_{t=[T\pi]+1}^{T} q_t(\tilde{\theta}_K(\pi))\right]' W_{T2}(\tilde{\theta}_K(\pi))\left[\frac{\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde{\theta}_K(\pi))}{T^{1/2}}\right] \\
& + \quad \frac{1}{2\pi}\left[\frac{\partial vec(W_{T1}(\tilde{\theta}_K(\pi)))}{\partial\theta'}\right]'\left[T^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi))\bigotimes T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi))\right] \\
& + \quad \frac{1}{2(1-\pi)}\left[\frac{\partial vec(W_{T2}(\tilde{\theta}_K(\pi)))}{\partial\theta'}\right]' \\
& \times \quad \left[T^{-1}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde{\theta}_K(\pi))\bigotimes T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde{\theta}_K(\pi))\right] \\
& = \quad 0,
\end{aligned}
\tag{79}
$$

where

$$
W_{Tj}(\tilde{\theta}_K(\pi)) = (\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)})^{-1/2} P_{(\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)})^{-1/2}\tilde{D}^j_T(\tilde{\theta}_K(\pi))} (\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)})^{-1/2},
$$

for $j = 1, 2$.

By Assumptions S1,S6 and consistency of $\tilde{\theta}_K(\pi)$

$$
\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)} \xrightarrow{p} \Omega_{\theta_0,\theta_0}. \tag{80}
$$

Note that by (76), and consistency of restricted K-estimator

$$
T^{-1/2}\tilde{D}^1_T(\tilde{\theta}_K(\pi), \pi) \xrightarrow{p} \pi D(\theta_0), \tag{81}
$$

where

$$
D(\theta_0) = [q_1(\theta_0), \cdots, q_n(\theta_0)],
$$

since $m_1(\theta_0) = 0$ by Assumption S2, and also note that by the definition of $R(\theta_0)$ in Assumption S2, in this standard identified case $D(\theta_0) = R(\theta_0)$. Similarly

$$
T^{-1/2}\hat{D}^2_T(\tilde{\theta}_K(\pi), \pi) \xrightarrow{p} (1-\pi)D(\theta_0) = (1-\pi)R(\theta_0). \tag{82}
$$

Then using continuous differentiability of $q_{it}(\theta)$ in Assumption S4, by Proposition 6.6.1 (Leibniz Rule) in Sohrab (2003), and consistency of $\tilde{\theta}_K(\pi)$ in combination with (80)-(82) shows

$$
\frac{\partial vec(W_{Tj}(\tilde{\theta}_K(\pi)))}{\partial\theta'} = O_p(1), \tag{83}
$$

for $j = 1, 2$.

Since

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) &= [T^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) - E\psi_t(\tilde{\theta}_K(\pi))] \\
&+ T^{-1}\sum_{t=1}^{[T\pi]}E\psi_t(\tilde{\theta}_K(\pi))
\end{aligned}
$$

In the above equation the first term in square brackets converges to zero in probability by Lemma 1 (which is obtainable by Assumption S1) and the second term goes to zero by Assumption S2, and consistency of the estimator. So

$$
T^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) \xrightarrow{p} 0. \tag{84}
$$

Then

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) &= [T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) - E\psi_t(\tilde{\theta}_K(\pi))] \\
&+ T^{-1/2}\sum_{t=1}^{[T\pi]}E\psi_t(\tilde{\theta}_K(\pi)).
\end{aligned} \tag{85}
$$

The first term on the right hand side of (85) weakly converges to a process by Assumptions S1 and S2, then having a Taylor series expansion around θ_0 for the second term

$$
ET^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) = \pi R(\theta_0)T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0) + o(1), \tag{86}
$$

by Assumption S2. So clearly by (86), consistency of the estimator, and (85) is

$$
T^{-1/2}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi)) = O_p(T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0)). \tag{87}
$$

Same results apply to the second part of the partial sample so we can say that the third and fourth terms on the partial derivative (79) are

$$
\begin{aligned}
\frac{1}{2\pi}\left[\frac{\partial vec(W_{T1}(\tilde{\theta}_K(\pi)))}{\partial\theta'}\right]' &\times \left[T^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi))\bigotimes T^{-1}\sum_{t=1}^{[T\pi]}\psi_t(\tilde{\theta}_K(\pi))\right] \\
&+ \frac{1}{2(1-\pi)}\left[\frac{\partial vec(W_{T2}(\tilde{\theta}_K(\pi)))}{\partial\theta'}\right]' \\
&\times \left[T^{-1}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde{\theta}_K(\pi))\bigotimes T^{-1/2}\sum_{t=[T\pi]+1}^{T}\psi_t(\tilde{\theta}_K(\pi))\right] \\
&= o_p(T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0)),
\end{aligned} \tag{88}
$$

by (83)(84)(87).

Now note that by (81)(82), and Assumption S6 with the consistency of the estimator

$$(\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)})^{-1/2} P_{(\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)})^{-1/2}\tilde{D}^j_T(\tilde{\theta}_K(\pi))} (\hat{\Omega}^j_{\tilde{\theta}_K(\pi)\tilde{\theta}_K(\pi)})^{-1/2} \xrightarrow{p} \Omega^{-1/2}_{\theta_0\theta_0} P_{\Omega^{-1/2}_{\theta_0\theta_0} R(\theta_0)} \Omega^{-1/2}_{\theta_0\theta_0}, \tag{89}$$

for $j = 1, 2$.

Then we can replace $W_{Tj}(\tilde{\theta}_K(\pi))$ in (79) by its limit

$$W(\theta_0) = \Omega^{-1/2}_{\theta_0\theta_0} P_{\Omega^{-1/2}_{\theta_0\theta_0} R(\theta_0)} \Omega^{-1/2}_{\theta_0\theta_0}. \tag{90}$$

Next,

$$T^{-1/2}\sum_{t=1}^{[T\pi]} \psi_t(\tilde{\theta}_K(\pi)) = T^{-1/2}\sum_{t=1}^{[T\pi]} \psi_t(\theta_0) + (T^{-1}\sum_{t=1}^{[T\pi]} q_t(\theta^*))T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0), \tag{91}$$

where $\theta^* \in (\theta_0, \tilde{\theta}_K(\pi))$.

We can rewrite the partial derivative equation in (79) in the following manner

$$\begin{aligned}
\frac{1}{\pi}[\frac{1}{T}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi))]'W(\theta_0)[\frac{1}{T^{1/2}}\sum_{t=1}^{[T\pi]} \psi_t(\theta_0)] \quad &+ \quad \frac{1}{\pi}[\frac{1}{T}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi))]'W(\theta_0)[\frac{1}{T}\sum_{t=1}^{[T\pi]} q_t(\theta^*)]T^{1/2}(\tilde{\theta}_K - \theta_0) \\
&+ \quad \frac{1}{1-\pi}[\frac{1}{T}\sum_{t=[T\pi]+1}^{T} q_t(\tilde{\theta}_K(\pi))]'W(\theta_0)[\frac{1}{T^{1/2}}\sum_{t=[T\pi]+1}^{T} \psi_t(\theta_0)] \\
&+ \quad \frac{1}{1-\pi}[\frac{1}{T}\sum_{t=[T\pi]+1}^{T} q_t(\tilde{\theta}_K(\pi))]'W(\theta_0) \\
&\times \quad [\frac{1}{T}\sum_{t=[T\pi]+1}^{T} q_t(\theta^*)]T^{1/2}(\tilde{\theta}_K - \theta_0) \\
&+ \quad o_p(T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0)) = 0.
\end{aligned} \tag{92}$$

Then consider

$$T^{-1}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi)) = [T^{-1}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi)) - Eq_t(\tilde{\theta}_K(\pi))] + [T^{-1}\sum_{t=1}^{[T\pi]} Eq_t(\tilde{\theta}_K(\pi))]. \tag{93}$$

Then by Assumption S4 (like Lemma 1), we can show that

$$T^{-1/2}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi)) - Eq_t(\tilde{\theta}_K(\pi)) = O_p(1).$$

So the first term on the right hand side of (93) converges in probability to zero. Then by Assumption S5, consistency and definition of $R(\theta_0)$ in Assumption S2 we have uniformly over $\theta \times \pi$

$$T^{-1}\sum_{t=1}^{[T\pi]} q_t(\tilde{\theta}_K(\pi)) \xrightarrow{p} \pi R(\theta_0). \tag{94}$$

Then by Assumption S1 and $E\psi_t(\theta_0) = 0$, following the proof of Lemma 1

$$T^{-1/2} \sum_{t=1}^{[T\pi]} \psi_t(\theta_0) \Longrightarrow \Omega_{\theta_0\theta_0}^{1/2} W_{GK}(\pi), \tag{95}$$

and

$$T^{-1/2} \sum_{t=[T\pi]+1}^{T} \psi_t(\theta_0) \Longrightarrow \Omega_{\theta_0\theta_0}^{1/2} [W_{GK}(1) - W_{GK}(\pi)]. \tag{96}$$

Use (90)(94)(95)(96) in (92) and simplify to have

$$T^{1/2}(\tilde{\theta}_K(\pi) - \theta_0) \xrightarrow{d} -[R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1} R(\theta_0)]^{-1} R(\theta_0)'\Omega_{\theta_0\theta_0}^{-1/2} W_{GK}(1).$$

□.