Efficient Inference with Time-Varying Identification Strength

PRELIMINARY

- NOT FOR CIRCULATION -

Bertille Antoine*
Simon Fraser University

Otilia Boldea†
Tilburg University

Abstract

In this paper, we develop inference procedures to detect both parameter instability and changes in the strength of identification, or strength of instrumental variables. If parameter instability is not linked to changes in instrument strength, we show the validity and efficiency of standard GMM inference methods, conditional on the break-points. On the other hand, if parameter instability is linked to changes in instrument strength, we develop methods to estimate the location and magnitude of such changes. We derive the asymptotic distribution of 2SLS and GMM structural parameter estimates in different subsamples, and show which inference methods are valid and efficient. We also provide a step-by-step guide for practitioners on how to detect and locate these changes.

We illustrate our methods via the New Keynesian Phillips Curve for the US. We find that the instrument strength changes twice between 1968 and 2005. We also show that our confidence intervals are more informative about the parameters of interest than the usual weak-instrument-robust confidence intervals. This allows us to confirm

* *Email: Bertille_Antoine@sfu.ca*
† *Email: O.Boldea@uvt.nl*

that agents are forward-looking, and to recover a positive trade-off between output gap and inflation.

Keywords: Weak identification; IV Estimation; Structural change; Multiple breakpoints.

JEL classification: C12, C13.

1 Introduction

The purpose of this paper is to develop reliable inference procedures to detect parameter instability as well as changes in the strength of identification. Both the strength of identification and the instability of parameters over time are widespread concerns in applied economics. Moreover, parameter instability in a structural model may induce changes in the strength of identification. This issue, often neglected by researchers, may lead to incorrect, or inefficient inference. Therefore it seems valuable for empirical researchers to detect such changes in order to provide reliable estimation in each subsample. To our knowledge, this is the first paper to consider these two issues in a unified framework, and provide a comprehensive treatment of the link between them. We focus on the empirically relevant linear regression model with endogenous variables.

Identification issues arise when the quality of statistical information about some, or all, parameters of an economic model is poor, such as when the correlation between endogenous variables and instrumental variables is small. This situation is broadly referred to as "weak identification". Over the last 30 years, many such cases have been reported in the empirical literature, especially in macroeconometrics: for instance the consumption-based capital asset pricing model (see Hansen and Singleton (1982), Stock and Wright (2000)). Staiger and Stock (1997) consider weak identification in a linear IV regression (see also Stock and Wright (2000) for a nonlinear generalization) and develop an alternative asymptotic framework where the moment conditions tend to zero at rate square-root of the sample size T around the true value of the parameters. In this weak instruments asymptotic framework, parameters cannot be consistently estimated and have nonstandard asymptotic distributions. The primary goal of the weak instruments asymptotics is to devise inference procedures robust to weak identification in the "worst case scenario", that is when statistical information disappears as fast as it accumulates at rate $\sqrt{T}$. More recently, Hansen, Hausman and Newey's (2008) survey of the applied literature suggests that such modeling techniques may not always be the most suitable. Hence, many authors consider instead a rate of decay to zero slower than $\sqrt{T}$. In this near-weak instruments asymptotic framework, one recovers consistency and asymptotic normality in estimation, though at a slower than parametric rate. Hahn and Kursteiner (2002) study IV estimators in a linear model, while Caner (2010) and Antoine and Renault (2009, 2012a) consider the nonlinear case. In addition, Antoine and Renault (2012b) develop a statistical test to assess identification strength in order to decide whether standard asymptotic theory based on asymptotic normality is reliable. More specifically, they propose a test to detect near-weak identification when the null hypothesis assumes weak identification.

This paper considers a framework where the exact identification strength is unknown and allowed to change over time.

All the aforementioned papers rely on the assumption, quite common in the literature, that the economic relationships of interest are stable over time. However, with samples covering extended periods, this assumption is often questionable. It can actually be argued that policy changes and/or exogenous shifts may cause realignments in the relationship between economic variables. As a result, statistical methods have been developed to detect structural instability: see inter alia Andrews and Fair (1988), Ghysels and Hall (1990a,b), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Bai and Perron (1998), Hall and Sen (1999), Bai and Perron (2003), Qu and Perron (2007). It is common practice to consider that the structural parameters of a model are subject to discrete shifts at unknown points in the sample. The main concern is then to estimate economic relationships in the different subsamples. An important precursor to this analysis is the identification of the points in the sample at which the parameters change: so-called break-points. For a linear regression model estimated by Ordinary Least Squares, Bai and Perron (1998) estimate simultaneously break-points and regression parameters. They establish the consistency and the limiting distribution of the resulting break-point fractions and propose a sequential procedure for selecting the number of break-points in the sample. Hall, Han, and Boldea (2012) extend this framework to allow for endogenous regressors. They show that estimation based on a GMM criterion yields inconsistent estimators of the break fractions (defined as the ratio of the break-point and the sample size), whereas consistent estimators are obtained when minimizing a Two Stage Least Squares (2SLS) criterion. They also develop a limiting distribution theory for estimated regression parameters conditional on the break-points, and a procedure to estimate the number of break-points. The asymptotic distribution theory for estimated break fractions is derived in Boldea, Hall and Han (2012). Despite empirical evidence suggesting that identification issues are widespread in applied economics, all the above methods assume strong identification within the subsamples which may lead to inaccurate, or incorrect inference.

The proposed paper combines the above two streams of literature to study how parameter instability may affect the strength of identification. We develop break-point methods to detect parameter instability and changes in the identification strength in order to provide valid inference in the resulting subsamples. More specifically, we extend the linear regression model with endogenous variables to allow structural parameters and identification strength to change at (unknown) points in the sample. In such a framework, we develop reliable methods to: i) estimate the number of break-points, detect and locate them; ii) estimate the structural parameters, and develop their asymptotic limiting distributions and associated inference procedures. To our knowledge, this is the first paper that considers both issues in a unified framework, and provide a comprehensive treatment of the link between them.

We apply the inference procedures developed in this paper to the New Keynesian Phillips Curve (NKPC) using quarterly data for the US over the period 1968:3 to 2005:4. The NKPC is of considerable theoretical importance in monetary policy analysis as it is used to model inflation, identify the mix of forward- and backward-looking components of inflation, as well as the trade-off between inflation and unemployment over the cycle. There is considerable evidence suggesting that macroeconomic models such as the NKPC are subject to parameter instability and to identification issues (see for instance Hall, Han and Boldea (2012), Canova and Sala (2009), Kleibergen and Mavroeidis (2009) and references therein). Our analysis indicates that there are shifts in the structural parameters of the NKPC itself, as well as changes in the identification strength of the reduced form. We show that not all subsamples are weakly identified, and identify the subsamples over which standard inference for NKPC can still be performed. As a result, our confidence intervals about the parameters of interest are more informative than the weak-instrument-robust confidence intervals. This allows us to confirm that agents are forward-looking and to recover a positive trade-off between output gap and inflation (at least over some subsample).

This paper is organized as follows. Section 2 reviews various identification frameworks for the stable linear IV regression model. Section 3 provides asymptotic results for efficient estimation and inference in the presence of breaks either in structural parameters, or in the identification strength. Section 4 presents our general framework and inference procedure, along with a comprehensive step-by-step guide for practitioners. In section 5, we apply our results to the estimation of the New Keynesian Phillips Curve. Section 6 concludes. Regularity assumptions and proofs of our theoretical results are discussed in the Appendix.

2 Identification in the stable linear IV regression model

In this section, we present an overview of the identification settings and associated asymptotic results commonly used in the stable linear IV model.

In many econometric models, the parameter value θ^0 of interest is identified through moment restrictions of the form

$$\mathrm{E}\left[g(X_t, \theta^0)\right] = \mathbf{0}\,, \tag{2.1}$$

where $g(.)$ is a known function of the random vector of observations $X_t \in \mathbb{R}^d$ and of the structural parameter of interest $\theta \in \Theta \subset \mathbb{R}^p$. Such restrictions are generally deduced from an underlying economic model. For example, Euler equations often provide relevant restrictions: see Hansen, Heaton and Yaron (1996) for an application of GMM estimation to the consumption based capital asset pricing model. We are specifically interested in the standard linear regression model with endogenous variables

$$\mathrm{E}\left[W_t\left(y_t - Y_t'\theta_y^0 - Z_t'\theta_z^0\right)\right] = \mathbf{0}\,, \tag{2.2}$$

with y_t the dependent variable, Y_t the vector of p_1 (endogenous) variables, Z_t the vector of p_2 exogenous variables, W_t the vector of q instrumental variables, $X_t = (y_t,\ Y_t',\ Z_t',\ W_t')'$, $\theta^{0\prime} = (\theta_y^{0\prime},\ \theta_z^{0\prime})$, and $p = p_1 + p_2$.

In such a setting, weak identification is often modeled by assuming that these unconditional moments flatten around θ^0 as the sample size T increases. Typically, Antoine and Renault

(2009), in the line of Staiger and Stock (1997), assume that, for any k between 1 and q,

$$E\left[W_{k,t}\left(y_t - Y_t'\theta_y - Z_t'\theta_z\right)\right] = \frac{m_k(\theta)}{r_{k,T}}\,, \tag{2.3}$$

where $m_k(.)$ is a constant function, $r_{k,T}$ is a deterministic real sequence such that $r_{k,T} = 1$ or $r_{k,T} \stackrel{T}{\to} \infty$. The faster the unknown sequence $r_{k,T}$ diverges to infinity, the weaker the associated instrumental variable (IV), or moment condition is. We can actually distinguish three cases of interest.

(i) When $r_{k,T} = 1$, the IV is strong. This is the standard case. When all the moment conditions are strong, standard estimation of the structural parameters at rate $\sqrt{T}$ and standard testing techniques are (asymptotically) valid.

(ii) When $r_{k,T} \stackrel{T}{\to} \infty$ and $r_{k,T} = o(\sqrt{T})$, the IV is near-weak. When all the moment conditions are near-weak at the same rate r_T, standard estimation of the structural parameters and standard testing techniques are still asymptotically valid, but at the slower rate $\sqrt{T}/r_T$. When moment conditions are associated with different near-weak rates, the structural parameters are usually identified at the slowest available rate, $[\sqrt{T}/\max_k(r_{k,T})]$. The interested reader is referred to sections 2.1 and 4.1 in Antoine and Renault (2010) for a thorough discussion of such cases.

(iii) When $r_{k,T} = \sqrt{T}$, the IV is weak. Consistent estimation of the structural parameters is not possible anymore and one must rely on so-called "identification-robust" inference techniques. See e.g. the surveys by Stock, Wright, and Yogo (2002), and Dufour (2003) and references therein.

As already mentioned in the introduction, weak identification may not always be best-suited in practice (see Hansen, Hausman and Newey (2008)). In addition, near-weak identification can be disentangled from weak identification through the test proposed by Antoine and Renault (2012b). As a result, this paper considers a framework where the exact identification pattern is unknown. In our application, we use the test of Antoine and Renault (2012b) to check whether near-weak identification is a sensible assumption[1]. In addition, we also allow the identification strength as well as the structural parameter value θ^0 to change over time. Indeed, with samples covering extended periods, economic relationships are unlikely to

[1] We also use the tests proposed by Stock and Yogo (2005) that are based on the bias and the size distortion of 2SLS.

remain stable over time. In the next section, we extend the model defined by (2.2) and (2.3) to allow parameter instability and changes in the identification strength. In order to build a comprehensive econometric procedure, we first look at two simplified models of interest before introducing our general model unstable structural and reduced form equations.

3 Two simplified frameworks of interest

In economics, we usually think about a reduced form as implicitly derived from a structural system, say

$$[y\ Y]\begin{bmatrix} 1 & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} = ZB + U\,.$$

Provided Γ is invertible, this implies that, with obvious notations,

$$\begin{aligned} y &= -Y\Gamma_{21} + ZB_1 + U_1 \\ Y &= [ZB\Gamma^{-1}]_{22} + [U\Gamma^{-1}]_{22} = Z\Pi + V\,. \end{aligned}$$

Thus, whenever either Γ_{12}, Γ_{22}, or B_2 changes, so do Π and $\mathtt{Var}(V)$, but not the structural equation. In light of this, we propose in section 3.1 a framework where reduced-form parameters can change while structural ones do not. Intuitively, this could also happen if parameters of one or more other structural equations changed. We also propose in section 3.2 a framework where structural parameters can change while reduced-form ones do not. In the above framework, this happens whenever $\Gamma_{12} = 0$. In such a case, provided Γ_{22} is invertible, the reduced form equation writes

$$Y = ZB_2\Gamma_{22}^{-1} + U_2\Gamma_{22}^{-1} = Z\Pi + V\,.$$

and breaks in the structural equation (whever Γ_{21} or B_1 changes) do not transmit to the reduced form. Finally, whenever $\Gamma_{12} \neq 0$, any break in Γ_{21} or B_1 will appear in both the structural and the reduced form equations. This case will be discussed in section 4 when we introduce our general framework. Additional examples may be found in Boldea and Hall (2011).

3.1 Unstable identification strength

Our first framework of interest extends the standard linear IV regression model to allow instability in the reduced form over time, while the structural parameters remain stable. More specifically, the (stable) structural equation still writes

$$y_t = Y_t'\theta_y^0 + Z_t'\theta_z^0 + u_t \quad , \quad E(u_t|\mathcal{F}_{t-1}) = 0\,. \tag{3.1}$$

For a given vector of q valid instruments W_t with $q \geq p_1$, the reduced form with one break-point now writes

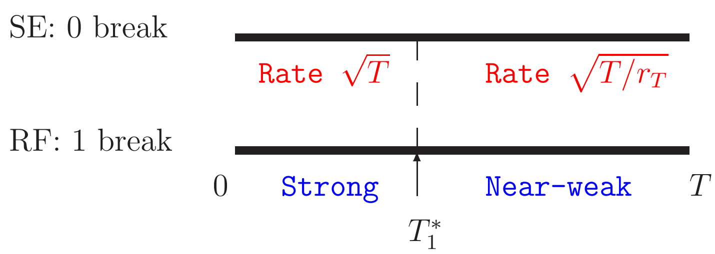
$$Y_{1t}' = \begin{cases} \frac{W_t'\Pi_1}{r_{1T}} + v_t' & , \ t \leq \lfloor T\nu_1^0 \rfloor \\ \frac{W_t'\Pi_2}{r_{2T}} + v_t' & , \ t > \lfloor T\nu_1^0 \rfloor \end{cases} , \tag{3.2}$$

where ν_1^0 is the break fraction[2], $r_{iT} = 1$, or $r_{iT} \to \infty$, $r_{iT} = o(\sqrt{T})$, Π_i is a full-rank matrix of size (q, p_1) for $i = 1, 2$, and W_t is not correlated with v_t.

The above reduced form (3.2) highlights the fact that the break-point $T_1^* \equiv \lfloor T\nu_1^0 \rfloor$ may capture two kinds of changes in the associated parameters. It is important to distinguish cases where the identification strength remains constant, that is[3] $r_{1T} \propto r_{2T}$ and $\Pi_1 \neq \Pi_2$, from cases where the identification strength changes, that is $r_{iT} = o(r_{jT})$, $i \neq j$. We are especially interested in identifying cases where the identification strength changes. Knowing about changes in the identification strength can lead to improved inference about the structural parameters, in the sense that faster convergence rates can be recovered. To illustrate this claim, consider the following example where the structural equation is stable over time, while the reduced form has one break. In the first subsample, the identification is strong, while in the second subsample the identification is near-weak at some rate r_T.

[2]For any break-fraction $0 < \nu_1^0 < 1$, the associated break point is $T_1^* \equiv \lfloor T\nu_1^0 \rfloor$ where $\lfloor w \rfloor$ represents the integer part of w.

[3]$r_{1T} \propto r_{2T} \Leftrightarrow r_{1T}/r_{2T} \stackrel{T}{\to} c$ with c a real number such that $0 < |c| < \infty$.



Conditional on knowing the break-point T_1^*, and given the results of Theorem 3.4, the (same) structural parameters can be estimated at rate $\sqrt{T}$ in the first subsample, but only at rate $\sqrt{T}/r_T$ in the second subsample. Of course, these results are only asymptotic, but we believe that in practice there are cases where such information can be used to draw sharper inference (or tighter confidence regions) on structural parameters of interest. It turns out that a similar result holds when the weakest subsample is actually weak with $r_T = \sqrt{T}$. We will show that consistent estimation of the structural parameters is possible through standard inference procedures and that (conservative) weak-identification robust procedures are not necessary.

For any given (candidate) break-point T_1, let $\hat{\Pi}_{1T}(T_1)$ and $\hat{\Pi}_{2T}(T_1)$ denote the OLS estimators in (4.2) over each associated subsample. The break-point estimator $\hat{T}_1^*$ of T_1^* is defined as the minimizer of the following OLS criterion,

$$\hat{T}_1^* = \arg\min_{T_1} \left[S_T(T_1, \hat{\Pi}_{1T}(T_1), \hat{\Pi}_{2T}(T_1)) \right]$$

with

$$\begin{aligned} S_T(T_1, \hat{\Pi}_{1T}(T_1), \hat{\Pi}_{2T}(T_1)) &= \frac{1}{T} \sum_{t=1}^{T_1} \Big(Y_t - W_t' \hat{\Pi}_{1T}(T_1) \Big)' \Big(Y_t - W_t' \hat{\Pi}_{1T}(T_1) \Big) \\ &+ \frac{1}{T} \sum_{t=T_1+1}^{T} \Big(Y_t - W_t' \hat{\Pi}_{2T}(T_1) \Big)' \Big(Y_t - W_t' \hat{\Pi}_{2T}(T_1) \Big) \end{aligned}$$

In order to discuss the consistency of the estimated break fraction, we formally consider the following two cases of interest:

- Case (a): no change in the identification strength,

$$r_{iT} \propto r_{jT}\,, \quad r_{iT} = o(\sqrt{T}) \quad \text{and} \quad \Pi_{1T} \neq \Pi_{2T} \quad \text{for } i \neq j\,.$$

In such a case, the identification strength is assumed to be at least near-weak.

- Case (b): stronger identification strength over subsample i,

$$r_{iT} = o(r_{jT})\,, \quad r_{iT} = o(\sqrt{T}) \quad \text{for } i \neq j\,.$$

In such a case, the identification strength of the strongest subsample has to be at leat near-weak, but the other subsample can be weakly identified. In other words, nothing prevents subsample j to be such that $r_{jT} = \sqrt{T}$.

Theorem 3.1. *(Consistency of estimated break fraction in RF)*
Under regularity assumptions 1 to 4(ii), under the assumptions of case (a) or (b), the estimated break fraction $\hat{\nu}_1 \equiv \hat{T}_1^*/T$ is a consistent estimator of ν_1^0 such that

$$\|\hat{\nu}_1 - \nu_1^0\| = \mathcal{O}_P(r_{iT}^2/T)\,.$$

The above theorem shows that the estimator of the break fraction is consistent at the rate inherited from the strongest subsample. This result even holds when the weakest subsample is actually weakly identified, $r_{jT} = \sqrt{T}$. Intuitively, only the magnitude of the break, or the difference between parameters in the subsamples before and after the break matter.
It is interesting to point out that efficient estimation of the structural parameters is provided by GMM without estimating the break-point of the reduced form. Such an estimator is defined as

$$\hat{\theta}_{GMM} \;=\; \arg\min_{\theta} \left[\overline{g}_T'(\theta) S_T^{-1} \overline{g}_T(\theta)\right] \quad \text{with} \quad \overline{g}_T(\theta) = \frac{1}{T} W'(y - Y\theta_y - Z\theta_z)\,,$$

and S_T a consistent estimator of $\texttt{AVar}\left(\overline{g}_T(\theta^0)\right)$.

Theorem 3.2. *(GMM estimation of the structural parameters)*
Under regularity assumptions 1 to 5, under the assumptions of case (a) or (b), the GMM estimators of the structural parameters are consistent, asymptotically normally distributed, and efficient at the fastest available rate, $\sqrt{T}/r_{iT}$.

In particular, this means that (consistent) estimation of the break-point in the reduced form is not required for GMM to provide consistent (and efficient) estimation of the (stable) structural parameters. In addition, the above result holds even when the weakest subsample is only weakly identified, $r_{jT} = \sqrt{T}$. However, it is essential to detect breaks in the reduced form equation and associated changes in the identification strength in order to achieve this result. In other words, testing for weak identification[4] must be done over each subsample which requires consistent estimation of the break fraction.

3.2 Unstable structural parameters

Our second framework of interest extends the standard linear IV regression model to allow instabilities in the structural parameters over time, while the identification strength remains stable. More specifically, with one break-point[5], the structural equation is

$$y_t = \begin{cases} Y_t'\theta^0_{y,1} + Z_t'\theta^0_{z,1} + u_t & , \quad t \le \lfloor T\lambda^0_1 \rfloor \\ Y_t'\theta^0_{y,2} + Z_t'\theta^0_{z,2} + u_t & , \quad t > \lfloor T\lambda^0_1 \rfloor \end{cases} \quad , \quad E(u_t|\mathcal{F}_{t-1}) = 0\,, \tag{3.3}$$

where λ^0_1 is the break fraction. We define $\theta^0_i = (\theta^{0\prime}_{y,i}\ \theta^{0\prime}_{z,i})'$ for $i = 1, 2$. For a given vector of q valid instruments W_t with $q \ge p_1$, the stable near-weakly identified reduced form is

$$Y_t' = \frac{W_t'\Pi}{r_T} + v_t'\,, \tag{3.4}$$

where $r_T \to \infty$ with $r_T = o(\sqrt{T})$, Π is a full-rank matrix of size (q, p_1), and W_t is uncorrelated with v_t.

[4]In our application, we use several versions of Stock and Yogo's (2005) tests of weak identification, as well as the test recently proposed by Antoine and Renault (2012). All these testing procedures are only valid over stable models.

[5]The model is written with one break-point to ease the exposition. Testing the number of break-points is discussed in Theorem 3.5 below.

We now extend the results of HHB to the near-weak IV regression framework described by equations (3.3) and (3.4). We show that minimizing a 2SLS criterion provides consistent estimators of both the break fraction λ_1^0 and the structural parameters θ_i^0, $i = 1, 2$. Our estimation procedure is as follows. In the first stage, the reduced form for Y_t is estimated by OLS. Let $\hat{Y}_t$ denote the resulting predicted value for Y_t,

$$\hat{Y}_t = W_t' \left(\sum_{t=1}^{T} W_t W_t' \right)^{-1} \left(\sum_{t=1}^{T} W_t Y_t' \right) .$$

In the second stage, we consider the following 2SLS criterion:

$$Q_{2SLS}(T_1, \theta_1, \theta_2) = \sum_{t=1}^{T_1} \left(y_t - \hat{Y}_t' \theta_{y,1} - Z_t' \theta_{z,1} \right)^2 + \sum_{t=T_1+1}^{T} \left(y_t - \hat{Y}_t' \theta_{y,2} - Z_t' \theta_{z,2} \right)^2 ,$$

where T_1 is a candidate break-point. We first concentrate with respect to (θ_1, θ_2) to get $(\hat{\theta}_1(T_1), \hat{\theta}_2(T_1))$, and then minimize $Q_{2SLS}(T_1, \hat{\theta}_1(T_1), \hat{\theta}_2(T_1))$ over all possible valid partitions of the sample[6] defined by T_1. The estimators of the break-point $\hat{T}_1$ and of the structural parameters $\hat{\theta}_i$, $i = 1, 2$, are then defined as follows:

$$\hat{T}_1 \;=\; \arg\min_{T_1} Q_{2SLS}(T_1, \hat{\theta}_1(T_1), \hat{\theta}_2(T_1)) \;\text{ and }\; (\hat{\theta}_1, \hat{\theta}_2) = (\hat{\theta}_1(\hat{T}_1), \hat{\theta}_2(\hat{T}_1))$$

In order to deduce the asymptotic properties of the above estimators, some regularity assumptions are needed. These assumptions are similar to the ones discussed in HHB, except for near-weak identification, and they are presented in the appendix.

Theorem 3.3. *(Asymptotic properties)*
Define the following diagonal matrix Λ_T that collects the different rates of convergence as

$$\Lambda_T = \begin{pmatrix} \sqrt{T}/r_T \mathbf{I}_{p_1} & \mathbf{0} \\ \mathbf{0} & \sqrt{T}\mathbf{I}_{p_2} \end{pmatrix} \quad \text{with } \mathbf{I}_k \text{ the identity matrix of size } k.$$

(i) Under regularity assumptions 1 to 4, $\hat{\lambda}_1 \equiv \hat{T}_1/T$ is a consistent estimator of λ_1^0 such that $\|\hat{\lambda}_1 - \lambda_1^0\| = \mathcal{O}_P(1/T)$.
(ii) Under regularity assumptions 1 to 5, $\Lambda_T \left(\hat{\theta}_i - \theta_i^0 \right) \xrightarrow{d} \mathcal{N}(0, S_i)$, for $i = 1, 2$, with S_i defined in the appendix.

[6] A partition is valid if each subsample contains enough observations for the estimation to be meaningful. A formal statement of this assumption is provided in regularity assumption 1 in the appendix.

We have just shown that the estimators of the structural parameters associated with the endogenous variables, $\theta^0_{y,i}$, are asymptotically normally distributed, but at a slower rate than usual, namely $\sqrt{T}/r_T$ that comes from the near-weak instrumental variables W_t. It is worth mentioning that the estimator of the structural parameters associated with the exogenous regressors, $\theta^0_{z,i}$, are not affected by the near-weak instruments: they are asymptotically normally distributed at the standard rate $\sqrt{T}$.

We now turn to GMM estimation. We know from HHB that minimizing a GMM criterion (instead of the above 2SLS criterion) does not yield consistent estimation of the break fraction. However, given the above consistent 2SLS estimator of the break fraction, we show that GMM estimators of the structural parameters are not only consistent, but also more efficient than $\hat{\theta}_i$. The fast rate of consistence of $\hat{\lambda}_1$, (namely T in comparison to the rate of consistency of the estimators, $\sqrt{T}$ at best), explains intuitively why the standard result of the efficiency of GMM applies in this framework. The GMM estimators given the (estimated) break-point $\hat{T}_1$ are defined as follows:

$$\begin{aligned} \begin{pmatrix} \hat{\theta}_{GMM,1} \\ \hat{\theta}_{GMM,2} \end{pmatrix} &= \arg\min_{\theta_1,\theta_2} \left[\overline{g}'_T(\theta_1,\theta_2) S_T^{-1} \overline{g}_T(\theta_1,\theta_2) \right] , \\ \text{where } \ \overline{g}_T(\theta_1,\theta_2) &= \begin{bmatrix} \sum_{t=1}^{\hat{T}_1} W_t \left(y_t - Y_t'\theta_{y,1} - Z_t'\theta_{z,1} \right) \\ \sum_{t=\hat{T}_1+1}^{T} W_t \left(y_t - Y_t'\theta_{y,2} - Z_t'\theta_{z,2} \right) \end{bmatrix} , \end{aligned}$$

and S_T is a consistent estimator of $\texttt{AVar}\left[\overline{g}_T(\theta^0_1,\theta^0_2)\right]$ which is defined in the appendix.

The following result provides the asymptotic distribution of our GMM estimator. Similar to 2SLS estimators, we actually show that the estimated structural parameters associated with the endogenous variables are asymptotically normally distributed at rate $\sqrt{T}/r_T$, whereas the estimated structural parameters associated with the exogenous regressors are asymptotically normally distributed at the standard rate $\sqrt{T}$.

Theorem 3.4. *(Efficient GMM estimation)*
Under regularity assumptions 1 to 5, we have:
(i) The (unfeasible) GMM estimator $\hat{\theta}_{GMM}(T^0_1)$ of θ^0 given the true break-point T^0_1 is such that $\Lambda_T\left(\hat{\theta}_{GMM,i}(T^0_1) - \theta^0_i\right)$, for $i = 1, 2$, is asymptotically normally distributed with mean 0 and variance $S_{GMM,i}$ which is defined in the appendix.

(ii) Given the 2SLS estimator of the break fraction $\hat{\lambda}_1$ provided in Theorem 3.3, the (feasible) GMM estimator $\hat{\theta}_{GMM}$ of θ^0 that uses the estimated break-point $\hat{T}_1$ is asymptotically equivalent to $\hat{\theta}(T_1^0)$ in the sense that $\Lambda_T\left(\hat{\theta}_{GMM,i} - \theta_i^0\right)$, for $i = 1, 2$, is asymptotically normally distributed with mean 0 and variance $S_{GMM,i}$.

(iii) Given the 2SLS estimator of the break fraction $\hat{\lambda}_1$ provided in Theorem 3.3, the (feasible) GMM estimator $\hat{\theta}_{GMM}$ of the structural parameters is asymptotically more efficient than the 2SLS estimator in the sense that $[S_i - S_{GMM,i}]$ is semi-positive definite for any $i = 1, 2$.

Structural equation (3.3) assumes one break-point. However, in practice, the number of breaks needs to be detected a priori. For simplicity, we start by assuming a maximum of one break. To develop a Wald test of zero versus one break, we write the null and alternative hypotheses in terms of linear restrictions on the parameters. Accordingly, we define $R = \tilde{R} \otimes I_p$ where $\tilde{R} = [1, -1]$, and $\theta^{0\prime} = [\theta_1^{0\prime}, \theta_2^{0\prime}]$. The null and alternative hypotheses can then be stated as: $H_0 : R\theta^0 = 0$ versus $H_1 : R\theta^0 \neq 0$. The associated Wald test statistics writes

$$Sup-Wald_T(0:1) \; = \; \sup_{\lambda_1 \in \Lambda_\epsilon} \left[T\hat{\theta}'(T_1) R' [R \hat{V}_W(T_1) R']^{-1} R \hat{\theta}(T_1) \right] , \tag{3.5}$$

where $\Lambda_\epsilon = [\epsilon, 1-\epsilon]$, for a small $\epsilon > 0$ chosen by the researcher[7], $\hat{\theta}'(T_1) = [\hat{\theta}_1'(T_1) \; \hat{\theta}_2'(T_1)]$ is the 2SLS estimator of θ^0 based on candidate break-point $T_1 = \lfloor T\lambda_1 \rfloor$ and first-stage least-square projection $\hat{Y}_t$, $\hat{V}_W(T_1) = \text{diag}\left[\hat{V}_W^{(1)}(T_1), \hat{V}_W^{(2)}(T_1)\right]$, where for $i = 1, 2$,

$$\hat{V}_W^{(i)}(T_1) \; = \; \left\{ T^{-1} \sum_{t \in I_i} \hat{X}_t \hat{X}_t' \right\}^{-1} \hat{H}_i(T_1) \left\{ T^{-1} \sum_{t \in I_i} \hat{X}_t \hat{X}_t' \right\}^{-1} ,$$

with $I_1 \equiv [1, T_1]$, $I_2 \equiv [T_1 + 1, T]$, $\hat{X}_t \equiv [\hat{Y}_t \; Z_t]$, and $\hat{H}_i(T_1)$ a consistent estimator of

$$H_i = \lim_{T \to \infty} \texttt{Var} \left[T^{-1/2} \sum_{t \in I_i} \hat{X}_t (v_t' \theta_{y,i}^0 + u_t) \right] .$$

Here, $\hat{H}_i(T_1)(i = 1, 2)$ can be constructed using a HAC estimator.

The following theorem provides the limiting distribution of the sup-Wald test statistic.

[7] See regularity assumption 1 in the appendix.

Theorem 3.5. *(Test 0 vs 1 break)*
Under regularity assumptions 1 to 5,

$$Sup - Wald_T(0:1) \;\Rightarrow\; \sup_{\lambda_1 \in \Lambda_\epsilon} \frac{||B_{p_1}(\lambda_1) - \lambda_1 B_{p_1}(1)||^2}{\lambda_1(1-\lambda_1)},$$

where $\Rightarrow$ indicates weak convergence in Skorohod metric, $||\cdot||$ is the Euclidean norm, and $B_{p_1}(\lambda_1)$ is a $p_1 \times 1$ vector of independent standard Brownian motions defined on $[0,1]$.

This test can be generalized to yield a sequential procedure for estimating the number of breaks, in a similar fashion to Hall, Han and Boldea (2012): the interested reader is referred to pp. 22-23 for a detailed account of this procedure.

4 General framework

In this section, we consider the general framework that combines the two frameworks of interest introduced in the previous section. It allows instability both in the structural equation, and in the reduced form. Assuming again one break-point in each equation for ease of exposition, the model is

$$y_t = \begin{cases} Y_t'\theta^0_{y,1} + Z_t'\theta^0_{z,1} + u_t & , \; t \le \lfloor T\lambda^0_1 \rfloor \\ Y_t'\theta^0_{y,2} + Z_t'\theta^0_{z,2} + u_t & , \; t > \lfloor T\lambda^0_1 \rfloor \end{cases} \quad , \quad E(u_t|\mathcal{F}_{t-1}) = 0\,, \tag{4.1}$$

where λ^0_1 is the break fraction and $\lfloor w \rfloor$ the integer part of w. We define $\theta^{0\prime}_i = (\theta^{0\prime}_{y,i}\ \theta^{0\prime}_{z,i})$ for $i = 1,2$. For a given set of q valid instruments W_t with $q \ge p_1$,

$$Y'_{1t} = \begin{cases} \frac{W_t'\Pi_1}{r_{1T}} + v_t' & , \; t \le \lfloor T\nu^0_1 \rfloor \\ \frac{W_t'\Pi_2}{r_{2T}} + v_t' & , \; t > \lfloor T\nu^0_1 \rfloor \end{cases} \;, \tag{4.2}$$

where $r_{iT} = 1$, or $r_{iT} \to \infty$, $r_{iT} = o(\sqrt{T})$, Π_i is a full-rank matrix of size (q, p_1) for $i = 1,2$, and W_t is not correlated with v_t. The break fraction ν^0_1 may, or may not, coincide with the break fraction λ^0_1 of the structural form (3.3).
If the degrees of weakness are equivalent in both subsamples, $r_{1T} \propto r_{2T}$, the reduced form (4.2) is the natural extension to near-weak identification of unstable reduced forms already considered in the literature; see for instance Hall, Han and Boldea (2012). If the degrees

of weakness are not equivalent, (4.2) will be able to capture changes in the strength of identification. The challenge will then be to detect and locate both parameter instability and changes in the strength of identification, as well as to provide sharp and correct inference in the different subsamples. Our general inference procedure (described below) combines the results developed in sections 3.2 and 3.1; it is described below.

One question that has not been addressed so far is whether the break-points are common to both equations, that is whether $\lambda_1^0 = \nu_1^0$. This is an important practical question because ignoring a break, or incorrectly imposing one, may lead to incorrect, or inefficient inference. We believe that going through our general inference procedure first will help better understand the issue at stake. Therefore, we start by presenting our general 4-stage inference procedure. The discussion and results about common breaks follow.

1. First stage: the reduced form.

We first use the results developed in section 3.1.

(a) Find the (estimated) number of break-points $\hat{m}^*$. These break-points are referred to as T_i^*, $i = 1, \cdots, \hat{m}^*$.

(b) Given the estimated number of break-points $\hat{m}^*$, minimize the appropriate OLS criterion over all admissible partitions to find the estimated break-points $\hat{T}_i^*$ (see Theorem 3.1).

(c) Impose the estimated break-points $\hat{T}_i^*$ and use OLS to get $\hat{Y}_t$.

Additional comments:

Detecting the above break-points is always possible, irrespective of the identification strength. However, consistent estimation of the break-points is only possible when the identification strength is at least near-weak. In this respect, the test for identification strength developed in Antoine and Renault (2012b) is helpful to distinguish between weak and near-weak identification. The test procedure is presented in Appendix for completeness.

2. Second stage: the structural equation.

We now impose the above estimated break-points $\hat{T}_i^*$, and work over the associated subsamples separately. Note that on each such subsample the reduced form is stable. We thus use the results developed in section 3.2.

(a) Replace Y_t by $\hat{Y}_t$ in (4.1).

(b) Find the number of break-points m_i in each subsample i with $i = 1, \cdots, \hat{m}^* + 1$ (see Theorem 3.5). These break-points are referred to as T_j, $j = 1, \cdots, m_n$ where $m_n = m_1 + m_2 + \cdots + m_{\hat{m}^*+1}$.
(c) For each subsample, given the estimated number of break-points $\hat{m}_i$, minimize the appropriate 2SLS criterion over all admissible partitions to get the estimated break-points $\hat{T}_j$ (see Theorem 3.3.)
Additional comments:
Note that estimation of the above break-points is only possible when the identification is at least near-weak.

3. Third stage: the common breaks.
We need to check whether the break-points are common. This means that we check whether the break-points $\hat{T}_i^*$ found in the first stage, and imposed in the structural equation (4.1) in the second stage are actual break-points. To this end, we now partition the sample according to $\hat{T}_j$, and work over the associated subsamples.
(a) For each subsample, test for one (common) break (see Theorem 4.1 below).
(b) If the break is not common, it should not be imposed in the structural equation.
(c) If the break is common, we re-estimate this break-point using the structural equation because it is possible to get an estimator with a faster consistency rate. This break-point is also included in the set of break-points $\hat{T}_j$.

4. Fourth stage: estimation of structural parameters.
We are now ready to estimate the structural parameters. To this end, we partition the sample according to the complete set of break-points $\hat{T}_j$ deduced from stages 1 and 3, and we work over the associated subsamples. Over such subsample, the structural equation is sable while the reduced form might be unstable. We then use the results developed in section 3.1 and estimate the structural parameters by standard GMM.

We now present our Wald test for common break. Consider the case that there is exactly one break in the reduced form, T_1^*, estimated by $\hat{T}_1^*$ via the methods described in Section 3.2. Suppose there are no breaks that are only specific to the structural equation. To test whether T_1^* is a common break to the reduced form and structural equation, we test whether the 2SLS parameter estimates in intervals $\hat{I}_1^* = [1, \hat{T}_1^*]$ and $\hat{I}_2^* = [\hat{T}_1^* + 1, T]$ are equal. These parameter estimates are defined as $\tilde{\theta}_i = (\sum_{\hat{I}_i^*} \hat{X}_{t,i}\hat{X}_{t,i}')^{-1} \sum_{\hat{I}_i^*} \hat{X}_{t,i} y_t$, $i = 1, 2$, where $\hat{X}_{t,i} = (\hat{Y}_{t,i}', Z_t')'$, with $\hat{Y}_{t,i}$ being the projected endogenous regressors from the first stage in intervals $\hat{I}_i^*$. Let $\tilde{\theta} = (\tilde{\theta}_1', \tilde{\theta}_2')'$. Then the Wald test for a common break is:

$$Wald_T = T\tilde{\theta}' R' [R\hat{V}R']^{-1} R\tilde{\theta},$$

with $\hat{V} = \text{diag}\left[\hat{V}_1, \hat{V}_2\right]$, and for $i = 1, 2$,

$$\hat{V}_i = \left\{ T^{-1} \sum_{\hat{I}_i^*} \hat{X}_{t,i}\hat{X}_{t,i}' \right\}^{-1} \hat{H}_i \left\{ T^{-1} \sum_{\hat{I}_i^*} \hat{X}_{t,i}\hat{X}_{t,i}' \right\}^{-1},$$

and $\hat{H}_i \xrightarrow{p} \lim_{T\to\infty} \texttt{Var}\left[T^{-1/2} \sum_{I_i^*} \hat{X}_{t,i} u_t\right]$, with $I_1^* = [1, T_1^*]$ and $I_2^* = [T_1^* + 1, T]$.

Theorem 4.1. *(Wald test for common break)*
Under the null of no break in the structural equation and Assumptions 1-5, $Wald_T \xrightarrow{d} \chi_p^2$.

5 Application to the New Keynesian Phillips Curve

For our empirical analysis, we consider the New Keynesian Phillips Curve (NKPC). The NKPC is a dynamic relationship resulting from a limited (or full-information) equilibrium model between inflation and driving variables such as the output gap, unemployment or real marginal costs. There is a large body of empirical evidence suggesting that the NKPC is both unstable and possibly not strongly identified. Instruments commonly used for empirical works on the NKPC include lags of each model's dependent and forcing variables[8].
A typical NKPC equation takes the following form

$$\pi_t = \alpha_f \pi_{t+1}^e + \alpha_b \pi_{t-1} + \alpha_y y_t + u_t \quad t = 1, \cdots, T$$

[8]More recently, researchers have identified additional useful instruments such as the long-short interest rate spread (Gali and Gertler (1999) and Gali, Gertler, and Lopez-Salido (2001)), lags of model dependent and forcing variables from various competing specifications (Dufour, Khalaf, and Kichian (2010)), and factors extracted via principle components from the 132 variables in Stock and Watson (2005) (Kapetanios and Marcellino (2010)). This is beyond the scope of this study.

where π_t is the inflation, π^e_{t+1} is inflation's expectations, and y_t is a driving variable measured via the output gap, unemployment, or real marginal costs.

We use quarterly data for the United States from 1968:3 to 2005:4. π_t is computed as the annualized quarterly growth rate of GDP implicit price deflator (i.e. 400 times the first difference of the log GDP deflator); y_t is the output gap, obtained as real log GDP minus real log potential GDP. The GDP price deflators and real GDP measures are taken from the freely available FRED2 dataset[9], while real potential GDP is taken from the estimates published by the Congressional Budget Office (CBO). We use as expected inflation, π^e_{t+1}, the quarterly Greenbook median inflation forecasts (GRB), available with a lag of five years from the Philadelphia Fed webpage[10], representing the median of inflation forecasts of several experts one quarter ahead.

We produce results using the following set of instruments w_t, commonly used in this literature (see e.g. Clarida, Gali, and Gertler (1998)):

$$\begin{aligned} \pi^e_{t+1} &= w_t\delta_1 + v_{1,t} \\ y_t &= w_t\delta_2 + v_{2,t} \end{aligned}$$

where w_t contains the intercept and the first lags of the output gap, inflation, expected inflation, the interest rate, the money growth and the unemployment rates. Here, the interest rate is measured as the real interest rate on three-month Treasury Bills, the money growth as the growth of the M2 aggregate, and the unemployment rate refers to the civilian unemployment rate, all available from the FRED2 dataset. Further details about the data can be found from their respective sources indicated above and from Zhang, Osborn and Kim (2008).[11]

- **Empirical strategy and results.**

First, we test for breaks in the reduced form. We find two break-points in π_t. The associated estimators are: 1974:2 and 1980:4. We do not find any break in y_t.

Second, we test for common and additional breaks in the structural equation. We find one common break, estimated at 1974:2 and one additional break at 1999:4.

[9]See http://research.stlouisfed.org/fred2/.

[10]See http://www.phil.frb.org/research-and-data/real-time-center/greenbook-data/.

[11]We would like to thank Denise Osborn for providing us with this dataset.

Third, we test the identification strength. We rely on the following procedures: (i) the test proposed by Antoine and Renault (2012b) that is based on a J-test calculated at a distorted GMM estimator (see the appendix for practical implementation details); (ii) the tests proposed by Stock and Yogo (2005) that are based on the 2SLS bias and size distortion[12]. Both tests consider the null hypothesis of weak identification. The advantage of the test of Antoine and Renault (2012b) is that it relies on the efficient GMM estimator (as shown and discussed in sections 3 and 4); as a result, we may expect it to be more powerful. We impose the breaks of the structural and reduced form equations before running the tests. Our results are displayed in Table 1 where 1 indicates "rejection of the null hypothesis of weak identification", whereas 0 indicates "no rejection of the null".

Focusing on the test of Stock and Yogo (2005) based on 10% bias and the test of Antoine and Renault (2012b), we conclude that there is only one subsample that is weakly identified, namely subsample 2a from 1974:3 to 1980:4. Such results also confirm *a posteriori* that the above estimated break points are likely consistent.

Fourth, we provide the estimators of the structural parameters after imposing the estimated break points of the structural equation only. The GMM estimators along with 95% confidence intervals and identification strength are presented in Table 2. It is important to point out that the estimators over subsample 2 (which is the combination of subsamples 2a and 2b in Table 1) are consistent even if the first part of the subsample is weakly identified. See Theorem 3.2.

• **Discussion of our results.**

Our results confirm previous findings of instability in the NKPC. However, we also show that the instrument strength changes from 1968 to 2005. While the first and third subsamples are quite small for reliable interpretation, we find that over the second subsample from 1974:3 to 1999:4, the strength of the instruments is not weak.[13] Our results are in contrast

[12] The test based on 10% bias is commonly used in practice.

[13] As already explained, the instruments are weak over subsample 2a and not weak over subsample 2b. According to Theorem 3.2, the GMM estimators of the structural parameters inherit the rate of convergence of the stronger subsample (here subsample 2b). This explains why the identification strength of subsample 2 is qualified as "not weak".

Subsample 1 - 1968:4 to 1974:2 (23 obs.)					
Antoine and Renault (2012)	*Stock and Yogo (2005)*				
	Bias 5%	Bias 10%	Bias 20%	Size dist. 10%	Size dist. 15%
1	0	1	1	0	0
Subsample 2a - 1974:3 to 1980:4 (26 obs.)					
Antoine and Renault (2012)	*Stock and Yogo (2005)*				
	Bias 5%	Bias 10%	Bias 20%	Size dist. 10%	Size dist. 15%
0	0	0	0	0	0
Subsample 2b - 1981:1 to 1999:4 (76 obs.)					
Antoine and Renault (2012)	*Stock and Yogo (2005)*				
	Bias 5%	Bias 10%	Bias 20%	Size dist. 10%	Size dist. 15%
1	1	1	1	1	1
Subsample 3 - 2000:1 to 2005:4 (24 obs.)					
Antoine and Renault (2012)	*Stock and Yogo (2005)*				
	Bias 5%	Bias 10%	Bias 20%	Size dist. 10%	Size dist. 15%
1	1	1	1	0	1

Table 1: Testing identification strength over stable subsamples (after imposing estimated breaks of structural and reduced form equation. 1 indicates ”rejection of the null hypothesis of weak identification”, and 0 ”no rejection”.

with previous findings. This contrast can actually be explained by the fact that we use identification strength tests over stable subsamples, so that the tests results are no longer biased by the presence of breaks. It is interesting to point out that when the above tests for identification strength are applied over the whole sample, the test of Antoine and Renault (2012b) concludes that the identification is weak, whereas all the tests of Stock and Yogo (2005) reject the weakness. This may indicate that the full-sample tests in Antoine and Renault (2012b) are more robust to breaks than the full-sample Stock and Yogo's (2005) tests, although a formal analysis of this statement is beyond the scope of the paper.

Our inference procedures allow us to recover a large and significant coefficient on output gap. These results are encouraging for practitioners and central bankers as they indicate

Subsample 1 - 1968:4 to 1974:2 (23 obs.)				
Identif. strength	*Parameter*	*Estimator*	*CI*	
Not weak	α_f	0.9649*	0.7007	1.2292
	α_b	0.2437*	0.0027	0.4847
	α_y	-2.1668*	-3.5347	-0.7988
Subsample 2 - 1974:3 to 1999:4 (102 obs.)				
Identif. strength	*Parameter*	*Estimator*	*CI*	
Not weak	α_f	0.7351*	0.3782	1.0919
	α_b	0.1994	-0.1554	0.5543
	α_y	3.7973*	1.5589	6.0357
Subsample 3 - 2000:1 to 2005:4 (24 obs.)				
Identif. strength	*Parameter*	*Estimator*	*CI*	
Not weak	α_f	0.8796*	0.4435	1.3156
	α_b	0.4789*	0.1575	0.8004
	α_y	-1.1362	-2.5110	0.2387

Table 2: Estimation results of the structural equation, $\pi_t = \alpha_f \pi^e_{t+1} + \alpha_b \pi_{t-1} + \alpha_y y_t + u_t$.

that there is a positive trade-off between output gap and inflation, at least for part of our sample. We also confirm previous findings that the agents are forward-looking, and that forward-looking behavior is more important than backward-looking behavior in explaining inflation.

6 Conclusion

There is considerable evidence suggesting that macroeconomic models such as the NKPC are subject to parameter instability and to identification issues. This is the first paper to our knowledge to consider these two problems in a unified framework, and provide a comprehensive treatment of the link between them.
It is important to realize that parameter instability can be linked or not to changes in identification strength (that is instrument strength). As a result, we start by considering

two simplified scenarios in a linear model with endogeneity. When the parameter instability occurs with no changes in instrument strength, we show that standard procedures to detect break-points, and estimate them, along with the structural parameters of interest, are asymptotically valid, though at a rate that can be slower than $\sqrt{T}$ due to weaker identification patterns. In addition, we show how to conduct sharp inference in these settings via GMM conditional on the break-points (rather than 2SLS as in Hall, Han and Boldea (2012)).

When the breaks are linked to changes in instrument strength, we prove that the sharpest inference is based on the full-sample GMM estimator, that recovers the fastest convergence rate possible. Combining these two scenarios, we show that detecting and locating changes in instrument strength is essential for correct and efficient asymptotic inference, and we provide a step-by-step guide for practitioners on detecting and locating these changes.

We apply our procedure to the US NKPC and confirm previous findings of instability. More importantly, we show that the instability is linked to changes in instrument strength and that over the sample 1968-2005, the instruments we use are both weak and not weak. These results are encouraging for practitioners as our methods are easier to implement than the weak-instrument-robust inference procedures.

References

[1] D. W. K. Andrews, *Tests for parameter instability and structural change with unknown change point*, Econometrica **61** (1993), 821–856.

[2] D. W. K. Andrews and R. Fair, *Inference in econometric models with structural change*, Review of Economic Studies **55** (1988), 615–640.

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

[4] B. Antoine and E. Renault, *Efficient GMM with Nearly-weak Instruments*, The Econometrics Journal, Tenth Anniversary Issue **12** (2009), 135–171.

[5] ———, *Efficient Inference with Poor Instruments: a General Framework*, ch. 2 in Handbook of Empirical Economics and Finance, Taylor & Francis, 2010.

[6] ———, *Efficient Minimum Distance Estimation with Multiple Rates of Convergence*, forthcoming Journal of Econometrics (2012a).

[7] ———, *Testing Identification Strength*, Working paper, Simon Fraser University (2012b).

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

[9] ———, *Critical values for multiple structural change tests*, Econometrics Journal **6** (2003), 72–78.

[10] O. Boldea and A. R. Hall, *Testing Structural Stability in Macroeconometric Models*, forthcoming in Handbook of Research Methods and Applications on Empirical Macroeconomics edited by N. Hashimzade and M. Thornton, to be published by Edward Elgar (2011).

[11] O. Boldea, A. R. Hall, and S. Han, *Asymptotic distribution theory for break point estimators in models estimated via 2SLS*, Econometric Reviews **31** (2012), 1–33.

[12] M. Caner, *Estimation in GMM and CUE with Nearly-Weak Instruments*, Econometrics Reviews **29** (2010), 330–363.

[13] F. Canova and L. Sala, *Back to Square One: Identification Issues in DSGE Models*, Journal of Monetary Economics **56** (2009), 431–449.

[14] R. Clarida, J. Galí, and M. Gertler, *Monetary policy rules in practice: Some international evidence*, European Economic Review **42** (1998), 1033–1067.

[15] J.-M. Dufour, L. Khalaf, and M. Kichian, *On the Precision of Calvo Parameter Estimates in Structural NKPC Models*, Journal of Economic Dynamics and Control **34** (2010), 1582–1595.

[16] J. Galí and M. Gertler, *Inflation Dynamics: A Structural Econometric Analysis*, Journal of Monetary Economics **44** (1999), 195–222.

[17] J. Galí, M. Gertler, and J.D. Lopez-Salido, *European Inflation Dynamics*, European Economic Review **45** (2001), 1237–1270.

[18] E. Ghysels and A. R. Hall, *Are consumption based intertemporal asset pricing models structural?*, Journal of Econometrics **45** (1990a), 121–139.

[19] ———, *A test for structural stability of Euler condition parameters estimated via the Generalized Method of Moments*, International Economic Review **31** (1990b), 355–364.

[20] J. Hahn and G. Kuersteiner, *Discontinuities of Weak Instruments limiting Distributions*, Economics Letters **75** (2002), 325–331.

[21] A. R. Hall, S. Han, and O. Boldea, *Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors*, forthcoming in Journal of Econometrics (2012).

[22] A. R. Hall and A. Sen, *Structural stability testing in models estimated by Generalized Method of Moments*, Journal of Business and Economic Statistics **17** (1999), 335–348.

[23] C. Hansen, J. Hausman, and W. Newey, *Estimation with Many Instrumental Variables*, Journal of Business and Economic Statistics **26** (2008), 398–422.

[24] L. P. Hansen, J. Heaton, and A. Yaron, *Finite Sample Properties of some Alternative GMM Estimators*, Journal of Business and Economic Statistics **14** (1996), 262–280.

[25] L. P. Hansen and J. Singleton, *Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models*, Econometrica **50** (1982), 1269–1286.

[26] L.P. Hansen, *Large Sample Properties of Generalized Method of Moments Estimators*, Econometrica **50** (1982), no. 4, 1029–1054.

[27] G. Kapetanios and M. Marcellino, *Factor-GMM estimation with large sets of possibly weak instruments*, Computational Statistics and Data Analysis **54** (2010), no. 11, 2655–2675.

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

[29] Z. Qu and P. Perron, *Estimating and testing structural changes in multivariate regressions*, Econometrica **75** (2007), 459–502.

[30] D. Staiger and J. Stock, *Instrumental Variables Regression with Weak instruments*, Econometrica **65** (1997), 557–586.

[31] J.H. Stock and M.W. Watson, *Implications of dynamic factor models for VAR analysis*, mimeo (2011).

[32] J.H. Stock and J.H. Wright, *GMM with Weak Identification*, Econometrica **68** (2000), no. 5, 1055–1096.

A Regularity assumptions

For simplicity, we restate the DGPs for different cases here. The stable structural equation (with no breaks) is

$$y_t = Y_t'\theta_y^0 + Z_t'\theta_z^0 + u_t, \tag{A.1}$$

and

$$y_t = \begin{cases} Y_t'\theta_{y,1}^0 + Z_t'\theta_{z,1}^0 + u_t, & t \leq T_1^0 \\ Y_t'\theta_{y,2}^0 + Z_t'\theta_{z,2}^0 + u_t, & t > T_1^0 \end{cases}, \tag{A.2}$$

where $T_1^0 = [T\lambda_1^0]$. The stable reduced form in section 3.2 is

$$Y_t' = \frac{W_t'\Pi}{r_T} + v_t'. \tag{A.3}$$

The unstable reduced form in section 3.1 is

$$Y_t' = \begin{cases} \frac{W_t'\Pi_1}{r_{1T}} + v_t', & t \leq T_1^* \\ \frac{W_t'\Pi_2}{r_{2T}} + v_t', & t > T_1^* \end{cases} \tag{A.4}$$

where $T_1^* = [T\nu_1^0]$.

We assume throughout that Z_t is a subset of W_t. The following regularity assumptions are essentially similar to the ones imposed in HHB, and are reproduced for completeness.

Assumption 1. *(Regularity of the break fractions)*
The break fractions are such that: $0 < \lambda_1^0 < 1$ *and* $0 < \nu_1^0 < 1$*. This means that for the minimizations of appropriate criterions (e.g. 2SLS or GMM), we consider the set of candidate break-points to be such that*

$$\max(T_1, T - T_1) > \max(q-1, \epsilon T) \ \textit{ for some } \epsilon > 0 \textit{ such that},$$

$$\epsilon < \min(\lambda_1^0, 1-\lambda_1^0)\,, \ \ \epsilon < \min(\nu_1^0, 1-\nu_1^0)\,.$$

Assumption 2. *(Regularity of the error terms)*
Define h_t *as the array of real-valued (n,1) random vectors (with* $n = (p_1+1) \times q$*) of the errors of the structural equation and reduced form times the instruments,* $h_t \equiv \begin{pmatrix} u_t \\ v_t \end{pmatrix} \otimes W_t$*. It is defined on the probability space* $(\Omega, \mathcal{F}, P)$*. The i-th element of* h_t *is denoted* $h_{t,i}$*.*

(i) Consider $V_T \equiv \texttt{Var}\left(\sum_{t=1}^{T} h_t\right)$ *and its eigenvalues* $(\xi_{T,1}, \ldots, \xi_{T,n})$*. Then the* (n,n)*-diagonal matrix with these eigenvalues along the main diagonal is* $\mathcal{O}(T^{-1})$*.*
(ii) $E(h_{t,i}) = 0$ *and for some* $d > 2$, $\|h_{t,i}\|_d < \Gamma < \infty$ *for* $t = 1, \cdots, T$ *and* $i = 1, \cdots, n$*.*
(iii) $\{h_{t,i}\}$ *is near epoch dependent with respect to some process* $\{g_t\}$ *such that* $\|h_t - E(h_t|\mathcal{G}_{t-m}^{t+m}\|_2 \leq \nu_m$ *with* $\nu_m = \mathcal{O}(m^{-1/2})$ *where* $\mathcal{G}_{t-m}^{t+m}$ *is a* σ*-algebra based on* $(g_{t-m}, \cdots, g_{t+m})$*.*
(iv) $\{g_t\}$ *is either* ϕ*-mixing of size* $m^{-d/[2(d-1)]}$ *or* α*-mixing of size* $m^{-d/(d-2)}$*.*

Assumption 3. *(Regularity of the reduced form)*
(i) If there is no break in the reduced form, as in equation (A.3)*,* $\texttt{Rank}(\Pi) = p_1$*.*
(ii) If there are breaks in the reduced form, as in equation (A.4)*, then* $\texttt{Rank}(\Pi_i) = p_1$*, for* $i = 1, 2$*.*

Assumption 4. *(Regularity of the instrumental variables)*
(i) There exists an $0 < l_0 < \min\{T_1^0, T - T_1^0\}$ *such that for all* $l = [\epsilon T] > l_0$*, for some* $\epsilon \in (0,1)$*, with* $l \leq \min\{T_1^0, T - T_1^0\}$*, the minimum eigenvalues of* $(1/l)\sum_{t=T_1^0+1}^{T_1^0+l} W_t W_t{}'$ *and of* $(1/l)\sum_{t=T_1^0-l}^{T_1^0} W_t W_t{}'$ *are bounded away from zero in probability.*
(ii)

$$T^{-1}\sum_{t=1}^{\lfloor Tr \rfloor} W_t W_t' \xrightarrow{P} Q_{WW}(r)\,,$$

uniformly in $r \in [0,1]$ *where* $Q_{WW}(r)$ *is positive definite (pd) for any* $r > 0$ *and strictly increasing in* r*.*
(iii) If there is a break in the reduced form at T_1^**, then:*

$$T^{-1}\sum_{t=1}^{T_1^*} W_t W_t' \xrightarrow{p} Q_1 \text{ and } T^{-1}\sum_{t=T_1^*+1}^{T} W_t W_t' \xrightarrow{p} Q_2,$$

two pd matrices of constants.

Assumption 5. *(i) If there is a break in the reduced form at* T_1^**, then:*

$$\texttt{Avar}\left[T^{-1/2}\sum_{t=1}^{T_1^*} h_t\right] = V_1^* \text{ and } \texttt{Avar}\left[T^{-1/2}\sum_{t=T_1^*+1}^{T} h_t\right] = V_2^*,$$

two pd matrices of constants. (ii) Define $V_T^*(r) = Var[T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} h_t]$*. Then* $\lim_{T\to\infty} V_T^*(r) = rV^*$ *uniformly in* $r \in [0,1]$ *where* V^* *is a positive definite matrix of constants. In addition, the matrix* $Q_{WW}(r)$ *defined in Assumption 4(ii) takes the form* $Q_{WW}(r) = rQ_{WW}$*.*

Assumption 6. *(Conditional homoskedasticity) Assumptions 4(iii) and 5(i) hold, with* $V_i^* = \sigma_u^2 Q_i$, *for* $i = 1, 2$, *and* $\sigma_u^2 = E[u_t^2|\mathcal{F}_{t-1}]$, *where* $\mathcal{F}_{t-1} = \{W_t, W_{t-1}, \ldots\}$.

Assumption 5 cannot be generalized to within-regime heteroskedasticity, since the distribution of the Wald test statistic under the null is no longer valid. In that case, a bootstrap procedure is recommended to mimic the asymptotic distribution of the Wald test statistic under the null.

To simplify exposition and provide better intuition of results, all the sections below are written for a single endogenous regressor ($p_1 = 1$), no exogenous regressors, $p_2 = 0$ (for the case of RF breaks) and one exogenous regressor, $p_2 = 1$, (for the case of SE breaks), and one break only in either the SE or RF. Their generalization to multiple endogenous and exogenous regressors, and to multiple breaks can be found in the Supplemental Appendix, available from the authors upon request.

B Restatement of Theorems proven for one break

Theorem B.1. *(Consistency of break fraction - Theorem 3.1)*
(i) Under regularity assumptions 1 to 4(ii), the break fraction estimator $\hat{\nu}_1 \equiv \hat{T}_1^*/T$ *is a consistent estimator of* ν_1^0.
(ii) Under regularity assumptions 1 to 4(ii), $\|\hat{\nu}_1 - \nu_1^0\| = \mathcal{O}_P(R_T^2/T)$, *where* $R_T = \min_i(r_{i,T})$.
(iii) Under regularity assumptions 1 to 5(i), consider the full sample 2SLS estimator $\hat{\theta}_y$ *constructed with first-stage* $\hat{Y}_t = \begin{cases} W_t'\hat{\Pi}_1, & t \leq \hat{T}_1^* \\ W_t'\hat{\Pi}_2, & t > \hat{T}_1^* \end{cases}$. *Then:*

$$T^{1/2} R_T^{-1}(\hat{\theta}_y - \theta_y^0) \xrightarrow{d} \mathcal{N}(0, V_{2SLS}),$$

with V_{2SLS} *defined below. Define* $R_T = \min(r_{1T}, r_{2T})$.
Case (a). For $R_T = r_{1T} = r_{2T}$, $V_{2SLS} = B^{-1}VB^{-1}$, $B = \Pi_1'Q_1\Pi_1 + \Pi_2'Q_2\Pi_2$, $V = \Pi_1'V_{u,1}^*\Pi_1 + \Pi_2'V_{u,2}^*\Pi_2$, *and* $V_{u,1}^*, V_{u,2}^*$ *are the* $q \times q$ *upper left blocks of the matrices* V_1^*, V_2^*.
Case (b). Without loss of generality (Wlog), assume $R_T = r_{1T} = o(r_{2T})$. *Then* $V_{2SLS} = B_1^{-1}V_1B_1^{-1}$, $B_1 = \Pi_1'Q_1\Pi_1$ *and* $V = \Pi_1'V_{u,1}^*\Pi_1$.

Theorem B.2. *(GMM estimation of the structural parameters - Theorem 3.2)*
Under regularity assumptions 1 to 5(i), consider the full sample GMM estimator $\hat{\theta}_{GMM,y}$ *with weighting matrix* Σ^{-1}*. Then:*

$$T^{1/2}R_T^{-1}(\hat{\theta}_{GMM,y} - \theta_y^0) \xrightarrow{d} \mathcal{N}(0, V_{GMM}),$$

with V_{GMM} *defined below.*
Case (a). For $R_T = r_{1T} = r_{2T}$*,* $V_{GMM} = \tilde{B}^{-1}\tilde{V}\tilde{B}^{-1}$*,* $\tilde{B} = (\Pi_1'Q_1 + \Pi_2'Q_2)\Sigma^{-1}(Q_1\Pi_1 + Q_2\Pi_2)$*,* $\tilde{V} = (\Pi_1'Q_1 + \Pi_2'Q_2)\Sigma^{-1}\ V_u^*\ \Sigma^{-1}(Q_1\Pi_1 + Q_2\Pi_2)$*, and* $V_u^* = V_{u,1}^* + V_{u,2}^*$*.*
Case (b). Without loss of generality (Wlog), assume $R_T = r_{1T} = o(r_{2T})$*. Then* $V_{GMM} = \left[\Pi_1'Q_1'V_{u,1}^*Q_1\Pi_1\right]^{-1}$*.*

Bertille: New Theorems

Theorem B.3. *(Efficiency of GMM vs. 2SLS - Theorem* **??***) Let regularity assumptions 1 to 5(i) hold. Also assume the special case of one endogenous regressor,* $p_1 = 1$*, no exogenous regressors* $p_2 = 0$*, one break in the RF, and no break in SE.*
Case (a): $R_T = r_{1T} = r_{2T}$*. Due to the RF break, in general, the full-sample 2SLS estimator is no longer a special case of the full-sample GMM estimator. As a consequence, even in the just-identified case,* $q = 1$*, efficiency of one estimator compared to the other depends on the parameter configuration.*
Case (b): Wlog, assume $R_T = r_{1T} = o(r_{2T})$*. In this case, both the full-sample GMM and 2SLS estimators reduce to their sub-sample counterparts pertaining to the strongest identified sub-sample. As a result, the optimal GMM estimator with* $\Sigma = Q_1$ *is more efficient than the 2SLS estimator.*

Theorem B.4. *(Test for 0 vs. 1 break in RF - Theorem* **??***)*

Define $\Lambda_\epsilon = [\epsilon, 1-\epsilon]$*, for some* $\epsilon > 0$*, and the null hypothesis of* $\Pi_T^0 = \Pi_1/r_{1T} = \Pi_2/r_{2T}$ *(i.e.* $\Pi_1 = \Pi_2$ *and* $r_{1T} = r_{2T}$*), versus the alternative that* $\Pi_1 \neq \Pi_2$ *or* $r_{iT} = o(r_{jT})$ *for some* $i \neq j, i, j \in \{1, 2\}$*. Then the Bai and Perron (1998) Wald test statistic for testing for one break in the reduced form is:*

$$Sup-Wald_T(0:1) \;=\; \sup_{\nu \in \Lambda_\epsilon}\left[T\hat{\Pi}'(\nu)R'[R\hat{G}_TR']^{-1}R\hat{\Pi}'(\nu)\right], \qquad \text{(B.1)}$$

where $R = (1, -1) \otimes I_q$, $\hat{\Pi}'(\nu) = [\hat{\Pi}_1'(\nu), \hat{\Pi}_2'(\nu)]'$, *with* $\hat{\Pi}_1'(\nu), \hat{\Pi}_2'(\nu)$ *the OLS estimators of* Π^0 *in sub-samples* $1, \ldots, [T\nu]$*, respectively* $[T\nu]+1, \ldots, T$*. We assume that* $\hat{G}_T \xrightarrow{p} G$*, where* $G = Q_{WW}^{-1} V^* Q_{WW}^{-1}$*. Then, under regularity assumptions 1 to 5 and the null hypothesis, this test has the same null asymptotic distribution as in Bai and Perron (1998), i.e.*

$$Sup - Wald_T(0:1) \ \Rightarrow \ \sup_{\nu \in \Lambda_\epsilon} \frac{||B_q(\nu) - \nu B_q(1)||^2}{\nu(1-\nu)},$$

where $\Rightarrow$ *indicates weak convergence in Skorohod metric,* $|| \cdot ||$ *is the Euclidean norm, and* $B_q(\nu)$ *is a* $q \times 1$ *vector of independent standard Brownian motions defined on* $[0,1]$*.*

C Proofs of the main results

Proof of Theorem B.1 *(Consistency of the break fractions).*
We assume throughout the proof that $\hat{T}_1^* < T_1^*$. The proof for $\hat{T}_1^* \geq T_1^*$ is similar and omitted for simplicity.
Part (i). Let $\hat{\Pi}_1, \hat{\Pi}_2$ be the sub-sample OLS estimators of Π_1/r_{1T}, Π_2/r_{2T} in intervals $[1, \hat{T}_1^*]$, respectively $[\hat{T}_1^*+1, T]$. Also let $\hat{v}_t = Y_t - W_t'\hat{\Pi}_1$ in interval $[1, \hat{T}_1^*]$, $\hat{v}_t = Y_t - W_t'\hat{\Pi}_2$ in interval $[\hat{T}_1^* + 1, T]$, and $d_t^* = \hat{v}_t - v_t$. We show consistency of the break-fraction by contradiction, in two steps. By definition of the sum of squared residuals, $\sum_{t=1}^T \hat{v}_t^2 \leq \sum_{t=1}^T v_t^2$, hence:

$$2\sum_{t=1}^T v_t d_t^* + \sum_{t=1}^T (d_t^*)^2 \leq 0. \tag{C.1}$$

In the first step, we show that:

$$\sum_{t=1}^T (d_t^*)^2 = \mathcal{O}_P(T/R_T^2) \text{ and } \sum_{t=1}^T v_t d_t^* = \mathcal{O}_P(T^{1/2}/R_T), \tag{C.2}$$

implying that $\sum_{t=1}^T (d_t^*)^2$ dominates $2\sum_{t=1}^T v_t d_t^*$ and so[14] from (C.1),

$$\text{plim}(R_T^2/T)\sum_{t=1}^T (d_t^*)^2 \leq 0. \tag{C.3}$$

[14]Note that if $R_T = T^{1/2}$, then $\sum_{t=1}^T (d_t^*)^2$ and $2\sum_{t=1}^T v_t d_t^*$ are of the same order, and the proof argument cannot be applied.

Because $(R_T^2/T)\sum_{t=1}^T (d_t^*)^2 \geq 0$, (C.3) cannot hold unless $\text{plim}(R_T^2/T)\sum_{t=1}^T (d_t^*)^2 = 0$. In the second step, we show that if $\hat{\nu}_1 \overset{p}{\nrightarrow} \nu_1^0$, then with positive probability, $(R_T^2/T)\sum_{t=1}^T d_t^2 > 0$, which contradicts $\text{plim}(R_T^2/T)\sum_{t=1}^T (d_t^*)^2 = 0$, so it cannot hold. Thus, consistency is proven if the two steps are estalished.

Step 1. Start by noting that:

$$d_t^* = \hat{v}_t - v_t = \begin{cases} Y_t - W_t'\hat{\Pi}_1, & t \leq \hat{T}_1^* \\ Y_t - W_t'\hat{\Pi}_2, & t > \hat{T}_1^* \end{cases} = \begin{cases} W_t'(\Pi_1/r_{1T} - \hat{\Pi}_1), & t \leq \hat{T}_1^* \\ W_t'(\Pi_1/r_{1T} - \hat{\Pi}_2), & \hat{T}_1^* + 1 \leq t \leq T_1^* \\ W_t'(\Pi_2/r_{2T} - \hat{\Pi}_2), & t > T_1^* \end{cases}.$$

It follows that:

$$\sum_{t=1}^T v_t d_t^* = (\Pi_1/r_{1T} - \hat{\Pi}_1)' \sum_{t=1}^{\hat{T}_1^*} W_t v_t + (\Pi_1/r_{1T} - \hat{\Pi}_2)' \sum_{\hat{T}_1^*+1}^{T_1^*} W_t v_t + (\Pi_2/r_{2T} - \hat{\Pi}_2)' \sum_{T_1^*+1}^{T} W_t v_t. \tag{C.4}$$

By Assumptions 1-2 and the functional CLT in Wooldridge and White (1988), Theorem 2.11, $\sum_{t=1}^{[Tr]} W_t v_t = \mathcal{O}_P(T^{1/2})$, uniformly in $r \in (0,1]$. Thus,

$$\sum_{t=1}^{\hat{T}_1^*} W_t v_t = \mathcal{O}_P(T^{1/2}), \sum_{\hat{T}_1^*+1}^{T_1^*} W_t v_t = \mathcal{O}_P(T^{1/2}), \sum_{T_1^*+1}^{T} W_t v_t = \mathcal{O}_P(T^{1/2}). \tag{C.5}$$

Also, by Assumption 4(ii),

$$\Pi_1/r_{1T} - \hat{\Pi}_1 = -\left(\sum_{t=1}^{\hat{T}_1^*} W_t W_t'\right)^{-1} \sum_{t=1}^{\hat{T}_1^*} W_t v_t = \mathcal{O}_P(T^{-1})\mathcal{O}_P(T^{1/2}) = \mathcal{O}_P(T^{-1/2}). \tag{C.6}$$

On the other hand,

$$\begin{aligned} \Pi_2/r_{2T} - \hat{\Pi}_2 = &-\left(\sum_{t=\hat{T}_1^*+1}^{T} W_t W_t'\right)^{-1} \sum_{t=\hat{T}_1^*+1}^{T} W_t v_t \\ &-\left(\sum_{t=\hat{T}_1^*+1}^{T} W_t W_t'\right)^{-1} \left(\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t W_t'\right)\left(\frac{\Pi_1}{r_{1T}} - \frac{\Pi_2}{r_{2T}}\right) \\ &= \mathcal{O}_P(T^{-1})\mathcal{O}_P(T^{1/2}) + \mathcal{O}_P(T^{-1})\mathcal{O}_P(T)\mathcal{O}_P(1/R_T) \\ &= \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(1/R_T) = \mathcal{O}_P(1/R_T). \end{aligned} \tag{C.7}$$

Similarly,

$$\Pi_1/r_{1T} - \hat{\Pi}_2 = (\Pi_1/r_{1T} - \Pi_2/r_{2T}) + (\Pi_2/r_{2T} - \hat{\Pi}_2) = \mathcal{O}_P(1/R_T) + \mathcal{O}_P(1/R_T) = \mathcal{O}_P(1/R_T). \tag{C.8}$$

Substituting (C.5)-(C.8) into (C.4) yields:

$$\sum_{t=1}^{T} v_t d_t^* = \mathcal{O}_P(T^{-1/2})\mathcal{O}_P(T^{1/2}) + \mathcal{O}_P(1/R_T)\mathcal{O}_P(T^{1/2}) + \mathcal{O}_P(1/R_T)\mathcal{O}_P(T^{1/2}) = \mathcal{O}_P(T^{1/2}/R_T). \tag{C.9}$$

Next, note that:

$$\begin{aligned}
\sum_{t=1}^{T}(d_t^*)^2 &= \sum_{t=1}^{\hat{T}_1^*}(d_t^*)^2 + \sum_{\hat{T}_1^*+1}^{T_1^*}(d_t^*)^2 + \sum_{t=T_1^*+1}^{T}(d_t^*)^2 = \left(\frac{\Pi_1}{r_{1T}} - \hat{\Pi}_1\right)' \sum_{t=1}^{\hat{T}_1^*} W_t W_t' \left(\frac{\Pi_1}{r_{1T}} - \hat{\Pi}_1\right) \\
&\quad + \left(\frac{\Pi_1}{r_{1T}} - \hat{\Pi}_2\right)' \sum_{\hat{T}_1^*+1}^{T_1^*} W_t W_t' \left(\frac{\Pi_1}{r_{1T}} - \hat{\Pi}_2\right) + \left(\frac{\Pi_2}{r_{2T}} - \hat{\Pi}_2\right)' \sum_{T_1^*+1}^{T} W_t W_t' \left(\frac{\Pi_2}{r_{2T}} - \hat{\Pi}_2\right) \\
&= \mathcal{O}_P(1) + \mathcal{O}_P(1/R_T)\mathcal{O}_P(T)\mathcal{O}_P(1/R_T) + \mathcal{O}_P(1/R_T)\mathcal{O}_P(T)\mathcal{O}_P(1/R_T) \\
&= \mathcal{O}_P(T/R_T^2).
\end{aligned} \tag{C.10}$$

Equations (C.9)-(C.10) coincide with (C.2), completing Step 1.

Step 2. We start by defining the norm $\|v\|$ as the Euclidean norm for vectors v, and $\|A\|$ as the square root of the maximum eigenvalue of $A'A$ for matrices (hence $\|A\| \leq (\text{trace } A'A)^{1/2}$ and $\|A\|$ is greater than the square root of the minimum eigenvalue (mineig) of $A'A$). If $\hat{\nu}_1^* \overset{p}{\nrightarrow} \nu_1^0$, then there exists $\eta \in (0,1)$, such that with positive probability ϵ, $T_1^* - \hat{T}_1^* = [T\hat{\nu}_1] - [T\nu_1^0] \geq T\eta$. Denote by γ_T the minimum eigenvalue of $A = T\eta^{-1}\sum_{t=[T\nu_1^0]-T\eta+1}^{[T\nu_1^0]} W_t W_t'$. Because the latter is a symmetric pd matrix, $\|A\| \geq [\text{mineig}(A'A)]^{1/2} = [\text{mineig}(A^2)]^{1/2} = \text{mineig}(A) = \gamma_T$. By Assumption 3(i), $\gamma_T > C + o_P(1)$ for some $C > 0$. So with positive probability ϵ,

$$\begin{aligned}
R_T^2/T\sum_{t=1}^{T}(d_t^*)^2 &\geq T^{-1}\sum_{t=T_1^*-T\eta+1}^{T_1^*}(d_t^*)^2 \\
&= R_T\left(\frac{\Pi_1}{r_{1T}}-\hat{\Pi}_2\right)'\left(T^{-1}\sum_{t=[T\nu_1^0]-T\eta+1}^{[T\nu_1^0]}W_tW_t'\right)R_T\left(\frac{\Pi_1}{r_{1T}}-\hat{\Pi}_2\right) \qquad \text{(C.11)} \\
&\geq R_T^2\left\|\frac{\Pi_1}{r_{1T}}-\hat{\Pi}_2\right\|^2\left\|T^{-1}\sum_{t=[T\nu_1^0]-T\eta+1}^{[T\nu_1^0]}W_tW_t'\right\| \\
&\geq \|R_T(\Pi_1/r_{1T}-\Pi_2/r_{2T})+R_T(\Pi_2/r_{2T}-\hat{\Pi}_2)\|^2\eta\gamma_T \\
&\geq \|R_T(\Pi_1/r_{1T}-\Pi_2/r_{2T})+o_P(1)\|^2\eta(C+o_P(1)) \\
&= \|R_T(\Pi_1/r_{1T}-\Pi_2/r_{2T})\|^2\eta C+\mathcal{O}_P(1). \qquad \text{(C.12)}
\end{aligned}$$

Now, if $r_{1T}=r_{2T}=R_T$, then (C.12) implies $R_T^2/T\sum_{t=1}^{T}(d_t^*)^2 \geq \|\Pi_1-\Pi_2\|^2\eta C+\mathcal{O}_P(1)$, so $R_T^2/T\sum_{t=1}^{T}(d_t^*)^2+\mathcal{O}_P(1)>0$. This is also true for $r_{1T}<r_{2T}$ or $r_{1T}>r_{2T}$, because then $R_T^2/T\sum_{t=1}^{T}(d_t^*)^2+\mathcal{O}_P(1)>\|\Pi_1\|^2\eta C>0$ or $R_T^2/T\sum_{t=1}^{T}(d_t^*)^2+\mathcal{O}_P(1)>\|\Pi_2\|^2\eta C>0$. Since this contradicts (C.3), the consistency proof is complete.

Part (ii). From part (i), it follows that any break-point estimator T_1 is such that $T_1^*-T_1\leq \epsilon^*T$, for some chosen $\epsilon^*>0$, different than the ϵ^* defined in part (i). We will prove part (ii) by contradiction as well. To that end, assume that for chosen $C^*>0$, $T_1^*-T_1>C^*R_T^2$. Define SSR_1^*, SSR_2^* and SSR_3^* as the sum of squared residuals in the reduced form obtained with break-point T_1, T_1^* and (T_1,T_1^*) respectively. Then by definition of OLS, with probability one,

$$\min_{T_1 s.t. C^*R_T^2<T_1^*-T_1\leq\epsilon^*T}(SSR_1^*-SSR_2^*)\leq 0.$$

We will show that if $C^*R_T^2<T_1^*-T_1\leq\epsilon^*T$ for some large but fixed C^* and small but fixed ϵ^*, then $\text{plim}(SSR_1^*-SSR_2^*)>0$, contradicting the above. It follows that $T_1^*-T_1\leq C^*R_T^2$, and by symmetry of the argument, if $T_1\geq T_1^*$, $T_1-T_1^*\leq C^*R_T^2$, establishing the desired convergence rate for the break-fraction estimator.

We show that $\text{plim}(SSR_1^*-SSR_2^*)>0$ in two steps. Denote by $(\hat{\Pi}_1,\hat{\Pi}_2)$ the OLS estimators based on sample partition $(1,T_1,T)$, $(\hat{\Pi}_1,\hat{\Pi}_\Delta,\tilde{\Pi}_2)$ the ones based on $(1,T_1,T_1^*,T)$, and

$(\tilde{\Pi}_1, \tilde{\Pi}_2)$ the ones based on $(1, T_1^*, T)$. The first step amounts to showing that:

$$\begin{aligned} SSR_1^* - SSR_3^* &= (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' \left(\sum_{T_1+1}^{T_1^*} W_t W_t' \right) (\tilde{\Pi}_2 - \hat{\Pi}_\Delta) \\ &\quad - (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' \left(\sum_{T_1+1}^{T_1^*} W_t W_t' \right) \left(\sum_{t=T_1+1}^{T} W_t W_t' \right)^{-1} \left(\sum_{T_1+1}^{T_1^*} W_t W_t' \right) (\tilde{\Pi}_2 - \hat{\Pi}_\Delta) \\ &\equiv N_1 - N_2 \end{aligned} \tag{C.13}$$

By symmetry of argument, we then also have:

$$\begin{aligned} SSR_2^* - SSR_3^* &= (\hat{\Pi}_1 - \hat{\Pi}_\Delta)' \left(\sum_{T_1+1}^{T_1^*} W_t W_t' \right) (\hat{\Pi}_1 - \hat{\Pi}_\Delta) \\ &\quad - (\hat{\Pi}_1 - \hat{\Pi}_\Delta)' \left(\sum_{T_1+1}^{T_1^*} W_t W_t' \right) \left(\sum_{t=1}^{T_1^*} W_t W_t' \right)^{-1} \left(\sum_{T_1+1}^{T_1^*} W_t W_t' \right) (\hat{\Pi}_1 - \hat{\Pi}_\Delta) \\ &\equiv N_3 - N_4 \end{aligned}$$

Step one is useful as it helps determine the orders of $N_1 - N_4$ by re-writing the problem in terms of OLS estimators that only use observations from one regime: $\hat{\Pi}_1, \tilde{\Pi}_2, \hat{\Pi}_\Delta$ rather than two regimes: $\hat{\Pi}_2$.[15] Using this, in the second step, we show that N_1 dominates N_2, N_3 for large C^* and small ϵ^*. We also show that $N_1 > 0$ with large probability, and because $N_4 \geq 0$ with probability one in the limit, by Assumption 4(ii), for large C^* and small ϵ^*, we establish the desired result:

$$\text{plim}(SSR_1^* - SSR_2^*) = N_1 - N_2 - N_3 + N_4 \geq N_1 - N_2 - N_3 > 0. \tag{C.14}$$

[15]Note that such a proof is similar to the one showing asymptotic equivalence of F- and Wald-tests.

Step 1. Note that:

$$SSR_1^* - SSR_3^* = \sum_{t=T_1+1}^{T_1^*} [(Y_t - W_t'\hat{\Pi}_2)^2 - (Y_t - W_t'\hat{\Pi}_\Delta)^2] + \sum_{t=T_1^*+1}^{T} [(Y_t - W_t'\hat{\Pi}_2)^2 - (Y_t - W_t'\tilde{\Pi}_2)^2]$$
$$= (\Pi_\Delta - \hat{\Pi}_2)' \left[2\sum_{t=T_1+1}^{T_1^*} W_t v_t + \sum_{t=T_1+1}^{T_1^*} W_t W_t' [(\Pi_1/r_{1T} - \hat{\Pi}_2) + (\Pi_1/r_{1T} - \hat{\Pi}_\Delta)] \right]$$
$$+ (\tilde{\Pi}_2 - \hat{\Pi}_2)' \left[2\sum_{t=T_1^*+1}^{T} W_t v_t + \sum_{t=T_1^*+1}^{T} W_t W_t' [(\Pi_2/r_{2T} - \hat{\Pi}_2) + (\Pi_2/r_{2T} - \tilde{\Pi}_2)] \right] \tag{C.15}$$

For simplicity, let $A = \sum_{t=T_1+1}^{T_1^*} W_t W_t'$ and $B = \sum_{t=T_1^*+1}^{T} W_t W_t'$. By definition of OLS,

$$(A+B)\hat{\Pi}_2 = \sum_{t=T_1+1}^{T} W_t v_t = \sum_{t=T_1+1}^{T_1^*} W_t v_t + \sum_{t=T_1^*+1}^{T} W_t v_t = A\hat{\Pi}_\Delta + B\tilde{\Pi}_2.$$

Thus, we have:

$$\Pi_\Delta - \hat{\Pi}_2 = (A+B)^{-1}[(A+B-A)\hat{\Pi}_\Delta - B\tilde{\Pi}_2] = (A+B)^{-1}B(\hat{\Pi}_\Delta - \tilde{\Pi}_2)$$
$$\tilde{\Pi}_2 - \hat{\Pi}_2 = (A+B)^{-1}[-A\hat{\Pi}_\Delta + (A+B-B)\tilde{\Pi}_2] = (A+B)^{-1}A(\tilde{\Pi}_2 - \hat{\Pi}_\Delta).$$

Substituting this into (C.15) and noting that A, B are symmetric, we obtain:

$$\begin{aligned}
&SSR_1^* - SSR_3^* \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' B(A+B)^{-1}[2A(\hat{\Pi}_\Delta - \Pi_1/r_1T) + A(\Pi_1/r_{1T} - \hat{\Pi}_2) + A(\Pi_1/r_{1T} - \hat{\Pi}_\Delta)] \\
&+ (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' A(A+B)^{-1}[2B(\tilde{\Pi}_2 - \Pi_2/r_1T) + B[(\Pi_2/r_{2T} - \hat{\Pi}_2) + B(\Pi_2/r_{2T} - \tilde{\Pi}_2)] \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' B(A+B)^{-1}A(\hat{\Pi}_\Delta - \hat{\Pi}_2) + (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' A(A+B)^{-1}B(\tilde{\Pi}_2 - \hat{\Pi}_2) \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' B(A+B)^{-1}A(A+B)^{-1}B(\hat{\Pi}_\Delta - \tilde{\Pi}_2) \\
&+ (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' A(A+B)^{-1}B(A+B)^{-1}A(\tilde{\Pi}_2 - \hat{\Pi}_\Delta) \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)'[B(A+B)^{-1}A(A+B)^{-1}B + A(A+B)^{-1}B(A+B)^{-1}A](\tilde{\Pi}_2 - \hat{\Pi}_\Delta).
\end{aligned}$$

To prove (C.13), we are left with showing that:

$$B(A+B)^{-1}A(A+B)^{-1}B + A(A+B)^{-1}B(A+B)^{-1}A = A - A(A+B)^{-1}A. \tag{C.16}$$

This can be shown as follows. Let $V_1 = B(A+B)^{-1}$ and $V_2 = A(A+B)^{-1}$. Then $V_1+V_2 = I$ and $V_1' + V_2' = I$, or $V_1 = I - V_2$, so:

$$\begin{aligned}
&B(A+B)^{-1}A(A+B)^{-1}B + A(A+B)^{-1}B(A+B)^{-1}A \\
&= (I-V_2)A(I-V_2') + V_2BV_2' = A - AV_2' - V_2A + V_2AV_2' + V_2BV_2' \\
&= A - A(A+B)^{-1}A - A(A+B)^{-1}A + A(A+B)^{-1}(A+B)(A+B)^{-1}A = A - A(A+B)^{-1}A.
\end{aligned}$$

This shows the desired results and completes Step 1.

Step 2. Recall:

$$SSR_1^* - SSR_3^* = (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)'A(\tilde{\Pi}_2 - \hat{\Pi}_\Delta) - (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)'A(A+B)^{-1}B(\tilde{\Pi}_2 - \hat{\Pi}_\Delta) = N_1 - N_2.$$

Since $T_1^* - T_1 \leq \epsilon^* T$, by Assumption 4(ii),

$$(A+B)^{-1}A = \left(T^{-1}\sum_{t=T_1+1}^{T} W_tW_t'\right)^{-1}\left(T^{-1}\sum_{T_1+1}^{T_1^*} W_tW_t'\right) = \mathcal{O}_P(1)\mathcal{O}_P(\epsilon^*T/T) = \mathcal{O}_P(\epsilon^*),$$

so N_1 dominates N_2 for ϵ^* small enough. Recall that $N_3 = (\hat{\Pi}_1 - \hat{\Pi}_\Delta)'A(\hat{\Pi}_1 - \hat{\Pi}_\Delta)$. To show that N_1 also dominates N_3, we need to compare the orders of $(\tilde{\Pi}_2 - \hat{\Pi}_\Delta)$ and $(\hat{\Pi}_1 - \hat{\Pi}_\Delta)$. Since $\hat{\Pi}_1$ and $\hat{\Pi}_\Delta$ are both sub-sample estimators of Π_1/r_{1T},

$$\hat{\Pi}_1 - \hat{\Pi}_\Delta = (\hat{\Pi}_1 - \Pi_1/r_{1T}) - (\hat{\Pi}_\Delta - \Pi_1/r_{1T}) = \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(T^{-1/2}) = \mathcal{O}_P(T^{-1/2}).$$

Since $\tilde{\Pi}_2$ estimates Π_2/r_{2T} with all observations from the second regime only,

$$\begin{aligned}
\tilde{\Pi}_2 - \hat{\Pi}_\Delta &= (\tilde{\Pi}_2 - \Pi_2/r_{2T}) - (\hat{\Pi}_\Delta - \Pi_1/r_{1T}) - (\Pi_1/r_{1T} - \Pi_2/r_{2T}) \\
&= \mathcal{O}_P(T^{-1/2}) - \mathcal{O}_P(T^{-1/2}) - (\Pi_1/r_{1T} - \Pi_2/r_{2T}) = \mathcal{O}_P(T^{-1/2}) + \mathcal{O}_P(1/R_T) = \mathcal{O}_P(1/R_T).
\end{aligned}$$

Thus, $(\tilde{\Pi}_2 - \hat{\Pi}_\Delta)$ dominates $(\hat{\Pi}_1 - \hat{\Pi}_\Delta)$, implying that N_1 dominates N_3. To complete Step 2 and the proof, it remains to show that $N_1 > 0$ with large probability for large T, small ϵ^* and large enough C^*. To that end, note:

$$\begin{aligned}
N_1 &= (T_1^* - T_1)[\mathcal{O}_P(T^{-1/2}) - (\Pi_1/r_{1T} - \Pi_2/r_{2T})]'\left(\frac{1}{T_1^* - T_1}A\right)[\mathcal{O}_P(T^{-1/2}) - (\Pi_1/r_{1T} - \Pi_2/r_{2T})] \\
&\geq C^* R_T(\Pi_1/r_{1T} - \Pi_2/r_{2T})'[Q_{WW}(\nu_1^0) - Q_{WW}(\nu_1)]\ R_T(\Pi_1/r_{1T} - \Pi_2/r_{2T}) + o_P(1).
\end{aligned}$$

Now by Assumption 4(ii), where $Q_{WW}(r)$ is pd and strictly increasing in r, $[Q_{WW}(\nu_1^0) - Q_{WW}(\nu_1)]$ is pd, and so $N_1 > 0$ in the limit if $R_T(\Pi_1/r_{1T} - \Pi_2/r_{2T}) \nrightarrow 0$. When $r_{1T} = r_{2T}$, $R_T(\Pi_1/r_{1T} - \Pi_2/r_{2T}) = \Pi_1 - \Pi_2 \neq 0$ by construction. Similarly, if $r_{1T} < r_{2T}$, $R_T(\Pi_1/r_{1T} - \Pi_2/r_{2T}) \to \Pi_1 \neq 0$, while if $r_{1T} > r_{2T}$, $R_T(\Pi_1/r_{1T} - \Pi_2/r_{2T}) \to -\Pi_2 \neq 0$.[16]

Part (iii). To derive the full-sample distribution of the 2SLS estimator in Theorem B.1, let:

$$\hat{W} = \begin{pmatrix} \hat{W}_1 & O_{\hat{T}_1^* \times q} \\ O_{(T-\hat{T}_1^*)\times q} & \hat{W}_2 \end{pmatrix} \qquad \text{and} \qquad \bar{W} = \begin{pmatrix} W_1 & O_{T_1^* \times q} \\ O_{(T-T_1^*)\times q} & W_2 \end{pmatrix}$$

where $\hat{W}_1$ is the $\hat{T}_1^* \times q$ matrix with rows $W_1', \ldots, W_{\hat{T}_1^*}'$, $\hat{W}_2$ is the $(T - \hat{T}_1^*) \times q$ matrix with rows $W_{\hat{T}_1^*+1}', \ldots, W_T'$, W_1 is the $T_1^* \times q$ with rows $W_1', \ldots, W_{T_1^*}'$, W_2 is the $(T - T_1^*) \times q$ matrix with rows $W_{T_1^*+1}', \ldots, W_T'$. Then $\hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_T)'$ can be written as $\hat{Y} = \hat{W}\hat{\Pi}$, with $\hat{\Pi} = (\hat{\Pi}_1', \hat{\Pi}_2')'$. Denoting $y = (y_1, \ldots, y_T)'$, $U = (u_1, \ldots, u_T)'$, and $Y = (Y_1, \ldots, Y_T)'$, the 2SLS estimator is:

$$\hat{\theta} = (\hat{Y}'\hat{Y})^{-1}\hat{Y}'y = (\hat{Y}'\hat{Y})^{-1}\hat{Y}(\hat{Y}'\theta^0 + (Y - \hat{Y})'\theta^0 + U) = \theta^0 + (\hat{Y}'\hat{Y})^{-1}\hat{Y}'\tilde{U},$$

with $\tilde{U} = (Y - \hat{Y})\theta^0 + U$. It follows that:

$$T^{1/2}R_T^{-1}(\hat{\theta}_y - \theta_y^0) = (R_T^2 T^{-1}\hat{Y}'\hat{Y})^{-1}R_T T^{-1/2}\hat{Y}'\tilde{U}. \tag{C.17}$$

We will first show that:

$$R_T^2 T^{-1}\hat{Y}'\hat{Y} = R_T^2(\Pi_1'/r_{1T}Q_1\Pi_1/r_{1T} + \Pi_2'/r_{1T}Q_2\Pi_2/r_{1T}) + o_P(1) \tag{C.18}$$

$$R_T T^{-1/2}\hat{Y}'\tilde{U} = R_T T^{-1/2}\left(\Pi_1'/r_{1T}\sum_{t=1}^{T_1^*} W_t u_t + \Pi_2'/r_{2T}\sum_{t=T_1^*+1}^{T} W_t u_t\right) + o_P(1). \tag{C.19}$$

Without loss of generality, assume $r_{1T} = o(r_{2T})$ or $r_{1T} = r_{2T}$. Then, in the next step, we derive the asymptotic limit of $R_T^2(\Pi_1'/r_{1T}Q_1\Pi_1/r_{1T} + \Pi_2'/r_{1T}Q_2\Pi_2/r_{1T})$, respectively $R_T T^{-1/2}\left(\Pi_1'/r_{1T}\sum_{t=1}^{T_1^*} W_t u_t + \Pi_2'/r_{2T}\sum_{t=T_1^*+1}^{T} W_t u_t\right)$ in both cases, completing the proof of Theorem B.1.

[16] Note that the case $R_T = \sqrt(T)$ is not considered, because since part (i) doesn't hold, neither does the rest of the theorem, which is derived assuming part (i) holds.

Consider first $R_T^2 T^{-1}\hat{Y}'\hat{Y} = R_T^2\hat{\Pi}'(T^{-1}\hat{W}'\hat{W})\hat{\Pi}$. By Theorem B.1(ii) and Assumption 4,

$$T^{-1}\sum_{\hat{T}_1^*+1}^{T_1^*} W_t W_t' = \frac{T_1^* - \hat{T}_1^*}{T}\left(\frac{1}{T_1^* - \hat{T}_1^*}\sum_{\hat{T}_1^*+1}^{T_1^*} W_t W_t'\right) = \mathcal{O}_P(R_T^2/T)\mathcal{O}_P(1) = o_P(1),$$

so:

$$T^{-1}\hat{W}'\hat{W} - T^{-1}\bar{W}'\bar{W} = -T^{-1}\sum_{\hat{T}_1^*+1}^{T_1^*} W_t W_t' = o_P(1). \tag{C.20}$$

By Assumption 5(i), $T^{-1}\bar{W}'\bar{W} \xrightarrow{p} Q$. On the other hand, by Assumptions 1-2 and the functional CLT in Wooldridge and White (1988), Theorem 2.11,

$$\begin{aligned} T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t v_t &= T^{-1/2}(T_1^* - \hat{T}_1^*)^{1/2}\left[(T_1^* - \hat{T}_1^*)^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t v_t\right] = \mathcal{O}_P(R_T/T^{1/2})\mathcal{O}_P(1) \\ &= o_P(1), \end{aligned}$$

implying that:

$$T^{-1/2}\sum_{t=1}^{\hat{T}_1^*} W_t v_t - T^{-1/2}\sum_{t=1}^{T_1^*} W_t v_t = -T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t v_t = o_P(1)$$

$$T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T} W_t v_t - T^{-1/2}\sum_{t=T_1^*+1}^{T} W_t v_t = T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t v_t = o_P(1).$$

From (C.20) and the last two equations,

$$\begin{aligned} T^{1/2}(\hat{\Pi}_1 - \Pi_1/r_{1T}) &= (T^{-1}\hat{W}_1'\hat{W}_1)^{-1}T^{-1/2}\sum_{t=1}^{\hat{T}_1^*} W_t v_t \\ &= (T^{-1}\bar{W}_1'\bar{W}_1)^{-1}T^{-1/2}\sum_{t=1}^{T_1^*} W_t v_t + o_P(1) \\ &= \mathcal{O}_P(1)\mathcal{O}_P(1) + o_P(1) = \mathcal{O}_P(1) \end{aligned}$$

$$T^{1/2}(\hat{\Pi}_2 - \Pi_2/r_{1T}) = (T^{-1}\hat{W}_2'\hat{W}_2)^{-1}\left(T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T} W_t v_t\right) +$$

$$+ (T^{-1}\hat{W}_2'\hat{W}_2)^{-1}\left(T^{-1}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t W_t'\right) T^{1/2}(\Pi_1/r_{1T} - \Pi_2/r_{2T})$$

$$= (T^{-1}\bar{W}_2'\bar{W}_2)^{-1}\left(T^{-1/2}\sum_{t=T_1^*+1}^{T} W_t v_t\right) + \mathcal{O}_P(1)\mathcal{O}_P(R_T^2/T)\mathcal{O}_P(T^{1/2}/R_T)$$

$$+ o_P(1) = \mathcal{O}_P(1) + \mathcal{O}_P(T^{-1/2}R_T) + o_P(1) = \mathcal{O}_P(1) + o_P(1) = \mathcal{O}_P(1).$$

It follows that $\hat{\Pi}_1 = \Pi_1/r_{1T} + o_P(1)$, $\hat{\Pi}_2 = \Pi_2/r_{2T} + o_P(1)$, and together with (C.20), these imply:

$$R_T^2 T^{-1}\hat{Y}'\hat{Y} = R_T^2(T^{-1}\sum_{t=1}^{T_1^*}\hat{Y}_t^2 + T^{-1}\sum_{t=T_1^*+1}^{T}\hat{Y}_t^2) + o_P(1)$$

$$= R_T^2(\Pi_1'/r_{1T}\left(T^{-1}\sum_{t=1}^{T_1^*} W_t W_t'\right)\Pi_1/r_{1T} + \Pi_2'/r_{1T}\left(T^{-1}\sum_{t=T_1^*+1}^{T} W_t W_t'\right)\Pi_2/r_{2T}) + o_P(1)$$

$$= R_T^2(\Pi_1'/r_{1T}Q_1\Pi_1/r_{1T} + \Pi_2'/r_{2T}Q_2\Pi_2/r_{2T}) + o_P(1).$$

The latter proves (C.18). Next, we show (C.19). By the above,

$$R_T T^{-1/2}\hat{Y}'\tilde{U} = R_T\hat{\Pi}'T^{-1/2}\hat{W}'\tilde{U} = R_T[(\Pi_1'/r_{1T}, \Pi_2'/r_{2T})' + o_P(1)]T^{-1/2}\hat{W}\tilde{U}. \qquad \text{(C.21)}$$

First, $T^{-1/2}\hat{W}\tilde{U} - T^{-1/2}\bar{W}\tilde{U} = -T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t\tilde{u}_t$. Since

$$\tilde{u}_t = u_t + (Y_t - \hat{Y}_t)\theta_y^0 = \begin{cases} u_t + v_t\theta_y^0 - W_t'(\hat{\Pi}_1 - \Pi_1/r_{1T})\theta_y^0, & t \le \hat{T}_1^* \\ u_t + v_t\theta_y^0 - W_t'(\hat{\Pi}_2 - \Pi_1/r_{1T})\theta_y^0, & \hat{T}_1^* + 1 \le t \le T_1^* \\ u_t + v_t\theta_y^0 - W_t'(\hat{\Pi}_2 - \Pi_2/r_{2T})\theta_y^0, & t > T_1^* \end{cases},$$

$$
\begin{aligned}
T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t\tilde{u}_t &= T^{-1/2}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t(u_t+v_t\theta_y^0) - \left(T^{-1}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_tW_t'\right) T^{1/2}(\hat{\Pi}_2-\Pi_1/r_{1T})\theta_y^0 \\
&= \frac{(T_1^*-\hat{T}_1^*)^{1/2}}{T^{1/2}}\left(\frac{1}{(T_1^*-\hat{T}_1^*)^{1/2}}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_t(u_t+v_t\theta_y^0)\right) - \\
&- \left(T^{-1}\sum_{t=\hat{T}_1^*+1}^{T_1^*} W_tW_t'\right)[T^{1/2}(\hat{\Pi}_2-\Pi_2/r_{2T})+T^{1/2}(\Pi_2/r_{1T}-\Pi_1/r_{1T})]\theta_y^0 \\
&= \mathcal{O}_P(R_T/T^{1/2}) - \mathcal{O}_P(R_T^2/T)[\mathcal{O}_P(1)+\mathcal{O}_P(T^{1/2}/R_T)] \\
&= o_P(1) - \mathcal{O}_P(R_T/T^{1/2}) = o_P(1).
\end{aligned}
$$

Hence,

$$
T^{-1/2}\hat{W}'\tilde{U} = T^{-1/2}\bar{W}'\tilde{U} + o_P(1). \qquad \text{(C.22)}
$$

Now, $\bar{W}'\tilde{U} = (\sum_{t=1}^{T_1^*} W_t'\tilde{u}_t, \sum_{t=T_1^*+1}^{T} W_t'\tilde{u}_t)'$, and

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T_1^*} W_t\tilde{u}_t &= T^{-1/2}\sum_{t=1}^{T_1^*} W_t(u_t+v_t\theta_y^0) - \left(T^{-1}\sum_{t=1}^{T_1^*} W_tW_t'\right)[T^{1/2}(\hat{\Pi}_1-\Pi_1/r_{1T})]\theta_y^0 \\
&= T^{-1/2}\sum_{t=1}^{T_1^*} W_t(u_t+v_t\theta_y^0) - T^{-1/2}\sum_{t=1}^{T_1^*} W_tv_t\theta_y^0 \\
&= T^{-1/2}\sum_{t=1}^{T_1^*} W_tu_t.
\end{aligned}
$$

Similarly, $T^{-1/2}\sum_{t=T_1^*+1}^{T} W_t\tilde{u}_t = T^{-1/2}\sum_{t=T_1^*+1}^{T} W_tu_t$, hence $T^{-1/2}\hat{W}'\tilde{U} = T^{-1/2}\bar{W}'U + o_P(1)$. From this and (C.21)-(C.22), it follows that:

$$
\begin{aligned}
R_TT^{-1/2}\hat{Y}'\tilde{U} &= R_T[(\Pi_1/r_{1T},\Pi_2/r_{2T})']T^{-1/2}\bar{W}'U + o_P(1) \\
&= R_T\left[\Pi_1'/r_{1T}\left(T^{-1/2}\sum_{t=1}^{T_1^*} W_tu_t\right) + \Pi_2'/r_{2T}\left(T^{-1/2}\sum_{t=T_1^*+1}^{T} W_tu_t\right)\right] + o_P(1),
\end{aligned}
$$

which coincides with (C.19). Substituting (C.18)-(C.19) into (C.17), we obtain:

$$T^{1/2}R_T^{-1}(\hat{\theta}_y-\theta_y^0)=[R_T^2(\Pi_1'/r_{1T}Q_1\Pi_1/r_{1T}+\Pi_2'/r_{2T}Q_2\Pi_2/r_{2T})]^{-1}\times$$
$$\times R_T\left[\Pi_1'/r_{1T}\left(T^{-1/2}\sum_{t=1}^{T_1^*}W_tu_t\right)+\Pi_2'/r_{2T}\left(T^{-1/2}\sum_{t=T_1^*+1}^{T}W_tu_t\right)\right]+o_P(1). \tag{C.23}$$

Case (a): $r_{1T}=r_{2T}$. In this case, $R_T=r_{1T}=r_{2T}$, and (C.23) becomes (Assumption 3 is employed to ensure invertibility):

$$T^{1/2}R_T^{-1}(\hat{\theta}_y-\theta_y^0)=[\Pi_1'Q_1\Pi_1+\Pi_2'Q_2\Pi_2]^{-1}\left[\Pi_1'\left(T^{-1/2}\sum_{t=1}^{T_1^*}W_tu_t\right)+\Pi_2'\left(T^{-1/2}\sum_{t=T_1^*+1}^{T}W_tu_t\right)\right]$$
$$+o_P(1)=B^{-1}\Pi'T^{-1/2}\bar{W}'U+o_P(1). \tag{C.24}$$

Note that $T^{-1/2}\bar{W}'U=(T^{-1/2}\sum_{t=1}^{T_1^*}W_t'u_t,T^{-1/2}\sum_{t=T_1^*+1}^{T}W_t'u_t)'$, and that under Assumptions 2-5(i), $T^{-1/2}\bar{W}'U\xrightarrow{d}\mathcal{N}\left(0,\begin{pmatrix}V_{u,1}^* & O_{q\times q}\\ O_{q\times q} & V_{u,2}^*\end{pmatrix}\right)$, where the block-diagonality of the asymptotic variance follows from asymptotic independence of sub-samples $[1,T_1^*]$ and $[T_1^*+1,T]$, by Assumption 2. Thus,

$$\Pi'T^{-1/2}\bar{W}'U\xrightarrow{d}\mathcal{N}(0,\Pi_1'V_{u,1}^*\Pi_1+\Pi_2'V_{u,1}^*\Pi_2)=\mathcal{N}(0,V),$$

so the desired result follows:

$$T^{1/2}R_T^{-1}(\hat{\theta})_y-\theta_y^0)\xrightarrow{d}\mathcal{N}(0,B^{-1}VB^{-1}).$$

Case (b): $r_{1T}=o(r_{2T})$. In this case, $R_T=r_{1T}$, $R_T(\Pi_1/r_{1T},\Pi_2/r_{2T})=(\Pi_1,O_{q\times 1})$, and the full sample 2SLS estimator is asymptotically equivalent to the 2SLS estimator obtained using only the first sub-sample:

$$\begin{aligned}T^{1/2}R_T^{-1}(\hat{\theta}_y-\theta_y^0)&=[(\Pi_1',O_{1\times q})'Q(\Pi,O_{q\times 1})]^{-1}(\Pi_1',O_{1\times q})'T^{-1/2}\bar{W}'U+o_P(1)\\&=[\Pi_1'Q_1\Pi_1]^{-1}\Pi_1'\left(T^{-1/2}\sum_{t=1}^{\hat{T}_1^*}W_tu_t\right)+o_P(1)\\&\xrightarrow{d}B_1^{-1}\mathcal{N}(0,\Pi_1'V_{u,1}^*\Pi_1)=\mathcal{N}(0,B_1^{-1}V_1B_1^{-1}).\end{aligned}$$

This is the distribution stated in Theorem 3.1, completing its proof.■

- Proof of Theorem B.2 *(GMM estimation)*

Letting Σ^{-1} be the weighting matrix, the full-sample GMM estimator using this weighting matrix is as usual:

$$\begin{aligned} \hat{\theta}_{GMM,y} &= \left[\frac{Y'W}{T}\Sigma^{-1}\frac{W'Y}{T}\right]^{-1}\frac{Y'W}{T}\Sigma^{-1}\frac{W'}{T}(Y\theta_y^0 + U) \\ \Leftrightarrow \quad \left(\hat{\theta}_{GMM,y} - \theta_y^0\right) &= \left[\frac{Y'W}{T}\Sigma^{-1}\frac{W'Y}{T}\right]^{-1}\frac{Y'W}{T}\Sigma^{-1}\frac{W'U}{T} \end{aligned}$$

Without loss of generality, assume sub-sample $i = 1$ is at least as strong as sub-sample $j = 2$, i.e. $r_{1T} = o(r_{2T}) = R_T$. Then, using the definitions of W_1, W_2 in the proof of Theorem B.1(iii), and letting $V_1 = (v_1, \ldots, v_{T_1^*})'$, respectively $V_2 = (v_{T_1^*+1}, \ldots, v_T)'$, we get:

$$\begin{aligned} \frac{Y'W}{T} &= \frac{Y_1'W_1}{T} + \frac{Y_2'W_2}{T} \\ &= \left[\frac{\Pi_1'}{r_{1T}}\frac{W_1'W_1}{T} + \frac{V_1'W_1}{T}\right] + \left[\frac{\Pi_2'}{r_{2T}}\frac{W_2'W_2}{T} + \frac{V_2'W_2}{T}\right] \\ &= \frac{1}{r_{1T}}\left[\Pi_1'\frac{W_1'W_1}{T} + \frac{r_{1T}}{\sqrt{T}}\frac{V_1'W_1}{\sqrt{T}} + \frac{r_{1T}}{r_{2T}}\Pi_2'\frac{W_2'W_2}{T} + \frac{r_{1T}}{\sqrt{T}}\frac{V_2'W_2}{\sqrt{T}}\right] \\ &= \begin{cases} \frac{1}{r_{1T}}\left[\Pi_1'\frac{W_1'W_1}{T} + o_P(1)\right] & \text{if } r_{1T} = o(r_{2T}) \\ \frac{1}{r_{1T}}\left[\Pi_1'\frac{W_1'W_1}{T} + \Pi_2'\frac{W_2'W_2}{T} + o_P(1)\right] & \text{if } r_{1T} = r_{2T} \end{cases} \\ &= \begin{cases} \frac{1}{r_{1T}}\left[\Pi_1'Q_1 + o_P(1)\right] & \text{if } r_{1T} = o(r_{2T}) \\ \frac{1}{r_{1T}}\left[\Pi_1'Q_1 + \Pi_2'Q_2 + o_P(1)\right] & \text{if } r_{1T} = r_{2T} \end{cases}. \end{aligned}$$

since $V_1'W_1/\sqrt{T} = \mathcal{O}_P(1)$, $r_{1T}/\sqrt{T} = o(1)$.

Case (a): $r_{1T} = r_{2T}$. Then $R_T = r_{1T} = r_{2T}$, and

$$\left[\frac{Y'W}{T}\Sigma^{-1}\frac{W'Y}{T}\right] = \frac{1}{R_T^2}\left(\Pi_1'Q_1 + \Pi_2'Q_2\right)\Sigma^{-1}\left(Q_1\Pi_1 + Q_2\Pi_2\right) + o_P(1) = \frac{1}{R_T^2}\tilde{B} + o_P(1),$$

so:

$$\frac{\sqrt{T}}{R_T}\left(\hat{\theta}_{GMM,y} - \theta_y^0\right) = \tilde{B}^{-1}\left(\Pi_1'Q_1 + \Pi_2'Q_2\right)\Sigma^{-1}\frac{W'\tilde{U}}{\sqrt{T}} + o_P(1).$$

From the proof of Theorem B.1(iii), $T^{-1/2}W'U \xrightarrow{d} \mathcal{N}(0, V^*_{u,1} + V^*_{u,2})$, so

$$\begin{aligned}\frac{\sqrt{T}}{R_T}\left(\hat{\theta}_{GMM,y} - \theta^0_y\right) &\xrightarrow{d} \mathcal{N}\left(0, \tilde{B}^{-1}(\Pi_1' Q_1 + \Pi_2' Q_2)\Sigma^{-1}(V^*_{u,1} + V^*_{u,2})\Sigma^{-1}(Q_1\Pi_1 + Q_2\Pi_2)\tilde{B}^{-1}\right) \\ &= \mathcal{N}\left(0, \tilde{B}^{-1}\tilde{V}\tilde{B}^{-1}\right).\end{aligned}$$

Case (b): $r_{1T} = o(r_{2T}) = R_T$. We have:

$$\begin{aligned}\left[\frac{Y'W}{T}\Sigma^{-1}\frac{W'Y}{T}\right] &= \frac{1}{r^2_{1T}}\left[\Pi_1'\frac{W_1'W_1}{T}\Sigma^{-1}\frac{W_1W_1'}{T}\Pi_1 + o_P(1)\right] \\ &= \frac{1}{r^2_{1T}}\left[\Pi_1' Q_1\Sigma^{-1}Q_1\Pi_1 + o_P(1)\right].\end{aligned}$$

It follows that the GMM estimator is equivalent to the estimator obtained from the strongest sub-sample, in this case $[1, T_1^*]$, and:

$$T^{1/2}R_T^{-1}(\hat{\theta}_{GMM,y} - \theta^0_y) = \left[\Pi_1' Q_1\Sigma^{-1}Q_1\Pi_1\right]^{-1}\Pi_1' Q_1\Sigma^{-1}\frac{\sum_{t=1}^{T_1^*} W_t u_t}{\sqrt{T}}.$$

From the proof of Theorem B.1(iii), $T^{-1/2}\sum_{t=1}^{T_1^*} W_t u_t \xrightarrow{d} \mathcal{N}(0, V^*_{u,1})$, so

$$T^{1/2}R_T^{-1}(\hat{\theta}_{GMM,y} - \theta^0_y) \xrightarrow{d} \mathcal{N}(0, \left[\Pi_1' Q_1\Sigma^{-1}Q_1\Pi_1\right]^{-1}\left[\Pi' Q_1\Sigma^{-1}V^*_{u,1}\Sigma^{-1}Q_1\Pi\right]\left[\Pi_1' Q_1\Sigma^{-1}Q_1\Pi_1\right]^{-1}). \blacksquare$$

- Proof of Theorem B.3 *(Efficiency of GMM vs. 2SLS)*

Case (a): $R_T = r_{1T} = r_{2T}$. Then:

$$V_{GMM} = \tilde{B}^{-1}\tilde{V}\tilde{B}^{-1} \text{ and } V_{2SLS} = B^{-1}VB^{-1},$$

with:

$$\begin{aligned}\tilde{B} &= (\Pi_1' Q_1 + \Pi_2' Q_2)\Sigma^{-1}(Q_1\Pi_1 + Q_2\Pi_2) \\ B &= \Pi_1' Q_1\Pi_1 + \Pi_2' Q_2\Pi_2 \\ \tilde{V} &= (\Pi_1' Q_1 + \Pi_2' Q_2)\Sigma^{-1}(V^*_{u,1} + V^*_{u,2})\Sigma^{-1}(Q_1\Pi_1 + Q_2\Pi_2) \\ V &= \Pi_1' V^*_{u,1}\Pi_1 + \Pi_2' V^*_{u,2}\Pi_2.\end{aligned}$$

Under conditional homoskedasticity, $V^*_{u,i} = \sigma^2_u Q_i$ for $i = 1, 2$. Hence, setting as usual $\Sigma = (V^*_{u,1} + V^*_{u,2}) = \sigma^2_u(Q_1 + Q_2)$, we obtain different variances for the GMM and 2SLS estimators, indicating that the 2SLS estimator is no longer equivalent to the GMM estimator under conditional homoskedasticity (and no autocorrelation):

$$\begin{aligned} V_{GMM} &= \sigma_u^2[(\Pi_1' Q_1 + \Pi_2' Q_2)(Q_1+Q_2)^{-1}(Q_1\Pi_1 + Q_2\Pi_2)]^{-1} \\ V_{2SLS} &= \sigma_u^2[\Pi_1' Q_1 \Pi_1 + \Pi_2' Q_2 \Pi_2]^{-1}. \end{aligned}$$

In general, the two don't seem to be comparable. Consider the case $q = 1$, implying that $\Pi_i, Q_i (i = 1, 2)$ are scalars. Then:

$$\begin{aligned} V_{GMM} &= \sigma_u^2[(\Pi_1' Q_1 + \Pi_2' Q_2)(Q_1+Q_2)^{-1}(Q_1\Pi_1 + Q_2\Pi_2)]^{-1} = \sigma_u^2 \frac{Q_1+Q_2}{(\Pi_1 Q_1 + \Pi_2 Q_2)^2} \\ V_{2SLS} &= \sigma_u^2[\Pi_1' Q_1 \Pi_1 + \Pi_2' Q_2 \Pi_2]^{-1} \equiv \sigma_u^2 \frac{1}{\Pi_1^2 Q_1 + \Pi_2^2 Q_2}. \end{aligned}$$

In the special case of $Q_1 = Q_2$, if we let wlog $\Pi_2 = \beta \Pi_1$, then:

$$\begin{aligned} V_{GMM} &= \sigma_u^2 \frac{2Q_1}{\Pi_1^2 Q_1^2 (1+\beta)^2} = \frac{\sigma_u^2}{\Pi_1^2 Q_1} \frac{2}{(1+\beta)^2} \\ V_{2SLS} &= \sigma_u^2 \frac{1}{(1+\beta^2)\Pi_1^2 Q_1} = \frac{\sigma_u^2}{\Pi_1^2 Q_1} \frac{1}{(1+\beta^2)}. \end{aligned}$$

It can be shown that $\frac{2}{(1+\beta)^2} > \frac{1}{(1+\beta^2)}$, and because $\beta \neq 1$ by definition of the RF break, it follows that $V_{GMM} > V_{2SLS}$, and the 2SLS estimator is more efficient. However, $Q_1 = Q_2 = Q_0$ is an arbitrary restriction, in the sense that under the usual Assumption 5(ii), it only holds for $\nu^0 = 0.5$. In general, either estimator can be more efficient than the other depending on the magnitude and location of the break.

Case (b). $r_{1T} = o(r_{2T})$. Set $\Sigma = V_{u,1}^*$, and optimal GMM estimator is:

$$T^{1/2} R_T^{-1}(\hat{\theta}_{GMM,y} - \theta_y^0) \xrightarrow{d} \mathcal{N}(0, \left[\Pi_1' Q_1 V_{u,1}^* Q_1 \Pi_1\right]^{-1}).$$

Thus, by usual arguments, it is more efficient than the 2SLS estimator, which is obtained with $\Sigma = Q_1$.

- Proof of Theorem B.4 *(Test for 0 vs. 1 break in RF)*

Start by noting that, for ι_2 a 2×1 vector of ones,

$$RT^{1/2}\hat{\Pi}(\nu) = R \begin{pmatrix} T^{1/2}(\hat{\Pi}_1(\nu) - \Pi_T^0) \\ T^{1/2}(\hat{\Pi}_2(\nu) - \Pi_T^0) \end{pmatrix} = RT^{1/2}(\hat{\Pi}(\nu) - \iota_2 \otimes \Pi_T^0).$$

Thus, the asymptotic distribution of the Wald test is determined by that of $T^{1/2}(\hat{\Pi}_1(\nu) - \Pi_T^0)$ and $T^{1/2}(\hat{\Pi}_1(\nu) - \Pi_T^0)$. Under the null hypothesis,

$$T^{1/2}(\hat{\Pi}_1(\nu) - \Pi_T^0) = \left(T^{-1} \sum_{t=1}^{[T\nu]} W_t W_t' \right)^{-1} \left(T^{-1/2} \sum_{t=1}^{[T\nu]} W_t v_t \right).$$

Under Assumptions 4-5, $T^{-1} \sum_{t=1}^{[T\nu]} W_t W_t' \xrightarrow{p} \nu Q_{WW}$. Under Assumptions 2 and 5, by the FCLT, $T^{-1/2} \sum_{t=1}^{[T\nu]} W_t v_t \Rightarrow (V^*)^{1/2} B_q(\nu)$. Hence,

$$T^{1/2}(\hat{\Pi}_1(\nu) - \Pi_T^0) \Rightarrow [Q_{WW}^{-1}(V^*)^{1/2}] B_q(\nu)/\nu = G^{1/2} B_q(\nu)/\nu$$

By similar arguments,

$$T^{1/2}(\hat{\Pi}_2(\nu) - \Pi_T^0) \Rightarrow [Q_{WW}^{-1}(V^*)^{1/2}][B_q(1) - B_q(\nu)]/(1-\nu) = G^{1/2}[B_q(1) - B_q(\nu)]/(1-\nu).$$

Let ι_2 be a 2×1 vector of ones, and $C = \begin{pmatrix} \nu & 0 \\ 0 & 1-\nu \end{pmatrix}$. Then the above implies:

$$T^{1/2}(\hat{\Pi}(\nu) - \iota_2 \otimes \Pi_T^0) \Rightarrow (C^{-1} \otimes G^{1/2}) \begin{pmatrix} B_q(\nu) \\ B_q(1) - B_q(\nu) \end{pmatrix} \equiv (C^{-1} \otimes G^{1/2}) \bar{B}(\nu). \quad \text{(C.25)}$$

Also,

$$\begin{aligned} \mathbf{Avar}[T^{1/2}(\hat{\Pi}(\nu) - \iota_2 \otimes \Pi_T^0)] &= (C^{-1} \otimes G^{1/2})\ (C \otimes I_q)\ (C^{-1} \otimes G^{1/2\prime}) \\ &= (C^{-1} C C^{-1})\,(G^{1/2} G^{1/2\prime}) = C^{-1} \otimes G. \end{aligned} \quad \text{(C.26)}$$

Using (C.25)-(C.26) and letting $r = (1, -1)'$, such that $R' = r \otimes I_q$, we obtain:

$$\begin{aligned} &Sup - Wald_T(0:1) = \sup_{\nu \in \Lambda_\epsilon} \left[T \hat{\Pi}'(\nu) R' [R \hat{G}_T R']^{-1} R \hat{\Pi}'(\nu) \right] \\ \Rightarrow &\sup_{\nu \in \Lambda_\epsilon} \bar{B}'(\nu)\ (C^{-1} \otimes G^{1/2\prime})(r \otimes I_q) \left[(r' \otimes I_q)(C^{-1} \otimes G^{1/2})(r \otimes I_q) \right]^{-1} (r' \otimes I_q)(C^{-1} \otimes G^{1/2})\ \bar{B}(\nu) \\ = &\sup_{\nu \in \Lambda_\epsilon} \bar{B}'(\nu)\ [(C^{-1} r) \otimes G^{1/2\prime}]\ [(r' C^{-1} r) \otimes G]^{-1} [(r' C^{-1}) \otimes G^{1/2}]\ \bar{B}(\nu) \\ = &\sup_{\nu \in \Lambda_\epsilon} \bar{B}'(\nu)\ [(C^{-1} r) \otimes G^{1/2\prime}]\ [(r' C^{-1} r)^{-1} \otimes G^{-1}]\ [(r' C^{-1}) \otimes G^{1/2}]\ \bar{B}(\nu) \\ = &\sup_{\nu \in \Lambda_\epsilon} \bar{B}'(\nu)\ \left\{ [C^{-1} r (r' C^{-1} r)^{-1} r' C^{-1}] \otimes [G^{1/2\prime} G^{-1} G^{1/2}] \right\}\ \bar{B}(\nu) \\ = &\sup_{\nu \in \Lambda_\epsilon} \bar{B}'(\nu)\ \left\{ [C^{-1} r (r' C^{-1} r)^{-1} r' C^{-1}] \otimes I_q \right\}\ \bar{B}(\nu). \end{aligned}$$

It can be shown that $C^{-1}r(r'C^{-1}r)^{-1}r'C^{-1} = \frac{1}{\nu(1-\nu)}\begin{pmatrix} (1-\nu)^2 & -\nu(1-\nu) \\ -\nu(1-\nu) & \nu^2 \end{pmatrix}$, implying the desired result:

$$\begin{aligned} \bar{B}'(\nu)\left\{[C^{-1}r(r'C^{-1}r)^{-1}r'C^{-1}] \otimes I_q\right\}\bar{B}(\nu) &= \frac{\|(1-\nu)B_q(\nu) - \nu(B_q(1) - B_q(\nu))\|^2}{\nu(1-\nu)} \\ &= \frac{\|B_q(\nu) - \nu B_q(1)\|^2}{\nu(1-\nu)}. \blacksquare \end{aligned}$$