

**EFFICIENT INFERENCE WITH TIME-VARYING
IDENTIFICATION STRENGTH**

By

Bertille Antoine and Otilia Boldea

June 18, 2014

RESEARCH INSTITUTE FOR ECONOMETRICS

DISCUSSION PAPER NO. 5-14



Research Institute for Econometrics

מכון מחקר לאקונומטריקה

DEPARTMENT OF ECONOMICS

BAR-ILAN UNIVERSITY

RAMAT-GAN 5290002, ISRAEL

<http://econ.biu.ac.il/en/node/2473>

Efficient Inference with Time-Varying Identification Strength*

Bertille Antoine
Simon Fraser University

Bertille_Antoine@sfu.ca

Otilia Boldea
Tilburg University

O.Boldea@uvt.nl

June 18, 2014

Abstract

In the last two decades, there has been a lot of empirical evidence suggesting that many macroeconomic and financial models (e.g. for inflation, interest rates, or exchange rates) are subject to both parameter instability and identification problems. In this paper, we address both issues in a unified framework, and provide a comprehensive treatment of the link between them. Changes in identification strength provide an additional source of information that is used to improve estimation. More generally, we show that detecting and locating changes in instrument strength is essential for efficient asymptotic inference, and we provide a step-by-step guide for practitioners. In our simulation studies, our global inference procedures show very good size and power properties.

Keywords: GMM; Identification; Weak instruments; Break point; Change in identification strength.

JEL classification: C13, C22, C26, C36, C51.

*We are grateful for useful comments to Jaap Abbring, Jörg Breitung, Bin Chen, Carolina Caetano, Xu Cheng, Valentina Corradi, Frank Diebold, Herman van Dijk, Dennis Fok, Alastair Hall, Lynda Khalaf, Maral Kichian, Frank Kleibergen, Sophocles Mavroeidis, Adam McCloskey, Nour Meddahi, Ulrich Müller, Serena Ng, Eric Renault, Bernard Salanié, Frank Schorfheide, Bas Werker; to all participants at the seminars at Boston College, Brown U, Columbia U, Emory, Erasmus U Rotterdam, Georgia State U, Guelph U, LSE, Rochester U, TI Amsterdam, Tilburg U, TSE, UPenn, U Washington, Western U, and at the conferences: EC2 (Maastricht, 2012), CESG (Kingston, 2012), ESEM (Malaga, 2012; Göteborg, 2013), NESG (Amsterdam, 2013), CIREQ (Montreal, 2014). Otilia Boldea acknowledges NWO VENI grant 415-11-001, and the hospitality of the UPenn Economics Department, where part of this research was conducted.

1 Introduction

Early work on identification through heteroskedasticity by Rigobon (2003) and Klein and Vella (2010), among others, shows that if a p -vector of parameters θ^0 is not identified by q full-sample moment conditions ($q < p$), but the variance of (structural) errors changes over the sample at a known break point, such information can be used to construct $2q$ valid and non-redundant moment conditions.¹ θ^0 can then be identified whenever $2q \geq p$.

In this paper, we focus on using these extra moment conditions for efficient estimation of the parameters of interest - rather than identification. We show that not only changes in the variance of structural errors, but, more generally, changes in the derivative of the moment conditions originating from changes in the parameters of the reduced form, changes in the identification strength, or in some other second moment matrices of the data provide additional information that can be used to construct more efficient estimators than the full-sample GMM estimator which ignores such changes.

As an illustration, consider the following example with one dependent variable, one endogenous regressor, and one instrument. The structural equation is stable over time, while the parameter of the reduced form changes once at time T^* . The identification is strong over the first subsample (for $t \leq T^*$), and semi-strong² over the second one (for $t > T^*$) at some rate \tilde{r}_T (with $\tilde{r}_T = o(\sqrt{T})$). Conditional on knowing the break point T^* , the (same) structural parameter can be estimated at rate \sqrt{T} when using data from the first subsample, but only at rate \sqrt{T}/\tilde{r}_T when using data from the second subsample. We propose a GMM-type estimator that combines information from both subsamples to deliver a more efficient estimator of the structural parameter that converges at the fast rate \sqrt{T} . A similar result holds when the weakest subsample is actually weak (with $\tilde{r}_T = \sqrt{T}$), as well as when the break is unknown and estimated. When part of the sample is weakly identified, the advantage is even more striking: consistent estimation of the structural parameter is now possible through GMM-type inference procedures while (conservative) confidence regions obtained by weak-identification robust procedures are not necessary anymore. In our simulations, we document how the identification may appear weak over the whole sample when the break is ignored, even when part of the sample is strongly identified.

A lot of empirical evidence suggests that many econometric and financial models are subject to parameter instability and identification issues. In section 2, we discuss such evidence for

¹ q conditions can be constructed over each of the (two) subsamples where the variance is constant.

²See additional discussions about identification strength in Appendix A.1.

our main example, the New Keynesian Phillips Curve (NKPC), along with other examples. We show how the frameworks and methods developed in this paper can improve the efficiency of estimated parameters of interest, and shorten their confidence intervals. This is especially relevant for confidence sets of NKPC parameters which are often wide and uninformative when using weak-instrument robust methods over the full sample.

In practice, the existence and location of the break point may be unknown. We thus approach the unknown change in identification strength as a break point estimation problem. In doing so, our framework extends not only to changes in identification strength, but also to stable identification strength with parameter change in the reduced form. In addition, we explicitly account for the potential weakness of the instruments. Specifically, instruments may not be strong, and their identification strength may even change over the sample: for example, the identification may be weak over part of the sample.

Our main contributions are threefold. First, we extend the standard linear regression model with endogenous regressors to allow parameters and identification strength to change over time. Second, we develop statistical methods to detect parameter instability and changes in the identification strength. Third, we introduce two new efficient estimation procedures for the main parameters of the model. As a by-product, one of our estimation procedure also delivers more efficient estimators of the reduced form parameters than the usual full-sample OLS, in the presence of breaks in the second moments of the data. To our knowledge, this paper is the first to explicitly allow the identification strength to change over time, and to show how to efficiently use such information.

The following three cases are discussed successively:

1. the case where the structural equation remains stable while the reduced form may change (both parameter change without change in identification strength, and changes in identification strength are allowed);
2. the case where the reduced form equation remains stable while the equation of interest may change³;
3. the case where both equations may change at the same time.

In each case, we propose methods to detect and estimate the location of the break. Our methods also incorporate weaker and changing identification strength.

³Illustrations of such cases are given in section 2, Example 4, and in Appendix A.2.

Our paper relates to the weak-identification literature: see the surveys by Stock, Wright, and Yogo (2002), Dufour (2003), and Andrews and Stock (2005), and the survey of the applied literature by Hansen, Hausman, and Newey (2008). Typically, we consider a linear framework where the exact identification strength is unknown and allowed to change over time. We rely on Staiger and Stock’s (1997) popular rule-of-thumb to decide which identification framework (weak vs ”not weak”) is appropriate.

Our methods also cover and extend existing methods in the break point literature. In a linear time series regression model, Bai and Perron (1998) are first to jointly estimate break points and regression parameters by minimizing the sum of squared residuals. Hall, Han and Boldea (2012) extend this framework to allow for endogenous regressors. We extend Hall, Han and Boldea’s (2012) results to allow for weaker identification patterns that may change over time; we also study GMM-type estimators rather than 2SLS.

Finally, our paper relates to the recent work of Caner (2011) and Magnusson and Mavroeidis (2014). Caner (2011) proposes structural change tests that are robust to weak identification. Magnusson and Mavroeidis (2014) use reduced form breaks to improve testing, but without allowing changes in identification strength over the sample. We use similar breaks, but, in contrast with Magnusson and Mavroeidis, our methods identify the subsamples over which the instruments are not entirely weak, and use this additional information to shorten confidence intervals on the parameters of interest. In other words, we provide an additional source of information to improve estimation. Our simulations reveal that there is valuable (and reliable) information contained in the break that can be used to improve estimation: for instance, if the break is ignored, the identification may appear weak over the whole sample, even when a subsample is strongly identified.

In our Monte-Carlo study, we consider the linear IV regression model with one break in the reduced form. When estimating the slope parameter, we show that our proposed estimator always displays the smallest RMSE irrespective of the location of the break, especially under conditional heteroskedasticity. In addition, we compare the power curves associated with two types of inference: one, our proposed inference procedure that relies on the detection (and estimation) of the break and weak identification; two, a weak-identification robust inference procedure that ignores the break. Overall, the power properties of our inference procedure are better, while a (simple) Bonferroni-type adjustment is sufficient to control the size across all our simulation designs.

The paper is organized as follows. Section 2 presents four motivating examples that illustrate the relevance of our framework. Section 3 provides asymptotic results for efficient

estimation and inference in the presence of a break in the reduced form equations, or in the main equation. Section 4 presents our general framework and inference procedure, along with a comprehensive step-by-step guide for practitioners. In section 5 we introduce more general characterizations of the identification strength that allow each instrument and direction in the parameter space to display their own identification pattern. We also discuss efficient estimation of the reduced form in the presence of breaks. Section 6 illustrates our theoretical results through simulation studies. Section 7 concludes. The appendix contains additional discussions about identification and parameter instability in the linear IV model, simplified proofs of the theoretical results, as well as the graphs and tables associated with the simulation studies. Complete proofs of the theoretical results can be found in the supplemental appendix.

2 Motivating examples

In the last two decades, there has been a lot of empirical evidence suggesting that many macroeconomic and financial models (e.g. for inflation, aggregate demand, interest rates, or exchange rates) are subject to parameter instability and identification issues. We present four examples that emphasize the relevance of the framework and inference methods proposed in this paper.

- **Example 1: Break in reduced form parameters**

The New Keynesian Phillips Curve (NKPC) has recently received a lot of attention. The NKPC is a dynamic relationship resulting from a limited (or full-information) equilibrium model between inflation and driving variables such as output gap, unemployment, or real marginal costs (see, among others, Taylor (1980), Rotemberg (1982), Calvo (1983), and Clarida, Gali and Gertler (1999)). The typical stylized NKPC equation writes

$$\pi_t = \theta_f \pi_{t+1}^e + \theta_b \pi_{t-1} + \theta_y y_t + u_t \quad t = 1, \dots, T$$

where π_t denotes the inflation, π_{t+1}^e the expected inflation at time $t+1$ based on information available up to time t , and y_t the chosen driving variable (e.g. output gap, unemployment, or real marginal costs). Following Clarida, Gali and Gertler (1999), the parameters $(\theta_f, \theta_b, \theta_y)$ are assumed to be stable over the sample period (functions of some underlying structural parameters). Since π_{t+1}^e is based on information up to time t , and since y_t may be correlated

with contemporaneous noise such as demand shocks at time t , both are endogenous. Instruments commonly used to correct for endogeneity are lags of inflation, inflation forecasts, output gap, the average labor share, short-term interest rates and unemployment rates.⁴ Several NKPC studies for US find weak instruments over the period 1960-2007 (see Mavroeidis (2005), Dufour, Khalaf and Kichian (2006), Nason and Smith (2008), Kleibergen and Mavroeidis (2009) and Magnusson and Mavroeidis (2014)). Others find that the instruments are strong for the sample 1969-2005 (see Zhang, Osborn and Kim (2008, 2009)). This suggests that instrument strength changes over the sample. A change in identification strength over the sample period is also supported by the results in Kleibergen and Mavroeidis (2009, Table 4): the weak-identification robust confidence sets for the same NKPC parameters are considerably larger for the period 1960-1983 than for the period 1984-2007, suggesting that identification is stronger in the latter period. Such changes in instrument strength could come from a break in a policy function that generates a break in the reduced form for endogenous regressors, but no break in the deep structural parameters. The methods developed in this paper allow us to detect and locate not only reduced form parameter breaks, but also breaks in instrument strength. These breaks are then used to improve efficiency of structural parameter estimates and shorten their confidence intervals. This is especially relevant for the confidence sets of NKPC parameters which are often wide and uninformative when using weak-instrument robust methods over the full sample.

In contrast to Magnusson and Mavroeidis (2014), who focus on constructing confidence sets that control size regardless of the strength of instruments, and that may therefore be wide, our methods identify the subsamples over which the instruments are not entirely weak, and use those subsamples to shorten confidence intervals on the parameters of interest. In other words, and as suggested by Kleibergen and Mavroeidis (2009), we provide an additional source of information - namely the change in identification strength - to improve estimation: *"A natural response to the current finding that the NKPC is not well identified [...] is to look for more information"*.

- **Example 2: Break in structural error variance with no parameter breaks**

Asset returns models provide another example where our methods prove useful. Suppose we are interested in modeling a financial return on an asset in the home country, r_t^h , as a

⁴More recently, researchers have identified additional useful instruments such as the long-short interest rate spread (Gali and Gertler (1999), 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 from Stock and Watson's (2005) 132 variables (Kapetanios and Marcellino (2010)).

function of the same asset returns in foreign countries, r_t^f , and some lags,

$$r_t^h = \theta_a + \theta_b' r_t^f + \sum_{i=1}^p \theta_{c,i} r_{t-i}^h + \sum_{i=1}^q \theta_{d,i}' r_{t-i}^f + \epsilon_t.$$

For example, in Rigobon (2003), r_t^h are sovereign bond yields in Argentina, r_t^f are sovereign bond yields in the intimately connected foreign bond markets of Mexico, Brazil and US, and ϵ_t are structural shocks to the bond market. As Rigobon (2003) forcefully argues, the structural shock volatility increases substantially in financial crises. Such a break in variance can then be used for identification purposes. In this paper, we show that even if the model is identified, a break in the variance of ϵ_t provides an additional (non-redundant) moment condition for the estimation of $(\theta_a, \theta_b, \theta_c, \theta_d)$, and thus delivers more efficient estimators of these parameters.⁵ More generally, our method produces more efficient estimates of any asset return model when there is a break in the volatility of structural shocks, such as in a crisis.

- **Example 3: Break in instrument variance with no parameter break**

Most macroeconomic variables such as output, consumption, inflation, unemployment, to mention a few, have experienced a decline in volatility in the mid-1980s. This decline is referred to as the Great Moderation (see among others Stock and Watson (2002), Bernanke (2004) and Gali and Gambetti (2008)), and is usually modeled as a break in variance. Since lags of these macroeconomic variables are often used as instruments for estimation of various structural parameters, the Great Moderation effect amounts to a change in instrument variance, which can be used to construct additional moment conditions, and improve the efficiency of estimated structural parameters.

- **Example 4: Break in parameters of the equation of interest**

Suppose that we want to model interest rate via the Taylor rule. Orphanides (2001) writes down the simplest Taylor rule as⁶

$$f_t = \theta_1 + \theta_2 \pi_t + \theta_3 y_t + \epsilon_t,$$

with f_t the federal funds real interest rate, π_t the inflation, and y_t the output gap. The parameters θ_2 and θ_3 reflect the weight monetary policy puts on targeting inflation and

⁵Our result holds whether r_t^f is endogenous or not: see section 5.2 for the exogenous case.

⁶Formulations with a backward looking component are also allowed (see Clarida, Gali and Gertler (1998)).

output, respectively. In this model, both π_t and y_t are endogenous, and their lags along with some other macroeconomic variables are usually employed to estimate $(\theta_1, \theta_2, \theta_3)$. Since θ_2 and θ_3 are policy parameters, they are not immune to Lucas critique, and there is considerable evidence of a break in monetary policy in early 1980s; see e.g. Stock and Watson (2002) and Ahmed, Kevin, and Wilson (2004). The seminal paper of Stock and Watson (2002) convincingly shows that since the 1980s, monetary policy is more committed to fighting inflation, which translates in a bigger θ_2 .^{7,8} This provides an example where the equation of interest has parameter breaks, while the reduced forms of π_t and y_t may not experience breaks over the sample period of interest. Our methods show how to improve efficiency of parameter estimates in the presence of such a break in the equation of interest.

3 Two simplified frameworks of interest

Our framework extends the linear regression model with endogenous variables to allow structural parameters and identification strength to change over time. In this section, we start with two simplified frameworks: first, the case where the parameters of interest remain stable; second, the case where the reduced form remains stable. In each case, we introduce a new GMM-type estimator that uses additional valid information from the model. We study its asymptotic properties, including its efficiency with respect to existing estimators.

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 with p_1 exogenous variables Z_t , p_2 endogenous variables Y_t , and $p = p_1 + p_2$ parameters of interest θ^0 , writes

$$y_t = Z_t' \theta_z^0 + Y_t' \theta_y^0 + u_t = X_t' \theta^0 + u_t, \quad \text{with } X_t' = [Z_t' \ Y_t'] \text{ and } \theta^{0'} = [\theta_z^{0'} \ \theta_y^{0'}]. \quad (3.1)$$

⁷This break is at the core of the debate "Good policy or good luck?", which aims at explaining the reasons for the Great Moderation. Many studies, including Stock and Watson's (2002) and Ahmed, Kevin, and Wilson's (2004), found that the Great Moderation was partly due to improved policy - that is, breaks in the Taylor rule parameters - and partly due to luck - or a break downward in the variance of shocks ϵ_t .

⁸It can be argued that θ_2 is smaller again in recent years, because the federal funds rate was kept low and constant during the crisis, and was not used to fight inflation.

For a given vector of q valid instruments W_t with $q \geq p_2$ that includes the exogenous regressors Z_t , the unstable reduced form now writes

$$Y'_t = \begin{cases} \frac{W'_t \Pi_1}{r_{1T}} + v'_t & , \quad t \leq T^* \\ \frac{W'_t \Pi_2}{r_{2T}} + v'_t & , \quad t > T^* \end{cases}, \quad T^* = [T\nu^0], \quad (3.2)$$

with T^* the break point, ν^0 the break fraction, $r_{iT} = 1$, or $r_{iT} \rightarrow \infty$, and Π_i a full-rank matrix of size (q, p_2) for $i = 1, 2$. For now, we consider the simplest framework where all the instruments have the same (unknown) identification strength over each subsample. More general identification patterns allowing the strength to vary across instruments and directions of the parameter space are discussed in section 5.1.

The above break point T^* may capture two kinds of changes in the associated parameters:

- the identification strength remains stable over the whole sample, that is⁹ $r_{1T} \propto r_{2T}$ and $\Pi_1 \neq \Pi_2$;
- the identification strength changes, that is $r_{1T} = o(r_{2T})$, or $r_{2T} = o(r_{1T})$.

We are especially interested in cases where the identification strength changes. We show that, in such a case, only the magnitude of the change matters for identifying the true break point (that is the change from rate r_{iT} to rate r_{jT}), and not the change in the value of the reduced form parameters Π_i . To our knowledge, this is the first paper that explicitly accounts for changes in identification strength. Such changes are important because they can lead to improved inference about the structural parameters with more efficient estimators converging at faster rates than full-sample estimators.

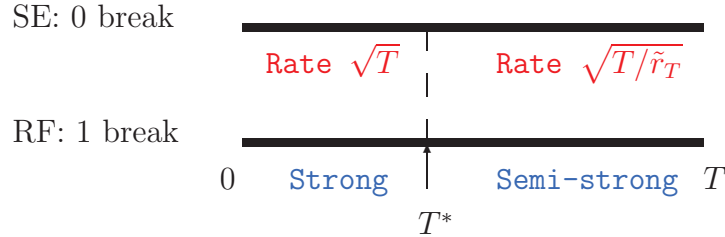
As an illustration, consider the following example where the structural equation is stable over time, while the reduced form has one break. In the first subsample, the identification is strong, while in the second subsample the identification is semi-strong¹⁰ at some (unknown) rate \tilde{r}_T (with $\tilde{r}_T = o(\sqrt{T})$). Conditional on knowing the break point T^* , the (same) structural parameters can be estimated at rate \sqrt{T} when using data from the first subsample, but only at rate \sqrt{T}/\tilde{r}_T when using data from the second subsample.¹¹ Our GMM-type estimator combines information from both subsamples to deliver more efficient estimators of the structural parameters (that converge at the fast rate \sqrt{T}); see Theorems 1 and 2. Of course, these results are asymptotic, but our simulations show that there are cases where such information can be used to draw sharper inference (or tighter confidence regions) on structural

⁹ $r_{1T} \propto r_{2T} \Leftrightarrow r_{1T}/r_{2T} \xrightarrow{T} c$ with c a real number such that $0 < |c| < \infty$.

¹⁰See additional discussions about identification strength in Appendix A.1.

¹¹Standard estimators (e.g. 2SLS or GMM) can be computed without knowing the rate of weakness \tilde{r}_T .

parameters of interest. A similar result holds when the break is unknown and when the weakest subsample is actually weak with $\tilde{r}_T = \sqrt{T}$. In such a case, the advantage is even more striking: consistent estimation of the structural parameters is now possible through GMM-type inference procedures and (conservative) weak-identification robust procedures are not necessary anymore.



Before introducing our GMM-type estimator of θ^0 , we first define the break point estimator. For any given (candidate) break point $\lfloor T\nu \rfloor$, $\hat{\Pi}_1(\nu)$ and $\hat{\Pi}_2(\nu)$ denote the OLS estimators computed in (3.2) over each associated subsample. The break point estimator $\hat{T}^* \equiv \lfloor T\hat{\nu} \rfloor$ of T^* is defined as in Bai and Perron (1998):

$$\hat{\nu} = \arg \min_{\nu} \left[Q_{OLS} \left(\nu, \hat{\Pi}_{vec}^s(\nu) \right) \right],$$

$$Q_{OLS} \left(\nu, \hat{\Pi}_{vec}^s(\nu) \right) = \frac{1}{T} \sum_{t=1}^{\lfloor T\nu \rfloor} \left(Y_t^s - W_t' \hat{\Pi}_1^s(\nu) \right)^2 + \frac{1}{T} \sum_{t=\lfloor T\nu \rfloor+1}^T \left(Y_t^s - W_t' \hat{\Pi}_2^s(\nu) \right)^2,$$

where, for a given choice of s , Y_t^s denotes the s^{th} element of Y_t , $\hat{\Pi}_i^s(\nu)$ the s^{th} column of $\hat{\Pi}_i(\nu)$ for $i = 1, 2$, and $\hat{\Pi}_{vec}^s(\nu) = \text{vec} \left(\hat{\Pi}_1^s(\nu), \hat{\Pi}_2^s(\nu) \right)$ with $s \in \{1, \dots, p_2\}$.^{12,13}

We now introduce three estimators of the structural parameters. These estimators will be considered in the simulation study in section 6.

- The full-sample 2SLS estimator uses first-stage predicted regressors $\hat{X}_t = \text{vec} (Z_t, \hat{Y}_t)$. It is defined as in Hall, Han, and Boldea (2012),

¹²The $\text{vec}(\cdot)$ notation is defined as follows: for any $\ell_1 \times \ell_2$ matrices A_1, \dots, A_{ℓ_3} , let $\text{vec} (A_1, \dots, A_{\ell_3})$ be the $(\ell_1 \ell_2 \ell_3) \times 1$ vector that stacks all ℓ_2 columns of each matrix A_1, \dots, A_{ℓ_3} , in order.

¹³For our purposes, only the consistency of the break point estimator and the associated rate stated in Theorem 1(i) are relevant, and not how the estimator is obtained. The asymptotic distribution of the OLS estimators $\hat{\Pi}_{vec}(\nu)$ is unaffected by the choice of s (or the precision of the break point estimator). In practice, one can also use the multivariate methods of Qu and Perron (2007) to estimate the break point common to all reduced forms; evaluation of the latter methods is beyond the scope of our paper.

$$\hat{\theta}_{2SLS} = \left(\sum_{t=1}^T \hat{X}_t \hat{X}_t' \right)^{-1} \sum_{t=1}^T \hat{X}_t y_t, \quad \hat{Y}_t' = \begin{cases} W_t' \hat{\Pi}_1, & t \leq \hat{T}^* \\ W_t' \hat{\Pi}_2, & t > \hat{T}^* \end{cases}, \quad \text{with } \hat{\Pi}_i = \hat{\Pi}_i(\hat{\nu}), i = 1, 2.$$

- The full-sample (or standard) GMM estimator is defined as:

$$\hat{\theta}_{GMM} = \arg \min_{\theta} \left[g_T'(\theta) \hat{S}_u^{-1} g_T(\theta) \right],$$

$$\text{with } g_T(\theta) = \frac{1}{T} \sum_{t=1}^T W_t (y_t - X_t' \theta) \quad \text{and} \quad \hat{S}_u \xrightarrow{p} \text{AVar} (T^{1/2} g_T(\theta^0)).$$

It ignores the break point in the reduced form, and is such that

$$\hat{\theta}_{GMM} = \left(X' W \hat{S}_u^{-1} W' X \right)^{-1} \left(X' W \hat{S}_u^{-1} W' y \right).$$

- The modified GMM estimator is defined as:

$$\hat{\theta}_{MOD} = \arg \min_{\theta} \left[\check{g}_T'(\theta) (\hat{S}_u^a)^{-1} \check{g}_T(\theta) \right],$$

$$\text{with } \check{g}_T(\theta) = \left(\begin{array}{c} \sum_{t=1}^{\hat{T}^*} W_t (y_t - X_t' \theta) / \hat{T}^* \\ \sum_{t=\hat{T}^*+1}^T W_t (y_t - X_t' \theta) / (T - \hat{T}^*) \end{array} \right) \quad \text{and} \quad \hat{S}_u^a \xrightarrow{p} \text{AVar} (T^{1/2} \check{g}_T(\theta^0)).$$

It uses information about the break point and is such that,

$$\hat{\theta}_{MOD} = \left(X' \check{W} (\hat{S}_u^a)^{-1} \check{W}' X \right)^{-1} \left(X' \check{W} (\hat{S}_u^a)^{-1} \check{W}' y \right),$$

$$\text{with } \check{W} \text{ the } T \times 2q \text{ matrix defined as } \check{W}' = \left(\begin{array}{cccc} W_1 & \cdots & W_{\hat{T}^*} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & W_{\hat{T}^*+1} & \cdots & 0 \end{array} \right).$$

These GMM-type estimators are known as "partial-sample GMM estimators". They were used in Andrews (1993) to derive the properties of a break point test (but for a break in θ^0 rather than in the reduced form) under local alternatives. Note that the above 2SLS is a special case of the GMM-MOD, but it is not the traditional 2SLS.

To derive asymptotic properties of the above estimators, we impose the following regularity assumptions.

Assumption 1. (*Regularity of the break fraction, error terms and reduced form*)

(i) $0 < \nu^0 < 1$, and the candidate break fractions ν are such that

$\max([T\nu], T - [T\nu]) \geq \max(q, \epsilon T)$, for some $\epsilon > 0$ such that $\epsilon < \min(\nu^0, 1 - \nu^0)$ and so

$\nu \in \Lambda_\epsilon = [\epsilon, 1 - \epsilon]$.

- (ii) Let $h_t = W_t \otimes \text{vec}(u_t, v_t)$ with i^{th} element $h_{t,i}$.
- The eigenvalues of $S = A\text{Var}\left(T^{-1/2} \sum_{t=1}^T h_t\right)$ are $\mathcal{O}(1)$.
- $E(h_{t,i}) = 0$ and for some $d > 2$, $\|h_{t,i}\|_d < \infty$ for $t = 1, \dots, T$ and $i = 1, \dots, (p_2 + 1)q$.
- $\{h_{t,i}\}$ is near epoch dependent with respect to some process $\{\xi_t\}$, $\|h_t - E(h_t | \mathcal{G}_{t-m}^{t+m})\|_2 \leq \nu_m$ with $\nu_m = \mathcal{O}(m^{-1/2})$ where \mathcal{G}_{t-m}^{t+m} is a σ -algebra based on $(\xi_{t-m}, \dots, \xi_{t+m})$.
- $\{\xi_t\}$ is either ϕ -mixing of size $m^{-d/[2(d-1)]}$ or α -mixing of size $m^{-d/(d-2)}$.
- (iii) Π_i is full column-rank equal to p_2 for $i = 1, 2$.

Assumption 1 is common for the break point literature, and is similar to Hall, Han and Boldea (2012). Part (i) ensures that there are enough observations in each subsample to identify the true break point. Part (ii) allows for general patterns of weak dependence in the data. Part (iii) ensures that the instruments are not redundant.

Assumption 2. *(Regularity of the identification strength)*

Let $r_T = \min(r_{1T}, r_{2T})$. We assume that $r_T = o(\sqrt{T})$.

Since the slowest sequence r_{iT} is associated with the subsample with the strongest identification, the sequence r_T corresponds to the strongest subsample. Assumption 2 prevents the identification strength to be weak over the whole sample. For instance, when there is a change in identification strength with stronger identification over subsample i , that is $r_{iT} = o(r_{jT})$, identification can be weak over the weakest subsample j , but not over subsample i . However, when there is no change in identification strength, $r_{iT} \propto r_{jT}$, the identification cannot be weak. We show in Theorem 1 below that such an assumption guarantees that the structural parameters can be consistently estimated.

Assumption 3. *(Regularity of the instrumental variables)*

Let $\hat{Q}_1(r) = T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} W_t W_t'$. Then $\hat{Q}_1(r) \xrightarrow{P} Q_1(r)$, uniformly in $r \in [0, 1]$ (u.r.) where $Q_1(r)$ is positive definite and strictly increasing in r .

Assumption 4. *(Regularity of the variances)*

$$A\text{Var} \left[T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} h_t \right] = S_1(r) = \begin{pmatrix} S_{u,1}(r) & S'_{uv,1}(r) \\ S_{uv,1}(r) & S_{v,1}(r) \end{pmatrix},$$

u.r, where $S_1(r)$ is positive definite and strictly increasing in r , with $S_{u,1}(r)$, $S_{v,1}(r)$ of size $q \times q$, respectively $(p_2q) \times (p_2q)$.

Assumptions 3 and 4 are typical for the break point literature. Assumption 3 ensures that there is enough variation in the instruments to identify the break point. It also allows for the variance of instruments to change over the sample, as in the Great Moderation Example 3 discussed in section 2. Assumption 4 allows for heteroskedasticity in the sample moments of the structural equation and the reduced form. It also allows for a break in the variance of structural errors u_t , as in Example 2 due to financial crises.

The following theorem collects asymptotic results about the above estimators of the break fraction and the structural parameters.

Theorem 1. *(Consistency of $\hat{\nu}$ and Asymptotic normality of $\hat{\theta}_{2SLS}$, $\hat{\theta}_{GMM}$, and $\hat{\theta}_{MOD}$)*

(i) Under Assumptions 1 to 3, we have $\|\hat{\nu} - \nu^0\| = \mathcal{O}_P(r_T^2/T)$.

(ii) Let $\Lambda_T = \text{diag}(T^{1/2}I_{p_1}, T^{1/2}r_T^{-1}I_{p_2})$ with r_T defined in assumption 2.

Under Assumptions 1 to 4, $\Lambda_T(\hat{\theta}_{2SLS} - \theta^0)$, $\Lambda_T(\hat{\theta}_{GMM} - \theta^0)$, and $\Lambda_T(\hat{\theta}_{MOD} - \theta^0)$ are asymptotically normally distributed with mean 0 and asymptotic variances, respectively,

$$\begin{aligned}
- \text{ when } r_{iT} = r_{jT}, & \begin{cases} V_{2SLS} &= (A_i + A_j)^{-1}(B_i + B_j)(A_i + A_j)^{-1} \\ V_{GMM} &= [(\Pi_i^{a'} Q_i + \Pi_j^{a'} Q_j)(S_{u,i} + S_{u,j})^{-1}(Q_i \Pi_i^a + Q_j \Pi_j^a)]^{-1} \\ V_{MOD} &= [\Pi_i^{a'} Q_i S_{u,i}^{-1} Q_i \Pi_i^a + \Pi_j^{a'} Q_j S_{u,j}^{-1} Q_j \Pi_j^a]^{-1} \end{cases} \\
- \text{ when } r_{iT} = o(r_{jT}), & \begin{cases} V_{2SLS} &= (A_i + A_j^z)^{-1}(B_i + B_j^z)(A_i + A_j^z)^{-1} \\ V_{GMM} &= [(\Pi_i^{a'} Q_i + \Pi_z^{a'} Q_j)(S_{u,i} + S_{u,j}^z)^{-1}(Q_i \Pi_i^a + Q_j \Pi_z^a)]^{-1} \\ V_{MOD} &= [\Pi_i^{a'} Q_i S_{u,i}^{-1} Q_i \Pi_i^a + \Pi_z^{a'} Q_j S_{u,j}^{-1} Q_j \Pi_z^a]^{-1} \end{cases}
\end{aligned}$$

with $Q_1 = Q(\nu^0)$, $Q_2 = Q_1(1) - Q_1(\nu^0)$, $S_{u,1} = S_{u,1}(\nu^0)$, $S_{u,2} = S_u(1) - S_{u,1}(\nu^0)$,

and $A_i = \Pi_i^{a'} Q_i \Pi_i^a$, $A_j^z = \Pi_z^{a'} Q_j \Pi_z^a$, $B_i = \Pi_i^{a'} S_{u,i} \Pi_i^a$, $B_j^z = \Pi_z^{a'} S_{u,i} \Pi_z^a$, $\Pi_i^a = (\Pi_z, \Pi_i)$, and $\Pi_z^a = (\Pi_z, O_{q \times p_2})$, for $i \neq j$ and either $i = 1$ or 2 .¹⁴

Comments:

(i) The consistency of the estimated break fraction extends results developed by Bai and Perron (1998) and Bai (1997) for $r_{1T} = r_{2T} = 1$. Here, we show that even with a change in identification strength, $r_{iT} = o(r_{jT})$ (for $i \neq j$ and either $i = 1$ or 2), the break fraction

¹⁴Note that while Π_i are the reduced form parameters of the endogenous regressors, Π_z denotes the implicit "reduced form" parameters for exogenous regressors, with elements equal to one for the correlation of exogenous regressors with themselves, and zero for the other instruments. Thus, Π_i^a is the matrix of all "reduced form" coefficients (including those on the exogenous regressors) and is only used to facilitate presentation of all results in a unified way across models with and without exogenous regressors.

estimator is consistent at the rate inherited from the strongest subsample. This holds even when the weakest subsample is weakly identified, $r_{jT} = \sqrt{T}$. Intuitively, only the magnitude of the break (in this case $o(\sqrt{T})$) matters for consistency of the break fraction estimator.

(ii) The rates of convergence of estimated parameters of the exogenous variable θ_z^0 (standard rate \sqrt{T}) and the estimated parameters of the endogenous variables θ_y^0 (slower rate \sqrt{T}/r_T) are extensions of the results developed by Antoine and Renault (2009) over stable reduced forms. The rate \sqrt{T}/r_T comes from the strongest subsample and holds even when the weakest subsample is genuinely weak.

(iii) To our knowledge, the consistency of both GMM-type estimators $\hat{\theta}_{GMM}$ and $\hat{\theta}_{MOD}$ - even when the weakest subsample is genuinely weak (that is $r_{jT} = \sqrt{T}$) - is new. Hence, ignoring the break point does not lead to a loss of consistency. However, using such information (to construct $2q$ valid sample moments) is crucial for efficiency as shown below in Theorem 2. The asymptotic normality of 2SLS-type estimator $\hat{\theta}$ is an extension of the results developed by Hall, Han, and Boldea (2012) for $r_{iT} = 1$.

The following assumptions are useful to derive some of our efficiency results.

Assumption 5. (*Homogeneity of the second moments*)

(i) $Q_1(r) = rQ$; (ii) $S_1(r) = rS$.

The above assumption prevents changes in the second moments of the instruments and in their correlation with the error terms of the reduced form. For example, a break in $\text{Var}(v_t W_t)$ at T^0 means that $\text{Var}(v_t W_t)$ changes once at some T^0 , so the homogeneity assumption 5(ii) is violated. In addition, a break in $E(W_t W_t')$ implies that $E(W_t W_t')$ changes once as $t = T^0$, so the homogeneity assumption 5(i) is violated.

Assumption 6. (*Conditional homoskedasticity in subsamples*)

$$S_i = \Phi \otimes Q_i, \text{ where } \Phi = \mathbb{E} \left[\begin{pmatrix} u_t^2 & u_t v_t' \\ u_t v_t & v_t v_t' \end{pmatrix} \middle| \mathcal{F}_t \right] = \begin{pmatrix} \Phi_u & \Phi'_{uv} \\ \Phi_{uv} & \Phi_v \end{pmatrix},$$

with \mathcal{F}_t the σ -algebra generated by $\{W_t, W_{t-1}, \dots\}$, and $i = 1, 2$.

Assumption 6 is only used in special cases to compare the above estimators - see e.g. Theorem 2(ii).

Theorem 2. (*Efficiency of estimated structural parameters*)

(i) Under Assumptions 1 to 4, $\hat{\theta}_{MOD}$ is always at least as efficient as $\hat{\theta}_{2SLS}$ and $\hat{\theta}_{GMM}$ asymptotically.

- (ii) In addition, $V_{MOD} = V_{GMM} \Leftrightarrow S_{u,i}^{-1}Q_i\Pi_i^a = \begin{cases} S_{u,j}^{-1}Q_j\Pi_j^a & \text{when } r_{iT} = r_{jT}, \\ S_{u,j}^{-1}Q_j\Pi_z^a & \text{when } r_{iT} = o(r_{jT}). \end{cases}$
- (iii) Under Assumptions 1 to 6, $V_{GMM} > V_{2SLS}$.

Comments:

(i) Using the additional information about the reduced form break leads to more efficient estimators (e.g. $\hat{\theta}_{MOD}$). Intuitively, the break does not impact the validity of traditional full-sample moment conditions, however more information can be extracted from the subsample moments due to their derivative changing over the sample. Condition (ii) above is hard to satisfy unless Assumption 5 holds. So, in general, $\hat{\theta}_{MOD}$ is strictly more efficient than $\hat{\theta}_{GMM}$.

(ii) The strict efficiency gain of $\hat{\theta}_{MOD}$ also occurs when there is no break in the reduced form (i.e. $\Pi_1 = \Pi_2 = \Pi$ and $r_{1T} = r_{2T} = r_T$), as long as one of the homogeneity assumption, 5(i) or 5(ii), is violated. We now discuss two such violations: a) a break in structural error variance as in Example 2; b) a break in instrument variance as in Example 3.

a) In the first case, Assumption 5(ii) is violated: for example, when the structural sample moment variance increases from $\text{Var}(W_t u_t) = S_{u,1}^*$ for $t \leq T^0$ to $\text{Var}(W_t u_t) = S_{u,2}^* > S_{u,1}^*$ for $t > T^0$. In Example 2, these breaks occur because a financial crisis can induce a break in the structural error variance u_t . Then, even with no break in the reduced form parameters, the variance break provides additional information that is used in $\hat{\theta}_{MOD}$ to obtain a strictly more efficient estimator. The strict efficiency gain occurs because the condition in Theorem 2(ii) can still be violated. For example, under Assumption 5(i), the condition writes $S_{u,1}^{*-1}Q\Pi^a = S_{u,2}^{*-1}Q\Pi^a$, and does not hold in general when $S_{u,2}^* > S_{u,1}^*$.

b) The second case refers to violations of Assumption 5(i). Example 3 discusses the Great Moderation, where the instrument variance decreases at T^0 , so that $E(W_t W_t') = Q_1^*$ for $t \leq T^0$, and $E(W_t W_t') = Q_2^* < Q_1^*$ for $t > T^0$. Then, under Assumption 5(ii), with no break in the reduced form parameters, the condition in Theorem 2(ii) writes $Q_1^* \Pi = Q_2^* \Pi$, and does not hold in general when $Q_1^* > Q_2^*$.

(iii) Finally, we consider a change in instrument strength, as in Example 1 for NKPC. Then, with no exogenous regressors, $\hat{\theta}_{MOD}$ is strictly more efficient than $\hat{\theta}_{GMM}$. With at least one exogenous regressor, $\hat{\theta}_{MOD}$ is still more efficient whenever Assumptions 5 and 6 hold.¹⁵

¹⁵Without exogenous regressors, see the proof of Theorem 2, case (b)(i) in the Appendix; with exogenous regressors, see the proof in the Supplemental Appendix.

For efficiency purposes, it is therefore essential to detect breaks in the reduced form equation and associated changes in the identification strength. The latter is done to ensure that at least one subsample is not weakly identified. Testing for weak identification¹⁶ must be done over each (stable) subsample, which requires consistent estimation of the break fraction.

Before testing for weak identification, one can test for a break in the reduced form using the sup-Wald statistic of Bai and Perron (1998). Define the null hypothesis for a certain endogenous regressor s as $H_0 : \Pi_1^s/r_{1T} = \Pi_2^s/r_{2T}$ (i.e. $\Pi_1^s = \Pi_2^s$ and $r_{1T} = r_{2T}$), versus the alternative H_A that $\Pi_1^s \neq \Pi_2^s$ or $r_{iT} = o(r_{jT})$ for $i \neq j$ and either $i = 1$ or 2 . Then, Bai and Perron's (1998) sup-Wald test statistic for testing for a break in the reduced form is:

$$Sup - Wald_T^{RF} = \sup_{\nu \in \Lambda_\epsilon} Wald_T^{RF}(\nu),$$

$$\begin{aligned} \text{where } Wald_T^{RF}(\nu) &= \left[T \hat{\Pi}_{vec}^{s'}(\nu) \mathcal{R}'_q (\mathcal{R}_q \hat{G}_{RF}^s(\nu) \mathcal{R}'_q)^{-1} \mathcal{R}_q \hat{\Pi}_{vec}^s(\nu) \right], \\ \hat{G}_{RF}^s(\nu) &= \text{diag} \left[\hat{Q}_1^{-1}(\nu) \hat{S}_{v,1}^s(\nu) \hat{Q}_1^{-1}(\nu), \hat{Q}_2^{-1}(\nu) \hat{S}_{v,2}^s(\nu) \hat{Q}_2^{-1}(\nu) \right], \\ G_{RF}^s(\nu) &= \text{diag} \left[Q_1^{-1}(\nu) S_{v,1}^s(\nu) Q_1^{-1}(\nu), Q_2^{-1}(\nu) S_{v,2}^s(\nu) Q_2^{-1}(\nu) \right], \\ \mathcal{R}_q &= (1, -1) \otimes I_q, \end{aligned}$$

and $S_{v,i}^s(\nu)$ the s^{th} diagonal ($q \times q$) block of $S_{v,i}(\nu)$ with $\hat{S}_{v,i}^s(\nu) \xrightarrow{p} S_{v,i}^s(\nu)$, uniformly in ν .

The asymptotic distribution of the sup-Wald test is stated below.

Theorem 3. (Test for a break in the reduced form)

(i) Under H_0 , and Assumptions 1 to 5, the above Sup-Wald test has the same null asymptotic distribution as in Bai and Perron (1998),

$$Sup - Wald_T^{RF} \Rightarrow \sup_{\nu \in \Lambda_\epsilon} \frac{\|\mathcal{B}_q(\nu) - \nu \mathcal{B}_q(1)\|^2}{\nu(1-\nu)},$$

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

(ii) Under H_A , and Assumptions 1 to 4, $Sup - Wald_T^{RF} \xrightarrow{p} \infty$. In addition, if Assumption 5 holds, then the implicit break fraction estimator $\hat{\nu}^W = \arg(Sup - Wald_T^{RF}) \xrightarrow{p} \nu^0$. Otherwise, $\hat{\nu}^W \xrightarrow{p} \nu^0$ is not guaranteed.

Comments:

(i) It is important to point out that result (i) holds even when $r_T = \sqrt{T}$, that is, when the instruments are weak over the whole sample.

¹⁶In our simulations, we rely on Staiger and Stock's (1997) rule-of-thumb to test weak identification.

(ii) While the above Sup-Wald test has power one against the alternative hypothesis H_A and should be employed because of its robustness to conditional heteroskedasticity and autocorrelation, the implicit break point estimator $\hat{\nu}^W$ is not always consistent unless the homogeneity assumption 5 holds. Thus, if the test rejects, it is still desirable to use $\hat{\nu}$ as a break point estimator, which is consistent even if assumption 5 is violated.

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. Example 4 shows that this can occur when the equation of interest is a policy function, while the reduced forms originate from a structural model that does not change over time.¹⁷ More specifically, the equation of interest with a break point is:

$$y_t = \begin{cases} Z_t' \theta_{z,1}^0 + Y_t' \theta_{y,1}^0 + u_t & , \quad t \leq [T\lambda^0] \\ Z_t' \theta_{z,2}^0 + Y_t' \theta_{y,2}^0 + u_t & , \quad t > [T\lambda^0] \end{cases} , \quad (3.3)$$

We also define the vector of all (unknown) parameters of interest, $\theta_{vec}^0 = \text{vec}(\theta_1^0, \theta_2^0)$ with $\theta_i^0 = \text{vec}(\theta_{z,i}^0, \theta_{y,i}^0)$ for $i = 1, 2$. The stable reduced form is

$$Y_t' = \frac{W_t' \Pi}{r_T} + v_t' , \quad (3.4)$$

where $r_T = 1$, or $r_T \rightarrow \infty$ with $r_T = o(\sqrt{T})$, and W_t is uncorrelated with v_t and u_t .

Before introducing our GMM-type estimator of θ_{vec}^0 , we first define the break point estimator. We extend results of Hall, Han, and Boldea (2012) to show that minimizing a 2SLS criterion provides consistent estimators of both the break fraction λ^0 and the structural parameters θ_{vec}^0 . In the first stage, the reduced form for Y_t is estimated over the full-sample by OLS to get $\hat{Y}_t = \hat{\Pi}' W_t$, with $\hat{\Pi}$ the OLS estimator in (3.4), and the augmented projected regressors $\hat{X}_t = \text{vec}(Z_t, \hat{Y}_t) = \hat{\Pi}^a' W_t$, where $\hat{\Pi}^a = (\Pi_z, \hat{\Pi})$. In the second stage, we define the following 2SLS criterion given a candidate break point $[T\lambda]$ and $\theta_{vec} = \text{vec}(\theta_1, \theta_2)$,

$$Q_{2SLS}(\lambda, \theta_{vec}) = \sum_{t=1}^{[T\lambda]} \left(y_t - \hat{X}_t' \theta_1 \right)^2 + \sum_{t=[T\lambda]+1}^T \left(y_t - \hat{X}_t' \theta_2 \right)^2 .$$

¹⁷To see another explicit system of equations with breaks that can give rise to a stable reduced form, see Appendix A.2.

We concentrate with respect to θ_{vec} to get the 2SLS estimators for each candidate break point, $\hat{\theta}_{vec}(\lambda) = \text{vec}(\hat{\theta}_1(\lambda), \hat{\theta}_2(\lambda))$, and then minimize $Q_{2SLS}(\lambda, \hat{\theta}_{vec}(\lambda))$ over all possible valid partitions of the sample defined by $[T\lambda]$. The 2SLS estimators of the break point $\hat{T} = [T\hat{\lambda}]$ and of the structural parameters $\hat{\theta}_{vec}$, are defined as

$$\hat{\lambda} = \arg \min_{\lambda} Q_{2SLS}(\lambda, \hat{\theta}_{vec}(\lambda)) \quad \text{and} \quad \hat{\theta}_{vec} = \text{vec}(\hat{\theta}_1(\hat{\lambda}), \hat{\theta}_2(\hat{\lambda})) = \text{vec}(\hat{\theta}_1, \hat{\theta}_2).$$

We now turn to GMM estimation of θ_{vec}^0 . We know from Hall, Han, and Boldea (2012) that minimizing a GMM criterion directly with respect to λ and θ_{vec}^0 (instead of the above 2SLS criterion) does not deliver consistent estimators of the break fraction. However, we show below that if we use the above 2SLS break point estimator $\hat{T} = [T\hat{\lambda}]$ to partition the sample, the resulting partial-sample GMM estimators are consistent. Intuitively, $\hat{\lambda}$ converges faster than the parameter estimates, explaining the consistency result for GMM. We now introduce two GMM-type estimators of θ_{vec}^0 .

- The partial-sample GMM estimators using \hat{T} are defined as follows:

$$\hat{\theta}_{GMM,vec} = \begin{pmatrix} \hat{\theta}_{GMM,1} \\ \hat{\theta}_{GMM,2} \end{pmatrix} = \arg \min_{\theta_1, \theta_2} \left[g'_T(\theta_1, \theta_2) \hat{S}_u^{-1} g_T(\theta_1, \theta_2) \right],$$

$$\text{where } g_T(\theta_1, \theta_2) = \begin{bmatrix} \hat{T}^{-1} \sum_{t=1}^{\hat{T}} W_t (y_t - Y'_t \theta_{y,1} - Z'_t \theta_{z,1}) \\ (T - \hat{T})^{-1} \sum_{t=\hat{T}+1}^T W_t (y_t - Y'_t \theta_{y,2} - Z'_t \theta_{z,2}) \end{bmatrix}, \quad (3.5)$$

and $\hat{S}_u = \text{diag}[\hat{S}_{u,1}, \hat{S}_{u,2}] \xrightarrow{p} \text{AVar}[T^{1/2} g_T(\theta_1^0, \theta_2^0)]$ with $\hat{S}_{u,i}$ of size $(q \times q)$.

These GMM estimators were defined in Andrews (1993) to study local properties of a break point test. Here, we derive their asymptotic distribution under a stable reduced form, and compare them with their 2SLS counterparts. Similar to 2SLS estimators, we show that the GMM estimators of the structural parameters associated with the endogenous variables are asymptotically normally distributed at rate \sqrt{T}/r_T , whereas those associated with the exogenous regressors are asymptotically normally distributed at the standard rate \sqrt{T} .

However, these GMM estimators, $\hat{\theta}_{GMM,vec}$, are not the most efficient, because they ignore the information that the reduced form is stable, while the 2SLS estimators $\hat{\theta}_{vec}$ take this information into account. This suggests that a (augmented) GMM estimator that takes this information into account will also be more efficient than $\hat{\theta}_{GMM,vec}$.

- The "augmented" GMM estimator is built by adding moment conditions from the reduced form. In the absence of breaks, the reduced form moments are redundant to the

structural form moments, because adding just-identified nuisance parameters leaves the GMM estimators asymptotically unaffected (see Ahn and Schmidt (1995)).

The same intuition applies when adding the just-identified q full-sample moments of the reduced form. However, it does not apply when we add the $2q$ subsample moments of the reduced form, before and after T^0 . In other words, we add $2qp_2$ moment conditions and qp_2 nuisance parameters Π (recall that Π is stable over the whole sample). These added reduced form moment conditions over-identify Π . In addition, we show in Theorem 5 that they are not redundant and deliver a more efficient estimator.

Our new (augmented) GMM estimator $\hat{\theta}_{MOD,vec} = \text{vec}(\hat{\theta}_{MOD,1}, \hat{\theta}_{MOD,2})$ (along with the new reduced form estimator $\hat{\Pi}_{MOD,vec} = \text{vec}(\hat{\Pi}_{MOD})$) are based on the following moments:

$$\check{g}_T(\theta_{vec}, \Pi_{vec}) = \begin{pmatrix} g_T(\theta_{vec}) \\ g_{T,2}(\Pi_{vec}) \end{pmatrix} \quad \text{with } g_T \text{ as in (3.5), and}$$

$g_{T,2}(\Pi_{vec})$ stacks the OLS moment conditions from the subsamples of the reduced form with $\Pi_{vec} = \text{vec}(\Pi)$ for any Π . More precisely, we have:

$$g_{T,2}(\Pi_{vec}) = \begin{bmatrix} \hat{T}^{-1} \sum_{t=1}^{\hat{T}} W_t(Y_t^1 - W_t' \Pi^1) \\ \dots \\ \hat{T}^{-1} \sum_{t=1}^{\hat{T}} W_t(Y_t^{p_2} - W_t' \Pi^{p_2}) \\ (T - \hat{T})^{-1} \sum_{t=\hat{T}+1}^T W_t(Y_t^1 - W_t' \Pi^1) \\ \dots \\ (T - \hat{T})^{-1} \sum_{t=\hat{T}+1}^T W_t(Y_t^{p_2} - W_t' \Pi^{p_2}) \end{bmatrix},$$

where Y_t^s and Π^s are the s^{th} columns of Y_t' and Π , for $s = 1, \dots, p_2$. Then, the optimal MOD estimators are defined as:

$$\begin{pmatrix} \hat{\theta}_{MOD,vec} \\ \hat{\Pi}_{MOD,vec} \end{pmatrix} = \arg \min_{\theta_{vec}, \Pi} \left[\check{g}_T'(\theta_{vec}, \Pi_{vec}) \hat{\mathcal{S}}^{-1} \check{g}_T(\theta_{vec}, \Pi_{vec}) \right],$$

where $\hat{\mathcal{S}} \xrightarrow{p} \text{AVar}[\sqrt{T} \check{g}_T(\theta_{vec}^0, \Pi_{vec}^0)]$ with $\Pi_{vec}^0 = \text{vec}(\Pi)$.

In order to discuss asymptotic properties of the above estimators, we impose the following regularity assumptions. Those are quite similar to the regularity assumptions of section 3.1.

Assumption 7. (Regularity of the break fraction λ^0 , and reduced form)

(i) $0 < \lambda^0 < 1$, and the candidate break points satisfy $\max([T\lambda], T - [T\lambda]) > \max(p - 1, \epsilon T)$ for some $\epsilon > 0$ such that $\epsilon < \min(\lambda^0, 1 - \lambda^0)$ and so $\lambda \in \Lambda_\epsilon = [\epsilon, 1 - \epsilon]$.

(ii) $\text{Rank}(\Pi) = p_2$.

The following theorem collects asymptotic results about the above estimators of the break fraction, the structural and reduced form parameters.

Theorem 4. (*Consistency of $\hat{\lambda}$ and Asymptotic normality of $\hat{\theta}_{vec}$, $\hat{\theta}_{GMM,vec}$ and $\hat{\theta}_{MOD,vec}$*)

(i) Under Assumptions 1(ii), 2 to 4, and 7, $\|\hat{\lambda} - \lambda^0\| = \mathcal{O}_P(T^{-1})$.

(ii) Define Λ_T as in Theorem 1. Under Assumptions 1(ii), 2 to 4, and 7, $[I_2 \otimes \Lambda_T](\hat{\theta}_{vec} - \theta_{vec}^0)$, $[I_2 \otimes \Lambda_T](\hat{\theta}_{GMM,vec} - \theta_{vec}^0)$, and $[I_2 \otimes \Lambda_T](\hat{\theta}_{MOD,vec} - \theta_{vec}^0)$ are asymptotically normally distributed with mean zero and asymptotic variances defined explicitly in the appendix.

(iii) Under Assumptions 1(ii), 2 to 4, and 7, $\text{diag}(I_2 \otimes \Lambda_T, T^{1/2}I_{p_2q}) \begin{bmatrix} \hat{\theta}_{MOD,vec} - \theta_{vec}^0 \\ \hat{\Pi}_{MOD,vec} - \Pi_{vec}^0/r_T \end{bmatrix}$ is asymptotically normally distributed with mean zero and asymptotic variance explicitly defined in the appendix.

Comments:

(i) The structural equation break fraction estimator converges faster to its true value than its reduced form counterpart; in fact it does so at the fastest available rate T . This stems from the presence of breaks in the exogenous regressor parameters; intuitively, the exogenous regressors are their own strong instruments, and so the strength of the endogenous regressor instruments determines the fast convergence rate. However, part (ii) shows that estimators of the endogenous regressor parameters, $\theta_{y,i}^0$, converge at rate \sqrt{T}/r_T , which is slower than usual: such slow rate is due to the presence of instruments W_t that are not strong. The other parameters, $\theta_{z,i}^0$, are not affected by the instruments, and their estimators are asymptotically normally distributed at the standard rate \sqrt{T} .

(ii) The explicit formulas for the asymptotic variance-covariance matrices are provided in the appendix because they require cumbersome notations.

The following theorem compares the asymptotic variance of the proposed estimators.

Theorem 5. (*Efficiency of estimated structural parameters*)

(i) Under Assumptions 1(ii), 2 to 4, and 7, $\hat{\theta}_{MOD,vec}$ is always as efficient as $\hat{\theta}_{GMM,vec}$.

(ii) Under Assumptions 1(ii), 2 to 4, and 5 to 7, for $i = 1, 2$,

$$\begin{aligned} AVar(\hat{\theta}_{2SLS,i}) \leq AVar(\hat{\theta}_{GMM,i}) &\Leftrightarrow 2\Phi'_{uv}\theta_{y,i}^0 + \theta_{y,i}^{0'}\Phi_v\theta_{y,i}^0 \leq 0 \\ AVar(\hat{\theta}_{MOD,i}) \leq AVar(\hat{\theta}_{2SLS,i}) &\Leftrightarrow 2\Phi'_{uv}\theta_{y,i}^0 + \theta_{y,i}^{0'}\Phi_v\theta_{y,i}^0 \geq -\frac{\delta\Phi_u^2}{1 + \delta\Phi_u}, \\ &\text{with } \delta = \Phi_u^{-2}\Phi'_{uv}(\Phi_v - \Phi_{uv}\Phi_u^{-1}\Phi'_{uv})^{-1}\Phi_{uv}. \end{aligned}$$

Comments:

(i) MOD estimators are more efficient than the usual partial sample GMM estimators, because they exploit the additional information that the reduced form is stable. In general, the added moment conditions $g_{T,2}(\cdot)$ are not redundant, because of two key reasons: first, Π is over-identified; second, the added moment conditions are correlated with the initial ones. Even when the first subsample moments of the reduced form first are redundant for the second subsample moments of the reduced form - as it is the case under Assumptions 5 and 6 - they are not overall redundant for the estimation of θ_{vec}^0 . This intuition is similar to Theorem 4 in Breusch, Qian, Schmidt, and Wyhowski (1999), where it is shown that with three moment conditions, say (g_1, g_2, g_3) , g_3 redundant given g_2 does not imply that g_3 is redundant given (g_1, g_2) . The above non-redundancy result is also related to recent results by Antoine and Renault (2014) who extends Breusch et al. (1999) to frameworks that allow different identification strengths.

(ii) 2SLS estimators are not a special case of partial-sample GMM estimators: in fact, the latter only use subsample information for estimation of parameters θ_1^0 and θ_2^0 , while 2SLS estimators use both subsample and full-sample information. As a result, the 2SLS estimators considered here can be more efficient than partial-sample GMM estimators. The above condition,

$$2\Phi'_{uv}\theta_{y,i}^0 + \theta_{y,i}^{0'}\Phi_v\theta_{y,i}^0 \leq 0 \Leftrightarrow \Phi_u + 2\Phi'_{uv}\theta_{y,i}^0 + \theta_{y,i}^{0'}\Phi_v\theta_{y,i}^0 \leq \Phi_u$$

is actually equivalent to the following condition on conditional variance of the errors,

$$\text{Var}(u_t + v_t'\theta_{y,i}^0 | \mathcal{F}_t) \leq \text{Var}(u_t | \mathcal{F}_t), \quad \text{with } \mathcal{F}_t \text{ the information set available at time } t.$$

Heuristically, it states that 2SLS estimators are more efficient when the second-stage error after using the full-sample reduced form for estimation on a subsample is less than the structural GMM error.

(iii) 2SLS estimators are not a special case of MOD estimators either.¹⁸ In general, our MOD estimators are more efficient even under homogeneity and conditional homoskedasticity, as long as

$$\Phi'_{uv}\theta_{y,i}^0 + \theta_{y,i}^{0'}\Phi_v\theta_{y,i}^0 \geq -\frac{\delta\Phi_u^2}{1 + \delta\Phi_u}.$$

This condition is harder to interpret, but it is automatically satisfied with a single endogenous regressor and no exogenous regressors, that is $p_2 = 1, p_1 = 0$. In such a case,

¹⁸This is related to results in Hall, Han and Boldea (2012) about the relationship between 2SLS and GMM estimators that breaks down in the presence of break points.

$\Phi_u \delta = \Phi_{uv}^2 (\Phi_v \Phi_u - \Phi_{uv}^2)^{-1}$, and the above condition becomes,

$$2\Phi_{uv}\theta_{y,i}^0 + (\theta_{y,i}^0)^2\Phi_v \geq -\frac{\Phi_{uv}^2}{\Phi_v} \Leftrightarrow \left(\theta_{y,i}^0 + \frac{\Phi_{uv}}{\Phi_v}\right)^2 \geq 0.$$

In this case, MOD is always at least as efficient as 2SLS. MOD and 2SLS estimators are asymptotically equivalent when $-\theta_{y,i}^0 = \frac{\Phi_{uv}}{\Phi_v}$ for $i = 1, 2$. To interpret this condition, assume the regressors are fixed. Then $\beta^0 = \frac{\Phi_{uv}}{\Phi_v}$ is the limiting coefficient of a regression of u_t on v_t , suggesting that the MOD estimator "purges" u_t of the true correlation with v_t . On the other hand, the 2SLS estimator transforms the error u_t into $(u_t + v_t'\theta_{y,i}^0)$ through an orthogonal projection, so that $-\theta_{y,i}^0$ plays the role of β^0 for each subsample. As a result, when these two are equal, say over subsample 1, the two associated estimators 2SLS and MOD are asymptotically equivalent over subsample 1, but the MOD estimator is more accurate over subsample 2.

So far, our analysis assumes that the existence of the break point in the structural equation (3.3) is known. In practice, this existence often needs to be established. To that end, we consider the sup-Wald test of Hall, Han and Boldea (2012) for which the null and alternative hypotheses are: $H_0 : \mathcal{R}_p\theta^0 = 0$ versus $H_A : \mathcal{R}_p\theta^0 \neq 0$, with $\mathcal{R}_p = (1, -1) \otimes I_p$. The test statistic is:

$$Sup - Wald_T = \sup_{\lambda \in \Lambda_\epsilon} Wald_T(\lambda), \quad (3.6)$$

where
$$Wald_T(\lambda) = T\hat{\theta}'_{vec}(\lambda)\mathcal{R}'_p \left[\mathcal{R}_p \hat{G}(\lambda) \mathcal{R}'_p \right]^{-1} \mathcal{R}_p \hat{\theta}_{vec}(\lambda),$$

$$\hat{G} = \text{diag}[\hat{G}_1(\lambda), \hat{G}_2(\lambda)] \quad \text{with} \quad \hat{G}_i(\lambda) = \hat{\mathcal{A}}_i^{-1}(\lambda)\hat{H}_i(\lambda)\hat{\mathcal{A}}_i^{-1}(\lambda),$$

$$\hat{\mathcal{A}}_i(\lambda) = T^{-1} \sum_{t \in I_i} \hat{X}_t \hat{X}'_t \quad , I_1 = \{1, \dots, [T\lambda]\}, I_2 = \{[T\lambda] + 1, \dots, T\},$$

and $\hat{H}_i(\lambda)$ is a HAC estimator such that

$$\Lambda_T^{-1} \hat{H}_i(\lambda) \Lambda_T^{-1} \xrightarrow{p} H_i(\lambda) = \text{AVar} \left[\sum_{t \in I_i} \Lambda_T^{-1} \hat{X}_t (u_t + v_t' \theta_{y,i}^0) \right].$$

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

Theorem 6. *(Test for a break in the structural equation)*

(i) Under $H_0 : \theta_1^0 = \theta_2^0$, Assumptions 1(ii), 2 to 4, 5, and 7,

$$Sup - Wald_T \Rightarrow \sup_{\lambda \in \Lambda_\epsilon} \frac{\|\mathcal{B}_p(\lambda) - \lambda \mathcal{B}_p(1)\|^2}{\lambda(1-\lambda)},$$

where $\mathcal{B}_p(\lambda)$ is a $p \times 1$ vector of independent standard Brownian motions defined on $[0, 1]$.

(ii) Under $H_A : \theta_\Delta^0 = \theta_1^0 - \theta_2^0 \neq 0$, and Assumptions 1(ii), 2 to 4, and 7, $\text{Sup-Wald}_T \xrightarrow{p} \infty$ such that

$$\text{Sup-Wald}_T = \begin{cases} \mathcal{O}_p(T/r_T^2) & \text{without exogenous regressors } (p_1 = 0) \\ \mathcal{O}_p(T) & \text{in presence of exogenous regressors } (p_1 \neq 0) \end{cases}$$

In addition, if Assumption 5 holds and either $p_1 = 0, p_2 = 1$ (only one endogenous regressor, no exogenous regressors) or $\theta_{y,1}^0 = \theta_{y,2}^0$, then the implicit break fraction estimator $\hat{\lambda}^W = \arg(\text{Sup-Wald}_T^{RF}) \xrightarrow{p} \lambda^0$. Otherwise, $\hat{\lambda}^W \xrightarrow{p} \lambda^0$ is not guaranteed.

Comments:

(i) The above Sup-Wald test should be used because of its robustness to conditional heteroskedasticity and autocorrelation. However, if one is not willing to impose the homogeneity assumption 5, then the implicit break fraction estimator may not be consistent under H_A , that is $\hat{\lambda}^W \not\xrightarrow{p} \lambda^0$. In such a case, the break point estimator $\hat{\lambda}$ should be used instead. This is similar to the properties of the Sup-Wald test in the reduced form highlighted in the previous section.

(ii) The rate of divergence under the alternative H_A depends on the presence of exogenous regressors: without exogenous regressors, the rate of divergence is affected by the identification strength of the instruments and is equal to T/r_T^2 ; in presence of exogenous regressors, the rate is standard equal to T , and not affected by the identification strength.

4 Common Break

In this section, we combine the two frameworks of interest introduced in the previous section, allowing for a common break in the equation of interest and in the reduced form:

$$y_t = \begin{cases} Z_t' \theta_{z,1}^0 + Y_t' \theta_{y,1}^0 + u_t & , \quad t \leq [T\lambda^0] \\ Z_t' \theta_{z,2}^0 + Y_t' \theta_{y,2}^0 + u_t & , \quad t > [T\lambda^0] \end{cases} \quad (4.1)$$

$$Y_t' = \begin{cases} \frac{W_t' \Pi_1}{r_{1T}} + v_t' & , \quad t \leq [T\lambda^0] \\ \frac{W_t' \Pi_2}{r_{2T}} + v_t' & , \quad t > [T\lambda^0] \end{cases} \quad (4.2)$$

where $r_{iT} = 1$, or $r_{iT} \rightarrow \infty$, with $i = 1, 2$, and W_t is not correlated with v_t and u_t .

When there is no change in the identification strength, $r_{1T} \propto r_{2T}$, (4.2) naturally extends the unstable reduced form models considered in Hall, Han and Boldea (2012) to weaker identification patterns. Otherwise, (4.2) captures changes in identification strength concomitant

to those in the parameter of interest θ_{vec}^0 . Our goal is to detect and locate both parameter instability and changes in the strength of identification, as well as to provide correct and sharp inference on θ_{vec}^0 . This goal is met by combining results developed in section 3.

1. First stage: the reduced form.

- (a) Test whether there is a break in the reduced form using $Sup - Wald_T^{RF}$.
- (b) If a break is detected, use the break point estimator \hat{T}^* and subsample OLS to construct \hat{Y}_t as discussed in section 3.1. Otherwise, construct \hat{Y}_t by full-sample OLS.

2. Second stage: the structural equation.

- (a) If no break has been detected in the reduced form, test for a break in the main equation using $Sup - Wald_T$. If a break is found, proceed with inference using $\hat{\lambda}$, $\hat{\theta}_{MOD,i}$ or $\hat{\theta}_i$.
- (b) If a break point has been detected in the reduced form, impose the above estimated break point \hat{T}^* and work over the associated subsamples separately. Since on each such subsample the reduced form is stable, we use the results developed in section 3.2.

3. Third stage: the common break structure.

- (a) If a break point has been detected in the reduced form, impose the break \hat{T}^* in (4.1) and test whether the break is common to the structural equation using the test $Wald_T^c$ described below.
- (b) If the test does not reject, the main equation is stable. Use $\hat{\theta}_{MOD}$ for inference as discussed in section 3.1. If the test rejects, the break is common to both equations, so proceed with inference using \hat{T} and the partial sample estimators $\hat{\theta}_{MOD,i}$ or $\hat{\theta}_i$.

It is important to mention that the estimation of the above break points and the inference described in this paper are feasible when the identification is "not weak" over at least one of the subsamples. Thus, one needs to test for weak versus "not weak" identification over each subsample. For simplicity, we rely on Staiger and Stock's (1997) popular rule-of-thumb in our simulations. When both subsamples have weak instruments, weak-identification robust inference procedures should be used.¹⁹

We now present our Wald test for common break. Consider the case where the reduced form break T^* has been detected and estimated by \hat{T}^* using the methods described in section 3.1. To test whether T^* is a common break to the main equation, we test whether the 2SLS parameter estimates defined over each subsample are equal to each other. These parameter

¹⁹See Magnusson and Mavroeidis (2014) for suggestions.

estimates are defined as

$$\hat{\theta}_i^c = \left(\sum_{\hat{I}_i^*} \hat{X}_{t,i} \hat{X}'_{t,i} \right)^{-1} \left(\sum_{\hat{I}_i^*} \hat{X}_{t,i} y_t \right), \quad \text{with} \quad \hat{X}_{t,i} = \text{vec} (Z_t, \hat{Y}_{t,i}), \quad \text{and} \quad \hat{Y}'_{t,i} = W'_t \hat{\Pi}_i,$$

$i = 1, 2$, and $\hat{I}_1^* = \{1, \dots, \hat{T}^*\}$ and $\hat{I}_2^* = \{\hat{T}^* + 1, \dots, T\}$. The Wald test for a common break is:

$$Wald_T^c = T \hat{\theta}_{vec}^c \mathcal{R}'_p (\mathcal{R}_p \hat{G}^c \mathcal{R}'_p)^{-1} \mathcal{R}'_p \hat{\theta}_{vec}^c,$$

$$\text{with} \quad \hat{\theta}_{vec}^c = \text{vec} (\hat{\theta}_1^c, \hat{\theta}_2^c), \quad \mathcal{R}_p = (1, -1) \otimes I_p,$$

$$\hat{G}^c = \text{diag} [\hat{G}_1^c, \hat{G}_2^c], \quad \hat{G}_i^c = (\hat{A}_i^c)^{-1} \hat{B}_i^c (\hat{A}_i^c)^{-1},$$

$$\hat{A}_i^c = T^{-1} \sum_{\hat{I}_i^*} \hat{X}_{t,i} \hat{X}'_{t,i},$$

$$\hat{B}_i^c \text{ such that } \Lambda_{iT}^{-1} \hat{B}_i^c \Lambda_{iT}^{-1} \xrightarrow{p} B_i^c = \text{AVar} [T^{-1/2} \sum_{\hat{I}_i^*} \Lambda_{iT}^{-1} \hat{X}_{t,i} u_t],$$

$$\text{and} \quad \Lambda_{iT} = \text{diag}(T^{1/2} I_{p_1}, T^{1/2} r_{iT}^{-1} I_{p_2}).$$

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

Theorem 7. (*Wald test for common break*)

(i) Under $H_0 : \theta_1^0 = \theta_2^0$, Assumptions 1(ii), 2 to 4, 5, and 7, $Wald_T^c \xrightarrow{d} \chi_p^2$.

(ii) Under $H_A : \theta_1^0 \neq \theta_2^0$, Assumptions 1(ii), 2 to 4, 5, and 7, we have

$$Wald_T^c = \begin{cases} \mathcal{O}_p(T) & \text{in presence of exogenous regressors } (p_1 \neq 0) \\ \mathcal{O}_p(T/\bar{r}_T^2) & \text{without exogenous regressors } (p_1 = 0) \text{ and } \bar{r}_T = \max_i(r_{iT}). \end{cases}$$

Comment:

The above test for common break is somewhat similar to the break point test defined in section 3.2 to detect a break in the main equation. There are two main differences: first, the above test is simpler than Sup-Wald test because it is computed directly at the estimated break point coming from the reduced form; second, the rate of divergence is different in absence of exogenous regressors.

5 Extensions and related results

5.1 General characterization of the identification strength

In this section, we consider a more general characterization of the identification strength by allowing each instrument and direction in the parameter space to display their own

identification pattern. More specifically, the reduced form equation (3.2) now writes,

$$Y'_t = \begin{cases} W'_t \Pi_T^{(1)} + v'_t & , \quad t \leq T^* \\ W'_t \Pi_T^{(2)} + v'_t & , \quad t > T^* \end{cases}$$

where each element (k, l) of matrices $\Pi_T^{(j)}$ ($j = 1, 2$) is allowed to display its own rate of convergence, that is $\pi_{kl,T}^{(j)} = \pi_{kl}^{(j)} / r_{kl,T}^{(j)}$ with $r_{kl,T}^{(j)} = 1$, or $r_{kl,T}^{(j)} \xrightarrow{T} \infty$ with $r_{kl,T}^{(j)} = \mathcal{O}(\sqrt{T})$. The break point T^* now captures changes in the identification strength of some instruments only, changes of different magnitudes, as well as changes in different directions of the parameter space. We focus on the following three special cases:

- Case a): the (overidentified) case with one endogenous variable and two instruments with different identification strengths;
 - Case b): the (just-identified) case with two endogenous variables and associated structural parameters identified at different rates;
 - Case c): the (just-identified) case with two endogenous variables where the instrument strength is the same for each reduced form, but differs across instruments.
- Case a): instruments with different identification strengths.

In practice, instruments often display different identification strengths. For instance, in Example 1 in section 2, lags of inflation are usually relatively strong instruments for inflation, but lags of output gap are not.

We consider the (overidentified) case with one endogenous variable, no additional exogenous variable, and two instruments associated with two different identification strengths:

$$y_t = Y_t \alpha^0 + u_t, \quad Y_t = \begin{cases} \frac{W_{1t} \pi_1^{(1)}}{r_{1T}^{(1)}} + \frac{W_{2t} \pi_2^{(1)}}{r_{2T}^{(1)}} + v'_t & , \quad t \leq T^* \\ \frac{W_{1t} \pi_1^{(2)}}{r_{1T}^{(2)}} + \frac{W_{2t} \pi_2^{(2)}}{r_{2T}^{(2)}} + v'_t & , \quad t > T^* \end{cases}, \quad r_{iT}^{(j)} = o(r_{kT}^{(j)}) \quad (i \neq k)$$

with $E(u_t) = 0$, $E(W_{1t} u_t) = E(W_{2t} u_t) = 0$, and $E(W_{1t} v_t) = E(W_{2t} v_t) = 0$.

The above framework now captures changes in the strength of one instrument only (e.g. when $r_{2T}^{(1)} = r_{2T}^{(2)}$), changes in the strength of both instruments but of different magnitudes, as well as the case where one instrument becomes weaker, while the other becomes stronger. From the results of section 3.1, when there is one break and two instruments with the same strength $r_T^{(j)}$ over subsample j , we know that with $r_T = \min(r_T^{(1)}, r_T^{(2)})$ and $r_T = o(\sqrt{T})$,

- (i) the estimated break fraction is consistent at rate r_T^2/T ;
- (ii) only the magnitude of the break matters;
- (iii) the MOD estimator of α^0 is always at least as efficient as the modified 2SLS and the full-sample GMM.

From Theorem 2 in Antoine and Renault (2014), we also know that when there is no break, and two instruments with different strengths \tilde{r}_{iT} such that $\min(\tilde{r}_{1T}, \tilde{r}_{2T}) = o(\sqrt{T})$,

- (iv) the (standard) over-identified GMM estimator of α^0 that relies on both instruments W_1 and W_2 is more efficient (in terms of asymptotic variance) even when W_2 is genuinely weak, as long as $(W_{1t}u_t)$ and $(W_{2t}u_t)$ are correlated.

These results and intuition directly transfer to the above framework after redefining r_T as the slowest rate over all possible rates displayed in both matrices $\Pi_T^{(j)}$ ($j = 1, 2$).²⁰

- Case b): structural parameters identified at different rates.

In practice, some parameters are often known to be more difficult to estimate accurately. For example, in the intertemporally separable consumption based capital asset pricing model with constant relative risk-aversion preferences, this is usually the case for the risk-aversion parameter, but not for the discount factor. In Stock and Wright (2000), the discount factor is modeled as strongly identified, whereas the risk-aversion is weakly identified.

Accordingly, we consider the (just-identified) case with two endogenous variables²¹, no additional exogenous variable, and two orthogonal instruments. Over each subsample, W_1 and W_2 are both strong instruments for Y_1 , but they are both weaker instruments for Y_2 :

$$\begin{aligned}
 y_t &= Y_{1t}\alpha^0 + Y_{2t}\beta^0 + u_t \\
 Y_{1t} &= \begin{cases} W_{1t}\pi_{11}^{(1)} + W_{2t}\pi_{21}^{(1)} + v_{1t} & , \quad t \leq T^* \\ W_{1t}\pi_{11}^{(2)} + W_{2t}\pi_{21}^{(2)} + v_{1t} & , \quad t > T^* \end{cases} \\
 Y_{2t} &= \begin{cases} W_{1t}\pi_{12}^{(1)}/r_{2,T}^{(1)} + W_{2t}\pi_{22}^{(1)}/r_{2,T}^{(1)} + v_{2t} & , \quad t \leq T^* \\ W_{1t}\pi_{12}^{(2)}/r_{2,T}^{(2)} + W_{2t}\pi_{22}^{(2)}/r_{2,T}^{(2)} + v_{2t} & , \quad t > T^* \end{cases}
 \end{aligned}$$

with $E(u_t) = 0$, $E(W_{1t}u_t) = E(W_{2t}u_t) = 0$, $E(W_{1t}v_t) = E(W_{2t}v_t) = 0$, and $E(W_{1t}W_{2t}) = 0$.

²⁰The current proof of Theorem 1 goes through with the new definition of r_T .

²¹This example also relates to earlier discussions with one exogenous regressor and one endogenous one.

Taken over each subsample j , this is the (linear) framework of Stock and Wright (2000) where the structural parameters α^0 and β^0 are identified at different rates. The associated moment functions over subsample j

$$\begin{aligned} & \begin{cases} g_1^{(j)}(\alpha, \beta) &= E[W_{1t}(y_t - Y_{1t}\alpha - Y_{2t}\beta)] \\ g_2^{(j)}(\alpha, \beta) &= E[W_{2t}(y_t - Y_{1t}\alpha - Y_{2t}\beta)] \end{cases} \\ \Leftrightarrow & \begin{cases} g_1^{(j)}(\alpha, \beta) &= (\alpha^0 - \alpha)E(W_{1t}^2)\pi_{11}^{(j)} + (\beta^0 - \beta)E(W_{1t}^2)\pi_{12}^{(j)}/r_{2,T}^{(j)} \\ g_2^{(j)}(\alpha, \beta) &= (\alpha^0 - \alpha)E(W_{2t}^2)\pi_{21}^{(j)} + (\beta^0 - \beta)E(W_{2t}^2)\pi_{22}^{(j)}/r_{2,T}^{(j)} \end{cases} \end{aligned}$$

contain a "strong part" that only depends on α and is not drifting toward zero. It follows that α^0 is strongly identified, while β^0 is not. In addition, when $r_{2,T}^{(j)} = o(\sqrt{T})$, the standard GMM estimator of (α^0, β^0) is such that $\begin{pmatrix} \sqrt{T}(\hat{\alpha} - \alpha^0) \\ \sqrt{T}/r_{2,T}^{(j)}(\hat{\beta} - \beta^0) \end{pmatrix}$ is asymptotically normal with mean 0; see Antoine and Renault (2009).

The above framework now captures changes in the identification strength of one parameter only, or even in both parameters but of different magnitudes.

Since the instrument strength is the same across the two reduced forms, the results developed in section 3.1 (e.g. estimation of the break point) apply equation by equation. For efficient estimation of the structural parameters, it is always better to consider the MOD estimator, as already discussed.

- Case c): each instrument has the same identification strength across all reduced forms.

For example, the intercept is always a strong instrument. We consider the (just-identified) case with two endogenous variables, no additional exogenous variable, and two orthogonal instruments. Over each subsample, W_1 is a strong instrument for both Y_1 and Y_2 , while W_2 is a weaker instrument for both Y_1 and Y_2 ,

$$\begin{aligned} y_t &= Y_{1t}\alpha^0 + Y_{2t}\beta^0 + u_t \\ Y_{1t} &= \begin{cases} W_{1t}\pi_{11}^{(1)} + W_{2t}\pi_{21}^{(1)}/r_{2,T}^{(1)} + v_{1t} & , t \leq T^* \\ W_{1t}\pi_{11}^{(2)} + W_{2t}\pi_{21}^{(2)}/r_{2,T}^{(2)} + v_{1t} & , t > T^* \end{cases} \\ Y_{2t} &= \begin{cases} W_{1t}\pi_{12}^{(1)} + W_{2t}\pi_{22}^{(1)}/r_{2,T}^{(1)} + v_{2t} & , t \leq T^* \\ W_{1t}\pi_{12}^{(2)} + W_{2t}\pi_{22}^{(2)}/r_{2,T}^{(2)} + v_{2t} & , t > T^* \end{cases} \end{aligned}$$

with $E(u_t) = 0$, $E(W_{1t}u_t) = E(W_{2t}u_t) = 0$, $E(W_{1t}v_t) = E(W_{2t}v_t) = 0$, and $E(W_{1t}W_{2t}) = 0$.

Different directions in the parameter space (or linear combinations of α^0 and β^0) can now be identified at different rates. The associated moment functions over subsample j write:

$$\begin{cases} g_1^{(j)}(\alpha, \beta) = (\alpha^0 - \alpha)E(W_{1t}^2)\pi_{11}^{(j)} + (\beta^0 - \beta)E(W_{1t}^2)\pi_{12}^{(j)} \\ g_2^{(j)}(\alpha, \beta) = (\alpha^0 - \alpha)E(W_{2t}^2)\pi_{21}^{(j)}/r_{2,T}^{(j)} + (\beta^0 - \beta)E(W_{2t}^2)\pi_{22}^{(j)}/r_{2,T}^{(j)} \end{cases}$$

The (strong) instrument W_1 delivers the strongly identified moment $g_1^{(j)}$, while W_2 delivers the weakly identified moment $g_2^{(j)}$. It follows that only one (specific) direction in the parameter space is strongly identified. However, unlike case b), the strong direction does not necessarily correspond to any structural parameter; e.g. the structural parameters will not be strongly identified. Following Antoine and Renault (2009), the strong direction can be found through a reparametrization that is based on the orthogonal of the null space of the Jacobian associated with the (strong) moment $g_1^{(j)}$. In our case, this Jacobian vector is

$$J_1^{(j)} = \frac{\partial g_1^{(j)}(\alpha^0, \beta^0)}{\partial[\alpha \ \beta]} = -[E(W_{1t}^2)\pi_{11}^{(j)} \ E(W_{1t}^2)\pi_{12}^{(j)}],$$

and the orthogonal of its null space is spanned by the vector $e_s^{(j)} = \begin{pmatrix} \pi_{11}^{(j)} \\ \pi_{12}^{(j)} \end{pmatrix}$.

It is interesting to realize that different strong directions are identified over each subsample whenever $e_s^{(1)}$ and $e_s^{(2)}$ are not parallel to each other, that is $e_s^{(1)} \neq ce_s^{(2)}$ for some constant $c \neq 0$. For example, this is the case when either π_{11} or π_{12} changes (but not both), or when they both change but not by a proportional amount: we expect this to happen more often than not in practice. It also means that when considering moments from both subsamples together, two (different) directions in the parameter space will be strongly identified, and therefore the entire parameter space (including α^0 and β^0 !).

The MOD estimator of (α^0, β^0) (see section 3.1) is defined by stacking the four moments obtained from both subsamples: $g_1^{(1)}$ and $g_1^{(2)}$ are the two strong moments that drive the strength of the identification of the parameter space, but for efficiency, the other two moments should also be included as already discussed in case a).

5.2 Efficient estimators for the reduced form

In this section, we show that in the presence of breaks, we can construct not only more efficient GMM estimators of the structural form, but also more efficient estimators of the reduced form. To formalize this, consider the following reduced form where we are interested

in efficiently estimating Π . For simplicity, we consider one endogenous regressor Y_t and no additional exogenous regressor ($p_1 = 0, p_2 = 1$), and we also impose strong identification.

$$Y_t = W_t' \Pi + v_t.$$

The parameter Π is stable, however we allow for potential breaks in $\text{Var}(v_t W_t)$, $\mathbf{E}(W_t W_t')$, or both at T^0 , which is assumed known for simplicity. A break in $\text{Var}(v_t W_t)$ at T^0 implies that the homogeneity assumption 5(ii) is violated. A break in $\mathbf{E}(W_t W_t')$ implies that the homogeneity assumption 5(i) is violated.

In section 3.2, we have introduced two estimators of Π : the usual OLS estimator, $\hat{\Pi} = (W'W)^{-1}W'Y$, and the MOD estimator $\hat{\Pi}_{MOD}$ based on the moment conditions,

$$\check{g}_T(\theta_{vec}, \Pi) = \begin{cases} g_T(\theta_{vec}) \\ g_{T,2}(\Pi) \end{cases}.$$

We now introduce a third estimator that ignores the structural form and relies on the subsample moment conditions before and after the break, $g_{T,2}(\Pi)$. We call this estimator $\hat{\Pi}_{GMM}$ because it is the optimal estimator that uses the $2q$ moments $g_{T,2}(\Pi)$ to estimate q parameters. The following theorem shows that $\hat{\Pi}_{MOD}$ is the most efficient.

Theorem 8. (*Efficiency of reduced form estimators*)

(i) Under Assumptions 1(ii), 2 to 4, and 7, $\sqrt{T}(\hat{\Pi} - \Pi)$, $\sqrt{T}(\hat{\Pi}_{GMM} - \Pi)$, and $\sqrt{T}(\hat{\Pi}_{MOD} - \Pi)$ are asymptotically normally distributed with mean zero and respective asymptotic variance-covariance matrices,

$$\begin{aligned} V_{OLS, \Pi} &= (Q_1 + Q_2)^{-1}(S_{v,1} + S_{v,2})(Q_1 + Q_2)^{-1} \\ V_{GMM, \Pi} &= (Q_1 S_{v,1}^{-1} Q_1 + Q_2 S_{v,2}^{-1} Q_2)^{-1} \\ V_{MOD, \Pi} &= (V_{GMM, \Pi}^{-1} + \mathcal{G}'_* \mathcal{G}_*)^{-1} \quad \text{with } \mathcal{G}_* \text{ defined in the appendix.} \end{aligned}$$

(ii) Under Assumptions 1(ii), 2 to 4, and 7: $V_{MOD, \Pi} \leq V_{GMM, \Pi} \leq V_{OLS, \Pi}$.

(iii) Under Assumptions 1(ii), 2 to 4, and 7, and 5 or 6: $V_{MOD, \Pi} \leq V_{GMM, \Pi} = V_{OLS, \Pi}$.

Comments:

(i) The inference developed to construct more efficient GMM estimators of the structural parameters carries over to provide more efficient estimators of the reduced form parameters in the presence of breaks. In other words, the OLS estimators of the reduced form parameters are no longer the most efficient in presence of breaks in second moments of the instruments,

or in their correlation with the error terms. It is important to note that OLS estimators remain the most efficient under conditional homoskedasticity, that is if the regressors W are independent of the errors v and can be treated as fixed. Consequently, our results do not conflict with the Gauss-Markov Theorem, which states that OLS estimators are the most efficient given W .

(ii) In addition, if the regressors are independent of the errors, $\hat{\Pi}_{GMM}$ and $\hat{\Pi}_{MOD}$ are more efficient than $\hat{\Pi}$, even under (unconditional) homoskedasticity. Thus, our estimators correct not only for heteroskedasticity across subsamples, but also for changes in the second moment of the regressors.

6 Monte-Carlo simulations

We consider the framework of section 3.1 with one endogenous regressor Y , q valid instruments (including the intercept), and one break in the reduced form:

$$y_t = \alpha + Y_t\beta + \sigma_t\epsilon_t, \quad Y_t = \begin{cases} 1 + W_t'\Pi_1 + v_t & t \leq T^* \\ 1 + W_t'\Pi_2 + v_t & t > T^* \end{cases}, \quad E[\epsilon_t W_t] = 0, \quad E[v_t W_t] = 0.$$

The errors (ϵ_t, v_t) are i.i.d. jointly normally distributed with mean 0, variance 1 and correlation ρ ; the instruments W_t are i.i.d jointly normally distributed with mean zero and variance-covariance matrix equal to the identity matrix, and independent of (ϵ_t, v_t) . The parameters of the model are such that, with ι_k denoting the vector of ones of size k ,

$$(\alpha \ \beta) = (0 \ 0), \quad \Pi_i = d_i \iota_{q-1}, \quad (i = 1, 2) \quad \text{with} \quad d_1 = \sqrt{\frac{R_1^2}{(q-1)(1-R_1^2)}}, \quad d_2 = d_1 + b.$$

We consider two versions of the model: homoskedasticity with $\sigma_t^2 = 1$; conditional heteroskedasticity (Garch) with $\sigma_t^2 = 0.1 + 0.6u_{t-1}^2 + 0.3\sigma_{t-1}^2$ and $u_t = \sigma_t\epsilon_t$.

We are interested in the slope parameter β . In experiments 1 and 2, we compare the performance of three estimators of β : (i) the new MOD estimator proposed in this paper (that relies on the break); (ii) the 2SLS estimator proposed by Hall, Han and Boldea (2012) (that also relies on the break); (iii) the standard full-sample GMM that ignores the break. In experiment 1, their performances are evaluated by computing the Monte-Carlo bias, standard deviation, root-mean squared errors (RMSE), as well as the length and coverage of corresponding 95% confidence intervals²², for various configurations of the model. In

²²The standard errors of each estimator are computed using the formulas in Theorem 1. We use HAC-type estimators under conditional heteroskedasticity.

experiment 2, we investigate these performance measures as a function of the location of the break. In experiment 3, we compare the power curves of two types of inference to test $H_0 : \beta = \beta_0$ over a range of values for β_0 . We consider the identification-robust inference procedure that ignores the break and the inference procedure proposed in this paper that relies on the detection and estimation of the break, as well as the detection of weak identification.

• **Experiment 1:**

Our benchmark model is such that the sample size is $T = 400$, the endogeneity parameter is $\rho = 0.5$, the true break is located at $T^* = 160$ with break size $b = 1$. We use $q = 4$ instruments (including the intercept), and the R-square over the first subsample is $R_1^2 = 0.2$, which corresponds to a first-stage F-test statistic equal to 13 and somewhat strong identification.²³ The implied reduced form parameters are $d_1 = 0.29$ and $d_2 = 1.29$.

We then explore different configurations of the model. First, we decrease R_1^2 to display weaker identification in the first subsample, while the second subsample remains strong: $R_1^2 = 0.05$ (and $F_1 = 2.7$), and $R_1^2 = 0.01$ (and $F_1 = 0.5$). The break size is still $b = 1$, but the implied reduced form parameters are now $d_1 = 0.13$ and $d_1 = 0.06$, respectively. Then, we consider larger sample size, $T = 800$, more instruments, $q = 6$, larger endogeneity parameter, $\rho = 0.75$. In all these experiments, the break is assumed to be known, and the results are displayed in Tables 1 and 2 (for the homoskedastic and Garch cases). The results for cases where the break location is unknown and estimated are displayed in Tables 3 and 4 (for the homoskedastic and Garch cases): three break sizes, 1, 0.5, and 0.2, are considered; $R_1^2 = 0.2$ and $d_1 = 0.29$ throughout, while $d_2 = 1.29, 0.79, \text{ and } 0.49$. All the results are based on 5,000 replications.

- When the break is known, the main results do not vary much over the different specifications. We then focus on the benchmark case. Under homoskedasticity, the performances of MOD and 2SLS are very close when considering the bias, the standard deviation, and the RMSE. And their RMSEs are significantly smaller than for GMM. It is worth mentioning that the biases of MOD and 2SLS tend to be larger than for GMM, but they are well-compensated by the gains in terms of standard deviation; in addition, when the sample size increases, such biases decrease as expected. When looking at the 95% confidence intervals of the slope parameter, MOD displays the shortest ones while maintaining good coverage

²³Recall the link between the R^2 and the first-stage F-statistic $F = \frac{R^2}{(1-R^2)} \times \frac{(T-q)}{(q-1)}$. Staiger and Stock's (1997) rule-of-thumb declares the instruments weak when the F-test statistic is below 10.

properties. Under conditional heteroskedasticity (Garch case), the standard deviation and RMSE of MOD are much smaller than for 2SLS as expected.

- When the break point is treated as unknown, the actual break size is important for the accuracy of the estimated break location. With a break size of 1, the estimated break is quite reliable with an average (over the estimated breaks) very close to the actual break: the average is 161.3 with a true break at 160. When the break size decreases, the quality of the estimator of the break location deteriorates: for instance, with a true break at 160 and a break size of 0.2, the average is 172.4. Reliable estimation of the location of the break is crucial for the bias properties of MOD and 2SLS. We can see that when the break is not accurately estimated, their biases increase, and the coverage properties of the confidence intervals also worsen.²⁴ This bias should not be too much of a concern, because it only appears when the break size is small, and, oftentimes, such small breaks cannot be detected; see also experiment 3 below.

• **Experiment 2:** Performance as a function of the true location of the break.

We have shown the asymptotic efficiency of MOD (compared to GMM and 2SLS). And, at least asymptotically, it is always efficient to "split" the sample in order to double the number of moments. Intuitively, it seems reasonable when the break is somewhat in the middle of the sample. We now investigate how the performance of the three estimators, MOD, 2SLS and GMM, varies with the (true) location of the break.

We consider three versions of the above model, all with $T = 400$, $\rho = 0.5$, $q = 4$, and $R_1^2 = 0.1$ (which corresponds to $d_1 = 0.1925$ and $F_1 = 5.8$):

- model (i): the R-square remains the same over both subsamples: $R_1^2 = R_2^2 = 0.1$. The associated break size is $b = -0.385$, and $d_2 = -0.1925$. The identification strength is borderline weak over the second subsample with $F_2 = 8.7$.
- model (ii): the R-square increases over the second subsample, $R_2^2 = 0.22$. The associated break size is $b = -0.5$, and $d_2 = -0.3075$. In this model, the identification is strong over the second subsample with $F_2 = 22.2$.
- model (iii): the R-square is smaller over the second subsample, $R_2^2 = 0.025$. The

²⁴One remedy consists in discarding the data around the estimated break (e.g. in a confidence interval for the break location). This simple strategy should mitigate the drawback from estimating the break inaccurately, and using partly misspecified moments. However, it does require the asymptotic distribution of the break, which is beyond the scope of this paper.

associated break size is $b = -0.1$, and $d_2 = 0.0925$. In this model, the identification is weak over the second subsample with $F_2 = 2$.

The results under homoskedasticity and conditional heteroskedasticity are presented in Figures 1 and 2: two measures of performance are considered, the Monte-Carlo RMSE (left), and the Monte-Carlo standard deviation (right). All results are based on 5,000 replications.

- In model (i), both Monte-Carlo RMSE and standard deviations for MOD and 2SLS are stable as the break location changes from (0.1×400) to (0.9×400) . This is quite different for GMM: first, both its RMSE and standard deviation are quite larger than those of MOD and 2SLS (as expected from the results of experiment 1); second, both are increasing as a function of the location of the break until it is in the middle of the sample, then they are decreasing to return to their original levels. Results for model (ii) are very similar, and available upon request.

- In model (iii), both Monte-Carlo RMSE and standard deviations for all inference procedures are decreasing functions of the location of the break. This is not very surprising since the explanatory power over the second subsample is quite smaller than over the first one ($R_1^2 = 0.1$ and $R_2^2 = 0.025$).

• **Experiment 3:** Power curves of the overall inference procedure.

We now compare the power curves of two types of inference procedures to test $H_0 : \beta = \beta_0$ for a range of β_0 values: (i) an identification-robust procedure (IdR hereafter) that ignores the break; (ii) our suggested procedure (MOD hereafter) (that tests for break and identification strength). When using IdR, we compute a 95% confidence interval for β , and check whether the tested value β_0 belongs to it. We consider two IdR procedures, Anderson-Rubin (hereafter AR) and Kleibergen (2005, hereafter K).²⁵ When using our suggested procedure, we first test 0 vs 1 break: if 0 break, we test for weak over the whole sample and use either GMM or IdR to compute the confidence interval for β ; if 1 break, we estimate it, and conditional on the estimated break, we test for weak over each subsample, and use either MOD or IdR. We test for the presence of break with the Sup-Wald test at 95%; we use Staiger and Stock's rule-of-thumb to test for weak identification. We consider two versions of MOD: one where we simply use a 95% confidence interval for β ; the other, MOD-adj, where we adjust the size of the test using a Bonferroni-type correction, and use

²⁵These procedures are more computationally-friendly than other IdR procedures, because their critical values are known and do not need to be simulated for each tested value.

a $(1 - 0.05/2)\%$ confidence interval. When MOD (or MOD-adj) relies on K, we denote it MOD-K (or MOD-K-adj); when it relies on AR, we denote it MOD-AR (or MOD-AR-adj). We consider the three versions of the homoskedastic model from experiment 2, models (i), (ii) and (iii). All results are based on 5,000 replications, and are reported in Figures 3, 4, and 5. We also report the rejection probability at the true value, the probability of detecting the break, and the probability of detecting weak identification when the break is ignored (full-sample F-test), and when the detection of the break is taken into account (full-sample F-test when no break is detected, and subsample F-test when a break is detected). The results for AR are very close to K (with a slight lack of power for AR as expected) and are not reproduced here; they are available upon request.

Comments:

(i) Test of weak identification in presence of a break.

Interestingly, accounting for the presence of a break matters a lot when testing weak identification. When the break is ignored, the sample is declared weak much more often, even though it is not necessarily weak over each subsample. For example, for model (ii), the identification is strong over the second subsample ($F_2 = 22$). Yet, the sample is almost always declared weak when the break is ignored. However, when the break is accounted for, the sample is never declared weak, as expected. This means that there is valuable and reliable information contained in the break that can be used to improve estimation as we discuss next. As a robustness check, it is also worth mentioning that for model (iii) with a second subsample that is quite weak ($F_2 = 2$), accounting for the break does not change how often the sample is declared weak: both probabilities are approximately 0.76.

(ii) IdR vs MOD.

Overall the power properties of MOD are better than K. This means that confidence intervals for β will be narrower when using MOD. Of course, MOD is slightly oversized due to the pretest: instead of 5%, the rejection probabilities at the true value are between 8% and 10% across all simulation designs. However, our simple Bonferroni-type adjustment is sufficient to control the size across all simulation designs without affecting the power properties much. Fully accounting for the error of pretesting is beyond the scope of this paper. We refer the interested reader to the powerful size-correction methods recently developed by McCloskey (2012); see also references therein.

In model (i), K does not have any power, while MOD (and MOD-adj) both display the usual well-shaped power curve achieving a power equal to 0.5 for tested values $|\beta_0| > 0.6$. In model (ii), all inference procedures have some power. However, MOD has much more

power: it is equal to 1 for tested values $|\beta_0| > 0.6$ for MOD, and less than 0.5 for K. Finally, in model (iii), the identification is quite weak over the second subsample, and all inference procedures behave very similarly. It is reassuring to see that when the identification is weak over the whole sample, there does not seem to be a cost in accounting for the break.

7 Conclusion

There is a lot of empirical evidence that macroeconomic models such as the NKPC are subject to parameter instability and identification issues. In this paper, we consider both issues in a unified framework, and provide a comprehensive treatment of the link between them. To our knowledge, it is the first paper that explicitly accounts for the connection between parameter instability and changes in identification strength. Such changes in identification strength provide an additional source of information that is used to improve estimation.

As long as at least one subsample is not weakly identified, we show that standard procedures can be used to detect and estimate break points. In addition, given the estimated break point, we propose a GMM-type estimator for the parameters of interest that is more efficient than competitors (e.g. the full-sample GMM and the 2SLS estimator of Hall, Han and Boldea (2012)). When parameter instability is confined to the main equation, we exploit the stability of the reduced form equation to propose another efficient GMM-type estimator when the identification is not weak. More generally, 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.

In our simulation study, our inference procedures rely on Staiger and Stock's (1997) popular rule-of-thumb (based on the first stage F-test) to distinguish weak and "not weak" identification, either on the whole sample, or on each subsample. The associated results, especially the power curves of our global inference procedures, are very promising. We expect such results to be even better with more elaborate and powerful tests of weak identification such as those proposed in Antoine and Renault (2013).

References

- [1] K. Abadir and J. Magnus, *Matrix algebra*, Econometric Exercises 1, Cambridge University Press, 2005.
- [2] S. Ahmed, A. Levin, and B.A. Wilson, *Recent U.S. macroeconomic stability: good policies, good practices, or good luck?*, The Review of Economics and Statistics **86** (2004), 824–832.
- [3] S.C. Ahn and P. Schmidt, *A separability result for GMM estimation, with applications to GLS prediction and conditional moment tests*, Econometric Reviews **14** (1995), 19–34.
- [4] D.W.K. Andrews, *Tests for parameter instability and structural change with unknown change point*, Econometrica **61** (1993), 821–856.
- [5] D.W.K. Andrews and X. Cheng, *Estimation and inference with weak, semi-strong, and strong Identification*, Econometrica **80** (2012), 2153–2211.
- [6] D.W.K. Andrews and J.H. Stock, *Inference with weak instruments*, Econometric Society Monograph Series, vol. 3, ch. 8 in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Cambridge University Press, Cambridge, 2007.
- [7] B. Antoine and E. Renault, *Efficient GMM with nearly-weak Instruments*, The Econometrics Journal, Tenth Anniversary Issue **12** (2009), 135–171.
- [8] ———, *Efficient inference with poor instruments: a general framework*, ch. 2 in *Handbook of Empirical Economics and Finance*, Taylor & Francis, 2010.
- [9] ———, *Testing identification strength*, Working paper, Simon Fraser University (2013).
- [10] ———, *On the relevance of weaker instruments*, Working Paper (2014).
- [11] J. Bai, *Estimation of a change point in multiple regression models*, Review of Economics and Statistics **79** (1997), 551–563.
- [12] J. Bai and P. Perron, *Estimating and testing linear models with multiple structural changes*, Econometrica **66** (1998), 47–78.

- [13] B. Bernanke, *The Great Moderation*, ch. 5 of 'The Taylor rule and the transformation of monetary policy', pp. 143–182, Hoover Institution, 2004.
- [14] T. Breusch, H. Qian, P. Schmidt, and D. Wyhowski, *Redundancy of moment conditions*, *Journal of Econometrics* **91** (1999), 89–111.
- [15] G.A. Calvo, *Staggered prices in a utility-maximizing framework*, *Journal of Monetary Economics* **12** (1983), 383–398.
- [16] M. Caner, *Pivotal structural change tests in linear simultaneous equations models with weak identification*, *Econometric Theory* **27** (2011), 413–427.
- [17] R. Clarida, J. Galí, and M. Gertler, *Monetary policy rules in practice: some international evidence*, *European Economic Review* **42** (1998), 1033–1067.
- [18] ———, *The science of monetary policy: a New Keynesian perspective*, *Journal of Economic Literature* **37** (1999), 1661–1707.
- [19] J.-M. Dufour, *Identification, weak instruments and statistical inference in econometrics*, *Canadian Journal of Economics* **36** (2003), 767–808.
- [20] J.-M. Dufour, L. Khalaf, and M. Kichian, *Inflation dynamics and the New Keynesian Phillips Curve: an identification robust econometric analysis*, *Journal of Economic Dynamics and Control* **30** (2006), 1707–1727.
- [21] ———, *On the precision of Calvo parameter estimates in structural NKPC models*, *Journal of Economic Dynamics and Control* **34** (2010), 1582–1595.
- [22] J. Galí and L. Gambetti, *On the sources of the Great Moderation*, NBER Working Paper No. 14171 (2008).
- [23] J. Galí, M. Gertler, and D. Lopez-Salido, *European inflation dynamics*, *European Economic Review* **45** (2001), 1237–1270.
- [24] J. Hahn and G. Kuersteiner, *Discontinuities of weak instruments limiting distributions*, *Economics Letters* **75** (2002), 325–331.
- [25] A. R. Hall, S. Han, and O. Boldea, *Inference regarding multiple structural changes in linear models with endogenous regressors*, *Journal of Econometrics* **170** (2012), 281–302.

- [26] C. Hansen, J. Hausman, and W. Newey, *Estimation with many instrumental variables*, Journal of Business and Economic Statistics **26** (2008), 398–422.
- [27] 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.
- [28] L.P. Hansen, *Large sample properties of Generalized Method of Moments Estimators*, Econometrica **50** (1982), no. 4, 1029–1054.
- [29] G. Kapetanios and M. Marcellino, *Factor-GMM estimation with large sets of possibly weak instruments*, Computational Statistics and Data Analysis **54** (2010), 2655–2675.
- [30] F. Kleibergen, *Testing parameters in GMM without assuming that they are identified*, Econometrica **73** (2005), 1103–1123.
- [31] 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.
- [32] R. Klein and F. Vella, *Estimating a class of triangular simultaneous equations models without exclusion restrictions*, Journal of Econometrics **154** (2010), 154–164.
- [33] L. Magnusson and S. Mavroeidis, *Identification using stability restrictions*, Forthcoming, Econometrica (2014).
- [34] S. Mavroeidis, *Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve*, Journal of Money, Credit and Banking **37** (2005), 421–448.
- [35] J. Nason and G.W. Smith, *Identifying the new Keynesian Phillips curve*, Journal of Applied Econometrics **23** (2008), 525–551.
- [36] A. Orphanides, *Monetary policy rules based on real-time data*, The American Economic Review **91** (2001), 964–985.
- [37] Z. Qu and P. Perron, *Estimating and testing structural changes in multivariate regressions*, Econometrica **75** (2007), 459–502.
- [38] R. Rigobon, *Identification through heteroskedasticity*, Review of Economics and Statistics **85** (2003), 777–792.

- [39] J.J. Rotemberg, *Sticky prices in the United States*, Journal of Political Economy **90** (1982), 1187–1211.
- [40] D. Staiger and J. Stock, *Instrumental variables regression with weak instruments*, Econometrica **65** (1997), 557–586.
- [41] J. Stock, J. Wright, and M. Yogo, *A survey of weak instruments and weak identification in Generalized Method of Moments*, Review of Economics and Statistics **20** (2002), 518–529.
- [42] J. H. Stock and M.W. Watson, *Implications of dynamic factor models for VAR analysis*, NBER Working Paper (2005).
- [43] J.H. Stock and M.W. Watson, *Has the business cycle changed and why?*, ch. in NBER Macroeconomics Annual, pp. 159–230, 2002.
- [44] J.H. Stock and J.H. Wright, *GMM with weak identification*, Econometrica **68** (2000), no. 5, 1055–1096.
- [45] J.B. Taylor, *Aggregate dynamics and staggered contracts*, Journal of Political Economy **88** (1980), 1–23.
- [46] C. Zhang, D. R. Osborn, and D. H. Kim, *The New Keynesian Phillips curve: from sticky inflation to sticky prices*, Journal of Money, Credit and Banking **40** (2008), 667–699.
- [47] C. Zhang, D.R. Osborn, and D.H. Kim, *Observed inflation forecasts and the New Keynesian Phillips curve*, Oxford Bulletin of Economics and Statistics **71** (2009), 375–398.

Appendix

This Appendix contains three parts (denoted A, B, and C). Appendix A provides additional discussions about identification and parameter instability in the linear IV regression model. Appendix B contains the proofs of the theoretical results of the paper. Appendix C contains the tables of results of the Monte-Carlo study.

A Identification and instability

We now discuss identification and parameter instability in the linear IV regression model.

A.1 Identification in the stable linear IV regression model

We start with an overview of the identification settings and associated asymptotic results commonly used in the stable linear IV model. The associated moment restrictions write

$$E [W_t (y_t - Z_t' \theta_z^0 - Y_t' \theta_y^0)] = \mathbf{0}, \quad (\text{A.1})$$

with y_t the dependent variable, Y_t the vector of p_2 endogenous variables, Z_t the vector of p_1 exogenous variables, W_t the vector of q (valid) instrumental variables including Z_t , $X_t = \text{vec} (Z_t, Y_t)$, $\theta^0 = \text{vec} (\theta_z^0, \theta_y^0)$, 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 ,

$$E [W_{k,t} (y_t - Z_t' \theta_z - Y_t' \theta_y)] = \frac{m_k(\theta)}{r_{k,T}}, \quad (\text{A.2})$$

where $\theta = \text{vec} (\theta_z, \theta_y)$, $m_k(\cdot)$ is a constant function, $r_{k,T}$ is a deterministic real sequence such that $r_{k,T} = 1$ or $r_{k,T} \xrightarrow{T} \infty$. The faster the unknown sequence $r_{k,T}$ diverges to infinity, the weaker the associated instrumental variable (IV), or moment condition is. Three cases of interest have been distinguished in the literature:

- When $r_{k,T} = 1$, the IV is strong. This is the standard case. When all the moment conditions are strong, standard inference procedures deliver \sqrt{T} -consistent estimators of the structural parameters.

- When $r_{k,T} \xrightarrow{T} \infty$ and $r_{k,T} = o(\sqrt{T})$, the IV is semi-strong.²⁶ When all the moment conditions are semi-strong at the same rate r_T , standard inference procedures are still asymptotically valid, and feasible without knowing the exact rate r_T . However, convergence rates of associated estimators are slower and depend on the degree of weakness, \sqrt{T}/r_T . When moment conditions are associated with different 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.
- 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), Dufour (2003), Andrews and Stock (2005), and references therein.

Our paper considers a framework where the exact identification pattern is unknown, and allowed to change over the sample. In sections 3 and 4, we consider cases where all instruments have the same identification strength. In section 5.1, we consider more general characterizations of identification strength.

A.2 Parameter instability in the stable linear IV regression model

In this section, we motivate why parameter instability may be relevant either in the reduced form, in the main equation of interest, or in both. Intuitively, if parameters in the main equation are "deep" parameters of an underlying structural model (such as preferences), they may not change in response to a change in policy specified by a reduced form. However, if these parameters are not "deep" parameters, they may change without any change in policy, or in response to the change in policy, which can lead to changes in the main equation that are either idiosyncratic or concomitant with the breaks in reduced form. Below we provide such an example.

Consider a reduced form as implicitly derived from a structural system, say

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

²⁶We use the terminology introduced by Andrews and Cheng (2012). Earlier literature referred to such cases as "near-weak" identification: see Hahn and Kuersteiner (2002).

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

$$\begin{aligned} y &= -Y\Gamma_{21} + Z\tilde{\Theta}_1 + U_1 \\ Y &= [Z\tilde{\Theta}\Gamma^{-1}]_{22} + [U\Gamma^{-1}]_{22} \stackrel{\text{def}}{=} Z\Pi + V. \end{aligned}$$

Thus, whenever either Γ_{12} , Γ_{22} , or $\tilde{\Theta}_2$ changes, so do Π and $\text{Var}(V)$, 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. We also propose in section 3.2 a framework where parameters in the main equation can change while reduced-form ones do not, which can happen if $\Gamma_{12} = 0$. In such a case, provided Γ_{22} is invertible, the reduced form equation writes

$$Y = Z\tilde{\Theta}_2\Gamma_{22}^{-1} + U_2\Gamma_{22}^{-1} \stackrel{\text{def}}{=} Z\Pi + V.$$

and breaks in the main equation (whenever Γ_{21} or $\tilde{\Theta}_1$ changes) do not transmit to the reduced form. Finally, whenever $\Gamma_{12} \neq 0$, any break in Γ_{21} or $\tilde{\Theta}_1$ will appear in both the main and the reduced form equations. This case is discussed in section 4.

B Proofs of the theoretical results

To simplify the exposition, all the proofs below are written for a single endogenous regressor ($p_2 = 1$) and no exogenous regressors, $p_1 = 0$. Complete proofs in presence of multiple endogenous regressors and exogenous regressors can be found in the Supplemental Appendix.

• **Proof of Theorem 1: Asymptotic properties of $\hat{\nu}$, $\hat{\theta}_{2SLS}$, $\hat{\theta}_{GMM}$, and $\hat{\theta}_{MOD}$.**

We assume that $\hat{T}^* < T^*$. The proof for $\hat{T}^* \geq T^*$ is similar and omitted for simplicity.

• (i) *Consistency of ν* . We drop the subscripts s on $\hat{\Pi}_1^s$, $\hat{\Pi}_2^s$ and Y_t^s , since for one endogenous regressor, $Y_t^s = Y_t$, both scalars. Let $\hat{v}_t \stackrel{\text{def}}{=} Y_t - W_t'\hat{\Pi}_1$ in interval $[1, \hat{T}^*]$, $\hat{v}_t \stackrel{\text{def}}{=} Y_t - W_t'\hat{\Pi}_2$ in interval $[\hat{T}^* + 1, T]$, and $d_t^* \stackrel{\text{def}}{=} \hat{v}_t - v_t$. By definition of the sum of squared residuals,

$$\sum_{t=1}^T \hat{v}_t^2 \leq \sum_{t=1}^T v_t^2 \Rightarrow 2 \sum_{t=1}^T v_t d_t^* + \sum_{t=1}^T (d_t^*)^2 \leq 0. \quad (\text{B.1})$$

We show consistency by contradiction in two steps. In step 1, we show that²⁷:

$$\sum_{t=1}^T (d_t^*)^2 = \mathcal{O}_P(Tr_T^{-2}) \text{ and } \sum_{t=1}^T v_t d_t^* = \mathcal{O}_P(T^{1/2}r_T^{-1}) \quad (\text{B.2})$$

²⁷If $r_T = T^{1/2}$, then $\sum_{t=1}^T (d_t^*)^2$ and $2 \sum_{t=1}^T v_t d_t^*$ are of the same order, and our argument does not apply.

Then, (B.2) $\Rightarrow \sum_{t=1}^T (d_t^*)^2 \gg 2 \sum_{t=1}^T v_t d_t^*$ (" \gg " means "dominates asymptotically")

$$\Rightarrow \text{plim}(r_T^2 T^{-1}) \sum_{t=1}^T (d_t^*)^2 \leq 0$$

$$\Rightarrow \text{plim}(r_T^2 T^{-1}) \sum_{t=1}^T (d_t^*)^2 = 0 \quad \text{by non-negativity} \quad (\text{B.3})$$

In step 2, we show that if $\hat{\nu} \not\xrightarrow{p} \nu^0$, then, with positive probability, $(r_T^2 T^{-1}) \sum_{t=1}^T (d_t^*)^2 > 0$, which contradicts (B.3).

- Step 1. Let $\Pi_{iT} \stackrel{\text{def}}{=} \Pi_i / r_{iT}$, and $\Psi_1^v(r) \stackrel{\text{def}}{=} T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} W_t v_t$, $\Psi_2^v(r) \stackrel{\text{def}}{=} T^{-1/2} \sum_{t=\lfloor Tr \rfloor+1}^T W_t v_t$, $\Psi_\Delta^v \stackrel{\text{def}}{=} T^{-1/2} \sum_{t=\lfloor T\hat{\nu} \rfloor+1}^{\lfloor T\nu^0 \rfloor} W_t v_t$, and $\Psi_i^v(\nu^0) \stackrel{\text{def}}{=} \Psi_i^v$, for $i = 1, 2$. Note that:

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

$$\sum_{t=1}^T v_t d_t^* = (\Pi_{1T} - \hat{\Pi}_1)' [T^{1/2} \Psi_1^v(\hat{\nu})] + (\Pi_{1T} - \hat{\Pi}_2)' [T^{1/2} \Psi_\Delta^v] + (\Pi_{2T} - \hat{\Pi}_2)' [T^{1/2} \Psi_2^v]. \quad (\text{B.4})$$

By Assumptions 1(i), (ii), and the functional CLT (FCLT) in Wooldridge and White (1988), Theorem 2.11, $\Psi_i^v(r) = \mathcal{O}_P(1)$, uniformly in $r \in [0, 1]$ (u.r. thereafter). Thus, $\Psi_1^v(\hat{\nu}) = \mathcal{O}_P(1)$, $\Psi_\Delta^v = \mathcal{O}_P(1)$, $\Psi_2^v = \mathcal{O}_P(1)$.

Recall $\hat{Q}_1(r) = T^{-1} \sum_{t=1}^{\lfloor Tr \rfloor} W_t W_t'$, $\hat{Q}_2(r) = \hat{Q} - \hat{Q}_1(r)$, and let $\hat{Q}_\Delta \stackrel{\text{def}}{=} T^{-1} \sum_{t=\lfloor T\hat{\nu} \rfloor+1}^{\lfloor T\nu^0 \rfloor} W_t W_t'$. Then, by Assumption 3, $\hat{Q}_i(r) = \mathcal{O}_P(1)$ and $Q_\Delta = \mathcal{O}_P(1)$, hence:

$$\Pi_{1T} - \hat{\Pi}_1 = -\hat{Q}_1^{-1}(\hat{\nu}) [T^{-1/2} \Psi_1^v(\hat{\nu})] = \mathcal{O}_P(1) \mathcal{O}_P(T^{-1/2}) = \mathcal{O}_P(T^{-1/2}). \quad (\text{B.5})$$

On the other hand, with $\Pi_T^\Delta \stackrel{\text{def}}{=} \Pi_{1T} - \Pi_{2T} = \mathcal{O}_P(r_T^{-1})$,

$$\Pi_{2T} - \hat{\Pi}_2 = -\hat{Q}_2^{-1}(\hat{\nu}) [T^{-1/2} \Psi_2^v(\hat{\nu})] - \hat{Q}_2^{-1}(\hat{\nu}) \hat{Q}_\Delta \Pi_T^\Delta = \mathcal{O}_P(r_T^{-1}) \quad (\text{B.6})$$

$$\Pi_{1T} - \hat{\Pi}_2 = \Pi_T^\Delta + (\Pi_{2T} - \hat{\Pi}_2) = \mathcal{O}_P(r_T^{-1}). \quad (\text{B.7})$$

Substituting (B.5)-(B.7) into (B.4) yields $\sum_{t=1}^T v_t d_t^* = \mathcal{O}_P(T^{1/2} r_T^{-1})$. Next, note that:

$$\begin{aligned} \sum_{t=1}^T (d_t^*)^2 &= \sum_{t=1}^{\hat{T}^*} (d_t^*)^2 + \sum_{t=\hat{T}^*+1}^{T^*} (d_t^*)^2 + \sum_{t=T^*+1}^T (d_t^*)^2 = (\Pi_{1T} - \hat{\Pi}_1)' T \hat{Q}_1(\hat{\nu}) (\Pi_{1T} - \hat{\Pi}_1) \\ &+ (\Pi_{1T} - \hat{\Pi}_2)' T \hat{Q}_\Delta (\Pi_{1T} - \hat{\Pi}_2) + (\Pi_{2T} - \hat{\Pi}_2)' T \hat{Q}_2(\nu^0) (\Pi_{2T} - \hat{\Pi}_2) \\ &= \mathcal{O}_P(1) + \mathcal{O}_P(r_T^{-1}) \mathcal{O}_P(T) \mathcal{O}_P(r_T^{-1}) + \mathcal{O}_P(r_T^{-1}) \mathcal{O}_P(T) \mathcal{O}_P(r_T^{-1}) = \mathcal{O}_P(T r_T^{-2}) \square \end{aligned}$$

- **Step 2.** Define $\|\cdot\|$ as the Euclidean norm for vectors, and $\|J\|$ as the square root of the maximum eigenvalue of $J'J$ for matrices. If $\hat{\nu} \xrightarrow{p} \nu^0$, then there exists $\eta \in (0, 1)$, such that with positive probability ϵ , $T^* - \hat{T}^* = [T\nu^0] - [T\hat{\nu}] \geq T\eta$. Because $\hat{Q}_{\nu^0, \eta} = T^{-1} \sum_{t=[T\nu^0]-T\eta+1}^{[T\nu^0]} W_t W_t'$ is a symmetric pd matrix, $\|\hat{Q}_{\nu^0, \eta}\| \geq \text{mineig}(\hat{Q}_{\nu^0, \eta})$. By Assumption 3, $\text{mineig}(\hat{Q}_{\nu^0, \eta}) > 0$ in probability limit. Let $\hat{\Pi}_2$ be the OLS estimator of the reduced form in interval $\{[T\nu^0 - T\eta] + 1, T\}$. Then, with positive probability ϵ ,

$$\begin{aligned} r_T^2 T^{-1} \sum_{t=1}^T (d_t^*)^2 &\geq r_T^2 T^{-1} \left(\sum_{t=T^*-T\eta+1}^{T^*} (d_t^*)^2 \right) = r_T (\Pi_{1T} - \hat{\Pi}_2)' \eta \hat{Q}_{\nu^0, \eta} r_T (\Pi_{1T} - \hat{\Pi}_2) \\ &\geq \|r_T \Pi_T^\Delta + r_T (\Pi_{2T} - \hat{\Pi}_2)\|^2 \eta \text{mineig}(\hat{Q}_{\nu^0, \eta}). \end{aligned} \quad (\text{B.8})$$

From (B.6), $r_T (\Pi_{2T} - \hat{\Pi}_2) = -[\hat{Q}_2(\nu^0 - \eta)]^{-1} \hat{Q}_{\nu^0, \eta} r_T \Pi_T^\Delta + o_P(1) \stackrel{\text{def}}{=} -\tilde{Q} r_T \Pi_T^\Delta + o_P(1)$, which can be substituted into (B.8).

- When $r_{1T} = r_{2T} = r_T$, we have:

$$\begin{aligned} r_T^2 T^{-1} \sum_{t=1}^T (d_t^*)^2 &\geq \|(I - \tilde{Q})(\Pi_1 - \Pi_2)\|^2 \eta \text{mineig}(\hat{Q}_{\nu^0, \eta}) \\ &\geq \|(\Pi_1 - \Pi_2)\|^2 [\text{mineig}^2(I - \tilde{Q})] \eta \text{mineig}(\hat{Q}_{\nu^0, \eta}) \end{aligned}$$

which is positive with probability ϵ for large T , because $Q_{\nu^0, \eta}$ is pd by Assumption 3, as well as $I - \tilde{Q} = I - [\hat{Q}_2(\nu^0 - \eta)]^{-1} \hat{Q}_{\nu^0, \eta} = [\hat{Q}_2(\nu^0 - \eta)]^{-1} Q_2(\nu^0)$ is pd.

- When $r_{iT} = o(r_{jT})$ ($i \neq j$), we have, using similar arguments,

$$r_T^2 T^{-1} \sum_{t=1}^T (d_t^*)^2 \geq \|(I - \tilde{Q})\Pi_i\|^2 \eta \text{mineig}(Q_{\nu^0, \eta})$$

which is positive with probability ϵ for large T , because $\text{rank}(\Pi_i) = 1$ by assumption 1(iii). \square

• (i) *Convergence rate of $\hat{\nu}$.* Since $\hat{\nu} \xrightarrow{p} \nu^0$, any break point estimator $\hat{T}^* = [T\hat{\nu}]$ is such that $T^* - \hat{T}^* \leq \epsilon^* T$, for some chosen $\epsilon^* > 0$ (thus uniformly over $\hat{\nu}$). We find the convergence rate by contradiction as well. For chosen $C^* > 0$, assume that $T^* - \hat{T}^* > C^* r_T^2$, uniformly over $\hat{\nu}$. Define SSR_1^* , SSR_2^* and SSR_3^* as the sum of squared residuals in the reduced form obtained with break points \hat{T}^* , T^* and (\hat{T}^*, T^*) respectively. Then, by definition of OLS,

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

We show that if $C^* r_T^2 < T^* - \hat{T}^* \leq \epsilon^* T$ 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 $T^* - \hat{T}^* \leq C^* r_T^2$,

and by symmetry of the argument, if $\hat{T}^* \geq T^*$, $\hat{T}^* - T^* \leq C^* r_T^2$, 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 $(1, \hat{T}^*, T)$, by $(\hat{\Pi}_1, \hat{\Pi}_\Delta, \tilde{\Pi}_2)$ the ones based on $(1, \hat{T}^*, T^*, T)$, and by $(\tilde{\Pi}_1, \tilde{\Pi}_2)$ the ones based on $(1, T^*, T)$. In step 1, we show that:

$$\begin{aligned} SSR_1^* - SSR_3^* &= (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' [T\hat{Q}_\Delta] (\tilde{\Pi}_2 - \hat{\Pi}_\Delta) - (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' [T\hat{Q}_\Delta \hat{Q}_2^{-1}(\nu^0) \hat{Q}_\Delta] (\tilde{\Pi}_2 - \hat{\Pi}_\Delta) \\ &\stackrel{\text{def}}{=} N_1^* - N_2^*. \end{aligned} \quad (\text{B.9})$$

By the same arguments, we also have:

$$\begin{aligned} SSR_2^* - SSR_3^* &= (\hat{\Pi}_1 - \hat{\Pi}_\Delta)' [T\hat{Q}_\Delta] (\hat{\Pi}_1 - \hat{\Pi}_\Delta) - (\hat{\Pi}_1 - \hat{\Pi}_\Delta)' [T\hat{Q}_\Delta \hat{Q}_1^{-1}(\nu^0) \hat{Q}_\Delta] (\hat{\Pi}_1 - \hat{\Pi}_\Delta) \\ &\stackrel{\text{def}}{=} N_3^* - N_4^*. \end{aligned}$$

In step 2, we show that N_1^* dominates N_2^*, N_3^*, N_4^* for large C^* and small ϵ^* . We also show that $N_1^* > 0$ at the limit, for large C^* and small ϵ^* , hence:

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

- **Step 1.** We have:

$$\begin{aligned} SSR_1^* - SSR_3^* &= \sum_{t=\hat{T}^*+1}^{T^*} [(Y_t - W_t' \hat{\Pi}_2)^2 - (Y_t - W_t' \hat{\Pi}_\Delta)^2] + \sum_{t=T^*+1}^T [(Y_t - W_t' \hat{\Pi}_2)^2 - (Y_t - W_t' \tilde{\Pi}_2)^2] \\ &= T^{1/2} (\hat{\Pi}_\Delta - \hat{\Pi}_2)' \left[2\Psi_\Delta^v + \hat{Q}_\Delta T^{1/2} [(\Pi_{1T} - \hat{\Pi}_2) + (\Pi_{1T} - \hat{\Pi}_\Delta)] \right] \\ &\quad + (\tilde{\Pi}_2 - \hat{\Pi}_2)' \left[\Psi_2^v(\nu^0) + \hat{Q}_2(\nu^0) T^{1/2} [(\Pi_{2T} - \hat{\Pi}_2) + (\Pi_{2T} - \tilde{\Pi}_2)] \right] \end{aligned} \quad (\text{B.10})$$

For simplicity, let $D \stackrel{\text{def}}{=} \hat{Q}_\Delta$ and $E \stackrel{\text{def}}{=} \hat{Q}_2(\nu^0)$. By definition of OLS,

$$(D + E) \hat{\Pi}_2 = T^{1/2} \Psi_2^v(\hat{\nu}) = T^{1/2} \Psi_\Delta^v + T^{1/2} \Psi_2^v(\nu^0) = D \hat{\Pi}_\Delta + E \tilde{\Pi}_2.$$

Thus, we have:

$$\begin{aligned} \hat{\Pi}_\Delta - \hat{\Pi}_2 &= \hat{\Pi}_\Delta - (D + E)^{-1} [D \hat{\Pi}_\Delta + E \tilde{\Pi}_2] = (D + E)^{-1} [(D + E) \hat{\Pi}_\Delta - D \hat{\Pi}_\Delta - E \tilde{\Pi}_2] \\ &= (D + E)^{-1} E (\hat{\Pi}_\Delta - \tilde{\Pi}_2), \\ \tilde{\Pi}_2 - \hat{\Pi}_2 &= (D + E)^{-1} [(D + E) \tilde{\Pi}_2 - D \hat{\Pi}_\Delta - E \tilde{\Pi}_2] = (D + E)^{-1} D (\tilde{\Pi}_2 - \hat{\Pi}_\Delta). \end{aligned}$$

Substituting this into (B.10) and noting that D, E are symmetric, we obtain:

$$\begin{aligned}
& T^{-1}(SSR_1^* - SSR_3^*) \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' E(D + E)^{-1} [2D(\hat{\Pi}_\Delta - \Pi_{1T}) + D(\Pi_{1T} - \hat{\Pi}_2) + D(\Pi_{1T} - \hat{\Pi}_\Delta)] \\
&+ (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' D(D + E)^{-1} [2E(\tilde{\Pi}_2 - \Pi_{2T}) + E[(\Pi_{2T} - \hat{\Pi}_2) + E(\Pi_{2T} - \tilde{\Pi}_2)]] \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' E(D + E)^{-1} D(\hat{\Pi}_\Delta - \hat{\Pi}_2) + (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' D(D + E)^{-1} E(\tilde{\Pi}_2 - \hat{\Pi}_2) \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' E(D + E)^{-1} D(D + E)^{-1} E(\hat{\Pi}_\Delta - \tilde{\Pi}_2) \\
&+ (\tilde{\Pi}_2 - \hat{\Pi}_\Delta)' D(D + E)^{-1} E(D + E)^{-1} D(\tilde{\Pi}_2 - \hat{\Pi}_\Delta) \\
&= (\hat{\Pi}_\Delta - \tilde{\Pi}_2)' [E(D + E)^{-1} D(D + E)^{-1} E + D(D + E)^{-1} E(D + E)^{-1} D] (\hat{\Pi}_\Delta - \tilde{\Pi}_2).
\end{aligned}$$

To prove (B.9), we are left with showing that:

$$E(D + E)^{-1} D(D + E)^{-1} E + D(D + E)^{-1} E(D + E)^{-1} D = D - D(D + E)^{-1} D.$$

Let $F_1 \stackrel{\text{def}}{=} E(D + E)^{-1}$ and $F_2 \stackrel{\text{def}}{=} D(D + E)^{-1}$. Then $F_1 + F_2 = I$, $F_1' + F_2' = I$ and $F_1 = I - F_2$, so:

$$\begin{aligned}
& E(D + E)^{-1} D(D + E)^{-1} E + D(D + E)^{-1} E(D + E)^{-1} D = F_1 D F_1' + F_2 E F_2' \\
&= (I - F_2) D (I - F_2') + F_2 E F_2' = D - DF_2' - F_2 D + F_2 D F_2' + F_2 E F_2' \\
&= D - D(D + E)^{-1} D - D(D + E)^{-1} D + D(D + E)^{-1} (D + E) (D + E)^{-1} D \\
&= D - D(D + E)^{-1} D.
\end{aligned}$$

- **Step 2.** Since $T^* - \hat{T}^* \leq \epsilon^* T$, by Assumption 3,

$$(D + E)^{-1} D = \hat{Q}_2^{-1}(\hat{\nu}) \hat{Q}_\Delta = \mathcal{O}_P(1) \mathcal{O}_P(\epsilon^*) = \mathcal{O}_P(\epsilon^*) \Rightarrow N_1^* \gg N_2^*,$$

for ϵ^* small enough. Similarly, $N_3^* \gg N_4^*$. To show that $N_1^* \gg N_3^*$, we need to compare $(\tilde{\Pi}_2 - \hat{\Pi}_\Delta)$ and $(\hat{\Pi}_1 - \hat{\Pi}_\Delta)$. Since $\hat{\Pi}_1$ and $\hat{\Pi}_\Delta$ are both subsample estimators of Π_{1T} ,

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

Since $\tilde{\Pi}_2$ is the estimator of Π_{2T} in subsample $[T^* + 1, T]$, $\tilde{\Pi}_2 - \Pi_{2T} = \mathcal{O}_P(T^{-1/2})$, so

$$\tilde{\Pi}_2 - \hat{\Pi}_\Delta = (\tilde{\Pi}_2 - \Pi_{2T}) - (\hat{\Pi}_\Delta - \Pi_{1T}) - \Pi_T^\Delta = \mathcal{O}_P(T^{-1/2}) - \Pi_T^\Delta = \mathcal{O}_P(r_T^{-1}).$$

Thus, $(\tilde{\Pi}_2 - \hat{\Pi}_\Delta) \gg (\hat{\Pi}_1 - \hat{\Pi}_\Delta)$, implying that $N_1^* \gg N_3^*$, and so $N_1^* \gg N_j^*$, for $j = 2, 3, 4$.

We show that $N_1^* > 0$ at the limit, for small ϵ^* and large enough C^* . For large enough C^* ,

$$T \hat{Q}_\Delta / (T^* - \hat{T}^*) = \sum_{t=\hat{T}^*+1}^{T^*} W_t W_t' / (T^* - \hat{T}^*) = \mathcal{O}_P(1)$$

$$\begin{aligned} \Rightarrow r_T^2 N_1^* &= r_T^2 (T^* - \hat{T}^*) [\mathcal{O}_P(T^{-1/2}) - \Pi_T^\Delta]' [T \hat{Q}_\Delta / (T^* - \hat{T}^*)] [\mathcal{O}_P(T^{-1/2}) - \Pi_T^\Delta] \\ &\geq C^* r_T \Pi_T^{\Delta'} [Q_1 - Q_1(\hat{\nu})] r_T \Pi_T^\Delta + o_P(1). \end{aligned}$$

By Assumption 3, $\hat{Q}_\Delta = \hat{Q}_1 - \hat{Q}_1(\hat{\nu})$ is pd uniformly in $\hat{\nu}$, and so $N_1^* > 0$ at the limit if $r_T \Pi_T^\Delta \not\rightarrow 0$. When $r_{1T} = r_{2T}$, $r_T \Pi_T^\Delta = \Pi_1 - \Pi_2 \neq 0$ by construction. Similarly, when $r_{iT} = o(r_{jT})$, $|r_T \Pi_T^\Delta| \rightarrow |\Pi_i| \neq 0$.

• (ii) *Asymptotic distribution of 2SLS.* Let \overline{W}_1 be the $(\hat{T}^* \times q)$ matrix with rows $W'_1, \dots, W'_{\hat{T}^*}$, \overline{W}_2 the $((T - \hat{T}^*) \times q)$ matrix with rows $W'_{\hat{T}^*+1}, \dots, W'_T$ and,

$$\overline{W} = \begin{pmatrix} \overline{W}_1 & O_{\hat{T}^* \times q} \\ O_{(T - \hat{T}^*) \times q} & \overline{W}_2 \end{pmatrix} \quad \text{and} \quad \overline{W}^0 = \begin{pmatrix} \overline{W}_1^0 & O_{T^* \times q} \\ O_{(T - T^*) \times q} & \overline{W}_2^0 \end{pmatrix},$$

\overline{W} and \overline{W}^0 are the diagonal partition at \hat{T}^* and T^* , respectively. Then $\hat{Y} \stackrel{\text{def}}{=} \text{vec}(\hat{Y}_1, \dots, \hat{Y}_T)$ can be written as $\hat{Y} = \overline{W} \hat{\Pi}_{vec}$, with $\hat{\Pi}_{vec} = \text{vec}(\hat{\Pi}_1, \hat{\Pi}_2)$. Let $y \stackrel{\text{def}}{=} \text{vec}(y_1, \dots, y_T)$, $Y \stackrel{\text{def}}{=} \text{vec}(Y_1, \dots, Y_T)$ and $U \stackrel{\text{def}}{=} \text{vec}(u_1, \dots, u_T)$, 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:

$$T^{1/2} r_T^{-1} (\hat{\theta} - \theta^0) = (r_T^2 T^{-1} \hat{Y}' \hat{Y})^{-1} r_T T^{-1/2} \hat{Y}' \tilde{U}. \quad (\text{B.11})$$

In step 1, we show that:

$$r_T^2 T^{-1} \hat{Y}' \hat{Y} = r_T^2 [\Pi'_{1T} Q_1 \Pi_{1T} + \Pi'_{2T} Q_2 \Pi_{2T}] + o_P(1) \quad (\text{B.12})$$

$$r_T T^{-1/2} \hat{Y}' \tilde{U} = r_T [\Pi'_{1T} \Psi_1^u + \Pi'_{2T} \Psi_2^u] + o_P(1), \quad (\text{B.13})$$

where $(\Psi_i^u(\cdot), \Psi_\Delta^u(\cdot))$ are defined as $(\Psi_i^v(\cdot), \Psi_\Delta^v(\cdot))$, with u_t replacing v_t , and $\Psi_i^j \stackrel{\text{def}}{=} \Psi_i^j(\nu^0)$, $j \in \{u, v\}$, $i = 1, 2$. Equations (B.12) and (B.13) imply that $T^{1/2} r_T^{-1} (\hat{\theta} - \theta^0)$ has the same asymptotic distribution as if ν^0 was known. In step 2, we derive this asymptotic distribution.

- **Step 1.** Consider $r_T^2 T^{-1} \hat{Y}' \hat{Y} = r_T^2 \hat{\Pi}'_{vec} (T^{-1} \overline{W}' \overline{W}) \hat{\Pi}_{vec}$.

By Theorem 1(i) and Assumption 3,

$$\hat{Q}_\Delta = T^{-1} \sum_{\hat{T}^*+1}^{T^*} W_t W_t' = \frac{T^* - \hat{T}^*}{T} \left(\frac{1}{T^* - \hat{T}^*} \sum_{\hat{T}^*+1}^{T^*} W_t W_t' \right) = \mathcal{O}_P(r_T^2/T) \mathcal{O}_P(1) = o_P(1),$$

$$\text{and so, } T^{-1} \overline{W}' \overline{W} - T^{-1} \overline{W}^0{}' \overline{W}^0 = o_P(1). \quad (\text{B.14})$$

Next, we analyze $\hat{\Pi}_{vec}$, or equivalently $\hat{\Pi}_i$ for $i = 1, 2$. Under Assumptions 1(i), (ii), and the FCLT in Wooldridge and White (1988), Theorem 2.11,

$$\Psi_{\Delta}^v = \left[\frac{T^* - \hat{T}^*}{T^*} \right]^{1/2} \left[(T^* - \hat{T}^*)^{-1/2} \sum_{t=\hat{T}^*+1}^{T^*} W_t v_t \right] = \mathcal{O}_P(r_T/T^{1/2})\mathcal{O}_P(1) = o_P(1),$$

so $\Psi_1^v(\hat{\nu}) - \Psi_1^v = -\Psi_{\Delta}^v = o_P(1)$ and $\Psi_2^v(\hat{\nu}) - \Psi_2^v = \Psi_{\Delta}^v = o_P(1)$. From these and (B.14),

$$\begin{aligned} T^{1/2}(\hat{\Pi}_1 - \Pi_{1T}) &= \hat{Q}_1^{-1}(\hat{\nu})\Psi_1^v(\hat{\nu}) = Q_1^{-1}\Psi_1^v + o_P(1) = \mathcal{O}_P(1) \\ T^{1/2}(\hat{\Pi}_2 - \Pi_{2T}) &= \hat{Q}_2^{-1}(\hat{\nu})\Psi_2^v(\hat{\nu}) + \hat{Q}_2^{-1}(\hat{\nu})\hat{Q}_{\Delta}T^{1/2}\Pi_T^{\Delta} \\ &= Q_2^{-1}\Psi_2^v + \mathcal{O}_P(1)\mathcal{O}_P(r_T^2T^{-1})\mathcal{O}_P(T^{1/2}r_T^{-1}) = \mathcal{O}_P(1). \end{aligned}$$

It follows that $\hat{\Pi}_1 = \Pi_{1T} + o_P(1)$, $\hat{\Pi}_2 = \Pi_{2T} + o_P(1)$, and with (B.14) these imply:

$$r_T^2T^{-1}\hat{Y}'\hat{Y} = r_T^2 \left(T^{-1} \sum_{t=1}^{T^*} \hat{Y}_t^2 + T^{-1} \sum_{t=T^*+1}^T \hat{Y}_t^2 \right) + o_P(1) = r_T^2[\Pi'_{1T}Q_1\Pi_{1T} + \Pi'_{2T}Q_2\Pi_{2T}] + o_P(1).$$

The latter proves (B.12). Next, we show (B.13). By the above, defining $\tilde{\Psi}_i, \tilde{\Psi}_i(\cdot), \tilde{\Psi}_{\Delta}$ as $\Psi_i^u, \Psi_i^u(\cdot), \Psi_{\Delta}^u$, but with \tilde{u}_t replacing u_t , for $i = 1, 2$,

$$r_T T^{-1/2} \hat{Y}' \tilde{U} = r_T \hat{\Pi}'_1 \tilde{\Psi}_1(\hat{\nu}) + r_T \hat{\Pi}'_2 \tilde{\Psi}_2(\hat{\nu}) = r_T \Pi'_{1T} \tilde{\Psi}_1(\hat{\nu}) + r_T \Pi'_{2T} \tilde{\Psi}_2(\hat{\nu}). \quad (\text{B.15})$$

First, $\|T^{-1/2}\overline{W}\tilde{U} - T^{-1/2}\overline{W}^0\tilde{U}\| = \|T^{-1/2}\sum_{t=\hat{T}^*+1}^{T^*} W_t \tilde{u}_t\|$. Also, note that:

$$\tilde{u}_t = u_t + (Y_t - \hat{Y}_t)\theta^0 = \begin{cases} u_t + v_t\theta^0 - W'_t(\hat{\Pi}_1 - \Pi_{1T})\theta^0, & t \leq \hat{T}^* \\ u_t + v_t\theta^0 - W'_t(\hat{\Pi}_2 - \Pi_{1T})\theta^0, & \hat{T}^* + 1 \leq t \leq T^* \\ u_t + v_t\theta_y^0 - W'_t(\hat{\Pi}_2 - \Pi_{2T})\theta^0, & t > T^* \end{cases}$$

$\Psi_i^{uv}, \Psi_i^{uv}(\cdot), \Psi_{\Delta}^{uv}$ are defined as before, but with u_t replaced by $(u_t + v_t\theta^0)$, for $i = 1, 2$, $\tilde{\Psi}_{\Delta} \stackrel{\text{def}}{=} T^{-1/2} \sum_{t=\hat{T}^*+1}^{T^*} W_t \tilde{u}_t$, and so,

$$\tilde{\Psi}_{\Delta} = \Psi_{\Delta}^{uv} - \hat{Q}_{\Delta} [T^{1/2}(\hat{\Pi}_2 - \Pi_{1T})] \theta^0 = \mathcal{O}_P(r_T T^{-1/2}) - \mathcal{O}_P(r_T^2 T^{-1})\mathcal{O}_P(T^{1/2} r_T^{-1}) = o_P(1).$$

Hence $\tilde{\Psi}_i(\hat{\nu}) - \tilde{\Psi}_i = -\tilde{\Psi}_{\Delta} = o_P(1)$. Next, we analyze $\tilde{\Psi}_i$. $\tilde{\Psi}_1 = \Psi_1^{uv} - \hat{Q}_1(\nu^0) [T^{1/2}(\hat{\Pi}_1 - \Pi_{1T})] \theta^0 + o_P(1) = \Psi_1^{uv} - \Psi_1^v \theta^0 + o_P(1) = \Psi_1^u + o_P(1)$. Similarly, $\tilde{\Psi}_2 = \Psi_2^u + o_P(1)$. From this, $\tilde{\Psi}_i(\hat{\nu}) = \tilde{\Psi}_i + o_P(1)$, and from (B.15), $r_T T^{-1/2} \hat{Y}' \tilde{U} = r_T [\Pi'_{1T} \Psi_1^u + \Pi'_{2T} \Psi_2^u] + o_P(1)$, which coincides with (B.13). In step 2, we derive the limits in (B.13).

- **Step 2.** When $r_{1T} = r_{2T} = r_T$. Then (B.13) becomes:

$$\begin{aligned} T^{1/2}r_T^{-1}(\hat{\theta} - \theta^0) &= [\Pi_1'Q_1\Pi_1 + \Pi_2'Q_2\Pi_2]^{-1}r_T [\Pi_{1T}'\Psi_1^u + \Pi_{2T}'\Psi_2^u] + o_P(1) \\ &= (A_1 + A_2)^{-1}r_T [\Pi_{1T}'\Psi_1^u + \Pi_{2T}'\Psi_2^u] + o_P(1). \end{aligned} \quad (\text{B.16})$$

By Assumption 1(ii), $\Psi_1^u \perp \Psi_2^u$ asymptotically, so by the CLT, $r_T[\Pi_{1T}'\Psi_1^u + \Pi_{2T}'\Psi_2^u] \xrightarrow{d} \mathcal{N}(0, B_1 + B_2)$. Thus, $T^{1/2}r_T^{-1}(\hat{\theta} - \theta^0) \xrightarrow{d} \mathcal{N}(0, (A_1 + A_2)^{-1}(B_1 + B_2)(A_1 + A_2)^{-1})$.

- When $r_{iT} = o(r_{jT})$ and wlog $i = 1$, we have $r_T = r_{1T}$, $r_T(\Pi_{1T}, \Pi_{2T}) \rightarrow (\Pi_1, O_{q \times 1})$, and $T^{1/2}r_T^{-1}(\hat{\theta} - \theta^0) = [\Pi_1'Q_1\Pi_1]^{-1}\Pi_1'\Psi_1^u + o_P(1) \xrightarrow{d} A_1^{-1}\mathcal{N}(0, B_1) = \mathcal{N}(0, A_1^{-1}B_1A_1^{-1})$.

• (ii) *Asymptotic distribution of GMM.*

$$\begin{aligned} \hat{\theta}_{GMM} &= \left[\frac{Y'W}{T} \hat{S}_u^{-1} \frac{W'Y}{T} \right]^{-1} \frac{Y'W}{T} \hat{S}_u^{-1} \frac{W'}{T} (Y\theta^0 + U) \\ \Leftrightarrow T^{1/2}r_T^{-1}(\hat{\theta}_{GMM} - \theta^0) &= \left[r_T \frac{Y'W}{T} \hat{S}_u^{-1} r_T \frac{W'Y}{T} \right]^{-1} r_T \frac{Y'W}{T} \hat{S}_u^{-1} \frac{W'U}{T^{1/2}} \end{aligned}$$

$$\begin{aligned} \text{with } r_T T^{-1} Y'W &= r_T \Pi_{1T}' \hat{Q}_1(\nu^0) + r_T \Pi_{2T}' \hat{Q}_2(\nu^0) + r_T T^{-1/2} \Psi_1^v + r_T T^{-1/2} r_T \Psi_2^v + o_P(1) \\ &= r_T \Pi_{1T}' Q_1 + r_T \Pi_{2T}' Q_2 + o_P(1) \quad \text{since } \Psi_i^v = \mathcal{O}_P(1). \end{aligned}$$

Also, as before, $T^{-1/2}W'U = \Psi_1^u + \Psi_2^u \xrightarrow{d} \mathcal{N}(0, S_u)$. Let $\mu_i \stackrel{\text{def}}{=} Q_i \Pi_i$ ($i = 1, 2$).

- When $r_{1T} = r_{2T} = r_T$. Then $r_T T^{-1} Y'W = \Pi_1' Q_1 + \Pi_2' Q_2 + o_P(1) = \mu_1' + \mu_2' + o_P(1)$. Hence, using the optimal GMM estimator with $\hat{S}_u \xrightarrow{p} \text{AVar}(T^{1/2}g_T(\theta^0)) = S_u$,

$$T^{1/2}r_T^{-1}(\hat{\theta}_{GMM} - \theta^0) \xrightarrow{d} \mathcal{N}(0, [(\mu_1 + \mu_2)' S_u^{-1} (\mu_1 + \mu_2)]^{-1}) = \mathcal{N}(0, V_{GMM}).$$

- When $r_{1T} = o(r_{2T}) = r_T$, $r_T \Pi_{2T} \rightarrow 0$, $r_T T^{-1} Y'W \xrightarrow{p} \mu_1'$, and $T^{1/2}r_T^{-1}(\hat{\theta}_{GMM} - \theta^0) \xrightarrow{d} \mathcal{N}(0, [\mu_1' S_u^{-1} \mu_1]^{-1})$.

• (ii) *Asymptotic distribution of MOD.* For the weighting matrix $(\hat{S}_u^a)^{-1}$, we have:

$$\hat{\theta}_{MOD} = \left[Y' \check{W} (\hat{S}_u^a)^{-1} \check{W}' Y \right]^{-1} Y' \check{W} (\hat{S}_u^a)^{-1} \check{W}' (Y\theta^0 + U),$$

where $\check{W}' = [\bar{W}_1' / \hat{T}^*, \bar{W}_2' / (T - \hat{T}^*)]$ and \bar{W}_1, \bar{W}_2 defined in part (ii). So,

$$T^{1/2}r_T^{-1}[\hat{\theta}_{MOD} - \theta^0] = \left[r_T Y' \check{W} (\hat{S}_u^a)^{-1} \check{W}' Y \right]^{-1} r_T Y' \check{W} (\hat{S}_u^a)^{-1} [\check{W}' U \sqrt{T}] \quad (\text{B.17})$$

Now, $Y' \check{W} = [\sum_{t=1}^{\hat{T}^*} Y_t W_t' / \hat{T}^*, \sum_{t=\hat{T}^*+1}^T Y_t W_t' / (T - \hat{T}^*)]$, and from the results for $\hat{\theta}_{GMM}$,

$$r_T \sum_{t=1}^{\hat{T}^*} Y_t W_t' / \hat{T}^* = r_T \frac{T}{\hat{T}^*} \left[T^{-1} \sum_{t=1}^{\hat{T}^*} Y_t W_t' \right] = r_T \Pi_{1T}' Q_1 / \nu^0 + o_P(1).$$

Similarly, $r_T \sum_{t=\hat{T}^*+1}^T Y_t W_t' / (T - \hat{T}^*) = r_T \Pi_{2T}' Q_2 / (1 - \nu^0) + o_P(1)$, hence:

$$r_T Y' \check{W} = [r_T \Pi_{1T}' Q_1 / \nu^0, r_T \Pi_{2T}' Q_2 / (1 - \nu^0)] + o_P(1) \stackrel{\text{def}}{=} K + o_P(1). \quad (\text{B.18})$$

In addition, $\check{W}' U \sqrt{T} = [\sqrt{T} \sum_{t=1}^{\hat{T}^*} u_t W_t' / \hat{T}^*, \sqrt{T} \sum_{t=\hat{T}^*+1}^T u_t W_t' / (T - \hat{T}^*)]$, and using arguments similar to the proof for 2SLS, the difference between the quantities above evaluated at the true break T^* and at the estimated break \hat{T}^* is $o_P(1)$, so:

$$\begin{aligned} \sqrt{T} \sum_{t=1}^{\hat{T}^*} u_t W_t' / \hat{T}^* &= T(\hat{T}^*)^{-1} \Psi_1^u + o_P(1) \xrightarrow{d} \mathcal{N}(0, S_{u,1} / (\nu^0)^2) \\ \sqrt{T} \sum_{t=\hat{T}^*+1}^T u_t W_t' / (T - \hat{T}^*) &\xrightarrow{d} \mathcal{N}(0, S_{u,2} / (1 - \nu^0)^2). \end{aligned}$$

Because $\sqrt{T} \sum_{t=1}^{\hat{T}^*} u_t W_t' / \hat{T}^* \perp \sqrt{T} \sum_{t=\hat{T}^*+1}^T u_t W_t' / (T - \hat{T}^*)$ asymptotically by Assumption 1(ii), $\check{W}' U \sqrt{T} \xrightarrow{d} \mathcal{N}(0, S_u^a)$, where $S_u^a = \text{diag}[S_{u,1} / (\nu^0)^2, S_{u,2} / (1 - \nu^0)^2]$. From (B.17)-(B.18), $T^{1/2} r_T^{-1} [\hat{\theta}_{MOD} - \theta^0]$ is asymptotically normally distributed with mean 0 and asymptotic variance-covariance matrix V_{MOD} with

$$\hat{V}_{MOD} = [K(\hat{S}_u^a)^{-1} K']^{-1} K(\hat{S}_u^a)^{-1} S_u^a (\hat{S}_u^a)^{-1} K' [K(\hat{S}_u^a)^{-1} K']^{-1} \xrightarrow{p} V_{MOD} \quad \text{and} \quad \hat{S}_u^a \xrightarrow{p} S_u^a.$$

- When $r_T = r_{1T} = r_{2T}$, $K = [\mu'_1 / \nu^0, \mu'_2 / (1 - \nu^0)] + o(1)$, and

$$V_{MOD} = [\mu'_1 (S_{u,1})^{-1} \mu_1 + \mu'_2 (S_{u,2})^{-1} \mu_2]^{-1}.$$

- When $r_T = r_{1T} = o(r_{2T})$, $K = [\mu'_1 / \nu^0, O_{1 \times q}]$, and $V_{MOD} = [\mu'_1 (S_{u,1})^{-1} \mu_1]^{-1}$. ■

• Proof of Theorem 2: Efficiency of estimated structural parameters

We prove the following equivalent results.

• Case (a): under Assumptions 1 to 4, and when $r_{1T} = r_{2T}$,

(i) $V_{MOD} \leq V_{2SLS}$; $V_{MOD} \leq V_{GMM}$, and $V_{MOD} = V_{GMM}$ iff $S_{u,1}^{-1} Q_1 \Pi_1^a = S_{u,2}^{-1} Q_2 \Pi_2^a$;

(ii) In general, $V_{GMM} \gg V_{2SLS}$. However, under Assumptions 5 and 6, $V_{GMM} > V_{2SLS}$.

• Case (b): under Assumptions 1 to 4, when $r_{iT} = o(r_{jT})$,

(i) $V_{MOD} \leq V_{2SLS}$; $V_{MOD} \leq V_{GMM}$, with equality only for $S_{u,i}^{-1} Q_i \Pi_i^a = S_{u,j}^{-1} Q_j \Pi_j^a$ (note that if $p_1 = 0$, then the inequality is strict);

(ii) In general, $V_{GMM} \gg V_{2SLS}$. However, under Assumptions 5 and 6, $V_{GMM} > V_{2SLS}$.

- Proof of case (a). Recall that:

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

(i) By construction, $\hat{\theta}_{MOD}$ is the optimal version of $\hat{\theta}_{2SLS}$, so $V_{MOD} \leq V_{2SLS}$. In step 1, we show that $V_{MOD}^{-1} \geq V_{GMM}^{-1}$. In step 2, we establish the conditions under which $V_{MOD} = V_{GMM}$.

- **Step 1.** Let $S_u = S_{u,1} + S_{u,2}$, $\mu = \text{vec}(\mu_1, \mu_2)$ (with $\mu_i = Q_i \Pi_i$), and $L = S_{u,2} S_{u,1}^{-1}$.

$$\begin{aligned} V_{MOD}^{-1} - V_{GMM}^{-1} &= \mu_1' S_{u,1}^{-1} \mu_1 + \mu_2' S_{u,2}^{-1} \mu_2 - (\mu_1 + \mu_2)' S_u^{-1} (\mu_1 + \mu_2) \\ &= \mu' \begin{pmatrix} S_{u,1}^{-1} & O \\ O & S_{u,2}^{-1} \end{pmatrix} \mu - \mu' \begin{pmatrix} S_u^{-1} & S_u^{-1} \\ S_u^{-1} & S_u^{-1} \end{pmatrix} \mu \\ &= \mu' \begin{pmatrix} S_{u,1}^{-1} - S_u^{-1} & -S_u^{-1} \\ -S_u^{-1} & S_{u,2}^{-1} - S_u^{-1} \end{pmatrix} \mu = \mu' \begin{pmatrix} S_u^{-1} L & -S_u^{-1} \\ -S_u^{-1} & S_u^{-1} L^{-1} \end{pmatrix} \mu \stackrel{\text{def}}{=} f(\mu), \end{aligned}$$

because $S_{u,1}^{-1} - S_u^{-1} = S_u^{-1} [S_u - S_{u,1}] S_{u,1}^{-1} = S_u^{-1} S_{u,2} S_{u,1}^{-1} = S_u^{-1} L$,
and $S_{u,2}^{-1} - S_u^{-1} = S_u^{-1} [S_u - S_{u,2}] S_{u,2}^{-1} = S_u^{-1} S_{u,1} S_{u,2}^{-1} = S_u^{-1} L^{-1}$.

It is well known that for any symmetric matrix such that $M = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$, with A, C square symmetric matrices of the same dimension, and C pd, M is psd (positive semi-definite) iff $(A - BC^{-1}B)$ (the Schur complement of C) is psd. Let $M = \begin{pmatrix} S_u^{-1} L & -S_u^{-1} \\ -S_u^{-1} & S_u^{-1} L^{-1} \end{pmatrix}$. $S_u^{-1} L^{-1} = S_{u,2}^{-1} - (S_{u,1} + S_{u,2})^{-1}$ is pd by construction, and its Schur's complement is $S_u^{-1} L - S_u^{-1} L S_u^{-1} = O$. Thus, M is psd, and $f(\mu) \geq 0$. This implies that $V_{MOD}^{-1} \geq V_{GMM}^{-1}$, and therefore $V_{MOD} \leq V_{GMM}$.

- **Step 2.** Since $f(\mu) \geq 0$ and convex, we solve the first-order conditions to find its minimum:

$$\frac{\partial f(\mu)}{\partial \mu_1} = 2S_u^{-1}(L\mu_1 - \mu_2) \quad \text{and} \quad \frac{\partial f(\mu)}{\partial \mu_2} = 2S_u^{-1}(L^{-1}\mu_2 - \mu_1) = 0.$$

Then $V_{GMM} = V_{MOD} \Leftrightarrow L\mu_1^* = \mu_2^* \Leftrightarrow S_{u,1}^{-1} Q_1 \Pi_1 = S_{u,2}^{-1} Q_2 \Pi_2$.

(ii) In general, $V_{GMM} >< V_{2SLS}$, but under Assumptions 5,6, we show $V_{GMM} > V_{2SLS}$.

Under Assumption 6, $S_{u,i} = \Phi_u Q_i$, with Φ_u scalar, and so

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

Under Assumption 5, $Q_1 = \nu^0 Q$ and $Q_2 = (1 - \nu^0)Q$, and

$$\begin{aligned} V_{2SLS}^{-1}/\Phi_u &= \nu^0 \Pi_1' Q \Pi_1 + (1 - \nu^0) \Pi_2' Q \Pi_2 \\ V_{GMM}^{-1}/\Phi_u &= [\Pi_1' \nu^0 + \Pi_2' (1 - \nu^0)] Q [\Pi_1 \nu^0 + \Pi_2 (1 - \nu^0)] \\ \Rightarrow V_{GMM}^{-1}/\Phi_u - V_{2SLS}^{-1}/\Phi_u &= -\nu^0 (1 - \nu^0) (\Pi_1 - \Pi_2)' Q (\Pi_1 - \Pi_2) < 0, \end{aligned}$$

and we get $V_{GMM} > V_{2SLS}$.

- Proof of Case (b): wlog let $r_T = r_{1T} = o(r_{2T})$.

$$\begin{aligned} V_{2SLS} &= (\Pi_1' Q_1 \Pi_1)^{-1} (\Pi_1' S_{u,1} \Pi_1) (\Pi_1' Q_1 \Pi_1)^{-1} \\ V_{GMM} &= [\Pi_1' Q_1 (S_{u,1} + S_{u,2})^{-1} Q_1 \Pi_1]^{-1} \\ V_{MOD} &= [\Pi_1' Q_1 (S_{u,1})^{-1} Q_1 \Pi_1]^{-1}. \end{aligned}$$

(i) By construction, $\hat{\theta}_{MOD}$ is the optimal version of $\hat{\theta}_{2SLS}$, so $V_{MOD} \leq V_{2SLS}$. Also, $V_{MOD}^{-1} - V_{GMM}^{-1} = \Pi_1' Q_1 [S_{u,1}^{-1} - (S_{u,1} + S_{u,2})^{-1}] Q_1 \Pi_1 > 0$, thus $V_{MOD} < V_{GMM}$.

(ii) In general, $V_{GMM} > V_{2SLS}$, but under Assumption 6, $S_{u,i} = \Phi_u Q_i$, and then,

$$V_{2SLS}^{-1}/\Phi_u = \Pi_1' Q_1 \Pi_1 \quad \text{and} \quad V_{GMM}^{-1}/\Phi_u = (\Pi_1' Q_1) (Q_1 + Q_2)^{-1} (Q_1 \Pi_1).$$

Thus, $V_{2SLS}^{-1} - V_{GMM}^{-1} = \Phi_u \Pi_1' Q_1 [Q_1^{-1} - (Q_1 + Q_2)^{-1}] Q_1 \Pi_1 \geq 0$, and $V_{2SLS} < V_{GMM}$. ■

• Proof of Theorem 3: Test for a break in the reduced form

- (i) Let $\Pi_T \stackrel{\text{def}}{=} \Pi_i / r_{iT}$ ($i = 1, 2$), and $\iota_2 \stackrel{\text{def}}{=} (1, 1)'$. Then,

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

Under H_0 , $T^{1/2} (\hat{\Pi}_1(\nu) - \Pi_T) = \hat{Q}_1(\nu) \Psi_1^v(\nu)$. Under Assumptions 3 and 5, $\hat{Q}_1(\nu) \xrightarrow{p} \nu Q$. Under Assumptions 1(ii) and 4, by the FCLT, $\Psi_1^v(\nu) \Rightarrow S_v^{1/2} \mathcal{B}_q(\nu)$. Hence,

$$T^{1/2} (\hat{\Pi}_1(\nu) - \Pi_T) \Rightarrow [Q^{-1} S_v^{1/2}] \mathcal{B}_q(\nu) / \nu \stackrel{\text{def}}{=} G_{RF}^{1/2} \mathcal{B}_q(\nu) / \nu \quad \text{with} \quad G_{RF} \stackrel{\text{def}}{=} Q^{-1} S_v Q^{-1}.$$

Using similar arguments, we also have:

$$T^{1/2} (\hat{\Pi}_2(\nu) - \Pi_T) \Rightarrow [Q^{-1} S_v^{1/2}] [\mathcal{B}_q(1) - \mathcal{B}_q(\nu)] / (1 - \nu) = G_{RF}^{1/2} [\mathcal{B}_q(1) - \mathcal{B}_q(\nu)] / (1 - \nu).$$

Let $\Upsilon \stackrel{\text{def}}{=} \text{diag}(\nu, 1 - \nu)$. Then, the above imply:

$$T^{1/2} (\hat{\Pi}_{vec}(\nu) - \iota_2 \otimes \Pi_T) \Rightarrow (\Upsilon^{-1} \otimes G_{RF}^{1/2}) \mathcal{B} \mathcal{B}_q(\nu) \tag{B.19}$$

$$\text{with } \mathcal{B} \mathcal{B}_q(\nu) \stackrel{\text{def}}{=} \begin{pmatrix} \mathcal{B}_q(\nu) \\ \mathcal{B}_q(1) - \mathcal{B}_q(\nu) \end{pmatrix}$$

$$\begin{aligned} \text{and } AVar[T^{1/2} (\hat{\Pi}_{vec}(\nu) - \iota_2 \otimes \Pi_T)] &= (\Upsilon^{-1} \otimes G_{RF}^{1/2}) (\Upsilon \otimes I_q) (\Upsilon^{-1} \otimes G_{RF}^{1/2}) \\ &= (\Upsilon^{-1} \Upsilon \Upsilon^{-1}) (G_{RF}^{1/2} G_{RF}^{1/2}) = \Upsilon^{-1} \otimes G_{RF} \tag{B.20} \end{aligned}$$

Using (B.19)-(B.20) and letting $r_q \stackrel{\text{def}}{=} (1, -1)'$, such that $\mathcal{R}'_q = r_q \otimes I_q$, we obtain:

$$\begin{aligned}
Wald_T^{RF}(\nu) &= T \hat{\Pi}'(\nu) \mathcal{R}'_q [\mathcal{R}_q \hat{G}_{RF} \mathcal{R}'_q]^{-1} \mathcal{R}_q \hat{\Pi}'(\nu) \\
&\Rightarrow \mathcal{B}\mathcal{B}'_q(\nu) (\Upsilon^{-1} \otimes G_{RF}^{1/2})(r_q \otimes I_q) \left[(r'_q \otimes I_q)(\Upsilon^{-1} \otimes G_{RF}^{1/2})(r_q \otimes I_q) \right]^{-1} (r'_q \otimes I_q)(\Upsilon^{-1} \otimes G_{RF}^{1/2}) \mathcal{B}\mathcal{B}_q(\nu) \\
&= \mathcal{B}\mathcal{B}'_q(\nu) [(\Upsilon^{-1} r_q) \otimes G_{RF}^{1/2}] [(r'_q \Upsilon^{-1} r_q) \otimes G_{RF}]^{-1} [(r'_q \Upsilon^{-1}) \otimes G_{RF}^{1/2}] \mathcal{B}\mathcal{B}_q(\nu) \\
&= \mathcal{B}\mathcal{B}'_q(\nu) [(\Upsilon^{-1} r_q) \otimes G_{RF}^{1/2}] [(r'_q \Upsilon^{-1} r_q)^{-1} \otimes G_{RF}^{-1}] [(r'_q \Upsilon^{-1}) \otimes G_{RF}^{1/2}] \mathcal{B}\mathcal{B}_q(\nu) \\
&= \mathcal{B}\mathcal{B}'_q(\nu) \{ [\Upsilon^{-1} r_q (r'_q \Upsilon^{-1} r_q)^{-1} r'_q \Upsilon^{-1}] \otimes I_q \} \mathcal{B}\mathcal{B}_q(\nu).
\end{aligned}$$

We can show that $\Upsilon^{-1} r_q (r'_q \Upsilon^{-1} r_q)^{-1} r'_q \Upsilon^{-1} = \frac{1}{\nu(1-\nu)} \begin{pmatrix} (1-\nu)^2 & -\nu(1-\nu) \\ -\nu(1-\nu) & \nu^2 \end{pmatrix}$, and so:

$$\mathcal{B}\mathcal{B}'_q(\nu) \{ [\Upsilon^{-1} r_q (r'_q \Upsilon^{-1} r_q)^{-1} r'_q \Upsilon^{-1}] \otimes I_q \} \mathcal{B}\mathcal{B}_q(\nu) = \frac{\|\mathcal{B}_q(\nu) - \nu \mathcal{B}_q(1)\|^2}{\nu(1-\nu)}.$$

• (ii) We now show that for $\Pi_\Delta = \lim_{T \rightarrow \infty} r_T(\Pi_1/r_{1T} - \Pi_2/r_{2T})$, we have:

$$T^{-1} r_T^2 Sup - Wald_T^{RF} \xrightarrow{p} \sup_{\nu \in \Lambda_\epsilon} \begin{cases} \Pi'_\Delta Q_2 Q_2^{-1}(\nu) [\overline{G}_{RF}(\nu)]^{-1} Q_2^{-1}(\nu) Q_2 \Pi_\Delta & \nu \leq \nu^0 \\ \Pi'_\Delta Q_1 Q_1^{-1}(\nu) [\overline{G}_{RF}(\nu)]^{-1} Q_1^{-1}(\nu) Q_1 \Pi_\Delta & \nu > \nu^0 \end{cases}$$

with $\overline{G}_{RF}(\nu) = Q_1^{-1}(\nu) S_{v,1}(\nu) Q_1^{-1} + Q_2^{-1}(\nu) S_{v,2}(\nu) Q_2^{-1}(\nu)$.

If $\nu \leq \nu^0$, we use similar arguments to part (i), to show that, uniformly in ν (u. ν),

$$\begin{aligned}
\hat{\Pi}_1(\nu) - \Pi_{1T} &= \hat{Q}_1^{-1}(\nu) T^{-1/2} \Psi_1^v(\nu) = \mathcal{O}_P(T^{-1/2}) \\
\hat{\Pi}_2(\nu) - \Pi_{2T} &= \hat{Q}_2^{-1}(\nu) T^{-1/2} \Psi_2^v(\nu) + \hat{Q}_2^{-1}(\nu) \hat{Q}_\Delta \Pi_T^\Delta \\
&= \mathcal{O}_P(T^{-1/2}) + Q_2^{-1}(\nu) [Q_2(\nu) - Q_2] \Pi_T^\Delta = \mathcal{O}_P(T^{-1/2}) + \{I - Q_2^{-1}(\nu) Q_2\} \Pi_T^\Delta.
\end{aligned}$$

It follows that u. ν ,

$$\begin{aligned}
\mathcal{R}_q \hat{\Pi}_{vec}(\lambda) &= \hat{\Pi}_1(\nu) - \hat{\Pi}_2(\nu) = (\hat{\Pi}_1(\nu) - \Pi_{1T}) - (\hat{\Pi}_2(\nu) - \Pi_{2T}) + \Pi_T^\Delta \\
&= \mathcal{O}_P(T^{-1/2}) + \{I - I + Q_2^{-1}(\nu) Q_2\} \Pi_T^\Delta = \mathcal{O}_P(T^{-1/2}) + [Q_2^{-1}(\nu) Q_2] \Pi_T^\Delta.
\end{aligned}$$

Hence, with $\hat{\overline{G}}_{RF}(\nu) = \hat{Q}_1^{-1}(\nu) \hat{S}_{v,1}(\nu) \hat{Q}_1^{-1} + \hat{Q}_2^{-1}(\nu) \hat{S}_{v,2}(\nu) \hat{Q}_2^{-1}(\nu) \xrightarrow{p} \overline{G}_{RF}(\nu)$,

$$Wald_T^{RF}(\nu) = \mathcal{O}_P(T^{1/2} r_T^{-1}) + T \Pi_T^{\Delta'} Q_2 Q_2^{-1}(\nu) \hat{\overline{G}}_{RF}(\nu) Q_2^{-1}(\nu) Q_2 \Pi_T^\Delta,$$

where the latter term dominates because it is $\mathcal{O}_P(T r_T^{-2})$. This implies that u. $\nu \leq \nu^0$,

$$Wald_T^{RF}(\nu) = \mathcal{O}_P(T^{1/2} r_T^{-1}) + T \Pi_T^{\Delta'} Q_2 Q_2^{-1}(\nu) \overline{G}_{RF}^{-1}(\nu) Q_2^{-1}(\nu) Q_2 \Pi_T^\Delta.$$

Similarly, it can be shown that u. $\nu > \nu^0$,

$$Wald_T^{RF}(\nu) = \mathcal{O}_P(T^{1/2} r_T^{-1}) + T \Pi_T^{\Delta'} Q_1 Q_1^{-1}(\nu) \overline{G}_{RF}^{-1}(\nu) Q_1^{-1}(\nu) Q_1 \Pi_T^\Delta.$$

Under Assumption 5, $Q_1(\nu) = \nu Q$, $S_{v,1}(\nu) = \nu S_v$, $Q_1 = \nu^0 Q$, thus u. ν ,

$$Wald_T^{RF}(\nu) = \mathcal{O}_P(T^{1/2}r_T^{-1}) + \begin{cases} (1 - \nu^0)^2 \frac{\nu}{1-\nu} [T\Pi_T^{\Delta'} Q S_v^{-1} Q \Pi_T^{\Delta}], & \nu \leq \nu^0 \\ (\nu^0)^2 \frac{1-\nu}{\nu} [T \Pi_T^{\Delta'} Q S_v^{-1} Q \Pi_T^{\Delta}], & \nu > \nu^0. \end{cases}$$

Since the probability limit of $T^{-1}r_T^2 Wald_T^{RF}(\nu)$ is continuous in ν , with the supremum attained at ν^0 , by the continuous mapping theorem, $\hat{\nu}^W = \arg [\sup_{\nu \in \Lambda_\epsilon} Wald_T^{RF}(\nu)] \xrightarrow{P} \nu^0$. However, if Assumption 5 is violated, $\hat{\nu}^W \xrightarrow{P} \nu^0$ is not always satisfied. As a counterexample, let $Q_2(\nu^*) = \sqrt{e}Q_1(\nu^*)$ for some $\nu^* < \nu^0$ and $0 < e < 1$. Then $r_T\Pi_T^{\Delta} \xrightarrow{P} \Pi_\Delta$, and:

$$T^{-1}r_T^2 Wald_T^{RF}(\nu^*) \xrightarrow{P} \Pi'_\Delta Q_2 [eS_{v,1}(\nu^*) + S_{v,2}(\nu^*)]^{-1} Q_2 \Pi'_\Delta = T\Pi'_\Delta Q_2 [(e-1)S_{v,1}(\nu^*) + S_v]^{-1} Q_2 \Pi_\Delta$$

$$T^{-1}r_T^2 Wald_T^{RF}(\nu^0) \xrightarrow{P} \Pi'_\Delta Q_2 S_v^{-1} Q_2 \Pi_\Delta.$$

Since $S_{v,1}(\nu^*)$ is symmetric, pd, $e < 1$ and S_v is symmetric, pd, $[(e-1)S_{v,1}(\nu^*) + S_v]^{-1} > S_v^{-1}$, so $\text{plim}[Wald_T^{RF}(\nu^*) - Wald_T^{RF}(\nu^0)] > 0$, meaning that $\hat{\nu}^W \not\xrightarrow{P} \nu^0$. ■

For Theorems 4-6, Q_i and $S_{u,i}, S_{v,i}, S_{uv,i}$ are as before, but with λ^0 replacing ν^0 .

• **Proof of Theorem 4: Asymptotic properties of $\hat{\lambda}$, $\hat{\theta}_{vec}$, $\hat{\theta}_{GMM,vec}$, and $\hat{\theta}_{MOD,vec}$**

- We prove the following statements (with explicit formulas for the asymptotic variances):

(i) Under Assumptions 1(ii), 2 to 4, 7, $\|\hat{\lambda} - \lambda^0\| = \mathcal{O}_P(T^{-1})$.

(ii) Under Assumptions 1(ii), 2 to 4, 7,

$$[I_2 \otimes \Lambda_T] (\hat{\theta}_{vec} - \theta_{vec}^0) \xrightarrow{d} \mathcal{N}(0, V_{2SLS}),$$

with $V_{2SLS} = \begin{bmatrix} D'_1 \Omega_1 D_1 & D'_1 \Omega_{12} D_2 \\ D'_2 \Omega'_{12} D_1 & D'_2 \Omega_2 D_2 \end{bmatrix}$, where $\Lambda_T = \text{diag}(T^{1/2}I_{p_1}, T^{1/2}r_T^{-1}I_{p_2})$, $A_i = \Pi^{a'} Q_i \Pi^a$,

$D_i = A_i^{-1}(\Pi^{a'} M'_i)$, $M'_1 = (I, Q_2 Q^{-1}, -Q_1 Q^{-1})$, $M'_2 = (I, Q_1 Q^{-1}, -Q_2 Q^{-1})$, $a_i = \theta_{y,i}^0 \otimes I_q$,

$$\Omega_i = \begin{pmatrix} S_{u,i} & a'_i S_{uv,i} & O \\ S'_{uv,i} a_i & a'_i S_{v,i} a_i & O \\ O & O & a'_i (S_v - S_{v,i}) a_i \end{pmatrix}, \quad \Omega_{12} = \begin{pmatrix} O & O & a'_2 S_{uv,1} \\ O & O & a'_1 S_{v,1} a_2 \\ S'_{uv,2} a_1 & a'_1 S_{v,2} a_2 & O \end{pmatrix}.$$

(iii) Under Assumptions 1(ii), 2 to 4, 7, with $r_T = o(\sqrt{T})$,

$$[I_2 \otimes \Lambda_T] (\hat{\theta}_{GMM,vec} - \theta_{vec}^0) \xrightarrow{d} \mathcal{N}(0, V_{GMM,vec}),$$

where $V_{GMM,vec} = \text{diag}(V_{GMM,1}, V_{GMM,2})$, and $V_{GMM,i} = [\Pi^{a'} Q_i (S_{u,i})^{-1} Q_i \Pi^a]^{-1}$.

(iv) Under Assumptions 1(ii), 2 to 4, 7, with $r_T = o(\sqrt{T})$,

$$\text{diag}(I_2 \otimes \Lambda_T, T^{1/2}I_{p_2q}) \begin{bmatrix} \hat{\theta}_{MOD,vec} - \theta_{vec}^0 \\ \hat{\Pi}_{MOD,vec} - \Pi_{vec}^0 / r_T \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, [\Gamma' \mathcal{S}^{-1} \Gamma]^{-1}), \quad \text{where:}$$

$$\mathcal{S} = \begin{bmatrix} S_{u,1} & O & S'_{uv,1} & O \\ O & S_{u,2} & O & S'_{uv,2} \\ S_{uv,1} & O & S_{v,1} & O \\ O & S_{uv,2} & O & S_{v,2} \end{bmatrix}, \quad \Gamma = - \begin{bmatrix} Q_1 \Pi^a & O & O \\ O & Q_2 \Pi^a & O \\ O & O & \begin{bmatrix} I_{p_2} \otimes Q_1 \\ I_{p_2} \otimes Q_2 \end{bmatrix} \end{bmatrix} = - \text{diag}(\Gamma_1, \Gamma_2),$$

with $\Gamma_1 = \text{diag}(Q_1 \Pi^a, Q_2 \Pi^a)$ and $\Gamma_2 = \begin{bmatrix} I_{p_2} \otimes Q_1 \\ I_{p_2} \otimes Q_2 \end{bmatrix}$. Thus,

$$[I_2 \otimes \Lambda_T] \left(\hat{\theta}_{MOD,vec} - \theta_{vec}^0 \right) \xrightarrow{d} \mathcal{N}(0, V_{MOD,vec}),$$

where $V_{MOD,vec} = (V_{GMM,vec}^{-1} + \mathcal{H}' \mathcal{E}^{-1/2} \mathcal{M}_{\mathcal{J}} \mathcal{E}^{-1/2} \mathcal{H})^{-1}$,

with $\mathcal{E} = \text{diag}(S_{v,1} - S_{uv,1} S_{u,1}^{-1} S'_{uv,1}, S_{v,2} - S_{uv,2} S_{u,2}^{-1} S'_{uv,2})$, $\mathcal{M}_{\mathcal{J}} = I - \mathcal{J}(\mathcal{J}' \mathcal{J})^{-1} \mathcal{J}'$,

$\mathcal{J} = \mathcal{E}^{-1/2} \Gamma_2$, and $\mathcal{H} = [\text{diag}(S_{uv,1} S_{u,1}^{-1}, S_{uv,2} S_{u,2}^{-1})] \Gamma_1$.

- To facilitate the proof, let $\theta^0(t) \stackrel{\text{def}}{=} \theta_1^0 \mathbf{1}[t \leq T^0] + \theta_2^0 \mathbf{1}[t > T^0]$ and $\tilde{u}_t \stackrel{\text{def}}{=} y_t - \hat{Y}_t \theta^0(t) = u_t + (Y_t - \hat{Y}_t) \theta^0(t) = u_t + v_t \theta^0(t) + W_t'(\Pi_T - \hat{\Pi})$, and $\Pi_T \stackrel{\text{def}}{=} \Pi/r_T$. Also, for simplicity, let $\hat{A}_1(r) \stackrel{\text{def}}{=} T^{-1} r_T^2 \sum_{t=1}^{[Tr]} \hat{Y}_t^2$, $\hat{A}_2(r) \stackrel{\text{def}}{=} \hat{A}_1(1) - \hat{A}_1(r)$, $\xi_1(r) \stackrel{\text{def}}{=} r_T T^{-1/2} \sum_{t=1}^{[Tr]} \hat{Y}_t \tilde{u}_t$, and $\xi_2(r) \stackrel{\text{def}}{=} \xi_1(1) - \xi_1(r)$. Let $A_i \stackrel{\text{def}}{=} \Pi' Q_i \Pi$ (as in Theorem 1(ii) but with $\Pi_1 = \Pi_2 = \Pi$, and ν^0 replaced by λ^0), and $\xi_i(\lambda^0) \stackrel{\text{def}}{=} \xi_i$. With this notation, $T^{1/2} r_T^{-1} (\hat{\theta}_i - \theta_i^0) = \hat{A}_i^{-1}(\hat{\lambda}) \xi_i(\hat{\lambda})$.

- We start by proving the following preliminary Lemma.

Lemma 1. *Under Assumptions 1(ii), 4, and 7, (i) $\hat{A}_1(r) = O_p(1)$, uniformly in r (u.r. thereafter); (ii) $\xi_1(r) = O_p(1)$ u.r.*

Proof of Lemma 1:

• (i) Note that $\hat{A}_1(r) = r_T^2 T^{-1} \sum_{t=1}^{[Tr]} \hat{Y}_t^2 = (r_T \hat{\Pi}') \hat{Q}_1(r) (r_T \hat{\Pi})$, and $\hat{Q}_1(r) \xrightarrow{p} Q_1(r)$, u.r., respectively $r_T \hat{\Pi} \xrightarrow{p} \Pi$. So, $\hat{A}_1(r) = \mathcal{O}_P(1) \times \mathcal{O}_P(1) \times \mathcal{O}_P(1) = \mathcal{O}_P(1)$ u.r.

• (ii) If we set $\Pi_{1T} = \Pi_{2T} = \Pi_T$ in the Proof of Theorem 1(i), $T^{1/2}(\hat{\Pi} - \Pi_T) = \mathcal{O}_P(1)$, independently of r . Also, by the FCLT in Wooldridge and White (1988), Theorem 2.11, $\Psi_1^{uv}(r) = T^{-1/2} \sum_{t=1}^{[Tr]} W_t(u_t + v_t \theta^0(t)) = \mathcal{O}_P(1)$ u.r. It follows that:

$$\begin{aligned} \xi_1(r) &= r_T \Pi_T' \tilde{\Psi}_1(r) + o_P(1) = \Pi' \Psi_1^{uv}(r) - \Pi' \hat{Q}_1(r) [T^{1/2}(\hat{\Pi} - \Pi_T)] + o_P(1) \\ &= \mathcal{O}_P(1) + \mathcal{O}_P(1) \times \mathcal{O}_P(1) + o_P(1) = \mathcal{O}_P(1) \text{ u.r.} \end{aligned}$$

□

- We now return to the proof of the main results. Assume wlog $\hat{T} < T$. The proof for $\hat{T} \geq T$ is similar and omitted for simplicity.

• (i) *Consistency of $\hat{\lambda}$.* Let $\hat{u}_t \stackrel{\text{def}}{=} y_t - Y_t \hat{\theta}_1$ for $t \in \{1, \dots, \hat{T}\}$, $\hat{u}_t \stackrel{\text{def}}{=} y_t - Y_t \hat{\theta}_2$ otherwise, and $d_t \stackrel{\text{def}}{=} \hat{u}_t - \tilde{u}_t$. We show consistency by contradiction, in two steps. By definition, $\sum_{t=1}^T \hat{u}_t^2 \leq \sum_{t=1}^T \tilde{u}_t^2$, hence $2 \sum_{t=1}^T \tilde{u}_t d_t + \sum_{t=1}^T d_t^2 \leq 0$. In step 1, we show that:

$$\sum_{t=1}^T d_t^2 = \mathcal{O}_P(T r_T^{-2}) \text{ and } \sum_{t=1}^T \tilde{u}_t d_t = \mathcal{O}_P(T^{1/2} r_T^{-1}), \quad (\text{B.21})$$

implying that $\sum_{t=1}^T d_t^2 \gg 2 \sum_{t=1}^T \tilde{u}_t d_t$ and so²⁸ $\text{plim}(r_T^2/T) \sum_{t=1}^T d_t^2 \leq 0$. Because we have $(r_T^2/T) \sum_{t=1}^T d_t^2 \geq 0$, it follows that $\text{plim}(r_T^2/T) \sum_{t=1}^T d_t^2 = 0$. In step 2, if $\hat{\lambda} \not\rightarrow \lambda^0$, then with positive probability, $(r_T^2/T) \sum_{t=1}^T d_t^2 > 0$, contradicting $\text{plim}(r_T^2/T) \sum_{t=1}^T d_t^2 = 0$, so $\hat{\lambda} \xrightarrow{P} \lambda^0$.

- **Step 1.** Note that:

$$d_t = \hat{u}_t - \tilde{u}_t = \begin{cases} y_t - \hat{Y}_t \hat{\theta}_1 - y_t + \hat{Y}_t \theta^0(t), & t \leq \hat{T} \\ y_t - \hat{Y}_t \hat{\theta}_2 - y_t + \hat{Y}_t \theta^0(t), & t > \hat{T} \end{cases} = \begin{cases} \hat{Y}_t(\theta_1^0 - \hat{\theta}_1), & t \leq \hat{T} \\ \hat{Y}_t(\theta_1^0 - \hat{\theta}_2), & \hat{T} + 1 \leq t \leq T^0 \\ \hat{Y}_t(\theta_2^0 - \hat{\theta}_2), & t > T^0 \end{cases}.$$

It follows that, for $\xi_\Delta = \xi_1(\lambda^0) - \xi_1(\hat{\lambda})$,

$$r_T T^{-1} \sum_{t=1}^T \tilde{u}_t d_t = T^{1/2}(\theta_1^0 - \hat{\theta}_1) \xi_1(\hat{\lambda}) + T^{1/2}(\theta_1^0 - \hat{\theta}_2) \xi_\Delta + T^{1/2}(\theta_2^0 - \hat{\theta}_2) \xi_2(\hat{\lambda}).$$

By Lemma 1, $\xi_1(\hat{\lambda}) = \mathcal{O}_P(1)$, $\xi_2(\hat{\lambda}) = \mathcal{O}_P(1)$, $\xi_\Delta = \mathcal{O}_P(1)$. Also, recall $\theta_\Delta^0 \stackrel{\text{def}}{=} \theta_1^0 - \theta_2^0$ and $\hat{A}_\Delta \stackrel{\text{def}}{=} \hat{A}_1(\hat{\lambda}) - \hat{A}_1$, then

$$\begin{aligned} T^{1/2} r_T^{-1} [\hat{\theta}_1 - \theta_1^0] &= \hat{A}_1^{-1}(\hat{\lambda}) \xi_1(\hat{\lambda}) = \mathcal{O}_P(1) \\ T^{1/2} r_T^{-1} [\hat{\theta}_2 - \theta_2^0] &= \hat{A}_2^{-1}(\hat{\lambda}) \xi_2(\hat{\lambda}) + \hat{A}_1(\hat{\lambda}) \hat{A}_\Delta T^{1/2} r_T^{-1} \theta_\Delta^0 \\ &= \mathcal{O}_P(1) + \mathcal{O}_P(T^{1/2} r_T^{-1}) = \mathcal{O}_P(T^{1/2} r_T^{-1}). \end{aligned}$$

Hence, $T^{1/2} r_T^{-1} [\hat{\theta}_2 - \theta_1^0] = \mathcal{O}_P(1) - T^{1/2} r_T^{-1} \theta_\Delta^0 = \mathcal{O}_P(T^{1/2} r_T^{-1})$. Adding these together, $\sum_{t=1}^T \tilde{u}_t d_t = \mathcal{O}_P(T^{1/2} r_T^{-1})$. Next, note that

$$\begin{aligned} \sum_{t=1}^T d_t^2 &= \sum_{t=1}^{\hat{T}} d_t^2 + \sum_{t=\hat{T}+1}^{T^0} d_t^2 + \sum_{t=T^0+1}^T d_t^2 \\ &= (\theta_1^0 - \hat{\theta}_1)^2 T r_T^{-2} \hat{A}_1(\hat{\lambda}) + (\theta_1^0 - \hat{\theta}_2)^2 T r_T^{-2} \hat{A}_\Delta + (\theta_2^0 - \hat{\theta}_2)^2 T r_T^{-2} \hat{A}_2(\hat{\lambda}) \\ &= \mathcal{O}_P(1) \mathcal{O}_P(T r_T^{-2}) + \mathcal{O}_P(1) \mathcal{O}_P(T r_T^{-2}) + \mathcal{O}_P(1) \mathcal{O}_P(T r_T^{-2}) = \mathcal{O}_P(T r_T^{-2}). \end{aligned}$$

²⁸If $r_T = T^{1/2}$, then $\sum_{t=1}^T d_t^2$ and $2 \sum_{t=1}^T \tilde{u}_t d_t$ are of the same order, and our argument does not apply.

- **Step 2.** If $\hat{\lambda} \not\xrightarrow{P} \lambda^0$, then there exists $\eta \in (0, 1)$, such that with positive probability ϵ , $T^0 - \hat{T} = [T\lambda^0] - [T\hat{\lambda}] \geq T\eta$. Then, with probability ϵ , for some $C > 0$,

$$r_T^2 T^{-1} \sum_{t=1}^T d_t^2 \geq r_T^2 T^{-1} \left(\sum_{t=T^0-T\eta+1}^{T^0} d_t^2 \right) = (\theta_1^0 - \hat{\theta}_2)^2 [\hat{A}_1(\lambda^0) - \hat{A}_1(\eta)] > C + o_P(1)$$

• (i) *Rate of convergence of $\hat{\lambda}$.* From above, any break point estimator $\hat{T} = [T\hat{\lambda}]$ is such that $T^0 - \hat{T} \leq \epsilon T$, for some chosen $\epsilon > 0$. Assume that for chosen $C > 0$, $T^0 - \hat{T} > Cr_T^2$. Define SSR_1 , SSR_2 and SSR_3 as the 2SLS sum of squared residuals in the structural equation, obtained with break points \hat{T} , T^0 and (\hat{T}, T^0) respectively. As for Theorem 1(i), it is sufficient to show that if $Cr_T^2 < T^0 - \hat{T} \leq \epsilon T$ for some large but fixed C and small but fixed ϵ , then $\text{plim}(SSR_1 - SSR_2) > 0$, which cannot hold by definition. It follows that $T^0 - \hat{T} \leq Cr_T^2$, and by symmetry of the argument, if $\hat{T} \geq T^0$, $\hat{T} - T^0 \leq Cr_T^2$, establishing the desired convergence rate for the break fraction estimator.

We now show that $\text{plim}(SSR_1 - SSR_2) > 0$. Denote by $(\hat{\theta}_1, \hat{\theta}_2)$ the 2SLS estimators based on sample partition $(1, \hat{T}, T)$, $(\hat{\theta}_1, \hat{\theta}_\Delta, \tilde{\theta}_2)$ the ones based on $(1, \hat{T}, T^0, T)$, and $(\tilde{\theta}_1, \tilde{\theta}_2)$ the ones based on $(1, T^0, T)$, all using the full-sample first stage predictor \hat{Y}_t . Then by similar arguments as for the proof of Theorem 1(i), we have:

$$\begin{aligned} SSR_1 - SSR_3 &= (\tilde{\theta}_2 - \hat{\theta}_\Delta)^2 Tr_T^{-2} \hat{A}_\Delta - (\tilde{\theta}_2 - \hat{\theta}_\Delta)^2 Tr_T^{-2} \hat{A}_\Delta \hat{A}_2^{-1}(\hat{\lambda}) \hat{A}_\Delta \stackrel{\text{def}}{=} N_1 - N_2 \\ SSR_2 - SSR_3 &= (\hat{\theta}_1 - \hat{\theta}_\Delta)^2 Tr_T^{-2} \hat{A}_\Delta - (\hat{\theta}_1 - \hat{\theta}_\Delta)^2 Tr_T^{-2} \hat{A}_\Delta \hat{A}_1^{-1}(\hat{\lambda}) \hat{A}_\Delta \stackrel{\text{def}}{=} N_3 - N_4 \end{aligned}$$

Since \hat{A}_Δ contains $T^0 - \hat{T} \leq [\epsilon T]$ terms, $\hat{A}_\Delta = \mathcal{O}_P(\epsilon)$, while $A_2(\hat{\lambda}) = \mathcal{O}_P(1)$ by Lemma 1. It follows that for small ϵ , $N_1 \gg N_2$. Because $\tilde{\theta}_2$ is estimating θ_2^0 with observations only in the second regime, $[T^0 + 1, T]$, it can be shown that $\tilde{\theta}_2 - \theta_2^0 = \mathcal{O}_P(T^{-1/2}r_T)$; on the other hand, $\hat{\theta}_\Delta$ is estimating θ_1^0 in subsample $[\hat{T} + 1, T^0]$, so for large enough C , it can be shown that $\hat{\theta}_\Delta - \theta_1^0 = \mathcal{O}_P(T^{-1/2}r_T)$, hence $\tilde{\theta}_2 - \hat{\theta}_\Delta = \theta_\Delta^0 + o_P(1)$, and so:

$$N_1 = [(\theta_\Delta^0)^2 + o_P(1)] \times \mathcal{O}_P(1) = \mathcal{O}_P(1).$$

For N_3 , it can be shown that $\hat{\theta}_1 - \theta_1^0 = \mathcal{O}_P(T^{-1/2}r_T)$, and since $\hat{\theta}_\Delta - \theta_1^0 = \mathcal{O}_P(T^{-1/2}r_T)$, $\hat{\theta}_1 - \hat{\theta}_\Delta = \mathcal{O}_P(T^{-1/2}r_T)$, and therefore $N_3 = \mathcal{O}_P(T^{-1}r_T^2) \times \mathcal{O}_P(1) = o_P(1)$. Similarly to $N_1 \gg N_2$, it can be shown that $N_3 \gg N_4$. It follows that $N_1 \gg N_j$, for $j = 2, 3, 4$ for chosen small ϵ and large C , and so $SSR_1 - SSR_2 = N_1 + o_P(1)$, hence

$$\text{plim} Tr_T^{-2}(SSR_1 - SSR_2) = (\theta_\Delta^0)^2 \text{plim} \hat{A}_\Delta = (\theta_\Delta^0)^2 \text{plim} \inf_t [\Pi' E(W_t W_t') \Pi] > 0,$$

because $\inf_t E(W_t W_t')$ is pd by Assumption 3. Thus, $\text{plim}(SSR_1 - SSR_2) > 0$.

• (ii) *Asymptotic distribution of 2SLS*. Recall that $T^{1/2}r_T^{-1}[\hat{\theta}_1 - \theta_1^0] = \hat{A}_1^{-1}(\hat{\lambda})\xi_1(\hat{\lambda})$. First analyze $\hat{A}_1^{-1}(\hat{\lambda})$ knowing that $\hat{T} - T^0 = \mathcal{O}_P(r_T^2)$. Since

$$\hat{A}_\Delta = T^{-1}r_T^2 \sum_{\hat{T}+1}^{T^0} \hat{Y}_t^2 = (T^0 - \hat{T})/T(r_T^2/(T^0 - \hat{T})) \sum_{\hat{T}+1}^{T^0} \hat{Y}_t^2 = \mathcal{O}_P(r_T^2 T^{-1}),$$

$\hat{A}_1(\hat{\lambda}) = \hat{A}_1(\lambda^0) - \hat{A}_\Delta = \hat{A}_1(\lambda^0) + o_P(1)$. Similarly, $\hat{A}_2(\hat{\lambda}) = \hat{A}_2(\lambda^0) + o_P(1)$, and $\hat{A}_i(\lambda^0) \xrightarrow{p} A_i$, with $A_i = \Pi'Q_i\Pi$, $Q_i = Q_i(\lambda^0)$, so:

$$T^{1/2}r_T^{-1}[\hat{\theta}_1 - \theta_1^0] = A_1^{-1}\xi_1(\hat{\lambda}) + o_P(1) = A_1^{-1}\Pi'\tilde{\Psi}_1(\hat{\lambda}) + o_P(1)$$

$$T^{1/2}r_T^{-1}[\hat{\theta}_2 - \theta_2^0] = A_2^{-1}\xi_2(\hat{\lambda}) + o_P(1) = A_2^{-1}\Pi'\tilde{\Psi}_2(\hat{\lambda}) + o_P(1).$$

It remains to analyze the asymptotic distributions of $\tilde{\Psi}_i(\hat{\lambda})$. Note that $\tilde{\Psi}_1(\hat{\lambda}) = \tilde{\Psi}_1(\lambda^0) - \tilde{\Psi}_\Delta$. It can be shown that $\tilde{\Psi}_\Delta = \mathcal{O}_P(r_T T^{-1/2})$ because it contains only observations in the subsample $\{\hat{T} + 1, \dots, T^0\}$, with $\hat{T} - T^0 = \mathcal{O}_P(r_T^2)$. Thus, $\tilde{\Psi}_i(\hat{\lambda}) = \tilde{\Psi}_i(\lambda^0) + o_P(1)$. As before, let $\tilde{\Psi}_i(\lambda^0) \stackrel{\text{def}}{=} \tilde{\Psi}_i$, $\Psi_i^j(\lambda^0) \stackrel{\text{def}}{=} \Psi_i^j$, where λ^0 replaces ν^0 in the notation for the Proof of Theorem 1(ii), and $j = u, v, uv$. Thus:

$$\begin{aligned} \tilde{\Psi}_1 &= \tilde{\Psi}^{uv} - \hat{Q}_1(\lambda^0)Q^{-1}(\Psi_1^v + \Psi_2^v)\theta_1^0 = \Psi_1^u + (I - Q_1Q^{-1})\Psi_1^v\theta_1^0 - (Q_1Q^{-1})\Psi_2^v\theta_1^0 + o_P(1) \\ &\stackrel{\text{def}}{=} M_1'\Psi_{1,vec} + o_P(1) \quad \text{with} \quad \Psi_{1,vec} = \text{vec}(\Psi_1^u, \Psi_1^v\theta_1^0, \Psi_2^v\theta_1^0). \end{aligned}$$

Recall that $M_1' \stackrel{\text{def}}{=} (I, Q_2Q^{-1}, -Q_1Q^{-1})$. Note that extra terms involving ν_t show up here because of the full-sample first-stage. They would not show up in the absence of breaks. From the CLT, $\Psi_{1,vec} \xrightarrow{d} \mathcal{N}(0, \Omega_1)$. Thus, $\tilde{\Psi}_1 \xrightarrow{d} \mathcal{N}(0, M_1'\Omega_1M_1)$, so $T^{1/2}r_T^{-1}[\hat{\theta}_1 - \theta_1^0] \xrightarrow{d} \mathcal{N}(0, D_1'\Omega_1D_1)$. Using similar arguments, $\tilde{\Psi}_2(\hat{\lambda}) \xrightarrow{d} \mathcal{N}(0, M_2'\Omega_2M_2)$. Hence, $T^{1/2}r_T^{-1}[\hat{\theta}_2 - \theta_2^0] \xrightarrow{d} \mathcal{N}(0, D_2'\Omega_2D_2)$. Moreover, because of the full-sample first-stage, asymptotically, $T^{1/2}r_T^{-1}[\hat{\theta}_1 - \theta_1^0] \not\perp T^{1/2}r_T^{-1}[\hat{\theta}_2 - \theta_2^0]$, and $\text{ACov}\{T^{1/2}r_T^{-1}[\hat{\theta}_1 - \theta_1^0], T^{1/2}r_T^{-1}[\hat{\theta}_2 - \theta_2^0]\} = D_1'\Omega_{12}D_2$.

• (ii) *Asymptotic distribution of GMM*.

We first prove that the asymptotic distribution of subsample GMM estimators is the same whether we use T^0 or \hat{T} to split the sample. Heuristically, it holds because we showed that $\hat{T} - T^0 = \mathcal{O}_P(r_T^2)$ uniformly in a r_T^2 -neighborhood. As in Theorem 1(ii), we denote the partition of W at \hat{T} and T^0 (rather than \hat{T}^*, T^*) as \bar{W} and \bar{W}^0 . We also partition $Y = \text{vec}(\bar{Y}_1, \bar{Y}_2) = \text{vec}(\bar{Y}_1^0, \bar{Y}_2^0)$ at \hat{T} , respectively T^0 , $y = \text{vec}(\bar{y}_1, \bar{y}_2) = \text{vec}(\bar{y}_1^0, \bar{y}_2^0)$, $U = \text{vec}(\bar{U}_1, \bar{U}_2) = \text{vec}(\bar{U}_1^0, \bar{U}_2^0)$, and $V = \text{vec}(\bar{V}_1, \bar{V}_2) = \text{vec}(\bar{V}_1^0, \bar{V}_2^0)$. Then for the weighting matrices $\hat{S}_{u,i}^{-1}$,

$$\hat{\theta}_{GMM,i} = (\bar{Y}'_i\bar{W}_i\hat{S}_{u,i}^{-1}\bar{W}'_i\bar{Y}_i)^{-1}\bar{Y}'_i\bar{W}_i\hat{S}_{u,i}^{-1}\bar{W}'_i\bar{y}_i.$$

Note that $T^{-1/2}\overline{W}'_1\overline{y}_1 = T^{-1/2}\sum_{t=1}^{T^0}W_t y_t + o_P(1) = T^{-1/2}\overline{W}'_1\overline{y}_1^0 + o_P(1)$, by similar arguments as for the previous proof. Also, $r_T T^{-1}\overline{W}'_1\overline{Y}_1 = r_T T^{-1}\overline{W}'_1\overline{Y}_1^0 + o_P(1)$. It follows that $\hat{\theta}_{GMM,1}$ is asymptotically equivalent to the estimator using T^0 instead of \hat{T} , and similarly for $\hat{\theta}_{GMM,2}$. The asymptotic distributions for the subsample GMM estimators using T^0 instead of \hat{T} follow from standard arguments. In particular, $T^{-1}r_T\overline{Y}'_i\overline{W}_i^0 \xrightarrow{p} \Pi'Q_i$, $T^{-1/2}\overline{W}'_i\overline{U}_i^0 \xrightarrow{d} \mathcal{N}(0, S_{u,i})$, $T^{1/2}r_T^{-1}[\hat{\theta}_{GMM,i} - \theta_i^0] \xrightarrow{d} \mathcal{N}(0, (\Pi'Q_i S_{u,i}^{-1}Q_i\Pi)^{-1})$ since $\hat{S}_{u,i} \xrightarrow{p} S_{u,i}$ ($i = 1, 2$); and, they are asymptotically independent by Assumption 1(ii).

• (iii) *Asymptotic distribution of MOD.* Let $\hat{\beta} \stackrel{\text{def}}{=} \begin{bmatrix} \hat{\theta}_{MOD,vec} \\ \hat{\Pi}_{MOD} \end{bmatrix}$, $\beta_T^0 \stackrel{\text{def}}{=} \begin{bmatrix} \theta_{vec}^0 \\ \Pi/r_T \end{bmatrix}$, and $\check{\Lambda}_T \stackrel{\text{def}}{=} \text{diag}(I_2 \otimes \Lambda_T, T^{1/2}I_q)$. Since $\hat{\theta}_{MOD,vec}$ is based on the same quantities as $\hat{\theta}_{GMM,vec}$ and $\hat{\theta}_{vec}$, we can show using similar arguments that the asymptotic distribution of $\check{\Lambda}_T(\hat{\beta} - \beta_T^0)$ is as if the break point T^0 was known. For the rest of the proof, assume wlog that the break point is known. Then, the estimator $\hat{\beta}$ is based on the following moment conditions²⁹:

$$\check{g}_T(\beta) = \begin{cases} g_{T,1}(\theta_{vec}) \stackrel{\text{def}}{=} T^{-1}\overline{W}'^0(y - \overline{Y}^0\theta_{vec}) \\ g_{T,2}(\Pi) \stackrel{\text{def}}{=} T^{-1} \begin{bmatrix} \overline{W}'_1(\overline{Y}_1 - \overline{W}_1\Pi) \\ \overline{W}'_2(\overline{Y}_2 - \overline{W}_2\Pi) \end{bmatrix} \end{cases},$$

where we use the notations introduced in part (ii). Then, for some $\beta = (\theta'_{vec} \Pi)'$,

$$\check{g}_T(\beta) = T^{-1} \begin{bmatrix} \overline{W}'^0(y - \overline{Y}^0\theta_{vec}) \\ \overline{W}^0(Y - W\Pi) \end{bmatrix} = T^{-1} \begin{bmatrix} \overline{W}^0 & O \\ O & \overline{W}^0 \end{bmatrix}' \left(\begin{bmatrix} y \\ Y \end{bmatrix} - \begin{bmatrix} \overline{Y}^0 & O \\ O & W \end{bmatrix} \beta \right).$$

Hence, the estimation of β_T^0 is written as a usual GMM problem. Therefore, from Assumptions 1(ii), 6, 7 and usual GMM asymptotics,

$$\check{\Lambda}_T(\hat{\beta} - \beta_T^0) \xrightarrow{d} \mathcal{N}(0, [\Gamma' \mathcal{S}^{-1} \Gamma]^{-1}), \quad (\text{B.22})$$

$$\begin{aligned} \text{with } \Gamma &= \text{plim} \frac{\partial \check{g}_T(\beta^0)}{\partial \beta'} (T^{-1/2} \check{\Lambda}_T^{-1}) = -T^{-1} \text{plim} \begin{bmatrix} \overline{W}^0 & O \\ O & \overline{W}^0 \end{bmatrix}' \begin{bmatrix} \overline{Y}^0 & O \\ O & W \end{bmatrix} (T^{-1/2} \check{\Lambda}_T^{-1}) \\ &= - \begin{bmatrix} \text{plim}[(T^{-1}\overline{W}^0\overline{Y}^0)(I_2 \otimes \text{diag}(I_{p_1}, r_T I_{p_2}))] & O \\ O & \text{plim}(T^{-1}\overline{W}^0 W) \end{bmatrix} = -\text{diag}(\Gamma_i), \\ \mathcal{S} &= \text{AVar} \left\{ T^{-1/2} \begin{bmatrix} \overline{W}^0 & O \\ O & \overline{W}^0 \end{bmatrix}' \begin{bmatrix} U \\ V \end{bmatrix} \right\} = \text{AVar} \begin{bmatrix} T^{-1/2}\overline{W}^0 U \\ T^{-1/2}\overline{W}^0 V \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{S}_{1,1} & \mathcal{S}_{1,2} \\ \mathcal{S}'_{1,2} & \mathcal{S}_{2,2} \end{bmatrix}, \end{aligned}$$

²⁹Note that for simplicity we scale the subsample moment conditions by T^{-1} instead of \hat{T}^{-1} and $(T - \hat{T})^{-1}$. The scaling is irrelevant since it cancels out in the formula for the GMM estimator.

where $\mathcal{S}_{1,1}$ and $\mathcal{S}_{2,2}$ are $(q \times q)$, and Γ_1 and Γ_2 are derived below. Note that the results above are obtained because under Assumption 3, we have:

$$\begin{aligned}\Gamma_1 &= -\text{diag}[\text{plim}(T^{-1}\overline{W}_1^{0'}\overline{Y}_1^0 \text{diag}(I_{p_1}, r_T I_{p_2})), \text{plim}(T^{-1}\overline{W}_2^{0'}\overline{Y}_2^0 \text{diag}(I_{p_1}, r_T I_{p_2}))] \\ &= -\text{diag}[\text{plim}(T^{-1}\overline{W}_1^{0'}\overline{W}_1^0)\Pi, \text{plim}(T^{-1}\overline{W}_2^{0'}\overline{W}_2^0)\Pi] = -\text{diag}(Q_i\Pi), \\ \Gamma_2 &= -\text{plim}(T^{-1}\overline{W}^{0'}W) = -\begin{bmatrix} \text{plim}(T^{-1}\overline{W}_1^{0'}\overline{W}_1^0) \\ \text{plim}(T^{-1}\overline{W}_2^{0'}\overline{W}_2^0) \end{bmatrix} = -\begin{bmatrix} I_{p_2} \otimes Q_1 \\ I_{p_2} \otimes Q_2 \end{bmatrix}.\end{aligned}$$

By Assumptions 1(ii), 4 and 6, $\mathcal{S} = \begin{bmatrix} \mathcal{S}_{1,1} & \mathcal{S}_{1,2} \\ \mathcal{S}_{1,2} & \mathcal{S}_{2,2} \end{bmatrix}$, with:

$$\begin{aligned}\mathcal{S}_{1,1} &= \text{AVar}(T^{-1/2}\overline{W}^{0'}U) = \text{diag}(S_{u,i}) \\ \mathcal{S}_{1,2} &= \text{ACov}(T^{-1/2}\overline{W}^{0'}U, T^{-1/2}\overline{W}^{0'}V) \\ &= \text{diag}\left[\text{ACov}\left(T^{-1/2}\sum_1^{T^0}W_t u_t, \sum_1^{T^0}W_t v_t\right), \text{ACov}\left(T^{-1/2}\sum_{T^0+1}^T W_t u_t, \sum_{T^0+1}^T W_t v_t\right)\right] \\ &= \text{diag}(S_{uv,i}), \\ \mathcal{S}_{2,2} &= \text{AVar}(T^{-1/2}\overline{W}^{0'}V) = \text{AVar}\begin{bmatrix} T^{-1/2}\overline{W}_1^{0'}\overline{V}_1^0 \\ T^{-1/2}\overline{W}_2^{0'}\overline{V}_2^0 \end{bmatrix} \\ &= \text{diag}[\text{AVar} T^{-1/2}\overline{W}_i^{0'}\overline{V}_i^0] \\ &= \text{diag}\left[\text{AVar}\left(T^{-1/2}\sum_1^{T^0}W_t v_t\right), \text{AVar}\left(T^{-1/2}\sum_{T^0+1}^T W_t v_t\right)\right] = \text{diag}(S_{v,i}).\end{aligned}$$

Given the above results, we use as in Antoine and Renault (2014) the partitioned inverse formula in Abadir and Magnus (2005), pp. 106, to get the desired result. In particular, for any A, B, C, D matrices, with nonsingular A, D , $E = D - CA^{-1}B$, and $F = A - BD^{-1}C$,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BE^{-1}CA^{-1} & -A^{-1}BE^{-1} \\ -E^{-1}CA^{-1} & E^{-1} \end{bmatrix} = \begin{bmatrix} F^{-1} & -F^{-1}BD^{-1} \\ -D^{-1}CF^{-1} & D^{-1} + D^{-1}CF^{-1}BD^{-1} \end{bmatrix}.$$

We use the first formula for \mathcal{S}^{-1} , and the second for $(\Gamma'\mathcal{S}^{-1}\Gamma)^{-1}$. Let $\mathcal{E} = \mathcal{S}_{2,2} - \mathcal{S}_{2,1}\mathcal{S}_{1,1}^{-1}\mathcal{S}_{1,2}$.

$$\begin{aligned}\mathcal{S}^{-1} &= \begin{bmatrix} \mathcal{S}_{1,1} & \mathcal{S}_{1,2} \\ \mathcal{S}'_{1,2} & \mathcal{S}_{2,2} \end{bmatrix}^{-1} = \begin{bmatrix} \mathcal{S}_{1,1}^{-1} + \mathcal{S}_{1,1}^{-1}\mathcal{S}_{1,2}\mathcal{E}^{-1}\mathcal{S}'_{1,2}\mathcal{S}_{1,1}^{-1} & -\mathcal{S}_{1,1}^{-1}\mathcal{S}_{1,2}\mathcal{E}^{-1} \\ -\mathcal{E}^{-1}\mathcal{S}'_{1,2}\mathcal{S}_{1,1}^{-1} & \mathcal{E}^{-1} \end{bmatrix}, \\ \Gamma'\mathcal{S}^{-1}\Gamma &= \begin{bmatrix} \Gamma'_1(\mathcal{S}_{1,1}^{-1} + \mathcal{S}_{1,1}^{-1}\mathcal{S}_{1,2}\mathcal{E}^{-1}\mathcal{S}'_{1,2}\mathcal{S}_{1,1}^{-1})\Gamma_1 & -\Gamma'_1\mathcal{S}_{1,1}^{-1}\mathcal{S}_{1,2}\mathcal{E}^{-1}\Gamma_2 \\ -\Gamma'_2\mathcal{E}^{-1}\mathcal{S}'_{1,2}\mathcal{S}_{1,1}^{-1}\Gamma_1 & \Gamma'_2\mathcal{E}^{-1}\Gamma_2 \end{bmatrix}\end{aligned}$$

Then, according to the second formula above, $V_{MOD,vec} = \mathcal{F}^{-1}$, with:

$$\mathcal{F} = \Gamma'_1(\mathcal{S}_{1,1}^{-1} + \mathcal{S}_{1,1}^{-1} \mathcal{S}_{1,2} \mathcal{E}^{-1} \mathcal{S}'_{1,2} \mathcal{S}_{1,1}^{-1})\Gamma_1 - (\Gamma'_1 \mathcal{S}_{1,1}^{-1} \mathcal{S}_{1,2} \mathcal{E}^{-1} \Gamma_2)(\Gamma'_2 \mathcal{E}^{-1} \Gamma_2)^{-1} \Gamma'_2 \mathcal{E}^{-1} \mathcal{S}'_{1,2} \mathcal{S}_{1,1}^{-1} \Gamma_1.$$

Since $\mathcal{J} = \mathcal{E}^{-1/2} \Gamma_2$ and $\mathcal{H} = \mathcal{S}'_{1,2} \mathcal{S}_{1,1}^{-1} \Gamma_1$, it follows that:

$$\mathcal{F} = \Gamma'_1 \mathcal{S}_{1,1}^{-1} \Gamma_1 + \mathcal{H}' \mathcal{E}^{-1/2} (I - \mathcal{J}(\mathcal{J}' \mathcal{J})^{-1} \mathcal{J}') \mathcal{E}^{-1/2} \mathcal{H} = \Gamma'_1 \mathcal{S}_{1,1}^{-1} \Gamma_1 + \mathcal{H}' \mathcal{E}^{-1/2} \mathcal{M}_{\mathcal{J}} \mathcal{E}^{-1/2} \mathcal{H}.$$

Note that from the above, $\Gamma'_1 \mathcal{S}_{1,1}^{-1} \Gamma_1 = V_{GMM,vec}^{-1}$, so that whenever $\mathcal{H}' \mathcal{E}^{-1/2} \mathcal{M}_{\mathcal{J}} \mathcal{E}^{-1/2} \mathcal{H} = 0$, the extra moment conditions $g_{T,2}(\Pi_T)$ are asymptotically redundant. ■

• **Proof of Theorem 5: Efficiency of estimated structural parameters**

• (i) Showing that $\hat{\theta}_{MOD,vec}$ is more efficient than $\hat{\theta}_{GMM,vec}$ is equivalent to showing that the additional moment conditions $g_{T,2}(\cdot)$ are asymptotically non-redundant for the estimation of θ_{vec}^0 (even though their derivative with respect to θ_{vec} is zero). The necessary and sufficient condition for non-redundancy follows from the proof of Theorem 4 and we can show that it is the same as Antoine and Renault's (2014) inequality on pp. 11:

$$(\mathcal{M}_{\mathcal{J}} \mathcal{E}^{-1/2} \mathcal{H})' \mathcal{M}_{\mathcal{J}} \mathcal{E}^{-1/2} \mathcal{H} \neq 0 \text{ and is positive semidefinite (psd).} \quad (\text{B.23})$$

With $\mathcal{G} = \mathcal{M}_{\mathcal{J}} \mathcal{E}^{-1/2} \mathcal{H}$, (B.23) can be written as $\mathcal{G}' \mathcal{G} \neq 0$ and psd. $\mathcal{G}' \mathcal{G}$ is psd by construction (see Abadir and Magnus (2005), pp. 214, Exercise 8), and we only need to show $\mathcal{G} \neq 0$.

$$\begin{aligned} \mathcal{H} &= \mathcal{S}'_{1,2} \mathcal{S}_{1,1}^{-1} \Gamma_1 = \text{diag}(S_{uv,i}) \text{diag}(S_{u,i}^{-1}) [-\text{diag}(Q_i \Pi)] \\ &= -\text{diag}[S_{uv,i} S_{u,i}^{-1} Q_i \Pi] \stackrel{\text{def}}{=} -\text{diag}[\mathbf{\Gamma}_i], \\ \mathcal{E} &= \text{diag}(S_{v,i} - S_{uv,i} S_{u,i}^{-1} S'_{uv,i}) \stackrel{\text{def}}{=} \text{diag}[\mathcal{E}_i], \\ \mathcal{E}^{-1/2} \mathcal{H} &= -\text{diag}[\mathcal{E}_i^{-1/2} \mathbf{\Gamma}_i] \\ \mathcal{J} &= -\mathcal{E}^{-1/2} \Gamma_2 = -\begin{bmatrix} \mathcal{E}_1^{-1/2} & O \\ O & \mathcal{E}_2^{-1/2} \end{bmatrix} \begin{bmatrix} I_{p_2} \otimes Q_1 \\ I_{p_2} \otimes Q_2 \end{bmatrix} = -\begin{bmatrix} \mathcal{E}_1^{-1/2} Q_1 \\ \mathcal{E}_2^{-1/2} Q_2 \end{bmatrix} \stackrel{\text{def}}{=} -\begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix}, \\ \mathcal{J}' \mathcal{J} &= \mathbf{E}'_1 \mathbf{E}_1 + \mathbf{E}'_2 \mathbf{E}_2, \quad (\mathcal{J}' \mathcal{J})^{-1} = (\mathbf{E}'_1 \mathbf{E}_1 + \mathbf{E}'_2 \mathbf{E}_2)^{-1}, \\ \mathcal{M}_{\mathcal{J}} &= I - \mathcal{J}(\mathcal{J}' \mathcal{J})^{-1} \mathcal{J}' = \begin{bmatrix} I - \mathbf{E}_1(\mathbf{E}'_1 \mathbf{E}_1 + \mathbf{E}'_2 \mathbf{E}_2)^{-1} \mathbf{E}'_1 & -\mathbf{E}_1(\mathbf{E}'_1 \mathbf{E}_1 + \mathbf{E}'_2 \mathbf{E}_2)^{-1} \mathbf{E}'_2 \\ -\mathbf{E}_2(\mathbf{E}'_1 \mathbf{E}_1 + \mathbf{E}'_2 \mathbf{E}_2)^{-1} \mathbf{E}'_1 & I - \mathbf{E}_2(\mathbf{E}'_1 \mathbf{E}_1 + \mathbf{E}'_2 \mathbf{E}_2)^{-1} \mathbf{E}'_2. \end{bmatrix} \end{aligned}$$

It is important to note that \mathcal{E}_i^{-1} exists, because \mathcal{E}_i is the Schur complement of $S_{u,i}$ in the variance matrix $S_i = \begin{bmatrix} S_{u,i} & S'_{uv,i} \\ S_{uv,i} & S_{v,i} \end{bmatrix}$. Since $S_{u,i}$ and S_i are pd by construction, so is \mathcal{E}_i .

Therefore, \mathcal{E}_i is invertible, and so $\mathcal{E}_i^{-1/2}$ and \mathbf{E}_i exist. It follows that:

$$\begin{aligned}
-\mathcal{G} &= -M_{\mathcal{J}}\mathcal{E}^{-1/2}\mathcal{H} \stackrel{\text{def}}{=} \begin{bmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} \\ \mathcal{G}_{21} & \mathcal{G}_{22} \end{bmatrix} \quad \text{with} \\
\mathcal{G}_{11} &= [I - \mathbf{E}_1(\mathbf{E}'_1\mathbf{E}_1 + \mathbf{E}'_2\mathbf{E}_2)^{-1}\mathbf{E}'_1] \mathcal{E}_1^{-1/2}\mathbf{\Gamma}_1, \quad \mathcal{G}_{12} = -\mathbf{E}_1(\mathbf{E}'_1\mathbf{E}_1 + \mathbf{E}'_2\mathbf{E}_2)^{-1}\mathbf{E}'_2 \mathcal{E}_2^{-1/2}\mathbf{\Gamma}_2 \\
\mathcal{G}_{22} &= [I - \mathbf{E}_2(\mathbf{E}'_1\mathbf{E}_1 + \mathbf{E}'_2\mathbf{E}_2)^{-1}\mathbf{E}'_2] \mathcal{E}_2^{-1/2}\mathbf{\Gamma}_2, \quad \mathcal{G}_{21} = -\mathbf{E}_2(\mathbf{E}'_1\mathbf{E}_1 + \mathbf{E}'_2\mathbf{E}_2)^{-1}\mathbf{E}'_1 \mathcal{E}_1^{-1/2}\mathbf{\Gamma}_1
\end{aligned}$$

To show $\mathcal{G} \neq 0$, we now show that $\mathcal{G}_{12} \neq 0$. Since \mathcal{E}_i and Q_i are pd, so is $\mathbf{E}_i = \mathcal{E}_i^{-1/2}Q_i$. It follows that $\mathbf{E}_1(\mathbf{E}'_1\mathbf{E}_1 + \mathbf{E}'_2\mathbf{E}_2)^{-1}\mathbf{E}'_2$ is pd and invertible, and

$$\mathcal{G}_{12} = 0 \Leftrightarrow \mathcal{E}_2^{-1/2}\mathbf{\Gamma}_2 = 0 \Leftrightarrow \mathbf{\Gamma}_2 = 0 \Leftrightarrow S_{uv,2} S_{u,2}^{-1} Q_2\Pi = 0 \Leftrightarrow \Pi = 0,$$

since $S_{uv,2}$ is a $q \times q$ pd covariance matrix by Assumptions 3 and 4. The latter cannot hold because it contradicts Assumption 7(ii). Hence, $\mathcal{G}_{12} \neq 0$, so $\mathcal{G} \neq 0$, and the additional moment conditions for $\hat{\theta}_{MOD,vec}$ are not redundant.

Note that under Assumptions 4 and 5, $\mathcal{G}'\mathcal{G}$ is of rank 1, thus rank deficient. This implies that some linear combinations of $\hat{\theta}_{MOD,vec}$ will be asymptotically equivalent to the same linear combinations of $\hat{\theta}_{GMM,vec}$, but in general they will be asymptotically more efficient.

• (ii) Let $c = \Phi_u + 2(1 - \lambda^0)\Phi_{uv}\theta_1^0 + (1 - \lambda^0)\Phi_v(\theta_1^0)^2$. Under Assumptions 5 and 6, we can show that $M'_1\Omega_1M_1 = \lambda^0cQ$. Since $A_i = \Pi'Q_1\Pi = \lambda^0\Pi'Q\Pi \stackrel{\text{def}}{=} \lambda^0A$, it follows that $V_{2SLS,1}^{-1} = \lambda^0A/c$. On the other hand, $V_{GMM,1}^{-1} = \lambda^0A/\Phi_u$, so we can compare the two by comparing c with Φ_u .

$$\begin{aligned}
c - \Phi_u &= (1 - \lambda^0)\theta_1^0(2\Phi_{uv} + \Phi_v\theta_1^0) = (1 - \lambda^0)\Phi_v\theta_1^0(2\Phi_{uv}/\Phi_v + \theta_1^0) \leq 0 \\
&\Leftrightarrow \theta_1^0(2\Phi_{uv}/\Phi_v + \theta_1^0) \leq 0.
\end{aligned}$$

This implies that $V_{2SLS,1}^{-1} \geq V_{GMM,1}^{-1}$, so $V_{2SLS,1} \leq V_{GMM,1}$. The proof for $V_{2SLS,2} \leq V_{GMM,2}$ is similar.

- From the calculations above, we have:

$$V_{2SLS,1} = \frac{c}{\lambda^0}A^{-1} = \frac{\Phi_u}{\lambda^0} + \frac{1 - \lambda^0}{\lambda^0}\theta_1^0(2\Phi_{uv} + \theta_1^0\Phi_v).$$

On the other hand, with $\delta \stackrel{\text{def}}{=} \Phi_u^{-2}\Phi_{uv}^2\bar{\mathcal{E}}^{-1}$, we have $V_{MOD,vec} = \mathcal{F}^{-1}$, with

$$\begin{aligned}
\mathcal{F} &= \Gamma'_1\mathcal{S}_{1,1}^{-1}\Gamma_1 + \mathcal{H}'\mathcal{E}^{-1/2}\mathcal{M}_{\mathcal{J}}\mathcal{E}^{-1/2}\mathcal{H} = \text{diag}(\lambda_i^0\Phi_u^{-1}A) + \mathcal{G}'\mathcal{G} \\
&= \begin{bmatrix} \lambda^0[\Phi_u^{-1} + (1 - \lambda^0)\delta] & -\lambda^0(1 - \lambda^0)\delta \\ -\lambda^0(1 - \lambda^0)\delta & (1 - \lambda^0)[\Phi_u^{-1} + \lambda^0\delta] \end{bmatrix} \otimes A \stackrel{\text{def}}{=} \mathcal{F} \otimes A, \\
V_{MOD,vec} &= \mathcal{F}^{-1} \stackrel{\text{def}}{=} \bar{\mathcal{F}}^{-1} \otimes A^{-1}.
\end{aligned}$$

Below we calculate $\overline{\mathcal{F}}^{-1}$:

$$\begin{aligned}\det \overline{\mathcal{F}} &= \lambda^0(1 - \lambda^0)\{[\Phi_u^{-1} + (1 - \lambda^0)\delta][\Phi_u^{-1} + \lambda^0\delta] - \lambda^0(1 - \lambda^0)\delta^2\} \\ &= \lambda^0(1 - \lambda^0)(\Phi_u^{-2} + \Phi_u^{-1}\delta) = \lambda^0(1 - \lambda^0)\Phi_u^{-2}(1 + \Phi_u\delta) \\ \overline{\mathcal{F}}^{-1} &= \frac{\Phi_u^2}{\lambda^0(1 - \lambda^0)(1 + \Phi_u\delta)} \begin{bmatrix} (1 - \lambda^0)[\Phi_u^{-1} + \lambda^0\delta] & \lambda^0(1 - \lambda^0)\delta \\ \lambda^0(1 - \lambda^0)\delta & \lambda^0[\Phi_u^{-1} + (1 - \lambda^0)\delta] \end{bmatrix} \\ &= \frac{\Phi_u^2}{(1 + \Phi_u\delta)} \begin{bmatrix} \frac{1}{\lambda^0} [\Phi_u^{-1} + \lambda^0\delta] & \delta \\ \delta & \frac{1}{1 - \lambda^0} [\Phi_u^{-1} + (1 - \lambda^0)\delta] \end{bmatrix}.\end{aligned}$$

It follows that $V_{MOD,1} = c_{MOD}A^{-1}$, $V_{2SLS} = cA^{-1}$, and

$$\begin{aligned}c_{MOD} &= \frac{\Phi_u(1 + \lambda^0\Phi_u\delta)}{\lambda^0(1 + \Phi_u\delta)} = \frac{\Phi_u}{\lambda^0} + \frac{\Phi_u(1 + \lambda^0\Phi_u\delta)}{\lambda^0(1 + \Phi_u\delta)} - \frac{\Phi_u}{\lambda^0} \\ &= \frac{\Phi_u}{\lambda^0} + \frac{\Phi_u(1 + \lambda^0\Phi_u\delta - 1 - \Phi_u\delta)}{\lambda^0(1 + \Phi_u\delta)} = \frac{\Phi_u}{\lambda^0} - \frac{1 - \lambda^0}{\lambda^0} \frac{\Phi_u^2\delta}{1 + \Phi_u\delta}, \\ c &= \frac{\Phi_u}{\lambda^0} + \frac{1 - \lambda^0}{\lambda^0} \theta_1^0(2\Phi_{uv} + \theta_{y,1}^0\Phi_v), \\ c - c_{MOD} &= \frac{1 - \lambda^0}{\lambda^0} \left(\theta_1^0(2\Phi_{uv} + \theta_1^0\Phi_v) + \frac{\Phi_u^2\delta}{1 + \Phi_u\delta} \right) \geq 0 \\ &\Leftrightarrow \theta_1^0(2\Phi_{uv} + \theta_1^0\Phi_v) \geq -\frac{\Phi_u^2\delta}{1 + \Phi_u\delta}.\end{aligned}$$

Hence, $\theta_1^0(2\Phi_{uv} + \theta_1^0\Phi_v) \geq -\frac{\Phi_u^2\delta}{1 + \Phi_u\delta} \equiv V_{2SLS,1} \geq V_{MOD,1}$. The proof for $V_{MOD,2}$ and $V_{2SLS,2}$ is similar and therefore omitted. ■

• **Proof of Theorem 6: Test for a break in the main equation**

• (i) Let θ^0 be the common value of (θ_i^0) under the null hypothesis, for $i = 1, 2$. The asymptotic distribution of the Wald test is determined by that of $T^{1/2}r_T^{-1}(\hat{\theta}_i(\lambda) - \theta^0) = \hat{A}_i^{-1}(\lambda) \xi_i(\lambda)$. From Lemma 1, the proof of Theorem 4 and Assumption 5, $\hat{A}_i(\lambda) \xrightarrow{P} \lambda A$. Also, under Assumptions 1(ii), 4 and 5, by the FCLT,

$$\begin{aligned}\xi_1(\lambda) &= o_P(1) + \Pi' \Psi_1^{uv}(\lambda) - \Pi'[\hat{Q}_1(\lambda)\hat{Q}_1^{-1}(1)]\Psi^v(1)\theta^0 \\ &= \Pi' \Psi_1^{uv}(\lambda) - \Pi' \lambda \Psi^v(1)\theta^0 + o_P(1) \Rightarrow \Pi'[P^{1/2}\mathcal{B}_q(\lambda) - \lambda(P^*)^{1/2}\mathcal{B}_q^*(1)],\end{aligned}\quad (\text{B.24})$$

where $P^{*1/2}\mathcal{B}_q^*(\lambda)$ and $P^{1/2}\mathcal{B}_q(\lambda)$ are two (dependent) $q \times 1$ Brownian motions generated by partial sums of $(W_t v_t \theta^0)$, respectively $[W_t(u_t + v_t \theta^0)]$, and $P = S_u + (S_{uv} + S'_{uv})\theta^0 + S_v(\theta^0)^2$, $P^* = S_v(\theta^0)^2$. Let $G^* = A^{-1}[\Pi' P^* \Pi] A^{-1} = A^{-2}[\Pi' P^* \Pi]$, $G = A^{-2}[\Pi' P \Pi]$, because A is a

scalar here. Also, $G^{1/2}G^{1/2'} = G$, and $(G^*)^{1/2}(G^*)^{1/2'} = G^*$, with $(G^*)^{1/2} = A^{-1}\Pi'(P^*)^{1/2}$ and $G^{1/2} = A^{-1}\Pi'P^{1/2}$. Then, from (B.24),

$$\begin{aligned} T^{1/2}r_T^{-1}(\hat{\theta}_1(\lambda) - \theta^0) &\Rightarrow G^{1/2}\frac{\mathcal{B}_q(\lambda)}{\lambda} - (G^*)^{1/2}\mathcal{B}_q^*(1) \\ T^{1/2}r_T^{-1}(\hat{\theta}_2(\lambda) - \theta^0) &\Rightarrow G^{1/2}\frac{\mathcal{B}_q(1) - \mathcal{B}_q(\lambda)}{1 - \lambda} - (G^*)^{1/2}\mathcal{B}_q^*(1) \\ \mathcal{R}_p T^{1/2}r_T^{-1}\hat{\theta}_{vec}(\lambda) &= T^{1/2}r_T^{-1}(\hat{\theta}_1(\lambda) - \theta^0) - T^{1/2}r_T^{-1}(\hat{\theta}_2(\lambda) - \theta^0) \\ &\Rightarrow G^{1/2}\frac{\mathcal{B}_q(\lambda)}{\lambda} - G^{1/2}\frac{\mathcal{B}_q(1) - \mathcal{B}_q(\lambda)}{1 - \lambda} = G^{1/2}\frac{\mathcal{B}_q(\lambda) - \lambda\mathcal{B}_q(1)}{\lambda(1 - \lambda)}. \end{aligned}$$

Recall that $\mathcal{B}\mathcal{B}_q(\lambda) \stackrel{\text{def}}{=} \mathcal{B}_q(\lambda) - \lambda\mathcal{B}_q(1)$. Since $\hat{H}_i(\lambda) \xrightarrow{p} \lambda_i\Pi'P\Pi$, for $\lambda_1 = \lambda$, $\lambda_2 = 1 - \lambda$, it follows that $\hat{G}_i(\lambda) \xrightarrow{p} \lambda_i^{-1}G$. Hence, $\mathcal{R}_p\hat{G}(\lambda)\mathcal{R}'_p = \mathcal{R}_p \text{diag}(\lambda^{-1}G, (1 - \lambda)^{-1}G)\mathcal{R}'_p = \frac{G}{\lambda(1 - \lambda)}$. Thus,

$$\left[T\hat{\theta}'(\lambda)\mathcal{R}'_p[\mathcal{R}_p\hat{G}_i(\lambda)\mathcal{R}'_p]^{-1}\mathcal{R}_p\hat{\theta}(\lambda) \right] \Rightarrow \frac{\mathcal{B}\mathcal{B}'_q(\lambda)[G^{1/2'}G^{-1}G^{1/2}]\mathcal{B}\mathcal{B}_q(\lambda)}{\lambda(1 - \lambda)} = \frac{\mathcal{B}\mathcal{B}'_1(\lambda)\mathcal{B}\mathcal{B}_1(\lambda)}{\lambda(1 - \lambda)},$$

because $G^{1/2'}(G)^{-1}G^{1/2}$ is a projection matrix of rank $p = p_2 = 1$, thus selecting only the first element of $\mathcal{B}\mathcal{B}_q(\lambda)$ (for an extensive proof, see Hall, Han and Boldea (2012, Supplemental Appendix, pp. 23-27)).

• (ii) We show that under $H_A : \theta_\Delta^0 = \theta_1^0 - \theta_2^0 \neq 0$, Assumptions 1(ii), 3, 2 and 7,

$$T^{-1}r_T^2 \text{Wald}_T(\lambda) \xrightarrow{p} \begin{cases} \{ A_2^{-1}A_2(\lambda)G^{-1}(\lambda)A_2(\lambda)A_2^{-1} \} (\theta_\Delta^0)^2 & \lambda \leq \lambda^0 \\ \{ A_1^{-1}A_1(\lambda)G^{-1}(\lambda)A_1(\lambda)A_1^{-1} \} (\theta_\Delta^0)^2 & \lambda > \lambda^0, \end{cases}$$

where $G(\lambda) = G_1(\lambda) + G_2(\lambda)$, $G_i(\lambda) = A_i^{-1}(\lambda)H_i(\lambda)A_i^{-1}(\lambda)$ ($i = 1, 2$).

If $\lambda \leq \lambda^0$, uniformly in λ (u. λ),

$$\begin{aligned} \hat{\theta}_1(\lambda) - \theta_1^0 &= \hat{A}_1^{-1}(\lambda)T^{-1/2}r_T \xi_1(\lambda) = \mathcal{O}_P(T^{-1/2}r_T) \\ \hat{\theta}_2(\lambda) - \theta_2^0 &= \hat{A}_2^{-1}(\lambda)T^{-1/2}r_T \xi_2(\lambda) + \hat{A}_2^{-1}(\lambda)\hat{A}_\Delta\theta_\Delta^0 \\ &= \mathcal{O}_P(T^{-1/2}r_T) + A_2^{-1}(\lambda)[A_2(\lambda) - A_2]\theta_\Delta^0 = \mathcal{O}_P(T^{-1/2}r_T) + \{I - A_2^{-1}(\lambda)A_2\}\theta_\Delta^0. \end{aligned}$$

It follows that u. λ , $\mathcal{R}_p\hat{\theta}_{vec}(\lambda) = \hat{\theta}_1(\lambda) - \hat{\theta}_2(\lambda) = \mathcal{O}_P(T^{-1/2}r_T) + [A_2^{-1}(\lambda)A_2]\theta_\Delta^0$. We have $\mathcal{R}_p\hat{G}(\lambda)\mathcal{R}'_p = \hat{G}_1(\lambda) + \hat{G}_2(\lambda) = \mathcal{O}_P(r_T^2)$, and $r_T^2[\hat{G}_1(\lambda) + \hat{G}_2(\lambda)] \xrightarrow{p} G_1(\lambda) + G_2(\lambda) \stackrel{\text{def}}{=} G(\lambda)$. Hence, u. $\lambda \leq \lambda^0$,

$$T^{-1}r_T^2 \text{Wald}_T(\lambda) \xrightarrow{p} \{A_2A_2^{-1}(\lambda)G^{-1}(\lambda)A_2^{-1}(\lambda)A_2\}(\theta_\Delta^0)^2.$$

Similarly, it can be shown that u. $\lambda > \lambda^0$,

$$T^{-1}r_T^2 \text{Wald}_T(\lambda) \xrightarrow{p} \{A_1A_1^{-1}(\lambda)G^{-1}(\lambda)A_1^{-1}(\lambda)A_1\}(\theta_\Delta^0)^2.$$

Under Assumption 5, $A_i(\lambda) = \lambda_i A$, $A_i A_i^{-1}(\lambda) = A_i^{-1}(\lambda) A_i = \lambda_i^{-1} \lambda_i^0$, where $\lambda_1 = \lambda$, $\lambda_2 = 1 - \lambda$, and $\lambda_2^0 = 1 - \lambda_1^0$, and

$$\begin{aligned} G_i(\lambda) &= A_i^{-1}(\lambda) H_i(\lambda) A_i^{-1}(\lambda) = \lambda_i^{-2} A^{-2} H_i(\lambda) = \lambda_i^{-2} A^{-2} (\lambda_i \Pi' (S_u + (S_{uv} + S'_{uv}) \theta_i^0 + S_v (\theta_i^0)^2) \Pi) \\ &= A^{-2} (\lambda_i^{-1} \Pi' (S_u + (S_{uv} + S'_{uv}) \theta_i^0 + S_v (\theta_i^0)^2) \Pi) \stackrel{\text{def}}{=} A^{-2} (\lambda_i^{-1} \mathbf{h}_i) \\ G(\lambda) &= G_1(\lambda) + G_2(\lambda) = A^{-2} \left(\frac{\mathbf{H}_1}{\lambda} + \frac{\mathbf{H}_2}{1-\lambda} \right) = A^{-2} [\lambda \mathbf{H}_2 + (1-\lambda) \mathbf{H}_1] \frac{1}{\lambda(1-\lambda)} \\ G^{-1}(\lambda) &= A^2 \lambda (1-\lambda) [\lambda \mathbf{H}_2 + (1-\lambda) \mathbf{H}_1]^{-1}. \end{aligned}$$

Note that because $S_u + (S_{uv} + S'_{uv}) \theta_i^0 + S_v (\theta_i^0)^2 = A \text{Var}(u_t + v_t \theta_i^0)$, it is pd, so $\mathbf{H}_i = \Pi' [S_u + (S_{uv} + S'_{uv}) \theta_i^0 + S_v (\theta_i^0)^2] \Pi > 0$ since $\Pi \neq 0$. It follows that:

$$T^{-1} r_T^2 \text{Wald}_T(\lambda) \xrightarrow{P} A^2 \begin{cases} (1 - \lambda^0)^2 \frac{\lambda/(1-\lambda)}{\lambda \mathbf{H}_2 + (1-\lambda) \mathbf{H}_1} & \lambda \leq \lambda^0 \\ (\lambda^0)^2 \frac{(1-\lambda)/\lambda}{\lambda \mathbf{H}_2 + (1-\lambda) \mathbf{H}_1} & \lambda > \lambda^0. \end{cases}$$

It can be shown that $\text{Wald}_T(\lambda)$ is asymptotically maximized at λ^0 , thus, by continuous mapping theorem, $\hat{\lambda}^W \xrightarrow{P} \lambda^0$. However, if Assumption 5 doesn't hold, and for some $\lambda^* < \lambda^0$, $Q_2(\lambda) = e Q_1(\lambda)$, for a scalar e , then it can be shown that for $0 < e < 1$, $\text{Wald}_T(\lambda^*) > \text{Wald}_T(\lambda^0) + o_P(1)$, so $\hat{\lambda}^W \not\xrightarrow{P} \lambda^0$. ■

• Proof of Theorem 7: Wald test for common break

• (i) Let θ^0 be the common value of θ_i^0 under the null hypothesis. By arguments similar to the Proof of Theorem 1(ii), the distribution of the subsample 2SLS estimators $\hat{\theta}_i^c$ is as if the break point ν^0 was known, so $T^{1/2} r_{iT} [\hat{\theta}_i^c - \theta^0] \xrightarrow{d} \mathcal{N}(0, G_i^c)$, where $G_i \stackrel{\text{def}}{=} (A_i^c)^{-1} B_i^c (A_i^c)^{-1}$, $A_i^c \stackrel{\text{def}}{=} \Pi_i^{a'} Q_i(\nu^0) \Pi_i^a$, and $B_i^c \stackrel{\text{def}}{=} \Pi_i^{a'} S_{u,i}(\nu^0) \Pi_i^a$. Moreover, $\hat{\theta}_1^c \perp \hat{\theta}_2^c$ asymptotically, because they are constructed with asymptotically independent subsamples in the first-stage. Also, by construction, $r_{iT}^2 \hat{B}_i^c \xrightarrow{P} B_i^c$, and $r_{iT}^{-2} \hat{G}_i^c \xrightarrow{P} G_i^c$.

- When $r_T = r_{1T} = r_{2T}$,

$$\begin{aligned} T^{1/2} r_T^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c &\xrightarrow{d} \mathcal{N}(0, G^c) \quad \text{with} \quad r_T^{-2} \mathcal{R}_p \hat{G}^c \mathcal{R}_p' \xrightarrow{P} G_1^c + G_2^c = G^c \\ \text{and} \quad \text{Wald}_T^c &= [T^{1/2} r_T^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c]' [r_T^{-2} \mathcal{R}_p \hat{G}^c \mathcal{R}_p']^{-1} [T^{1/2} r_T^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c] \xrightarrow{P} \chi_1^2 \end{aligned}$$

- When $r_{iT} = o(r_{jT})$, and wlog $i = 1$,

$$\begin{aligned} T^{1/2} r_{2T}^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c &= [r_{1T}/r_{2T}] T^{1/2} r_{1T}^{-1} (\hat{\theta}_1^c - \theta^0) - T^{1/2} r_{2T}^{-1} (\hat{\theta}_2^c - \theta^0) \\ &= o_P(1) - T^{1/2} r_{2T}^{-1} (\hat{\theta}_2^c - \theta^0) \xrightarrow{d} \mathcal{N}(0, G_2^c) \\ r_{2T}^{-2} \mathcal{R}_p \hat{G}^c \mathcal{R}_p' &\xrightarrow{P} G_2^c \\ \text{Wald}_T^c &= [T^{1/2} r_{2T}^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c]' [r_{2T}^{-2} \mathcal{R}_p \hat{G}^c \mathcal{R}_p']^{-1} [T^{1/2} r_{2T}^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c] \xrightarrow{P} \chi_1^2, \end{aligned}$$

and the test reduces to a J-test on the weaker subsample. This is not the case if $p_1 \neq 0$ (see the Supplemental Appendix for a detailed proof).

- (ii) - We now show that when $r_{1T} = r_{2T} = r_T$, $T^{-1}r_T^2 \text{Wald}_T^c \xrightarrow{P} \theta_\Delta^{0'}(G^c)^{-1} \theta_\Delta^0$. We have:

$$T^{1/2}r_T(\hat{\theta}_1^c - \theta_1^0) = \hat{A}_1^{-1}(\hat{\nu}) \xi_1(\hat{\nu}) = \mathcal{O}_P(1) \quad \text{and} \quad T^{1/2}r_T(\hat{\theta}_2^c - \theta_2^0) = \hat{A}_2^{-1}(\hat{\nu}) \xi_2(\hat{\nu}) = \mathcal{O}_P(1)$$

It follows that $T^{1/2}r_T^{-1}\mathcal{R}_p\hat{\theta}_{vec}^c = T^{1/2}r_T^{-1}[\hat{\theta}_1^c - \hat{\theta}_2^c] = \mathcal{O}_P(1) + T^{1/2}r_T^{-1}\theta_\Delta^0$.

Also, $r_T^{-2}\mathcal{R}_p\hat{G}^c\mathcal{R}_p' \xrightarrow{P} G^c$, hence $T^{-1}r_T^2 \text{Wald}_T^c \xrightarrow{P} \theta_\Delta^{0'}(G^c)^{-1}\theta_\Delta^0$.

- We now show that when $r_{iT} = o(r_{jT})$,

$$T^{-1}r_{jT}^2 \text{Wald}_T^c \xrightarrow{P} \theta_\Delta^{0'}(G_j^c)^{-1} \theta_\Delta^0.$$

We assume wlog that $r_T = r_{1T} \ll r_{2T}$. Then:

$$\begin{aligned} T^{1/2}r_{2T}^{-1} \mathcal{R}_p \hat{\theta}_{vec}^c &= [r_{1T}/r_{2T}] T^{1/2}r_{1T}^{-1}(\hat{\theta}_1^c - \theta_1^0) - T^{1/2}r_{2T}^{-1}(\hat{\theta}_2^c - \theta_2^0) + T^{1/2}r_{2T}^{-1}\theta_\Delta^0 \\ &= o_P(1) - \mathcal{O}_P(1) + T^{1/2}r_{2T}^{-1}\theta_\Delta^0 \\ r_{2T}^{-2}\hat{G}^c &\xrightarrow{P} \text{diag}[O, G_2^c] \Leftrightarrow r_{2T}^{-2} \mathcal{R}_p \hat{G}^c \mathcal{R}_p' \xrightarrow{P} G_2^c, \text{ so } T^{-1}r_{2T}^2 \text{Wald}_T^c \xrightarrow{P} \theta_\Delta^{0'}(G_2^c)^{-1} \theta_\Delta^0. \blacksquare \end{aligned}$$

• **Proof of Theorem 8: Efficiency of reduced form estimators**

$V_{MOD,\Pi}$ is explicitly defined as in the Theorem with $\mathcal{G}'_*\mathcal{G}_* = \mathcal{H}'_*\mathcal{E}_*^{-1/2}\mathcal{M}_{\mathcal{J}_*}\mathcal{E}_*^{-1/2}\mathcal{H}_*)^{-1}$ and,

$$\mathcal{E}_* = \text{diag}(S_{u,1} - S'_{uv,1}S_{v,1}^{-1}S_{uv,1}, S_{u,2} - S'_{uv,2}S_{v,2}^{-1}S_{uv,2}), \quad \mathcal{M}_{\mathcal{J}_*} = I - \mathcal{J}_*(\mathcal{J}'_*\mathcal{J}_*)^{-1}\mathcal{J}'_*,$$

$$\mathcal{J}_* = \mathcal{E}_*^{-1/2}\Gamma_1, \quad \mathcal{H} = \text{diag}(S'_{uv,1}S_{v,1}^{-1}, S'_{uv,2}S_{v,2}^{-1}) \Gamma_2, \quad \Gamma_1 = \text{diag}(Q_1\Pi^a, Q_2\Pi^a), \quad \Gamma_2 = \begin{bmatrix} I_{p_2} \otimes Q_1 \\ I_{p_2} \otimes Q_2. \end{bmatrix}$$

- (i) The distribution of $\hat{\Pi}$ is derived by usual OLS asymptotics; for $\hat{\Pi}_{GMM}$ with the optimal weighting matrix $(\hat{S}_v^a)^{-1} \xrightarrow{P} \text{diag}(S_{v,1}, S_{v,2})$, we have:

$$\begin{aligned} T^{1/2}(\hat{\Pi}_{GMM} - \Pi) &= [(T^{-1}W'\overline{W}^0)(\hat{S}_v^a)^{-1}(T^{-1}\overline{W}^{0'}W)]^{-1}(T^{-1}W'\overline{W}^0)(\hat{S}_v^a)^{-1}T^{-1/2}\overline{W}^0v \\ &\quad T^{-1/2}\overline{W}^0v \xrightarrow{d} \mathcal{N}(0, \text{diag}(S_{v,1}, S_{v,2})), \quad T^{-1}\overline{W}^{0'}W \xrightarrow{P} \text{vec}(Q_1, Q_2), \\ \Rightarrow T^{1/2}(\hat{\Pi}_{GMM} - \Pi) &\xrightarrow{d} \mathcal{N}(0, (Q_1S_{v,1}^{-1}Q_1 + Q_2S_{v,2}^{-1}Q_2)^{-1}). \end{aligned}$$

The distribution of $\hat{\Pi}_{MOD}$ follows from Theorem 4, by similar arguments as for $V_{MOD,vec}$.

- (ii) $V_{OLS,\Pi} = (Q_1 + Q_2)^{-1}(S_{v,1} + S_{v,2})(Q_1 + Q_2)^{-1}$ and $V_{GMM,\Pi} = (Q_1S_{v,1}^{-1}Q_1 + Q_2S_{v,2}^{-1}Q_2)^{-1}$. Using similar arguments as for Theorem 2(ii), but replacing μ_i with $Q_i a$, for any $q \times 1$ vector a , it follows that $V_{GMM,\Pi} \leq V_{OLS,\Pi}$, with equality iff $S_{v,1}^{-1}Q_1 a = S_{v,2}^{-1}Q_2 a$ for all a . Similarly to Theorem 5(i), $V_{MOD,\Pi} \leq V_{GMM,\Pi}$, because:

$$\begin{aligned} \mathcal{H}_* &= \mathcal{S}_{1,2}\mathcal{S}_{2,2}^{-1}\Gamma_2 = -\text{vec}(S'_{uv,i}S_{v,i}^{-1}Q_i) \stackrel{\text{def}}{=} -\text{vec}(\Gamma_{i*}), \\ \mathcal{E}_* &= \text{diag}(S_{u,i} - S'_{uv,i}S_{v,i}^{-1}S_{uv,i}) \stackrel{\text{def}}{=} \text{diag}(\mathcal{E}_{i*}^{-1/2}). \end{aligned}$$

So $\mathcal{E}_{i^*}, \mathbf{\Gamma}_{i^*}$ play the role of $\mathcal{E}_i, \mathbf{\Gamma}_i$ in the proof of Theorem 5(i), and $V_{MOD,\Pi} \leq V_{GMM,\Pi}$ because $V_{MOD,\Pi} = V_{GMM,\Pi} \Rightarrow \mathcal{E}_{2^*}^{-1/2} \mathbf{\Gamma}_{2^*} = 0 \Leftrightarrow \mathbf{\Gamma}_{2^*} = 0 \Leftrightarrow S'_{uv,2} S_{v,2}^{-1} Q_2 = 0$, which cannot hold.

• (iii) From the above, it follows that even under Assumptions 5-6, $V_{MOD,\Pi} \leq V_{GMM,\Pi}$. However, from (ii), $V_{GMM,\Pi} = V_{OLS,\Pi}$ iff $S_{v,1}^{-1} Q_1 a = S_{v,2}^{-1} Q_2 a$ for all vectors a . Under Assumption 5, $S_{v,i}^{-1} Q_i = (\lambda_i^0 S_v)^{-1} \lambda_i^0 Q = S_v^{-1} Q$, so $V_{GMM,\Pi} = V_{OLS,\Pi}$. Also, under Assumption 6, $S_{v,i} = \Phi_v Q_i$, so $S_{v,i}^{-1} Q_i = \Phi_v I_q$. ■.

C Results of the Monte-Carlo experiments

Benchmark case:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0023	0.0297	0.0298	0.1123	0.9340
2SLS	0.0023	0.0293	0.0294	0.1480	0.9860
GMM	0.0005	0.0342	0.0342	0.1305	0.9416
Decrease the R_1^2 from 0.2 to 0.05:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0031	0.0341	0.0342	0.1294	0.9342
2SLS	0.0030	0.0337	0.0338	0.1694	0.9852
GMM	0.0008	0.0416	0.0416	0.1590	0.9432
Decrease the R_1^2 from 0.2 to 0.01:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0035	0.0366	0.0367	0.1389	0.9356
2SLS	0.0034	0.0361	0.0363	0.1815	0.9850
GMM	0.0010	0.0464	0.0464	0.1775	0.9422
Increase sample size from 400 to 800:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0010	0.0193	0.0193	0.0755	0.9478
2SLS	0.0010	0.0192	0.0192	0.0988	0.9882
GMM	0.0003	0.0219	0.0219	0.0862	0.9496
Increase the number of IV from 3 to 6:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0025	0.0209	0.0211	0.0778	0.9322
2SLS	0.0026	0.0205	0.0206	0.1032	0.9830
GMM	0.0014	0.0240	0.0241	0.0917	0.9416
Increase endogeneity from 0.5 to 0.75:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0034	0.0296	0.0298	0.1122	0.9332
2SLS	0.0033	0.0293	0.0294	0.1478	0.9840
GMM	0.0008	0.0342	0.0342	0.1305	0.9400

Table 1: Experiment 1 with known break location in the homoskedastic case.

Benchmark case:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0015	0.0245	0.0246	0.0876	0.9342
2SLS	0.0015	0.0284	0.0284	0.1282	0.9894
GMM	0.0004	0.0305	0.0305	0.1106	0.9490
Decrease the R_1^2 from 0.2 to 0.05:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0019	0.0285	0.0286	0.1018	0.9344
2SLS	0.0020	0.0325	0.0326	0.1459	0.9884
GMM	0.0006	0.0372	0.0372	0.1350	0.9508
Decrease the R_1^2 from 0.2 to 0.01:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0022	0.0309	0.0309	0.1097	0.9358
2SLS	0.0023	0.0349	0.0350	0.1561	0.9874
GMM	0.0007	0.0414	0.0414	0.1508	0.9522
Increase sample size from 400 to 800:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0006	0.0171	0.0171	0.0621	0.9340
2SLS	0.0007	0.0203	0.0203	0.0895	0.9880
GMM	-0.0001	0.0201	0.0201	0.0746	0.9502
Increase the number of IV from 3 to 6:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0014	0.0165	0.0166	0.0578	0.9198
2SLS	0.0015	0.0211	0.0211	0.0899	0.9874
GMM	0.0005	0.0196	0.0196	0.0728	0.9408
Increase endogeneity from 0.5 to 0.75:					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0023	0.0245	0.0247	0.0875	0.9314
2SLS	0.0023	0.0284	0.0285	0.1281	0.9884
GMM	0.0006	0.0306	0.0306	0.1106	0.9498

Table 2: Experiment 1 with known break location in the Garch case.

Break size is equal to 1					
Monte-Carlo average of estimated break location is $\hat{T} = 161.35$					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0030	0.0290	0.0292	0.1124	0.9380
2SLS	0.0029	0.0287	0.0288	0.3827	1.0000
GMM	-0.0003	0.0338	0.0338	0.1307	0.9490
Break size is equal to 0.5					
Monte-Carlo average of estimated break location is $\hat{T} = 162.2$					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0083	0.0461	0.0468	0.1771	0.9310
2SLS	0.0080	0.0454	0.0460	0.2865	0.9970
GMM	-0.0000	0.0508	0.0508	0.1964	0.9470
Break size is equal to 0.2					
Monte-Carlo average of estimated break location is $\hat{T} = 172.4$					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0229	0.0686	0.0723	0.2619	0.9190
2SLS	0.0222	0.0675	0.0710	0.2430	0.9095
GMM	0.0008	0.0729	0.0729	0.2815	0.9475

Table 3: Experiment 1 with unknown location of the break in the benchmark homoskedastic case. The location of the break is estimated for three different break sizes (1, 0.5 and 0.2), and the true break location is $T^* = 160$.

Break size is equal to 1					
Monte-Carlo average of estimated break location is $\hat{T} = 161.35$					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0021	0.0245	0.0246	0.0876	0.9346
2SLS	0.0021	0.0284	0.0285	0.4288	0.9978
GMM	0.0004	0.0305	0.0305	0.1106	0.9490
Break size is equal to 0.5					
Monte-Carlo average of estimated break location is $\hat{T} = 162.2$					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0057	0.0382	0.0386	0.1362	0.9290
2SLS	0.0062	0.0454	0.0458	0.3189	0.9846
GMM	0.0009	0.0459	0.0459	0.1660	0.9482
Break size is equal to 0.2					
Monte-Carlo average of estimated break location is $\hat{T} = 172.4$					
Estimator	Bias	Std dev	RMSE	Length	Coverage
MOD	0.0158	0.0557	0.0579	0.1976	0.9058
2SLS	0.0176	0.0696	0.0717	0.2687	0.8996
GMM	0.0018	0.0658	0.0658	0.2375	0.9458

Table 4: Experiment 1 with unknown location of the break in the benchmark Garch case. The location of the break is estimated for three different break sizes (1, 0.5 and 0.2), and the true break location is $T^* = 160$.

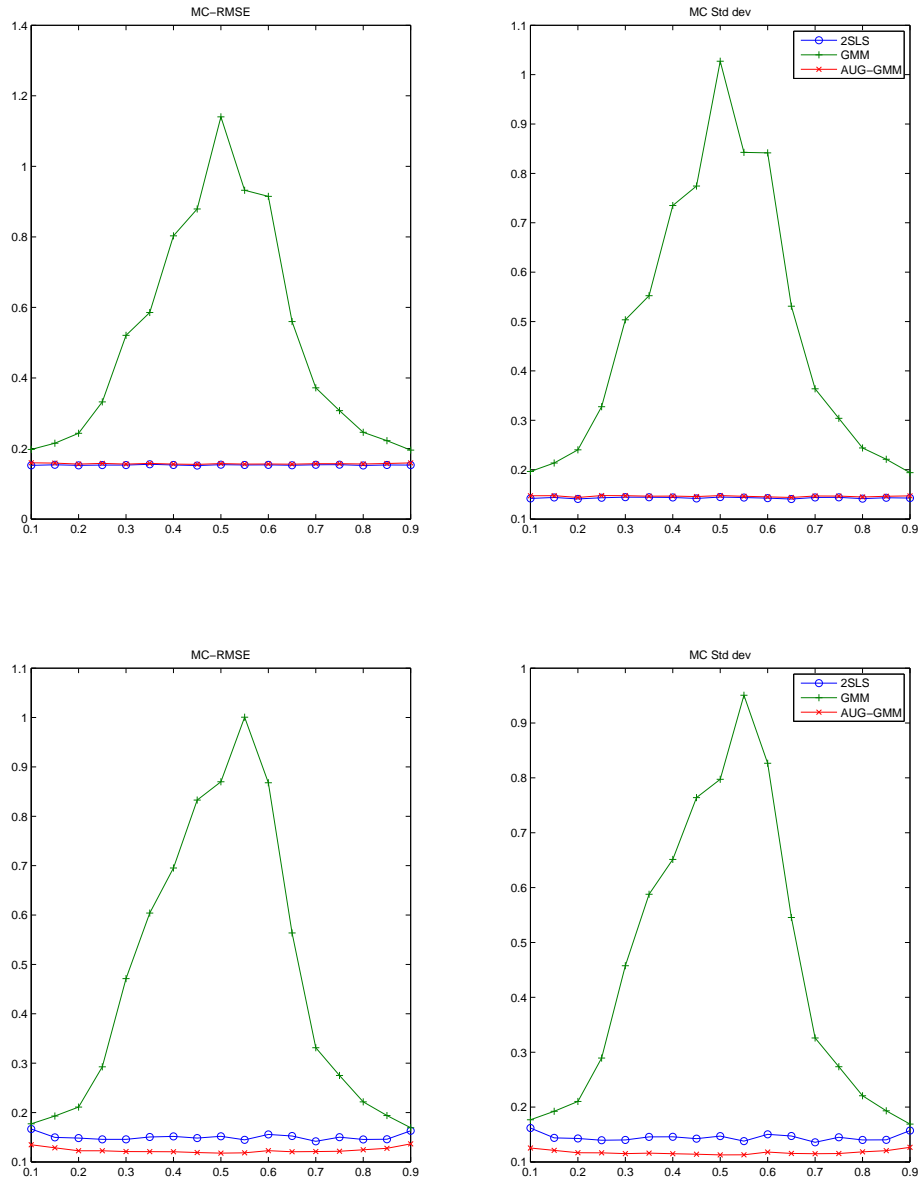


Figure 1: Experiment 2 for model (i) in the homoskedastic case (top) and in the Garch case (bottom). Left panel is RMSE and right panel is Standard deviation for MOD (red x), 2SLS (blue o), and GMM (green +).

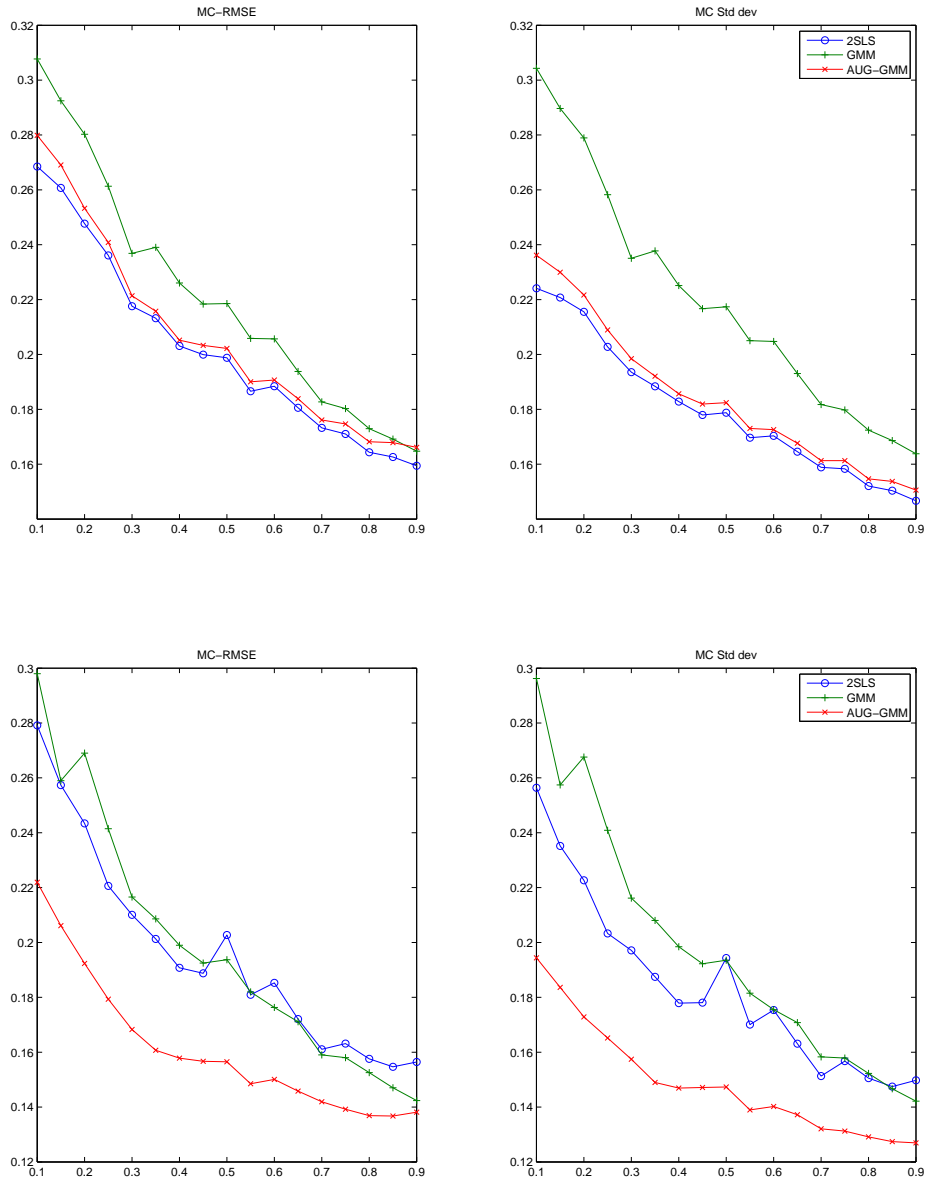
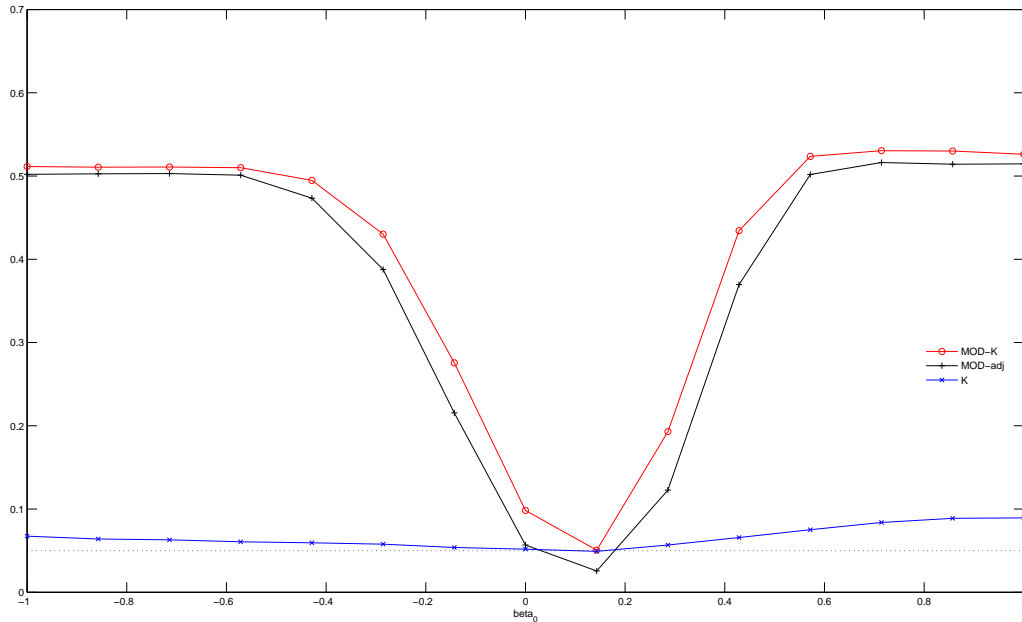


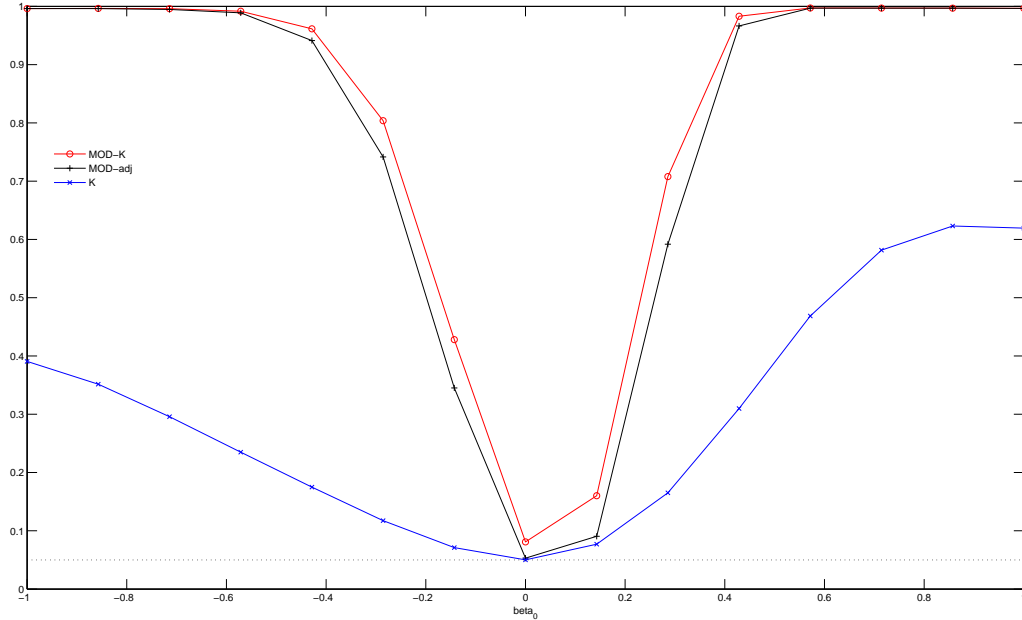
Figure 2: Experiment 2 for model (iii) in the homoskedastic case (top) and in the Garch case (bottom). Left panel is RMSE and right panel is Standard deviation for MOD (red x), 2SLS (blue o), and GMM (green +).



Rejection probabilities (at the true value $\beta_0 = 0$):					
MOD-K	MOD-K-adj	K	MOD-AR	MOD-AR-adj	AR
<i>(represented on the graph)</i>			<i>(not represented on the graph)</i>		
0.0982	0.0568	0.0518	0.0858	0.0496	0.0240

Probability of detecting the break	0.9994
Probability of weak identification* (ignoring the break info)	1
Probability of weak identification* (with break info)	0.5056
* We use Staiger and Stock's rule-of-thumb to test weak identification.	

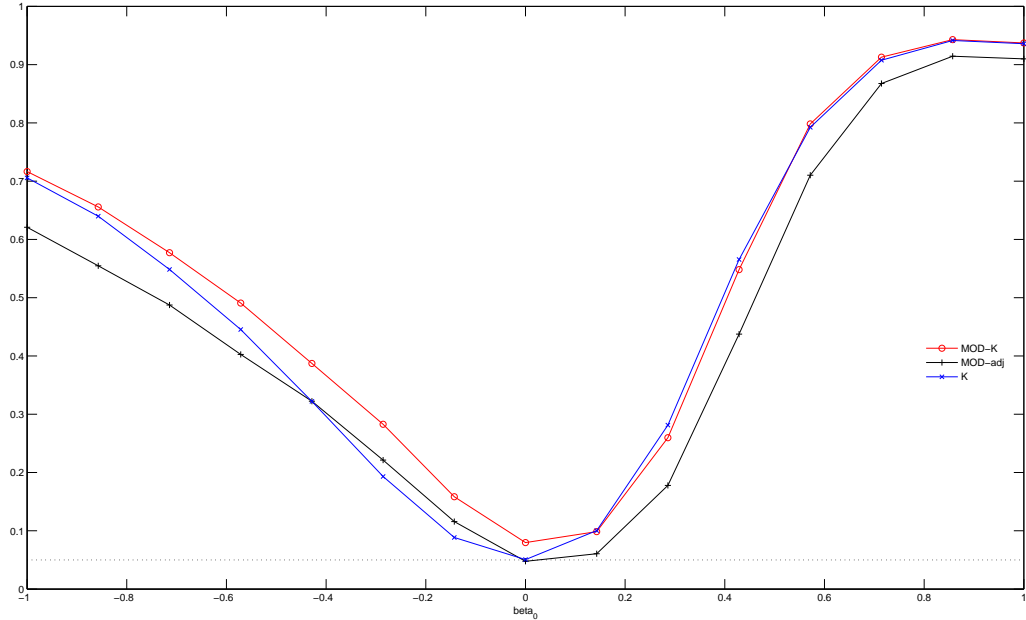
Figure 3: Experiment 3 for model (i) in the homoskedastic case, no change in R^2 : $R_1^2 = R_2^2 = 0.1$. We represent the power curves when testing $H_0 : \beta = \beta_0$ at $\alpha = 5\%$ using (a) MOD-K (red o) where we either use MOD or K depending on the tests for break and weakness; (b) MOD-adj (black +) after adjusting the size of the test using a Bonferroni-type correction, $\alpha_{adj} = \alpha/2 = 0.025$; (c) K (blue x) where we use Kleibergen's procedure ignoring the break.



Rejection probabilities (at the true value $\beta_0 = 0$):					
MOD-K	MOD-K-adj	K	MOD-AR	MOD-AR-adj	AR
<i>(represented on the graph)</i>			<i>(not represented on the graph)</i>		
0.0810	0.0530	0.0502	0.0808	0.0530	0.0244

Probability of detecting the break	1
Probability of weak identification* (ignoring the break info)	0.9858
Probability of weak identification* (with break info)	0.0038
* We use Staiger and Stock's rule-of-thumb to test weak identification.	

Figure 4: Experiment 3 for model (ii) in the homoskedastic case, R^2 larger over the second subsample: $R_1^2 = 0.1$ and $R_2^2 = 0.22$. We represent the power curves when testing $H_0 : \beta = \beta_0$ at $\alpha = 5\%$ using (a) MOD-K (red o) where we either use MOD or K depending on the tests for break and weakness; (b) MOD-adj (black +) is MOD-K after adjusting the size of the test using a Bonferroni-type correction, $\alpha_{adj} = \alpha/2 = 0.025$; (c) K (blue x) where we use Kleibergen's procedure ignoring the break.



Rejection probabilities (at the true value $\beta_0 = 0$):					
MOD-K	MOD-K-adj	K	MOD-AR	MOD-AR-adj	AR
<i>(represented on the graph)</i>			<i>(not represented on the graph)</i>		
0.0798	0.0476	0.0506	0.0620	0.0372	0.0238

Probability of detecting the break	0.2298
Probability of weak identification* (ignoring the break info)	0.7764
Probability of weak identification* (with break info)	0.7510
* We use Staiger and Stock's rule-of-thumb to test weak identification.	

Figure 5: Experiment 3 for model (iii) in the homoskedastic case, R^2 smaller over the second subsample: $R_1^2 = 0.1$ and $R_2^2 = 0.025$. We represent the power curves when testing $H_0 : \beta = \beta_0$ at $\alpha = 5\%$ using (a) MOD-K (red o) where we either use MOD or K depending on the tests for break and weakness; (b) MOD-adj (black +) is MOD-K after adjusting the size of the test using a Bonferroni-type correction, $\alpha_{adj} = \alpha/2 = 0.025$; (c) K (blue x) where we use Kleibergen's procedure ignoring the break.