Testing for a Threshold in Models with Endogenous Regressors

Rothfelder, Mario; Boldea, Otilia

Publication date:
2016

Document Version
Early version, also known as pre-print

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Rothfelder, M., & Boldea, O. (2016). *Testing for a Threshold in Models with Endogenous Regressors*. (CentER Discussion Paper; Vol. 2016-029). CentER, Center for Economic Research.

No. 2016-029

TESTING FOR A THRESHOLD IN MODELS WITH ENDOGENOUS REGRESSORS

By

Mario P. Rothfelder, Otilia Boldea

3 August 2016

ISSN 0924-7815
ISSN 2213-9532

Testing for a Threshold in Models with Endogenous Regressors*

Mario P. Rothfelder† Otilia Boldea‡

July 27, 2016

Using 2SLS estimation, we propose two tests for a threshold in models with endogenous regressors: a sup LR test and a sup Wald test. Here, the 2SLS estimation is not conventional because it uses additional information about the first-stage being linear or not. Because of this additional information, our tests can be more accurate than the threshold test in Caner and Hansen (2004) which is based on conventional GMM estimation.

We derive the asymptotic distributions of the two tests for a linear and for a threshold reduced form. In both cases, the distributions are non-pivotal, and we propose obtaining critical values via a fixed regressor wild bootstrap. Our simulations show that in small samples, the GMM test of Caner and Hansen (2004) can be severely oversized under heteroskedasticity, while the 2SLS tests we propose are much closer to their nominal size.

We use our tests to investigate the common claim that the government spending multiplier is larger close to the zero lower bound, and therefore that the governments should have spent more in the recent crisis. We find no empirical support for this claim.

Keywords: 2SLS, GMM, threshold tests, wild bootstrap
JEL Classification: C12, C13, C21, C22

*We are grateful for useful comments to Inna Filippenko, Bruce Hansen, Siem Jan Koopman, David Pacini and Hashem Pesaran, as well as to the participants at the Tilburg Brown Bag Seminar 2014, the Netherlands Econometrics Study Group 2014-2015, IAAE 2015, Bristol Econometrics Study Group 2015, NBER-NSF 2015, and NASM 2016. The second author gratefully acknowledges the NWO VENI Grant 415-11-001.

†Tilburg University, Dept. of Econometrics and CentER, Email: m.p.rothfelder@uvt.nl
‡Tilburg University, Dept. of Econometrics and CentER, Email: o.boldea@uvt.nl

1. Introduction

Threshold models are widely used in economics to model unemployment, output, growth, bank profits, asset prices, exchange rates, and interest rates. See Hansen (2011) for a survey of economic applications.

Pioneered by Howell Tong - see e.g. Tong (1990), threshold models with exogenous regressors have been widely studied and their asymptotic theory is well known.[1] Even though exogeneity is violated in many economic applications, papers on threshold regression with *endogenous regressors* remain relatively scarce. They were pioneered by Caner and Hansen (2004), who show that when regressors are endogenous but the threshold variable is exogenous, the threshold parameter can be estimated by minimizing a two stage least squares (2SLS) criterion over values of the threshold variable encountered in the sample.

In general, the applied researcher needs to decide whether there is a threshold to begin with. This can be done via testing for an unknown threshold. For example, the government spending multiplier is often conjectured to be larger in regimes where the nominal interest rate is close to the zero lower bound - see Eggertsson (2010) and Christiano et al. (2011).[2] This conjecture can be validated by testing whether there is a threshold driven by low interest rates (which we do in this paper). Another example is testing whether growth slows down when the debt to GDP ratio is high - see Reinhart and Rogoff (2010) (tests for this conjecture albeit using exogenous regressors can be found in Lee et al. (2014) and Hansen (2016) a.o.). Many more examples can be found in Hansen (2011).

In this paper, we develop 2SLS tests for no threshold against the alternative of one unknown threshold for models with endogenous regressors. Caner and Hansen (2004) already proposed a GMM sup Wald test for the same hypothesis. Here, we show that this test is severely oversized in small, heteroskedastic samples. We propose instead two 2SLS tests (a 2SLS sup LR test and a 2SLS sup Wald test), which we show have superior size properties in finite samples. The superior size stems from how the 2SLS estimators are constructed. They are not conventional, because they use additional information about the first stage, while the conventional GMM estimators in Caner and Hansen (2004) do not use any information about the first stage. With this additional information, we show that the 2SLS estimators can be more accurate than the conventional GMM estimators, and that they lead to better sized tests in finite samples.[3]

The additional information we use is whether there is a threshold in the first stage. We consider two cases: the first stage is a linear model and the first stage is a threshold

[1] See a.o. Hansen (1996, 1999, 2000) and Gonzalo and Wolf (2005) for inference, Gonzalo and Pitarakis (2002) for multiple threshold regression and model selection, Caner and Hansen (2001) and Gonzalo and Pitarakis (2006) for threshold regression with unit roots, Seo and Linton (2007) for smoothed estimators of threshold models, Lee et al. (2011) for testing for thresholds, and Hansen (2016) for threshold regressions with a kink.

[2] This can happen because when the monetary policy is less effective, fiscal stimulus can quickly lower real interest rates by raising inflation, resulting in potentially large multiplier effects.

[3] These unconventional 2SLS estimators were already proposed in Caner and Hansen (2004), but not for constructing tests for a threshold.

model.[4] We compute the 2SLS tests for each case separately, and show that their null asymptotic distributions depend on the data and on the case considered. Nevertheless, critical values are straightforward to compute via the wild bootstrap, so these tests are easily implemented in practice. To our knowledge, this is the first paper to propose and analyze 2SLS tests for a threshold.

We study the properties of both tests via simulation. We generate critical values via a fixed regressor wild bootstrap that we describe in this paper. We find that the 2SLS sup LR and the 2SLS sup Wald test are either correctly sized or slightly undersized. In contrast, the GMM sup Wald test is correctly sized under homoskedasticity, but under heteroskedasticity, it is severely oversized.[5] This holds for both linear and threshold reduced forms. As the sample size grows large, all tests approach their nominal sizes. We conclude that the 2SLS tests are valuable complementary diagnostics to the GMM test for a threshold, especially under heteroskedasticity.

Our paper is closely related to two papers in the break-point literature - Hall et al. (2012) and Boldea et al. (2016). Both papers study the 2SLS sup LR and 2SLS sup Wald tests for a break, the first one for a linear first stage, the second one for a first stage with a break. The asymptotic distributions for the break-point tests are pivotal in the first paper and depend on the break in the first stage in the second paper. In contrast, we find that the asymptotic distributions of the threshold tests are non-pivotal in both cases, a linear or a threshold first stage. Moreover, they are very different than the break-point distributions, and we show that they only coincide in unrealistic threshold models.

The paper is also related to Magnusson and Mavroeidis (2014), who use information about break-points in the first stage (and in general break-points in the derivative of the moment conditions) to improve efficiency of tests for moment conditions. It is also related to Antoine and Boldea (2015a) and Antoine and Boldea (2015b): the first uses breaks in the Hessian of the GMM minimand and the second uses full sample RF information. Both papers focus on more efficient estimation, while we focus on improved testing.

It should be noted that we allow for endogenous regressors, but not for endogenous threshold variables. For the latter, see Kourtellos et al. (2015). Also, to account for regressor endogeneity, we make use of instruments for constructing parametric test statistics for thresholds. As a result, our tests have nontrivial local power for $O(T^{-1/2})$ threshold shifts. This is in contrast with Yu and Phillips (2014), who does not use instruments, but rather local shifts around the threshold to construct a nonparametric threshold test. As a result, his test covers more general models, at the cost of losing

[4]Caner and Hansen (2004) consider the same cases for estimating the threshold parameter, but not for testing for a threshold. One can distinguish between the two cases by testing for a threshold in the first stage, using currently available tests such as the OLS sup Wald test in Hansen (1996).

[5]Note that unlike the Wald test for classical hypotheses, the (heteroskedasticity-robust) sup Wald test for the null hypothesis for an unknown threshold does not have a pivotal null distribution. That means that correcting for heteroskedasticity (and therefore using Wald tests instead of LR tests) does not necessarily result in better size properties for the sup Wald test compared to the sup LR test; this is indeed what we find in the simulations.

power in $O(T^{-1/2})$ neighborhoods.

We apply our tests to check whether the US government spending multiplier is larger in regimes where the nominal interest rate is close to the zero lower bound. We find strong evidence for a threshold in the first stage, but not in the second stage. Therefore, we find no empirical evidence that in the sample considered, the government spending multiplier is larger at the zero lower bound.

This paper is organized as follows. Section 2 introduces the threshold model. Section 3 defines the 2SLS and GMM estimators, and theoretically and numerically motivates the use of 2SLS estimators. Section 4 defines the new 2SLS test statistics and derives their asymptotic distributions. Section 5 describes the existing GMM test of Caner and Hansen (2004). Section 6 describes the fixed regressor wild bootstrap, and illustrates the small sample properties of all tests via simulations. Section 7 contains the empirical application. Section 8 concludes and Section 9 contains all tables and graphs. All the proofs are relegated to the Appendix, together with additional notation.

2. Threshold Model

Our framework is a linear model with a possible threshold at γ^0:

$$\begin{aligned} y_t &= \left(z_t^\top \theta_{1z}^0 + x_{1t}^\top \theta_{1x}^0\right) \mathbb{1}_{\{q_t \leq \gamma^0\}} + \left(z_t^\top \theta_{2z}^0 + x_{1t}^\top \theta_{2x}^0\right) \mathbb{1}_{\{q_t > \gamma^0\}} + \epsilon_t \\ &= w_t^\top \theta_1^0 \mathbb{1}_{\{q_t \leq \gamma^0\}} + w_t^\top \theta_2^0 \mathbb{1}_{\{q_t > \gamma^0\}} + \epsilon_t. \end{aligned}$$

Here, y_t is the dependent variable, z_t is a $p_1 \times 1$-vector of endogenous variables and x_{1t} a $p_2 \times 1$-vector of exogenous variables containing the intercept, and $w_t = (z_t^\top, x_{1t}^\top)^\top$. We set $p_1 + p_2 = p$. Also, q_t is the exogenous threshold variable (which can be a function of the exogenous regressors) and $\mathbb{1}_{\{\mathcal{A}\}}$ denotes the indicator function on the set $\mathcal{A}$. Furthermore, for $i = 1, 2$, θ_{iz}^0 are $p_1 \times 1$-vectors of slope parameters associated with z_t, θ_{ix}^0 are $p_2 \times 1$-vectors of the slope parameters associated with x_{1t} and $\gamma^0 \in \Gamma^0 = [\gamma_{min}, \gamma_{max}]$, its compact support.[6] The second equation is just a more compact way of writing the first, with $w_t = (z_t^\top, x_{1t}^\top)^\top$ being the augmented regressors, and $\theta_i^0 = (\theta_{iz}^{0\top}, \theta_{ix}^{0\top})^\top$ being $p \times 1$-vectors of the slope parameters, for $i = 1, 2$.

We assume that z_t is endogenous ($\mathbb{E}[\epsilon_t] = 0$; $\mathbb{E}[z_t \epsilon_t] \neq 0$) and strong instruments x_t are available; these instruments include x_{1t}, the exogenous regressors.

As in Caner and Hansen (2004), we consider two different specifications for the first stage (which we call *reduced form or RF* for lack of better terminology): a linear reduced form (LRF), given by

$$z_t = \Pi^{0\top} x_t + u_t,$$

and a threshold reduced form (TRF) given by

$$z_t = \Pi_1^{0\top} x_t \mathbb{1}_{\{q_t \leq \rho^0\}} + \Pi_2^{0\top} x_t \mathbb{1}_{\{q_t > \rho^0\}} + u_t.$$

[6] We can allow for $\Gamma^0 = \mathbb{R}$. However, the end-points of the support of q_t, even when infinite, are relevant for simulating asymptotic $p-$values. Without further information, the only end-points we observe are those in the sample: the minimum and maximum value of q_t, which we call $\gamma_{min}, \gamma_{max}$; therefore, we fix $\Gamma^0 = [\gamma_{min}, \gamma_{max}]$.

In both specifications for the RF, $x_t = (x_{1t}^\top, x_{2t}^\top)^\top$ is a $q \times 1$-vector with $q \geq p$, $q = p_2 + q_1$; Π^0, Π_1^0 and Π_2^0 are $q \times p_1$-matrices of the RF slope parameters; $\rho^0 \in \Gamma^0$ is the RF threshold parameter, possibly different than γ^0, with the same support Γ^0.

As common in the threshold literature, we assume that ϵ_t and u_t are martingale differences, i.e. $\mathbb{E}[\epsilon_t|\mathfrak{F}_t] = 0$ and $\mathbb{E}[u_t|\mathfrak{F}_t] = \mathbf{0}$, $\mathfrak{F}_t = \sigma\{q_{t-s}, x_{t-s}, u_{t-s-1}, \epsilon_{t-s-1}|s \geq 0\}$, and $(x_t^\top, z_t^\top)^\top$ is measurable with respect to $\mathfrak{F}_t$. This assumption implies that the threshold variable q_t is exogenous, and so are the instruments x_t.

Next, we write the equations above in matrix form. To do so, stack all observations in the following T-row matrices:

$$\begin{aligned} X_1^\rho &= \left(x_t^\top \mathbb{1}_{\{q_t \leq \rho\}}\right)_{t=1,\ldots,T} & X_2^\rho &= \left(x_t^\top \mathbb{1}_{\{q_t > \rho\}}\right)_{t=1,\ldots,T} \\ W_1^\gamma &= \left(w_t^\top \mathbb{1}_{\{q_t \leq \gamma\}}\right)_{t=1,\ldots,T} & W_2^\gamma &= \left(w_t^\top \mathbb{1}_{\{q_t > \gamma\}}\right)_{t=1,\ldots,T}. \end{aligned}$$

Let Y, X, Z, W, ϵ and u be the matrices stacking observations $t = 1, \ldots, T$. Then the LRF is:

$$Z = X\Pi^0 + u \tag{2.1}$$

and the TRF is:

$$Z = X_1^{\rho^0}\Pi_1^0 + X_2^{\rho^0}\Pi_2^0 + u. \tag{2.2}$$

The equation of interest - which can arise from a structural model and for lack of better terminology is called the structural form (SF) - is, for a threshold parameter γ^0:

$$Y = W_1^{\gamma^0}\theta_1^0 + W_2^{\gamma^0}\theta_2^0 + \epsilon. \tag{2.3}$$

If there is no SF threshold, $\theta_1^0 = \theta_2^0 = \theta^0$, and the SF is $Y = W\theta^0 + \epsilon$.

3. 2SLS versus GMM estimation

In this section, we motivate the use of 2SLS estimation for constructing test statistics.

We are interested in testing for a SF threshold, the null hypothesis being $\mathbb{H}_0 : \theta_1^0 = \theta_2^0$ in (2.3). Because γ^0 is usually unknown and it is a nuisance parameter under the null hypothesis, a common practice is to calculate a series of test statistics, each for a given $\gamma \in \Gamma$ (where $\Gamma \subset \Gamma^0$), and then to take the supremum over these quantities to obtain a single test statistic for the null of no threshold against the alternative of one threshold. For example, Hansen (1996) and Caner and Hansen (2004) construct such tests.

In the presence of endogenous regressor, to test for $\mathbb{H}_0$, Caner and Hansen (2004) defines two-step GMM estimators of $\theta_i^0, (i = 1, 2)$ for each γ. These are conventional in the sense that by construction, they ignore any information about the RF. Specifically, for each $\gamma \in \Gamma$, where Γ is a closed interval in the support Γ^0, bounded away from the end-points of this support, and $i = 1, 2$:

$$\hat{\theta}_{i,GMM}^\gamma = \left(W_i^{\gamma\top} X_i^\gamma \hat{H}_{i,GMM}^{\epsilon^{-1}}(\gamma) X_i^{\gamma\top} W_i^\gamma\right)^{-1} \left(W_i^{\gamma\top} X_i^\gamma \hat{H}_{i,GMM}^{\epsilon}(\gamma) X_i^{\gamma\top} Y\right),$$

with estimated long-run variances:

$$\hat{H}^{\epsilon}_{1,GMM}(\gamma) = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}^2_{t,GMM}x_t x_t^{\top}\mathbb{1}_{\{q_t\leq\gamma\}}, \hat{H}^{\epsilon}_{2,GMM}(\gamma) = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}^2_{t,GMM}x_t x_t^{\top}\mathbb{1}_{\{q_t>\gamma\}},$$

where $\hat{\epsilon}_{t,GMM}$ is the t^{th} element of the $T\times 1$ vector $\hat{\epsilon}_{GMM} = y - W_1^{\gamma}\tilde{\theta}_{1,GMM}(\gamma) - W_2^{\gamma}\tilde{\theta}_{2,GMM}(\gamma)$, and $\tilde{\theta}_{i,GMM}(\gamma)$ are some preliminary first step GMM estimators of (2.3) for a given γ and $i = 1, 2$.[7]

If instead, we estimate (2.3) by 2SLS, we have no choice but to take into account the nature of the RF - linear model or threshold model - otherwise the resulting estimator of θ_i^0 may be inconsistent. These two cases - linear or threshold RF - have also been considered in Caner and Hansen (2004) for 2SLS slope estimators, but with the purpose of defining a consistent estimator the threshold parameter γ^0. If we do not know which case applies, we can test it via the sup Wald test in Hansen (1996).

For a linear RF (LRF), let:

$$\hat{Z} = X\hat{\Pi},\ \hat{W} = \left(\hat{Z}, X_1\right), \tag{3.1}$$

with $X_1 = (x_{1t}^{\top})_{t=1,\dots,T}$.

For a threshold RF (TRF), first estimate the threshold parameter ρ as in Caner and Hansen (2004):

$$\hat{\rho} = \underset{\rho\in\Gamma}{\text{argmin}}\ \det\left(\hat{u}(\rho)^{\top}\hat{u}(\rho)\right), \tag{3.2}$$

where $\hat{u}(\rho) = Z - X_1^{\rho}\hat{\Pi}_1(\rho) - X_2^{\rho}\hat{\Pi}_2(\rho)$ and $\hat{\Pi}_1(\rho), \hat{\Pi}_2(\rho)$ are the OLS estimators of Π_1^0, Π_2^0 in (2.2) for a given ρ:

$$\hat{\Pi}_1(\rho) = \left(X_1^{\rho\top}X_1^{\rho}\right)^{-1}X_1^{\rho\top}Z \tag{3.3}$$

$$\hat{\Pi}_2(\rho) = \left(X_2^{\rho\top}X_2^{\rho}\right)^{-1}X_2^{\rho\top}Z. \tag{3.4}$$

With $\hat{\rho}$, the TRF slope parameter estimates are $\hat{\Pi}_1 = \hat{\Pi}_1(\hat{\rho})$, $\hat{\Pi}_2 = \hat{\Pi}_2(\hat{\rho})$.

Then:

$$\hat{Z} = \hat{\Pi}_1 X_1^{\hat{\rho}} + \hat{\Pi}_2 X_2^{\hat{\rho}}. \tag{3.5}$$

The second-stage of the 2SLS is standard. Construct $\hat{W} = \left(\hat{Z}, X_1\right)$, with $\hat{Z}$ defined in (3.1) for a LRF and (3.5) for a TRF, and the 2SLS estimators of θ_1^0, θ_2^0 for a given $\gamma\in\Gamma$ are for $i = 1, 2$.

$$\hat{\theta}_1^{\gamma} = \left(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma}\right)^{-1}\left(\hat{W}_1^{\gamma\top}Y\right) \tag{3.6}$$

$$\hat{\theta}_2^{\gamma} = \left(\hat{W}_2^{\gamma\top}\hat{W}_2^{\gamma}\right)^{-1}\left(\hat{W}_2^{\gamma\top}Y\right). \tag{3.7}$$

[7]Note that because W are already partitioned according to $\mathbb{1}_{\{q_t\leq\gamma\}}$, we have $W_i^{\gamma\top}Y = W_i^{\gamma\top}Y_i$.

Both the 2SLS and the GMM estimators defined here are consistent under standard assumptions, as shown in Caner and Hansen (2004). But the GMM estimators ignore potentially valid information about the RF. As a result, the GMM estimators can be less efficient than the 2SLS estimators. This result is formalized below.

Theorem 1 (2SLS versus GMM).
Assume the SF is (2.3) with the TRF (2.2), one endogenous regressor, one instrument and no exogenous regressors ($p = q = p_1 = 1$), and impose $\mathbb{H}_0$: $\theta^0_z = \theta^0_{1z} = \theta^0_{2z}$. Let ρ^0 be known and let Assumptions A.1–A.4 of Section 3 hold, with $\sigma^2_\epsilon = \mathrm{Var}(\epsilon_t)$, $\sigma^2 = \mathrm{Var}(\epsilon_t + u_t\theta^0_z)$, $\pi^0_1 = \Pi^0_1$ and $\pi^0_2 = \Pi^0_2$. Define $\lambda = P(q_t \leq \gamma)$ and $\mu^0 = \mathbb{P}(q_t \leq \rho^0)$. Then, for a given γ,
(i) For both $i = 1, 2$,

$$\sqrt{T}(\hat{\theta}^\gamma_i - \theta^0)] \xrightarrow{d} \mathcal{N}(0, V^*_{A,i}(\gamma)) \text{ and } \sqrt{T}(\hat{\theta}^\gamma_{i,GMM} - \theta^0)] \xrightarrow{d} \mathcal{N}(0, V^*_{i,GMM}(\gamma)),$$

where $V^*_{A,i}(\gamma)$ and $V^*_{i,GMM}(\gamma)$ are defined in Lemma 9 of the Appendix.
(ii)

$$\sigma^2 \leq \sigma^2_\epsilon \iff \left\{V^*_{i,GMM}(\gamma) \geq V^*_{A,i}(\gamma) \text{ for both } i = 1, 2 \text{ simultaneously}\right\}.$$

(iii) If the RF is in fact linear, that is, if $\pi^0_1 = \pi^0_2$, then:

$$\sigma^2 \leq \sigma^2_\epsilon \iff V^*_{1,GMM}(\gamma) \geq V^*_{A,1}(\gamma)$$
$$\sigma^2 \leq \sigma^2_\epsilon \iff V^*_{2,GMM}(\gamma) \geq V_{A,2}(\gamma).$$

(iv) $V^*_{i,GMM}(\rho^0) = V^*_{A,i}(\rho^0)$.

Note that Theorem 1 is derived under conditional homoskedasticity (imposed in Assumption A.2) and under independence of q_t and x_t (imposed in Assumption A.3).[8]

The intuition for the results in Theorem 1 is as follows. If *the sample* $\{t : q_t \leq \gamma\}$ *is used for both the RF and the SF* to compute 2SLS estimators, and the *same sample* is used for GMM estimators, then both these estimators are conventional. Therefore, the two-step GMM is asymptotically more efficient than the 2SLS, and asymptotically equivalent in the just-identified case. This is shown in Theorem 1(iv) where we set $\gamma = \rho^0$. However, when $\gamma \neq \rho^0$ the 2SLS estimators are not conventional. For example, if $\gamma \leq \rho^0$, in computing the 2SLS estimator over the sample $\{t : q_t \leq \gamma\}$, we *use information from the RF over a larger sample* $\{t : q_t \leq \rho^0\}$. Theorem 1 (ii) shows that this additional information leads to more efficient estimators if the 2SLS errors ($\epsilon_t + u_t\theta^0_z$) have smaller variance than the GMM errors ϵ_t. This efficiency result also holds if instead the RF is linear, as shown in Theorem 1(iii).

Theorem 1 is not just a theoretical result, as shown in the example below.

Example 1. *Suppose that $\pi^0_1 = 1$, $\pi^0_2 = 1.25$, $\rho^0 = 0.25$. Let $q_t \overset{iid}{\sim} \mathcal{N}(0, 1)$, $x_t \overset{iid}{\sim} \mathcal{N}(0, 1)$ and $\begin{bmatrix}\epsilon_t \\ u_t\end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(0, \begin{bmatrix}1 & 0.5 \\ 0.5 & 1\end{bmatrix}\right)$. Let $f_i(\lambda, \theta^0_z) = V^*_{A,i}(\gamma) - V^*_{GMM,i}(\gamma)$, and $\gamma \leq \rho^0$ (if $\gamma > \rho^0$, the first plot becomes the second and viceversa).*

[8] In more general cases, it is much harder to obtain a similar result analytically.

Note that in this case, $\sigma^2 - \sigma_\epsilon^2 = (\theta_z^0)(1 + \theta_z^0)$. From Theorem 1, if $\theta_z^0(1 + \theta_z^0) < 0$, $f_i(\lambda, \theta_z^0) < 0$ and both 2SLS estimators are more efficient.[9] From Example 1, $\mu^0 = \mathbb{E}\mathbb{1}_{\{q_t \leq \rho^0\}} = 0.5981$. In Figures 3.1 and 3.2 we plot $f_1(\lambda, \theta_z^0)$ and $f_2(\lambda, \theta_z^0)$ as functions of $\theta_z^0 \in [-1.5, 0.5]$ and $\lambda = P(q_t \leq \gamma) \in (0, \mu^0]$. The purple areas indicate parameter configurations where 2SLS is more efficient than GMM, and these are sizable areas of the parameter space.

Figure 3.1: Plot and Contour Plot of $f_1(\cdot)$

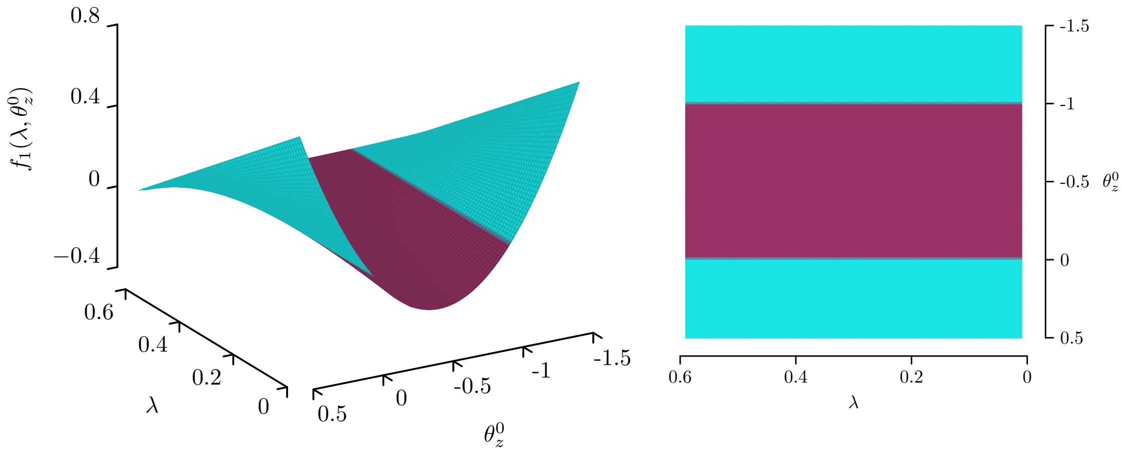
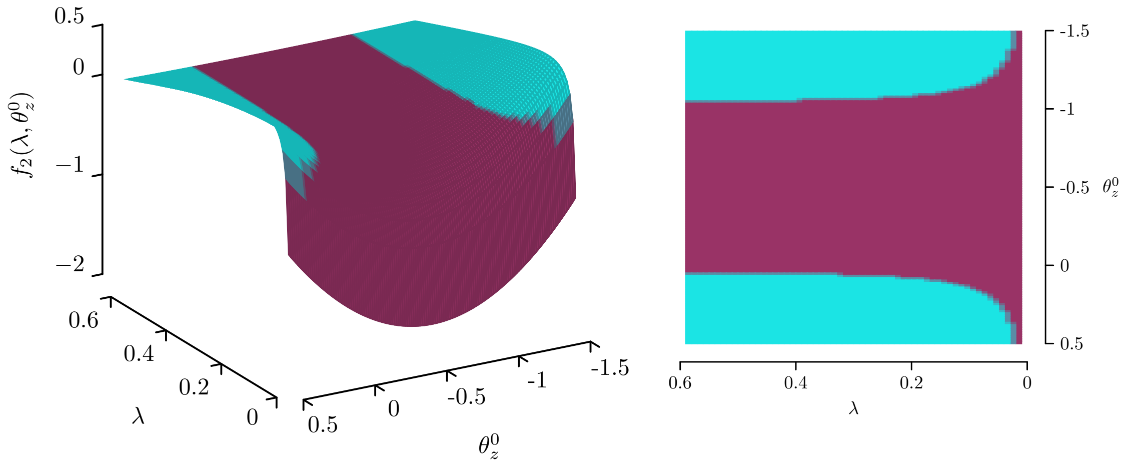


Figure 3.2: Plot and Contour Plot of $f_2(\cdot)$



Because 2SLS can be more accurate than GMM, so can be test statistics based on them. Tables 9.3 and 9.4 show that the GMM sup Wald test of Caner and Hansen (2004),

[9] As shown in the proof of Theorem 1, when $\sigma^2 > \sigma_\epsilon^2$, $\hat{\theta}_1^\gamma$ is less efficient than $\hat{\theta}_{1,GMM}^\gamma$, but $\hat{\theta}_2^\gamma$ can still be more efficient than $\hat{\theta}_{2,GMM}^\gamma$ depending on the DGP.

based on the GMM estimators, is severely oversized in small samples; at a nominal size of 5%, the empirical sizes reach up to 21.8% for 100 observations; they decrease with the sample size increasing, but they are still around $7-8\%$ for 1000 observations. Since many applications of threshold tests are macroeconomic applications, where a representative sample is around 500 observations, these size distortions are worrisome, as they will often lead to favor a threshold model when there is none. The same tables show that the 2SLS tests are either correctly sized or slightly undersized, but not oversized. This motivates us to propose 2SLS tests as complementary threshold diagnostics.

4. 2SLS Tests

4.1. Test Statistics

For a LRF, the first test statistic we propose is a sup LR test in the spirit of Davies (1977):

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma) = \sup_{\gamma\in\Gamma} \frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/(T-2p)}, \tag{4.1}$$

where SSR_0 and $SSR_1(\gamma)$ are the 2SLS sum of squared residuals under the null and the alternative hypotheses:

$$SSR_0 = (Y - \hat{W}\hat{\theta})^\top (Y - \hat{W}\hat{\theta}),$$
$$SSR_1(\gamma) = (Y_1^\gamma - \hat{W}_1^\gamma \hat{\theta}_1^\gamma)^\top (Y_1^\gamma - \hat{W}_1^\gamma \hat{\theta}_1^\gamma) + (Y_2^\gamma - \hat{W}_2^\gamma \hat{\theta}_2^\gamma)^\top (Y_2^\gamma - \hat{W}_2^\gamma \hat{\theta}_2^\gamma),$$

and where $\hat{\theta} = (\hat{W}^\top \hat{W})^{-1}\hat{W}^\top Y$ is the full-sample 2SLS estimator, and $\hat{W}, \hat{\theta}_1^\gamma, \hat{\theta}_2^\gamma$ are defined in Section 3 for a LRF.

A scaled version of this test is known as the sup F test in the break-point literature - see Bai and Perron (1998) for OLS and Hall et al. (2012) for 2SLS.

We also propose the sup Wald test:

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma) = \sup_{\gamma\in\Gamma} T \left[\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma\right]^\top \hat{V}^{-1}(\gamma) \left[\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma\right], \tag{4.2}$$

where $\hat{V}(\gamma)$ is defined in Definition A.2 of the Appendix, and unlike the 2SLS sup Wald test in Hall et al. (2012), it takes into account that the 2SLS estimators $\hat{\theta}_1^\gamma$ and $\hat{\theta}_2^\gamma$ are correlated through a full-sample first-stage.

For a TRF, the test statistics are calculated exactly as above, but taking into account the TRF when computing the first stage of the 2SLS estimation, as in (3.5). Therefore, $\sup_{\gamma\in\Gamma} W^{2SLS}_{T,TRF}(\gamma)$ is computed with $\hat{V}_A(\gamma)$ instead of $\hat{V}(\gamma)$, and $\hat{V}_A(\gamma)$ is defined in Definition A.3 of the Appendix.

4.2. Assumptions

Define

$$M_1(\gamma) = \mathbb{E}[x_t x_t^\top \mathbb{1}_{\{q_t\le\gamma\}}],\ \ M = M(\gamma_{\max}) = \mathbb{E}[x_t x_t^\top], \text{and}\ \ M_2(\gamma) = M - M_1(\gamma)$$

as the second moment functionals of the instruments x_t, where $\gamma \in \Gamma$. We impose the same assumptions as in Caner and Hansen (2004) below:

Assumption A.1.

1. *Let* $v_t = (\epsilon_t, u_t^\top)^\top$ *denote the compound error term. Then*

$$\mathbb{E}[v_t | \mathfrak{F}_t] = 0$$

   with $\mathfrak{F}_t = \sigma\{x_{t-s}, v_{t-s-1}, q_{t-s} | s \geq 0\}$.

2. *The series* $(\epsilon_t, u_t^\top, x_t^\top, z_t^\top, q_t)^\top$ *is strictly stationary and ergodic with* ρ*-mixing coefficient* $\rho(m) = \mathcal{O}(m^{-A})$ *for some* $A > \frac{a}{a-1}$ *and* $1 < a \leq 2$. *Also, for some* $b > a$,

$$\sup_t \mathbb{E}\|x_t\|_2^{4b} < \infty, \ \sup_t \mathbb{E}\|v_t\|_2^{4b} < \infty,$$

   with $\|\cdot\|_2$ *being the Euclidean norm, and* $\inf_{\gamma \in \Gamma} \det M_1(\gamma) > 0$.

3. *The density of* v_t *is absolutely continuous, bounded and positive everywhere.*

4. *The threshold variable* q_t *has a continuous pdf* $f(q_t)$ *with* $\sup_t |f(q_t)| < \infty$.

5. *The variance of the compound error term* v_t *is given by*

$$\mathbb{E}[v_t v_t^\top] = \Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \Sigma_{\epsilon,u}^\top \\ \Sigma_{\epsilon,u} & \Sigma_u \end{pmatrix},$$

   which is positive definite.

6. *Assume* Π^0 *(LRF) or* Π_1^0, Π_2^0 *(TRF) are full rank.*

A.1.1 is needed for threshold models, and it excludes autocorrelation in the errors. However, lagged regressors can enter both the SF and the RF. A.1.2 is standard for time series and is trivially satisfied for many cross-section models (note that even though we use the time series notation with index t, our results equally apply to cross section models). However, it precludes nonstationary processes. A.1.3 is needed in the TRF case in order to make asymptotic statements about the RF parameters in the spirit of Chan (1993). A.1.4 requires the support of q_t to be continuous; if it is discrete, the search over Γ is much easier to perform. A.1.5 allows conditional heteroskedastic errors and finally, A.1.6 says that x_t is a strong instrument.

Assumption A.2.

$$\mathbb{E}[v_t v_t^\top | \mathfrak{F}_{t-1}] = \Sigma = \begin{pmatrix} \Sigma_\epsilon & \Sigma_{\epsilon,u}^\top \\ \Sigma_{\epsilon,u} & \Sigma_u \end{pmatrix}.$$

Assumption A.2 is a conditional homoskedasticity assumption, which we only use for special case derivations.

Assumption A.3. *The threshold variable q_t and the vector of exogenous variables x_t are independent. i.e.*

$$q_t \perp x_t \ \forall t = 1, 2, ..., T.$$

Assumption A.3 is also quite strong and is only used to relate the results in this paper to those on break-point tests, not for the main results of the paper. It doesn't allow the threshold variable q_t to be one of the instrumental variables or exogenous regressors x_t, and is quite restrictive. However, it mimics break-point models, where the threshold is time, or more exactly, a fraction of the sample size, t/T.

Assumption A.4 (Identifiability)**.** *If we have a TRF as in (2.2), $\Pi_1^0 \neq \Pi_2^0$.*

Assumption A.4 states that if there is a TRF, the threshold effect is large. It is imposed for simplicity.

4.3. Asymptotic distributions with a LRF

To write the asymptotic distributions, define the "ratios"

$$R_i(\gamma) = M_i(\gamma)M^{-1}, i = 1, 2.$$

Also, let

$$\mathcal{GP}_{\text{mat},1}(\gamma) \quad \text{and} \quad \mathcal{GP}_{\text{mat}}$$

as $q \times (p_1 + 1)$-matrices where all columns are $q \times 1$ zero mean Gaussian processes, and the covariance kernels of $\mathcal{GP}_1(\gamma) = \text{vec}(\mathcal{GP}_{\text{mat},1}(\gamma))$ and $\mathcal{GP} = \text{vec}(\mathcal{GP}_{\text{mat}})$ are given by $\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \leq \gamma\}}]$ and $\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)]$. Let $\mathcal{GP}_{\text{mat}} = \mathcal{GP}_{\text{mat},1}(\gamma_{\max})$.
Also, let

$$A^0 = [\Pi^0, S^\top]^\top$$

be the augmented matrix of the RF slope parameters, where $S = [I_{p_2}, \mathbf{0}_{p_2 \times q_1}]$, I_{p_2} is the $p_2 \times p_2$ identity matrix and $\mathbf{0}_{p_2 \times q_1}$ a $p_2 \times q_1$ null matrix ($p_2 + q_1 = q$). Hence, $x_{1t} = Sx_t$ and $w_t = A^0 x_t + \bar{u}_t$, where $\bar{u}_t = (u_t^\top, \mathbf{0}_{1 \times q_1})^\top$. Define the matrices

$$C_1(\gamma) = A^0 M_1(\gamma) A^{0\top}, \quad C = C_1(\gamma_{\max}) = A^0 M A^{0\top}, \quad \text{and} \quad C_2(\gamma) = C - C_1(\gamma)$$

and the Gaussian process:

$$\mathcal{B}_1(\gamma) = A^0 \left[\mathcal{GP}_{\text{mat},1}(\gamma)\ \tilde{\theta}_z^0 - R_1(\gamma)\mathcal{GP}_{\text{mat}}\ \check{\theta}_z^0\right]$$

where $\tilde{\theta}_z^0 = (1, \theta_z^{0\top})^\top$ and $\check{\theta}_z^0 = (0, \theta_z^{0\top})^\top$. Finally, let:

$$\mathcal{E}(\gamma) = C_1^{-1}(\gamma)\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)$$

where $\mathcal{B}_2(\gamma) = \mathcal{B} - \mathcal{B}_1(\gamma)$ with $\mathcal{B} = \mathcal{B}_1(\gamma_{\max})$. Let

$$\sigma^2 = \sigma_\epsilon^2 + 2\Sigma_{\epsilon,u}^\top \theta_z^0 + \theta_z^{0\top} \Sigma_u \theta_z^0.$$

With this notation, the null distributions for a LRF are stated below.

Theorem 2 (Asymptotic Distributions LRF)**.** *Let Z be generated by* (2.1)*, Y be generated by* (2.3)*, and $\hat{Z}$ be calculated by* (3.1)*. Then under $\mathbb{H}_0$ and Assumption A.1,*
(i)

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \mathcal{E}^\top(\gamma)Q^{-1}(\gamma)\mathcal{E}(\gamma),$$

where $Q(\gamma) = \sigma^2 C_1^{-1}(\gamma)\ C\ C_2^{-1}(\gamma)$;
(ii)

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \mathcal{E}^\top(\gamma)V^{-1}(\gamma)\mathcal{E}(\gamma),$$

where $V(\gamma)$ is defined in Definition A.2 in the Appendix, and, in general, $V(\gamma) \neq Q(\gamma)$.

In both cases, the suprema taken are over $\gamma \in \Gamma$ and this deserves some explanation. For theoretical derivations, it suffices that Γ is a closed interval in the support Γ^0 and that it is bounded away from the end-points of $\Gamma^0 = [\gamma_{min}, \gamma_{max}]$. But in practice, searching over γ includes calculations over the subsamples $\{t : \mathbb{1}_{\{q_t\leq\gamma\}}\}$ and $\{t : \mathbb{1}_{\{q_t>\gamma\}}\}$, which means that the data needs to be sorted into quantiles of q_t. Therefore, in practice, Γ is a set that contains ordered values of q_t encountered in the sample, from a pre-defined lower quantile $\underline{\gamma}$ to predefined upper quantile $\overline{\gamma}$, where $\underline{\gamma} > \gamma_{min}$ and $\overline{\gamma} < \gamma_{max}$. We refer to these upper and lower quantiles as "cut-offs" in the simulation section, and in practice they are chosen so that the subsamples $\{t : \gamma_{min} \leq q_t \leq \underline{\gamma}\}$ and $\{t : \gamma_{max} \geq q_t \geq \overline{\gamma}\}$ are large enough to produce reliable estimates; example cut-offs are the 15% and the 85% quantiles of q_t.

Both asymptotic distributions depend on second moment functionals of the data and the parameters in the RF. But critical values can be calculated by the bootstrap described in Section 6.

As shown in Corollary A1 in the Appendix, the asymptotic distributions remain non-pivotal for both tests even when the errors are conditional homoskedastic. More importantly, because the 2SLS estimators are not conventional, the sup Wald and sup LR tests are in general NOT asymptotically equivalent under conditional homoskedasticity. However, they are equivalent in the just-identified case as shown in Corollary A1. They are also equivalent in the overidentified case, when x_t and q_t are independent, as stated below and proven in the Appendix.

Corollary 1 (to Theorem 2)**.** *Let Z be generated by* (2.1)*, Y be generated by* (2.3)*, and $\hat{Z}$ be calculated by* (3.1)*.Then, under $\mathbb{H}_0$ and Assumptions A.1-A.3,*

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma) \Rightarrow \sup_{\lambda\in\Lambda_\epsilon} \frac{\mathcal{BB}_p^\top(\lambda)\mathcal{BB}_p(\lambda)}{\lambda(1-\lambda)}, \qquad \sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma) \Rightarrow \sup_{\lambda\in\Lambda_\epsilon} \frac{\mathcal{BB}_p^\top(\lambda)\mathcal{BB}_p(\lambda)}{\lambda(1-\lambda)},$$

where $\mathcal{BB}_p(\lambda) = \mathcal{BM}_p(\lambda) - \lambda\mathcal{BM}_p(1)$, $\mathcal{BM}_p(\cdot)$ is a $p\times 1$-vector of independent standard Brownian motions, $\lambda = \text{Prob}(q_t \leq \gamma)$, $\Lambda_\epsilon = [\epsilon_1, 1-\epsilon_2]$, where $\epsilon_1 = \text{Prob}(q_t \leq \underline{\gamma})$, $\epsilon_2 = \text{Prob}(q_t \leq \overline{\gamma})$.

The distribution in Corollary 1 is identical that of the sup F and sup Wald break-point tests - see Andrews (1993), Bai and Perron (1998) and Hall et al. (2012) among others. This is due to similarities between threshold and break point models; a break-point model is a special case of a threshold model when $q_t = t/T$.[10] Critical values for these distributions can be found in Andrews (1993) and Bai and Perron (1998). However, $x_t \perp q_t$ is a case rarely encountered in practice, and we do not consider this case in our simulations.

4.4. Asymptotic distributions with a TRF

For this section, we assume that the RF has a threshold ρ^0 (TRF). For stating the asymptotic distributions, similar to A^0 in the previous section, we define

$$A_1^0 = [\Pi_1^0, S^\top]^\top \quad \text{and} \quad A_2^0 = [\Pi_2^0, S^\top]^\top. \tag{4.3}$$

Also, let $a \wedge b = \min(a, b)$ for generic scalars a, b, and define the matrices:

$$C_{A,1}(\gamma) = A_1^0 M_1(\gamma \wedge \rho^0) A_1^{0\top} + A_2^0 \left[M_1(\gamma) - M_1(\gamma \wedge \rho^0)\right] A_2^{0\top}, \tag{4.4}$$

and $C_{A,2} = C_A - C_{A,1}(\gamma)$, where:

$$C_A = C_{A,1}(\gamma_{\max}) = A_1^0 M_1(\rho^0) A_1^{0\top} + A_2^0 M_2(\rho^0) A_2^{0\top},$$

as well as, in line with Section 4, the "ratios"

$$R_i(\gamma; \rho^0) = M_i(\gamma) M_i^{-1}(\rho^0).$$

The TRF analogs to the LRF processes $B_1(\gamma)$ and $\mathcal{E}(\gamma)$ are defined as:

$$\begin{aligned}
\mathcal{B}_{A,1}(\gamma) = A_1^0 &\left[\mathcal{GP}_{\text{mat},1}(\gamma \wedge \rho^0)\tilde{\theta}_z^0 - R_1(\gamma \wedge \rho^0; \rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right] \\
&+ A_2^0 \left[\left(\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\gamma \wedge \rho^0)\right)\tilde{\theta}_z^0\right] \\
&- A_2^0 \left[\left(R_2(\gamma \wedge \rho^0; \rho^0) - R_2(\gamma; \rho^0)\right)\mathcal{GP}_{\text{mat},2}(\rho^0)\check{\theta}_z^0\right].
\end{aligned} \tag{4.5}$$

and

$$\mathcal{E}_A(\gamma) = C_{A,1}^{-1}(\gamma)\mathcal{B}_{A,1}(\gamma) - C_{A,2}^{-1}(\gamma)\mathcal{B}_{A,2}(\gamma) \tag{4.6}$$

where

$$\mathcal{B}_{A,2}(\gamma) = \mathcal{B}_A - \mathcal{B}_{A,1}(\gamma)$$

with

$$\mathcal{B}_A = \mathcal{B}_A(\gamma_{\max}) = A_1^0 \mathcal{GP}_{\text{mat},1}(\rho^0)(\tilde{\theta}_z^0 - \check{\theta}_z^0) + A_2^0 \mathcal{GP}_{\text{mat},2}(\rho^0)(\tilde{\theta}_z^0 - \check{\theta}_z^0).$$

The more complicated expressions in this case stem from the fact that the relative location of γ and ρ^0 influences the asymptotic distribution of our tests, as Theorem 3 shows.

[10]Note, however, that the asymptotics for break-point tests cannot be obtained as a special case of our results here because in general, break-point models are not strictly stationary.

Theorem 3 (Asymptotic Distributions TRF)**.** *Let Z be generated by* (2.2)*, Y be generated by* (2.3)*, and $\hat{Z}$ be calculated by* (3.5)*. Under $\mathbb{H}_0$ and Assumptions A.1 and A.4,*
(i)

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,TRF}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \mathcal{E}_A^\top(\gamma) Q_A^{-1}(\gamma)\mathcal{E}_A(\gamma),$$

where $Q_A(\gamma) = \sigma^2 C^{-1}_{A,1}(\gamma)\ C_A\ C^{-1}_{A,2}(\gamma)$;
(ii)

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,TRF}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \mathcal{E}_A^\top(\gamma) V_A^{-1}(\gamma)\mathcal{E}_A(\gamma),$$

where $V_A(\gamma)$ is defined in Definition A.3 of the Appendix, and in general, $V_A(\gamma) \neq Q_A(\gamma)$.

Under conditional homoskedasticity, Corollary A2 in the Appendix shows that, as for a LRF, the sup Wald and sup LR tests are not asymptotically equivalent for a TRF, except for the just identified case $p = q$.

As in Boldea et al. (2016), in this section, the asymptotic distributions are non-pivotal, and don't simplify to the usual break-point distributions expressed in Corollary 1. This is not an issue in practice, because critical values can still be obtained by bootstrap, as we discuss in Section 6.

5. GMM test

In contrast to our paper, Caner and Hansen (2004) propose testing for a threshold using a GMM sup Wald test. To calculate this test, they use the conventional two-step GMM estimators defined in Section 3, with estimated variance-covariances:

$$\hat{V}_{i,GMM}(\gamma) = \left(T^{-1} W_i^{\gamma\top} X_i^\gamma \hat{H}^{\epsilon^{-1}}_{i,GMM}(\gamma) X_i^{\gamma\top} W_i^\gamma\right)^{-1}.$$

The Wald test statistic in Caner and Hansen (2004) for $\mathbb{H}_0$ at each γ is:

$$W_T^{\text{GMM}}(\gamma) = T[\hat{\theta}^\gamma_{1,GMM} - \hat{\theta}^\gamma_{2,GMM}]^\top \{\hat{V}_{1,GMM}(\gamma) + \hat{V}_{2,GMM}(\gamma)\}^{-1} [\hat{\theta}^\gamma_{1,GMM} - \hat{\theta}^\gamma_{2,GMM}],$$

and the sup Wald test is $\sup_{\gamma\in\Gamma} W_T^{GMM}(\gamma)$.

For clarity, we reproduce below the asymptotic distribution of this test, which was already derived in Caner and Hansen (2004). Assume that $\mathbb{H}_0$ holds, and let $V_{i,GMM}(\gamma) = \left[N_i(\gamma) H_i^{\epsilon^{-1}}(\gamma) N_i^\top(\gamma)\right]^{-1}$, where $H_i^\epsilon(\gamma)$ is defined in Definition A.1 of the Appendix. Also, let $N_i(\gamma) = A_i^{0\top} M_i(\gamma)$, and let $\overline{\mathcal{GP}}_1(\gamma)$, be a $q \times 1$ zero mean Gaussian process with covariance kernel equal to $\mathbb{E}[\overline{\mathcal{GP}}_1(\gamma_1)\overline{\mathcal{GP}}_1^\top(\gamma_2)] = H_i^\epsilon(\gamma_1 \wedge \gamma_2)$. Let $\overline{\mathcal{GP}} = \overline{\mathcal{GP}}_1(\gamma_{max})$ and $\overline{\mathcal{GP}}_2(\gamma) = \overline{\mathcal{GP}} - \overline{\mathcal{GP}}_1(\gamma)$.[11] Then Caner and Hansen (2004) show:

[11] In Caner and Hansen (2004), $\overline{\mathcal{GP}} = \lim_{\gamma\to\infty} \overline{\mathcal{GP}}_1(\gamma)$, to account for an unbounded support Γ^0; as discussed before, for all practical purposes, including calculation of critical values, it makes sense to impose $\Gamma^0 = [\gamma_{min}, \gamma_{max}]$, treat $\gamma_{min}, \gamma_{max}$ as fixed values, and therefore define $\overline{\mathcal{GP}} = \overline{\mathcal{GP}}_1(\gamma_{max})$.

Theorem 4 (Asymptotic distribution sup Wald GMM). *Let Z be generated by* (2.1) *or* (2.2)*, and Y be generated by* (2.3)*. Under $\mathbb{H}_0$ and Assumptions A.1 and A.4,*

$$\sup_{\gamma\in\Gamma} W_T^{GMM}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma}\left[V_{1,GMM}(\gamma)N_1(\gamma)H_1^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_1(\gamma) - V_{2,GMM}(\gamma)N_2(\gamma)H_2^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_2(\gamma)\right]^\top \\ \times \left[V_{1,GMM}(\gamma)+V_{2,GMM}(\gamma)\right]^{-1} \\ \times \left[V_{1,GMM}(\gamma)N_1(\gamma)H_1^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_1(\gamma) - V_{2,GMM}(\gamma)N_2(\gamma)H_2^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_2(\gamma)\right].$$

The proof is in Caner and Hansen (2004). Theorems 2-4 show that the 2SLS and GMM tests have different asymptotic distributions in general, but there are two notable exceptions, both for a LRF. First, under conditional homoskedasticity and just identification, a comparison of Corollaries A1 and A3 in the Appendix shows that the GMM test distribution looks just like the 2SLS distributions for a LRF, with the difference that the Gaussian processes are generated by ϵ_t rather than $(\epsilon_t + u_t\theta_z^0)$. Second, under Assumptions A.1-A.3 and a LRF, all the distributions are the same, and identical to the break-point sup F and sup Wald test distributions. This latter result is stated below and proven in the Appendix.

Corollary 2 (Corollary to Theorem 4). *Let Z be generated by* (2.1) *and Y be generated by* (2.3)*. Then, under $\mathbb{H}_0$, and Assumptions A.1-A.3,*

$$\sup_{\gamma\in\Gamma} W_T^{GMM}(\gamma) \Rightarrow \sup_{\lambda\in\Lambda_\epsilon} \frac{\mathcal{BB}_p^\top(\lambda)\mathcal{BB}_p(\lambda)}{\lambda(1-\lambda)}$$

Note that for a TRF and the same assumptions, the distribution in Corollary 2 does not apply.

6. Simulations

In this chapter, we investigate the small sample properties of the 2SLS tests and the GMM test. We first introduce the wild fixed-regresssor bootstrap.

6.1. Bootstrap and DGP

Bootstrap As shown in Section 4, the asymptotic distributions of the proposed test statistics are non-standard and therefore need to be either simulated or bootstrapped.

Simulating the asymptotic distributions involves, for example, simulating the Gaussian processes $\mathcal{E}(\cdot)$ and $\mathcal{E}_A(\cdot)$ in Theorems 2-4, while keeping x_t, q_t fixed. On the other hand, in simulations, usually $Q(\gamma), V(\gamma), Q_A(\gamma), V_A(\gamma)$ are replaced with consistent estimators based on the initial sample, $\hat{Q}(\gamma), \hat{V}(\gamma), \hat{Q}_A(\gamma), \hat{V}_A(\gamma)$, and are kept fixed across simulations. Using similar arguments to Hansen (1996), Theorem 2, one can show that the critical value simulated in this way converges to the true critical value of the test.

However, the randomness of $\hat{Q}(\gamma), \hat{V}(\gamma), \hat{Q}_A(\gamma), \hat{V}_A(\gamma)$ may affect the critical value approximation in finite samples. Therefore, we propose bootstrapping the critical values instead.

Below, we describe the **fixed regressor wild bootstrap** we used for simulating critical values. We first describe it for the 2SLS test and then for the GMM test.

Bootstrap for 2SLS tests:

1. based on the original sample, compute the test statistics in Section 3, gathered under the generic name $\hat{G}$:

$$\hat{G} : \sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma),\ \sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma),\ \sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma),\ \sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma)$$

2. compute the full-sample 2SLS parameter estimates $\hat{\theta} = (\hat{\theta}_z^\top, \hat{\theta}_x^\top)^\top$ for a LRF or for a TRF, using (3.1) or (3.5), and the corresponding residuals for these estimates:

$$\hat{v}_t = (\hat{\epsilon}_t^\top, \hat{u}_t^\top)^\top$$

3. for each bootstrap sample j, draw a random sample $t = 1, \ldots, T$ from $\eta_t \sim iid\ \mathcal{N}(0,1)$, and compute the **wild bootstrap** residuals:

$$\hat{v}_t^{(j)} = \hat{v}_t \eta_t$$

4. keeping x_t, q_t **fixed**, calculate a new bootstrap sample $(y_t^{(j)}, z_t^{(j)})$

$z_t^{(j)} = \hat{\Pi}^\top x_t + \hat{u}_t^{(j)}$ for a LRF or $z_t^{(j)} = \hat{\Pi}_1^\top x_t \mathbb{1}_{\{q_t \le \hat{\rho}\}} + \hat{\Pi}_2^\top x_t \mathbb{1}_{\{q_t > \hat{\rho}\}} + \hat{u}_t^{(j)}$ for a TRF
$y_t^{(j)} = \hat{z}_t^{(j)\top} \hat{\theta}_z + x_{1t}^\top \hat{\theta}_x + \hat{u}_t^{(j)}$

5. using the new sample $(y_t^{(j)}, z_t^{(j)}, x_t, q_t)$ with **fixed regressors** x_t, q_t, recalculate all 2SLS test statistics, gathered under the generic name $\hat{G}^{(j)}$

$$\hat{G}^{(j)} : \sup_{\gamma\in\Gamma} LR^{2SLS,(j)}_{T,LRF}(\gamma),\ \sup_{\gamma\in\Gamma} LR^{2SLS,(j)}_{T,LRF}(\gamma),\ \sup_{\gamma\in\Gamma} W^{2SLS,(j)}_{T,LRF}(\gamma),\ \sup_{\gamma\in\Gamma} LR^{2SLS,(j)}_{T,LRF}(\gamma)$$

6. repeat this procedure for $j = 1, \ldots, J$ times

7. the 5% bootstrap critical value for each test statistic is equal to the 95% quantile from the empirical distribution $(\hat{G}^{(1)}, \ldots, \hat{G}^{(J)})$, call it $\hat{G}_{0.95}$

8. if $\hat{G} > \hat{G}_{0.95}$ we reject, else we don't reject.

Bootstrap for the GMM test:

1. based on the original sample, compute the GMM test statistic:

$$\hat{G} = \sup_{\gamma\in\Gamma} W_T^{GMM}(\gamma)$$

2. compute the full-sample two-step GMM parameter estimates $\hat{\theta}_{GMM}$ using the 2SLS estimator $\hat{\theta}$ for a LRF as the first-step GMM estimator; calculate the corresponding residuals:

$$\tilde{\epsilon}_t = y_t - w_t^\top \hat{\theta}_{GMM}$$

3. for each bootstrap sample j, draw a random sample $t = 1, \dots, T$ from $\eta_t \sim iid\ \mathcal{N}(0,1)$, and compute the **wild bootstrap** residuals:

$$\tilde{\epsilon}_t^{(j)} = \tilde{\epsilon}_t \eta_t$$

4. keeping z_t, x_t, q_t fixed, calculate a new bootstrap sample $y_t^{(j)}$

$$y_t^{(j)} = w_t^\top \hat{\theta}_{GMM} + \tilde{\epsilon}_t^{(j)}$$

5. using the new sample $(y_t^{(j)}, z_t, x_t, q_t)$ with **fixed regressors** z_t, x_t, q_t, recalculate the GMM test statistic $\hat{G}^{(j)}$

$$\hat{G}^{(j)} = \sup_{\gamma \in \Gamma} W_T^{GMM,(j)}$$

6. the 5% bootstrap critical value for each test statistic is equal to the 95% quantile from the empirical distribution $(\hat{G}^{(1)}, \dots, \hat{G}^{(J)})$, call it $\hat{G}_{0.95}$

7. if $\hat{G} > \hat{G}_{0.95}$ we reject, else we don't reject.

Our bootstrap is slightly different than the one suggested in Caner and Hansen (2004) for the same test statistic. They suggested setting $y_i^{(j)} = \tilde{\epsilon}_t \eta_t$, therefore computing a "pseudo-sample" that ignores the predictable part of y_t under $\mathbb{H}_0$, which is $(w_t^\top \theta^0)$. Presumably, they do so because the value of θ^0 is irrelevant for the asymptotic distribution of their test statistic. However, θ^0 shows up in the asymptotic distribution of our test statistics, and for the sake of comparison, we compute $y_t^{(j)}$ as suggested in Step 5. Computing $y_t^{(j)}$ as we suggested is a proper wild bootstrap. Compared to Caner and Hansen (2004), it should replicate more closely the sample null behavior of the test.

To calculate the empirical sizes $\hat{\alpha}$ for a nominal significance level α, we repeat the bootstrap procedure MC times, for a certain fixed $\mathbb{H}_0$ DGP but with the original sample redrawn in each simulation draw $s = 1, \dots, MC$, and set:

$$\hat{\alpha} = \frac{1}{MC} \sum_{s=1}^{MC} \mathbb{1}_{\hat{G}_s > \hat{G}_{0.95,s}}, \tag{6.1}$$

where the subscript s in $\hat{G}_s, \hat{G}_{0.95,s}$ refers to the s^{th} simulated value of $\hat{G}, \hat{G}_{0.95}$. The empirical power is obtained analogously with the DGP under $\mathbb{H}_A$:

$$\hat{\beta} = \frac{1}{MC} \sum_{s=1}^{MC} \mathbb{1}_{\hat{G}_s > \hat{G}_{0.95,s}}. \tag{6.2}$$

The size adjusted power is $\hat{\beta} - \hat{\alpha}$, where the $\mathbb{H}_0$ for calculating $\hat{\alpha}$ is chosen to mirror $\mathbb{H}_A$, as explained in Section 6.3.

DGP The $\mathbb{H}_0$ DGP used in the simulations for calculating empirical sizes is:

$$y_t = \theta^0_{x_1} + z_t\theta^0_z + \epsilon_t = w_t^\top \theta^0 + \epsilon_t \tag{6.3}$$

$$z_t = (\Pi^0_{1,1} + \Pi^0_{1,2}x_t)\mathbb{1}_{\{q_t \leq \rho^0\}} + (\Pi^0_{2,1} + \Pi^0_{2,2}x_t)\mathbb{1}_{\{q_t > \rho^0\}} + u_t \tag{6.4}$$

where $x_t \overset{iid}{\sim} \mathcal{N}(1, 1)$, $q_t = x_t + 1$, and x_t, z_t, q_t are scalars. We set:

- $\theta^0_z = \theta^0_{x_1} = 1$.
- $\Pi^0_1 = (\Pi^0_{1,1}, \Pi^0_{1,2})^\top = (1, 1)^\top$.
- $\Pi^0_2 = (\Pi^0_{2,1}, \Pi^0_{2,2})^\top = (1, b)^\top$, where we allow $b \in \{0.5, 1, 1.5, 2, 2.5\}$. Note that $b = 1$ corresponds to a LRF, and $b \neq 1$ to a TRF.
- $\rho^0 = 1.75$.

We consider two cases: homoskedasticity and heteroskedasticity. For homoskedasticity, $\epsilon_t = \nu_t$, and for conditional heteroskedasticity, $\epsilon_t = \nu_t \cdot x_t/\sqrt{2}$ with

$$\begin{pmatrix} \nu_t \\ u_t \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right). \tag{6.5}$$

We set $J = 500$ and $MC = 1000$.

6.2. Size

This section presents empirical sizes for all tests using the DGP in the previous section. In all simulations, the nature of the RF (LRF) or (TRF) is taken as given.

For conditional homoskedastic errors, Tables 9.1-9.2 show that in small samples, our tests tend to be slightly undersized, but stay below the nominal level, while the GMM test is correctly sized or very slightly oversized. Here, it seems that the additional RF information does not result in better small sample properties. In large samples of about $T = 1000$, all tests are close to their nominal size.

Tables 9.3-9.4 show results for a LRF and a TRF and conditional heteroskedasticity. In both LRF and TRF, the GMM test is severely oversized - up to a size of 21.8% for a nominal size of 5%. Moreover, its oversize seems to increase as the magnitude of the RF threshold increases. Even around sample sizes of $T = 1000$, its size is still above the nominal size. In practice, this means that the GMM test frequently detects a threshold when there is none. This is in stark contrast to both our tests, which stay below or very close to the nominal size, for both a LRF and a TRF, with homoskedasticity or heteroskedasticity.

The results under heteroskedasticity confirm the intuition that the 2SLS tests can be more accurate when they use additional valid information compared to the GMM sup Wald test. Note however that as in the DGP of Theorem 1, in our simulated DGP, there should still be areas of the parameter space where the GMM estimation is more accurate, so the GMM test should not be discarded, but used in conjunction with the 2SLS tests to identify a threshold.

6.3. Power

In this section, we present the size adjusted power of the three tests. We slightly alter the DGP in (6.3) while leaving everything else equal. In particular we set

$$y_t = w_t^\top \theta_1^0 \mathbb{1}_{\{q_t \le \gamma^0\}} + w_t^\top \theta_2^0 \mathbb{1}_{\{q_t > \gamma^0\}} + \epsilon_t \tag{6.6}$$

with $\theta_1^0 = (1,\, 1)^\top$ as before and $\theta_2^0 = (a, c)^\top$ with $a \in \{1, 2\}$, $c \in \{1.25, 1.5, 1.75, 2\}$, and $\delta = c - 1$ the slope threshold size. This allows us to investigate how the power varies with the threshold size, measured by $a - 1$ and δ. Finally, we set $\gamma^0 = 2.25$.
We follow Davidson and MacKinnon (1998, Section 6) and plot size-power curves. That is, we plot all possible sizes between 0 and 1 on the x-axis. The sizes used for size-adjusted powers are true empirical sizes in the sense that they are computed based on (simulated) empirical critical values and the empirical distribution function of the test statistics[12]. On the y-axis we plot the size adjusted power which is calculated using the empirical critical values.

Figures 9.1–9.4 show the result of this exercise for a LRF, a TRF, with homoskedastic and heteroskedastic errors, and a slope threshold ($\delta = c - 1 \neq 0$) or an intercept and slope threshold ($a - 1 \neq 0$ and $\delta \neq 0$). In general, holding the sample size fixed, as the threshold sizes $|a - 1|$ or $|\delta|$ increase, the powers of all three tests increase. An increase in the sample size, given a fixed threshold size, increases the power of the tests as well. Furthermore, for moderate threshold sizes the tests have very similar power.
However, there are is also an interesting difference for small threshold size and/or small sample size: across most cases, the sup Wald tests outperform the sup LR test. Of course, this difference vanishes as the sample size and/or the threshold size increases. Moreover, it seems that the GMM sup Wald test has better power than the 2SLS sup Wald but only in the case of small samples and small threshold values.

Even though our simulations indicate that the sup Wald tests are better than the 2SLS sup LR test in terms of power, we know from Caner and Hansen (2004) that under the alternative, the γ at which the supremum is obtained for the sup LR test is a consistent threshold estimator whether we have an LRF or a TRF, so it is useful to compute the 2SLS sup LR test anyway.

7. Empirical Application

In this section, we test whether the government spending multiplier - measured as the percentage increase in output when government spending increases by 1% - changes in the presence of different interest rate regimes. For example, the multiplier is expected to be larger in the recent crisis if the transmission mechanism is largely demand-driven - see e.g. Eggertsson (2010) and Christiano et al. (2011). When the nominal interest rates are

[12] The empirical critical values are computed under the DGP of Section 6.2. Of course, other $\mathbb{H}_0$-DGPs are possible (e.g. averaging over θ_1^0 and θ_2^0) but it seems natural to take that of Section 6.2 for easy comparison.

close to the zero lower bound (ZLB) or in general below a certain threshold, government spending should be more effective in increasing growth, since higher consumption and investment are facilitated by a low real interest rate (potentially through higher inflation). On the other hand, if in the present crisis, the transmission mechanism is driven by supply, and despite the low nominal interest rate, government spending crowds out private investment, the multiplier is small. We use the following specification, in line with Hall (2009) and Kraay (2012), but allowing for a threshold:

$$\frac{y_t - y_{t-1}}{y_{t-1}} = \left(\alpha_1 + \beta_1 \frac{g_t - g_{t-1}}{y_{t-1}}\right) \mathbb{1}_{\{r_{t-1} \leq \gamma^0\}} + \left(\alpha_2 + \beta_2 \frac{g_t - g_{t-1}}{y_{t-1}}\right) \mathbb{1}_{\{r_{t-1} > \gamma^0\}} + \epsilon_t, \quad (7.1)$$

where y_t and g_t denote real GDP and government spending per capita, respectively, α_1, α_2 are constants, β_1, β_2 are the multipliers in the two regimes, and ϵ_t is an error term that satisfies Assumption A.1; r_t denotes the Federal Funds Rate and γ^0 is the unknown potential threshold-parameter.

We are interested in testing whether the multipliers in (7.1) are indeed different in different interest rate regimes, that is, whether we have a interest-rate driven threshold γ^0. Since $z_t = \frac{g_t - g_{t-1}}{y_{t-1}}$ is endogenous as output shocks can influence spending in the same quarter, we instrument it as in Ramey (2011), with one quarter-ahead government spending forecast errors, SPF_t, from the Survey of Professional Forecasters.[13] Thus, we specify the RF (with a potential threshold at ρ^0) as:

$$\frac{g_t - g_{t-1}}{y_{t-1}} = (\Pi_{1,1} + \Pi_{1,2} SPF_t)\mathbb{1}_{\{r_{t-1} \leq \rho^0\}} + (\Pi_{2,1} + \Pi_{2,2} SPF_t)\mathbb{1}_{\{r_{t-1} > \rho^0\}} + u_t. \quad (7.2)$$

We use quarterly US data spanning 1969Q1-2014Q4, with the real GDP and government spending from the Bureau of Economic Analysis[14], the federal funds rate from the Fed St. Louis[15] and the government spending forecasts from the Philadelphia Fed.[16] The data includes the current ZLB regime, as can be seen from the federal funds rate plot in Figure 9.5.

Since our sample includes the Volcker period, part of which is characterized by unusually high interest rates and volatile economic conditions, we consider three samples: 1969Q2-2014Q4, 1969Q2–1984Q4 and 1985Q1–2014Q4. Since low interest rates are mostly, but not exclusively near the end of our sample, we consider two cut-off points for testing for a threshold in (7.1): the 15% and the 5% quantiles of the empirical distribution of r_{t-1}.

We first test whether the RF is a threshold model (TRF) or a linear model (LRF) by the methods proposed in Hansen (1996). Based on the results, we estimate the LRF or TRF and test for a threshold in (7.1) using the 2SLS and GMM tests. Tables

[13]See Ramey (2011) for more discussion on instrument validity of SPF_t, and a description of how the forecast errors were calculated.
[14]Accessed February 2015.
[15]Accessed February 2015.
[16]Accessed February 2015.

9.5-9.7 present results for all the three samples considered. In these tables, we report bootstrapped p-values of the test statistics instead of bootstrapped critical values. Bootstrapped p-values are simply obtained by counting how many times in J bootstrap samples the original test statistic is larger than its bootstrapped equivalent.

Concerning the RF, regardless of the cut-off, or whether we use the full-sample or the post 1985 sample, we find that the RF has a threshold at ρ^0 below 7.

For the SF in (7.1), estimated on the whole sample, we find no clear evidence for a threshold effect based on 2SLS tests: none of the 2SLS threshold tests rejects at the 1% level. But the p-value of the GMM test is below the 1% level for a 5% cut-off. So relying on the GMM test alone, one would tend to include a threshold. Since the p-values of the 2SLS sup LR test are above the 5% level, the applied researcher can be more confident about the absence of a SF threshold.

Furthermore, the threshold estimator reflects a very high interest rate regime, sensitive to the cut-off choice of 5% or 15%, and not close to the ZLB. This prompts us to investigate the samples 1968Q1-1984Q4 (unusually high interest rates) and 1985Q1-2014Q4 (not so high interest rates) separately.[17]

Table 9.6 shows that all three threshold tests do not reject the null of no threshold regime for the period of 1985 onwards. That is, we find no evidence that a ZLB or any other interest rate regime in our sample changes the government spending multiplier or the effectiveness of the government spending on output growth.

We find that the government spending multiplier 2SLS and GMM estimators are close to each other, significant, and around 0.12. Thus, an increase in government spending of 1% of real GDP will increase growth by 0.12%. Our estimates are small and in line with Hall (2009) (who used a sample period from 1960–2008) . They are much smaller than in Nakamura and Steinsson (2014), who find an (open economy) multiplier of about 1.5. Eggertsson (2010) and Christiano et al. (2011) argue that in the neighborhood of the ZLB, when monetary policy is less effective, fiscal stimulus lowers real interest rates by raising inflation, resulting in potentially large multipliers. However, in the recent crisis, the US inflation has remained low and stable, which may explain why we don't find a larger multiplier near the ZLB.

8. Conclusion

In this paper, we propose two novel threshold tests for linear models with endogenous regressors, a sup LR and a sup Wald test. These tests are based on 2SLS estimation and explicitly account for a possible threshold effect in the RF. We derive the asymptotic distributions of our tests, which are non-pivotal but whose critical values or p-values can easily be bootstrapped. Our simulation study shows that the tests behaves well in small samples, and their size and power compare favorably to an existing GMM based sup Wald test. We therefore recommend using them in conjunction with the GMM test for threshold testing.

[17] The Volcker period results in Table 9.5 are presented for completeness, but the sample size is small, and care should be used in interpreting those results.

We apply our method to assess whether the US government spending multiplier is larger near the zero lower bound. All tests, but more conclusively the 2SLS tests, suggest that the US government spending multiplier for output growth did not change near the zero lower bound or any other interest rate regime.

References

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856.

Antoine, B. and O. Boldea (2015a): "Efficient Inference with Time-Varying Information and the New Keynesian Phillips Curve," SFU Working Paper 15-04.

——— (2015b): "Inference in Linear Models with Structural Changes and Mixed Identification Strength," SFU Working Paper 15-5.

Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

Boldea, O., A. Cornea, and A. Hall (2016): "Bootstrapping Structural Change Tests," Working Paper.

Caner, M. and B. Hansen (2001): "Threshold Autoregression with a Unit Root," *Econometrica*, 69, 1555–1596.

Caner, M. and B. E. Hansen (2004): "Instrumental Variable Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843.

Chan, K. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *Annals of Statistics*, 21, 520–533.

Christiano, L., M. Eichenbaum, and S. Rebelo (2011): "When is the Government Spending Multiplier Large?" *Journal of Political Economy*, 119, 78–121.

Davidson, R. and J. G. MacKinnon (1998): "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *The Manchester School*, 66, 1–26.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247–254.

Eggertsson, G. (2010): "What Fiscal Policy Is Effective at Zero Interest Rates?" in *NBER Macroeconomic Annual*, ed. by D. Acemogly and M. Woodford, University of Chicago Press.

Gonzalo, J. and J.-Y. Pitarakis (2002): "Estimation and Model Selection Based Inference in Single and Multiple Threshold Models," *Journal of Econometrics*, 110, 319–352.

——— (2006): "Threshold Effects in Cointegrating Regressions," *Oxford Bulletin of Economics and Statistics*, 68, 813–833.

Gonzalo, J. and M. Wolf (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*, 127, 201–224.

Hall, A. R., S. Han, and O. Boldea (2012): "Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors," *Journal of Econometrics*, 170, 281–302.

Hall, R. E. (2009): "By how much Does GDP Rise if the Government Buys More Output?" *Brookings Papers on Economic Activity*, 2(2009), 183–231.

Hansen, B. (2016): "Regression Kink with an Unknown Threshold," *Journal of Business and Economic Statistics*, forthcoming.

Hansen, B. E. (1996): "Inference when a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

——— (1999): "Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference," *Journal of Econometrics*, 93, 345–368.

——— (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.

——— (2011): "Threshold Autoregression in Economics," *Statistics and Its Interface*, 4, 123–127.

Kourtellos, A., T. Stengos, and C. Tan (2015): "Structural Threshold Regression," *Econometric Theory*, 1–34.

Kraay, A. (2012): "How Large is the Government Spending Multiplier? Ecvidence from World Bank Lending," *The Quarterly Journal of Economics*, 127, 1–59.

Lee, S., H. Park, M. Seo, and Y. Shin (2014): "A Contribution to the Reinhart and Rogoff Debate: not 90% but maybe 30%," *CEMMAP Working Paper CWP 39-14*.

Lee, S., M. H. Seo, and Y. Shin (2011): "Testing for Threshold Effects in Regression Models," *Journal of the American Statistical Association*, 106, 220–231.

Magnusson, L. and S. Mavroeidis (2014): "Identification Using Stability Restrictions," *Econometrica*, 82, 1799–1851.

Nakamura, E. and J. Steinsson (2014): "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions," *American Economic Review*, 104, 753–792.

Ramey, V. A. (2011): "Identifying Government Spending Shocks: It's All in the Timing," *Quarterly Journal of Economics*, 1–50.

Reinhart, C. M. and K. S. Rogoff (2010): "Growth in a Time of Debt," *American Economic Review: Papers and Proceedings*, 100, 573–578.

Seo, M. and O. Linton (2007): "A Smoothed Least Squares Estimator for Threshold Regression Models," *Journal of Econometrics*, 141, 704–735.

Tong, H. (1990): *Nonlinear Time Series - A Dynamical System Approach*, Oxford: Clarendon Press.

Yu, P. and P. Phillips (2014): "Threshold Regression with Endogeneity," Cowles Foundation Working Paper 1966.

9. Tables and Figures

Table 9.1: Empirical sizes for 5% nominal size, a LRF and homoskedastic errors

T	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$
100	0.047	0.030	0.052
250	0.048	0.037	0.040
500	0.057	0.053	0.047
1000	0.048	0.048	0.046

Table 9.2: Empirical sizes for 5% nominal size, a TRF and homoskedastic errors

	b=0.5			b=1.5		
T	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$
100	0.023	0.028	0.047	0.017	0.022	0.052
250	0.029	0.030	0.051	0.024	0.022	0.040
500	0.027	0.026	0.046	0.038	0.031	0.045
1000	0.042	0.040	0.042	0.033	0.037	0.043
	b=2.0			b=2.5		
T	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$
100	0.026	0.024	0.052	0.032	0.016	0.055
250	0.039	0.031	0.049	0.049	0.036	0.051
500	0.052	0.041	0.046	0.053	0.040	0.047
1000	0.045	0.049	0.047	0.048	0.050	0.043

Table 9.3: Empirical sizes for 5% nominal size, a LRF and heteroskedastic errors

T	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$
100	0.046	0.038	0.159
250	0.048	0.037	0.104
500	0.057	0.052	0.092
1000	0.048	0.045	0.080

Table 9.4: Empirical sizes for 5% nominal size, a TRF and heteroskedastic errors

	b=0.5			b=1.5		
T	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$
100	0.034	0.030	0.110	0.037	0.018	0.188
250	0.050	0.032	0.080	0.043	0.022	0.127
500	0.053	0.032	0.077	0.046	0.036	0.090
1000	0.061	0.049	0.070	0.052	0.036	0.080
	b=2.0			b=2.5		
T	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$	$LR_T^{2SLS}(\gamma)$	$W_T^{2SLS}(\gamma)$	$W_T^{GMM}(\gamma)$
100	0.050	0.021	0.203	0.046	0.024	0.218
250	0.055	0.032	0.130	0.056	0.026	0.135
500	0.053	0.043	0.091	0.060	0.045	0.096
1000	0.056	0.042	0.078	0.058	0.039	0.075

We also tried $J = 1000$ instead of $J = 500$, and set $T = 100$, keeping everything else the same. The size of the GMM test improves (e.g. for $b = 0.5$ empirical size is roughly 8%) but the 2SLS tests improve as well.

Table 9.5: Estimation Results Full Sample

	Sample 1969Q2–2014Q4					
	Cut-Off 15% $\underline{\gamma} = 1.95$ and $\overline{\gamma} = 9.46$			Cut-Off 5% $\underline{\gamma} = 0.15$ and $\overline{\gamma} = 13.58$		
	RF Results					
p-value W^{OLS}	0.0000			0.0000		
$\hat{\Pi}_{1,1}$	0.0029			0.0029		
$\hat{\Pi}_{1,2}$	0.4226			0.4226		
$\hat{\Pi}_{2,1}$	0.0068			0.0068		
$\hat{\Pi}_{2,2}$	0.5253			0.5253		
$\hat{\rho}$	6.7000			6.7000		
No. of obs.						
total	183	183	183	183	183	183
$r_{t-1} \leq \hat{\rho}$	124	124	124	124	124	124
$r_{t-1} > \hat{\rho}$	59	59	59	59	59	59
	SF Results					
Tests	LR^{2SLS}	W^{2SLS}	W^{GMM}	LR^{2SLS}	W^{2SLS}	W^{GMM}
Statistic	14.9129	9.8724	10.4121	21.6958	9.8724	10.4121
p-value	0.0650	0.0310	0.0240	0.0920	0.0430	0.0050
$\hat{\beta}_1$	0.0522	0.0522	0.0853	0.0522	0.0522	0.0853
	(0.0806)	(0.0806)	(0.0060)	(0.0806)	(0.0806)	(0.0060)
$\hat{\alpha}_1$	0.0066	0.0066	0.0065	0.0066	0.0066	0.0065
	(0.0006)	(0.0006)	(0.00004)	(0.0006)	(0.0006)	(0.00004)
$\hat{\beta}_2$	–	–	–	–	–	–
	–	–	–	–	–	–
$\hat{\alpha}_2$	–	–	–	–	–	–
	–	–	–	–	–	–
$\hat{\gamma}$	–	–	–	–	–	–
95%-CI for $\hat{\gamma}$	–	–	–	–	–	–
No. of obs.						
total	183	183	183	183	183	183
$r_{t-1} \leq \hat{\gamma}$	–	–	–	–	–	–
$r_{t-1} > \hat{\gamma}$	–	–	–	–	–	–

1 LR stands for the sup LR test and W for the sup Wald test for a threshold. Their superscripts indicates the estimation method used (OLS, 2SLS, or GMM)
2 Standard errors in parentheses.
3 $\Gamma = [\underline{\gamma}, \overline{\gamma}]$.

Table 9.6: Estimation Results Subsample 1985Q1–2014Q4

	Sample 1985Q1–2014Q4					
	Cut-Off 15% $\underline{\gamma} = 1.02$ and $\overline{\gamma} = 7.74$			Cut-Off 5% $\underline{\gamma} = 0.12$ and $\overline{\gamma} = 8.48$		
	RF Results					
p-value W^{OLS}	0.0000			0.0000		
$\hat{\Pi}_{1,1}$	0.0027			0.0027		
$\hat{\Pi}_{1,2}$	0.4285			0.4285		
$\hat{\Pi}_{2,1}$	0.0078			0.0078		
$\hat{\Pi}_{2,2}$	0.5713			0.5713		
$\hat{\rho}$	6.4700			6.4700		
No. of obs.						
total	120	120	120	120	120	120
$r_{t-1} \leq \hat{\rho}$	97	97	97	97	97	97
$r_{t-1} > \hat{\rho}$	23	23	23	23	23	23
	SF Results					
Tests	LR^{2SLS}	W^{2SLS}	W^{GMM}	LR^{2SLS}	W^{2SLS}	W^{GMM}
Statistic	4.7812	3.4531	3.1544	4.7182	3.4531	3.1544
p-value	0.4960	0.7250	0.7650	0.5460	0.8950	0.7080
$\hat{\beta}_1$	0.1146	0.1146	0.1242	0.1146	0.1146	0.1242
	(0.0672)	(0.0672)	(0.0069)	(0.0672)	(0.0672)	(0.0069)
$\hat{\alpha}_1$	0.0061	0.0061	0.0061	0.0061	0.0061	0.0061
	(0.0005)	(0.0005)	(0.00005)	(0.0005)	(0.0005)	(0.00005)
$\hat{\beta}_2$	–	–	–	–	–	–
	–	–	–	–	–	–
$\hat{\alpha}_2$	–	–	–	–	–	–
	–	–	–	–	–	–
$\hat{\gamma}$	–	–	–	–	–	–
95%-CI for $\hat{\gamma}$	–	–	–	–	–	–
No. of obs.						
total	120	120	120	120	120	120
$r_{t-1} \leq \hat{\gamma}$	–	–	–	–	–	–
$r_{t-1} > \hat{\gamma}$	–	–	–	–	–	–

[1] LR stands for the sup LR test and W for the sup Wald test for a threshold. Their superscripts indicates the method used (OLS, 2SLS, or GMM)

[2] Standard errors in parentheses.

[3] $\Gamma = [\underline{\gamma}, \overline{\gamma}]$.

Table 9.7: Estimation Results Subsample 1969Q2–1984Q4

	Sample 1969Q2–1984Q4					
	Cut-Off 15% $\underline{\gamma} = 4.87$ and $\overline{\gamma} = 12.69$			Cut-Off 5% $\underline{\gamma} = 4.30$ and $\overline{\gamma} = 15.85$		
	RF Results					
p-value W^{OLS}		0.8800			0.9060	
$\hat{\Pi}_{1,1}$		0.0052			0.0052	
$\hat{\Pi}_{1,2}$		0.4739			0.4739	
$\hat{\Pi}_{2,1}$		–			–	
$\hat{\Pi}_{2,2}$		–			–	
$\hat{\rho}$		–			–	
No. of obs.						
total	63	63	63	63	63	63
$r_{t-1} \leq \hat{\rho}$	–	–	–	–	–	–
$r_{t-1} > \hat{\rho}$	–	–	–	–	–	–
	SF Results					
Test	LR^{2SLS}	W^{2SLS}	W^{GMM}	LR^{2SLS}	W^{2SLS}	W^{GMM}
Statistic	18.3092	13.2520	22.1056	18.3092	13.2520	22.1056
p-value	0.0200	0.0030	0.0000	0.0380	0.0000	0.0000
$\hat{\beta}_1$	0.3850	0.3850	0.3874	0.3850	0.3850	0.3874
	(0.1873)	(0.1873)	(0.0233)	(0.1873)	(0.1873)	(0.0233)
$\hat{\alpha}_1$	0.0109	0.0109	0.0110	0.0109	0.0109	0.0110
	(0.0016)	(0.0016)	(0.0002)	(0.0016)	(0.0016)	(0.0002)
$\hat{\beta}_2$	-0.3049	-0.3049	-0.3061	-0.3049	-0.3049	-0.3061
	(0.2681)	(0.2681)	(0.0343)	(0.2681)	(0.2681)	(0.0343)
$\hat{\alpha}_2$	0.0040	0.0040	0.0041	0.0040	0.0040	0.0041
	(0.0027)	(0.0027)	(0.0004)	(0.0027)	(0.0027)	(0.0004)
$\hat{\gamma}$	8.8000	8.8000	8.800	8.8000	8.8000	8.8000
95%-CI for $\hat{\gamma}$		[5.5700; 11.3900]			[5.5700; 11.3900]	
No. of obs.						
total	63	63	63	63	63	63
$r_{t-1} \leq \hat{\gamma}$	32	32	32	32	32	32
$r_{t-1} > \hat{\gamma}$	31	31	31	31	31	31

[1] LR stands for the sup LR test and W for the sup Wald test for a threshold. Their superscripts indicates the method used (OLS, 2SLS, or GMM)
[2] Standard errors in parentheses.
[3] $\Gamma = [\underline{\gamma}, \overline{\gamma}]$.

Figure 9.1: Size-adjusted power curves - homoskedastic errors and no change in intercept

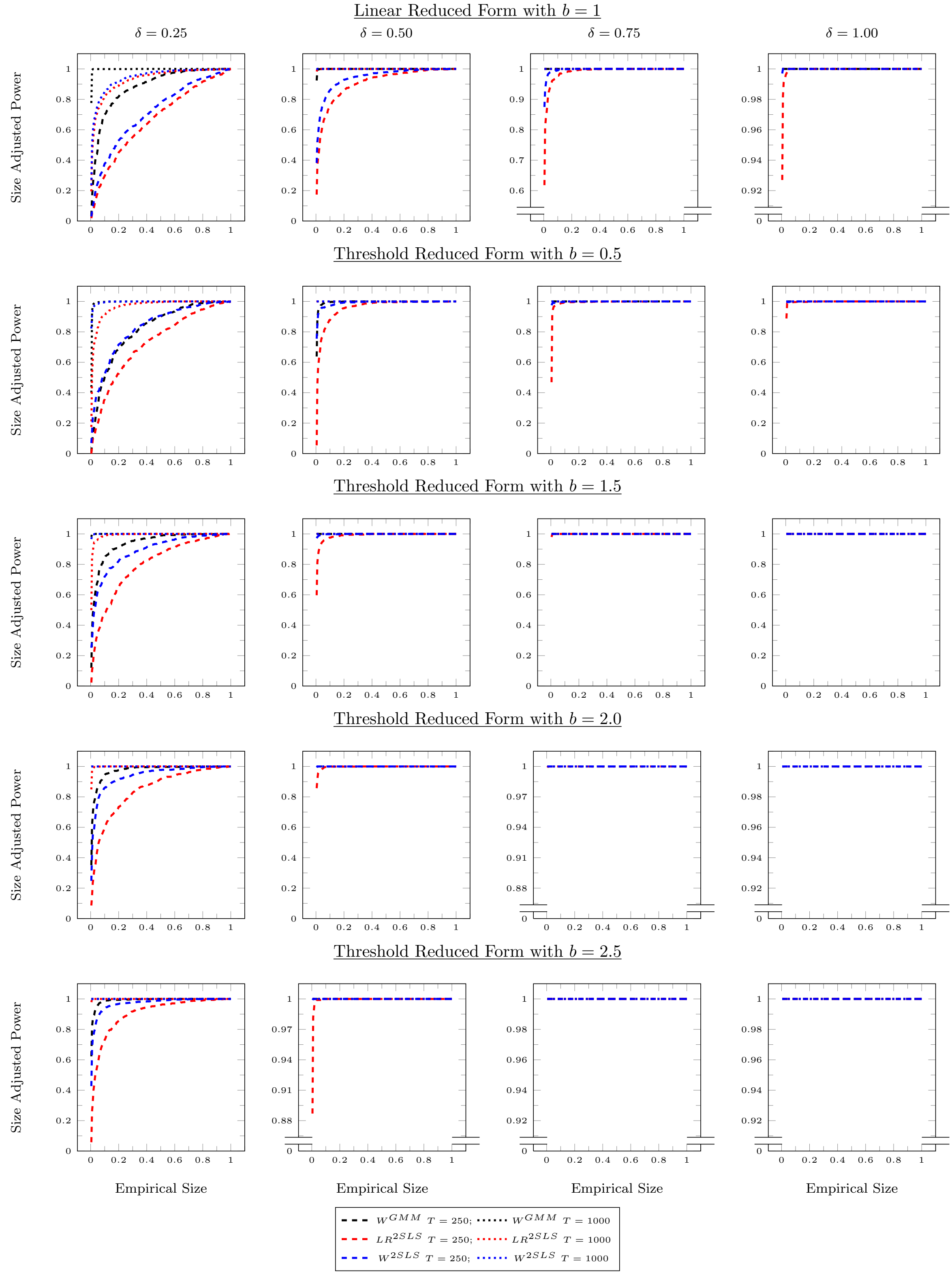


Figure 9.2: Size-adjusted power curves - homoskedastic errors and change in intercept

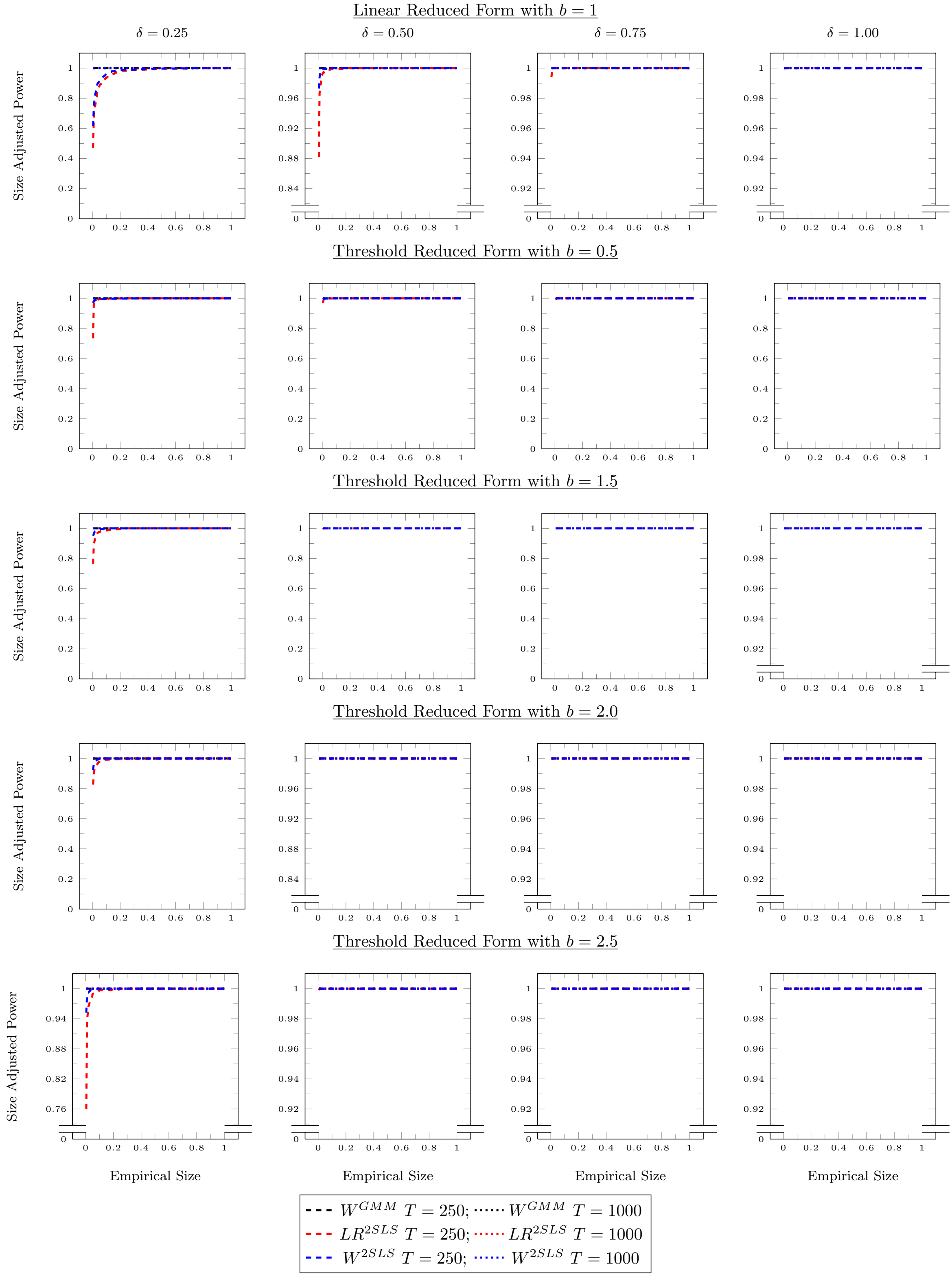


Figure 9.3: Size-adjusted power curves - heteroskedastic errors and no change in intercept

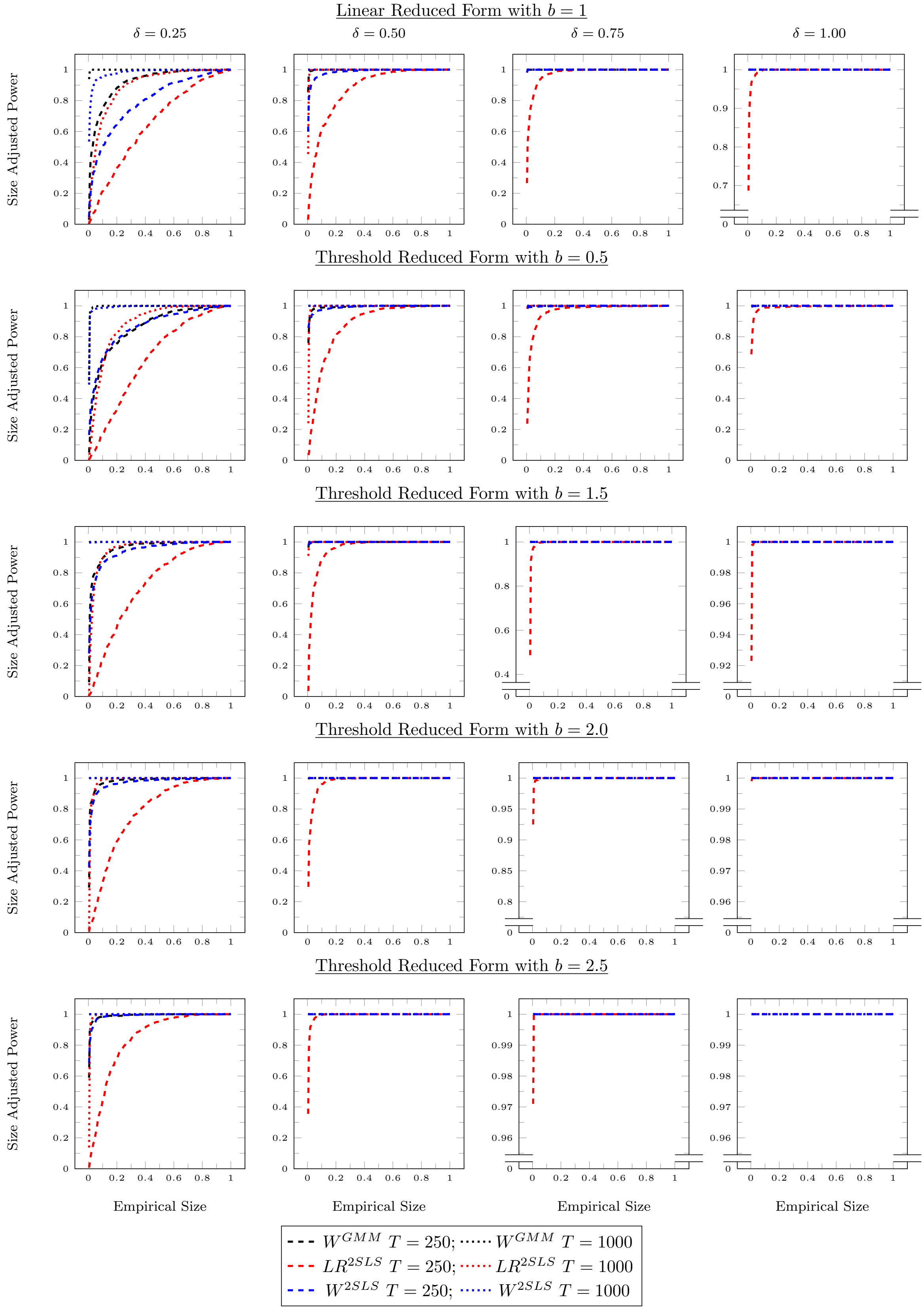


Figure 9.4: Size-adjusted power curves - heteroskedastic errors and change in intercept

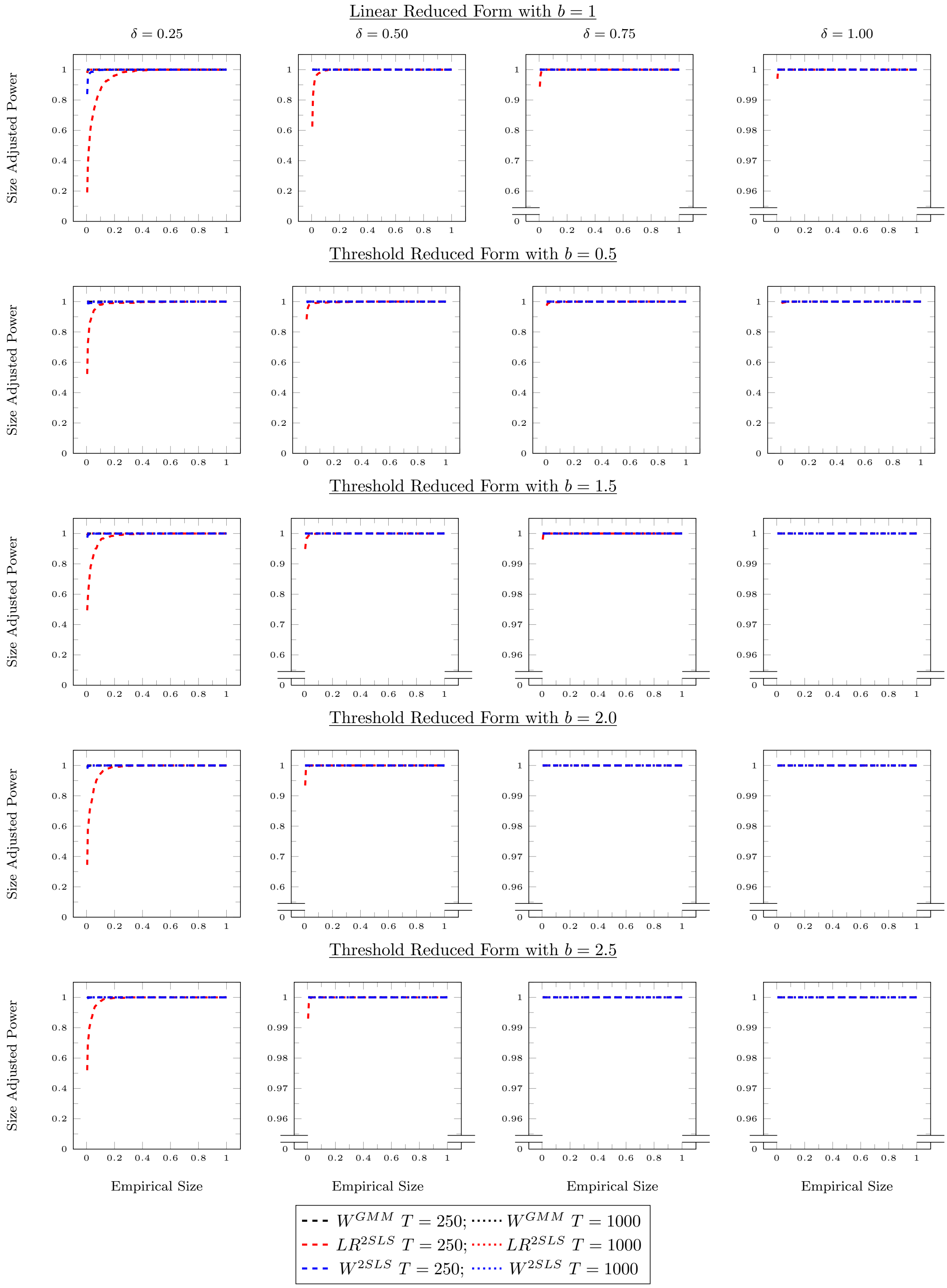
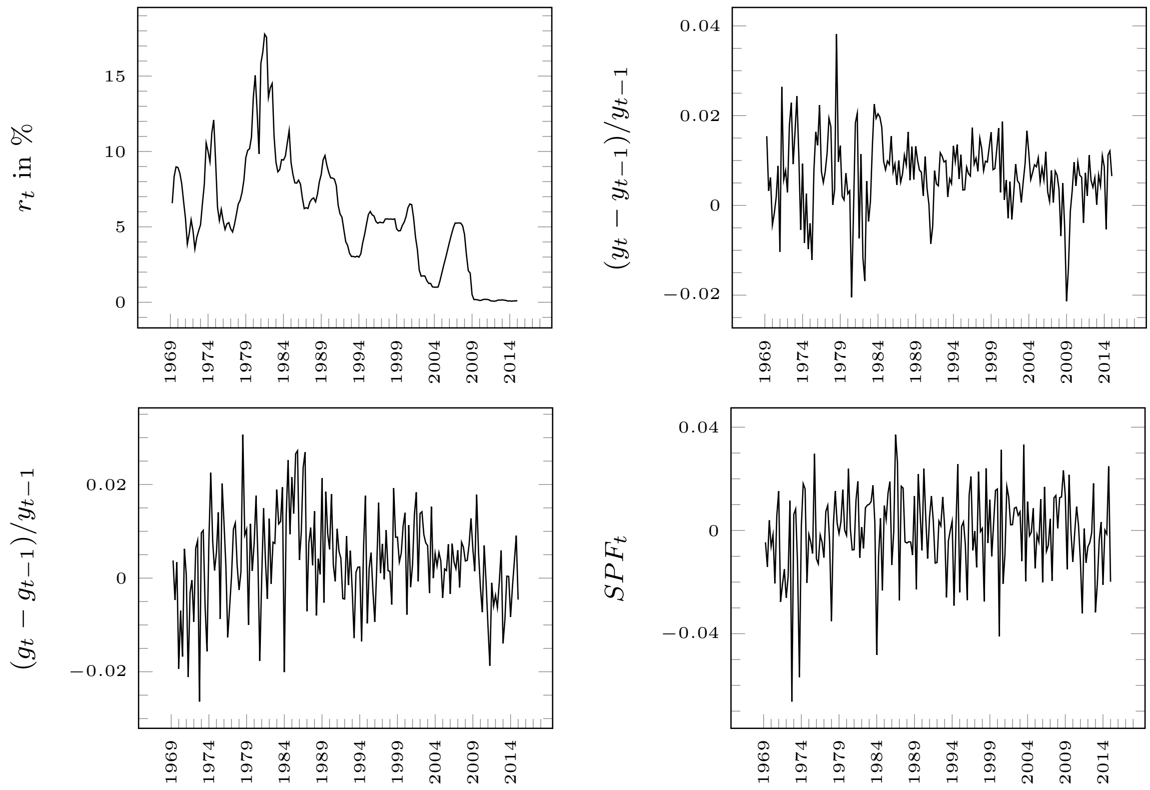


Figure 9.5: Data empirical application – 1969Q2–2014Q4



Mathematical Appendix

A. Definitions

Definition A.1 (*H* **and** $\hat{H}$ **matrices).**

$$
\begin{aligned}
H_1^u(\gamma) &= \mathbb{E}[x_t x_t^\top (u_t^\top \theta_z^0)^2 \mathbb{1}_{\{q_t \le \gamma\}}] & H_2^u(\gamma) &= \mathbb{E}[x_t x_t^\top (u_t^\top \theta_z^0)^2 \mathbb{1}_{\{q_t > \gamma\}}] \\
H_1^\epsilon(\gamma) &= \mathbb{E}[x_t x_t^\top \epsilon_t^2 \mathbb{1}_{\{q_t \le \gamma\}}] & H_2^\epsilon(\gamma) &= \mathbb{E}[x_t x_t^\top \epsilon_t^2 \mathbb{1}_{\{q_t > \gamma\}}] \\
H_1^{\epsilon,u}(\gamma) &= \mathbb{E}[x_t x_t^\top \epsilon_t u_t^\top \theta_z^0 \mathbb{1}_{\{q_t \le \gamma\}}] & H_2^{\epsilon,u}(\gamma) &= \mathbb{E}[x_t x_t^\top \epsilon_t u_t^\top \theta_z^0 \mathbb{1}_{\{q_t > \gamma\}}] \\
H_1(\gamma) &= H_1^u(\gamma) + 2H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma) & H_2(\gamma) &= H_2^u(\gamma) + 2H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma).
\end{aligned}
$$

Also, let $H = H_1(\gamma_{\max}) = \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2]$ *and* $H^u = H_1^u(\gamma_{\max}) = \mathbb{E}[x_t x_t^\top (u_t^\top \theta_z^0)^2]$.

Their estimators are constructed under $\mathbb{H}_0$*. Let* $\hat{z}_t$ *and therefore* $\hat{w}_t = (\hat{z}_t^\top, x_{1t}^\top)^\top$ *be calculated by* (3.1) *for a LRF and by* (3.5) *for a TRF. Let* $\hat{u}_t = z_t - \hat{z}_t$ *and* $\hat{\epsilon}_t = y_t - w_t'\hat{\theta}$*, where* $\hat{\theta} = (\hat{W}^\top \hat{W})^{-1} \hat{W}^\top Y$*, the full sample 2SLS estimator, partitioned as* $\hat{\theta} = (\hat{\theta}_z^\top, \hat{\theta}_x^\top)^\top$*. The sample analogs of all H matrices above are denoted with a hat accent* $\hat{H}$*, and replace* $\mathbb{E}$ *with* $T^{-1}\sum_{t=1}^T$*, and* $\epsilon_t, u_t, \theta_z^0$ *with* $\hat{\epsilon}_t, \hat{u}_t, \hat{\theta}_z$*; for example,* $\hat{H}_1^\epsilon(\gamma) = T^{-1}\sum_{t=1}^T x_t x_t^\top \hat{\epsilon}_t^2 \mathbb{1}_{\{q_t \le \gamma\}}$.

Definition A.2 ($V(\gamma)$ **and** $\hat{V}(\gamma)$**).** *We have a LRF as in* (2.1)*. Then:*

$$
\begin{aligned}
V(\gamma) &= V_1(\gamma) + V_2(\gamma) - V_{12}(\gamma) - V_{12}^\top(\gamma) \\
V_i(\gamma) &= C_i^{-1}(\gamma) A^0 \Big[H_i(\gamma) + R_i(\gamma) H^u R_i^\top(\gamma) - [H_i^{\epsilon,u}(\gamma) + H_i^u(\gamma)] R_i^\top(\gamma) \\
&\qquad - R_i(\gamma)[H_i^{\epsilon,u}(\gamma) + H_i^u(\gamma)] \Big] A^{0\top} C_i^{-1}(\gamma), \quad i = 1,2 \\
V_{12}(\gamma) &= -C_1^{-1}(\gamma) A^0 \Big[[H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)] R_2^\top(\gamma) + R_1(\gamma)[H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)] \\
&\qquad - R_1(\gamma) H^u R_2^\top(\gamma) \Big] A^{0\top} C_2^{-1}(\gamma).
\end{aligned}
$$

$V_A(\gamma)$ *is constructed by replacing all quantities in the definition of* $V_A(\gamma)$ *by their sample analogs, denoted with a hat accent. For example,* $\hat{V}_i(\gamma) = \hat{C}_i^{-1}(\gamma)\hat{A}\Big[\hat{H}_i(\gamma) + \hat{R}_i(\gamma)\hat{H}^u \hat{R}_i^\top(\gamma) - [\hat{H}_i^{\epsilon,u}(\gamma) + \hat{H}_i^u(\gamma)]\hat{R}_i^\top(\gamma) - \hat{R}_i(\gamma)[\hat{H}_i^{\epsilon,u}(\gamma) + \hat{H}_i^u(\gamma)]\Big]\hat{A}^\top \hat{C}_i^{-1}(\gamma)$*, with* $\hat{A} = [\hat{\Pi}, S^\top]^\top$, $\hat{C}_i(\gamma) = \hat{A}\hat{M}_i(\gamma)\hat{A}^\top$, $\hat{M}_1(\gamma) = T^{-1}\sum_{t=1}^T x_t x_t^\top \mathbb{1}_{\{q_t \le \gamma\}}$,
$\hat{M}_2(\gamma) = T^{-1}\sum_{t=1}^T x_t x_t^\top \mathbb{1}_{\{q_t > \gamma\}}$, $\hat{M} = \hat{M}_1(\gamma_{max})$, $\hat{R}_i(\gamma) = \hat{M}_i(\gamma)\hat{M}^{-1}$.

Definition A.3 ($V_A(\gamma)$ **and** $\hat{V}_A(\gamma)$)**.** *We have a TRF as in (2.2). Then:*

$$
\begin{aligned}
V_A(\gamma) &= V_{A,1}(\gamma) + V_{A,2}(\gamma) - V_{A,12}(\gamma) - V_{A,12}^\top(\gamma) \\
V_{A,1}(\gamma) &= C_{A,1}^{-1}(\gamma)A_1^0\Big[H_1(\gamma) + R_1(\gamma;\rho^0)H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) - [H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)]R_1^\top(\gamma;\rho^0) \\
&\qquad - R_1(\gamma;\rho^0)[H^{\epsilon,u}(\gamma) + H_1^u(\gamma)]\Big]A_1^{0\top}C_{A,1}^{-1}(\gamma) \\
V_{A,2}(\gamma) &= C_{A,2}^{-1}(\gamma)\Big[A_2^0\ H_2^\epsilon(\rho^0)\ A_2^0 + A_1^0\ [H_1^\epsilon(\rho^0) - H_1^\epsilon(\gamma) + H_1^u(\gamma) + R_1(\gamma;\rho^0)H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) \\
&\quad + R_1(\gamma;\rho^0)[H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)] \\
&\quad + [H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)]R_1^\top(\gamma;\rho^0)]\ A_1^{0\top}\Big]C_{A,2}^{-1}(\gamma) \\
V_{A,12}(\gamma) &= -C_{A,1}^{-1}(\gamma)A_1^0\ [H_1^u(\gamma) + H_1^{\epsilon,u}(\gamma) + R_1(\gamma;\rho^0)(H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)) \\
&\quad - (H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma))R_1^\top(\gamma;\rho^0) \\
&\quad + R_1(\gamma;\rho^0)H_1^u(\rho^0)R_1^\top(\gamma;\rho^0)]\ A_1^{0\top}C_{A,2}^{-1}(\gamma)
\end{aligned}
$$

whenever $\gamma \leq \rho^0$*. When* $\gamma > \rho^0$*, then*

$$
\begin{aligned}
V_{A,1}(\gamma) = C_{A,1}^{-1}(\gamma)\Big[&A_1^0H_1^\epsilon(\rho^0)A_1^{0\top} + A_2^0H_2^\epsilon(\rho^0)A_2^{0\top} + A_2^0[H_2^u(\gamma) - H_2^\epsilon(\gamma)]A_2^{0\top} \\
&+ A_2^0R_2(\gamma;\rho^0)H_2^u(\rho^0)R_2^\top(\gamma;\rho^0)A_2^{0\top} \\
&+ A_2^0H_2^{\epsilon,u}(\rho^0)R_2^\top(\gamma;\rho^0)A_2^{0\top} \\
&+ A_2^0R_2(\gamma;\rho^0)H_2^{\epsilon,u}(\rho^0)A_2^{0\top} \\
&- A_2^0[H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)]R_2^\top(\gamma;\rho^0)A_2^{0\top} \\
&- A_2^0R_2(\gamma;\rho^0)[H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)]A_2^{0\top}\Big]C_{A,1}^{-1}(\gamma) \\
V_{A,2}(\gamma) = C_{A,2}^{-1}(\gamma)A_2^0\Big[&H_2(\gamma) + R_2(\gamma;\rho^0)H_2^u(\rho^0)R_2^\top(\gamma;\rho^0) \\
&- [H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)]R_2^\top(\gamma;\rho^0) \\
&- R_2(\gamma;\rho^0)[H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)]\Big]A_2^{0\top}C_{A,2}^{-1}(\gamma) \\
V_{A,12}(\gamma) = -C_{A,1}^{-1}(\gamma)A_2^0\Big[&[H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)] + H_2^{\epsilon,u}(\rho^0)R_2^\top(\gamma;\rho^0) \\
&+ R_2(\gamma;\rho^0)H_2^u(\rho^0)R_2^\top(\gamma;\rho^0) \\
&- [H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)]R_2^\top(\gamma;\rho^0) \\
&- R_2(\gamma;\rho^0)[H_2^{\epsilon,u}(\gamma) + H_2^u(\gamma)]\Big]A_2^{0\top}C_{A,2}^{-1}(\gamma).
\end{aligned}
$$

$\hat{V}_A(\gamma)$ *is constructed by replacing all quantities in the definition of* $V_A(\gamma)$ *by their sample analogs, denoted with a hat accent. For example,* $\hat{C}_{A,1} = \hat{A}_1\hat{M}_1(\gamma \wedge \rho)\hat{A}_1^\top + \hat{A}_2[\hat{M}_1(\gamma) - \hat{M}_1(\gamma \wedge \rho)]\hat{A}_2^\top$*,* $\hat{A}_i = [\hat{\Pi}_i, S^\top]^\top$ *and* $\hat{R}_i(\gamma;\hat{\rho}) = \hat{M}_i(\gamma)\hat{M}_i^{-1}(\hat{\rho})$*.*

B. Proofs

In what follows, we use the symbol K to denote a strictly positive constant. Whenever needed, we use a subscript to distinguish among different constants.

For any $m \times 1$-vector x we denote by $\|x\|_2 = \sqrt{\sum_{i=1}^m x_i^2}$ the Euclidean norm. Moreover, for any real $m \times n$-matrix X we denote by $\|X\|_F = \sqrt{\operatorname{tr}(X^\top X)} = \sqrt{\operatorname{tr}(XX^\top)}$ the Frobenius matrix-norm which is submultiplicative, i.e. for two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times l}$ it holds that $\|AB\|_F \leq \|A\|_F\|B\|_F$, and is compatible with the Euclidean norm, i.e. for a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $x \in \mathbb{R}^{n\times 1}$ it holds that $\|Ax\|_2 \leq \|A\|_F\|x\|_2$. Also note that, for two vectors u, $v \in \mathbb{R}^{n\times 1}$ it holds that $\|uv^\top\|_F = \sqrt{\sum_i\sum_j |u_iv_j|^2} = \sqrt{\sum_i |u_i|^2 \sum_j |v_j|^2} = \sqrt{\sum_i |u_i|^2}\sqrt{\sum_j |v_j|^2} = \|u\|_2 \cdot \|v\|_2$. Furthermore, we denote by I_m the $m \times m$-identity matrix and by $\mathbf{0}_{m\times n}$ an $m \times n$-matrix of zeros.

To simplify notation, we define the following sets $\mathcal{T}_1(\gamma) = \{t : \mathbb{1}_{\{q_t\leq\gamma\}}\}$ and $\mathcal{T}_2(\gamma) = \{: \mathbb{1}_{\{q_t>\gamma\}}\}$. These sets partition the data according to the decision rules $\mathbb{1}_{\{q_t\leq\gamma\}}$ and $\mathbb{1}_{\{q_t>\gamma\}}$, respectively, and will be convenient to display sums.
Moreover, we define $\tilde{\epsilon} = \epsilon + (Z - \hat{Z})\theta_z^0$ and $s = \epsilon + u\theta_z^0$. Also, let $\bar{u}_t = \operatorname{vec}(u_t^\top, \mathbf{0}_{1\times p_2})^\top$ denote the augmented RF error. This way, we can write $w_t = A^0x_t + \bar{u}_t$ for a LRF. Note that $\tilde{\epsilon}$ can also be partitioned into regimes, with $\tilde{\epsilon}_1^\gamma = \epsilon_1^\gamma + (Z - \hat{Z})_1^\gamma\theta_z^0$ and $\tilde{\epsilon}_2^\gamma = \epsilon_2^\gamma + (Z - \hat{Z})_2^\gamma\theta_z^0$.

All convergence results, if not otherwise stated, are uniformly in γ. Moreover, $\xrightarrow{p}$ denotes convergence in probability and $\Rightarrow$ denotes weak convergence in the Skorokhod-metric.

Proofs for Section 4.3: 2SLS tests and a LRF

To prove Theorem 2, we first provide four Lemmata and their proofs.

Lemma 1. *Suppose Assumption A.1 holds. Then*

$$T^{-1/2}\operatorname{vec}(X_1^{\gamma\top}v) \Rightarrow \mathcal{GP}_1(\gamma)$$

where $\mathcal{GP}_1(\gamma)$ is a zero-mean Gaussian Process with covariance function

$$\mathcal{C}_{\mathcal{GP}}(\gamma_1,\gamma_2) = \mathbb{E}[\mathcal{GP}_1(\gamma_1)\mathcal{GP}_1^\top(\gamma_2)] = \mathbb{E}[(v_tv_t^\top \otimes x_tx_t^\top)\mathbb{1}_{\{q_t\leq(\gamma_1\wedge\gamma_2)\}}]$$

Proof of Lemma 1. Recall that X is a $T \times q$-matrix and v is a $T \times (1 + p_1)$-matrix, both satisfying Assumption 1. Further, let $v_{:,i}$ denote the i-th column of the matrix v. Then, by Hansen (1996, Theorem 1)

$$T^{-1/2}X_1^{\gamma\top}v_{:,i} \Rightarrow \mathcal{GP}_1^i(\gamma)$$

and therefore

$$T^{-1/2}\operatorname{vec}(X_1^{\gamma\top}v) \Rightarrow \begin{pmatrix} \mathcal{GP}_1^1(\gamma) \\ \vdots \\ \mathcal{GP}_1^{1+p_1}(\gamma) \end{pmatrix}. \tag{B.1}$$

By Hansen (1996, Theorem 1), $\mathcal{GP}_1^i(\gamma)$ is a zero-mean Gaussian Process with covariance function

$$\mathcal{C}^i_{\mathcal{GP}}(\gamma_1,\gamma_2) = \mathbb{E}[x_t x_t^\top v_{i,t}^2 \mathbb{1}_{\{q_t\le(\gamma_1\wedge\gamma_2)\}}]. \tag{B.2}$$

Similarly, it holds that

$$\mathcal{C}^{i,j}_{\mathcal{GP}}(\gamma_1,\gamma_2) = \mathbb{E}[\mathcal{GP}_1^i(\gamma_1)\mathcal{GP}_1^{j\top}(\gamma_2)] = \mathbb{E}[x_t x_t^\top v_{i,t} v_{j,t} \mathbb{1}_{\{q_t\le(\gamma_1\wedge\gamma_2)\}}]. \tag{B.3}$$

Combining (B.2) and (B.3),

$$\mathcal{C}_{\mathcal{GP}}(\gamma_1,\gamma_2) = \mathbb{E}[\mathcal{GP}_1(\gamma_1)\mathcal{GP}_1^{\top}(\gamma_2)] = \mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t\le(\gamma_1\wedge\gamma_2)\}}]. \tag{B.4}$$

Results (B.1) and (B.4) complete the proof. □

Lemma 2. *Suppose Assumption A.1 holds. Then*

(*i*) $T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} \xrightarrow{p} A^0 M_1(\gamma) A^{0\top} \equiv C_1(\gamma)$

(*ii*) $T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Rightarrow A^0\left(\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma)M^{-1}\mathcal{GP}_{\text{mat},1}\check{\theta}_z^0\right).$

Proof of Lemma 2. First, we prove claim (*i*) and then claim (*ii*).
Claim (*i*)**:** The RF predicted values are

$$\hat{Z} = X\hat{\Pi} \tag{B.5}$$

and

$$T^{1/2}(\hat{\Pi}-\Pi^0) = \left(T^{-1}X^\top X\right)^{-1}\left(T^{-1/2}X^\top u\right). \tag{B.6}$$

By Hansen (1996, Theorem 1), it holds uniformly in γ that

$$T^{-1}X_1^{\gamma\top}X_1^{\gamma} \xrightarrow{a.s.} M_1(\gamma), \text{ and } T^{-1}X^\top X \xrightarrow{a.s.} M. \tag{B.7}$$

This implies that $T^{-1}X^\top X = \mathcal{O}_p(1)$. By Lemma 1, $T^{-1/2}X^\top u = \mathcal{O}_p(1)$. Therefore, $T^{1/2}(\hat{\Pi}-\Pi^0) = \mathcal{O}_p(1)$ and so $\hat{\Pi}-\Pi^0 = o_p(1)$. Therefore, uniformly in γ,

$$T^{-1}\hat{Z}_1^{\gamma\top}\hat{Z}_1^{\gamma} = \hat{\Pi}^\top\left(T^{-1}X_1^{\gamma\top}X_1^{\gamma}\right)\hat{\Pi} \xrightarrow{p} \Pi^{0\top}M_1(\gamma)\Pi^0. \tag{B.8}$$

Last, with S the selection matrix such that $x_{1t} = x_t S$, it holds that

$$\hat{W}_1^{\gamma} = \left[\hat{Z}_1^{\gamma}\quad X_{1,1}^{\gamma}\right] = \left[X_1^{\gamma}\hat{\Pi}\quad X_{1,1}^{\gamma}\right] = X_1^{\gamma}\left[\hat{\Pi}\quad S\right] = X_1^{\gamma}\hat{A}^\top. \tag{B.9}$$

Therefore, by (B.8) and (B.9) and uniformly in γ,

$$T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} = \hat{A}\left(T^{-1}X_1^{\gamma\top}X_1^{\gamma}\right)\hat{A}^\top \xrightarrow{p} A^0 M_1(\gamma)A^{0\top} \equiv C_1(\gamma).$$

Claim (*ii*)**:** By (B.5) it follows that

$$T^{-1/2}\hat{Z}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} = \hat{\Pi}^\top(\underbrace{T^{-1/2}X_1^{\gamma\top}(\epsilon_1^{\gamma}+u_1^{\gamma}\theta_z^0)}_{=(\text{I})} - \underbrace{T^{-1/2}X_1^{\gamma\top}X_1^{\gamma}(\hat{\Pi}-\Pi^0)\theta_z^0}_{=(\text{II})}). \tag{B.10}$$

Next, we analyze the limiting behavior of (I) and (II). Recalling that $\tilde{\theta}_z^0 = (1, \theta_z^{0\top})^\top$,

$$I = T^{-1/2} X_1^{\gamma\top}(\epsilon_1^\gamma + u_1^\gamma \theta_z^0) = T^{-1/2}[X_1^{\gamma\top}\epsilon_1^\gamma, X_1^{\gamma\top} u_1^\gamma]\tilde{\theta}_z^0$$

and thus, by Lemma 1, uniformly in γ:

$$T^{-1/2}[X_1^{\gamma\top}\epsilon_1^\gamma, X_1^{\gamma\top} u_1^\gamma]\tilde{\theta}_z^0 \Rightarrow \mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0. \tag{B.11}$$

By (B.6), term (II) in (B.10) satisfies

$$II = T^{-1/2} X_1^{\gamma\top} X_1^\gamma(\hat{\Pi} - \Pi^0)\theta_z^0 = \left(T^{-1} X_1^{\gamma\top} X_1^\gamma\right)\left(T^{-1} X^\top X\right)^{-1}\left(T^{-1/2} X^\top u\theta_z^0\right). \tag{B.12}$$

Recalling that $\check{\theta}_z^0 = (0, \theta_z^{0\top})^\top$,

$$T^{-1/2} X^\top u\theta_z^0 = T^{-1/2} X^\top \epsilon \cdot 0 + T^{-1/2} X^\top u\theta_z^0 = T^{-1/2}[X^\top \epsilon, X^\top u]\check{\theta}_z^0 \tag{B.13}$$

So, by (B.7), (B.12)–(B.13) and Lemma 1, uniformly in γ,

$$T^{-1/2} X_1^{\gamma\top} X_1^\gamma(\hat{\Pi} - \Pi^0)\theta_z^0 \Rightarrow M_1(\gamma) M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0. \tag{B.14}$$

Next, because for any $a, b = (1)$, $\hat{\Pi}^\top(a - b) = \Pi^{0\top}(a - b) + o_p(1)$, (B.11) and (B.14) together with (B.10) yield, uniformly in γ,

$$T^{-1/2}\hat{Z}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma \Rightarrow \Pi^{0\top}\left(\mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma) M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0\right). \tag{B.15}$$

Last, because $\hat{W}_1^{\gamma\top} = \begin{bmatrix}\hat{Z}_1^\gamma & X_{1,1}^\gamma\end{bmatrix} = X_1^\gamma \hat{A}^\top$ (see (B.9)) it immediately follows with (B.15) that, uniformly in γ,

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma \Rightarrow \mathcal{B}_1(\gamma), \tag{B.16}$$

proving claim (ii). □

Lemma 3. *Suppose Assumption A.1 holds and define* $\hat{\theta}^\gamma = \mathrm{vec}(\hat{\theta}_1^\gamma, \hat{\theta}_2^\gamma)$, *and* $\bar{\theta}^0 = \mathrm{vec}(\theta^0, \theta^0)$. *Then, under* $\mathbb{H}_0$ *and for a fixed* γ*:*

$$T^{1/2}(\hat{\theta}^\gamma - \bar{\theta}^0) \Rightarrow \mathcal{N}(0, \Sigma^\gamma)$$

with

$$\Sigma^\gamma = \begin{bmatrix} V_1(\gamma) & V_{12}(\gamma) \\ V_{12}^\top(\gamma) & V_2(\gamma), \end{bmatrix}$$

where $V_1(\gamma), V_2(\gamma)$ *and* $V_{12}(\gamma)$ *are defined in Definition A.2.*

Proof of Lemma 3. First, we define the following quantities

$$\bar{W} = \begin{bmatrix} \hat{W}_1^\gamma & \mathbf{0} \\ \mathbf{0} & \hat{W}_2^\gamma \end{bmatrix}, \; \bar{Y} = \begin{bmatrix} Y_1^\gamma \\ Y_2^\gamma \end{bmatrix}, \; \hat{\theta}^\gamma = \begin{bmatrix} \hat{\theta}_1^\gamma \\ \hat{\theta}_2^\gamma \end{bmatrix}.$$

Thus, the 2SLS estimator is given by

$$\hat{\theta}^{\gamma} = (\bar{W}^{\top}\bar{W})^{-1}\bar{W}^{\top}\bar{Y} = \bar{\theta}^{0} + (\bar{W}^{\top}\bar{W})^{-1}\bar{W}^{\top}\bar{\tilde{\epsilon}}.$$

where

$$\bar{\tilde{\epsilon}} = \begin{bmatrix} \tilde{\epsilon}_1^{\gamma} \\ \tilde{\epsilon}_2^{\gamma} \end{bmatrix} = \begin{bmatrix} \epsilon_1^{\gamma} + (Z - \hat{Z})_1^{\gamma}\theta_z^0 \\ \epsilon_2^{\gamma} + (Z - \hat{Z})_2^{\gamma}\theta_z^0 \end{bmatrix}.$$

By Lemma 2,

$$T^{1/2}(\hat{\theta}^{\gamma} - \bar{\theta}^0) \Rightarrow \begin{bmatrix} C_1^{-1}(\gamma)\mathcal{B}_1(\gamma) \\ C_2^{-1}(\gamma)\mathcal{B}_2(\gamma) \end{bmatrix}.$$

Thus, we are left to derive

$$\Sigma^{\gamma} = \begin{bmatrix} \mathrm{Var}[C_1^{-1}(\gamma)\mathcal{B}_1(\gamma)] & \mathrm{Cov}[C_1^{-1}(\gamma)\mathcal{B}_1(\gamma), C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)] \\ \mathrm{Cov}[C_2^{-1}(\gamma)\mathcal{B}_2(\gamma), C_1^{-1}(\gamma)\mathcal{B}_1(\gamma)] & \mathrm{Var}[C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)] \end{bmatrix}.$$

Start with $\mathrm{Var}[\mathcal{B}_1(\gamma)]$. Write $v_t v_t^{\top} \otimes x_t x_t^{\top}$ as a short-cut for $(v_t v_t^{\top}) \otimes (x_t x_t^{\top})$, and $\check{\theta}_z^{0\top} \otimes A^0 M_1(\gamma) M^{-1}$ as a short-cut for $\check{\theta}_z^{0\top} \otimes (A^0 M_1(\gamma) M^{-1})$. Then:

$$\begin{aligned}
\mathrm{Var}[\mathcal{B}_1(\gamma)] &= \mathrm{Var}[A^0\mathcal{G}\mathcal{P}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0 - A^0 M_1(\gamma) M^{-1}\mathcal{G}\mathcal{P}_{\mathrm{mat},1}\check{\theta}_z^0] \\
&= \mathrm{Var}[(\tilde{\theta}_z^{0\top} \otimes A^0)\mathcal{G}\mathcal{P}_1(\gamma)] + \mathrm{Var}[(\check{\theta}_z^{0\top} \otimes A^0 M_1(\gamma) M^{-1})\mathcal{G}\mathcal{P}] \\
&- \mathrm{Cov}[(\tilde{\theta}_z^{0\top} \otimes A^0)\mathcal{G}\mathcal{P}_1(\gamma), (\check{\theta}_z^{0\top} \otimes A^0 M_1(\gamma) M^{-1})\mathcal{G}\mathcal{P}] \\
&- \mathrm{Cov}[(\check{\theta}_z^{0\top} \otimes A^0 M_1(\gamma) M^{-1})\mathcal{G}\mathcal{P}, (\tilde{\theta}_z^{0\top} \otimes A^0)\mathcal{G}\mathcal{P}_1(\gamma)] \\
&= (\tilde{\theta}_z^{0\top} \otimes A^0)\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t \le \gamma\}}](\tilde{\theta}_z^0 \otimes A^{0\top}) \\
&+ (\check{\theta}_z^{0\top} \otimes A^0 M_1(\gamma) M^{-1})\mathbb{E}[v_t v_t^{\top} \otimes x_t x_t^{\top}](\check{\theta}_z^0 \otimes M^{-1} M_1(\gamma) A^{0\top}) \\
&- (\tilde{\theta}_z^{0\top} \otimes A^0)\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t \le \gamma\}}](\check{\theta}_z^0 \otimes M^{-1} M_1(\gamma) A^{0\top}) \\
&- (\check{\theta}_z^{0\top} \otimes A^0 M_1(\gamma) M^{-1})\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t \le \gamma\}}](\tilde{\theta}_z^0 \otimes A^{0\top}) \\
&= A^0\mathbb{E}[x_t x_t^{\top}(\epsilon_t + u_t^{\top}\theta_z^0)^2\mathbb{1}_{\{q_t \le \gamma\}}]A^{0\top} \\
&+ A^0 M_1(\gamma) M^{-1}\mathbb{E}[x_t x_t^{\top}(u_t^{\top}\theta_z^0)^2] M^{-1} M_1(\gamma) A^{0\top} \\
&- A^0\mathbb{E}[x_t x_t^{\top}(\epsilon_t u_t^{\top}\theta_z^0 + \theta_z^{0\top} u_t u_t^{\top}\theta_z^0)\mathbb{1}_{\{q_t \le \gamma\}}] M^{-1} M_1(\gamma) A^{0\top} \\
&- A^0 M_1(\gamma) M^{-1}\mathbb{E}[x_t x_t^{\top}](\epsilon_t u_t^{\top}\theta_z^0 + \theta_z^{0\top} u_t u_t^{\top}\theta_z^0)\mathbb{1}_{\{q_t \le \gamma\}}]A^{0\top},
\end{aligned}$$

which yields the claim for $V_1(\gamma)$, when pre- and post-multiplied by $C_1^{-1}(\gamma)$.

Next, we consider $\mathrm{Var}[\mathcal{B}_2(\gamma)]$. First, note that

$$\begin{aligned}
\mathcal{B}_2(\gamma) &= A^0\mathcal{G}\mathcal{P}_{\mathrm{mat},1}\tilde{\theta}_z^0 - A^0\mathcal{G}\mathcal{P}_{\mathrm{mat},1}\check{\theta}_z^0 - A^0\mathcal{G}\mathcal{P}_{\mathrm{mat},1}(\gamma) + A^0 M_1(\gamma) M^{-1}\mathcal{G}\mathcal{P}_{\mathrm{mat},1}\check{\theta}_z^0 \\
&= A^0\mathcal{G}\mathcal{P}_{\mathrm{mat},2}(\gamma)\tilde{\theta}_z^0 - A^0 M_2(\gamma) M^{-1}\mathcal{G}\mathcal{P}_{\mathrm{mat},1}\check{\theta}_z^0
\end{aligned}$$

By similar arguments as for $\mathrm{Var}[\mathcal{B}_1(\gamma)]$,

$$\begin{aligned}\mathrm{Var}[\mathcal{B}_2(\gamma)] = &A^0\mathbb{E}[x_t x_t^\top(\epsilon_t + u_t^\top\theta_z^0)^2\mathbb{1}_{\{q_t>\gamma\}}]A^{0\top}\\ &+ A^0 M_2(\gamma)M^{-1}\mathbb{E}[x_t x_t^\top(u_t^\top\theta_z^0)^2]M^{-1}M_2(\gamma)A^{0\top}\\ &- A^0\mathbb{E}[x_t x_t^\top(\epsilon_t u_t^\top\theta_z^0 + \theta_z^{0\top}u_t u_t^\top\theta_z^0)\mathbb{1}_{\{q_t>\gamma\}}]M^{-1}M_2(\gamma)A^{0\top}\\ &- A^0 M_2(\gamma)M^{-1}\mathbb{E}[x_t x_t^\top(\epsilon_t u_t^\top\theta_z^0 + \theta_z^{0\top}u_t u_t^\top\theta_z^0)\mathbb{1}_{\{q_t>\gamma\}}]A^{0\top}\end{aligned}$$

which yields the claim for $V_2(\gamma)$, when pre- and post-multiplied by $C_2^{-1}(\gamma)$.

Finally, we derive an expression for $\mathrm{Cov}[\mathcal{B}_1(\gamma), \mathcal{B}_2(\gamma)]$[18]:

$$\begin{aligned}\mathrm{Cov}[\mathcal{B}_1(\gamma), \mathcal{B}_2(\gamma)] &= \mathrm{Cov}[A^0\mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0 - A^0M_1(\gamma)M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0,\\ &\qquad A^0\mathcal{GP}_{\mathrm{mat},2}(\gamma)\tilde{\theta}_z^0 - A^0M_2(\gamma)M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0]\\ &= -\mathrm{Cov}[A^0\mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0, A^0M_2(\gamma)M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0]\\ &\quad - \mathrm{Cov}[A^0M_1(\gamma)M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0, A^0\mathcal{GP}_{\mathrm{mat},2}(\gamma)\tilde{\theta}_z^0]\\ &\quad + \mathrm{Cov}[A^0M_1(\gamma)M^{-1}\mathcal{GP}_{\mathrm{mat},1}\check{\theta}_z^0, A^0M_2(\gamma)M^{-1}\mathcal{GP}_{\mathrm{mat},1}(\gamma)\check{\theta}_z^0]\\ &= -A^0\mathbb{E}[x_t x_t^\top(\epsilon_t u_t^\top\theta_z^0 + \theta_z^{0\top}u_t u_t^\top\theta_z^0)\mathbb{1}_{\{q_t\le\gamma\}}]M^{-1}M_2(\gamma)A^{0\top}\\ &\quad - A^0M_1(\gamma)M^{-1}\mathbb{E}[x_t x_t^\top(\epsilon_t u_t^\top\theta_z^0 + \theta_z^{0\top}u_t u_t^\top\theta_z^0)\mathbb{1}_{\{q_t>\gamma\}}]A^{0\top}\\ &\quad + AM_1(\gamma)M^{-1}\mathbb{E}[x_t x_t^\top(u_t^\top\theta_z^0)^2]M^{-1}M_2(\gamma)A^{0\top}\end{aligned}$$

which yields the claim for $V_{12}(\gamma)$ when pre-multiplied by $C_1^{-1}(\gamma)$ and post-multiplied by $C_2^{-1}(\gamma)$. □

Lemma 4. *Suppose Assumption A.1 holds. Under $\mathbb{H}_0$ and uniformly in γ for $i = 1, 2$,*

(i) $\hat{H}_i^\epsilon(\gamma) \xrightarrow{p} H_i^\epsilon(\gamma)$ *(ii)* $\hat{H}_i^{\epsilon,u}(\gamma) \xrightarrow{p} H_i^{\epsilon,u}(\gamma)$

(iii) $\hat{H}_i^{u}(\gamma) \xrightarrow{p} H_i^{u}(\gamma)$ *(iv)* $\hat{H}_i(\gamma) \xrightarrow{p} H_i(\gamma)$

Proof of Lemma 4. **Claim** (*i*)**:** Note that, under $\mathbb{H}_0$, $\hat{\epsilon}_t = y_t - w_t^\top\hat{\theta}$ and start with

$$\begin{aligned}\hat{H}_i^\epsilon(\gamma) &= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \hat{\epsilon}_t^2\\ &= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (y_t - w_t^\top\hat{\theta})^2\\ &= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top [w_t^\top(\theta^0 - \hat{\theta}) + \epsilon_t]^2\\ &= \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top [w_t^\top(\theta^0 - \hat{\theta})]^2}_{\text{(I)}} + \underbrace{2T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \epsilon_t w_t^\top(\theta^0 - \hat{\theta})}_{\text{(II)}} + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \epsilon_t^2}_{\text{(III)}}.\end{aligned}$$

[18]Note that $\mathrm{Cov}[\mathcal{GP}_1(\gamma), \mathcal{GP}_2(\gamma)] = \mathbb{E}[\mathcal{GP}_1(\gamma)\mathcal{GP}_2^\top(\gamma)] = \mathbf{0}$.

We are left to show the limiting behavior of (I), (II), and (III).

$$
\begin{aligned}
\|(\mathrm{I})\|_F &\leq T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t x_t^\top [w_t^\top(\theta^0 - \hat{\theta})]^2\|_F \\
&\leq \Big(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|w_t\|_2^2\Big) \|\theta^0 - \hat{\theta}\|_2^2 \\
&= \Big(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|A^0 x_t + \bar{u}_t\|_2^2\Big) \|\theta^0 - \hat{\theta}\|_2^2 \\
&\leq \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \Big[\|A^0\|_F \|x_t\|_2 + \|u_t\|_2\Big]^2\right) \|\theta^0 - \hat{\theta}\|_2^2 \\
&= \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4 \|A^0\|_F^2 + 2\|x_t\|_2^3 \|u_t\|_2 \|A^0\|_F + \|x_t\|_2^2 \|u_t\|_2^2\right) \|\theta^0 - \hat{\theta}\|_2^2 \\
&= o_p(1)
\end{aligned}
\tag{B.17}
$$

where the last equality holds because $\|\theta^0 - \hat{\theta}\| = o_p(1)$ under $\mathbb{H}_0$ (follows directly from Lemma 2 by dropping γ) and the term in paranthesis is $\mathcal{O}_p(1)$. To see this latter claim, note that $\|A^0\|_F = \mathcal{O}_p(1)$ by Assumption A.1 and consider

$$
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4 > K_1\right) \leq \frac{\mathbb{E} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4}{T K_1} \leq \frac{\sup_t \mathbb{E}\|x_t\|_2^4}{K_1}, \tag{B.18a}
$$

$$
\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2 > K_2\right) &\leq \frac{\mathbb{E} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2}{T K_2} \leq \frac{\sup_t \mathbb{E}\|x_t\|_2^3 \|u_t\|_2}{K_2} \\
&\leq \frac{\sup_t \Big[\mathbb{E}\|x_t\|_2^4\Big]^{3/4} \Big[\mathbb{E}\|u_t\|_2^4\Big]^{1/4}}{K_2} \\
&\leq \frac{\sup_t \Big[\mathbb{E}\|x_t\|_2^4\Big]^{3/4} \sup_t \Big[\mathbb{E}\|u_t\|_2^4\Big]^{1/4}}{K_2}
\end{aligned}
\tag{B.18b}
$$

and

$$
\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|u_t\|_2^2\right) &\leq \frac{\mathbb{E} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|u_t\|_2^2}{T K_3} \leq \frac{\sup_t \mathbb{E}\|x_t\|_2^2 \|u_t\|_2^2}{K_3} \\
&\leq \frac{\sup_t \Big[\mathbb{E}\|x_t\|_2^4 \mathbb{E}\|u_t\|_2^4\Big]^{1/2}}{K_3} \leq \frac{\sup_t \Big[\mathbb{E}\|x_t\|_2^4\Big]^{1/2} \sup_t \Big[\mathbb{E}\|u_t\|_2^4\Big]^{1/2}}{K_3}.
\end{aligned}
\tag{B.18c}
$$

Now, by Assumption A.1.2 it follows that all three terms (B.18a)–(B.18c) are $\mathcal{O}_p(1)$ and therefore, (B.17) follows.
For (II) it follows that

$$\begin{aligned}
\|(\mathrm{II})\|_F &\leq T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t x_t^\top w_t^\top (\theta^0 - \hat{\theta}) \epsilon_t\|_F \\
&\leq \Big(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|A^0 x_t + \bar{u}_t\|_2 |\epsilon_t| \Big) \|\theta^0 - \hat{\theta}\|_2 \\
&\leq \Big(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|A^0\|_F |\epsilon_t| + \|x_t\|_2^2 \|u_t\|_2 |\epsilon_t| \Big) \|\theta^0 - \hat{\theta}\|_2 \\
&= o_p(1). \qquad \text{(B.19)}
\end{aligned}$$

To see why this statement holds, consider

$$\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| > K_4 \right) &\leq \frac{\mathbb{E} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t|}{T K_4} \leq \frac{\sup_t \mathbb{E} \|x_t\|_2^3 |\epsilon_t|}{K_4} \\
&\leq \frac{\sup_t \left[\mathbb{E} \|x_t\|_2^4 \right]^{3/4} \left[\mathbb{E} |\epsilon_t|^4 \right]^{1/4}}{K_4} \\
&\leq \frac{\sup_t \left[\mathbb{E} \|x_t\|_2^4 \right]^{3/4} \sup_t \left[\mathbb{E} |\epsilon_t|^4 \right]^{1/4}}{K_4} \qquad \text{(B.20a)}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|u_t\|_2 |\epsilon_t| > K_5 \right) &\leq \frac{\mathbb{E} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \|u_t\|_2 |\epsilon_t|}{T K_5} \leq \frac{\sup_t \mathbb{E} \|x_t\|_2^2 \|u_t \epsilon_t\|_2}{K_5} \\
&\leq \frac{\sup_t \left[\mathbb{E} \|x_t\|_2^4 \right]^{1/2} \left[\mathbb{E} \|u_t \epsilon_t\|_2^2 \right]^{1/2}}{K_5} \\
&\leq \frac{\sup_t \left[\mathbb{E} \|x_t\|_2^4 \right]^{1/2} \left[\mathbb{E} \|u_t\|_2^4 \right]^{1/4} \left[\mathbb{E} |\epsilon_t|^4 \right]^{1/4}}{K_5} \\
&\leq \frac{\sup_t \left[\mathbb{E} \|x_t\|_2^4 \right]^{1/2} \sup_t \left[\mathbb{E} \|u_t\|_2^4 \right]^{1/4} \sup_t \left[\mathbb{E} |\epsilon_t|^4 \right]^{1/4}}{K_5}. \qquad \text{(B.20b)}
\end{aligned}$$

Thus, the fact that $\|\theta^0 - \hat{\theta}\|_2 = o_p(1)$ together with (B.20a) and (B.20b) yield (B.19). Finally, because (III) $\xrightarrow{p} H_1^\epsilon(\gamma)$, uniformly in γ, by Hansen (1996, Lemma 1), Claim (i) follows.

Claim (ii)**:** Under $\mathbb{H}_0$ and a LRF it holds that $\hat{\epsilon}_t = y_t - w_t^\top \hat{\theta}$ and $\hat{u}_t = z_t - \hat{\Pi}^\top x_t$. Therefore,

$$\begin{aligned}
\hat{H}_i^{\epsilon,u}(\gamma) &= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z = T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top [w_t^\top(\theta^0-\hat{\theta})+\epsilon_t][x_t^\top(\Pi^0-\hat{\Pi})+u_t^\top]\hat{\theta}_z \\
&= \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top x_t^\top A^0(\theta^0-\hat{\theta}) x_t^\top(\Pi^0-\hat{\Pi})\hat{\theta}_z}_{\text{(IV)}} + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top x_t^\top A^0(\theta^0-\hat{\theta}) u_t^\top \hat{\theta}_z}_{\text{(V)}} \\
&\quad + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \bar{u}_t^\top(\theta^0-\hat{\theta}) x_t^\top(\Pi^0-\hat{\Pi})\hat{\theta}_z}_{\text{(VI)}} + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \bar{u}_t^\top(\theta^0-\hat{\theta}) u_t^\top \hat{\theta}_z}_{\text{(VII)}} \\
&\quad + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \epsilon_t x_t^\top(\Pi^0-\hat{\Pi})\hat{\theta}_z}_{\text{(VIII)}} + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \epsilon_t^\top u_t^\top \hat{\theta}_z}_{\text{(IX)}} . \qquad \text{(B.21)}
\end{aligned}$$

Now, it immediately follows, by similar arguments as for Claim (i), that

$$\begin{aligned}
\|\text{(IV)}\|_F &\leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4\Big)\|A^0\|_F\|\Pi^0-\hat{\Pi}\|_F\|\theta^0-\hat{\theta}\|_2\|\hat{\theta}_z\|_2 \\
&= \mathcal{O}_p(1)\mathcal{O}_p(1)o_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1) \qquad \text{(B.22a)} \\
\|\text{(V)}\|_F &\leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3\|u_t\|_2\Big)\|A^0\|_F\|\theta^0-\hat{\theta}\|_2\|\hat{\theta}_z\|_2 \\
&= \mathcal{O}_p(1)\mathcal{O}_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1) \qquad \text{(B.22b)} \\
\|\text{(VI)}\|_F &\leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3\|u_t\|_2\Big)\|\Pi^0-\hat{\Pi}\|_F\|\theta^0-\hat{\theta}\|_2\|\hat{\theta}_z\|_2 \\
&= \mathcal{O}_p(1)o_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1) \qquad \text{(B.22c)} \\
\|\text{(VII)}\|_F &\leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2\|u_t\|_2^2\Big)\|\theta^0-\hat{\theta}\|_2\|\hat{\theta}_z\|_2 \\
&= \mathcal{O}_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1) \qquad \text{(B.22d)} \\
\|\text{(VIII)}\|_F &\leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3|\epsilon_t|\Big)\|\Pi^0-\hat{\Pi}\|_F\|\hat{\theta}_z\|_2 \\
&= \mathcal{O}_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1). \qquad \text{(B.22e)}
\end{aligned}$$

For the last term in (B.21) it holds, uniformly in γ by Hansen (1996, Lemma 1), that (IX) $= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \epsilon_t u_t^\top(\theta^0 + o_p(1)) = T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \epsilon_t u_t^\top \theta^0 + o_p(1) \xrightarrow{p} H_1^{\epsilon,u}(\gamma)$. Thus, Claim (ii) follows together with (B.22a)–(B.22e).

Claim (iii)**:** As before, $\hat{u}_t = z_t - \hat{\Pi}^\top x_t$. Then, under $\mathbb{H}_0$, it follows that

$$
\begin{aligned}
\hat{H}_i^u(\gamma) &= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 = T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \{[x_t^\top(\Pi^0 - \hat{\Pi}) + u_t^\top]\hat{\theta}_z\}^2 \\
&= \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top [x_t^\top(\Pi^0 - \hat{\Pi})\hat{\theta}_z]^2}_{\text{(X)}} + \underbrace{2\,T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top x_t^\top(\Pi^0 - \hat{\Pi})\hat{\theta}_z u_t^\top \hat{\theta}_z}_{\text{(XI)}} \\
&\quad + \underbrace{T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (u_t^\top \hat{\theta}_z)^2}_{\text{(XII)}} \qquad \text{(B.23)}
\end{aligned}
$$

Next,

$$
\|(\text{X})\|_F \leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4\Big)\|\Pi^0 - \hat{\Pi}\|_F^2 \|\hat{\theta}_z\|_2^2 = \mathcal{O}_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1) \qquad \text{(B.24a)}
$$

$$
\|(\text{XI})\|_F \leq \Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2\Big)\|\Pi^0 - \hat{\Pi}\|_F \|\hat{\theta}_z\|_2^2 = \mathcal{O}_p(1)o_p(1)\mathcal{O}_p(1) = o_p(1). \qquad \text{(B.24b)}
$$

For the last term in (B.23) it holds, uniformly in γ by Hansen (1996, Lemma 1), that (X) $= T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (u_t^\top \hat{\theta}_z)^2 = T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (u_t^\top \theta_z^0)^2 + o_p(1) \xrightarrow{p} H^u(\gamma)$. Thus, Claim (iii) follows together with (B.24a) and (B.24b).

Claim (iv)**:** This claim follows by noting that $\hat{H}_i(\gamma) = \hat{H}_i^\epsilon(\gamma) + 2\hat{H}_i^{\epsilon,u}(\gamma) + \hat{H}_i^u(\gamma)$, using Claims (i)–(iii) and the continuous mapping theorem. □

Proof of Theorem 2.

(i) sup LR Test: This proof is done in two parts: part (A) shows that $T^{-1}SSR_1(\gamma) \xrightarrow{p} \sigma^2$ and part (B) shows that $SSR_0 - SSR_1(\gamma) \Rightarrow \mathcal{E}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\mathcal{E}(\gamma)$.

Part (A)**.** The scaled sum of squared residuals of the restricted model, $SSR_1(\gamma)$, is

$$
\begin{aligned}
T^{-1}SSR_1(\gamma) &= T^{-1}[Y_1^\gamma - \hat{W}_1^\gamma \hat{\theta}_1^\gamma]^\top [Y_1^\gamma - \hat{W}_1^\gamma \hat{\theta}_1^\gamma] \\
&\quad + T^{-1}[Y_2^\gamma - \hat{W}_2^\gamma \hat{\theta}_2^\gamma]^\top [Y_2^\gamma - \hat{W}_2^\gamma \hat{\theta}_2^\gamma] \\
&= T^{-1}[\hat{W}_1^\gamma(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_1^\gamma]^\top [\hat{W}_1^\gamma(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_1^\gamma] \\
&\quad + T^{-1}[\hat{W}_2^\gamma(\theta^0 - \hat{\theta}_2^\gamma) + \tilde{\epsilon}_2^\gamma]^\top [\hat{W}_2^\gamma(\theta^0 - \hat{\theta}_2^\gamma) + \tilde{\epsilon}_2^\gamma] \\
&= T^{-1}\tilde{\epsilon}^\top \tilde{\epsilon} \\
&\quad + 2(T^{-1}\tilde{\epsilon}_1^{\gamma\top}\hat{W}_1^\gamma)(\theta^0 - \hat{\theta}_1^\gamma) + (\theta^0 - \hat{\theta}_1^\gamma)^\top (T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma)(\theta^0 - \hat{\theta}_1^\gamma) \\
&\quad + 2(T^{-1}\tilde{\epsilon}_2^{\gamma\top}\hat{W}_2^\gamma)(\theta^0 - \hat{\theta}_2^\gamma) + (\theta^0 - \hat{\theta}_2^\gamma)^\top (T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^\gamma)(\theta^0 - \hat{\theta}_2^\gamma).
\end{aligned}
\qquad \text{(B.25)}
$$

Next, by Lemma 2, for $i = 1, 2$, $T^{-1}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^{\gamma} = o_p(1)$ and $T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma} = \mathcal{O}_p(1)$ uniformly in γ. This implies that

$$\hat{\theta}_i^{\gamma} - \theta^0 = (T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma})^{-1}(\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^{\gamma}) = \mathcal{O}_p(1)o_p(1) = o_p(1)$$

and therefore, (B.25) simplifies to

$$\begin{aligned} T^{-1}SSR_1(\gamma) &= T^{-1}\tilde{\epsilon}^{\top}\tilde{\epsilon} + o_p(1) \\ &= T^{-1}s^{\top}s + 2(T^{-1}s^{\top}X)(\Pi^0 - \hat{\Pi})\theta_z^0 \\ &+ \theta_z^{0\top}(\Pi^0 - \hat{\Pi})X^{\top}X(\Pi^0 - \hat{\Pi})\theta_z^0 + o_p(1) \end{aligned} \tag{B.26}$$

where $s = \epsilon + u^{\top}\theta_z^0$, $\hat{\Pi} - \Pi^0 = o_p(1)$ and $T^{-1}s^{\top}X = o_p(1)$ by Lemma 2, uniformly in γ. Thus, (B.26) simplifies to

$$\begin{aligned} T^{-1}SSR_1(\gamma) &= T^{-1}s^{\top}s + o_p(1) \\ &= T^{-1}\epsilon^{\top}\epsilon + 2(T^{-1}\epsilon^{\top}u)\theta_z^0 + \theta_z^{0\top}(T^{-1}u^{\top}u)\theta_z^0 + o_p(1) \\ &\xrightarrow{p} \sigma_\epsilon^2 + 2\Sigma_{\epsilon,u}^{\top}\theta_z^0 + \theta_z^{0\top}\Sigma_u\theta_z^0 = \sigma^2 \end{aligned}$$

uniformly in γ. This proves part (i).

Part (B)**.** We have that

$$\begin{aligned} SSR_0 - SSR_1(\gamma) = &[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}]^{\top}[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}] - [Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}_1^{\gamma}]^{\top}[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}_1^{\gamma}] \\ &+ [Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}]^{\top}[Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}] - [Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}_2^{\gamma}]^{\top}[Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}_2^{\gamma}] \end{aligned} \tag{B.27}$$

Now, for $i = 1, 2$,

$$\begin{aligned} &[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}]^{\top}[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}] \\ &-[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma}]^{\top}[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma}] = Y_i^{\gamma\top}Y_i^{\gamma} - 2\hat{\theta}^{\top}\hat{W}_i^{\gamma\top}Y_i^{\gamma} + \hat{\theta}^{\top}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma}\hat{\theta} \\ &\qquad - Y_i^{\gamma\top}Y_i^{\gamma} + 2\hat{\theta}_i^{\gamma\top}\hat{W}_i^{\gamma} - \hat{\theta}_i^{\gamma\top}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma} \\ &\qquad = [\hat{\theta}_i^{\gamma} - \hat{\theta}]^{\top}\hat{W}_i^{\gamma\top}[2Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta} - \hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma}] \\ &\qquad = T^{1/2}[\hat{\theta}_i^{\gamma} - \hat{\theta}]^{\top}\Big[2(T^{-1/2}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^{\gamma}) \\ &\qquad -(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma})(T^{1/2}(\hat{\theta} - \theta^0)) \\ &\qquad -(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma})(T^{1/2}(\hat{\theta}_i^{\gamma} - \theta^0))\Big]. \end{aligned} \tag{B.28}$$

Next, we show the asymptotic behavior of the terms on the right hand side of (B.28) which then concludes the proof together with Part (i), (B.27), the continuous mapping theorem and weak convergence (uniformly in γ). It holds that

$$\begin{aligned} &(T^{-1}\hat{W}^{\top}\hat{W})(T^{1/2}(\hat{\theta} - \theta^0)) \\ =&T^{-1/2}\hat{W}^{\top}\tilde{\epsilon} \\ =&T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} + T^{-1/2}\hat{W}_2^{\gamma\top}\tilde{\epsilon}_2^{\gamma} \\ =&(T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma})(T^{1/2}(\hat{\theta}_1^{\gamma} - \theta^0)) + (T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^{\gamma})(T^{1/2}(\hat{\theta}_2^{\gamma} - \theta^0)) \end{aligned} \tag{B.29}$$

and by Lemma 2 that, uniformly in γ for $i = 1, 2$,

$$T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma} \xrightarrow{p} C_i(\gamma). \tag{B.30}$$

Define $\hat{\beta} \equiv T^{1/2}(\hat{\theta} - \theta^0)$, $\hat{\beta}_i \equiv T^{1/2}(\hat{\theta}_i^{\gamma} - \theta^0)$ and $D_i(\gamma) \equiv C^{-1}C_i(\gamma)$ $(i = 1, 2)$. Then, (B.30) can be restated as

$$\hat{\beta} = D_1(\gamma)\hat{\beta}_1 + D_2(\gamma)\hat{\beta}_2 + o_p(1). \tag{B.31}$$

Moreover, note that because $D_1(\gamma) + D_2(\gamma) = I$,

$$T^{1/2}(\hat{\theta}_1^{\gamma} - \hat{\theta}) = \hat{\beta}_1 - \hat{\beta} = D_2(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \tag{B.32a}$$

$$T^{1/2}(\hat{\theta}_2^{\gamma} - \hat{\theta}) = \hat{\beta}_2 - \hat{\beta} = -D_1(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \tag{B.32b}$$

$$T^{-1/2}\hat{W}_i^{\gamma\top}\hat{\epsilon}_i^{\gamma} = C_i(\gamma)\hat{\beta}_i + o_p(1) \tag{B.32c}$$

by (B.31) and Lemma 2.
So, using (B.29)–(B.32c), quantity (B.28) can be written, for $i = 1$, as

$$\begin{aligned} &(\hat{\beta}_1 - \hat{\beta}_2)^\top D_2^\top(\gamma)\left[2C_1(\gamma)\hat{\beta}_1 - C_1(\gamma)\hat{\beta} - C_1(\gamma)\hat{\beta}_1\right] + o_p(1) \\ =&(\hat{\beta}_1 - \hat{\beta}_2)^\top D_2^\top(\gamma)C_1(\gamma)(\hat{\beta}_1 - \hat{\beta}) + o_p(1) \\ =&(\hat{\beta}_1 - \hat{\beta}_2)^\top D_2^\top(\gamma)C_1(\gamma)D_2(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1). \end{aligned} \tag{B.33}$$

Similarly, using (B.29)–(B.31) and (B.32b), and(B.32c), quantity (B.28) can be stated, for $i = 2$, as

$$(\hat{\beta}_1 - \hat{\beta}_2)^\top D_1^\top(\gamma)C_2(\gamma)D_1(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1). \tag{B.34}$$

So, using (B.28), (B.33) and (B.34), quantity (B.27) can be restated as

$$\begin{aligned} SSR_0 - SSR_1(\gamma) &= (\hat{\beta}_1 - \hat{\beta}_2)^\top D_2^\top(\gamma)C_1(\gamma)D_2(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) \\ &+ (\hat{\beta}_1 - \hat{\beta}_2)^\top D_1^\top(\gamma)C_2(\gamma)D_1(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \\ &= (\hat{\beta}_1 - \hat{\beta}_2)^\top \left[(I_p - D_1^\top(\gamma))C_1(\gamma)(I_p - D_1(\gamma))\right. \\ &\quad \left.+D_1^\top(\gamma)(C - C_1(\gamma))D_1(\gamma)\right](\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \\ &= (\hat{\beta}_1 - \hat{\beta}_2)^\top \left[C_1(\gamma) - 2C_1(\gamma)D_1(\gamma) + D_1^\top(\gamma)C_1(\gamma)D_1(\gamma)\right. \\ &\quad \left.+D_1^\top(\gamma)CD_1(\gamma) - D_1^\top(\gamma)C_1(\gamma)D_1(\gamma)\right](\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \\ &= (\hat{\beta}_1 - \hat{\beta}_2)^\top \left[C_1(\gamma) - C_1(\gamma)D_1(\gamma)\right](\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \\ &= (\hat{\beta}_1 - \hat{\beta}_2)^\top C_2(\gamma)D_1(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1). \end{aligned} \tag{B.35}$$

Last, by Lemma 2 it holds, uniformly in γ, that

$$\begin{aligned} \hat{\beta}_1 - \hat{\beta}_2 &= (T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma})^{-1}(T^{-1/2}\hat{W}_1^{\gamma\top}\hat{\epsilon}_1^{\gamma}) - (T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^{\gamma})^{-1}(T^{-1/2}\hat{W}_2^{\gamma\top}\hat{\epsilon}_2^{\gamma}) \\ &\Rightarrow C_1^{-1}(\gamma)A^0\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)A^0\mathcal{B}_2(\gamma) \equiv \mathcal{E}(\gamma). \end{aligned} \tag{B.36}$$

So, combining (B.35) and (B.36) yields

$$SSR_0 - SSR_1(\gamma) \Rightarrow \mathcal{E}^\top(\gamma)C_2(\gamma)D_1(\gamma)\mathcal{E}(\gamma)$$

which in turn with Part (A), the continuous mapping theorem and weak convergence (uniformly in γ) proves the claim.

(ii) sup Wald Test: From Equation (B.28) it follows that

$$T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma) \Rightarrow \mathcal{E}(\gamma). \tag{B.37}$$

Moreover, from Definition A.2,

$$\begin{aligned}\hat{V}_i(\gamma) = \hat{C}_i^{-1}(\gamma)\hat{A}\Big[\hat{H}_i(\gamma) + \hat{R}_i(\gamma)\hat{H}^u\hat{R}_i^\top(\gamma) - [\hat{H}_i^{\epsilon,u}(\gamma) + \hat{H}_i^u(\gamma)]\hat{R}_i^\top(\gamma) \\ - \hat{R}_i(\gamma)[\hat{H}_i^{\epsilon,u}(\gamma) + \hat{H}_i^u(\gamma)]\Big]\hat{A}^\top\hat{C}_i^{-1}(\gamma)\end{aligned} \tag{B.38a}$$

and

$$\begin{aligned}\hat{V}_{12}(\gamma) = \hat{C}_1^{-1}(\gamma)\hat{A}\Big[(\hat{H}_1^{\epsilon,u}(\gamma) + \hat{H}_1^u(\gamma))\hat{R}_2^\top(\gamma) - \hat{R}_1(\gamma)(\hat{H}_2^{\epsilon,u}(\gamma) + \hat{H}_2^u(\gamma)) \\ + \hat{R}_1(\gamma)\hat{H}^u\hat{R}_2^\top(\gamma))\Big]\hat{A}^\top\hat{C}_2^{-1}(\gamma)\end{aligned} \tag{B.38b}$$

Now, by (B.7) and the continuous mapping theorem it immediately follows, uniformly in γ, that

$$\hat{R}_i(\gamma) = \hat{M}_i(\gamma)\hat{M}^{-1} = \left(T^{-1}X_i^{\gamma\top}X_i^\gamma\right)\left(T^{-1}X^\top X\right)^{-1} \xrightarrow{p} M_i(\gamma)M^{-1} = R_i(\gamma). \tag{B.39}$$

Moreover, by Lemma 2 and the continuous mapping theorem it also holds, uniformly in γ, that

$$\hat{C}_i^{-1}(\gamma) = (T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1} = \left(\hat{A}\hat{M}_i(\gamma)\hat{A}^\top\right)^{-1} \xrightarrow{p} C_i^{-1}(\gamma), \quad \text{and} \quad \hat{A} = [\hat{\Pi},\, S^\top]^\top \xrightarrow{p} A^0. \tag{B.40}$$

Finally, in Lemma 4 we derived the limits of $\hat{H}_i^\epsilon(\gamma)$, $\hat{H}_i^u(\gamma)$ and $\hat{H}_i^{\epsilon,u}(\gamma)$ concluding the proof together with (B.37)–(B.40). □

Corollary A 1 (to Theorem 2)**.** *Let Z be generated by (2.1), Y be generated by (2.3), and $\hat{Z}$ be calculated by (3.1). Under $\mathbb{H}_0$ and Assumptions A.1-A.2,*
(i)

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \tilde{\mathcal{E}}^\top(\gamma)Q^{-1}(\gamma)\tilde{\mathcal{E}}(\gamma)$$

(ii)

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \tilde{\mathcal{E}}^\top(\gamma)\tilde{V}^{-1}(\gamma)\tilde{\mathcal{E}}(\gamma)$$

where $\tilde{V}(\gamma) = \tilde{V}_1(\gamma) + \tilde{V}_2(\gamma) - \tilde{V}_{12}(\gamma) - \tilde{V}_{12}^\top(\gamma)$,

$$\begin{aligned}\tilde{V}_i(\gamma) &= C_i^{-1}(\gamma)\ A_0\Big[\sigma^2 I_q - (\sigma^2 - \sigma_\epsilon^2)R_i(\gamma)\Big]M_i(\gamma)A_0^\top\ C_i^{-1}(\gamma) \\ \tilde{V}_{12}(\gamma) &= -C_1^{-1}(\gamma)(\sigma^2 - \sigma_\epsilon^2)A^0R_1(\gamma)M_2(\gamma)A^{0\top}C_2^{-1}(\gamma),\end{aligned}$$

and $\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)$ is a $q\times(p_1+1)$-matrix where all columns are independent $q\times 1$ zero mean Gaussian processes with covariance kernel[19] *$M_1(\gamma_1 \wedge \gamma_2)$, $\Sigma^{1/2}$ is the principal square*

[19] Thus, the only difference between the two Gaussian processes $\widetilde{\mathcal{GP}}_{\text{mat},1}(\gamma)$ and $\mathcal{GP}_{\text{mat},1}(\gamma)$ lies in their covariance functions.

root of Σ, $\tilde{\mathcal{E}}(\gamma) = C_1^{-1}(\gamma)\tilde{\mathcal{B}}_1(\gamma) - C_2^{-1}(\gamma)\tilde{\mathcal{B}}_2(\gamma)$, *and* $\tilde{\mathcal{B}}_1(\gamma) = A^0[\tilde{\mathcal{GP}}_{\mathrm{mat},1}(\gamma)\Sigma^{1/2}\tilde{\theta}_z^0 - R_1(\gamma)\tilde{\mathcal{GP}}_{\mathrm{mat}}\Sigma^{1/2}\check{\theta}_z^0]$, $\tilde{\mathcal{B}}_2(\gamma) = \tilde{\mathcal{B}}_1(\gamma_{max}) - \tilde{\mathcal{B}}_1(\gamma)$.

(iii) If the system is just-identified, i.e. if $p = q$, *then the two test statistics are asymptotically equivalent with asymptotic distribution given by* $\sup_{\gamma\in\Gamma} J_1(\gamma)$, *where:*

$$\begin{aligned}
J_1(\gamma) = \frac{1}{\sigma^2}(\Sigma^{1/2}\tilde{\theta}_z^0)^\top &[M_1^{-1}(\gamma)\tilde{\mathcal{GP}}_{\mathrm{mat},1}(\gamma) - M_2^{-1}(\gamma)\tilde{\mathcal{GP}}_{\mathrm{mat},2}(\gamma)]^\top \\
&\times [M_2(\gamma)M^{-1}M_1(\gamma)] \\
&\times [M_1^{-1}(\gamma)\tilde{\mathcal{GP}}_{\mathrm{mat},1}(\gamma) - M_2^{-1}(\gamma)\tilde{\mathcal{GP}}_{\mathrm{mat},2}(\gamma)]\Sigma^{1/2}\tilde{\theta}_z^0.
\end{aligned}$$

Proof of Corollary A1.

(i) sup LR-test: We only need to show $\mathcal{E}(\gamma) = \tilde{\mathcal{E}}(\gamma)$ under Assumptions A.1 and A.2; in other words, that $\mathcal{GP}_{\mathrm{mat},1}(\gamma) = \tilde{\mathcal{GP}}_{\mathrm{mat},1}(\gamma)\Sigma^{1/2}$.[20] The covariance kernel of $\mathcal{GP}_1(\gamma)$ is given as $\mathbb{E}[\mathcal{GP}_1(\gamma_1)\mathcal{GP}_1^\top(\gamma_2)] = \mathbb{E}[(v_tv_t^\top \otimes x_tx_t^\top)\mathbb{1}_{\{q_t\leq\gamma_1\wedge\gamma_2\}}]$ by Lemma 1, using the shortcut notation $v_tv_t^\top \otimes x_tx_t^\top = (v_tv_t^\top)\otimes(x_tx_t^\top)$. Under Assumption A.2 this expression can be simplified to

$$\begin{aligned}
\mathbb{E}[(v_tv_t^\top \otimes x_tx_t^\top)\mathbb{1}_{\{q_t\leq\gamma_1\wedge\gamma_2\}}] &= \mathbb{E}\left[\mathbb{E}[(v_tv_t^\top \otimes x_tx_t^\top)\mathbb{1}_{\{q_t\leq(\gamma_1\wedge\gamma_2)\}}|x_t, q_t]\right] \\
&= \mathbb{E}\left[\mathbb{E}[v_tv_t^\top|x_t, q_t] \otimes (x_tx_t^\top\mathbb{1}_{\{q_t\leq(\gamma_1\wedge\gamma_2)\}})\right] \\
&= \mathbb{E}\left[\Sigma \otimes (x_tx_t^\top\mathbb{1}_{\{q_t\leq(\gamma_1\wedge\gamma_2)\}})\right] \\
&= \Sigma \otimes M_1(\gamma_1 \wedge \gamma_2).
\end{aligned}$$

Next, the principal square root of Σ, i.e. $\Sigma^{1/2}$ that satisfies $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$, exists since Σ is positive definite by Assumption A.1.5. Thus,

$$\begin{aligned}
\mathbb{E}[(v_tv_t^\top \otimes x_tx_t^\top)\mathbb{1}_{\{q_t\leq\gamma_1\wedge\gamma_2\}}] &= \Sigma \otimes M_1(\gamma_1 \wedge \gamma_2) \\
&= (\Sigma^{1/2} \otimes M_1(\gamma_1 \wedge \gamma_2))(\Sigma^{1/2} \otimes I_q) \\
&= (\Sigma^{1/2} \otimes I_q)(I_{p_1+1} \otimes M_1(\gamma_1 \wedge \gamma_2))(\Sigma^{1/2} \otimes I_q). \qquad \text{(B.41)}
\end{aligned}$$

The covariance kernel of $(\Sigma^{1/2} \otimes I_q)\tilde{\mathcal{GP}}_1(\gamma) = \mathrm{vec}(\tilde{\mathcal{GP}}_{\mathrm{mat},1}(\gamma)\Sigma^{1/2})$ is given by

$$\begin{aligned}
\mathbb{E}[(\Sigma^{1/2} \otimes I_q)\tilde{\mathcal{GP}}_1(\gamma_1)\tilde{\mathcal{GP}}_1^\top(\gamma_2)(\Sigma^{1/2} \otimes I_q)] &= (\Sigma^{1/2} \otimes I_q)\mathbb{E}[\tilde{\mathcal{GP}}_1(\gamma_1)\tilde{\mathcal{GP}}_1^\top(\gamma_2)](\Sigma^{1/2} \otimes I_q) \\
&= (\Sigma^{1/2} \otimes I_q)(I_{p_1+1} \otimes M_1(\gamma_1 \wedge \gamma_2))(\Sigma^{1/2} \otimes I_q) \qquad \text{(B.42)}
\end{aligned}$$

because $\mathbb{E}[\tilde{\mathcal{GP}}_1(\gamma_1)\tilde{\mathcal{GP}}_1^\top(\gamma_2)] = I_{p_1+1} \otimes M_1(\gamma_1 \wedge \gamma_2)$ by definition of $\tilde{\mathcal{GP}}_1(\gamma)$. Combining (B.41) and (B.42) yields the desired result since Gaussian processes are uniquely defined through their mean and covariance functions.

[20] We will do this by showing that their covariance functions are the same. Hence, because both processes have mean zero, equality follows due to the fact that Gaussian processes are uniquely defined through their mean and covariance functions.

(ii) sup Wald-test: Conditional homoskedasticity implies that $H_i^{\epsilon}(\gamma) = \sigma_\epsilon^2 M_i(\gamma)$, $H_i^u(\gamma) = (\theta_z^{0\top}\Sigma_u\theta_z^0)\ M_i(\gamma)$, and $H_i^{\epsilon,u}(\gamma) = (\Sigma_{\epsilon,u}^\top\theta_z^0)\ M_i(\gamma)$. Plugging these results into the expression for $V(\gamma)$ in Definition A.2 and simplifying yields the asymptotic distribution of the sup-Wald test for the overidentified case.

(iii) To show the asymptotic equivalence for $p = q$, define $\Delta_\sigma = \sigma^2 - \sigma_\epsilon^2$. Then:

$$\begin{aligned}
\tilde{V}(\gamma) &= \tilde{V}_1(\gamma) + \tilde{V}_2(\gamma) - \tilde{V}_{12}(\gamma) - \tilde{V}_{12}^\top(\gamma) \\
&= \sigma^2 C_1^{-1}(\gamma) - \Delta_\sigma C_1^{-1}(\gamma)A^0 R_1(\gamma)M_1(\gamma)A^{0\top}C_1^{-1}(\gamma) \\
&\quad + \sigma^2 C_2^{-1}(\gamma) - \Delta_\sigma C_2^{-1}(\gamma)A^0 R_2(\gamma)M_2(\gamma)A^{0\top}C_2^{-1}(\gamma) \\
&\quad + \Delta_\sigma C_1^{-1}(\gamma)A^0 R_1(\gamma)M_2(\gamma)A^{0\top}C_2^{-1}(\gamma) \\
&\quad + \Delta_\sigma C_2^{-1}(\gamma)A^0 R_2(\gamma)M_1(\gamma)A^{0\top}C_1^{-1}(\gamma) \\
&= \sigma^2(C_1^{-1}(\gamma) + C_2^{-1}(\gamma)) \\
&\quad + \Delta_\sigma C_1^{-1}(\gamma)A^0 R_1(\gamma)[M_2(\gamma)A^{0\top}C_2^{-1}(\gamma) - M_1(\gamma)A^{0\top}C_1^{-1}(\gamma)] \\
&\quad + \Delta_\sigma C_2^{-1}(\gamma)A^0 R_2(\gamma)[M_1(\gamma)A^{0\top}C_1^{-1}(\gamma) - M_2(\gamma)A^{0\top}C_2^{-1}(\gamma)].
\end{aligned} \tag{B.43}$$

In general, $A^0 \in \mathbb{R}^{p\times q}$. Thus, for the just-identified case, i.e. whenever $p = q$, $A^0 \in \mathbb{R}^{p\times p}$. Moreover, since $\Pi^0 \in \mathbb{R}^{q\times p_1}$, $q \geq p_1$, is of full (column) rank by Assumption A.1.6, A^0 is also of full rank and thus, invertible. Denote by $A^{0^{-1}}$ the inverse of A^0. Hence, it follows that $(A^0 M_i(\gamma)A^{0\top})^{-1} = A^{0\top^{-1}}M_i^{-1}(\gamma)A^{0^{-1}}$. Therefore,

$$M_2(\gamma)A^{0\top}C_2^{-1}(\gamma) - M_1(\gamma)A^{0\top}C_1^{-1}(\gamma) = 0. \tag{B.44}$$

By equations (B.43)-(B.44), $\tilde{V}(\gamma) = \sigma^2(C_1^{-1}(\gamma) + C_2^{-1}(\gamma))$. Finally,

$$\tilde{V}^{-1}(\gamma) = \frac{(C_1^{-1}(\gamma) + C_2^{-1}(\gamma))^{-1}}{\sigma^2} = \frac{C_1(\gamma)C^{-1}C_2(\gamma)}{\sigma^2}$$

which yields the asymptotic equivalence of both, sup-LR and sup-Wald tests in the just-identified case under conditional homoskedasticity.

Note that in this setting, $C_1^{-1}(\gamma)A^0 M_1(\gamma)M^{-1} = (A^{0\top})^{-1}M^{-1}$, which implies that:

$$\begin{aligned}
\tilde{\mathcal{E}}(\gamma) &= C_1^{-1}(\gamma)A^0[\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)\Sigma^{1/2}\tilde{\theta}_z^0 - M_1(\gamma)M^{-1}\tilde{\mathcal{GP}}_{\text{mat}}\Sigma^{1/2}\check{\theta}_z^0] \\
&\quad - C_2^{-1}(\gamma)A^0[\tilde{\mathcal{GP}}_{\text{mat},2}(\gamma)\Sigma^{1/2}\tilde{\theta}_z^0 - M_2(\gamma)M^{-1}\tilde{\mathcal{GP}}_{\text{mat}}\Sigma^{1/2}\check{\theta}_z^0] \\
&= C_1^{-1}(\gamma)A^0\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)\Sigma^{1/2}\tilde{\theta}_z^0 - C_2^{-1}(\gamma)A^0\tilde{\mathcal{GP}}_{\text{mat},2}(\gamma)\Sigma^{1/2}\tilde{\theta}_z^0 \\
&= (A^{0\top})^{-1}[M_1^{-1}(\gamma)\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma) - M_2^{-1}(\gamma)\tilde{\mathcal{GP}}_{\text{mat},2}(\gamma)]\Sigma^{1/2}\tilde{\theta}_z^0.
\end{aligned}$$

Also,

$$\begin{aligned}
C_2(\gamma)C^{-1}C_1(\gamma) &= A^0 M_2(\gamma)A^{0\top}(A^{0\top})^{-1}M(A^0)^{-1}A^0 M_2(\gamma)A^{0\top} \\
&= A^0 M_2(\gamma)M^{-1}M_1(\gamma)A^{0\top}.
\end{aligned}$$

Therefore, the asymptotic distribution under conditional homoskedasticity and just-identification is:

$$\begin{aligned}J_1(\gamma) = \frac{1}{\sigma^2}(\Sigma^{1/2}\tilde{\theta}_z^0)^\top &[M_1^{-1}(\gamma)\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma) - M_2^{-1}(\gamma)\tilde{\mathcal{GP}}_{\text{mat},2}(\gamma)]^\top \\ &\times [M_2(\gamma)M^{-1}M_1(\gamma)] \\ &\times [M_1^{-1}(\gamma)\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma) - M_2^{-1}(\gamma)\tilde{\mathcal{GP}}_{\text{mat},2}(\gamma)]\Sigma^{1/2}\tilde{\theta}_z^0.\end{aligned}$$

□

Proof of Corollary 1. First, by Assumption A.1.4 , $\text{Prob}(q_t \leq \gamma)$ is continuous in γ. We will replace the sup over the threshold parameter γ by sup over an equivalent value, $\text{Prob}(q_t \leq \gamma) = \lambda$. To see how this works, note first that $\Gamma \subset \Gamma^0$. Then, $\text{Prob}(q_t \leq \gamma_{\min}) = 0$ and $\text{Prob}(q_t \leq \gamma_{\max}) = 1$ in the sample. Suppose now, that Γ can be defined in terms of a cut-off value, say the κ-th quantile, i.e. $\Gamma = [\gamma_\kappa, \gamma_{1-\kappa}]$. Then equivalently, we have $\text{Prob}(q_t \leq \gamma) = \lambda$ for all $\gamma \in \Gamma$ where λ is uniformly distributed on $\Lambda_\kappa = (\kappa; 1-\kappa)$, i.e $\lambda \sim U(\Lambda_\kappa)$.
Now, by Assumption A.3, we have that

$$M_1(\gamma_1 \wedge \gamma_2) = \mathbb{E}[x_t x_t^\top \mathbb{1}_{\{q_t \leq \gamma_1 \wedge \gamma_2\}}] = \mathbb{E}[x_t x_t^\top]\mathbb{E}[\mathbb{1}_{\{q_t \leq \gamma_1 \wedge \gamma_2\}}] = \min\{\lambda_1, \lambda_2\}M. \tag{B.45}$$

This also implies that

$$M_1(\gamma) = \lambda M \tag{B.46a}$$
$$C_1(\gamma) = A^0 M_1(\gamma) A^{0\top} = \lambda A^0 M A^{0\top} = \lambda C \tag{B.46b}$$
$$M_2(\gamma) = (1-\lambda)M \tag{B.46c}$$
$$C_2(\gamma) = A^0 M_2(\gamma) A^{0\top} = (1-\lambda)A^0 M A^{0\top} = (1-\lambda)C. \tag{B.46d}$$

Therefore,

$$\begin{aligned}\tilde{V}_{12}(\gamma) &= -C_1^{-1}(\gamma)\Delta_\sigma A^0 M_1(\gamma) M^{-1} M_2(\gamma) A^{0\top} C_2^{-1}(\gamma) \\ &= -\Delta_\sigma \lambda^{-1} C^{-1} \lambda(1-\lambda) C (1-\lambda)^{-1} C^{-1} \\ &= -\Delta_\sigma C^{-1} \\ \tilde{V}_1(\gamma) &= \lambda^{-1} C^{-1}[\sigma^2 \lambda C - \Delta_\sigma \lambda^2 C]\lambda^{-1} C^{-1} \\ &= \sigma^2 \lambda^{-1} C^{-1} - \Delta_\sigma C^{-1} \\ \tilde{V}_2(\gamma) &= \sigma^2 (1-\lambda)^{-1} C^{-1} - \Delta_\sigma C^{-1} \\ \tilde{V}(\gamma) &= \tilde{V}_1(\gamma) + \tilde{V}_2(\gamma) - \tilde{V}_{12}(\gamma) - \tilde{V}_{12}^\top(\gamma) \\ &= \sigma^2 \lambda^{-1} C^{-1} + \sigma^2 (1-\lambda)^{-1} C^{-1} \\ &= \sigma^2 \frac{C^{-1}}{\lambda(1-\lambda)},\end{aligned}$$

implying that

$$\sup_{\gamma \in \Gamma} W_T^{2SLS}(\gamma), \sup_{\gamma \in \Gamma} LR_T^{2SLS}(\gamma) \Rightarrow \sup_{\lambda \in \Lambda_\kappa} \frac{\lambda(1-\lambda)}{\sigma^2} \tilde{\mathcal{E}}^\top(\gamma) C \tilde{\mathcal{E}}(\gamma)$$

Hence, in this situation, the sup Wald and sup LR-test are asymptotically equivalent, no matter whether the system is just- or overidentified.
Next, (B.45) implies that – under Assumptions A.2 and A.3 – the Gaussian process $\mathcal{GP}_1(\gamma)$ can be restated as

$$\begin{aligned} \mathcal{GP}_1(\gamma) &= (\Sigma^{1/2} \otimes I_q)\tilde{\mathcal{GP}}_1(\gamma) = (\Sigma^{1/2} \otimes M^{1/2})\mathcal{BM}_{q(p_1+1)}(\lambda) \\ \iff \mathcal{GP}_{\text{mat},1}(\gamma) &= M^{1/2}\mathcal{BM}_{\text{mat},q(p_1+1)}(\lambda)\Sigma^{1/2} \end{aligned} \tag{B.47}$$

where $\mathcal{BM}_{q(p_1+1)}$ denotes a $q(p_1+1)\times 1$ vector of independent Brownian motions on the unit interval, and $\mathcal{BM}_{\text{mat},q(p_1+1)}(\lambda)$ is the $q\times(p_1+1)$ matrix with $\text{vec}(\mathcal{BM}_{\text{mat},q(p_1+1)}(\lambda)) = \mathcal{BM}_{q(p_1+1)}(\lambda)$. Equation (B.47) in turn implies that $\mathcal{B}_1(\gamma)$ can be rewritten as $\mathcal{B}_1(\lambda)$. Therefore, we obtain

$$\begin{aligned} \tilde{\mathcal{E}}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\tilde{\mathcal{E}}(\gamma) &= [C_1^{-1}(\gamma)\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)]^\top \\ &\times C_2(\gamma)C^{-1}C_1(\gamma) \\ &\times [C_1^{-1}(\gamma)\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)] \\ &= \frac{1}{\lambda(1-\lambda)}[C^{-1}\mathcal{B}_1(\lambda) - \lambda C^{-1}\mathcal{B}_1(1)]^\top \\ &\times C[C^{-1}\mathcal{B}_1(\lambda) - \lambda C^{-1}\mathcal{B}_1(1)] \\ &= \frac{1}{\lambda(1-\lambda)}[C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1)]^\top \\ &\times [C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1)]. \end{aligned} \tag{B.48}$$

Next, we show that the term $C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1) \overset{\mathcal{D}}{=} [(\Sigma^{1/2}\tilde{\theta}_z^0)^\top \otimes I_p][\mathcal{BM}_{p(p_1+1)}(\lambda) - \lambda\mathcal{BM}_{p(p_1+1)}(1)]$, where $\mathcal{BM}_{p(p_1+1)}(\lambda)$ collects in a vector the first p out of each q block of elements of $\mathcal{BM}_{q(p_1+1)}(\lambda)$. Because of (B.46a) and (B.47) it follows that

$$\begin{aligned} \mathcal{B}_1(\lambda) &= A^0[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma)M^{-1}\mathcal{GP}_{\text{mat},1}\check{\theta}_z^0] \\ &= A^0M^{1/2}[\mathcal{BM}_{\text{mat},q(p_1+1)}(\lambda)\Sigma^{1/2}\tilde{\theta}_z^0 - \lambda\mathcal{BM}_{\text{mat},q(p_1+1)}(1)\Sigma^{1/2}\check{\theta}_z^0]. \end{aligned}$$

Furthermore, recall that $C = A^0MA^{0\top}$. Thus:

$$\begin{aligned} C^{-1/2}\mathcal{B}_1(\lambda) &= (A^0MA^{0\top})^{-1/2}A^0M^{1/2}\mathcal{BM}_{\text{mat},q(p_1+1)}(\lambda)\Sigma^{1/2}\tilde{\theta}_z^0 \\ &- \lambda(A^0MA^{0\top})^{-1/2}A^0M^{1/2}\mathcal{BM}_{\text{mat},q(p_1+1)}(1)\Sigma^{1/2}\check{\theta}_z^0. \end{aligned}$$

Note that because $(A^0MA^{0\top})^{-1/2}(A^0MA^{0\top})(A^0MA^{0\top})^{-1/2}$ is a $p\times p$ projection matrix, pre-multiplying with $(A^0MA^{0\top})^{-1/2}A^0M^{1/2}$ is without loss of generality equal in distribution to selecting the first p rows of the q rows of $BM_{\text{mat},q(p_1+1)}(\lambda)$ (this can be seen by writing down the eigenvalue decomposition of the projection matrix as in Hall et al. (2012), supplemental appendix, page 22-23), yielding

$$\begin{aligned} \mathcal{BM}_{\text{mat},p(p_1+1)}(\lambda) &\overset{\mathcal{D}}{=} (A^0MA^{0\top})^{-1/2}A^0M^{1/2}\mathcal{BM}_{\text{mat},q(p_1+1)}(\lambda) \\ \mathcal{BM}_{\text{mat},p(p_1+1)}(1) &\overset{\mathcal{D}}{=} (A^0MA^{0\top})^{-1/2}A^0M^{1/2}\mathcal{BM}_{\text{mat},q(p_1+1)}(1), \end{aligned}$$

where $\mathcal{BM}_{p(p_1+1)}(\lambda) = vec(\mathcal{BM}_{\text{mat},p(p_1+1)}(\lambda))$. From the last statement, using the fact that for generic matrices A, B, we have $vec(AB) = (B' \otimes I)vec(A)$,

$$\begin{aligned}
C^{-1/2}\mathcal{B}_1(\lambda) &\overset{\mathcal{D}}{=} \mathcal{BM}_{\text{mat},p(p_1+1)}(\lambda)\Sigma^{1/2}\tilde{\theta}_z^0 - \lambda\mathcal{BM}_{\text{mat},p(p_1+1)}(1)\Sigma^{1/2}\check{\theta}_z^0 \\
\lambda C^{-1/2}\mathcal{B}_1(1) &\overset{\mathcal{D}}{=} \lambda\mathcal{BM}_{\text{mat},p(p_1+1)}(1)\Sigma^{1/2}\tilde{\theta}_z^0 - \lambda\mathcal{BM}_{\text{mat},p(p_1+1)}(1)\Sigma^{1/2}\check{\theta}_z^0 \\
C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1) &\overset{\mathcal{D}}{=} \mathcal{BM}_{\text{mat},p(p_1+1)}(\lambda)\Sigma^{1/2}\tilde{\theta}_z^0 - \lambda\mathcal{BM}_{\text{mat},p(p_1+1)}(1)\Sigma^{1/2}\tilde{\theta}_z^0 \\
&= [(\Sigma^{1/2}\tilde{\theta}_z^0)^\top \otimes I_p]\,[\mathcal{BM}_{p(p_1+1)}(\lambda) - \lambda\mathcal{BM}_{p(p_1+1)}(1)] \\
&\equiv [(\Sigma^{1/2}\tilde{\theta}_z^0)^\top \otimes I_p]\mathcal{BB}_{p(p_1+1)}(\lambda). \qquad \text{(B.49)}
\end{aligned}$$

Using (B.49),

$$\begin{aligned}
\frac{\mathcal{E}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\mathcal{E}(\gamma)}{\sigma^2} &\overset{\mathcal{D}}{=} \frac{\left\{[(\Sigma^{1/2}\tilde{\theta}_z^0)^\top \otimes I_p]\mathcal{BB}_{p(p_1+1)}(\lambda)\right\}^\top \left\{[(\Sigma^{1/2}\tilde{\theta}_z^0)^\top \otimes I_p]\mathcal{BB}_{p(p_1+1)}(\lambda)\right\}}{\lambda(1-\lambda)(\Sigma^{1/2}\tilde{\theta}_z^0)^\top(\Sigma^{1/2}\tilde{\theta}_z^0)} \\
&= \frac{\mathcal{BB}_{p(p_1+1)}^\top \left\{[(\Sigma^{1/2}\tilde{\theta}_z^0)[(\Sigma^{1/2}\tilde{\theta}_z^0)^\top(\Sigma^{1/2}\tilde{\theta}_z^0)]^{-1}(\Sigma^{1/2}\tilde{\theta}_z^0)^\top] \otimes I_p\right\} \mathcal{BB}_{p(p_1+1)}}{\lambda(1-\lambda)}.
\end{aligned}$$

Since $F = (\Sigma^{1/2}\tilde{\theta}_z^0)[(\Sigma^{1/2}\tilde{\theta}_z^0)^\top(\Sigma^{1/2}\tilde{\theta}_z^0)]^{-1}(\Sigma^{1/2}\tilde{\theta}_z^0)^\top$ is a projection matrix, as before, pre-multiplying with $F \otimes I_p$ involves, without loss of generality, selecting the first p elements of $\mathcal{BB}_{p(p_1+1)}$, yielding $\mathcal{BB}_p(\lambda)$. Therefore,

$$\frac{\tilde{\mathcal{E}}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\tilde{\mathcal{E}}(\gamma)}{\sigma^2} \overset{\mathcal{D}}{=} \frac{\mathcal{BB}_p^\top(\lambda)\mathcal{BB}_p(\lambda)}{\lambda(1-\lambda)},$$

proving the claim. □

Proofs for Section 4.4: 2SLS tests and a TRF

Lemma 5. *Under Assumption A.1, $T(\hat{\rho} - \rho^0) = \mathcal{O}_p(1)$, $T^{1/2}(\hat{\Pi}_i - \Pi_i^0) = \mathcal{O}_p(1)$, $i = 1, 2$ and it holds that the distribution is as if ρ^0 was known:*

$$T^{1/2}\,\text{vec}(\hat{\Pi}_i(\rho^0) - \Pi_i^0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_i),$$

where $S_1 = (I_{p_1} \otimes M_1^{-1}(\rho^0))\mathbb{E}[(u_t u_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \rho^0\}}](I_{p_1} \otimes M_1^{-1}(\rho^0))$ and $S_2 = (I_{p_1} \otimes M_2^{-1}(\rho^0))\mathbb{E}[(u_t u_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t > \rho^0\}}](I_{p_1} \otimes M_2^{-1}(\rho^0))$.

Proof of Lemma 5. The results $T(\hat{\rho} - \rho^0) = \mathcal{O}_p(1)$, $T^{1/2}(\hat{\Pi}_i - \Pi_i^0) = \mathcal{O}_p(1)$, $i = 1, 2$ directly follow from Caner and Hansen (2004), Theorems 1 and 2. We will prove the statement for $T^{1/2}\,\text{vec}(\hat{\Pi}_1(\rho^0) - \Pi_1^0)$. The proof for $T^{1/2}\,\text{vec}(\hat{\Pi}_2(\rho^0) - \Pi_2^0)$ is similar and omitted for brevity.

By construction

$$\begin{aligned}\hat{\Pi}_1(\rho^0) &= (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} Z)\\ &= (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} X_1^{\rho^0}\Pi_1^0 + X_1^{\rho^0\top} X_2^{\rho^0}\Pi_2^0 + X_1^{\rho^0\top} u)\\ &= \Pi_1^0 + (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} u_1^{\rho^0})\end{aligned}$$

where the last equality holds because $X_1^{\rho^0\top} X_2^{\rho^0} = \mathbf{0}$. Hence,

$$\begin{aligned}T^{1/2}\operatorname{vec}(\hat{\Pi}_1(\rho^0) - \Pi_1^0) &= \operatorname{vec}\Big((T^{-1}X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(T^{-1/2}X_1^{\rho^0\top} u)\Big)\\ &= (I_{p_1}\otimes (T^{-1}X_1^{\rho^0\top} X_1^{\rho^0})^{-1})\operatorname{vec}(T^{1/2}(X_1^{\rho^0\top} u_1^{\rho^0})).\end{aligned}$$

Next, $(T^{-1}X_1^{\rho^0\top} X_1^{\rho^0})^{-1} \xrightarrow{p} M_1^{-1}(\rho^0)$ and by Lemma 1

$$T^{1/2}\operatorname{vec}(X_1^{\rho\top} u_1^{\rho}) \Rightarrow \mathcal{GP}_1(\rho).$$

Note that $\mathcal{GP}_1(\rho)$ is a zero-mean Gaussian process with covariance function $\mathcal{C}_{\mathcal{GP}}(\rho_1,\rho_2) = \mathbb{E}[(u_t u_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t\le\rho_1\wedge\rho_2\}}]$. Therefore,

$$T^{1/2}\operatorname{vec}(\hat{\Pi}_1(\rho^0) - \Pi_1^0) \Rightarrow (I_{p_1}\otimes M_1^{-1}(\rho^0))\mathcal{GP}_1(\rho^0).$$

Because $\mathcal{GP}_1(\rho^0)$ denotes the Gaussian process at a particular value ρ^0 it follows that $\mathcal{GP}_1(\rho^0) \sim \mathcal{N}(0, \mathbb{E}[u_t u_t^\top \otimes x_t x_t^\top \mathbb{1}_{\{q_t\le\rho^0\}}])$ and therefore,

$$T^{1/2}\operatorname{vec}(\hat{\Pi}_1(\rho^0) - \Pi_1^0) \xrightarrow{\mathcal{D}} (I_{p_1}\otimes M_1^{-1}(\rho^0))\mathcal{N}(0, \mathbb{E}[u_t u_t^\top \otimes x_t x_t^\top \mathbb{1}_{\{q_t\le\rho^0\}}]),$$

which concludes the proof. □

Lemma 6. *Suppose Assumption A.1 holds. Then, under $\mathbb{H}_0$,*

$$T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} \xrightarrow{p} A_1^0 M_1(\gamma\wedge\rho^0)A_1^{0\top} + A_2^0(M_1(\gamma) - M_1(\gamma\wedge\rho^0))A_2^{0\top} = C_{A,1}(\gamma)$$

and

$$\begin{aligned}T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Rightarrow\ & A_1^0\Big[\mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0 - R_1(\gamma\wedge\rho^0;\rho^0)\mathcal{GP}_{\mathrm{mat},1}(\rho^0)\check{\theta}_z^0\Big]\\ &+ A_2^0\Big[(\mathcal{GP}_{\mathrm{mat},1}(\gamma) - \mathcal{GP}_{\mathrm{mat},1}(\gamma\wedge\rho^0))\tilde{\theta}_z^0\\ &\qquad - (R_2(\gamma\wedge\rho^0;\rho^0) - R_2(\gamma;\rho^0))\mathcal{GP}_{\mathrm{mat},2}(\rho^0)\check{\theta}_z^0\Big]\\ &= \mathcal{B}_{A,1}(\gamma)\end{aligned}$$

Proof of Lemma 6. This proof is done in two parts: First, we show the asymptotic behavior of $T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma}$ and afterwards the asymptotic behavior of $T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma}$. Also, it will be helpful during the proofs to consider three cases: Case (a) assumes that $\gamma < \rho^0$, Case (b) that $\gamma = \rho^0$ and Case (c) that $\gamma > \rho^0$. There are two sub-cases within each case:

- In case (a) it follows that $\gamma < \hat{\rho}$ because $\hat{\rho} = \rho^0 + o_p(1)$ by Lemma 2 and $\gamma - \rho^0$ is a fixed strictly negative number by construction. This implies two sub-cases: $(a.1)$ with $\gamma < \hat{\rho} \leq \rho^0$ and $(a.2)$ with $\gamma < \rho^0 < \hat{\rho}$.
- In case (b) there are two sub-cases: $(b.1)$ with $\gamma = \rho^0 \leq \hat{\rho}$ and $(b.2)$ with $\hat{\rho} < \gamma = \rho^0$
- In case (c) it follows that $\gamma > \hat{\rho}$ because $\hat{\rho} = \rho^0 + o_p(1)$ by Lemma 2 and $\gamma - \rho^0$ is a fixed strictly positive number by construction. This implies two sub-cases: $(c.1)$ with $\hat{\rho} \leq \rho^0 < \gamma$ and $(c.2)$ with $\rho^0 < \hat{\rho} < \gamma$.

Claim (i)**.** Starting with case (a), because $\gamma < \hat{\rho}$ for both possible sub-cases, it holds uniformly in γ that

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= \hat{A}_1(T^{-1}X_1^{\gamma\top}X_1^{\gamma})\hat{A}_1^{\top} \\ &= A_1^0(T^{-1}X_1^{\gamma\top}X_1^{\gamma})A_1^{0\top} + o_p(1) \\ &\xrightarrow{p} A_1^0 M_1(\gamma) A_1^{0\top} \end{aligned} \tag{B.50}$$

by Lemma 2.
In case (b), we first consider sub-case $(b.1)$. Because $\gamma \leq \hat{\rho}$, it holds uniformly in γ that

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= \hat{A}_1(T^{-1}X_1^{\gamma\top}X_1^{\gamma})\hat{A}_1^{\top} \\ &= A_1^0(T^{-1}X_1^{\gamma\top}X_1^{\gamma})A_1^{0\top} + o_p(1) \\ &\xrightarrow{p} A_1^0 M_1(\gamma) A_1^{0\top} \end{aligned} \tag{B.51}$$

by Lemma 2. In sub-case $(b.2)$ it follows that

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= T^{-1}\hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}} + T^{-1}(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} - \hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}}) \\ &= \hat{A}_1(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_1^{\top} + \hat{A}_2(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} - T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_2^{\top}, \end{aligned} \tag{B.52}$$

because $\hat{\rho} < \gamma = \rho^0$. By Lemma 5 we have that $\hat{\rho} = \rho^0 + \mathcal{O}_p(T^{-1})$ and therefore,

$$\begin{aligned} T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} &= T^{-1}\sum_{t=1}^{T} x_t x_t^{\top}\mathbb{1}_{\{q_t \leq \hat{\rho}\}} \\ &= T^{-1}\sum_{t=1}^{T} x_t x_t^{\top}\mathbb{1}_{\{q_t \leq \rho^0\}} + T^{-1}\sum_{t=1}^{T} x_t x_t^{\top}(\mathbb{1}_{\{q_t \leq \hat{\rho}\}} - \mathbb{1}_{\{q_t \leq \rho^0\}}) \\ &= T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} + \mathcal{O}_p(T^{-1}) \\ &= T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} + o_p(1). \end{aligned} \tag{B.53}$$

So, (B.52), (B.53) and Lemma 2 imply, uniformly in γ,

$$T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} \xrightarrow{p} A_1^0 M_1(\rho^0) A_1^{0\top} = A_1^0 M_1(\gamma) A_1^{0\top}. \tag{B.54}$$

Last, we consider case (c). In sub-case $(c.1)$ we have uniformly in γ that

$$\begin{aligned}
T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= T^{-1}\hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}} + T^{-1}(\hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0} - \hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}}) \\
&\quad + T^{-1}(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} - \hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0}) \\
&= \hat{A}_1(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_1^\top + \hat{A}_2(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} - T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_2^\top \\
&\quad + \hat{A}_2(T^{-1}X_1^{\gamma\top}X_1^{\gamma} - T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})\hat{A}_2^\top \\
&\xrightarrow{p} A_1^0 M_1(\rho^0)A_1^{0\top} + A_2^0(M_1(\gamma) - M_1(\rho^0))A_2^{0\top} \qquad \text{(B.55)}
\end{aligned}$$

by Lemma 2. In sub-case $(c.2)$ it follows uniformly in γ that

$$\begin{aligned}
T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= T^{-1}\hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0} + T^{-1}(\hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}} - \hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0}) \\
&\quad + T^{-1}(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} - \hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}}) \\
&= \hat{A}_1(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})\hat{A}_1^\top + \hat{A}_1(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} - T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})\hat{A}_1^\top \\
&\quad + \hat{A}_2(T^{-1}X_1^{\gamma\top}X_1^{\gamma} - T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_2^\top \\
&\xrightarrow{p} A_1^0 M_1(\rho^0)A_1^{0\top} + A_2^0(M_1(\gamma) - M_1(\rho^0))A_2^{0\top}. \qquad \text{(B.56)}
\end{aligned}$$

Finally, putting results (B.50), (B.51), (B.54)–(B.56) together yields the claim.

Claim (ii)**.** To show this claim, we present a full proof for case (a). Since cases (b) and (c) follow similar reasoning we only state the most important intermediate results to conclude the claim.

Starting with sub-case $(a.1)$ of (a) it holds that

$$\begin{aligned}
T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1(T^{-1/2}X_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma}) \\
&= \hat{A}_1(T^{-1/2}X_1^{\gamma\top}(\epsilon_1^{\gamma} + (Z_1^{\gamma} - \hat{Z}_1^{\gamma})\theta_z^0) \\
&= \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}(\epsilon_1^{\gamma} + (X_1^{\gamma}\Pi_1^0 + u_1^{\gamma} - X_1^{\gamma}\hat{\Pi}_1)\theta_z^0)\right] \\
&= \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} - (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right], \qquad \text{(B.57)}
\end{aligned}$$

By Lemma 1 it follows that $T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} \Rightarrow \mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0$ uniformly in γ where $\mathrm{vec}(\mathcal{GP}_{\mathrm{mat},1}(\gamma)) = \mathcal{GP}_1(\gamma)$ with $\mathcal{GP}_1(\gamma)$ as in Lemma 1 and $\tilde{\theta}_z^0 = (1, \theta_z^{0\top})^\top$. Moreover, uniformly in γ

$$\begin{aligned}
(T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0 &= (T^{-1}X_1^{\gamma\top}X_1^{\gamma})(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})^{-1}(T^{-1/2}X_1^{\hat{\rho}\top}u_1^{\hat{\rho}})\theta_z^0 \\
&\Rightarrow M_1(\gamma)M_1^{-1}(\rho^0)\mathcal{GP}_{\mathrm{mat},1}(\rho^0)\check{\theta}_z^0
\end{aligned}$$

Therefore, (B.57) behaves uniformly in γ as

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Rightarrow A_1^0\left[\mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0 - R_1(\gamma;\rho^0)\mathcal{GP}_{\mathrm{mat},1}(\rho^0)\check{\theta}_z^0\right]. \qquad \text{(B.58)}$$

As in sub-case $(a.1)$, for sub-case $(a.2)$ it follows that

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} = \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} - (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right]. \qquad \text{(B.59)}$$

We now have[21]

$$\hat{\Pi}_1 - \Pi_1^0 = (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} u_1^{\rho^0}) + o_p(1)$$

because

$$\begin{aligned}\hat{\Pi}_1 &= (X_1^{\hat{\rho}\top} X_1^{\hat{\rho}})^{-1}(X_1^{\hat{\rho}\top} Z_1^{\hat{\rho}}) \\ &= (X_1^{\hat{\rho}\top} X_1^{\hat{\rho}})^{-1}(X_1^{\rho^0\top} X_1^{\rho^0}\Pi_1^0 + X_1^{\rho^0\top} X_1^{\rho^0}\Pi_2^0 - X_1^{\hat{\rho}\top} X_1^{\hat{\rho}}\Pi_2^0 + X_1^{\hat{\rho}\top} u_1^{\hat{\rho}}) \\ &= \Pi_1^0 + (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} u_1^{\rho^0}) + o_p(1) \end{aligned} \tag{B.60}$$

by Lemma 5. So, putting (B.59) and (B.60) together yields uniformly in γ that

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Rightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - R_1(\gamma;\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\breve{\theta}_z^0\right]. \tag{B.61}$$

For case (b), sub-case $(b.1)$, it follows, as for sub-case $(a.2)$, uniformly in γ that

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} = \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} - (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right]$$

with

$$\hat{\Pi}_1 - \Pi_1^0 = (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} u_1^{\rho^0}) + o_p(1).$$

So, as for sub-case $(a.2)$, uniformly in γ

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Rightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - R_1(\gamma;\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\breve{\theta}_z^0\right], \tag{B.62}$$

where $R_1(\gamma;\rho^0) = I_q$ whenever $\gamma = \rho^0$.
For sub-case $(b.2)$ it holds uniformly in γ that

$$\begin{aligned} T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1\left[T^{-1/2}X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - (T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \\ &\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\rho^0\top}s_1^{\rho^0} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) - T^{-1}(X_1^{\rho^0\top}X_1^{\rho^0} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\ &\Rightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\gamma)\breve{\theta}_z^0\right] \end{aligned} \tag{B.63}$$

by Lemmata 1, 2 and Equation (B.53).

[21]Note that in sub-case $(a.1)$ we could also write $\hat{\Pi}_1 - \Pi_1^0 = (X_1^{\rho^0\top} X_1^{\rho^0})^{-1}(X_1^{\rho^0\top} u_1^{\rho^0}) + o_p(1)$. However, the composition of the $o_p(1)$-term is different in both cases, as illustrated in (B.60). E.g. in (B.60) also $X_1^{\rho^0\top} X_1^{\rho^0}\Pi_2^0 - X_1^{\hat{\rho}\top} X_1^{\hat{\rho}}\Pi_2^0$ is included in the $o_p(1)$-term, whereas in $(a.1)$ this term completely vanishes already in samples (rather than only asymptotically) because of the relative locations of γ, ρ^0 and $\hat{\rho}$.

Last, we show the claim for case (c). In sub-case $(c.1)$ it holds uniformly in γ that

$$
\begin{aligned}
T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1\left[T^{-1/2}X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - (T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \\
&\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\rho^0\top}s_1^{\rho^0} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) - T^{-1}(X_1^{\rho^0\top}X_1^{\rho^0} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\gamma\top}s_1^{\gamma} - X_1^{\rho^0\top}s_1^{\rho^0}) - T^{-1}(X_1^{\gamma\top}X_1^{\gamma} - X_1^{\rho^0\top}X_1^{\rho^0})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\quad \Rightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right] \\
&\quad + A_2^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - (I_q - R_2(\gamma;\rho^0))\mathcal{GP}_{\text{mat},2}(\rho^0)\check{\theta}_z^0\right],
\end{aligned} \tag{B.64}
$$

where the middle term drops because $T^{-1/2}(X_1^{\rho^0\top}s_1^{\rho^0} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) = o_p(1)$, $T^{-1}(X_1^{\rho^0\top}X_1^{\rho^0} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}}) = o_p(1)$ and $T^{1/2}(\hat{\Pi}_2 - \Pi_2^0) = \mathcal{O}_p(1)$ by Lemma 5.
Last, sub-case $(c.2)$ yields uniformly in γ

$$
\begin{aligned}
T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1\left[T^{-1/2}X_1^{\rho^0\top}s_1^{\rho^0} - (T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \\
&\quad + \hat{A}_1\left[T^{-1/2}(X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - X_1^{\rho^0\top}s_1^{\rho^0}) - T^{-1}(X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} - X_1^{\rho^0\top}X_1^{\rho^0})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\gamma\top}s_1^{\gamma} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) - T^{-1}(X_1^{\gamma\top}X_1^{\gamma} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\quad \Rightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right] \\
&\quad + A_2^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - (I_q - R_2(\gamma;\rho^0))\mathcal{GP}_{\text{mat},2}(\rho^0)\check{\theta}_z^0\right],
\end{aligned} \tag{B.65}
$$

where the middle term drops because $T^{-1/2}(X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - X_1^{\rho^0\top}s_1^{\rho^0}) = o_p(1)$, $T^{-1}(X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} - X_1^{\rho^0\top}X_1^{\rho^0}) = o_p(1)$ and $T^{1/2}(\hat{\Pi}_2 - \Pi_2^0) = \mathcal{O}_p(1)$ by Lemma 5.
Finally, putting (B.58), (B.61)–(B.65) together immediately yields the claim. □

Lemma 7. *Suppose Assumption A.1 holds and define* $\hat{\theta}^{\gamma} = \text{vec}(\hat{\theta}_1^{\gamma}, \hat{\theta}_2^{\gamma})$, *and* $\bar{\theta}^0 = \text{vec}(\theta^0, \theta^0)$. *Then, under* $\mathbb{H}_0$ *and for a fixed* γ,

$$
T^{1/2}(\hat{\theta}^{\gamma} - \bar{\theta}^0) \Rightarrow \mathcal{N}(0, \Sigma_A^{\gamma})
$$

with

$$
\Sigma_A^{\gamma} = \begin{bmatrix} V_{A,1}(\gamma) & V_{A,12}(\gamma) \\ V_{A,12}^{\top}(\gamma) & V_{A,2}(\gamma) \end{bmatrix}
$$

where $V_{A,1}(\gamma), V_{A,2}(\gamma)$ *and* $V_{A,12}(\gamma)$ *are defined in Definition A.3.*

Proof of Lemma 7. First, we define the following quantities

$$
\bar{W} = \begin{bmatrix} \hat{W}_1^{\gamma} & \mathbf{0} \\ \mathbf{0} & \hat{W}_2^{\gamma} \end{bmatrix}, \quad \bar{Y} = \begin{bmatrix} Y_1^{\gamma} \\ Y_2^{\gamma} \end{bmatrix}, \quad \hat{\theta}^{\gamma} = \begin{bmatrix} \hat{\theta}_1^{\gamma} \\ \hat{\theta}_2^{\gamma} \end{bmatrix}.
$$

With this notation, the 2SLS estimator is

$$\hat{\theta}^{\gamma} = (\bar{W}^{\top}\bar{W})^{-1}\bar{W}^{\top}\bar{Y}$$
$$= \bar{\theta}^{0} + (\bar{W}^{\top}\bar{W})^{-1}\bar{W}^{\top}\tilde{\bar{\epsilon}}$$

where the last equality follows from Equation (A.20) in our paper and

$$\tilde{\bar{\epsilon}} = \begin{bmatrix} \tilde{\epsilon}_1^{\gamma} \\ \tilde{\epsilon}_2^{\gamma} \end{bmatrix} = \begin{bmatrix} \epsilon_1^{\gamma} + (Z-\hat{Z})_1^{\gamma}\theta_z^0 \\ \epsilon_2^{\gamma} + (Z-\hat{Z})_2^{\gamma}\theta_z^0 \end{bmatrix}.$$

Hence, by Lemma 6 it immediately follows that

$$T^{1/2}(\hat{\theta}^{\gamma} - \bar{\theta}^{0}) \Rightarrow \begin{bmatrix} C_{A,1}^{-1}(\gamma)\mathcal{B}_{A,1}(\gamma) \\ C_{A,2}^{-1}(\gamma)\mathcal{B}_{A,2}(\gamma) \end{bmatrix} \sim \mathcal{N}(0, \Sigma_A^{\gamma}).$$

for fixed γ. Thus, we are left to derive Σ_A^{γ}. Start with $V_{A,1}(\gamma)$:

$$\begin{aligned}
\text{Var}[\mathcal{B}_{A,1}(\gamma)] &= \text{Var}[A_1^0\mathcal{G}\mathcal{P}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)\mathcal{G}\mathcal{P}_{\text{mat},1}(\rho^0)\check{\theta}_z^0] \\
&= \text{Var}[(\tilde{\theta}_z^{0\top} \otimes A_1^0)\mathcal{G}\mathcal{P}_1(\gamma)] + \text{Var}[(\check{\theta}_z^{0\top} \otimes [A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)])\mathcal{G}\mathcal{P}_1(\rho^0)] \\
&- \text{Cov}[(\tilde{\theta}_z^{0\top} \otimes A_1^0)\mathcal{G}\mathcal{P}_1(\gamma), (\check{\theta}_z^{0\top} \otimes [A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)])\mathcal{G}\mathcal{P}_1(\rho^0)] \\
&- \text{Cov}[(\check{\theta}_z^{0\top} \otimes [A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)])\mathcal{G}\mathcal{P}_1(\rho^0), (\tilde{\theta}_z^{0\top} \otimes A_1^0)\mathcal{G}\mathcal{P}_1(\gamma)] \\
&= (\tilde{\theta}_z^{0\top} \otimes A_1^0)\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t\le\gamma\}}](\tilde{\theta}_z^{0} \otimes A_1^{0\top}) \\
&+ (\check{\theta}_z^{0\top} \otimes [A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)])\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t\le\rho^0\}}](\check{\theta}_z^{0} \otimes [M_1^{-1}(\rho^0)M_1(\gamma)A_1^{0\top}]) \\
&- (\tilde{\theta}_z^{0\top} \otimes A_1^0)\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t\le\gamma\}}](\check{\theta}_z^{0} \otimes [M_1^{-1}(\rho^0)M_1(\gamma)A_1^{0\top}]) \\
&- (\check{\theta}_z^{0\top} \otimes [A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)])\mathbb{E}[(v_t v_t^{\top} \otimes x_t x_t^{\top})\mathbb{1}_{\{q_t\le\gamma\}}](\tilde{\theta}_z^{0} \otimes A_1^{0\top}) \\
&= A_1^0\mathbb{E}[x_t x_t^{\top}(\epsilon_t + u_t^{\top}\theta_z^0)^2\mathbb{1}_{\{q_t\le\gamma\}}]A_1^{0\top} \\
&+ A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)\mathbb{E}[x_t x_t^{\top}(u_t^{\top}\theta_z^0)^2\mathbb{1}_{\{q_t\le\rho^0\}}]M_1^{-1}(\rho^0)M_1(\gamma)A_1^{0\top} \\
&- A_1^0\mathbb{E}[x_t x_t^{\top}(\epsilon_t u_t^{\top}\theta_z^0 + \theta_z^{0\top}u_t u_t^{\top}\theta_z^0)\mathbb{1}_{\{q_t\le\gamma\}}]M_1^{-1}(\rho^0)M_1(\gamma)A_1^{0\top} \\
&- A_1^0 M_1(\gamma)M_1^{-1}(\rho^0)\mathbb{E}[x_t x_t^{\top}(\epsilon_t u_t^{\top}\theta_z^0 + \theta_z^{0\top}u_t u_t^{\top}\theta_z^0)\mathbb{1}_{\{q_t\le\gamma\}}]A_1^{0\top} \\
&= A_1^0\ [H_1(\gamma) + R_1(\gamma;\rho^0)H_1^u(\rho^0)R_1^{\top}(\gamma;\rho^0) - R_1(\gamma;\rho^0)(H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)) \\
&- (H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma))R_1^{\top}(\gamma;\rho^0)]\ A_1^{0\top}
\end{aligned} \tag{B.66}$$

which yields the claim for $\gamma \le \rho^0$ when pre- and post-multiplied with $C_{A,1}^{-1}(\gamma)$.

Next, we consider $\text{Var}[\mathcal{B}_{A,2}(\gamma)]$. First, note that

$$\text{Var}[\mathcal{B}_{A,2}(\gamma)] = \text{Var}[\mathcal{B}_A] + \text{Var}[\mathcal{B}_{A,1}(\gamma)] - \text{Cov}[\mathcal{B}_A, \mathcal{B}_{A,1}(\gamma)] - \text{Cov}[\mathcal{B}_{A,1}(\gamma), \mathcal{B}_A] \tag{B.67}$$

where $\text{Var}[\mathcal{B}_{A,1}(\gamma)]$ was already derived in Equation (B.66), and $\mathcal{B}_A = \mathcal{B}_A(\gamma_{max}) = A_1^0\mathcal{G}\mathcal{P}_{\text{mat},1}(\rho^0)e_1 + A_2^0\mathcal{G}\mathcal{P}_{\text{mat},2}(\rho^0)e_1$ was defined right before Theorem 3 and $e_1 = \tilde{\theta}_z^0 - \check{\theta}_z^0 = (1, 0, \ldots, 0)^{\top}$. Because

$$\text{Var}[\mathcal{B}_A] = \text{Var}[A_1^0\mathcal{G}\mathcal{P}_{\text{mat},1}(\rho^0)e_1] + \text{Var}[A_2^0\mathcal{G}\mathcal{P}_{\text{mat},2}(\rho^0)e_1] \tag{B.68}$$

where we used the fact that $\text{Cov}[\mathcal{GP}_1(\rho^0), \mathcal{GP}_2(\rho^0)] = \mathbb{E}[\mathcal{GP}_1(\rho^0)\mathcal{GP}_2^\top(\rho^0)] = \mathbb{E}[\mathcal{GP}_1(\rho^0)\mathcal{GP}_1^\top] - \mathbb{E}[\mathcal{GP}_1(\rho^0)\mathcal{GP}_1^\top(\rho^0)] = \mathbf{0}$. Equation (B.68) thus reads as

$$\begin{aligned}\text{Var}[\mathcal{B}_A] &= (e_1^\top \otimes A_1^0)\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \rho^0\}}](e_1 \otimes A_1^{0\top}) \\ &+ (e_1^\top \otimes A_2^0)\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t > \rho^0\}}](e_1 \otimes A_2^{0\top}) \\ &= A_1^0 \mathbb{E}[x_t x_t^\top \epsilon_t^2 \mathbb{1}_{\{q_t \le \rho^0\}}]A_1^{0\top} + A_2^0 \mathbb{E}[x_t x_t^\top \epsilon_t^2 \mathbb{1}_{\{q_t > \rho^0\}}]A_2^{0\top} \\ &= A_1^0 H_1^\epsilon(\rho^0) A_1^{0\top} + A_2^0 H_2^\epsilon(\rho^0) A_2^{0\top}. \end{aligned} \tag{B.69}$$

From (B.67), we still need to derive $\text{Cov}[\mathcal{B}_A, \mathcal{B}_{A,1}(\gamma)]$:

$$\begin{aligned}\text{Cov}[\mathcal{B}_A, \mathcal{B}_{A,1}(\gamma)] &= \text{Cov}[A_1^0 \mathcal{GP}_{\text{mat},1}(\rho^0)e_1 + A_2^0 \mathcal{GP}_{\text{mat},2}(\rho^0)e_1, \\ &\qquad A_1^0 \mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - A_1^0 M_1(\gamma) M_1^{-1}(\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\breve{\theta}_z^0] \\ &= \text{Cov}[A_1^0 \mathcal{GP}_{\text{mat},1}(\rho^0)e_1, A_1^0 \mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0] \\ &- \text{Cov}[A_1^0 \mathcal{GP}_{\text{mat},1}(\rho^0)e_1, A_1^0 M_1(\gamma) M_1^{-1}(\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\breve{\theta}_z^0] \end{aligned} \tag{B.70}$$

where the last equality holds since $\gamma \le \rho^0$ implies that $\text{Cov}[\mathcal{GP}_1(\gamma), \mathcal{GP}_2(\rho^0)] = \text{Cov}[\mathcal{GP}_1(\rho^0), \mathcal{GP}_2(\rho^0)] = \mathbf{0}$. Thus, equation (B.70) can then be stated as

$$\begin{aligned}\text{Cov}[\mathcal{B}_A, \mathcal{B}_{A,1}(\gamma)] &= (e_1^\top \otimes A_1^0)\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \gamma\}}](\tilde{\theta}_z^0 \otimes A_1^{0\top}) \\ &- (e_1^\top \otimes A_1^0)\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \rho^0\}}](\breve{\theta}_z^0 \otimes M_1^{-1}(\rho^0) M_1(\gamma) A_1^{0\top}) \\ &= A_1^0 \mathbb{E}[x_t x_t^\top (\epsilon_t^2 + \epsilon_t u_t^\top \theta_z^0)\mathbb{1}_{\{q_t \le \gamma\}}]A_1^{0\top} \\ &- A_1^0 \mathbb{E}[x_t x_t^\top (\epsilon_t u_t^\top \theta_z^0)\mathbb{1}_{\{q_t \le \rho^0\}}]M_1^{-1}(\rho^0) M_1(\gamma) A_1^{0\top} \\ &= A_1^0\ [H_1^\epsilon(\gamma) + H_1^{\epsilon,u}(\gamma) - H_1^{\epsilon,u}(\rho^0) R_1^\top(\gamma;\rho^0)]\ A_1^{0\top}. \end{aligned} \tag{B.71}$$

Note that $\text{Cov}[\mathcal{B}_{A,1}(\gamma), \mathcal{B}_A] = \text{Cov}[\mathcal{B}_A, \mathcal{B}_{A,1}(\gamma)]^\top$. Hence, putting (B.66), (B.67), (B.69) and (B.71) together yields

$$\begin{aligned}\text{Var}[\mathcal{B}_{A,2}(\gamma)] &= A_1^0 H_1^\epsilon(\rho^0) A_1^{0\top} + A_2^0 H_2^\epsilon(\rho^0) A_2^{0\top} \\ &+ A_1^0\ [H_1(\gamma) + R_1(\gamma;\rho^0) H_1^u(\rho^0) R_1^\top(\gamma;\rho^0) - R_1(\gamma;\rho^0)(H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)) \\ &- (H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)) R_1^\top(\gamma;\rho^0)]\ A_1^{0\top} \\ &- A_1^0\ [H_1^\epsilon(\gamma) + H_1^{\epsilon,u}(\gamma) - H_1^{\epsilon,u}(\rho^0) R_1^\top(\gamma;\rho^0)]\ A_1^{0\top} \\ &- A_1^0\ [H_1^\epsilon(\gamma) + H_1^{\epsilon,u}(\gamma) - R_1(\gamma;\rho^0) H_1^{\epsilon,u}(\rho^0)]\ A_1^{0\top} \\ &= A_2^0\ H_2^\epsilon(\rho^0)\ A_2^{0\top} \\ &+ A_1^0\ [H_1^\epsilon(\rho^0) + H_1(\gamma) - 2H_1^{\epsilon,u}(\gamma) - 2H_1^\epsilon(\gamma) \\ &+ R_1(\gamma;\rho^0) H_1^u(\rho^0) R_1^\top(\gamma;\rho^0) \\ &+ R_1(\gamma;\rho^0)[-H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma) + H_1^{\epsilon,u}(\rho^0)] \\ &+ [-H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma) + H_1^{\epsilon,u}(\rho^0)] R_1^\top(\gamma;\rho^0)]\ A_1^{0\top} \\ &= A_2^0\ H_2^\epsilon(\rho^0)\ A_2^{0\top} \\ &+ A_1^0\ [H_1^\epsilon(\rho^0) - H_1^\epsilon(\gamma) + H_1^u(\gamma) \end{aligned}$$

$$
\begin{aligned}
&+ R_1(\gamma;\rho^0)H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) \\
&+ R_1(\gamma;\rho^0)[H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)] \\
&+ [H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)]R_1^\top(\gamma;\rho^0)]\ A_1^{0\top}.
\end{aligned}
$$

Pre- and post-multiplication with $C_{A,2}^{-1}(\gamma)$ then yields the claim when $\gamma \leq \rho^0$. Finally, we derive an expression for:

$$
\mathrm{Cov}[\mathcal{B}_{A,1}(\gamma), \mathcal{B}_{A,2}(\gamma)] = \mathrm{Cov}[\mathcal{B}_{A,1}(\gamma), \mathcal{B}_A] - \mathrm{Cov}[\mathcal{B}_{A,1}(\gamma), \mathcal{B}_{A,1}(\gamma)].
$$

Using results (B.71) and (B.66) immediately yields:

$$
\begin{aligned}
\mathrm{Cov}(\mathcal{B}_{A,1}(\gamma), \mathcal{B}_{A,2}(\gamma)) &= \mathrm{Cov}(\mathcal{B}_{A,1}(\gamma), \mathcal{B}_A) - \mathrm{Cov}(\mathcal{B}_{A,1}(\gamma), \mathcal{B}_{A,1}(\gamma)) \\
&= A_1^0\Big[H_1^\epsilon(\gamma) + H_1^{\epsilon,u}(\gamma) - R_1(\gamma;\rho^0)H_1^{\epsilon,u}(\gamma)\Big]A_1^{0^\top} \\
&- A_1^0\Big[H_1(\gamma) + R_1(\gamma;\rho^0)H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) - R_1(\gamma;\rho^0)[H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)] \\
&- [H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)]R_1^\top(\gamma;\rho^0)]\Big]A_1^{0^\top} \\
&= A_1^0\Big[H_1^\epsilon(\gamma) + H_1^{\epsilon,u}(\gamma) - H_1(\gamma) - R_1(\gamma;\rho^0)[H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)] \\
&- R_1(\gamma;\rho^0)[H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) + [H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)]R_1^\top(\gamma;\rho^0)]\Big]A_1^{0^\top} \\
&= A_1^0\Big[- H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma) - R_1(\gamma;\rho^0)[H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)] \\
&- R_1(\gamma;\rho^0)[H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) + [H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)]R_1^\top(\gamma;\rho^0)]\Big]A_1^{0^\top} \\
&= -A_1^0\Big[H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma) + R_1(\gamma;\rho^0)[H_1^{\epsilon,u}(\rho^0) - H_1^{\epsilon,u}(\gamma) - H_1^u(\gamma)] \\
&+ R_1(\gamma;\rho^0)[H_1^u(\rho^0)R_1^\top(\gamma;\rho^0) - [H_1^{\epsilon,u}(\gamma) + H_1^u(\gamma)]R_1^\top(\gamma;\rho^0)]\Big]A_1^{0^\top}
\end{aligned}
$$

Pre-, respectively post-multiplication with $C_{A,1}^{-1}(\gamma)$, respectively $C_{A,2}^{-1}(\gamma)$ yields the claim for $\mathrm{Cov}(\hat{\theta}_1^\gamma, \hat{\theta}_2^\gamma)$ when $\gamma \leq \rho^0$.

The case $\gamma > \rho^0$ is derived in a similar fashion and thus omitted for brevity. □

Lemma 8. *Suppose Assumption A.1 holds. Then, under $\mathbb{H}_0$ and uniformly in γ and for $i = 1, 2$,*

$$
(i)\ \hat{H}_i^\epsilon(\gamma) \xrightarrow{p} H_i^\epsilon(\gamma) \qquad (ii)\ \hat{H}_i^{\epsilon,u}(\gamma) \xrightarrow{p} H_i^{\epsilon,u}(\gamma)
$$
$$
(iii)\ \hat{H}_i^u(\gamma) \xrightarrow{p} H_i^u(\gamma) \qquad (iv)\ \hat{H}_i(\gamma) \xrightarrow{p} H_i(\gamma)
$$

Proof. **Claim** (i)**:** Let $\tilde{A}$ be the one of the two matrices A_1^0 and A_2^0 with larger Frobenius-norm. Then

$$
\begin{aligned}
\|w_t\|_2 &= \|A_1^0 x_t \mathbb{1}_{\{q_t \leq \rho^0\}} + A_2^0 x_t \mathbb{1}_{\{q_t > \rho^0\}} + \bar{u}_t\|_2 \\
&\leq \|A_1^0\|_F \|x_t\|_2 \mathbb{1}_{\{q_t \leq \rho^0\}} + \|A_2^0\|_F \|x_t\|_2 \mathbb{1}_{\{q_t > \rho^0\}} + \|u_t\|_2 \\
&\leq \|\tilde{A}\|_F \|x_t\|_2 + \|u_t\|_2
\end{aligned}
$$

Using this expression along the lines of the proof of Lemma 4 then yields the claim.

To show Claims (ii)–(iv), we consider the three cases, and their sub-cases, of Lemma 6 again.
Claim (ii)**:** *Case a:* In both sub-cases we obtain

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) &= T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top(\hat{\epsilon}_t[(\Pi_1^0-\hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\xrightarrow{p} H_1^{\epsilon,u}(\gamma) \end{aligned}$$

where convergence follows along the same lines as in the proof of Lemma 4.

Case b: In sub-case $b.1$ it holds, as for *Case a*, that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) &= T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top([\hat{\epsilon}_t(\Pi_1^0-\hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\xrightarrow{p} H_1^{\epsilon,u}(\gamma). \end{aligned}$$

In sub-case $b.2$ it follows that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) + T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) \\ &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t[(\Pi_1^0-\hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\quad + T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t[(\Pi_1^0-\hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\xrightarrow{p} H_1^{\epsilon,u}(\gamma), \end{aligned}$$

where the second term converges to 0 in probability since the sum is of order $\mathcal{O}_p(1)$ and $\hat{\rho} \xrightarrow{p} \rho^0$ implying $\mathcal{T}_1(\rho^0)\setminus\mathcal{T}_1(\hat{\rho}) \xrightarrow{p} \emptyset$.

Case c: In sub-case $c.1$ it holds that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) + T^{-1}\sum_{\mathcal{T}_1(\rho^0)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) \\ &\quad + T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\rho^0)} x_t x_t^\top(\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) \\ &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t[(\Pi_1^0-\hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\quad + T^{-1}\sum_{\mathcal{T}_1(\rho^0)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top(\hat{\epsilon}_t[(\Pi_1^0-\hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\quad + T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\rho^0)} x_t x_t^\top(\hat{\epsilon}_t[(\Pi_2^0-\hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z) \\ &\xrightarrow{p} H_1^{\epsilon,u}(\rho^0) + H_1^{\epsilon,u}(\gamma) - H_1^{\epsilon,u}(\rho^0) = H_1^{\epsilon,u}(\gamma), \end{aligned}$$

where the first and third term converge by similar arguments as in the proof of Lemma 4. The second term converges to 0 in probability since the sum is of order $\mathcal{O}_p(1)$ and $\hat{\rho} \xrightarrow{p} \rho^0$ implying $\mathcal{T}_1(\rho^0) \setminus \mathcal{T}_1(\hat{\rho}) \xrightarrow{p} \emptyset$ (this notation means that the number of elements in the set $\mathcal{T}_1(\rho^0) \setminus \mathcal{T}_1(\hat{\rho})$ is negligible in the limit as $T \to \infty$).
For sub-case $c.2$ it holds that

$$\begin{aligned}
T^{-1} \sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top (\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) &= T^{-1} \sum_{\mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) + T^{-1} \sum_{\mathcal{T}_1(\hat{\rho}) \setminus \mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) \\
&\quad + T^{-1} \sum_{\mathcal{T}_1(\gamma) \setminus \mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{\epsilon}_t \hat{u}_t^\top \hat{\theta}_z) \\
&= T^{-1} \sum_{\mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{\epsilon}_t [(\Pi_1^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z) \\
&\quad + T^{-1} \sum_{\mathcal{T}_1(\hat{\rho}) \setminus \mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{\epsilon}_t [(\Pi_2^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z) \\
&\quad + T^{-1} \sum_{\mathcal{T}_1(\gamma) \setminus \mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{\epsilon}_t [(\Pi_2^0 - \hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z) \\
&\xrightarrow{p} H_1^{\epsilon,u}(\rho^0) + H_1^{\epsilon,u}(\gamma) - H_1^{\epsilon,u}(\rho^0) = H_1^{\epsilon,u}(\gamma),
\end{aligned}$$

where the first and third term converge by similar arguments as in the proof of Lemma 4. The second term converges to 0 in probability since the sum is of order $\mathcal{O}_p(1)$ and $\hat{\rho} \xrightarrow{p} \rho^0$ implying $\mathcal{T}_1(\rho^0) \setminus \mathcal{T}_1(\hat{\rho}) \xrightarrow{p} \emptyset$.

Claim (iii)**:** *Case a:* In both sub-cases we obtain

$$\begin{aligned}
T^{-1} \sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 &= T^{-1} \sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\
&\xrightarrow{p} H_1^u(\gamma)
\end{aligned}$$

where convergence follows along the same lines as in the proof of Lemma 4.

Case b: In sub-case $b.1$ it also holds, as before, that

$$\begin{aligned}
T^{-1} \sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 &= T^{-1} \sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\
&\xrightarrow{p} H_1^u(\gamma).
\end{aligned}$$

In sub-case $b.2$ it follows that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 + T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 \\ &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &\xrightarrow{p} H_1^u(\gamma), \end{aligned}$$

where the second term converges to 0 in probability since the sum is of order $\mathcal{O}_p(1)$ and $\hat{\rho} \xrightarrow{p} \rho^0$ implying $\mathcal{T}_1(\rho^0) \setminus \mathcal{T}_1(\hat{\rho}) \xrightarrow{p} \emptyset$.

Case c: In sub-case $c.1$ it holds that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 + T^{-1}\sum_{\mathcal{T}_1(\rho^0)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 \\ &= T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\rho^0)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\rho^0)} x_t x_t^\top ([(\Pi_2^0 - \hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &\xrightarrow{p} H_1^u(\rho^0) + H_1^u(\gamma) - H_1^u(\rho^0) = H_1^u(\gamma), \end{aligned}$$

where the first and third term converge by similar arguments as in the proof of Lemma 4. The second term converges to 0 in probability since the sum is of order $\mathcal{O}_p(1)$ and $\hat{\rho} \xrightarrow{p} \rho^0$ implying $\mathcal{T}_1(\rho^0) \setminus \mathcal{T}_1(\hat{\rho}) \xrightarrow{p} \emptyset$.
For sub-case $c.2$ it holds that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 &= T^{-1}\sum_{\mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 + T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})\setminus\mathcal{T}_1(\rho^0)} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top (\hat{u}_t^\top \hat{\theta}_z)^2 \\ &= T^{-1}\sum_{\mathcal{T}_1(\rho^0)} x_t x_t^\top ([(\Pi_1^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\hat{\rho})\setminus\mathcal{T}_1(\rho^0)} x_t x_t^\top ([(\Pi_2^0 - \hat{\Pi}_1)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \\ &+ T^{-1}\sum_{\mathcal{T}_1(\gamma)\setminus\mathcal{T}_1(\hat{\rho})} x_t x_t^\top ([(\Pi_2^0 - \hat{\Pi}_2)^\top x_t + u_t]^\top \hat{\theta}_z)^2 \end{aligned}$$

$$\xrightarrow{p} H_1^u(\rho^0) + H_1^u(\gamma) - H_1^u(\rho^0) = H_1^u(\gamma),$$

where the first and third term converge by similar arguments as in the proof of Lemma 4. The second term converges to 0 in probability since the sum is of order $\mathcal{O}_p(1)$ and $\hat{\rho} \xrightarrow{p} \rho^0$ implying $\mathcal{T}_1(\rho^0) \setminus \mathcal{T}_1(\hat{\rho}) \xrightarrow{p} \emptyset$.

Claim (iv)**:** As in Lemma 4.

For $i = 2$, the proof follows similar steps and omitted for brevity. □

Proof of Theorem 3.

(i) sup LR Test: The proof of this result follows the same arguments as in the LRF case. For brevity, we will only display the major differences to the LRF case. As in the LRF case, we split the proof into two parts: in part (i) we will show that $T^{-1}SSR_1(\gamma) \xrightarrow{p} \sigma^2$ and in part (ii) that $SSR_0 - SSR_1(\gamma) \Rightarrow \mathcal{E}_A^\top(\gamma)C_{A,2}(\gamma)C_A^{-1}C_{A,1}(\gamma)\mathcal{E}(\gamma)$.

Part (i)**.** As in the LRF proof (cf. equation (B.26)) it holds uniformly in γ that

$$\begin{aligned}
T^{-1}SSR_1(\gamma) &= T^{-1}[Y_1^\gamma - \hat{W}_1^\gamma\hat{\theta}_1^\gamma]^\top[Y_1^\gamma - \hat{W}_1^\gamma\hat{\theta}_1^\gamma] \\
&+ T^{-1}[Y_2^\gamma - \hat{W}_2^\gamma\hat{\theta}_2^\gamma]^\top[Y_2^\gamma - \hat{W}_2^\gamma\hat{\theta}_2^\gamma] \\
&= T^{-1}[\hat{W}_1^\gamma(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_1^\gamma]^\top[\hat{W}_1^\gamma(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_1^\gamma] \\
&+ T^{-1}[\hat{W}_2^\gamma(\theta^0 - \hat{\theta}_2^\gamma) + \tilde{\epsilon}_2^\gamma]^\top[\hat{W}_2^\gamma(\theta^0 - \hat{\theta}_2^\gamma) + \tilde{\epsilon}_2^\gamma] \\
&= T^{-1}\tilde{\epsilon}^\top\tilde{\epsilon} \\
&+ 2(T^{-1}\tilde{\epsilon}_1^\gamma\hat{W}_1^\gamma)(\theta^0 - \hat{\theta}_1^\gamma) + (\theta^0 - \hat{\theta}_1^\gamma)^\top(T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma)(\theta^0 - \hat{\theta}_1^\gamma) \\
&+ 2(T^{-1}\tilde{\epsilon}_2^\gamma\hat{W}_2^\gamma)(\theta^0 - \hat{\theta}_2^\gamma) + (\theta^0 - \hat{\theta}_2^\gamma)^\top(T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^\gamma)(\theta^0 - \hat{\theta}_2^\gamma) \\
&= T^{-1}\tilde{\epsilon}^\top\tilde{\epsilon} + o_p(1),
\end{aligned} \tag{B.72}$$

where the last equality holds because, for $i = 1, 2$, $T^{-1}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^\gamma = o_p(1)$, $T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma = \mathcal{O}_p(1)$ and $\theta^0 - \hat{\theta}_i^\gamma = (T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1}(T^{-1}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^\gamma) = \mathcal{O}_p(1)o_p(1) = o_p(1)$ uniformly in γ by Lemma 3.
Next, rewrite (B.72) as

$$T^{-1}SSR_1(\gamma) = T^{-1}\tilde{\epsilon}_1^{\rho^0\top}\tilde{\epsilon}_1^{\rho^0} + T^{-1}\tilde{\epsilon}_2^{\rho^0\top}\tilde{\epsilon}_2^{\rho^0} + o_p(1). \tag{B.73}$$

By construction

$$\tilde{\epsilon}_1^{\rho^0} = \epsilon_1^{\rho^0} + (Z_1^{\rho^0} - \hat{Z}_1^{\rho^0})\theta_z^0$$

and thus

$$\tilde{\epsilon}_1^{\rho^0} = \begin{cases} s_1^{\rho^0} + X_1^{\rho^0}(\Pi_1^0 - \hat{\Pi}_1) & \text{if } \rho^0 \leq \hat{\rho} \\ s_1^{\rho^0} + X_1^{\rho^0}(\Pi_1^0 - \hat{\Pi}_1) + o_p(1) & \text{if } \rho^0 > \hat{\rho} \end{cases}$$

where $s_1^{\rho^0} = \epsilon_1^{\rho^0} + u_1^{\rho^0}\theta_z^0$. It can be shown that:

$$\begin{aligned} T^{-1}\tilde{\epsilon}_1^{\rho^0\top}\tilde{\epsilon}_1^{\rho^0} &= T^{-1}s_1^{\rho^0\top}s_1^{\rho^0} + 2(T^{-1}s_1^{\rho^0\top}X_1^{\rho^0})(\Pi_1^0 - \hat{\Pi}_1) \\ &+ (\Pi_1^0 - \hat{\Pi}_1)^\top(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})(\Pi_1^0 - \hat{\Pi}_1) \\ &= T^{-1}s_1^{\rho^0\top}s_1^{\rho^0} + o_p(1) \end{aligned}$$

because $T^{-1}s_1^{\rho^0\top}X_1^{\rho^0} = o_p(1)$ and $T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} = \mathcal{O}_p(1)$ and $\Pi_1^0 - \hat{\Pi}_1 = o_p(1)$ by Lemma 3.
Similarly, we obtain

$$T^{-1}\tilde{\epsilon}_2^{\rho^0\top}\tilde{\epsilon}_2^{\rho^0} = T^{-1}s_2^{\rho^0\top}s_2^{\rho^0} + o_p(1).$$

Therefore, (B.73) reads as

$$\begin{aligned} T^{-1}SSR_1(\gamma) &= T^{-1}s_1^{\rho^0\top}s_1^{\rho^0} + T^{-1}s_2^{\rho^0\top}s_2^{\rho^0} + o_p(1) \\ &= T^{-1}s^\top s + o_p(1) \\ &\xrightarrow{p} \sigma_\epsilon^2 + 2\Sigma_{\epsilon,u}^\top\theta_z^0 + \theta_z^{0\top}\Sigma_u\theta_z^0 \equiv \sigma^2, \end{aligned}$$

uniformly in γ, proving part (i).

Part (ii)**.** For this part, derivations remain as in the LRF case (up to equation (B.21)). Utilizing Lemma 4, expressions (B.30) and (B.31) in the LRF proof become

$$T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma \xrightarrow{p} C_{A,i}(\gamma)$$

and

$$\hat{\beta} = D_{A,1}(\gamma)\hat{\beta}_1 + D_{A,2}(\gamma)\hat{\beta}_2 + o_p(1)$$

by Lemma 3 with $D_{A,1}(\gamma) \equiv C_A^{-1}C_{A,1}(\gamma)$ and therefore, $D_{A,2}(\gamma) = C_A^{-1}C_{A,2}(\gamma) = I_p - D_{A,1}(\gamma)$. Consequently, equations (B.30)–(B.32a) in the LRF proof are adjusted in this fashion as well. The following derivations then remain the same.
Last, equation (B.36) from the LRF case now reads as [22]

$$\hat{\beta}_1 - \hat{\beta}_2 = C_{A,1}^{-1}(\gamma)\mathcal{B}_{A,1}(\gamma) - C_{A,2}^{-1}(\gamma)\mathcal{B}_{A,2}(\gamma) \equiv \mathcal{E}_A(\gamma).$$

Thus, as in the LRF case, it follows that

$$\begin{aligned} SSR_0 - SSR_1(\gamma) &= (\hat{\beta}_1 - \hat{\beta}_2)^\top C_{A,2}(\gamma)D_{A,1}(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \\ &\Rightarrow \mathcal{E}_A^\top(\gamma)C_{A,2}(\gamma)D_{A,1}(\gamma)\mathcal{E}_A(\gamma) \end{aligned}$$

uniformly in γ. Together with Part (i), $(a.s.)$ continuity of the process $\mathcal{E}_A(\gamma)$, the continuous mapping theorem and weak convergence (uniformly in γ) it then follows that

$$\sup_{\gamma\in\Gamma}\frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/T} \Rightarrow \sup_{\gamma\in\Gamma}\frac{\mathcal{E}_A^\top(\gamma)C_{A,2}(\gamma)C_A^{-1}C_{A,1}(\gamma)\mathcal{E}_A(\gamma)}{\sigma^2}$$

proving the claim of the theorem.

[22] A^0 is replaced with A_i^0, $i = 1, 2$, absorbed in the definition of $\mathcal{B}_{A,1}(\gamma)$.

(ii) sup Wald Test: The proof follows the exact same arguments as the proof of Theorem 2 by replacing the LRF quantities with the according TRF quantities. □

To write down Corollary A2 to Theorem 3 below, which derives the asymptotic distributions of the 2SLS tests under conditional homoskedasticity, we define the Gaussian processes

$$\tilde{\mathcal{E}}_A(\gamma) = C_{A,1}^{-1}(\gamma)\tilde{\mathcal{B}}_{A,1}(\gamma) - C_{A,2}^{-1}(\gamma)\tilde{\mathcal{B}}_{A,2}(\gamma)$$

and

$$\begin{aligned}
\tilde{\mathcal{B}}_{A,1}(\gamma) &= A_1^0\Big[\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma\wedge\rho^0)\Sigma^{1/2}\tilde{\theta}_z^0 - R_1(\gamma\wedge\rho^0;\rho^0)\tilde{\mathcal{GP}}_{\text{mat},1}(\rho^0)\Sigma^{1/2}\breve{\theta}_z^0\Big] \\
&\quad + A_2^0\Big[\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)\Sigma^{1/2}\tilde{\theta}_z^0 - \tilde{\mathcal{GP}}_{\text{mat},1}(\gamma\wedge\rho^0)\Sigma^{1/2}\tilde{\theta}_z^0\Big] \\
&\quad - A_2^0\Big[(R_2(\gamma\wedge\rho^0;\rho^0) - R_2(\gamma;\rho^0))\tilde{\mathcal{GP}}_{\text{mat},2}(\rho^0)\Sigma^{1/2}\breve{\theta}_z^0\Big] \\
\tilde{\mathcal{B}}_{A,2}(\gamma) &= \tilde{\mathcal{B}}_A(\gamma_{max}) - \tilde{\mathcal{B}}_{A,1}(\gamma),
\end{aligned}$$

and $\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)$ is a $q\times(p_1+1)$ matrix where all columns are independent $q\times 1$ zero mean Gaussian processes with covariance kernel $M_1(\gamma)$.[23] Then we have:

Corollary A 2 (to Theorem 3)**.** *Let Z be generated by (2.2), Y be generated by (2.3), and $\hat{Z}$ be calculated by (3.5). Then, under $\mathbb{H}_0$, and Assumptions A.1, A.2 and A.4,*
(i)

$$\sup_{\gamma\in\Gamma} LR_{T,TRF}^{2SLS}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \tilde{\mathcal{E}}_A^\top(\gamma)Q_A^{-1}(\gamma)\tilde{\mathcal{E}}_A(\gamma),$$

(ii)

$$\sup_{\gamma\in\Gamma} W_{T,TRF}^{2SLS}(\gamma) \Rightarrow \sup_{\gamma\in\Gamma} \tilde{\mathcal{E}}_A^\top(\gamma)\tilde{V}_A^{-1}(\gamma)\tilde{\mathcal{E}}_A(\gamma).$$

where $\tilde{V}_A(\gamma) = \tilde{V}_{A,1}(\gamma) + \tilde{V}_{A,2}(\gamma) - \tilde{V}_{A,12}(\gamma) - \tilde{V}_{A,12}^\top(\gamma)$ and:

$$\begin{aligned}
\tilde{V}_{A,1}(\gamma) &= C_{A,1}^{-1}(\gamma)\Big[\sigma^2 C_{A,1}(\gamma) - (\sigma^2-\sigma_\epsilon^2)A_1^0R_1(\gamma;\rho^0)M_1(\gamma)A_1^{0\top}\Big]C_{A,1}^{-1}(\gamma) \\
\tilde{V}_{A,2}(\gamma) &= C_{A,2}^{-1}(\gamma)\Big[\sigma_\epsilon^2 C_{A,2}(\gamma) + (\sigma^2-\sigma_\epsilon^2)(C_{A,1}(\gamma) - A_1^0R_1(\gamma;\rho^0)M_1(\gamma)A_1^{0\top})\Big]C_{A,2}^{-1}(\gamma) \\
\tilde{V}_{A,12}(\gamma) &= -(\sigma^2-\sigma_\epsilon^2)C_{A,1}^{-1}(\gamma)\Big[C_{A,1}(\gamma) - A_1^0R_1(\gamma;\rho^0)M_1(\gamma)A_1^{0\top}\Big]C_{A,2}^{-1}(\gamma)
\end{aligned}$$

whenever $\gamma\le\rho^0$. If $\gamma>\rho^0$, then

$$\begin{aligned}
\tilde{V}_{A,1}(\gamma) &= C_{A,1}^{-1}(\gamma)\Big[\sigma_\epsilon^2 C_{A,1}(\gamma) + (\sigma^2-\sigma_\epsilon^2)(C_{A,2}(\gamma) - A_2^0R_2(\gamma;\rho^0)M_2(\gamma)A_2^{0\top})\Big]C_{A,1}^{-1}(\gamma) \\
\tilde{V}_{A,2}(\gamma) &= C_{A,2}^{-1}(\gamma)\Big[\sigma^2 C_{A,2}(\gamma) - (\sigma^2-\sigma_\epsilon^2)A_2^0R_2(\gamma;\rho^0)M_2(\gamma)A_2^{0\top}\Big]C_{A,2}^{-1}(\gamma) \\
\tilde{V}_{A,12}(\gamma) &= -(\sigma^2-\sigma_\epsilon^2)C_{A,1}^{-1}(\gamma)\Big[C_{A,2}(\gamma) - A_2^0R_2(\gamma;\rho^0)M_2(\gamma)A_2^{0\top}\Big]C_{A,2}^{-1}(\gamma)
\end{aligned}$$

[23] Thus, the only difference between the two Gaussian processes $\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)$ and $\mathcal{GP}_{\text{mat},1}(\gamma)$ lies again in their covariance functions.

Moreover, if the system is just-identified, i.e. if $p = q$, then the two test statistics are asymptotically equivalent with asymptotic distribution given by

$$\sup_{\gamma \in \Gamma} \frac{\tilde{\mathcal{E}}_{\mathcal{A}}^{\top}(\gamma) C_2(\gamma) C^{-1} C_1(\gamma) \tilde{\mathcal{E}}_{\mathcal{A}}(\gamma)}{\sigma^2}.$$

Proof of Corollary A2: The proof follows the exact same arguments as the proof of Corollary A1. Note that when $p = q$, $\tilde{\mathcal{E}}_{\mathcal{A}}(\gamma)$ does not simplify to $\tilde{\mathcal{E}}(\gamma)$ from the LRF, because ρ^0 does not disappear from the definition of $\tilde{\mathcal{E}}_{\mathcal{A}}(\gamma)$. □

Proofs for Section 5: GMM tests

Corollary A 3 (to Theorem 4)**.** *Let Z be generated by (2.1) and Y be generated by (2.3). Then, under $\mathbb{H}_0$, Assumptions A.1, A.2 and $p = q$,*

$$\sup_{\gamma \in \Gamma} W_T^{GMM}(\gamma) \Rightarrow \sup_{\gamma \in \Gamma} J_2(\gamma),$$

where

$$\begin{aligned} J_2(\gamma) = &\left[M_1^{-1}(\gamma) \tilde{\mathcal{GP}}_{mat,1}^{(r1)}(\gamma) - M_2^{-1}(\gamma) \tilde{\mathcal{GP}}_{mat,2}^{(r1)}(\gamma) \right]^{\top} \\ &\times [M_1(\gamma) M^{-1} M_2(\gamma)] \qquad \text{(B.74)} \\ &\times \left[M_1^{-1}(\gamma) \tilde{\mathcal{GP}}_{mat,1}^{(r1)}(\gamma) - M_2^{-1}(\gamma) \tilde{\mathcal{GP}}_{mat,2}^{(r1)}(\gamma) \right] \qquad \text{(B.75)} \end{aligned}$$

and $\tilde{\mathcal{GP}}_{mat,i}^{(r1)}$ is the first row of the $q \times (p_1 + 1)$ matrix $\tilde{\mathcal{GP}}_{mat,1}^{(r2)}$ defined in Corollary A1.

Proof of Corollary A3. We have $H_i^{\epsilon}(\gamma) = \sigma_{\epsilon}^2 M_i(\gamma)$, $N_i(\gamma) = A^0 M_i(\gamma)$, $V_{i,GMM} = \sigma_{\epsilon}^2 (A^0 M_i A^{0\top})^{-1} = (A^{0\top})^{-1} M_i^{-1}(\gamma) (A^0)^{-1}$. The rest follows by plugging these into Theorem 4. □

Proof of Corollary 2. As for Corollary 1, we can equivalently write $\text{Prob}(q_t \le \gamma) = \lambda$ for all $\gamma \in \Gamma$ where λ is uniformly distributed on $\Lambda_{\kappa} = (\kappa; 1 - \kappa)$, i.e $\lambda \sim U(\Lambda_{\kappa})$.
Now, by Assumption A.3, we have that

$$\begin{aligned} H_1^{\epsilon}(\gamma) = \lambda H^{\epsilon}, \quad H_2^{\epsilon}(\gamma) = (1 - \lambda) H^{\epsilon} \qquad \text{(B.76)} \\ N_1(\gamma) = \lambda N, \quad N_2(\gamma) = (1 - \lambda) N \\ V_{GMM,1}(\gamma) = \lambda^{-1} \left[N H^{\epsilon^{-1}} N^{\top} \right]^{-1} \\ V_{GMM,2}(\gamma) = (1 - \lambda)^{-1} \left[N H^{\epsilon^{-1}} N^{\top} \right]^{-1} \\ V_{GMM,1}(\gamma) + V_{GMM,2}(\gamma) = \frac{\left[N H^{\epsilon^{-1}} N^{\top} \right]^{-1}}{\lambda(1 - \lambda)} \qquad \text{(B.77)} \end{aligned}$$

$$V_{GMM,1}(\gamma)N_1(\gamma)H_1^{\epsilon^{-1}}(\gamma) = \lambda^{-1}\left[NH^{\epsilon^{-1}}N^\top\right]^{-1} NH^{\epsilon^{-1}}$$
$$V_{GMM,2}(\gamma)N_2(\gamma)H_2^{\epsilon^{-1}}(\gamma) = (1-\lambda)^{-1}\left[NH^{\epsilon^{-1}}N^\top\right]^{-1} NH^{\epsilon^{-1}}$$

Moreover, (B.76) implies that –under Assumptions A.2 and A.3– the Gaussian process $\overline{\mathcal{GP}}_1(\gamma)$ can be restated as

$$\overline{\mathcal{GP}}_1(\gamma) = H^{\epsilon^{1/2}}\overline{\mathcal{BM}}(\lambda)$$
$$\overline{\mathcal{GP}} = H^{\epsilon^{1/2}}\overline{\mathcal{BM}}(1)$$

where $\overline{\mathcal{BM}}(\cdot)$ is a $q \times 1$-vector of independent Brownian motions on the unit interval. Thus, the term $V_{GMM,1}(\gamma)N_1(\gamma)H_1^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_1(\gamma) - V_{GMM,2}(\gamma)N_2(\gamma)H_2^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_2(\gamma)$ can be restated in terms of λ: as

$$V_{GMM,1}(\gamma)N_1(\gamma)H_1^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_1(\gamma) - V_{GMM,2}(\gamma)N_2(\gamma)H_2^{\epsilon^{-1}}(\gamma)\overline{\mathcal{GP}}_2(\gamma) \tag{B.78}$$
$$\lambda^{-1}\left[NH^{\epsilon^{-1}}N^\top\right]^{-1} NH^{\epsilon^{-1/2}}\overline{\mathcal{BM}}(\lambda) - (1-\lambda)^{-1}\left[NH^{\epsilon^{-1}}N^\top\right]^{-1} NH^{\epsilon^{-1/2}}(\overline{\mathcal{BM}}(1) - \overline{\mathcal{BM}}(\lambda)).$$

Because $\left[NH^{\epsilon^{-1}}N^\top\right]^{-1} NH^{\epsilon^{-1/2}}$ is half of a projection matrix, by similar arguments as for the proof of Corollary 1, we obtain the desired result. □

Proofs for Section 3: 2SLS versus GMM estimators

Lemma 9. *Suppose Assumptions A.1–A.4 hold and that $p = q = 1$. Define as in Theorem 1, $\lambda = Prob(q_t \le \gamma)$, $\mu^0 = Prob(q_t \le \rho^0)$, $\alpha = (\mu^0 - \lambda)/(1-\lambda)$, $\beta = \mu^0/\lambda$, and let $E(x_t^2) = m$. Then, under $\mathbb{H}_0$:*

$$V^*_{1,GMM}(\gamma) = \begin{cases} \frac{\sigma_\epsilon^2}{\lambda m\ \pi_1^{0^2}} & \text{if } \gamma \le \rho^0 \\ \frac{\sigma_\epsilon^2}{\lambda m\ [\beta\pi_1^0 + (1-\beta)\pi_2^0]^2} & \text{if } \gamma > \rho^0 \end{cases}$$
$$V^*_{2,GMM}(\gamma) = \begin{cases} \frac{\sigma_\epsilon^2}{(1-\lambda) m\ [\alpha\pi_1^0 + (1-\alpha)\pi_2^0]^2} & \text{if } \gamma \le \rho^0 \\ \frac{\sigma_\epsilon^2}{(1-\lambda) m\ \pi_2^{0^2}} & \text{if } \gamma > \rho^0. \end{cases}$$

Moreover,

$$V^*_{A,1}(\gamma) = \begin{cases} \frac{\sigma_\epsilon^2}{\lambda m\ \pi_1^{0^2}} + \frac{\sigma^2 - \sigma_\epsilon^2}{\lambda m\ \pi_1^{0^2}}\left(1 - \frac{\lambda}{\mu^0}\right) & \text{if } \gamma \le \rho^0 \\ \frac{\sigma_\epsilon^2}{\lambda m\ [\beta\pi_1^{0^2} + (1-\beta)\pi_2^{0^2}]} + \frac{\pi_2^{0^2}(1-\lambda)(\sigma^2-\sigma_\epsilon^2)}{\lambda^2 m\ [\beta\pi_1^{0^2} + (1-\beta)\pi_2^{0^2}]^2}\left(1 - \frac{1-\lambda}{1-\mu^0}\right) & \text{if } \gamma > \rho^0 \end{cases}$$
$$V^*_{A,2}(\gamma) = \begin{cases} \frac{\sigma_\epsilon^2}{(1-\lambda) m\ [\alpha\pi_1^{0^2} + (1-\alpha)\pi_2^{0^2}]} + \frac{\pi_1^{0^2}\lambda(\sigma^2-\sigma_\epsilon^2)}{(1-\lambda)^2 m\ [\alpha\pi_1^{0^2} + (1-\alpha)\pi_2^{0^2}]^2}\left(1 - \frac{\lambda}{\mu^0}\right) & \text{if } \gamma \le \rho^0 \\ \frac{\sigma_\epsilon^2}{(1-\lambda) m\ \pi_2^{0^2}} + \frac{\sigma^2-\sigma_\epsilon^2}{(1-\lambda) m\ \pi_2^{0^2}}\left(1 - \frac{1-\lambda}{1-\mu^0}\right) & \text{if } \gamma > \rho^0 \end{cases}$$

Proof of Lemma 9. First, we show the claim for the GMM case and afterwards for the 2SLS case.

GMM Variances: Let $\gamma \le \rho^0$. Then, if Assumptions A.1–A.4 hold, it follows that $H_1^\epsilon(\gamma) = \mathbb{E}[x_t^2\epsilon_t^2 \mathbb{1}_{\{q_t\le\gamma\}}] = \mathbb{E}[\mathbb{1}_{\{q_t\le\gamma\}}]\cdot\mathbb{E}[x_t^2]\sigma_\epsilon^2 = \lambda\sigma_\epsilon^2 m$, $H_1^\epsilon(\gamma)(\gamma) = \mathbb{E}[x_t^2\epsilon_t^2\mathbb{1}_{\{q_t>\gamma\}}] = (1-\lambda)\sigma_\epsilon^2 m$, $N_1(\gamma) = \mathbb{E}[x_t z_t \mathbb{1}_{\{q_t\le\gamma\}}] = \mathbb{E}[x_t^2\pi_1^0\mathbb{1}_{\{q_t\le\gamma\}}] = \lambda\pi_1^0 m$, and $N_2(\gamma) = \mathbb{E}[x_t z_t\mathbb{1}_{\{q_t>\gamma\}}] = \mathbb{E}[x_t^2\pi_1^0(\mathbb{1}_{\{q_t\le\rho^0\}} - \mathbb{1}_{\{q_t\le\gamma\}})] + \mathbb{E}[x_t^2\pi_2^0\mathbb{1}_{\{q_t>\rho^0\}}] = (\mu^0-\lambda)\pi_1^0 m + (1-\mu^0)\pi_2^0 m$. Plugging these results into the expressions for $V_{i,GMM}(\gamma)$ defined just before Theorem 4 directly yields the claim. The case $\gamma > \rho^0$ is omitted for brevity but follows similar arguments.

2SLS Variances: Let $\gamma \le \rho^0$. Then, if Assumptions A.1–A.4 hold, it follows that $M_1(\gamma) = \mathbb{E}[x_t^2\mathbb{1}_{\{q_t\le\gamma\}})] = \lambda m$, $C_{A,1}(\gamma) = \lambda\pi_1^{0^2} m$, and also that $\Psi_1(\gamma) \equiv \mathbb{E}[v_t v_t^\top x_t^2 \mathbb{1}_{\{q_t\le\gamma\}}] = \lambda m\Sigma$. Hence, $(\tilde\theta_z^{0\top}\otimes A_1^0)\Psi_1(\gamma)(\tilde\theta_z^0\otimes A_1^{0\top}) = \lambda\pi_1^{0^2} m\tilde\theta_z^{0\top}\Sigma\tilde\theta_z^0 = \lambda\pi_1^{0^2}m\sigma^2$, for example. Similar derivations apply for all the other quantities in $V_{A,1}(\gamma)$ defined in Definition A.3. Thus, it follows that

$$
\begin{aligned}
V_{A,1}(\gamma) &= \frac{1}{\lambda^2\pi_1^{0^4}m^2}\Big[\lambda\pi_1^{0^2}m\tilde\theta_z^{0\top}\Sigma\tilde\theta_z^0 + \frac{\lambda^2}{\mu^0}\pi_1^{0^2}m\check\theta_z^{0\top}\Sigma\check\theta_z^0 - 2\frac{\lambda^2}{\mu^0}\pi_1^{0^2}m\tilde\theta_z^{0\top}\Sigma\check\theta_z^0\Big] \\
&= \frac{1}{\pi_1^{0^2}m}\Big[\frac{\tilde\theta_z^{0\top}\Sigma\tilde\theta_z^0}{\lambda} + \frac{\check\theta_z^{0\top}\Sigma\check\theta_z^0}{\mu^0} - 2\frac{\tilde\theta_z^{0\top}\Sigma\check\theta_z^0}{\mu^0}\Big] \\
&= \frac{1}{\pi_1^{0^2}m}\Big[\frac{\mu^0\sigma_\epsilon^2 + 2\mu^0\theta_z^0\sigma_{\epsilon,u} + \mu^0\theta_z^{0^2}\sigma_u^2 + \lambda\theta_z^{0^2}\sigma_u^2 - 2\lambda\theta_z^0\sigma_{\epsilon,u} - 2\lambda\theta_z^{0^2}\sigma_u^2}{\lambda\mu^0}\Big] \\
&= \frac{1}{\pi_1^{0^2}m}\left[\frac{\sigma_\epsilon^2}{\lambda} + \theta_z^0(\sigma^2-\sigma_\epsilon^2)\Big(\frac{1}{\lambda} - \frac{1}{\mu^0}\Big)\right] = V_{A,1}^*(\gamma),
\end{aligned}
$$

where $\sigma^2 - \sigma_\epsilon^2 = 2\sigma_{\epsilon,u}\theta_z^0 + \sigma_u^2(\theta_z^0)^2$, and $\sigma_{\epsilon,u} = \Sigma_{\epsilon,u}$, $\sigma_u^2 = \Sigma_u$, proving the claim for $V_{A,1}^*(\gamma)$.

Next, we derive the desired result for $\mathrm{V}_{A,2}^*(\gamma)$. By the same arguments as above it immediately follows that

$$
\begin{aligned}
V_{A,2}(\gamma) = &\frac{1}{[(\mu^0-\lambda)\pi_1^{0^2} + (1-\mu^0)\pi_2^0]^2m^2} \\
&\times\Big[\mu^0\pi_1^{0^2}Me_1^\top\Sigma e_1 + (1-\mu^0)\pi_2^{0^2}Me_1^\top\Sigma e_1 + \lambda\pi_1^{0^2}m\tilde\theta_z^{0\top}\Sigma\tilde\theta_z^0 \\
&+ \frac{\lambda^2}{\mu^0}\pi_1^{0^2}m\check\theta_z^{0\top}\Sigma\check\theta_z^0 - 2\frac{\lambda^2}{\mu^0}\pi_1^{0^2}m\tilde\theta_z^{0\top}\Sigma\check\theta_z^0 - 2\frac{\lambda^2}{\mu^0}\pi_1^{0^2}Me_1^\top\Sigma\tilde\theta_z^0 \\
&+ 2\lambda\pi_1^{0^2}Me_1^\top\Sigma\check\theta_z^0\Big]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{[(\mu^0-\lambda)\pi_1^{0^2}+(1-\mu^0)\pi_2^0]^2 m} \\
&\times \Big[(1-\mu^0)\pi_2^{0^2}\sigma_\epsilon^2+\mu^0\pi_1^{0^2}\sigma_\epsilon^2+\lambda\pi_1^{0^2}\sigma_\epsilon^2+2\lambda\pi_1^{0^2}\theta_z^0\sigma_{\epsilon,u}+\lambda\pi_1^{0^2}\theta_z^{0^2}\sigma_u^2 \\
&\quad+\frac{\lambda^2}{\mu^0}\pi_1^{0^2}\theta_z^{0^2}\sigma_u^2-2\frac{\lambda^2}{\mu^0}\pi_1^{0^2}\theta_z^0\sigma_{\epsilon,u}-2\frac{\lambda^2}{\mu^0}\pi_1^{0^2}\theta_z^{0^2}\sigma_u^2-2\lambda\pi_1^{0^2}\sigma_\epsilon^2-2\lambda\pi_1^{0^2}\theta_z^0\sigma_{\epsilon,u} \\
&\quad+2\lambda\pi_1^{0^2}\theta_z^0\sigma_{\epsilon,u}\Big] \\
&= \frac{\sigma_\epsilon^2}{[(\mu^0-\lambda)\pi_1^{0^2}+(1-\mu^0)\pi_2^0]m}+\frac{\pi_1^{0^2}\lambda(1-\frac{\lambda}{\mu^0})\theta_z^0(2\sigma_{\epsilon,u}+\theta_z^0\sigma_u^2)}{[(\mu^0-\lambda)A_1^{0^2}+(1-\mu^0)A_2^0]^2 m}=V_{A,2}^*(\gamma)
\end{aligned}
$$

proving the claim for $V_{A,2}^*(\gamma)$. By a symmetry argument the claim follows for $\gamma>\rho^0$. □

Proof of Theorem 1.

Part (i): Limiting distributions. This follows from Caner and Hansen (2004) and Lemma 9 for GMM and Lemma 7 and Lemma 9 for 2SLS.

Part (ii): Variance comparisons for TRF. We only analyze the case $\gamma\le\rho^0$; by symmetry, the claim for $\gamma>\rho^0$ follows. From Lemma 9 it follows that:

$$
V_{1,GMM}^*(\gamma)-V_{A,1}^*(\gamma)=-\frac{1}{\lambda\pi_1^{0^2}m}\left[(\sigma^2-\sigma_\epsilon^2)\Big(1-\frac{\lambda}{\mu^0}\Big)\right].
$$

Hence,

$$
V_{1,GMM}^*(\gamma)\ge V_{A,1}^*(\gamma)\iff\sigma^2\le\sigma_\epsilon^2.
$$

For the second subsample,

$$
\begin{aligned}
V_{2,GMM}^*(\gamma)&=\frac{\sigma_\epsilon^2}{(1-\lambda)m\,[\alpha\pi_1^0+(1-\alpha)\pi_2^0]^2} \\
V_{A,2}^*(\gamma)&=\frac{\sigma_\epsilon^2}{(1-\lambda)m\,[\alpha\pi_1^{0^2}+(1-\alpha)\pi_2^{0^2}]}+\frac{\pi_1^{0^2}\lambda(1-\frac{\lambda}{\mu^0})(\sigma^2-\sigma_\epsilon^2)}{(1-\lambda)^2 m\,[\alpha\pi_1^{0^2}+(1-\alpha)\pi_2^{0^2}]^2}.
\end{aligned}
$$

From this,

$$
\begin{aligned}
&V_{2,GMM}^*(\gamma)-V_{A,2}^*(\gamma)\ge 0 \\
&\iff \frac{\sigma_\epsilon^2}{(1-\lambda)m\,[\alpha\pi_1^0+(1-\alpha)\pi_2^0]^2}-\frac{\sigma_\epsilon^2}{(1-\lambda)m\,[\alpha\pi_1^{0^2}+(1-\alpha)\pi_2^{0^2}]} \\
&\qquad-\frac{\pi_1^{0^2}\lambda(1-\frac{\lambda}{\mu^0})(\sigma^2-\sigma_\epsilon^2)}{(1-\lambda)^2 m\,[\alpha\pi_1^{0^2}+(1-\alpha)\pi_2^{0^2}]^2}\ge 0.
\end{aligned}
$$

Since $[\alpha\pi_1^0+(1-\alpha)\pi_2^0]^2-[\alpha\pi_1^{0^2}+(1-\alpha)\pi_2^{0^2}]=-\alpha(1-\alpha)(\pi_1^0-\pi_2^0)^2\le 0$,

$$
\frac{\sigma_\epsilon^2}{(1-\lambda)m\,[\alpha\pi_1^0+(1-\alpha)\pi_2^0]^2}\ge\frac{\sigma_\epsilon^2}{(1-\lambda)m\,[\alpha\pi_1^{0^2}+(1-\alpha)\pi_2^{0^2}]},
$$

implying that a sufficient condition for $V^*_{2,GMM}(\gamma) - V^*_{A,2}(\gamma) \geq 0$ is $\sigma^2 \leq \sigma^2_\epsilon$, the same condition that is necessary and sufficient for $V^*_{1,GMM}(\gamma) - V^*_{A,1}(\gamma) \geq 0$.

Part (iii). Variance comparisons for LRF. Here, $\pi^0_1 = \pi^0_2 = \pi^0$, and because there is no threshold in the RF, wlog we let $\rho^0 = \gamma_{max} \iff \mu^0 = 1$, and we calculate the variances from $\gamma \leq \rho^0 = \gamma_{max}$. Plugging these into the results of part (ii), we have:

$$\begin{aligned}
V^*_{1,GMM}(\gamma) - V^*_{A,1}(\gamma) &= -\frac{(1-\lambda)(\sigma^2 - \sigma^2_\epsilon)}{\lambda \pi^{02} m} \geq 0 \iff \sigma^2 \leq \sigma^2_\epsilon \\
V^*_{2,GMM}(\gamma) - V^*_{A,2}(\gamma) &= \frac{\sigma^2_\epsilon}{(1-\lambda)m\ \pi^{02}} - \frac{\sigma^2_\epsilon}{(1-\lambda)m\ \pi^{02}} - \frac{\lambda(\sigma^2 - \sigma^2_\epsilon)}{(1-\lambda)m\ \pi^{02}} \\
&= -\frac{\lambda(\sigma^2 - \sigma^2_\epsilon)}{(1-\lambda)m\ \pi^{02}} \geq 0 \iff \sigma^2 \leq \sigma^2_\epsilon.
\end{aligned}$$

Part (iv). We obtain the claim by plugging in $\gamma = \rho^0$ into the variance expressions of Lemma 9. □