Testing for a Threshold in Models with Endogenous Regressors

Mario P. Rothfelder* Otilia Boldea†

April 22, 2015

PRELIMINARY AND INCOMPLETE – DO NOT CITE OR CIRCULATE!

In the presence of endogenous regressors, threshold parameters are usually estimated by 2SLS. In this paper, we propose two tests for a threshold, both based on 2SLS estimation: a sup LR test and a sup Wald test. We derive their asymptotic distributions for a linear and a threshold reduced form. In both cases, the asymptotic distributions are non-pivotal, and we propose obtaining p-values via methods similar to the wild bootstrap.

We show through simulations that our 2SLS-based tests behave well in large samples, both in terms of size and power, whether we have a linear reduced form or not. In small samples, both sup Wald tests are oversized, while the sup LR is correctly sized. The size-adjusted power of all tests in small samples is nontrivial for a moderate threshold size, both for a linear and for a threshold reduced form. We illustrate our tests via an empirical application in which we investigate whether the government spending multiplier is larger in regimes where the nominal interest rate is close to the zero lower bound. All threshold tests suggest no evidence of an interest rate threshold.

Keywords: 2SLS, GMM, Thresholds, Testing, Bootstrap
JEL Classification: C12, C13, C21, C22

*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, output growth, bank profits, asset prices, exchange rates, interest rates, to mention just a few. See Hansen (2011) for a survey of many economic applications.

Pioneered by Howell Tong - see e.g. Tong (1990), threshold models with exogenous regressors have been widely studied for more than two decades, and their asymptotic theory is well known - 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, and Seo and Linton (2007), Lee et al. (2011) for discrete choice models.

Regressor exogeneity is known to be violated in many economic applications. Nevertheless, papers on threshold regression with endogeneity are relatively scarce. They were pioneered by Caner and Hansen (2004), who show that when the threshold variable is exogenous, but the regressors are endogenous, the threshold parameter can be estimated by maximizing a two stage least squares (2SLS) criterion over the values of the threshold variable encountered in the sample. Caner and Hansen (2004) also propose a 2SLS likelihood ratio (LR) test of the null hypothesis that threshold is equal to a certain fixed value.

In this paper, we are interested in a different null hypothesis, of no threshold against the alternative of one threshold. Since threshold estimation is done via 2SLS, we propose a 2SLS-based LR test for the presence of a threshold. For each possible threshold value γ, we compute an LR test of the null of no threshold against the alternative of one threshold at γ, and the resulting test statistic, sup LR, maximizes the sequence of tests over γ. This test is the 2SLS equivalent of the sup LR test of Davies (1977) and Hansen (1996), and is also known as the sup F test in the break-point literature.

We also propose a 2SLS counterpart that is robust to heteroskedasticity, the sup Wald test. To our knowledge, this is the first paper that proposes such tests and derives their asymptotic distributions in threshold models with endogenous regressors.

As in Caner and Hansen (2004), we consider two cases: case 1, the first-stage of the 2SLS regression (call it reduced form) is a linear model, and case 2, it is a threshold model itself, possibly with a different threshold parameter than the second stage. We show that the null asymptotic distributions of the sup LR and sup Wald tests depend on the data and on the case considered. In both cases, p-values can be computed by straightforward simulations. Unlike in the classical hypotheses tests, when testing for an unknown threshold, heteroskedasticity-robust sup Wald test do not have a pivotal null distribution. That means that correcting for heteroskedasticity does not necessarily result in better size properties for the sup Wald test compared to the sup LR test.

The properties of both tests are studied via a simulation study. p-values are generated via a simulation method that resembles wild bootstrap and is similar to the one proposed in Caner and Hansen (2004) for GMM sup Wald tests, and in Hansen (1996) for OLS sup Wald tests. We find that even in the presence of heteroskedasticity, the 2SLS sup LR has better size than both the 2SLS sup Wald and the GMM sup Wald proposed in

Caner and Hansen (2004). This holds for both linear and threshold reduced forms. As the sample size grows large, all tests approach their nominal sizes. They all exhibit good power properties for moderate threshold shifts.

Our paper is closely related to two papers in the break-point literature - Hall et al. (2012) and Boldea et al. (2014). Both papers study 2SLS-based sup LR and sup Wald tests for a break, the first one for a linear reduced form, the second one for a reduced form with a break. The asymptotic distributions for the break-point tests are pivotal in the first paper and depend on the break in the reduced form 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 reduced form.

It should be noted that we allow for endogenous regressors, but not for endogenous threshold variables. For the latter, see Kourtellos et al. (2013). 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 (2013), 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 power in $O(T^{-1/2})$ neighborhoods.

We illustrate our tests via an empirical application in which we investigate whether the 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 reduced form threshold, but all threshold tests for the equation of interest suggest no evidence of an interest rate threshold that would induce different government spending multiplier regimes.

This paper is organized as follows. In Section 2 we present the model as well as estimation and testing strategies. Sections 3 and 4 provide an asymptotic framework for the proposed test statistics. Section 5 compares our tests to the existing GMM sup Wald test. In Section 6 we illustrate the small sample properties of all the tests via simulation. Section 7 contains the empirical application. Section 8 conclude and Section 9 contains all tables and graphs. All proofs are relegated to the Appendix.

2. Model

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

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

Here, y_t is the dependent variable, z_t is a $p_1 \times 1$-vector of endogenous 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 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 $\gamma^0 \in \Gamma$ is the threshold parameter where $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ is its compact support. For testing,

Γ is assumed to be a strict compact subset of the support of the random variable q_t. Note that the second equation is just a more compact way of writing the first, with $w_t = (z_t^\top, x_{1t}^\top)^\top$ and $\theta_i^0 = (\theta_{iz}^{0\top}, \theta_{ix}^{0\top})^\top$ being $p \times 1$-vectors of the structural form slope parameters, for $i = 1, 2$. We assume that z_t is endogenous, that is, $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[z_t \epsilon_t] \neq 0$, and strong instruments x_t are available.

As in Caner and Hansen (2004), we consider two different specifications for the reduced form equation: a linear reduced form (LRF), given by

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

and a threshold reduced form (TRF) given by

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

In both specifications for the reduced form, $x_t = (x_{1t}^\top, x_{2t}^\top)^\top$ is a $q \times 1$-vector containing x_{1t}, with $q \geq p$, so let $q = p_2 + q_1$. That is, x_t includes the exogenous variables from the structural form (or the equation of interest) above. Here, Π^0, Π_1^0 and Π_2^0 are $q \times p_1$-matrices of the reduced form slope parameters and $\rho^0 \in \Gamma$ is the reduced form threshold parameter, possibly different than γ^0, with the same support Γ.

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.

Write the equations above in matrix form. To do so, define the partitioned matrices:

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

Also, let Y, X, Z, ϵ and u be the matrices stacking observations $t = 1, \ldots, T$. Then the linear reduced form can be written as:

$$Z = X\Pi^0 + u,$$

and the threshold reduced form as:

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

The structural form equation (the equation of interest), in the presence of a threshold parameter γ^0, is:

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

3. Tests for Linear Reduced Form

We are interested in testing for the presence of a threshold, $\mathbb{H}_0 : \theta_1^0 = \theta_2^0$. The fact that γ^0 is usually estimated by a 2SLS procedure in parametric models - see Caner and Hansen

(2004) - motivates us to look at 2SLS-based test statistics, rather than the GMM-based test statistic for the same hypothesis in Caner and Hansen (2004). We defer an extensive discussion of the relative advantages of the 2SLS-based versus the GMM-based tests to Section 5. For now, it suffices to notice that the GMM-based tests ignore information about the reduced form (whether it is a threshold or not), while the 2SLS-based tests use this information. Using valid information can result in an increase in efficiency of parameter estimates and accuracy of 2SLS-based tests. This intuition is confirmed by simulations in Section 6.

We propose two test statistics for $\mathbb{H}_0$, the sup LR and the sup Wald tests. For both tests, we assume in this section a linear reduced form (LRF):

$$Z = X\Pi^0 + u, \tag{3.1}$$

which we estimate by ordinary least-squares (OLS), so $\hat{\Pi} = (X^\top X)^{-1} X^\top Z$. We estimate the structural form:

$$Y = W_1^{\gamma^0}\theta_1^0 + W_2^{\gamma^0}\theta_2^0 + \epsilon, \tag{3.2}$$

with γ^0 replaced by each possible $\gamma \in \Gamma$. Note that γ^0 is a nuisance parameter that shows up only under the alternative $\mathbb{H}_A : \theta_1^0 - \theta_2^0 \neq 0$.

As mentioned, for constructing the test statistics, (3.2) is estimated by 2SLS. To that end, we define the predicted endogenous regressors as:

$$\hat{Z} = X\hat{\Pi},\ \hat{W} = \left(\hat{Z}, X_1\right), \tag{3.3}$$

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

In the second stage, for each $\gamma \in \Gamma$, we obtain the 2SLS estimators of θ_1^0, θ_2^0:

$$\hat{\theta}_1^\gamma = \left(\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma\right)^{-1}\left(\hat{W}_1^{\gamma\top}Y\right) \tag{3.4}$$

$$\hat{\theta}_2^\gamma = \left(\hat{W}_2^{\gamma\top}\hat{W}_2^\gamma\right)^{-1}\left(\hat{W}_2^{\gamma\top}Y\right). \tag{3.5}$$

Note that because the predicted regressors $\hat{W}$ are already partitioned according to $\mathbb{1}_{\{q_t \leq \gamma\}}$, we have $\hat{W}_i^{\gamma\top}Y = \hat{W}_i^{\gamma\top}Y_i$, for $i = 1, 2$.

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

$$\sup_{\gamma\in\Gamma} LR_{T,LRF}^{2SLS}(\gamma) = \sup_{\gamma\in\Gamma} \frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/(T-2p)}, \tag{3.6}$$

where SSR_0 is the 2SLS sum of squared residuals under the null hypothesis and $SSR_1(\gamma)$ the 2SLS sum of squared residuals under the alternative, that is:

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

where $\hat{\theta} = (\hat{W}^\top\hat{W})^{-1}\hat{W}^\top Y$ is the full-sample 2SLS estimator.

Note that if γ was known, this would be the usual LR test. The only difference is that because γ is unknown, we consider the maximum of a sequence of test statistics depending on γ, the sup LR test.

A similar OLS-based sup LR test was proposed for threshold models with exogenous regressors in Hansen (1996). A scaled version of it 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.

For testing $\mathbb{H}_0$, we also propose the heteroskedasticity-robust version of this test, the sup Wald test:

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma) = \sup_{\gamma\in\Gamma} \left[\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma\right]^\top \left[\widehat{\mathrm{Var}}(\gamma)\right]^{-1} \left[\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma\right], \tag{3.7}$$

where:

$$\begin{aligned}
\widehat{\mathrm{Var}}(\gamma) &= \widehat{\mathrm{Var}}(\hat{\theta}_1^\gamma) + \widehat{\mathrm{Var}}(\hat{\theta}_2^\gamma) \\
\widehat{\mathrm{Var}}(\hat{\theta}_i^\gamma) &= (\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1}\hat{H}_i(\gamma)(\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1} \text{ for } i = 1, 2, \\
\hat{H}_1(\gamma) &= \sum_{t=1}^T \hat{e}_t^2 \hat{w}_t\hat{w}_t^\top \mathbb{1}_{\{q_t\le\gamma\}},\ \hat{H}_2(\gamma) = \sum_{t=1}^T \hat{e}_t^2 \hat{w}_t\hat{w}_t^\top \mathbb{1}_{\{q_t>\gamma\}}.
\end{aligned}$$

Note that here, $\hat{e}_t = y_t - \hat{w}_t^\top\hat{\theta}_1^\gamma\mathbb{1}_{\{q_t\le\gamma\}} - \hat{w}_t^\top\hat{\theta}_2^\gamma\mathbb{1}_{\{q_t\le\gamma\}}$ are estimating the true 2SLS error terms $e_t = \epsilon_t + (w_t - \hat{w}_t)^\top\theta_1^0\mathbb{1}_{\{q_t\le\gamma\}} + (w_t - \hat{w}_t)^\top\theta_2^0\mathbb{1}_{\{q_t>\gamma\}}$, and not the original (structural) errors ϵ_t. We do not robustify against autocorrelation, because errors ϵ_t and u_t are assumed to be martingale differences. As mentioned in Caner and Hansen (2004, pp. 815), the martingale difference assumption is needed for identification of nonlinear models such as threshold models, and cannot be easily dropped.

Note that the sup is taken over $\gamma \in \Gamma$, a strict compact subset of the support of the random variable q_t, which we take to be included in $[\gamma_{\min}, \gamma_{\max}]$. The end-points of this interval can be (minus and plus) infinite. But we take Γ to be bounded away from the minimum and maximum q_t values observed in the sample. In practice, the end-points of Γ are the upper and lower 15% quantiles of the empirical distribution of q_t, to have enough observations to compute $\hat{\theta}_i^\gamma$.

The sup Wald test above was proposed in the break-point literature with endogenous regressors by Hall et al. (2012). However, in Hall et al. (2012), its asymptotic distribution is pivotal in the case of a LRF. Below, we show that for thresholds, the asymptotic distributions are non-pivotal even for LRF. To see that, 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.

We impose similar assumptions to 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 $\mathfrak{F}_t = \sigma\{x_{t-s}, v_{t-s-1}, q_{t-s}|s \ge 0\}$.

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* $b > a$*,*

$$\sup_t \mathbb{E}\|x_t\|_2^{6b} < \infty, \ \sup_t \mathbb{E}\|v_t\|_2^{4b} < \infty,$$

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*

$$\mathbb{E}[v_t v_t^\top] = \Sigma,$$

which is positive definite.

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

As discussed before, A.1.1 is needed for threshold models, and it excludes autocorrelation in the errors. However, general linear dynamics are allowed in both the equation of interest and the reduced form. A.1.2 is standard for time series and is trivially satisfied for many 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 reduced form 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 easier to perform, and in practice it is likely that the researcher would know the threshold. A.1.5 allows the errors to be conditional heteroskedastic and finally, A.1.6 is the usual strong instrument assumption on X.

Although the assumptions we use are standard, to our knowledge, this is the first paper that derives the asymptotic distribution of the sup LR and sup Wald tests above. To write the asymptotic distribution, define

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

as $q\times(p_1+1)$-matrices where all columns are $q\times 1$ zero mean Gaussian processes such that 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 $\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)]$, respectively, and $\mathcal{GP}_{\text{mat}} = \mathcal{GP}_{\text{mat},1}(\gamma_{\max})$. Next, let

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

be the augmented matrix of the reduced form 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 ($p_2 + q_1 = q$). Thus, $x_{1t} = Sx_t$. Define the matrices

$$C_1(\gamma) = A^0 M_1(\gamma) A^{0\top}, \quad C = C_1(\gamma_{\max}) = A^0 M A^{0\top} \quad \text{and} \quad D_1(\gamma) = C^{-1} C_1(\gamma),$$

and the Gaussian process:

$$\mathcal{B}_1(\gamma) = A^0 \left[\mathcal{GP}_{\text{mat},1}(\gamma)\ \tilde{\theta}_z^0 - M_1(\gamma) M^{-1} \mathcal{GP}_{\text{mat}}\ \check{\theta}_z^0 \right]$$

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 $\mathcal{B} = \mathcal{B}_1(\gamma_{\max})$ and $C_2(\gamma) = C - C_1(\gamma)$.
Then the null distribution of both tests with LRF is stated below.

Theorem 1 (Asymptotic Distributions LRF)**.** *Let Z be generated by (3.1), Y be generated by (3.2), and $\hat{Z}$ be calculated by (3.3). Then*[1]*, under $\mathbb{H}_0$ and Assumption A.1,*
(i)

$$\sup_{\gamma\in\Gamma} LR_{T,LRF}^{2SLS}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\mathcal{E}^\top(\gamma) C_2(\gamma) C^{-1} C_1(\gamma) \mathcal{E}(\gamma)}{\sigma^2},$$

where $\sigma^2 = \sigma_\epsilon^2 + 2\Sigma_{\epsilon,u}^\top \theta_z^0 + \theta_z^{0\top} \Sigma_u \theta_z^0$, and
(ii)

$$\sup_{\gamma\in\Gamma} W_{T,LRF}^{2SLS}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \mathcal{E}^\top(\gamma)[C_1^{-1}(\gamma)H_1(\gamma)C_1^{-1}(\gamma) + C_2^{-1}(\gamma)H_2(\gamma)C_2^{-1}(\gamma)]^{-1}\mathcal{E}(\gamma)$$

where

$$H_1(\gamma) = A^0 \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2 \mathbb{1}_{\{q_t \le \gamma\}}] A^{0\top}$$

and

$$H_2(\gamma) = A^0 \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2 \mathbb{1}_{\{q_t > \gamma\}}] A^{0\top}.$$

We note that in both cases, the asymptotic distribution depends on second moment functionals of the data and the parameters in the reduced form. Therefore, the asymptotic distributions are not pivotal, but can be simulated via similar techniques as proposed in Caner and Hansen (2004).

We also derive the distribution of sup LR and sup Wald under the usual conditional homoskedasticity assumption:

Assumption A.2.

$$\mathbb{E}[v_t v_t^\top | x_t, q_t] = \Sigma = \begin{pmatrix} \Sigma_\epsilon & \Sigma_{\epsilon,u}^\top \\ \Sigma_{\epsilon,u} & \Sigma_u \end{pmatrix}.$$

As illustrated below, we still obtain non-pivotal asymptotic distributions even when the errors are conditional homoscedastic. But, as expected, under conditional homoscedasticity, the Wald and LR tests and asymptotically equivalent.

[1] $\Longrightarrow$ denotes weak convergence in the Skorokhod-metric.

Corollary 1 (to Theorem 1)**.** *Let Z be generated by (3.1), Y be generated by (3.2), and $\hat{Z}$ be calculated by (3.3). Then, under $\mathbb{H}_0$, and Assumptions A.1 and A.2,*
(i)

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,LRF}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\tilde{\mathcal{E}}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\tilde{\mathcal{E}}(\gamma)}{\sigma^2}$$

where $\sigma^2 = \sigma^2_\epsilon + 2\Sigma^\top_{\epsilon,u}\theta^0_z + \theta^{0\top}_z\Sigma_u\theta^0_z$, and
(ii)

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,LRF}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\tilde{\mathcal{E}}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\tilde{\mathcal{E}}(\gamma)}{\sigma^2}$$

where $\widetilde{\mathcal{GP}}_{\text{mat},1}(\gamma)$ is a $q\times(p_1+1)$-matrix where all columns are independent $q\times 1$ zero mean Gaussian processes with covariance kernel[2] *$M_1(\gamma)$, Q is the principal square 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 $\tilde{\mathcal{B}}_1(\gamma) = A^0[\widetilde{\mathcal{GP}}_{\text{mat},1}(\gamma)Q\tilde{\theta}^0_z - M_1(\gamma)M^{-1}\widetilde{\mathcal{GP}}_{\text{mat}}Q\check{\theta}^0_z]$.*

Only when the regressors and threshold variables are independent, the asymptotic distributions are pivotal.

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 excludes, for example, cases in which the threshold variable q_t is part of the set of 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.

Corollary 2 (to Theorem 1)**.** *Let Z be generated by (3.1), Y be generated by (3.2), and $\hat{Z}$ be calculated by (3.3).Then, under $\mathbb{H}_0$ and Assumptions A.1, A.2, and A.3,*
(i),

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

and (ii),

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

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

[2] 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.

Thus, under Assumptions A.2 and A.3, the asymptotic distribution is pivotal, and it is identical to the one obtained for break-point models - 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 $q_t = t/T$.[3] Critical values for these pivotal distributions can be found in Andrews (1993) and Bai and Perron (1998). However, note that this result only holds for conditional homoscedastic errors and threshold variables independent of regressors, an unlikely case in practice.

4. Tests for Threshold Reduced Form

For this section, we assume that the reduced form has a threshold ρ^0 (TRF), possibly different than the structural form threshold γ^0. The test statistics are computed taking into account the TRF when computing the first stage of the 2SLS estimation.

Let the threshold reduced form (TRF) be:

$$Z = X_1^{\rho^0}\Pi_1^0 + X_2^{\rho^0}\Pi_2^0 + u, \tag{4.1}$$

and the structural form be as before:

$$Y = W_1^{\gamma^0}\theta_1^0 + W_2^{\gamma^0}\theta_2^0 + \epsilon. \tag{4.2}$$

We estimate the TRF via least-squares threshold methods, as in e.g. Caner and Hansen (2004):

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

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 (4.1) for a given ρ:

$$\hat{\Pi}_1(\rho) = \left(X_1^{\rho\top}X_1^\rho\right)^{-1} X_1^{\rho\top}Z$$
$$\hat{\Pi}_2(\rho) = \left(X_2^{\rho\top}X_2^\rho\right)^{-1} X_2^{\rho\top}Z.$$

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

Then we let

$$\hat{Z} = \hat{\Pi}_1 Z_1^{\hat{\rho}} + \hat{\Pi}_2 Z_2^{\hat{\rho}},\ \hat{W} = \left(\hat{Z}, X_1\right), \tag{4.3}$$

where $X_1^{\hat{\rho}}$ and $X_2^{\hat{\rho}}$ are defined as above with $\hat{\rho}$ instead of ρ. The computation of $\hat{\theta}_1^\gamma, \hat{\theta}_2^\gamma$ is then the same as in (3.4) and (3.5), but with $\hat{W}$ defined here. Similarly, the new test statistics:

$$\sup_{\gamma\in\Gamma} LR_{T,TRF}^{2SLS}(\gamma),\ \sup_{\gamma\in\Gamma} W_{T,TRF}^{2SLS}(\gamma)$$

are computed exactly as their LRF counterparts, but with $\hat{W}$ defined in (4.3). We assume:

[3] Note, however, that break-point asymptotics cannot be obtained as a special case of threshold asymptotics, because generally, Assumptions A.1.2 and A.1.4 do not hold for dynamic break-point models.

Assumption A.4 (Identifiability 1)**.** $\Pi_1^0 \neq \Pi_2^0$.

Assumption A.4 states that there is a (large) threshold effect, expressed by different slope parameters Π_1^0 and Π_2^0, present in the reduced form equation. For stating the asymptotic distributions, similar to A^0 in the previous section, we define

$$A_1^0 = [\Pi_1^0, S^\top]^\top \quad \text{and} \quad A_2^0 = [\Pi_2^0, S^\top]^\top \tag{4.4}$$

as the augmented matrices of the reduced form slope parameters. Also, define the matrices

$$C_{A,1}(\gamma) = A_1^0 M_1(\gamma \wedge \rho^0) A_1^{0\top} + A_2^0 \left[M_1(\gamma) - M_1(\gamma \wedge \rho^0)\right] A_2^{0\top}, \tag{4.5}$$

and

$$\begin{aligned} C_A &= C_{A,1}(\gamma_{\max}) \\ &= A_1^0 M_1(\rho^0) A_1^{0\top} + A_2^0 M_2(\rho^0) A_2^{0\top}. \end{aligned} \tag{4.6}$$

Thus, the TRF analogs to the LRF processes $B_1(\gamma)$ and $\mathcal{E}(\gamma)$ are:

$$\begin{aligned} \mathcal{B}_{A,1}(\gamma) = {} & A_1^0 \left[\mathcal{GP}_{\text{mat},1}(\gamma \wedge \rho^0)\tilde{\theta}_z^0 - M_1(\gamma \wedge \rho^0) M_1^{-1}(\rho^0) \mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right] \\ & + A_2^0 \left[\left(\mathcal{GP}_{\text{mat},1}(\gamma) - \mathcal{GP}_{\text{mat},1}(\gamma \wedge \rho^0)\right)\tilde{\theta}_z^0\right] \\ & - A_2^0 \left[\left(M_1(\gamma) - M_1(\gamma \wedge \rho^0)\right) M_2^{-1}(\rho^0) \mathcal{GP}_{\text{mat},2}(\rho^0)\right]. \end{aligned} \tag{4.7}$$

and

$$\mathcal{E}_A(\gamma) = C_{A,1}^{-1}(\gamma)\mathcal{B}_{A,1}(\gamma) - C_{A,2}^{-1}(\gamma)\mathcal{B}_{A,2}(\gamma) \tag{4.8}$$

where

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

with

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

Last, the TRF analogs to the LRF covariance matrices $H_1(\gamma)$ and $H_2(\gamma)$ are:

$$\begin{aligned} H_{A,1}(\gamma) = {} & A_1^0 \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2 \mathbb{1}_{\{q_t \leq \gamma \wedge \rho^0\}}] A_1^{0\top} \\ & + A_2^0 \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2 (\mathbb{1}_{\{q_t \leq \gamma\}} - \mathbb{1}_{\{q_t \leq \gamma \wedge \rho^0\}})] A_2^{0\top} \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} H_{A,2}(\gamma) = {} & A_2^0 \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2 \mathbb{1}_{\{q_t \geq \gamma \vee \rho^0\}}] A_2^{0\top} \\ & + A_1^0 \mathbb{E}[x_t x_t^\top (\epsilon_t + u_t^\top \theta_z^0)^2 (\mathbb{1}_{\{q_t > \gamma\}} - \mathbb{1}_{\{q_t \geq \gamma \vee \rho^0\}})] A_1^{0\top} \end{aligned} \tag{4.10}$$

The more complicated expressions in this case stem from the fact that the relative location between each potential structural form threshold parameter γ and the true reduced form threshold parameter ρ^0 influences the asymptotic distribution of our tests.

Theorem 2 (Asymptotic Distributions TRF). *Let Z be generated by (4.1), Y be generated by (4.2), and $\hat{Z}$ be calculated by (4.3). Then, under $\mathbb{H}_0$, and Assumptions A.1 and A.4,*
(i)

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,TRF}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\mathcal{E}_A^\top(\gamma)C_{A,2}(\gamma)C_A^{-1}C_{A,1}(\gamma)\mathcal{E}_A(\gamma)}{\sigma^2},$$

and (ii)

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,TRF}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \mathcal{E}_A^\top(\gamma)\Big[C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma) + C_{A,2}^{-1}(\gamma)H_{A,2}(\gamma)C_{A,2}^{-1}(\gamma)\Big]^{-1}\mathcal{E}_A(\gamma).$$

Under conditional homoscedasticity, we show below that the classical asymptotic equivalence between Wald and LR tests still holds, even for a TRF. To this extend 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

$$\begin{aligned}\tilde{\mathcal{B}}_{A,1}(\gamma) = {} & A_1^0\Big[\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma\wedge\rho^0)Q\tilde{\theta}_z^0 - M_1(\gamma\wedge\rho^0)M_1^{-1}(\rho^0)\tilde{\mathcal{GP}}_{\text{mat},1}(\rho^0)Q\check{\theta}_z^0\Big] \\ & + A_2^0\Big[\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)Q - \tilde{\mathcal{GP}}_{\text{mat},1}(\gamma\wedge\rho^0)Q\tilde{\theta}_z^0\Big] \\ & + A_2^0\Big[(M_1(\gamma) - M_1(\gamma\wedge\rho^0))M_2^{-1}(\rho^0)\tilde{\mathcal{GP}}_{\text{mat},2}(\rho^0)\Big]\end{aligned}$$

where $\widetilde{\mathcal{GP}}_{\text{mat},1}(\gamma)$ is a $q\times(p_1+1)$ matrix where all columns are independent $q\times 1$ zero mean Gaussian processes with covariance kernel $M_1(\gamma)$.[4] Then we have:

Corollary 3 (to Theorem 2). *Let Z be generated by (4.1), Y be generated by (4.2), and $\hat{Z}$ be calculated by (4.3). Then, under $\mathbb{H}_0$, and Assumptions A.1, A.2 and A.4,*
(i)

$$\sup_{\gamma\in\Gamma} LR^{2SLS}_{T,TRF}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\tilde{\mathcal{E}}_A^\top(\gamma)C_{A,2}(\gamma)D_{A,1}(\gamma)\tilde{\mathcal{E}}_A(\gamma)}{\sigma^2},$$

and (ii)

$$\sup_{\gamma\in\Gamma} W^{2SLS}_{T,TRF}(\gamma) \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\tilde{\mathcal{E}}_A^\top(\gamma)C_{A,2}(\gamma)D_{A,1}(\gamma)\tilde{\mathcal{E}}_A(\gamma)}{\sigma^2}.$$

As in Boldea et al. (2014), the asymptotic distributions are non-pivotal, and don't simplify under Assumption A.3. This is not an issue in practice, because the p-values can still be obtained by simulation techniques, as we discuss in Section 6.

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

5. Comparison to sup Wald GMM

In contrast to our paper, Caner and Hansen (2004) propose testing for a threshold using a GMM-based Wald test. In particular, instead of using 2SLS, they use two-step GMM estimation. Specifically, for each $\gamma \in \Gamma$, one estimates θ_i^0 in (4.2), for $i = 1, 2$, as:

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

with estimated variance-covariances:

$$\widehat{\mathrm{Var}}\left(\hat{\theta}_{i,GMM}^{\gamma}\right) = \left(W_i^{\gamma\top} X_i^{\gamma} \hat{\Omega}_{i,GMM}^{-1}(\gamma) X_i^{\gamma\top} W_i^{\gamma}\right)^{-1},$$

and estimated long-run variances:

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

where $\hat{\epsilon} = y - \hat{W}_1^{\gamma} \tilde{\theta}_{1,GMM}(\gamma) - \hat{W}_2^{\gamma} \tilde{\theta}_{2,GMM}(\gamma)$, and $\tilde{\theta}_{i,GMM}(\gamma)$ are some preliminary first step GMM estimators of (4.2) for a given γ.

With these definitions, the sup GMM Wald test statistic in Caner and Hansen (2004) for $\mathbb{H}_0$, at each γ, is:

$$W_{T,\mathrm{GMM}}(\gamma) = [\hat{\theta}_{1,GMM}^{\gamma} - \hat{\theta}_{2,GMM}^{\gamma}]^{\top} \{\widehat{\mathrm{Var}}[\hat{\theta}_{1,GMM}^{\gamma}] + \widehat{\mathrm{Var}}[\hat{\theta}_{2,GMM}^{\gamma}]\}^{-1} [\hat{\theta}_{1,GMM}^{\gamma} - \hat{\theta}_{2,GMM}^{\gamma}].$$

The sup Wald test is then $\sup_{\gamma \in \Gamma} W_{T,GMM}(\gamma)$. Define

$$\Omega_{1,GMM}(\gamma) = \mathbb{E}[x_t x_t^{\top} \epsilon_t^2 \mathbb{1}_{\{q_t \leq \gamma\}}], \qquad \Omega_{2,GMM}(\gamma) = \mathbb{E}[x_t x_t^{\top} \epsilon_t^2 \mathbb{1}_{\{q_t > \gamma\}}],$$

and $V_i(\gamma) = \left[N_i(\gamma) \Omega_{i,GMM}^{-1}(\gamma) N_i^{\top}(\gamma)\right]^{-1}$, $N_i(\gamma) = \mathrm{plim}[T^{-1} W_i^{\gamma\top} X_i^{\gamma}]$. Also, let $\mathcal{GP}_1(\gamma)$ be a $q \times 1$ zero mean Gaussian process with covariance kernel $\mathbb{E}[\mathcal{GP}_1(\gamma_1) \mathcal{GP}_1^{\top}(\gamma_2)] = \Omega_1(\gamma_1 \wedge \gamma_2)$, and $\mathcal{GP} = \lim_{\gamma \to \infty} \mathcal{GP}_1(\gamma)$. Then, Caner and Hansen (2004) show:

Theorem 3 (Asymptotic distribution sup Wald GMM)**.** *Let Z be generated by (3.1) or (4.1), and Y be generated by (4.2). Then, under $\mathbb{H}_0$, and Assumptions A.1 and A.4,*

$$\begin{aligned}\sup_{\gamma \in \Gamma} W_T^{GMM}(\gamma) \Longrightarrow \sup_{\gamma \in \Gamma} & \left[V_1(\gamma) N_1(\gamma) \Omega_1^{-1}(\gamma) \mathcal{GP}_1(\gamma) - V_2(\gamma) N_2(\gamma) \Omega_2^{-1}(\gamma) \mathcal{GP}_2(\gamma)\right]^{\top} \\ & \times [V_1(\gamma) + V_2(\gamma)]^{-1} \\ & \times \left[V_1(\gamma) N_1(\gamma) \Omega_1^{-1}(\gamma) \mathcal{GP}_1(\gamma) - V_2(\gamma) N_2(\gamma) \Omega_2^{-1}(\gamma) \mathcal{GP}_2(\gamma)\right].\end{aligned}$$

The proof is in Caner and Hansen (2004), and presumably this test is proposed there for two reasons.

The first reason may be that the asymptotic distribution of the GMM-based test does not depend on whether the reduced form is a linear or a threshold model, while our

asymptotic distributions do. This may be fine if the tests behave well and the only purpose is to test for a threshold. However, researchers are only interested in pre-testing for a threshold so that they can later estimate the structural form, and so if the test rejects, one still has to estimates the threshold parameter γ^0 by 2SLS. Moreover, two-step GMM estimators are routinely computed with first-step 2SLS estimates. In this case, a decision between using LRF and TRF may still be desirable.

The second reason may be that the GMM estimates are known to be more efficient compared to 2SLS in regular (smooth) models. So for example, for a linear regression, if a sample is used to pre-estimate a reduced form to compute 2SLS, and the same sample is used for GMM estimates, then the latter are asymptotically more efficient. However, in this paper we do not have a smooth model. We don't use the same sample to compute 2SLS and GMM estimates. For example, for LRF, we use the whole sample to compute the first-stage of the 2SLS estimators $\hat{\theta}_i^{\gamma}$, but only a sub-sample to compute GMM estimates $\hat{\theta}_{i,GMM}^{\gamma}$. Similarly, for TRF, we use a different sample to compute the first-stage of the 2SLS estimates then for $\hat{\theta}_{i,GMM}^{\gamma}$. Using different sub-samples makes the connection between 2SLS and GMM estimates break down; they are no longer a special case of the other, and so this motivates us to look at both GMM- and 2SLS-based tests. Moreover, as emphasized in Antoine and Boldea (2014) and Magnusson and Mavroeidis (2013), knowledge of reduced form discontinuities is valid information that can lead to more efficient estimation and testing.

Note that the 2SLS and GMM-based tests in general have different asymptotic distributions; the 2SLS distributions are not special cases of the GMM ones. But under Assumption A.2 and LRF, all the distributions are the same.

Corollary 4 (Corollary to Theorem 3)**.** *Let Z be generated by* (3.1) *or* (4.1)*, and Y be generated by* (4.2)*. Then, under $\mathbb{H}_0$, and Assumptions A.1, A.3 and A.4,*

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

This corollary shows that in the special case of independence of the regressors and threshold variables, the sup Wald GMM asymptotic distribution is pivotal. For our tests, the same holds for linear reduced form. For threshold reduced form, we observe that the critical values are lower, and it may be because our tests are using additional valid information about the TRF.

6. Simulations

In this chapter, we investigate the small sample properties of our proposed test statistics. To do so, we describe how to simulate the p-values of our tests based on the asymptotic distribution, and we describe our data generating processes in 6.1. In sections 6.2 and 6.3 we present results on empirical sizes and powers of our tests.

6.1. P-value Simulation and Data Generating Process

First, we lay out our approach to simulate empirical sizes and powers. Afterwards, we present the data generating process (DGP).

P-value Simulation As shown in Sections 3–4, the asymptotic distributions of the proposed test statistics are non-standard and thus, need to be simulated. For that, we follow Hansen (1996): Because of almost sure convergence of the moment functionals the 'middle' parts of the proposed test statistics (i.e. all except $\mathcal{E}(\gamma)$ for a LRF, respectively $\mathcal{E}_A(\gamma)$ for a TRF) are replaced by their sample analogs and kept fixed conditionally on the sample. So, e.g. we replace $M_1(\gamma)$ by $\frac{1}{T}\sum_{t=1}^{T} x_t x_t^\top \mathbb{1}_{\{q_t \leq \gamma\}}$. Moreover, unknown parameters are replaced by their 2SLS estimators as described in sections 3 and 4. However, the Gaussian processes need to be simulated. We approximate them by

$$\widehat{\mathcal{GP}}_{\text{mat},1}(\gamma) = T^{-1/2} X_1^{\gamma\top} \hat{v}^* \approx \mathcal{GP}_{\text{mat},1}(\gamma) \tag{6.1}$$

where $\hat{v}^* = \hat{v} \odot \eta$ with $\eta = (\eta_t, ..., \eta_t)_{t=1,...,T}$, $\eta_t \overset{iid}{\sim} \mathcal{N}(0,1)$, being a $T \times p_1$-matrix, $\hat{v} = (\hat{\epsilon}, \hat{u}^\top)^\top$ where $\hat{\epsilon}$ and $\hat{u}$ are the estimated original (structural) and reduced form errors, and $\odot$ denotes the Hadamard product.

The approximation in (6.1) is justified by Hansen (1996, Theorem 2) and Lemma 1 in the Appendix (in the sense that the approximation is a weak convergence result when the bootstrap sample and the sample size grows large). It resembles the wild parametric bootstrap technique for heteroskedastic errors, and we apply it to homoskedastic errors as well.

Instead of simulating the asymptotic distributions directly and computing critical values we rather simulate p-values following Hansen (1996) by applying the p-value transformation. This works as follows:[5]:

1. draw η as described above
2. set $\hat{v}^* = \hat{v} \odot \eta$
3. set $\widehat{\mathcal{GP}}_{\text{mat},1}(\gamma) = T^{-1/2} X_1^{\gamma\top} \hat{v}^*$
4. set $g(\gamma)$ as described in Theorem 1, respectively Theorem 2. $g(\gamma)$ denotes the (simulated) asymptotic point-wise test statistic under consideration
5. set $G = \sup_{\gamma \in \Gamma} g(\gamma)$
6. repeat steps 1–5, say J times and obtain a sample $(G_1, ..., G_J)$ of observations from the simulated asymptotic distribution of the test statistic.

[5] In contrast to the suggestion of an *i.i.d.* bootstrap in Hansen (1996), we do a wild bootstrap, because we find this translates in better empirical sizes.

7. set the simulated p-value $\hat{p}$ as

$$\hat{p} = \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}_{\{G_j > \hat{G}\}} \tag{6.2}$$

where $\hat{G}$ denotes the corresponding sample test statistic

In order to obtain empirical sizes $\hat{\alpha}$ for a nominal significance level α we repeat the above procedure under $\mathbb{H}_0$, say MC, times and set

$$\hat{\alpha} = \frac{1}{MC} \sum_{k=1}^{MC} \mathbb{1}_{\{\hat{p}_k \leq \alpha\}}. \tag{6.3}$$

The empirical power is obtained analogously with the DGP under $\mathbb{H}_A$.

Data Generating Process The DGP used in our simulations is defined as follows:

$$\begin{aligned} y_t &= z_t \theta_z^0 + \theta_{x_1}^0 + \epsilon_t \\ &= w_t^\top \theta^0 + \epsilon_t \qquad (6.4) \\ z_t &= (\Pi_{1,1}^0 + \Pi_{1,2}^0 x_t) \mathbb{1}_{\{q_t \leq \rho^0\}} + (\Pi_{2,1}^0 + \Pi_{2,2}^0 x_t) \mathbb{1}_{\{q_t > \rho^0\}} + u_t \qquad (6.5) \end{aligned}$$

where $x_t \overset{iid}{\sim} \mathcal{N}(1,1)$ and $q_t = x_t + 1$. We set the parameters as follows:

- $\theta_z^0 = \theta_{x_1}^0 = 1$.
- $\Pi_1^0 = (\Pi_{1,1}^0, \Pi_{1,2}^0)^\top = (1, 1)^\top$.
- $\Pi_2^0 = (\Pi_{2,1}^0, \Pi_{2,2}^0)^\top = (1, b)^\top$, where we allow $b \in \{0.5, 1, 1.5, 2, 2.5\}$. Note that $b = 1$ corresponds to the LRF case.
- $\rho^0 = 1.75$.

Lastly, we set $\epsilon_t = \nu_t \cdot x_t / \sqrt{2}$ with

$$\begin{pmatrix} \nu_t \\ u_t \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right). \tag{6.6}$$

This specification induces conditional heteroskedasticity. Moreover, we divide in the construction of ϵ_t by $\sqrt{2}$ in order to have an *unconditional* variance of 1 for ϵ_t. We do this to be able to ompare to the case of conditional homoskedasticity in which we simply set $\epsilon_t = \nu_t$.[6]
Last, due to extensive computational burden, we carry out 500 simulations to obtain single p-values and 1000 Monte Carlo repetitions (i.e. we simulate 1000 p-values).

[6]Results without the division with $\sqrt{2}$ are similar and available upon request.

6.2. Size

This section presents empirical sizes obtained for our tests as well as for the GMM sup Wald test in Caner and Hansen (2004) under the DGP and p-value simulations described above. Table 9.1 displays the results.

In case of conditional heteroskedasticity, the results show that in small samples, the 2SLS sup LR test is correctly sized, and clearly superior to both the 2SLS sup Wald and GMM sup Wald test. That the 2SLS sup LR test has better size properties than the GMM sup tests should not come as a surprise, because the 2SLS estimator use more information about the reduced form compared to the GMM sup Wald test. In particular, it uses TRF and LRF to compute the first-stage, while GMM ignores this information altogether. However, the result that the sup LR test also outperforms the 2SLS sup Wald test is somewhat surprising.

In case of conditional homoskedastic errors the results are similar. Thus, in finite samples, the estimates are less accurate, resulting in poorer size properties. Again, the 2SLS sup LR seems to be correctly sized, while the 2SLS sup Wald is oversized. It seems that robustifying against heteroskedasticity backfires both in the case of heteroskedasticity and homoskedasticity, resulting in small sample size distortions for the Wald test. However, notice that the Wald size distortions are smaller in the homoskedastic case. In both cases, these distortions go away as the sample size grows large ($T = 1000$).

6.3. Power

In this section, we present the empirical (size corrected) power of the three tests. We slightly alter the DGP in (6.4) while leaving everything else equal. In particular we set

$$y_t = w_t^\top \theta_1^0 \mathbb{1}_{\{q_t \leq \gamma^0\}} + w_t^\top \theta_2^0 \mathbb{1}_{\{q_t > \gamma^0\}} + \epsilon_t \tag{6.7}$$

with $\theta_1^0 = (1,\, 1)^\top$ as before and $\theta_2^0 = (a, c)^\top$ with $a \in \{1, 2\}$ and $c \in \{1.25, 1.5, 2\}$. This allows us to investigate how the power varies with the threshold size, measured by a and c. We set $\gamma^0 = 2.25$ and we only consider conditionally homoskedastic errors with LRF and two cases of a TRF for brevity.[7]

In order to parsimoniously display our results 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. Those sizes 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. On the y-axis we plot the size adjusted power which is calculated using the empirical critical values.

6.3.1. Linear Reduced Form

In this subsection, we consider the LRF case. Figure 9.1 displays size corrected power curves for the case of no threshold effect in the intercept of the structural form, i.e. for

[7] Simulations in progress for other cases.

$a = 1$, and Figure 9.2 for a threshold effect in the intercept of the structural form, i.e. for $a = 2$.

As expected, all three tests have larger power as the sample size increases and as the threshold size increases. The first effect can, e.g. be seen in the upper left subplot of Figure 9.1. The second effect is evident in the lower right suplot, as all power curves jump close to one as the size becomes strictly positive.

So, as expected, the power increases with the threshold size. Furthermore, a threshold effect in the intercept (Figure 9.2), which can be viewed as a larger threshold, leads to an increase in size-adjusted power of all three tests. The GMM Wald test seems to have larger size-corrected power but this effect vanishes as (i) the sample size increases or (ii) as the threshold effect in the slope parameter increases.

6.3.2. Threshold Reduced Form

In this subsection, we consider the TRF case. Figures 9.3 and 9.4 display size corrected power curves for the case of no threshold effect in the intercept of the structural form, i.e. for $a = 1$, and a small, respectively big threshold effect in the reduced form, i.e. for $b = 1.5$, respectively $b = 2.5$. Moreover, Figures 9.5 and 9.6 display power curves for the case of an additional threshold effect in the structural form intercept given both cases of a TRF.

Comparing Figures 9.3 and 9.4, respectively Figures 9.5 and 9.6 suggest that the magnitude of the reduced form threshold effect does not affect the power properties of the considered tests since power patterns between Figures 9.3 and 9.4, respectively Figures 9.5 and 9.6 are quite similar. Moreover, comparing Figures 9.3 and 9.4 with Figures 9.5 and 9.6 it seems that the size of the structural form threshold effect has a much bigger influence on the power of the tests. This is also corroborated if we compare the TRF results to the LRF results. Overall, all three tests seem to have reasonably large size-corrected power if the threshold is moderate to large.

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 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 two potential interest rate regimes:

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

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.[8] Thus, we specify the reduced form (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[9], the federal funds rate from the Fed St. Louis[10] and the government spending forecasts from the Philadelphia Fed.[11] The data includes the current ZLB regime, as can be seen from the federal funds rate plot in Figure 9.7.

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 for our analysis: the full-sample 1969Q2-2014Q4, and the sub-samples 1969Q2–1984Q4, respectively 1985Q1–2014Q4. Since low interest rates are mostly near the end of our sample, but we wish to allow for a low interest rate regime, 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 reduced form 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 LR and Wald threshold tests proposed in this paper, as well as the GMM Wald test in Caner and Hansen (2004).

Tables 9.2-9.4 present results for all the three samples considered. Regardless of the cut-off, or whether we use the full-sample or the post 1985 sample, we find that the reduced form has a threshold at ρ^0 below 7.

For the structural form in (7.1), estimated on the whole sample, we find weak evidence for a threshold effect: none of the threshold tests reject at the 1% level. At the 5% level,

[8] See Ramey (2011) for more discussion on instrument validity of SPF_t, and a description of how the forecast errors were quantified.

[9] Accessed February 2015.

[10] Accessed February 2015.

[11] Accessed February 2015.

all tests reject, but the 2SLS LR test, which we found in simulations to be closest to its nominal size, is close to the 5% level, whether we use a 15% cut-off or 5% cut-off in calculating our tests. Furthermore, the threshold estimator reflects a very high interest rate regime, sensitive to the cut-off choice, 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.[12]

Table 9.3 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.[13] 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, Sample 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 and explicitly account for a possible threshold effect in the reduced form. We derive the asymptotic distributions of our tests, which are non-pivotal but can be computed by methods similar to the wild bootstrap. Our simulation study shows that our sup LR test behaves well in small samples, and its size and power compare favorably to an existing GMM based sup Wald test. We find that the sup LR is correctly sized for small samples, compared to the Wald tests which are both oversized. In terms of power, all tests have comparable and large power when the threshold size is moderate (about half of the variance of the error). Moreover, the power properties of the proposed tests are in line with the findings of Hall et al. (2012) for break-point models.

We apply our method to assess whether the US government spending multiplier is larger in regimes with nominal interest rates that are low or near the zero lower bound.

[12] The Volcker period results in Table 9.4 are presented for completeness, but the sample size is small, and care should be used in interpreting those results.

[13] Note that the 2SLS standard errors we obtain, based on asymptotic results, are larger than the GMM standard errors (formulae for the asymptotic 2SLS standard errors for a TRF are available upon request). Preliminary simulation results (also available upon request) indicate that this might be due to poor asymptotic approximations to the 2SLS standard errors in finite samples and not the classical result that GMM is more efficient. As already mentioned earlier, Antoine and Boldea (2014) find that 2SLS can be more efficient than GMM in threshold models.

All the threshold tests we employ suggest that the US government spending multiplier for output growth did not change near the zero lower bound or any other interest rate regime.

9. Tables and Figures

Table 9.1: Empirical Sizes on nominal 5% significance level

Conditional Heteroskedasticity

	b=0.05			b=1			b=1.5		
T	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$	$LR^{2SLS}_{T,LRF}(\gamma)$	$W^{2SLS}_{T,LRF}(\gamma)$	$W^{GMM}_{T,LRF}(\gamma)$	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$
100	0.033	0.245	0.110	0.046	0.269	0.159	0.029	0.244	0.188
250	0.051	0.128	0.080	0.055	0.129	0.104	0.044	0.122	0.127
500	0.051	0.086	0.077	0.065	0.103	0.092	0.053	0.096	0.090
1000	0.060	0.073	0.070	0.061	0.068	0.080	0.057	0.070	0.080

	b=2			b=2.5		
T	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$
100	0.039	0.264	0.203	0.044	0.294	0.218
250	0.053	0.145	0.130	0.056	0.153	0.135
500	0.059	0.106	0.091	0.060	0.114	0.096
1000	0.060	0.083	0.078	0.060	0.085	0.075

Conditional Homoskedasticty

	b=0.05			b=1			b=1.5		
T	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$	$LR^{2SLS}_{T,LRF}(\gamma)$	$W^{2SLS}_{T,LRF}(\gamma)$	$W^{GMM}_{T,LRF}(\gamma)$	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$
100	0.023	0.173	0.092	0.044	0.186	0.146	0.017	0.149	0.167
250	0.028	0.092	0.094	0.041	0.104	0.114	0.023	0.069	0.125
500	0.028	0.067	0.070	0.053	0.086	0.079	0.037	0.062	0.079
1000	0.042	0.064	0.062	0.049	0.069	0.064	0.034	0.051	0.060

	b=2			b=2.5		
T	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$	$LR^{2SLS}_{T,TRF}(\gamma)$	$W^{2SLS}_{T,TRF}(\gamma)$	$W^{GMM}_{T,TRF}(\gamma)$
100	0.026	0.170	0.168	0.032	0.190	0.182
250	0.039	0.092	0.126	0.050	0.105	0.126
500	0.052	0.086	0.079	0.053	0.082	0.078
1000	0.045	0.061	0.064	0.048	0.067	0.063

[1] Note that we used a wild bootstrap for the homoskedastic case. We also experimented with an *i.i.d.* bootstrap, as suggested in Hansen (1996), yielding heavily undersized test statistics.

Table 9.2: Estimation Results Full Sample

	Sample 1969Q2–2014Q4					
	Cut-Off 15% $\gamma_{\min} = 1.95$ and $\gamma_{\max} = 9.46$			Cut-Off 5% $\gamma_{\min} = 0.15$ and $\gamma_{\max} = 13.58$		
	RF Results					
p-Wald		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					
	LR^{2SLS}	W^{2SLS}	W^{GMM}	LR^{2SLS}	W^{2SLS}	W^{GMM}
p-value	0.0430	0.0220	0.0540	0.0480	0.0240	0.0180
$\hat{\beta}_1$	0.1715	0.1715	0.0853	0.0878	0.0878	0.0904
	(0.12460)	(0.12460)	(0.00600)	(0.11510)	(0.11510)	(0.0061)
$\hat{\alpha}_1$	0.0072	0.0072	0.0065	0.0072	0.0072	0.0061
	(0.00100)	(0.00100)	(0.00005)	(0.00100)	(0.00100)	(0.00005)
$\hat{\beta}_2$	-0.2049	-0.2049	–	0.5434	0.5434	0.5172
	(0.27170)	(0.27170)	–	(0.56290)	(0.56290)	(0.0273)
$\hat{\alpha}_2$	0.0044	0.0044	–	-0.0065	-0.0065	-0.0053
	(0.00310)	(0.00310)	–	(0.00670)	(0.00670)	(0.0003)
$\hat{\gamma}$	8.8000	8.8000	–	10.9500	10.9500	10.9500
95%-CI for $\hat{\gamma}$	$[6.8400; 9.3500]$		–	$[8.8000 : 12.6900]$		
No. of obs.						
total	183	183	183	183	183	183
$r_{t-1} \leq \hat{\gamma}$	148	148	148	169	169	169
$r_{t-1} > \hat{\gamma}$	35	35	35	14	14	14

[1] Standard errors in parentheses.
[2] $\gamma_{\min}$ and $\gamma_{\max}$ are the min. and max. r_{t-1} over which the threshold models are estimated.

Table 9.3: Estimation Results Subsample 1985Q1–2014Q4

	Sample 1985Q1–2014Q4					
	Cut-Off 15% $\gamma_{\min} = 1.02$ and $\gamma_{\max} = 7.74$			Cut-Off 5% $\gamma_{\min} = 0.12$ and $\gamma_{\max} = 8.48$		
	RF Results					
p-Wald		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					
	LR^{2SLS}	W^{2SLS}	W^{GMM}	LR^{2SLS}	W^{2SLS}	W^{GMM}
p-value	0.4290	0.6010	0.7230	0.4830	0.8320	0.7180
$\hat{\beta}_1$	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] Standard errors in parentheses.
[2] $\gamma_{\min}$ and $\gamma_{\max}$ are the min. and max. r_{t-1} over which the threshold models are estimated.

Table 9.4: Estimation Results Subsample 1969Q2–1984Q4

	Sample 1969Q2–1984Q4					
	Cut-Off 15% $\gamma_{\min} = 4.87$ and $\gamma_{\max} = 12.69$			Cut-Off 5% $\gamma_{\min} = 4.30$ and $\gamma_{\max} = 15.85$		
	RF Results					
p-Wald		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					
	LR^{2SLS}	W^{2SLS}	W^{GMM}	LR^{2SLS}	W^{2SLS}	W^{GMM}
p-value	0.0040	0.0000	0.0000	0.0020	0.0000	0.0000
$\hat{\beta}_1$	0.3850	0.3850	0.3874	0.3850	0.3850	0.3874
	(0.18730)	(0.18730)	(0.02330)	(0.18730)	(0.18730)	(0.02330)
$\hat{\alpha}_1$	0.0109	0.0109	0.0110	0.0109	0.0109	0.0110
	(0.00160)	(0.00160)	(0.00020)	(0.00160)	(0.00160)	(0.00020)
$\hat{\beta}_2$	-0.3049	-0.3049	-0.3061	-0.3049	-0.3049	-0.3061
	(0.26810)	(0.26810)	(0.03430)	(0.26810)	(0.26810)	(0.03430)
$\hat{\alpha}_2$	0.0040	0.0040	0.0041	0.0040	0.0040	0.0041
	(0.00270)	(0.00270)	(0.00040)	(0.00270)	(0.00270)	(0.00040)
$\hat{\gamma}$	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] Standard errors in parentheses.
[2] $\gamma_{\min}$ and $\gamma_{\max}$ are the min. and max. r_{t-1} over which the threshold models are estimated.

Figure 9.1: Size Adjusted Power Curves – No Threshold Effect in SF Intercept and LRF

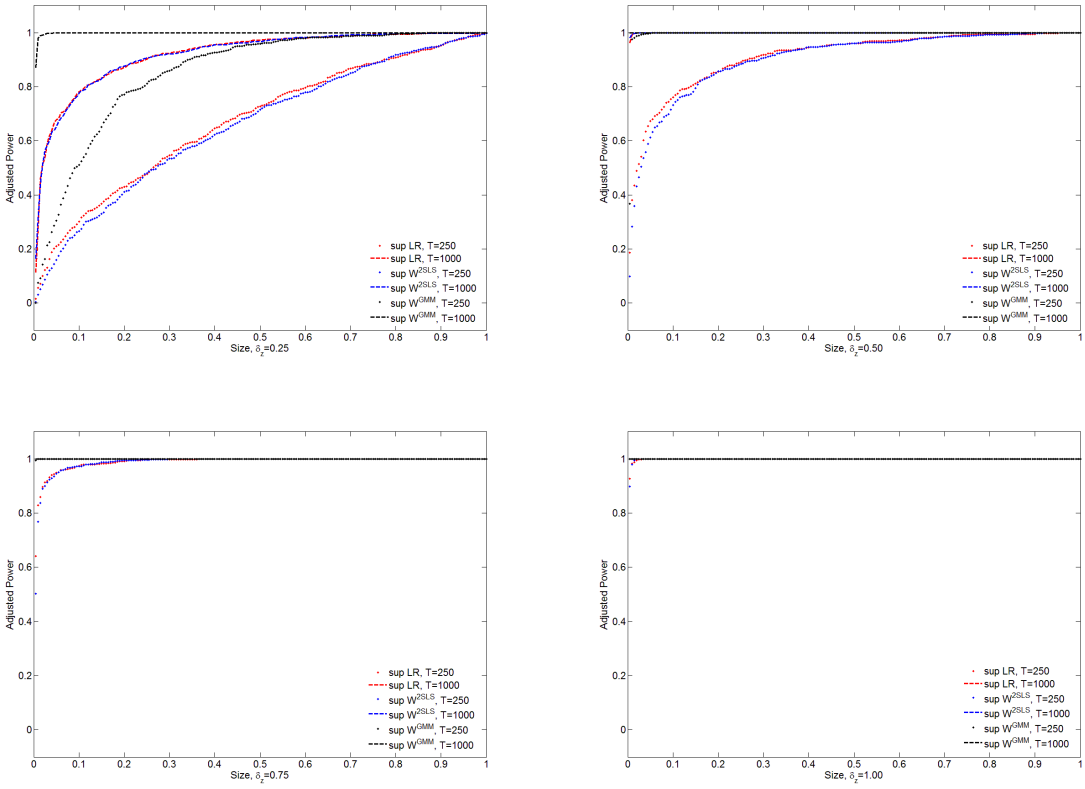


Figure 9.2: Size Adjusted Power Curves – Threshold Effect in SF Intercept and LRF

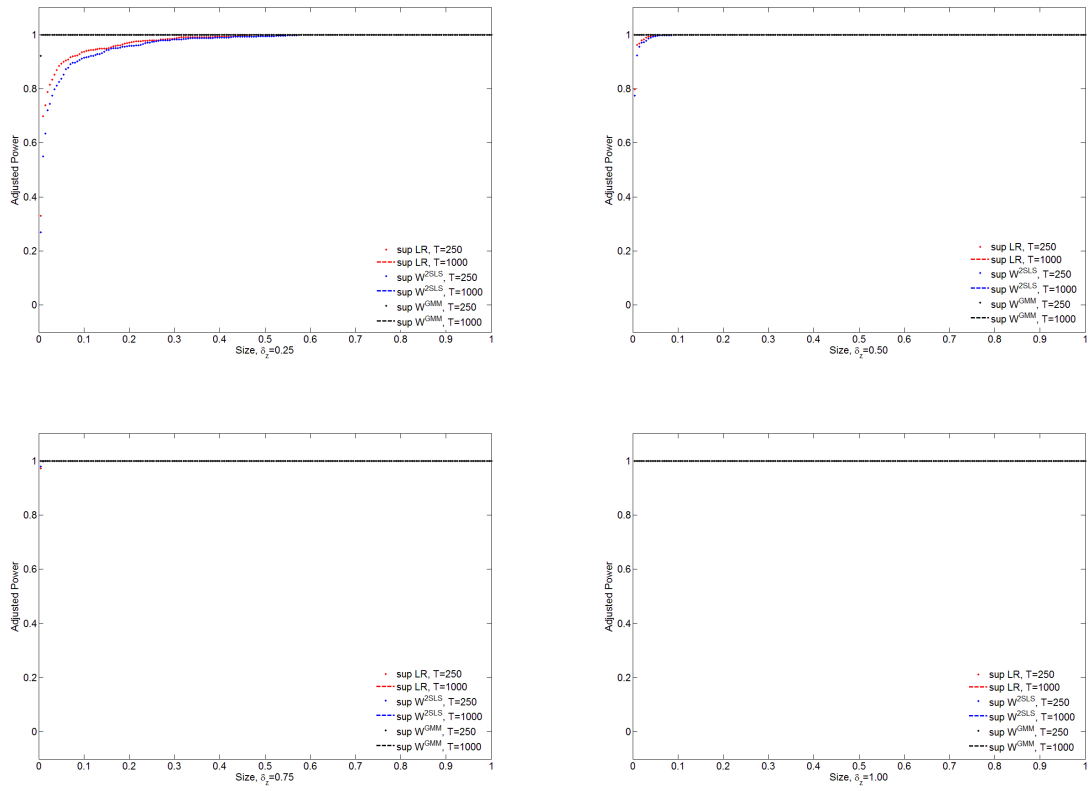


Figure 9.3: Size Adjusted Power Curves – No Threshold Effect in SF Intercept and TRF with Small Threshold Effect

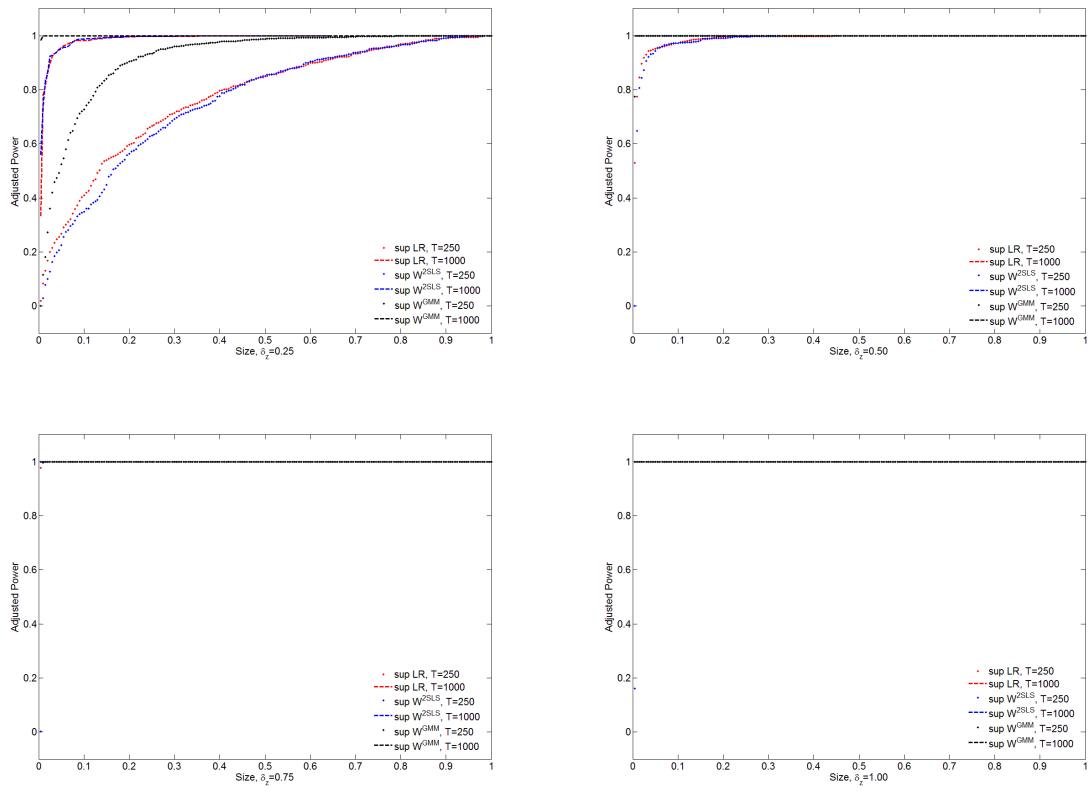


Figure 9.4: Size Adjusted Power Curves – No Threshold Effect in SF Intercept and TRF with Big Threshold Effect

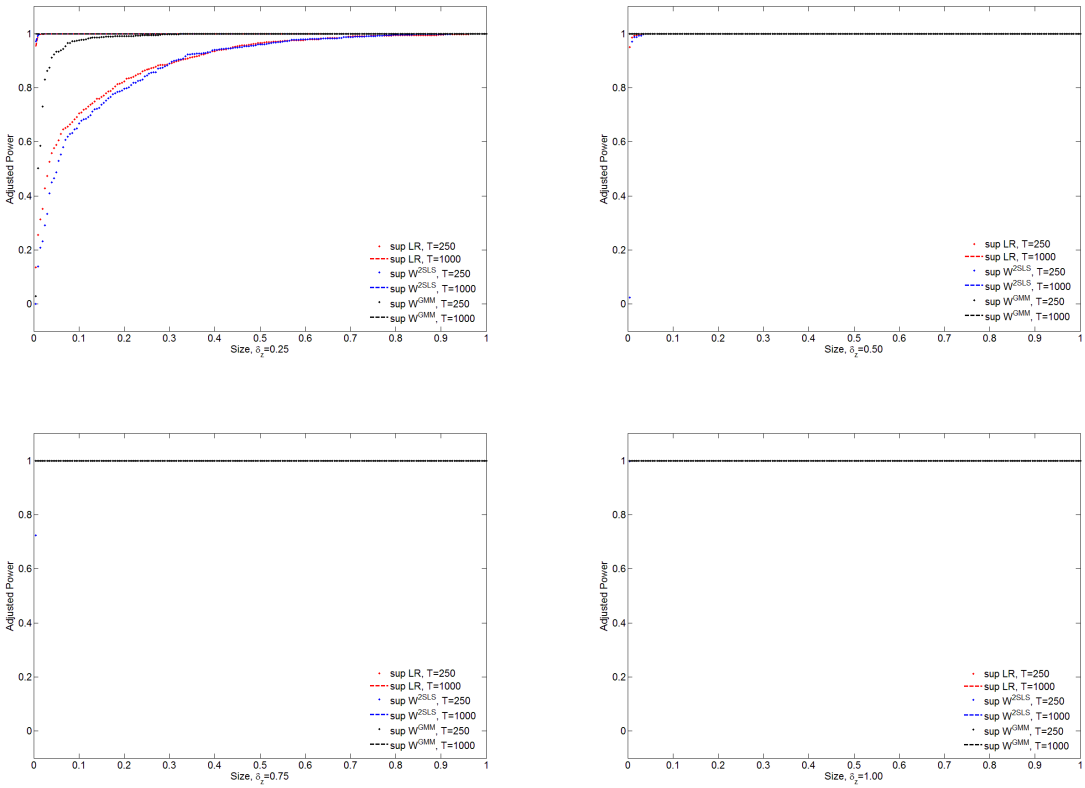


Figure 9.5: Size Adjusted Power Curves – Threshold Effect in SF Intercept and TRF with Small Threshold Effect

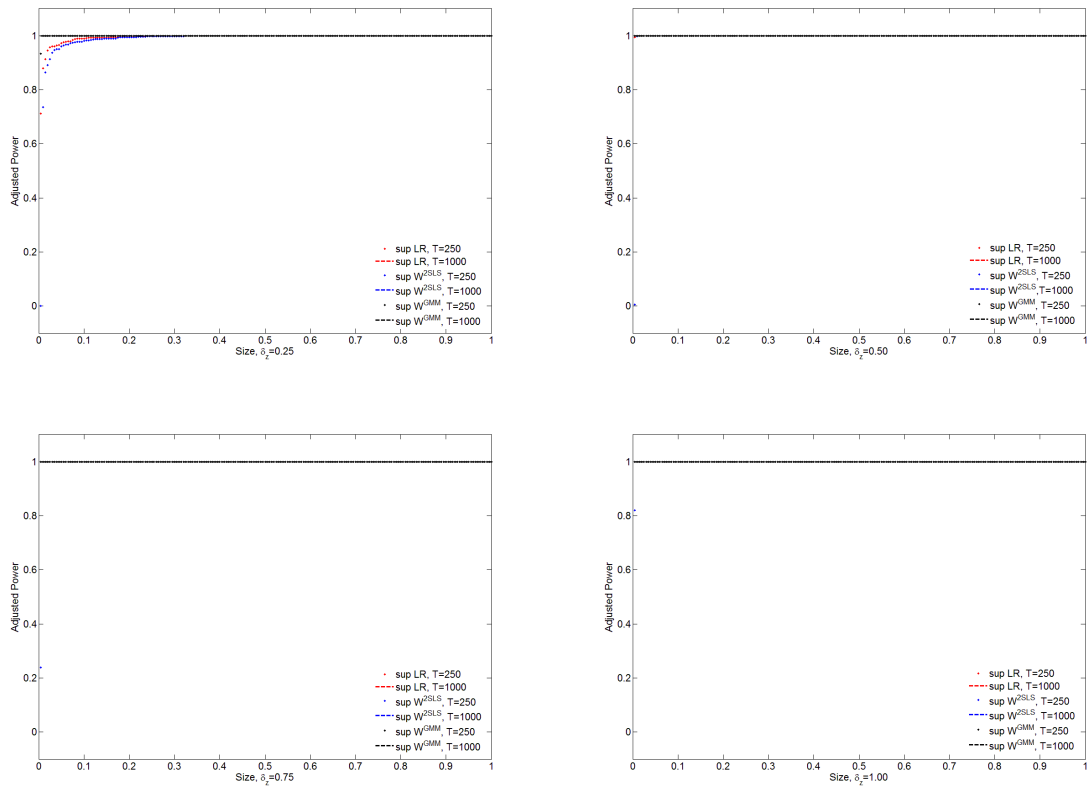


Figure 9.6: Size Adjusted Power Curves – Threshold Effect in SF Intercept and TRF with Big Threshold Effect

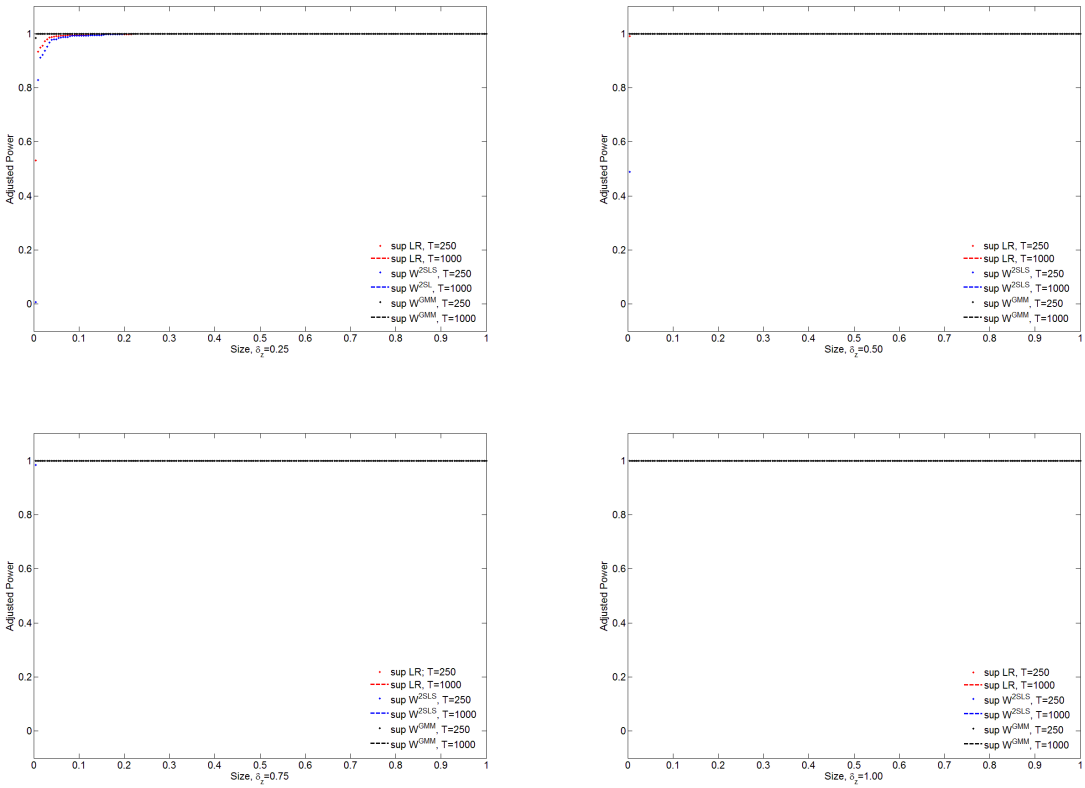
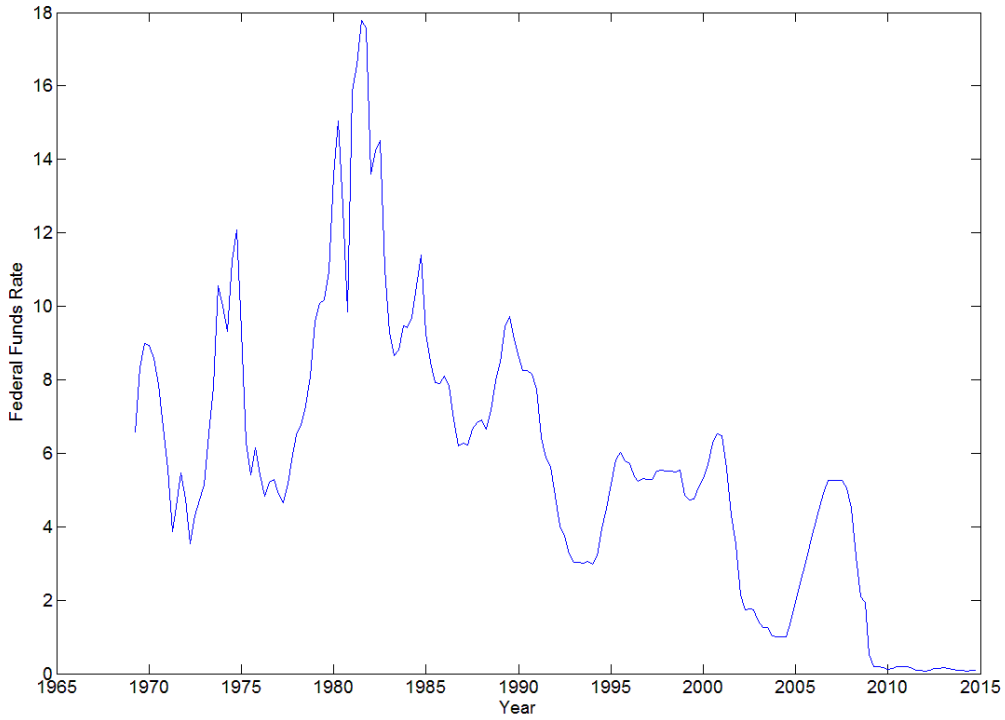


Figure 9.7: Federal Funds Rate – 1969Q2–2014Q4



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 (2014): "Efficient Inference with Time-Varying Identification Strength," Working Paper.

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 (2014): "Bootstrap-based Tests for Multiple Structural Changes in Linear Models with Endogenous Regressors," 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 theGovernment Buys More Output?" *Brookings Papers on Economic Activity*, 2(2009), 183–231.

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.

Kailath, T. (1980): *Linear Systems*, NJ: Prentice Hall.

Kourtellos, A., T. Stengos, and C. Tan (2013): "Structural Threshold Regression," Working Paper.

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., 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 (2013): "Identification Using Stability Restrictions," Working Paper.

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.

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. (2013): "Threshold Regression with Endogeneity," Working Paper,.

Mathematical Appendix

Notation. In what follows we use the symbol K to denote a strictly positive constant. Note that K does not need to be the same from line to line.

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 $m \times n$-matrix X we denote by $\|X\|_F = \sqrt{\text{tr}(X^\top X)}$ the Frobenius matrix-norm. 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.
Moreove, we define $\tilde{\epsilon} = \epsilon + (Z - \hat{Z})\theta_z^0$ and $s = \epsilon + u\theta_z^0$. Note that those quantities can also be partitioned as $\tilde{\epsilon}_1^\gamma = \epsilon_1^\gamma + (Z - \hat{Z})_1^\gamma \theta_z^0$, for example.

A. 2SLS Results involving a Linear Reduced Form

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

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

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

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

Proof of Lemma 1. Let X be a $T \times q$-matrix and v be 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)

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

and therefore

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

Next, by Hansen (1996, Theorem 1) it holds that every process $\mathcal{GP}_1^i(\gamma)$ is a zero-mean Gaussian Process with covariance function

$$\mathcal{C}_{\mathcal{GP}}^i(\gamma_1, \gamma_2) = \mathbb{E}[x_t x_t^\top v_{i,t}^2 \mathbb{1}_{\{q_t \leq (\gamma_1 \wedge \gamma_2)\}}]. \tag{A.3}$$

Similarly, it holds that

$$\mathcal{C}_{\mathcal{GP}}^{i,j}(\gamma_1, \gamma_2) = \mathbb{E}[\mathcal{GP}_1^i(\gamma_1)\mathcal{GP}_1^{j\top}(\gamma_2)] = \mathbb{E}[x_t x_t^\top v_{i,t} v_{j,t} \mathbb{1}_{\{q_t \leq (\gamma_1 \wedge \gamma_2)\}}]. \tag{A.4}$$

Thus, combinig (A.3) and (A.4) directly yields

$$\mathcal{C}_{\mathcal{GP}}(\gamma_1,\gamma_2) = \mathbb{E}[\mathcal{GP}_1(\gamma_1)\mathcal{GP}_1^\top(\gamma_2)] = \mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le (\gamma_1\wedge\gamma_2)\}}]. \tag{A.5}$$

Finally, results (A.2) and (A.5) directly prove the Lemma. □

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

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

$$\hat{Z} = X\hat{\Pi} \tag{A.6}$$

and it holds that

$$T^{1/2}(\hat{\Pi} - \Pi^0) = \left(T^{-1}X^\top X\right)^{-1}\left(T^{-1/2}X^\top u\right) \tag{A.7}$$

by standard OLS-derivations. By Hansen (1996, Theorem 1) it holds uniformly in γ that

$$T^{-1}X_1^{\gamma\top}X_1^\gamma \xrightarrow{a.s.} M_1(\gamma), \text{ and } T^{-1}X^\top X \xrightarrow{a.s.} M. \tag{A.8}$$

This implies that $T^{-1}X^\top X = \mathcal{O}_p(1)$ and thus, 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 thus, $\hat{\Pi} - \Pi^0 = o_p(1)$. So, uniformly in γ,

$$T^{-1}\hat{Z}_1^{\gamma\top}\hat{Z}_1^\gamma = \hat{\Pi}^\top\left(T^{-1}X_1^{\gamma\top}X_1^\gamma\right)\hat{\Pi} \xrightarrow{p} \Pi^{0\top}M_1(\gamma)\Pi^0. \tag{A.9}$$

Last, it holds that

$$\hat{W}_1^\gamma = \begin{bmatrix}\hat{Z}_1^\gamma & X_{1,1}^\gamma\end{bmatrix} = \begin{bmatrix}X_1^\gamma\hat{\Pi} & X_{1,1}^\gamma\end{bmatrix} = X_1^\gamma\begin{bmatrix}\hat{\Pi} & S\end{bmatrix} = X_1^\gamma\hat{A}^\top. \tag{A.10}$$

Therefore, by (A.9) and (A.10) and uniformly in γ,

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma &= \hat{A}\left(T^{-1}X_1^{\gamma\top}X_1^\gamma\right)\hat{A}^\top \\ &\xrightarrow{p} A^0M_1(\gamma)A^{0\top} \equiv C_1(\gamma), \end{aligned} \tag{A.11}$$

proving the claim.
Claim (*ii*)**:** By (A.6) it follows that

$$T^{-1/2}\hat{Z}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma = \hat{\Pi}^\top\left(\underbrace{T^{-1/2}X_1^{\gamma\top}(\epsilon_1^\gamma + u_1^\gamma\theta_z^0)}_{=a} - \underbrace{T^{-1/2}X_1^{\gamma\top}X_1^\gamma(\hat{\Pi} - \Pi^0)\theta_z^0}_{=b}\right). \tag{A.12}$$

Next, we show the limiting behavior of the term denoted by a and afterwards the limiting behavior of b.
Rewriting term a directly yields

$$T^{-1/2}X_1^{\gamma\top}(\epsilon_1^\gamma + u_1^\gamma\theta_z^0) = T^{-1/2}[X_1^{\gamma\top}\epsilon_1^\gamma, X_1^{\gamma\top}u_1^\gamma]\tilde{\theta}_z^0 \tag{A.13}$$

and thus, by Lemma 1, uniformly in γ:

$$T^{-1/2}[X_1^{\gamma\top}\epsilon_1^\gamma, X_1^{\gamma\top}u_1^\gamma]\tilde{\theta}_z^0 \Longrightarrow \mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0. \tag{A.14}$$

By (A.7) term b in (A.12) satisfies

$$T^{-1/2}X_1^{\gamma\top}X_1^\gamma(\hat{\Pi}-\Pi^0)\theta_z^0 = \left(T^{-1}X_1^{\gamma\top}X_1^\gamma\right)\left(T^{-1}X^\top X\right)^{-1}\left(T^{-1/2}X^\top u\theta_z^0\right). \tag{A.15}$$

Next, note that

$$\begin{aligned} T^{-1/2}X^\top u\theta_z^0 &= T^{-1/2}X^\top\epsilon\cdot 0 + T^{-1/2}X^\top u\theta_z^0 \\ &= T^{-1/2}[X^\top\epsilon, X^\top u]\check{\theta}_z^0 \end{aligned} \tag{A.16}$$

So, by (A.8) and (A.14)–(A.16), uniformly in γ,

$$T^{-1/2}X_1^{\gamma\top}X_1^\gamma(\hat{\Pi}-\Pi^0)\theta_z^0 \Longrightarrow M_1(\gamma)M^{-1}\mathcal{GP}_{\text{mat},1}\check{\theta}_z^0. \tag{A.17}$$

Last, because $\hat{\Pi}^\top(a-b) = \Pi^{0\top}(a-b) + o_p(1)$, (A.14) and (A.17) together with (A.12) yield uniformly in γ

$$T^{-1/2}\hat{Z}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma \Longrightarrow \Pi^{0\top}\left(\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma)M^{-1}\mathcal{GP}_{\text{mat},1}\check{\theta}_z^0\right) \equiv \Pi^{0\top}\mathcal{B}_1(\gamma). \tag{A.18}$$

Last, because $\hat{W}_1^\gamma = \begin{bmatrix}\hat{Z}_1^\gamma & X_{1,1}^\gamma\end{bmatrix} = X_1^\gamma\hat{A}^\top$ (see (A.10)) it immediately follows with (A.18) that, uniformly in γ,

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma \Longrightarrow A^0\mathcal{B}_1(\gamma), \tag{A.19}$$

proving claim (ii). □

Proof of Theorem 1. In order to show the statement of Theorem 1, we first prove it for the sup Wald test and afterwards for the sup LR test.

sup Wald Test: The Wald-type test statistic is given by

$$W_T^{2\text{SLS}}(\gamma) = T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma)^\top[T\operatorname{Var}(\hat{\theta}_1^\gamma) + T\operatorname{Var}(\hat{\theta}_2^\gamma)]^{-1}T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma)$$

This proof is done in two parts: In part (i) we prove that $T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma) \Longrightarrow C_1^{-1}(\gamma)A^0\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)A^0\mathcal{B}_2(\gamma) \equiv \mathcal{E}(\gamma)$ uniformly in γ and in part (ii) that $T\operatorname{Var}(\hat{\theta}_1^\gamma) + T\operatorname{Var}(\hat{\theta}_2^\gamma) \xrightarrow{p} C_1^{-1}(\gamma)H_1(\gamma)C_1^{-1}(\gamma) + C_2^{-1}(\gamma)H_2(\gamma)C_2^{-1}(\gamma)$ uniformly in γ. Parts (i) and (ii) together with the continuous mapping theorem and weak convergence (uniformly in γ) then immediately yield the claim.

Part (i)**.** For $i = 1, 2$, we have by construction

$$\begin{aligned} \hat{\theta}_i^\gamma &= \left(\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}\left(\hat{W}_i^{\gamma\top}Y\right) \\ &= \left(\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}\left(\hat{W}_i^{\gamma\top}(W\theta^0+\epsilon)\right) \\ &= \left(\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}\left(\hat{W}_i^{\gamma\top}(W\theta^0+\epsilon+\hat{Z}\theta_z^0-\hat{Z}\theta_z^0)\right) \\ &= \left(\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}\left(\hat{W}_i^{\gamma\top}(\hat{W}\theta^0+\epsilon+Z\theta_z^0-\hat{Z}\theta_z^0)\right) \\ &= \theta^0 + \left(\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}\left(\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^\gamma\right). \end{aligned} \tag{A.20}$$

Therefore,

$$T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma) = \left(T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma\right)^{-1}\left(T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma\right) - \left(T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^\gamma\right)^{-1}\left(T^{-1/2}\hat{W}_2^{\gamma\top}\tilde{\epsilon}_2^\gamma\right). \tag{A.21}$$

By Lemma 2 it holds, uniformly in γ, that

$$T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma \xrightarrow{p} C_i(\gamma) \text{ and } T^{-1/2}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^\gamma \Longrightarrow A^0\mathcal{B}_i(\gamma). \tag{A.22}$$

Thus, uniformly in γ,

$$\begin{aligned} T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma) &= \left(T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma\right)^{-1}\left(T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma\right) \\ &\quad - \left(T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^\gamma\right)^{-1}\left(T^{-1/2}\hat{W}_2^{\gamma\top}\tilde{\epsilon}_2^\gamma\right) \\ &\Longrightarrow C_1^{-1}(\gamma)A^0\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)A^0\mathcal{B}_2(\gamma) \equiv \mathcal{E}(\gamma). \end{aligned} \tag{A.23}$$

Part (ii)**.** Note that

$$T\,\mathrm{Var}(\hat{\theta}_i^\gamma) = \left(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)}\hat{w}_t\hat{w}_t^\top\hat{e}_t^2\right)\left(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1}.$$

By Lemma 2 and the continuous mapping theorem it holds uniformly in γ that

$$\left(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma\right)^{-1} \xrightarrow{p} C_i^{-1}(\gamma). \tag{A.24}$$

So, we are left to show that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_1(\gamma)}\hat{w}_t\hat{w}_t^\top\hat{e}_t^2 &\xrightarrow{p} A^0\mathbb{E}[x_tx_t^\top\{\epsilon_t + u_t^\top\theta_z^0\}^2\mathbb{1}_{\{q_t\le\gamma\}}]A^{0\top} \\ T^{-1}\sum_{\mathcal{T}_2(\gamma)}\hat{w}_t\hat{w}_t^\top\hat{e}_t^2 &\xrightarrow{p} A^0\mathbb{E}[x_tx_t^\top\{\epsilon_t + u_t^\top\theta_z^0\}^2\mathbb{1}_{\{q_t>\gamma\}}]A^{0\top} \end{aligned} \tag{A.25}$$

.

To do so, we will show that

$$T^{-1}\sum_{\mathcal{T}_i(\gamma)}\hat{w}_t\hat{w}_t^\top\hat{e}_t^2 = A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)}x_tx_t^\top\{\epsilon_t + u_t^\top\theta_z^0\}^2\right)A^{0\top} + o_p(1). \tag{A.26}$$

Based on the statement in (A.26) the claim then follows immediately.
First, note that $\hat{w}_t = \hat{A}x_t = A^0x_t + o_p(1)$ and thus that

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_i(\gamma)}\hat{w}_t\hat{w}_t^\top\hat{e}_t^2 &= \hat{A}\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)}x_tx_t^\top\hat{e}_t^2\right)\hat{A}^\top \\ &= A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)}x_tx_t^\top\hat{e}_t^2\right)A^{0\top} + o_p(1), \end{aligned} \tag{A.27}$$

since $\hat{A} = A^0 + o_p(1)$ uniformly in γ, and we expect the whole sum to be at most of order $\mathcal{O}_p(1)$.
Next, we rewrite the expression for the residuals, $\hat{e}_t$, in such a way that (A.26) follows. By construction

$$\hat{e}_t = (y_t - \hat{w}_t^\top \hat{\theta}_1^\gamma)\mathbb{1}_{\{q_t \le \gamma\}} + (y_t - \hat{w}_t^\top \hat{\theta}_2^\gamma)\mathbb{1}_{\{q_t > \gamma\}} \tag{A.28}$$

and therefore,

$$\hat{e}_t^2 = (y_t - \hat{w}_t^\top \hat{\theta}_1^\gamma)^2\mathbb{1}_{\{q_t \le \gamma\}} + (y_t - \hat{w}_t^\top \hat{\theta}_2^\gamma)^2\mathbb{1}_{\{q_t > \gamma\}} \tag{A.29}$$

because the mixed term includes $\mathbb{1}_{\{q_t \le \gamma\}} \cdot \mathbb{1}_{\{q_t > \gamma\}} = 0$.
Using (A.29), quantity (A.27) can be written as

$$\begin{aligned} T^{-1}\sum_{\mathcal{T}_i(\gamma)} \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 &= A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \hat{e}_t^2\right)A^{0\top} + o_p(1) \\ &= A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (y_t - \hat{w}_t^\top \hat{\theta}_i^\gamma)^2\right)A^{0\top} + o_p(1) \\ &\equiv A^0 B A^{0\top} + o_p(1). \end{aligned} \tag{A.30}$$

Next, under $\mathbb{H}_0$, $(y_t - \hat{w}_t^\top \hat{\theta}_i^\gamma)^2$ can be further rewritten as

$$\begin{aligned} (y_t - \hat{w}_t^\top \hat{\theta}_i^\gamma)^2 &= (w_t^\top \theta^0 + \epsilon_t - \hat{w}_t^\top \hat{\theta}_i^\gamma)^2 \\ &= (\hat{w}_t^\top(\theta^0 - \hat{\theta}_i^\gamma) + \epsilon_t + (z_t - \hat{z}_t)^\top \theta_z^0)^2 \\ &= (\hat{w}_t^\top(\theta^0 - \hat{\theta}_i^\gamma) + \tilde{\epsilon}_t)^2 \\ &= \tilde{\epsilon}_t^2 + 2\tilde{\epsilon}_t \hat{w}_t^\top(\theta^0 - \hat{\theta}_i^\gamma) + (\theta^0 - \hat{\theta}_i^\gamma)^\top \hat{w}_t\hat{w}_t^\top(\theta^0 - \hat{\theta}_i^\gamma). \end{aligned} \tag{A.31}$$

Therefore, quantity B from (A.30) reads as

$$\begin{aligned} B = {} & T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \tilde{\epsilon}_t^2 \\ & + 2T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \tilde{\epsilon}_t x_t^\top A^{0\top}(\theta^0 - \hat{\theta}_i^\gamma) \\ & + T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (\theta^0 - \hat{\theta}_i^\gamma)^\top A^0 x_t x_t^\top A^{0\top}(\theta^0 - \hat{\theta}_i^\gamma) + o_p(1) \end{aligned} \tag{A.32}$$

since $\hat{A} = A^0 + o_p(1)$ uniformly in γ, and we expect the sums to be at most of oder $\mathcal{O}_p(1)$.
Next, we show that the last two terms on the right-hand side of (A.32) converge in probability to zero, uniformly in γ:

The last term in (A.32) is bounded by[14]

$$
\begin{aligned}
&\left\| T^{-1} \sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top (\theta^0 - \hat{\theta}_i^\gamma)^\top A^0 x_t x_t^\top A^{0\top} (\theta^0 - \hat{\theta}_i^\gamma) \right\|_F \\
\leq & T^{-1} \sum_{\mathcal{T}_i(\gamma)} \| x_t x_t^\top (\theta^0 - \hat{\theta}_i^\gamma)^\top A^0 x_t x_t^\top A^{0\top} (\theta^0 - \hat{\theta}_i^\gamma) \|_F \\
= & T^{-1} \sum_{\mathcal{T}_i(\gamma)} \| x_t x_t^\top \|_F \cdot |(\theta^0 - \hat{\theta}_i^\gamma)^\top A^0 x_t|^2 \\
= & T^{-1} \sum_{\mathcal{T}_i(\gamma)} \| x_t \|_2^2 \cdot |(\theta^0 - \hat{\theta}_i^\gamma)^\top A^0 x_t|^2 \\
\leq & T^{-1} \sum_{\mathcal{T}_i(\gamma)} \| x_t \|_2^2 \cdot \| \theta^0 - \hat{\theta}_i^\gamma \|_2^2 \cdot \| A^0 x_t \|_2^2 \\
\leq & \| \theta^0 - \hat{\theta}_i^\gamma \|_2^2 \cdot \| A^0 \|_F^2 \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \| x_t \|_2^4 \right) \\
\xrightarrow{p} & 0,
\end{aligned}
$$

where convergence holds because $\|\theta^0 - \hat{\theta}_i^\gamma\|_2 = o_p(1)$ by Lemma 2, $\|A^0\|_F = \mathcal{O}(1)$ by Assumption A.1.2 and $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4 = \mathcal{O}_p(1)$, uniformly in γ. To see the last statement consider

$$
\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4 > K \right) &\leq \frac{\mathbb{E}\left[\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4 \right]}{TK} \\
&= \frac{\sum_{\mathcal{T}_i(\gamma)} \mathbb{E}\|x_t\|_2^4}{TK} \\
&\leq \frac{\sum_{\mathcal{T}_i(\gamma)} \sup_t \mathbb{E}\|x_t\|_2^4}{TK} \\
&\leq \frac{\sup_t \mathbb{E}\|x_t\|_2^4}{K}
\end{aligned}
$$

where $\sup_t \mathbb{E}\|x_t\|_2^4 \leq \sup_t \mathbb{E}\|x_t\|_2^{4b} < \infty$ for some $b > 1$ by Assumption A.1.2. Thus, the last expression can be made arbitrarily small uniformly in T by choosing K sufficiently large. Hence, $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4$ is uniformly tight, or equivalently, of order $\mathcal{O}_p(1)$.
By the same arguments, the second term on the RHS of equation (A.32) can be bounded

[14]Note that for $u, v \in \mathbb{R}^{n \times 1}$ it holds that $\|uv^\top\|_F = \|u\|_2 \cdot \|v\|_2$ because $\|uv^\top\|_F = \sqrt{\sum_i \sum_j |u_i v_j|^2} = \sqrt{\sum_i \sum_j |u_i|^2 |v_j|^2} = \sqrt{\sum_i |u_i|^2 \sum_j |v_j|^2} = \sqrt{\sum_i |u_i|^2} \sqrt{\sum_j |v_j|^2} = \|u\|_2 \cdot \|v\|_2$. Moreover, the Frobenius matrix norm is compatible with the Euclidean vector norm, i.e. for some $m \times n$-matrix A and some $n \times 1$-vector x holds that $\|Ax\|_2 \leq \|A\|_F \|x\|_2$. This can easily be shown by applying the definitions of the respective norms and the Cauchy-Schwartz inequality.

by

$$
\begin{aligned}
&\left\| T^{-1} \sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \tilde{\epsilon}_t x_t^\top A^{0\top} (\theta^0 - \hat{\theta}_i^\gamma) \right\|_F \\
\leq & \|\theta^0 - \hat{\theta}_i^\gamma\|_2 \cdot \|A^0\|_F \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\tilde{\epsilon}_t| \right) \\
\leq & \|\theta^0 - \hat{\theta}_i^\gamma\|_2 \cdot \|A^0\|_F \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| + \|\theta_z^0\|_2 T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|z_t - \hat{z}_t\|_2 \right) \\
\leq & \|\theta^0 - \hat{\theta}_i^\gamma\|_2 \cdot \|A^0\|_F \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| + \|\theta_z^0\|_2 \cdot \|\Pi^0 - \hat{\Pi}\|_F T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2 \right) \\
\xrightarrow{p} & 0
\end{aligned}
$$

where the last statement holds because $\|\theta^0 - \hat{\theta}_i^\gamma\|_2 = o_p(1)$ and $\|\Pi^0 - \hat{\Pi}\|_F = o_p(1)$ by Lemma 2, $\|A^0\|_F = O_p(1)$ by construction, $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| = \mathcal{O}_p(1)$, and $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |u_t| = \mathcal{O}_p(1)$, uniformly in γ. To see the last two statements consider:

$$
\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| > K \right) &\leq \frac{\mathbb{E}\left[\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| \right]}{TK} \leq \frac{\sup_t \mathbb{E}[\|x_t\|_2^3 |\epsilon_t|]}{K} \\
&\leq \frac{\sup_t \left(\mathbb{E}[\|x_t\|_2^6] \mathbb{E}[|\epsilon_t|^2] \right)^{1/2}}{K}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}\left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2 > K \right) &\leq \frac{\mathbb{E}\left[\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2 \right]}{TK} \leq \frac{\sup_t \mathbb{E}[\|x_t\|_2^3 \|u_t\|_2]}{K} \\
&\leq \frac{\sup_t \left(\mathbb{E}[\|x_t\|_2^6] \mathbb{E}[\|u_t\|_2^2] \right)^{1/2}}{K}
\end{aligned}
$$

where $\sup_t \mathbb{E}\|x_t\|_2^6 \leq \sup_t \mathbb{E}\|x_t\|_2^{6b} < \infty$, $\sup_t \mathbb{E}|\epsilon_t|^2 \leq \sup_t \mathbb{E}|\epsilon_t|^{4b} < \infty$ and $\sup_t \mathbb{E}\|u_t\|_2^2 \leq \sup_t \mathbb{E}\|u_t\|_2^{4b} < \infty$ by Assumption A.1.2. Thus, the last expressions can be made arbitrarily small uniformly in T by choosing K sufficiently large. Hence, $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\tilde{\epsilon}_t|$ is of order $\mathcal{O}_p(1)$.
Therefore, (A.30) simplifies to

$$
T^{-1} \sum_{\mathcal{T}_i(\gamma)} \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 = A^0 \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \tilde{\epsilon}_t^2 \right) A^{0\top} + o_p(1). \tag{A.33}
$$

Last,

$$\begin{aligned}\tilde{\epsilon}_t^2 &= \left[\epsilon_t + (z_t - \hat{z}_t)^\top \theta_z^0\right]^2 \\ &= \left[\epsilon_t + (\Pi^{0\top} x_t + u_t - \hat{\Pi}^\top x_t)^\top \theta_z^0\right]^2 \\ &= \left[\epsilon_t + u_t^\top \theta_z^0 + x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0\right]^2 \\ &= \left[s_t + x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0\right]^2 \end{aligned} \tag{A.34}$$

Therefore,

$$\begin{aligned}T^{-1}\sum_{\mathcal{T}_i(\gamma)} \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 = {} & A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top s_t^2\right)A^{0\top} \\ & + 2A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top s_t x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0\right)A^{0\top} \\ & + A^0\left(T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \theta_z^{0\top}(\Pi^0 - \hat{\Pi}) x_t x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0\right)A^{0\top} + o_p(1). \end{aligned} \tag{A.35}$$

It can now be shown that the last two expressions on the right-hand-side of (A.35) converge in probability to zero, uniformly in γ. For the thirs term we find:

$$\begin{aligned}&\left\| T^{-1}\sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top \theta_z^{0\top}(\Pi^0 - \hat{\Pi}) x_t x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0 \right\|_F \\ \leq & T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t x_t^\top \theta_z^{0\top}(\Pi^0 - \hat{\Pi}) x_t x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0\|_F \\ = & T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t x_t^\top\|_F \cdot |x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0|^2 \\ = & T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \cdot |x_t^\top (\Pi^0 - \hat{\Pi})\theta_z^0|^2 \\ \leq & T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^2 \cdot \|x_t\|_2^2 \cdot \|(\Pi^0 - \hat{\Pi})\theta_z^0\|_2^2 \\ \leq & \left(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4\right)\|\Pi^0 - \hat{\Pi}\|_F^2 \cdot \|\theta_z^0\|_2^2 \end{aligned}$$

where $\|\theta_z^0\|_2 = \mathcal{O}(1)$ by construction, $\|\Pi^0 - \hat{\Pi}\| = o_p(1)$ by Lemma 2 and, as already shown, $T^{-1}\sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^4 = \mathcal{O}_p(1)$ uniformly in γ.

Similarly, we find for the second term on the RHS of (A.35):

$$\begin{aligned}
&\left\| T^{-1} \sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top s_t x_t^\top (\Pi^0 - \hat{\Pi}) \theta_z^0 \right\|_F \\
\leq & \|\Pi^0 - \hat{\Pi}\|_F \cdot \|\theta_z^0\|_2 T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |s_t| \\
\leq & \|\Pi^0 - \hat{\Pi}\|_F \cdot \|\theta_z^0\|_2 \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| + \|\theta_z^0\|_2 T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2 \right) \\
\xrightarrow{p} & 0
\end{aligned}$$

where $\|\theta_z^0\|_2 = \mathcal{O}(1)$ by construction, $\|\Pi^0 - \hat{\Pi}\|_F = o_p(1)$ by Lemma 2 and, as already shown, $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 |\epsilon_t| = \mathcal{O}_p(1)$ and $T^{-1} \sum_{\mathcal{T}_i(\gamma)} \|x_t\|_2^3 \|u_t\|_2 = \mathcal{O}_p(1)$ uniformly in γ. Thus, (A.35) can be restated as

$$T^{-1} \sum_{\mathcal{T}_i(\gamma)} \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 = A^0 \left(T^{-1} \sum_{\mathcal{T}_i(\gamma)} x_t x_t^\top s_t^2 \right) A^{0\top} + o_p(1), \tag{A.36}$$

and therefore, uniformly in γ

$$T^{-1} \sum_{\mathcal{T}_1(\gamma)} \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 \xrightarrow{p} A^0 \mathbb{E}[x_t x_t^\top \{\epsilon_t + u_t^\top \theta_z^0\}^2 \mathbb{1}_{\{q_t \leq \gamma\}}] A^{0\top} \tag{A.37}$$

$$T^{-1} \sum_{\mathcal{T}_2(\gamma)} \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 \xrightarrow{p} A^0 \mathbb{E}[x_t x_t^\top \{\epsilon_t + u_t^\top \theta_z^0\}^2 \mathbb{1}_{\{q_t > \gamma\}}] A^{0\top}. \tag{A.38}$$

Results (A.23), (A.24), (A.37) and (A.38) then yield the claim by continuity of the involved terms and weak convergence (uniformly in γ) using the continuous mapping theorem.

sup LR Test: This proof is done in two parts: Part (i) shows that $T^{-1} SSR_1(\gamma) \xrightarrow{p} \sigma^2$ and part (ii) shows that $SSR_0 - SSR_1(\gamma) \Longrightarrow \mathcal{E}^\top(\gamma) C_2(\gamma) D_1(\gamma) \mathcal{E}(\gamma)$.

Part (i)**.** The scaled sum of squared residuals of the restricted model, $SSR_1(\gamma)$, are given by

$$\begin{aligned}
T^{-1} SSR_1(\gamma) &= T^{-1} [Y_1^\gamma - \hat{W}_1^\gamma \hat{\theta}_1^\gamma]^\top [Y_1^\gamma - \hat{W}_1^\gamma \hat{\theta}_1^\gamma] \\
&+ T^{-1} [Y_2^\gamma - \hat{W}_2^\gamma \hat{\theta}_2^\gamma]^\top [Y_2^\gamma - \hat{W}_2^\gamma \hat{\theta}_2^\gamma] \\
&= T^{-1} [\hat{W}_1^\gamma (\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_1^\gamma]^\top [\hat{W}_1^\gamma (\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_1^\gamma] \\
&+ T^{-1} [\hat{W}_2^\gamma (\theta^0 - \hat{\theta}_2^\gamma) + \tilde{\epsilon}_2^\gamma]^\top [\hat{W}_2^\gamma (\theta^0 - \hat{\theta}_2^\gamma) + \tilde{\epsilon}_2^\gamma] \\
&= T^{-1} \tilde{\epsilon}^\top \tilde{\epsilon} \\
&+ 2 (T^{-1} \tilde{\epsilon}_1^{\gamma\top} \hat{W}_1^\gamma)(\theta^0 - \hat{\theta}_1^\gamma) + (\theta^0 - \hat{\theta}_1^\gamma)^\top (T^{-1} \hat{W}_1^{\gamma\top} \hat{W}_1^\gamma)(\theta^0 - \hat{\theta}_1^\gamma) \\
&+ 2 (T^{-1} \tilde{\epsilon}_2^{\gamma\top} \hat{W}_2^\gamma)(\theta^0 - \hat{\theta}_2^\gamma) + (\theta^0 - \hat{\theta}_2^\gamma)^\top (T^{-1} \hat{W}_2^{\gamma\top} \hat{W}_2^\gamma)(\theta^0 - \hat{\theta}_2^\gamma).
\end{aligned} \tag{A.39}$$

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

$$\hat{\theta}_i^{\gamma} - \theta^0 = (T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma})^{-1}(\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^{\gamma}) = \mathcal{O}_p(1)o_p(1) = o_p(1) \tag{A.40}$$

and therefore, (A.39) simplifies to

$$T^{-1}SSR_1(\gamma) = T^{-1}\tilde{\epsilon}^{\top}\tilde{\epsilon} + o_p(1). \tag{A.41}$$

Thus, (A.41) can be written as

$$\begin{aligned} T^{-1}SSR_1(\gamma) = T^{-1}s^{\top}s &+ 2(T^{-1}s^{\top}X)(\Pi^0 - \hat{\Pi})\theta_z^0 \\ &+ \theta_z^{0\top}(\Pi^0 - \hat{\Pi})X^{\top}X(\Pi^0 - \hat{\Pi})\theta_z^0 + o_p(1) \end{aligned} \tag{A.42}$$

where $\hat{\Pi} - \Pi^0 = (T^{-1}X^{\top}X)^{-1}(T^{-1}X^{\top}u) = \mathcal{O}_p(1)o_p(1) = o_p(1)$ and $T^{-1}s^{\top}X = o_p(1)$ by Lemma 2, uniformly in γ. Thus, (A.42) simplifies to

$$\begin{aligned} T^{-1}SSR_1(\gamma) &= T^{-1}s^{\top}s + o_p(1) \\ &= T^{-1}\epsilon^{\top}\epsilon + 2(T^{-1}\epsilon^{\top}u)\theta_z^0 + \theta_z^{0\top}(T^{-1}u^{\top}u)\theta_z^0 + o_p(1) \\ &\xrightarrow{p} \sigma_{\epsilon}^2 + 2\Sigma_{\epsilon,u}^{\top}\theta_z^0 + \theta_z^{0\top}\Sigma_u\theta_z^0 \equiv \sigma^2 \end{aligned} \tag{A.43}$$

uniformly in γ. This proves part (i).

Part (ii)**.** We have

$$\begin{aligned} SSR_0 - SSR_1(\gamma) = [Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}]^{\top}[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}] &- [Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}_1^{\gamma}]^{\top}[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}_1^{\gamma}] \\ + [Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}]^{\top}[Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}] &- [Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}_2^{\gamma}]^{\top}[Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}_2^{\gamma}] \end{aligned} \tag{A.44}$$

Now, for $i = 1, 2$,

$$\begin{aligned} &[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}]^{\top}[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}] \\ &-[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma}]^{\top}[Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma}] = Y_i^{\gamma\top}Y_i^{\gamma} - 2\hat{\theta}^{\top}\hat{W}_i^{\gamma\top}Y_i^{\gamma} + \hat{\theta}^{\top}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma}\hat{\theta} \\ &\qquad - Y_i^{\gamma\top}Y_i^{\gamma} + 2\hat{\theta}_i^{\gamma\top}\hat{W}_i^{\gamma} - \hat{\theta}_i^{\gamma\top}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma} \\ &\qquad = [\hat{\theta}_i^{\gamma} - \hat{\theta}]^{\top}\hat{W}_i^{\gamma\top}[2Y_i^{\gamma} - \hat{W}_i^{\gamma}\hat{\theta} - \hat{W}_i^{\gamma}\hat{\theta}_i^{\gamma}] \\ &\qquad = T^{1/2}[\hat{\theta}_i^{\gamma} - \hat{\theta}]^{\top}\Big[2(T^{-1/2}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^{\gamma}) \\ &\qquad\quad -(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma})(T^{1/2}(\hat{\theta} - \theta^0)) \\ &\qquad\quad -(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^{\gamma})(T^{1/2}(\hat{\theta}_i^{\gamma} - \theta^0))\Big]. \end{aligned} \tag{A.45}$$

Next, we show the asymptotic behavior of the terms on the right hand side of (A.45) which then concludes the proof together with (A.43), (A.44), the continuous mapping theorem and weak convergence (uniformly in γ).

It holds that

$$
\begin{aligned}
&(T^{-1}\hat{W}^\top\hat{W})(T^{1/2}(\hat{\theta}-\theta^0))\\
=&T^{-1/2}\hat{W}^\top\tilde{\epsilon}\\
=&T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma + T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^\gamma\\
=&(T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^\gamma)(T^{1/2}(\hat{\theta}_1^\gamma-\theta^0)) + (T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^\gamma)(T^{1/2}(\hat{\theta}_2^\gamma-\theta^0))
\end{aligned}
\tag{A.46}
$$

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

$$
T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma \xrightarrow{p} C_i(\gamma). \tag{A.47}
$$

Futhermore, define $\hat{\beta}=T^{1/2}(\hat{\theta}-\theta^0)$, $\hat{\beta}_i=T^{1/2}(\hat{\theta}_i^\gamma-\theta^0)$ and $D_i(\gamma)=C^{-1}C_i(\gamma)$ $(i=1,2)$. Then, (A.46) can be restated as

$$
\hat{\beta}=D_1(\gamma)\hat{\beta}_1+D_2(\gamma)\hat{\beta}_2+o_p(1). \tag{A.48}
$$

Moreover, note that

$$
T^{1/2}(\hat{\theta}_1^\gamma-\hat{\theta})=\hat{\beta}_1-\hat{\beta}=D_2(\gamma)(\hat{\beta}_1-\hat{\beta}_2)+o_p(1) \tag{A.49}
$$

$$
T^{1/2}(\hat{\theta}_2^\gamma-\hat{\theta})=\hat{\beta}_2-\hat{\beta}=-D_1(\gamma)(\hat{\beta}_1-\hat{\beta}_2)+o_p(1) \tag{A.50}
$$

$$
T^{-1/2}\hat{W}_i^{\gamma\top}\tilde{\epsilon}_i^\gamma=C_i(\gamma)\hat{\beta}_i+o_p(1) \tag{A.51}
$$

by (A.48) and Lemma 2.
So, using (A.46)–(A.49) and (A.51), for $i=1$ quantity (A.45) can be written as

$$
\begin{aligned}
&(\hat{\beta}_1-\hat{\beta}_2)^\top D_2^\top(\gamma)\left[2C_1(\gamma)\hat{\beta}_1-C_1(\gamma)\hat{\beta}-C_1(\gamma)\hat{\beta}_1\right]+o_p(1)\\
=&(\hat{\beta}_1-\hat{\beta}_2)^\top D_2^\top(\gamma)C_1(\gamma)(\hat{\beta}_1-\hat{\beta})+o_p(1)\\
=&(\hat{\beta}_1-\hat{\beta}_2)^\top D_2^\top(\gamma)C_1(\gamma)D_2(\gamma)(\hat{\beta}_1-\hat{\beta}_2)+o_p(1).
\end{aligned}
\tag{A.52}
$$

Similarly, using (A.46)–(A.48) and (A.50)–(A.51), for $i=2$ quantity (A.45) can be stated as

$$
(\hat{\beta}_1-\hat{\beta}_2)^\top D_1^\top(\gamma)C_2(\gamma)D_1(\gamma)(\hat{\beta}_1-\hat{\beta}_2)+o_p(1). \tag{A.53}
$$

So, using (A.45), (A.52) and (A.53), quantity (A.44) can be restated as

$$
\begin{aligned}
SSR_0-SSR_1(\gamma)=&(\hat{\beta}_1-\hat{\beta}_2)^\top D_2^\top(\gamma)C_1(\gamma)D_2(\gamma)(\hat{\beta}_1-\hat{\beta}_2)\\
&+(\hat{\beta}_1-\hat{\beta}_2)^\top D_1^\top(\gamma)C_2(\gamma)D_1(\gamma)(\hat{\beta}_1-\hat{\beta}_2)+o_p(1)\\
=&(\hat{\beta}_1-\hat{\beta}_2)^\top\left[(I_p-D_1^\top(\gamma))C_1(\gamma)(I_p-D_1(\gamma))\right.\\
&\left.+D_1^\top(\gamma)(C-C_1(\gamma))D_1(\gamma)\right](\hat{\beta}_1-\hat{\beta}_2)+o_p(1)\\
=&(\hat{\beta}_1-\hat{\beta}_2)^\top\left[C_1(\gamma)-2C_1(\gamma)D_1(\gamma)+D_1^\top(\gamma)C_1(\gamma)D_1(\gamma)\right.\\
&\left.+D_1^\top(\gamma)CD_1(\gamma)-D_1^\top(\gamma)C_1(\gamma)D_1(\gamma)\right](\hat{\beta}_1-\hat{\beta}_2)+o_p(1)\\
=&(\hat{\beta}_1-\hat{\beta}_2)^\top\left[C_1(\gamma)-C_1(\gamma)D_1(\gamma)\right](\hat{\beta}_1-\hat{\beta}_2)+o_p(1)\\
=&(\hat{\beta}_1-\hat{\beta}_2)^\top C_2(\gamma)D_1(\gamma)(\hat{\beta}_1-\hat{\beta}_2)+o_p(1).
\end{aligned}
\tag{A.54}
$$

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

$$\begin{aligned}\hat{\beta}_1 - \hat{\beta}_2 &= (T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma})^{-1}(T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma}) - (T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^{\gamma})^{-1}(T^{-1/2}\hat{W}_2^{\gamma\top}\tilde{\epsilon}_2^{\gamma}) \\ &\Longrightarrow C_1^{-1}(\gamma)A^0\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)A^0\mathcal{B}_2(\gamma) \equiv \mathcal{E}(\gamma). \end{aligned} \tag{A.55}$$

So, combining (A.54) and (A.55) yields

$$SSR_0 - SSR_1(\gamma) \Longrightarrow \mathcal{E}^\top(\gamma)C_2(\gamma)D_1(\gamma)\mathcal{E}(\gamma) \tag{A.56}$$

which in turn with (A.43), the continuous mapping theorem and weak convergence (uniformly in γ) proves the claim. □

Proof of Corollary 1. We will first show the claim for the sup LR-test and afterwards for the sup Wald-test.

sup LR-test. In order to show the claim in this, we only need to show that $\mathcal{E}(\gamma) = \tilde{\mathcal{E}}(\gamma)$ under Assumptions A.1 and A.2. Or in other words, that $\mathcal{GP}_{\text{mat},1}(\gamma) = \tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)Q$[15]:
The covariance kernel of $\mathcal{GP}_1(\gamma)$ is given as $\mathbb{E}[\mathcal{GP}_1(\gamma_1)\mathcal{GP}_1^\top(\gamma_2)] = \mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \gamma_1 \wedge \gamma_2\}}]$ by Lemma 1. Under Assumption A.2 this expression can be simplified to

$$\begin{aligned}\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \gamma_1 \wedge \gamma_2\}}] &= \mathbb{E}\left[\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \gamma_1 \wedge \gamma_2\}}|x_t, q_t]\right] \\ &= \mathbb{E}\left[\mathbb{E}[v_t v_t^\top|x_t, q_t] \otimes x_t x_t^\top \mathbb{1}_{\{q_t \le \gamma_1 \wedge \gamma_2\}}\right] \\ &= \mathbb{E}\left[\Sigma \otimes x_t x_t^\top \mathbb{1}_{\{q_t \le \gamma_1 \wedge \gamma_2\}}\right] \\ &= \Sigma \otimes M_1(\gamma_1, \gamma_2). \end{aligned} \tag{A.57}$$

Note that Q denotes the principal square root of Σ, i.e. $Q^{1/2}Q^{1/2} = \Sigma$. Then, (A.57) can be restated as

$$\begin{aligned}\mathbb{E}[(v_t v_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \gamma_1 \wedge \gamma_2\}}] &= \Sigma \otimes M_1(\gamma_1 \wedge \gamma_2) \\ &= (Q \otimes M_1(\gamma_1 \wedge \gamma_2))(Q \otimes I) \\ &= (Q \otimes I)(I \otimes M_1(\gamma_1 \wedge \gamma_2))(Q \otimes I). \end{aligned} \tag{A.58}$$

On the other hand, the covariance kernel of $(Q \otimes I)\tilde{\mathcal{GP}}_1(\gamma) = \text{vec}(\tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)Q)$ is given by

$$\begin{aligned}\mathbb{E}[(Q \otimes I)\tilde{\mathcal{GP}}_1(\gamma_1)\tilde{\mathcal{GP}}_1^\top(\gamma_2)(Q \otimes I)] &= (Q \otimes I)\mathbb{E}[\tilde{\mathcal{GP}}_1(\gamma_1)\tilde{\mathcal{GP}}_1^\top(\gamma_2)](Q \otimes I) \\ &= (Q \otimes I)(I \otimes M_1(\gamma_1 \wedge \gamma_2))(Q \otimes I) \end{aligned} \tag{A.59}$$

because $\mathbb{E}[\tilde{\mathcal{GP}}_1(\gamma_1)\tilde{\mathcal{GP}}_1^\top(\gamma_2)] = I \otimes M_1(\gamma_1 \wedge \gamma_2)$ by definition of $\tilde{\mathcal{GP}}_1(\gamma)$. Thus, combing (A.58) and (A.59) yields the desired result since Gaussian processes are uniquely defined through their mean and covariance functions.

[15] 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.

sup Wald-test. In order to show the claim for this case, we are left to show that $[C_1^{-1}(\gamma)H_1(\gamma)C_1^{-1}(\gamma)+C_2^{-1}(\gamma)H_2(\gamma)C_2^{-1}(\gamma)]^{-1} = \sigma^{-2}C_2(\gamma)C^{-1}C_1(\gamma)$ under Assumptions A.1 and A.2. The equality of $\mathcal{E}(\gamma)$ and $\tilde{\mathcal{E}}(\gamma)$ has already been shown in the previous part of the proof.
Under assumption A.2 it follows for $H_1(\gamma)$ that

$$\begin{aligned} H_1(\gamma) &= A^0\mathbb{E}[x_t x_t^\top \{\epsilon_t + u_t^\top \theta_z^0\}^2 \mathbb{1}_{\{q_t \le \gamma\}}]A^{0\top} \\ &= A^0\mathbb{E}\left[\mathbb{E}[x_t x_t^\top \{\epsilon_t + u_t^\top \theta_z^0\}^2 \mathbb{1}_{\{q_t \le \gamma\}}|x_t, q_t]\right] A^{0\top} \\ &= A^0\mathbb{E}\left[x_t x_t^\top \mathbb{1}_{\{q_t \le \gamma\}}\mathbb{E}[(\epsilon_t + u_t^\top \theta_z^0)^2|x_t, q_t]\right] A^{0\top} \\ &= A^0\mathbb{E}[x_t x_t^\top \mathbb{1}_{\{q_t \le \gamma\}}(\sigma_\epsilon^2 + 2\Sigma_{\epsilon,u}^\top \theta_z^0 + \theta_z^{0\top}\Sigma_u\theta_z^0)]A^{0\top} \\ &= \sigma^2 A^0 M_1(\gamma)A^{0\top} \\ &= \sigma^2 C_1(\gamma). \end{aligned} \tag{A.60}$$

Similarly, for $H_2(\gamma)$ it holds under Assumption A.2 that

$$H_2(\gamma) = \sigma^2 C_2(\gamma). \tag{A.61}$$

Therefore, using (A.60) and (A.61), it follows that

$$\begin{aligned} [C_1^{-1}(\gamma)H_1(\gamma)C_1^{-1}(\gamma) + C_2^{-1}(\gamma)H_2(\gamma)C_2^{-1}(\gamma)]^{-1} &= [\sigma^2 C_1^{-1}(\gamma) + \sigma^2 C_2^{-1}(\gamma)]^{-1} \\ &= \sigma^{-2}[C_1^{-1}(\gamma) + C_2^{-1}(\gamma)]^{-1}. \end{aligned} \tag{A.62}$$

Having (A.62) we are left to show that $[C_1^{-1}(\gamma) + C_2^{-1}(\gamma)]^{-1} = C_2(\gamma)C^{-1}C_1(\gamma)$ or equivalently that $C_1^{-1}(\gamma) + C_2^{-1}(\gamma) = C_1^{-1}(\gamma)CC_2^{-1}(\gamma)$. To do so, we will rewrite expression (A.62) and make use of the Kailath identity in line 5[16] (cf. Kailath (1980)):

$$\begin{aligned} C_1^{-1}(\gamma) + C_2^{-1}(\gamma) &= C_1^{-1}(\gamma) + [C - CC^{-1}C_1(\gamma)]^{-1} \\ &= C_1^{-1}(\gamma) + [C(I - C^{-1}C_1(\gamma))]^{-1} \\ &= C_1^{-1}(\gamma) + [I - C^{-1}C_1(\gamma)]^{-1}C^{-1} \\ &= C_1^{-1}(\gamma) - C_1^{-1}(\gamma)(-C_1(\gamma))[I + C^{-1}C_1(\gamma)C_1^{-1}(\gamma)(-C_1(\gamma))]C^{-1}C_1(\gamma)C_1^{-1}(\gamma) \\ &= [C_1(\gamma) - C_1(\gamma)C^{-1}C_1(\gamma)]^{-1} \\ &= [(I - C_1(\gamma)C^{-1})C_1(\gamma)]^{-1} \\ &= [(C - C_1(\gamma))C^{-1}C_1(\gamma)]^{-1} \\ &= C_1^{-1}(\gamma)C^{-1}C_2^{-1}(\gamma). \end{aligned} \tag{A.63}$$

Finally, equations (A.60)–(A.63) yield the desired result that

$$\mathcal{E}^\top(\gamma)[C_1^{-1}(\gamma)H_1(\gamma)C_1^{-1}(\gamma) + C_2^{-1}(\gamma)H_2(\gamma)C_2^{-1}(\gamma)]^{-1}\mathcal{E}(\gamma) = \frac{\tilde{\mathcal{E}}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\tilde{\mathcal{E}}(\gamma)}{\sigma^2}$$

[16] The Kailath identity reads as follows:

Let A be a non-singular $n \times n$, B an $n \times k$ and D an $k \times n$ matrix such that $A + BD$ is non-singular. Then,

$$(A + BD)^{-1} = A^{-1} - A^{-1}B(I + DA^{-1}B)^{-1}DA^{-1}.$$

In our case set $A = C_1(\gamma)$, $B = -C_1(\gamma)$ and $D = C^{-1}C_1(\gamma)$ to apply the identity.

under Assumption A.2 and therefore, proving the claim. □

Proof of Corollary 2. First we note that by Assumption A.1.4 it follows that $\text{Prob}(q_t \leq \gamma) = G(\gamma)$ is continuous. Further, we will replace the threshold parameter γ by an equivalent value, say λ, defined on the open unit interval $(0;1)$. To see how this works, note first that $\Gamma \subset (\gamma_{\min}, \gamma_{\max})$. Then, $\text{Prob}(q_t \leq \gamma_{\min}) = 0$ and $\text{Prob}(q_t \leq \gamma_{\max}) = 1$. 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

$$M_1(\gamma_1 \wedge \gamma_2) = \mathbb{E}[x_t x_t^\top \mathbb{1}_{\{q_t \leq \gamma_1 \wedge \gamma_2\}}] = \mathbb{E}[x_t x_t^\top]\mathbb{E}[\mathbb{1}_{\{q_t \leq \gamma_1 \wedge \gamma_2\}}] = \min\{\lambda_1, \lambda_2\} M. \tag{A.64}$$

This also implies that

$$M_1(\gamma) = \lambda M \tag{A.65a}$$
$$C_1(\gamma) = A^0 M_1(\gamma) A^{0\top} = \lambda A^0 M A^{0\top} = \lambda C \tag{A.65b}$$
$$M_2(\gamma) = (1-\lambda) M \tag{A.65c}$$
$$C_2(\gamma) = A^0 M_2(\gamma) A^{0\top} = (1-\lambda) A^0 M A^{0\top} = (1-\lambda) C. \tag{A.65d}$$

Moreover, (A.64) implies that –under Assumptions A.2 and A.3– the Gaussian process $\mathcal{GP}_1(\gamma)$ can be restated as

$$\mathcal{GP}_1(\gamma) = (Q \otimes I)\tilde{\mathcal{GP}}_1(\gamma) = (Q \otimes M^{1/2})\mathcal{BM}_1(\lambda) \iff \mathcal{GP}_{\text{mat},1}(\gamma) = M^{1/2}\mathcal{BM}_{\text{mat},1}(\lambda) Q \tag{A.66}$$

where $\mathcal{BM}_1(\lambda)$ denotes a $q(p_1+1) \times 1$-Brownian motion on the open unit interval, implying that $\mathcal{B}_1(\gamma)$ can be restated in terms of λ as $\mathcal{B}_1(\lambda)$ (we will see the exact expression later in this proof).
Therefore, we obtain

$$\begin{aligned}
\mathcal{E}^\top(\gamma) C_2(\gamma) C^{-1} C_1(\gamma) \mathcal{E}(\gamma) &= [C_1^{-1}(\gamma)\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)]^\top \\
&\times C_2(\gamma) C^{-1} C_1(\gamma) \\
&\times [C_1^{-1}(\gamma)\mathcal{B}_1(\gamma) - C_2^{-1}(\gamma)\mathcal{B}_2(\gamma)] \\
&= \frac{1}{\lambda(1-\lambda)}[C^{-1}\mathcal{B}_1(\lambda) - \lambda C^{-1}\mathcal{B}_1(1)]^\top \\
&\times C[C^{-1}\mathcal{B}_1(\lambda) - \lambda C^{-1}\mathcal{B}_1(1)] \\
&= \frac{1}{\lambda(1-\lambda)}[C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1)]^\top \\
&\times [C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1)].
\end{aligned} \tag{A.67}$$

Next, we show that the term $C^{-1/2}\mathcal{B}_1(\lambda)$ equals in distribution $[(Q\tilde{\theta}_z^0)^\top \otimes I]\overline{\mathcal{BM}}_1(\lambda) - \lambda[(\check{\theta}_z^0)^\top \otimes I]\overline{\mathcal{BM}}_1(1)$. Because of (A.65a) and (A.66) it follows that $\mathcal{B}_1(\lambda) = A^0[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 -$

$M_1(\gamma)M^{-1}\mathcal{GP}_{\text{mat},1}\check{\theta}_z^0] = A^0[M^{1/2}\mathcal{BM}_{\text{mat},1}(\lambda)Q\tilde{\theta}_z^0 - \lambda M^{1/2}\mathcal{BM}_{\text{mat},1}(1)Q\check{\theta}_z^0]$. Further, recall that $C = A^0MA^{0\top}$. Thus:

$$\begin{aligned}
C^{-1/2}\mathcal{B}_1(\lambda) &= (A^0MA^{0\top})^{-1/2}A^0M^{1/2}\mathcal{BM}_{\text{mat},1}(\lambda)Q\tilde{\theta}_z^0 \\
&- \lambda(A^0MA^{0\top})^{-1/2}A^0M^{1/2}\mathcal{BM}_{\text{mat},1}(1)Q\check{\theta}_z^0 \\
&= [(Q\tilde{\theta}_z^0)^\top \otimes (A^0MA^{0\top})^{-1/2}AM^{1/2}]\mathcal{BM}_1(\lambda) \\
&- \lambda[(Q\check{\theta}_z^0)^\top \otimes (A^0MA^{0\top})^{-1/2}AM^{1/2}]\mathcal{BM}_1(1) \\
&= \Phi - \lambda\Psi
\end{aligned} \tag{A.68}$$

where

$$\begin{aligned}
\Phi &\sim \mathcal{N}\left(0, [(Q\tilde{\theta}_z^0)^\top \otimes (A^0MA^{0\top})^{-1/2}AM^{1/2}][(Q\tilde{\theta}_z^0) \otimes M^{1/2}A^{0\top}(A^0MA^{0\top})^{-1/2}]\right) \\
&= \mathcal{N}\left(0, [(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)] \otimes I\right) \qquad \text{(A.69a)} \\
\Psi &\sim \mathcal{N}\left(0, [(Q\check{\theta}_z^0)^\top \otimes (A^0MA^{0\top})^{-1/2}AM^{1/2}][(Q\check{\theta}_z^0) \otimes M^{1/2}A^{0\top}(A^0MA^{0\top})^{-1/2}]\right) \\
&= \mathcal{N}\left(0, [(Q\check{\theta}_z^0)^\top(Q\check{\theta}_z^0)] \otimes I\right). \qquad \text{(A.69b)}
\end{aligned}$$

Thus, it holds that

$$\Phi \overset{\mathcal{D}}{=} [(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(\lambda) \tag{A.70a}$$

$$\Psi \overset{\mathcal{D}}{=} [(Q\check{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(1) \tag{A.70b}$$

implying for (A.68) that

$$C^{-1/2}\mathcal{B}_1(\lambda) \overset{\mathcal{D}}{=} [(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(\lambda) - \lambda[(Q\check{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(1). \tag{A.71}$$

So, we have that

$$\begin{aligned}
C^{-1/2}\mathcal{B}_1(\lambda) - \lambda C^{-1/2}\mathcal{B}_1(1) &\overset{\mathcal{D}}{=} [(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(\lambda) - \lambda[(Q\check{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(1) \\
&- \lambda[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(1) + \lambda[(Q\check{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(1) \\
&= [(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(\lambda) - \lambda[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BM}_1(1)
\end{aligned} \tag{A.72}$$

Using (A.72), (A.67) can be written as

$$\mathcal{E}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\mathcal{E}(\gamma) \overset{\mathcal{D}}{=} \frac{\left\{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)\right\}^\top \left\{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)\right\}}{\lambda(1-\lambda)} \tag{A.73}$$

where $\mathcal{BB}(\lambda) = \mathcal{BM}_1(\lambda) - \lambda\mathcal{BM}_1(1)$.
Finally, to show the claim we need to divide (A.73) by $\sigma^2 = \tilde{\theta}_z^{0\top}\Sigma\tilde{\theta}_z^0 = (Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)$ and obtain

$$\begin{aligned}
\frac{\mathcal{E}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\mathcal{E}(\gamma)}{\sigma^2} &\overset{\mathcal{D}}{=} \frac{\left\{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)\right\}^\top \left\{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)\right\}}{\lambda(1-\lambda)(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)} \\
&= \frac{\left\{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)\right\}^\top \left\{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)\right\}}{\lambda(1-\lambda)(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)}.
\end{aligned} \tag{A.74}$$

Further, note that

$$\begin{aligned}\frac{[(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda)}{[(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)]^{1/2}} &= [[(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)]^{-1/2}(Q\tilde{\theta}_z^0)^\top \otimes I]\mathcal{BB}(\lambda) \\ &\sim \mathcal{N}\left(0, [(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)]^{-1/2}(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)[(Q\tilde{\theta}_z^0)^\top(Q\tilde{\theta}_z^0)]^{-1/2} \otimes I\right) \\ &= \mathcal{N}(0, I). \end{aligned} \tag{A.75}$$

Therefore, it holds that

$$\frac{\mathcal{E}^\top(\gamma)C_2(\gamma)C^{-1}C_1(\gamma)\mathcal{E}(\gamma)}{\sigma^2} \stackrel{\mathcal{D}}{=} \frac{\mathcal{BB}^\top(\lambda)\mathcal{BB}(\lambda)}{\lambda(1-\lambda)}, \tag{A.76}$$

proving the claim. □

B. 2SLS Results involving a Threshold Reduced Form

Lemma 3. *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 that the distribution is as if ρ^0 was known we have that*

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

holds. Further, $V_1 = (I_{p_1} \otimes M_1^{-1}(\rho^0))\mathbb{E}[(u_t u_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t \le \rho^0\}}](I_{p_1} \otimes M_1^{-1}(\rho^0))$ and $V_2 = (I_{p_1} \otimes M_2^{-1}(\rho^0))\mathbb{E}[(u_t u_t^\top \otimes x_t x_t^\top)\mathbb{1}_{\{q_t > \rho^0\}}](I_{p_1} \otimes M_2^{-1}(\rho^0))$.

Proof of Lemma 3. We will prove the statement for $T^{1/2}\operatorname{vec}(\hat{\Pi}_1(\rho^0) - \Pi_1^0)$. The proof for $T^{1/2}\operatorname{vec}(\hat{\Pi}_2(\rho^0) - \Pi_2^0)$ follows similar arguments and is therefore 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 $X_1^{\rho^0\top}X_2^{\rho^0} = \mathbf{0}$. So, the "cross"-terms cancel. Hence,

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

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

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

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,

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

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,

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

which concludes the proof. □

Lemma 4. *Suppose Assumption A.1 holds. Then, under* $\mathbb{H}_0$*,*

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

and

$$\begin{aligned} T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Longrightarrow \quad & A_1^0\Big[\mathcal{GP}_{\mathrm{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma \wedge \rho^0)M_1^{-1}(\rho^0)\mathcal{GP}_{\mathrm{mat},1}(\rho^0)\check{\theta}_z^0\Big] \\ & + A_2^0\Big[(\mathcal{GP}_{\mathrm{mat},1}(\gamma) - \mathcal{GP}_{\mathrm{mat},1}(\gamma \wedge \rho^0))\tilde{\theta}_z^0 \\ & \qquad - (M_1(\gamma) - M_1(\gamma \wedge \rho^0))M_2^{-1}(\rho^0)\mathcal{GP}_{\mathrm{mat},2}(\rho^0)\check{\theta}_z^0\Big] \\ & \equiv \mathcal{B}_{A,1}(\gamma) \end{aligned}$$

Proof of Lemma 4. 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 $\gamma = \rho^0 + \mathcal{O}(1)$ by construction. This implies two sub-cases: $(a.1)$ with $\gamma < \hat{\rho} \le \rho^0$ and $(a.2)$ with $\gamma < \rho^0 < \hat{\rho}$.
- In case (b) there are two sub-cases: $(b.1)$ with $\gamma = \rho^0 \le \hat{\rho}$ 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 $\gamma = \rho^0 + \mathcal{O}(1)$ by construction. This implies two sub-cases: $(c.1)$ with $\hat{\rho} \le \rho^0 < \gamma$ 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

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

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

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1(T^{-1}X_1^{\gamma\top}X_1^{\gamma})\hat{A}_1^{\top} \\ &= A_1^0(T^{-1}X_1^{\gamma\top}X_1^{\gamma})A_1^{0\top} + o_p(1) \\ &\xrightarrow{p} A_1^0 M_1(\gamma)A_1^{0\top} \end{aligned} \tag{B.2}$$

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

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= T^{-1}\hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}} + T^{-1}(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} - \hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}}) \\ &= \hat{A}_1(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_1^{\top} + \hat{A}_2(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} - T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_2^{\top}, \end{aligned} \tag{B.3}$$

because $\hat{\rho} < \gamma = \rho^0$. By Lemma 2 we have that $\hat{\rho} = \rho^0 + \mathcal{O}_p(T^{-1})$ and therefore,

$$\begin{aligned} T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} &= T^{-1}\sum_{t=1}^{T} x_t x_t^{\top}\mathbb{1}_{\{q_t \leq \hat{\rho}\}} \\ &= T^{-1}\sum_{t=1}^{T} x_t x_t^{\top}\mathbb{1}_{\{q_t \leq \rho^0\}} + T^{-1}\sum_{t=1}^{T} x_t x_t^{\top}(\mathbb{1}_{\{q_t \leq \hat{\rho}\}} - \mathbb{1}_{\{q_t \leq \rho^0\}}) \\ &= T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} + \mathcal{O}_p(T^{-1}) \\ &= T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} + o_p(1). \end{aligned} \tag{B.4}$$

So, (B.3), (B.4) and Lemma 2 imply, uniformly in γ,

$$T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} \xrightarrow{p} A_1^0 M_1(\rho^0)A_1^{0\top} = A_1^0 M_1(\gamma)A_1^{0\top}. \tag{B.5}$$

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

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= T^{-1}\hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}} + T^{-1}(\hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0} - \hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}}) \\ &\quad + T^{-1}(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} - \hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0}) \\ &= \hat{A}_1(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_1^{\top} + \hat{A}_2(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} - T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_2^{\top} \\ &\quad + \hat{A}_2(T^{-1}X_1^{\gamma\top}X_1^{\gamma} - T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})\hat{A}_2^{\top} \\ &\xrightarrow{p} A_1^0 M_1(\rho^0)A_1^{0\top} + A_2^0(M_1(\gamma) - M_1(\rho^0))A_2^{0\top} \end{aligned} \tag{B.6}$$

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

$$\begin{aligned} T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} &= T^{-1}\hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0} + T^{-1}(\hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}} - \hat{W}_1^{\rho^0\top}\hat{W}_1^{\rho^0}) \\ &\quad + T^{-1}(\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma} - \hat{W}_1^{\hat{\rho}\top}\hat{W}_1^{\hat{\rho}}) \\ &= \hat{A}_1(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})\hat{A}_1^{\top} + \hat{A}_1(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} - T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})\hat{A}_1^{\top} \\ &\quad + \hat{A}_2(T^{-1}X_1^{\gamma\top}X_1^{\gamma} - T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})\hat{A}_2^{\top} \\ &\xrightarrow{p} A_1^0 M_1(\rho^0)A_1^{0\top} + A_2^0(M_1(\gamma) - M_1(\rho^0))A_2^{0\top}. \end{aligned} \tag{B.7}$$

So, putting results (B.1), (B.2), (B.5)–(B.7) together yields the claim.

Claim (ii)**.** To show this claim, we present the underlying ideas for case (a). Since cases (b) and (c) follows 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

$$\begin{aligned} T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1(T^{-1/2}X_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma}) \\ &= \hat{A}_1(T^{-1/2}X_1^{\gamma\top}(\epsilon_1^{\gamma} + (Z_1^{\gamma} - \hat{Z}_1^{\gamma})\theta_z^0) \\ &= \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}(\epsilon_1^{\gamma} + (X_1^{\gamma}\Pi_1^0 + u_1^{\gamma} - X_1^{\gamma}\hat{\Pi}_1)\theta_z^0)\right] \\ &= \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} - (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right], \end{aligned} \tag{B.8}$$

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

$$\begin{aligned} (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0 &= (T^{-1}X_1^{\gamma\top}X_1^{\gamma})(T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})^{-1}(T^{-1/2}X_1^{\hat{\rho}\top}u_1^{\hat{\rho}})\theta_z^0 \\ &\xrightarrow{p} M_1(\gamma)M_1^{-1}(\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0 \end{aligned} \tag{B.9}$$

Therefore, (B.8) behaves uniformly in γ as

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Longrightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma)M_1^{-1}(\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right]. \tag{B.10}$$

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

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} = \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} - (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right]. \tag{B.11}$$

However, in contrast to sub-case $(a.1)$, it now holds that[17]

$$\hat{\Pi}_1 - \Pi_1^0 = (X_1^{\rho^0\top}X_1^{\rho^0})^{-1}(X_1^{\rho^0\top}u_1^{\rho^0}) + o_p(1) \tag{B.12}$$

because

$$\begin{aligned} \hat{\Pi}_1 &= (X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})^{-1}(X_1^{\hat{\rho}\top}Z_1^{\hat{\rho}}) \\ &= (X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})^{-1}(X_1^{\rho^0\top}X_1^{\rho^0}\Pi_1^0 + X_1^{\rho^0\top}X_1^{\rho^0}\Pi_2^0 - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}}\Pi_2^0 + X_1^{\hat{\rho}\top}u_1^{\hat{\rho}}) \\ &= \Pi_1^0 + (X_1^{\rho^0\top}X_1^{\rho^0})^{-1}(X_1^{\rho^0\top}u_1^{\rho^0}) + o_p(1) \end{aligned} \tag{B.13}$$

[17]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.13). E.g. in (B.13) 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 $\hat{\rho}$.

by Lemma 2. So, putting (B.11) and (B.13) together yields uniformly in γ that

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Longrightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma)M_1^{-1}(\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0\right]. \tag{B.14}$$

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

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} = \hat{A}_1\left[T^{-1/2}X_1^{\gamma\top}s_1^{\gamma} - (T^{-1}X_1^{\gamma\top}X_1^{\gamma})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \tag{B.15}$$

with

$$\hat{\Pi}_1 - \Pi_1^0 = (X_1^{\rho^0\top}X_1^{\rho^0})^{-1}(X_1^{\rho^0\top}u_1^{\rho^0}) + o_p(1). \tag{B.16}$$

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

$$T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} \Longrightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - M_1(\gamma)M_1^{-1}(\rho^0)\mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0\right], \tag{B.17}$$

where $M_1(\gamma)M_1^{-1}(\rho^0)$ cancels whenever $\gamma = \rho^0$.
For sub-case $(b.2)$ it holds uniformly in γ that

$$\begin{aligned}
T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1\left[T^{-1/2}X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - (T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \\
&\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\rho^0\top}s_1^{\rho^0} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) - T^{-1}(X_1^{\rho^0\top}X_1^{\rho^0} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\Longrightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\gamma)\check{\theta}_z^0\right]
\end{aligned} \tag{B.18}$$

by Lemmata 1, 2 and (B.4).
Last, we show the claim for case (c). In sub-case $(c.1)$ it holds uniformly in γ that

$$\begin{aligned}
T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1\left[T^{-1/2}X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - (T^{-1}X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \\
&\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\rho^0\top}s_1^{\rho^0} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) - T^{-1}(X_1^{\rho^0\top}X_1^{\rho^0} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\quad + \hat{A}_2\left[T^{-1/2}(X_1^{\gamma\top}s_1^{\gamma} - X_1^{\rho^0\top}s_1^{\rho^0}) - T^{-1}(X_1^{\gamma\top}X_1^{\gamma} - X_1^{\rho^0\top}X_1^{\rho^0})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\Longrightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right] \\
&\quad + A_2^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - (M_1(\gamma) - M_1(\rho^0))M_2^{-1}(\rho^0)\mathcal{GP}_{\text{mat},2}(\rho^0)\check{\theta}_z^0\right],
\end{aligned} \tag{B.19}$$

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 2.

Last, sub-case $(c.2)$ yields uniformly in γ

$$
\begin{aligned}
T^{-1/2}\hat{W}_1^{\gamma\top}\tilde{\epsilon}_1^{\gamma} &= \hat{A}_1\left[T^{-1/2}X_1^{\rho^0\top}s_1^{\rho^0} - (T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})T^{1/2}(\hat{\Pi}_1 - \Pi_1^0)\theta_z^0\right] \\
&+ \hat{A}_1\left[T^{-1/2}(X_1^{\hat{\rho}\top}s_1^{\hat{\rho}} - X_1^{\rho^0\top}s_1^{\rho^0}) - T^{-1}(X_1^{\hat{\rho}\top}X_1^{\hat{\rho}} - X_1^{\rho^0\top}X_1^{\rho^0})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&+ \hat{A}_2\left[T^{-1/2}(X_1^{\gamma\top}s_1^{\gamma} - X_1^{\hat{\rho}\top}s_1^{\hat{\rho}}) - T^{-1}(X_1^{\gamma\top}X_1^{\gamma} - X_1^{\hat{\rho}\top}X_1^{\hat{\rho}})T^{1/2}(\hat{\Pi}_2 - \Pi_2^0)\theta_z^0\right] \\
&\Longrightarrow A_1^0\left[\mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\check{\theta}_z^0\right] \\
&+ A_2^0\left[\mathcal{GP}_{\text{mat},1}(\gamma)\tilde{\theta}_z^0 - \mathcal{GP}_{\text{mat},1}(\rho^0)\tilde{\theta}_z^0 - (M_1(\gamma) - M_1(\rho^0))M_2^{-1}(\rho^0)\mathcal{GP}_{\text{mat},2}(\rho^0)\check{\theta}_z^0\right],
\end{aligned}
\tag{B.20}
$$

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 2.
Finally, putting (B.10), (B.14), (B.17)–(B.20) together immediately yields the claim. □

Proof of Theorem 2. In order to show the statement of Theorem 2, we first prove it for the sup LR test and afterwards for the sup Wald test.

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 $SSR_0 - SSR_1(\gamma) \Longrightarrow \mathcal{E}_A^\top(\gamma)D_{A,2}(\gamma)C_{A,1}(\gamma)\mathcal{E}(\gamma)$.

Part (i)**.** As in the LRF proof (cf. equation (A.41)) it holds uniformly in γ that

$$
\begin{aligned}
T^{-1}SSR_1(\gamma) &= T^{-1}[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}_1^{\gamma}]^\top[Y_1^{\gamma} - \hat{W}_1^{\gamma}\hat{\theta}_1^{\gamma}] \\
&+ T^{-1}[Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}_2^{\gamma}]^\top[Y_2^{\gamma} - \hat{W}_2^{\gamma}\hat{\theta}_2^{\gamma}] \\
&= T^{-1}[\hat{W}_1^{\gamma}(\theta^0 - \hat{\theta}_1^{\gamma}) + \tilde{\epsilon}_1^{\gamma}]^\top[\hat{W}_1^{\gamma}(\theta^0 - \hat{\theta}_1^{\gamma}) + \tilde{\epsilon}_1^{\gamma}] \\
&+ T^{-1}[\hat{W}_2^{\gamma}(\theta^0 - \hat{\theta}_2^{\gamma}) + \tilde{\epsilon}_2^{\gamma}]^\top[\hat{W}_2^{\gamma}(\theta^0 - \hat{\theta}_2^{\gamma}) + \tilde{\epsilon}_2^{\gamma}] \\
&= T^{-1}\tilde{\epsilon}^\top\tilde{\epsilon} \\
&+ 2(T^{-1}\tilde{\epsilon}_1^{\gamma}\hat{W}_1^{\gamma})(\theta^0 - \hat{\theta}_1^{\gamma}) + (\theta^0 - \hat{\theta}_1^{\gamma})^\top(T^{-1}\hat{W}_1^{\gamma\top}\hat{W}_1^{\gamma})(\theta^0 - \hat{\theta}_1^{\gamma}) \\
&+ 2(T^{-1}\tilde{\epsilon}_2^{\gamma}\hat{W}_2^{\gamma})(\theta^0 - \hat{\theta}_2^{\gamma}) + (\theta^0 - \hat{\theta}_2^{\gamma})^\top(T^{-1}\hat{W}_2^{\gamma\top}\hat{W}_2^{\gamma})(\theta^0 - \hat{\theta}_2^{\gamma}) \\
&= T^{-1}\tilde{\epsilon}^\top\tilde{\epsilon} + o_p(1),
\end{aligned}
\tag{B.21}
$$

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.21) as

$$
T^{-1}SSR_1(\gamma) = T^{-1}\tilde{\epsilon}_1^{\rho^0\top}\tilde{\epsilon}_1^{\rho^0} + T^{-1}\tilde{\epsilon}_2^{\rho^0\top}\tilde{\epsilon}_2^{\rho^0} + o_p(1). \tag{B.22}
$$

By construction

$$\tilde{\epsilon}_1^{\rho^0} = \epsilon_1^{\rho^0} + (Z_1^{\rho^0} - \hat{Z}_1^{\rho^0})\theta_z^0 \tag{B.23}$$

and thus

$$\tilde{\epsilon}_1^{\rho^0} = \begin{cases} s_1^{\rho^0} + X_1^{\rho^0}(\Pi_1^0 - \hat{\Pi}_1) & \text{if } \rho^0 \leq \hat{\rho} \\ s_1^{\rho^0} + X_1^{\rho^0}(\Pi_1^0 - \hat{\Pi}_1) + o_p(1) & \text{if } \rho^0 > \hat{\rho} \end{cases} \tag{B.24}$$

Therefore, we obtain

$$\begin{aligned} T^{-1}\tilde{\epsilon}_1^{\rho^0\top}\tilde{\epsilon}_1^{\rho^0} &= T^{-1}s_1^{\rho^0\top}s_1^{\rho^0} + 2(T^{-1}s_1^{\rho^0\top}X_1^{\rho^0})(\Pi_1^0 - \hat{\Pi}_1) \\ &\quad + (\Pi_1^0 - \hat{\Pi}_1)^\top(T^{-1}X_1^{\rho^0\top}X_1^{\rho^0})(\Pi_1^0 - \hat{\Pi}_1) \\ &= T^{-1}s_1^{\rho^0\top}s_1^{\rho^0} + o_p(1) \end{aligned} \tag{B.25}$$

because $T^{-1}s_1^{\rho^0\top}X_1^{\rho^0} = o_p(1)$ and $T^{-1}X_1^{\rho^0\top}X_1^{\rho^0} = \mathcal{O}_p(1)$ by Lemma 3 and $\Pi_1^0 - \hat{\Pi}_1 = o_p(1)$ by Lemma 2.
Similarly, we obtain

$$T^{-1}\tilde{\epsilon}_2^{\rho^0\top}\tilde{\epsilon}_2^{\rho^0} = T^{-1}s_2^{\rho^0\top}s_2^{\rho^0} + o_p(1). \tag{B.26}$$

Therefore, (B.22) reads as

$$\begin{aligned} T^{-1}SSR_1(\gamma) &= T^{-1}s_1^{\rho^0\top}s_1^{\rho^0} + T^{-1}s_2^{\rho^0\top}s_2^{\rho^0} + o_p(1) \\ &= T^{-1}s^\top s + o_p(1) \\ &\xrightarrow{p} \sigma_\epsilon^2 + 2\Sigma_{\epsilon,u}^\top\theta_z^0 + \theta_z^{0\top}\Sigma_u\theta_z^0 \equiv \sigma^2, \end{aligned} \tag{B.27}$$

uniformly in γ, proving part (i).

Part (ii)**.** For this part, derivations remain as in the LRF case (up to equation (A.46)). Utilizing Lemma 4, expressions (A.47) and (A.48) in the LRF proof become

$$T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma \xrightarrow{p} C_{A,i}(\gamma) \tag{B.28}$$

and

$$\hat{\beta} = D_{A,1}(\gamma)\hat{\beta}_1 + D_{A,2}(\gamma)\hat{\beta}_2 + o_p(1) \tag{B.29}$$

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 (A.47)–(A.49) in the LRF proof are adjusted in this fashion as well. The following derivations then remain the same.
Last, equation (A.55) from the LRF case now reads as [18]

$$\hat{\beta}_1 - \hat{\beta}_2 = C_{A,1}^{-1}(\gamma)\mathcal{B}_{A,1}(\gamma) - C_{A,2}^{-1}(\gamma)\mathcal{B}_{A,2}(\gamma) \equiv \mathcal{E}_A(\gamma). \tag{B.30}$$

Thus, as in the LRF case, it follows that

$$\begin{aligned} SSR_0 - SSR_1(\gamma) &= (\hat{\beta}_1 - \hat{\beta}_2)^\top C_{A,2}(\gamma)D_{A,1}(\gamma)(\hat{\beta}_1 - \hat{\beta}_2) + o_p(1) \\ &\Longrightarrow \mathcal{E}_A^\top(\gamma)C_{A,2}(\gamma)D_{A,1}(\gamma)\mathcal{E}_A(\gamma) \end{aligned} \tag{B.31}$$

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

uniformly in γ. Together with (B.27), $(a.s.)$ continuity of the process $\mathcal{E}_A(\gamma)$, the continuous mapping theorem and weak convergence (uniformly in γ) it then follows that

$$\sup_{\gamma\in\Gamma} \frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/T} \Longrightarrow \sup_{\gamma\in\Gamma} \frac{\mathcal{E}_A^\top(\gamma) C_{A,2}(\gamma) D_{A,1}(\gamma) \mathcal{E}_A(\gamma)}{\sigma^2} \tag{B.32}$$

proving the claim of the theorem.

sup Wald Test: The Wald-type test statistic reads as

$$\sup_{\gamma\in\Gamma} W_{T,\mathrm{TRF}}^{\mathrm{2SLS}}(\gamma) = \sup_{\gamma\in\Gamma} T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma)^\top [T\,\mathrm{Var}(\hat{\theta}_1^\gamma) + T\,\mathrm{Var}(\hat{\theta}_2^\gamma)]^{-1} T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma).$$

As already shown above, (pp. 28, cf. equation (B.30) with $\hat{\beta}_i = T^{1/2}(\hat{\theta}_i^\gamma - \theta^0)$ for $i = 1, 2$) it holds uniformly in γ that

$$T^{1/2}(\hat{\theta}_1^\gamma - \hat{\theta}_2^\gamma) \Longrightarrow \mathcal{E}_A(\gamma) \tag{B.33}$$

with $\mathcal{E}_A(\gamma)$ as defined in (B.30).
So, we are left to derive the asymptotic behavior of

$$T\,\mathrm{Var}(\hat{\theta}_i^\gamma) = (T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1}\Big(T^{-1}\sum_{\mathcal{T}_i(\gamma)} \hat{w}_t\hat{w}_t^\top \hat{e}_t^2\Big)(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1}, \quad i = 1, 2.$$

Lemma 3 implies that

$$(T^{-1}\hat{W}_i^{\gamma\top}\hat{W}_i^\gamma)^{-1} \xrightarrow{p} C_{A,i}^{-1}(\gamma) \tag{B.34}$$

uniformly in γ and thus, the behavior of $T^{-1}\sum_{\mathcal{T}_i(\gamma)} \hat{w}_t\hat{w}_t^\top \hat{e}_t^2$ is left to show.
To do so, as in the proof of Lemma 3, we consider 3 cases to facilitate exposition: 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 $\gamma = \rho^0 + \mathcal{O}(1)$ 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 $\gamma = \rho^0 + \mathcal{O}(1)$ by construction. This implies two sub-cases: $(c.1)$ with $\hat{\rho} \leq \rho^0 < \gamma$ and $(c.2)$ with $\rho^0 < \hat{\rho} < \gamma$.

Moreover, we only provide derivations for $T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t \leq \gamma\}}$ for the sake of brevity. The asymptotic behavior of $T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t > \gamma\}}$ follows similar arguments and is therefore omitted.

Case (a). Because[19] $\gamma < \hat{\rho}$ in both sub-cases and thus, $\hat{w}_t = \hat{A}_1 x_t$ it follows that

$$\begin{aligned} T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} &= \hat{A}_1\Big(T^{-1}\sum_t x_t x_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}}\Big)\hat{A}_1^\top \\ &= \hat{A}_1\Big(T^{-1}\sum_t x_t x_t^\top [y_t - \hat{w}_t^\top \hat{\theta}_1^\gamma]^2 \mathbb{1}_{\{q_t\le\gamma\}}\Big)\hat{A}_1^\top \\ &= \hat{A}_1\Big(T^{-1}\sum_t x_t x_t^\top [\hat{w}_t^\top(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_t]^2 \mathbb{1}_{\{q_t\le\gamma\}}\Big)\hat{A}_1^\top \end{aligned} \tag{B.35}$$

with $\tilde{\epsilon}_t = \epsilon_t + (z_t - \hat{z}_z)^\top \theta_z^0$. Let $B = T^{-1}\sum_t x_t x_t^\top [\hat{w}_t^\top(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_t]^2 \mathbb{1}_{\{q_t\le\gamma\}}$. Then

$$\begin{aligned} B = &\underbrace{T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 \mathbb{1}_{\{q_t\le\gamma\}}}_{=B_1} + 2\underbrace{T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t x_t^\top \hat{A}_1^\top(\theta^0 - \hat{\theta}_1^\gamma)\mathbb{1}_{\{q_t\le\gamma\}}}_{=B_2} \\ &+ \underbrace{T^{-1}\sum_t x_t x_t^\top(\theta^0 - \hat{\theta}_1^\gamma)\hat{A}_1 x_t x_t^\top \hat{A}_1^\top(\theta^0 - \hat{\theta}_1^\gamma)\mathbb{1}_{\{q_t\le\gamma\}}}_{=B_3}. \end{aligned} \tag{B.36}$$

Next, it can easily be shown that the terms B_2 and B_3 are both $o_p(1)$ uniformly in γ (cf. equation (A.32) and the subsequent arguments in the LRF case). Together with $\hat{A}_1 = A_1^0 + o_p(1) = \mathcal{O}_p(1)$ (as a consequence of Lemma 2) it follows that

$$T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} = A_1^0 B_1 A_1^{0\top} + o_p(1). \tag{B.37}$$

So, we are left to derive the asymptotic behavior of B_1 for this case. First note that

$$\begin{aligned} \tilde{\epsilon}_t \mathbb{1}_{\{q_t\le\gamma\}} &= [\epsilon_t + (z_t - \hat{z}_t)^\top \theta_z^0]\mathbb{1}_{\{q_t\le\gamma\}} \\ &= [\epsilon_t + (\Pi_1^{0\top} x_t \mathbb{1}_{\{q_t\le\rho^0\}} + \Pi_2^{0\top} x_t \mathbb{1}_{\{q_t>\rho^0\}} + u_t - \hat{\Pi}_1^\top x_t \mathbb{1}_{\{q_t\le\hat{\rho}\}} - \hat{\Pi}_2^\top x_t \mathbb{1}_{\{q_t>\hat{\rho}\}})]\mathbb{1}_{\{q_t\le\gamma\}} \\ &= [s_t + x_t^\top(\Pi_1^0 - \hat{\Pi}_1)\theta_z^0]\mathbb{1}_{\{q_t\le\gamma\}} \end{aligned} \tag{B.38}$$

because $\gamma \le \{\hat{\rho}\wedge\rho^0\}$ implies $\mathbb{1}_{\{q_t\le\gamma\}}\mathbb{1}_{\{q_t\le\hat{\rho}\}} = \mathbb{1}_{\{q_t\le\gamma\}}\mathbb{1}_{\{q_t\le\rho^0\}} = \mathbb{1}_{\{q_t\le\gamma\}}$ and $\mathbb{1}_{\{q_t\le\gamma\}}\mathbb{1}_{\{q_t>\hat{\rho}\}} = \mathbb{1}_{\{q_t\le\gamma\}}\mathbb{1}_{\{q_t>\rho^0\}} = 0$. Further, $s_t = \epsilon_t + u_t^\top \theta_z^0$. Thus, B_1 reads as

$$\begin{aligned} B_1 = &T^{-1}\sum_t x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\gamma\}} + 2T^{-1}\sum_t x_t x_t^\top s_t x_t^\top(\Pi_1^0 - \hat{\Pi}_1)\theta_z^0 \mathbb{1}_{\{q_t\le\gamma\}} \\ &+ T^{-1}\sum_t x_t x_t^\top \theta_z^{0\top}(\Pi_1^0 - \hat{\Pi}_1)^\top x_t x_t^\top(\Pi_1^0 - \hat{\Pi}_1)\theta_z^0 \mathbb{1}_{\{q_t\le\gamma\}}. \end{aligned} \tag{B.39}$$

It can easily be shown that the last two terms on the right-hand side of this expression are of order $o_p(1)$ (cf. Equation (A.35) in the proof of Theorem 1 and the subsequent arguments), uniformly in γ. Hence,

$$B_1 = T^{-1}\sum_t x_t x_t s_t^2 \mathbb{1}_{\{q_t\le\gamma\}} + o_p(1). \tag{B.40}$$

[19] Implying that the structural form partition $\mathcal{T}_1(\gamma) = \{t : \mathbb{1}_{\{q_t\le\gamma\}} = 1\}$ is a subset of the reduced form partition $\mathcal{T}_1(\hat{\rho}) = \{t : \mathbb{1}_{\{q_t\le\hat{\rho}\}} = 1\}$.

Hence, (B.36) reads as

$$T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} = A_1^0\Big(T^{-1}\sum_t x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\gamma\}}\Big)A_1^{0\top} + o_p(1) \tag{B.41}$$

and thus, uniformly in γ,

$$T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} \xrightarrow{p} A_1^0\mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\gamma\}}]A_1^{0\top} \tag{B.42}$$

because the data has bounded fourth moments (at least) by Assumption A.1.2.

Case (b). Here, sub-case $(b.1)$ with $\gamma = \rho^0 \le \hat{\rho}$ follows the same arguments as the derivations of case (a). Hence, in this sub-case it also holds uniformly in γ that

$$T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} \xrightarrow{p} A_1^0\mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\gamma\}}]A_1^{0\top}. \tag{B.43}$$

In sub-case $(b.2)$ with $\hat{\rho} < \gamma = \rho^0$ however, things look rather different. This is due the fact the structural form partition $\mathcal{T}_1(\gamma) = \{t : \mathbb{1}_{\{q_t\le\gamma\}} = 1\}$ is no more a subset of the reduced form partition $\mathcal{T}_1(\hat{\rho}) = \{t : \mathbb{1}_{\{q_t\le\hat{\rho}\}} = 1\}$. Thus, we decompose the partition of the sum based on $\mathbb{1}_{\{q_t\le\gamma\}}$ by a partition based on $\mathbb{1}_{\{q_t\le\hat{\rho}\}}$ and one based on $\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}$, retaining the original partition:

$$T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} = T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\hat{\rho}\}} + T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}). \tag{B.44}$$

From this it is obvious that $\hat{w}_t\mathbb{1}_{\{q_t\le\hat{\rho}\}} = \hat{A}_1 x_t \mathbb{1}_{\{q_t\le\hat{\rho}\}}$ and that $\hat{w}_t(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) = \hat{w}_t\mathbb{1}_{\{q_t\le\gamma\}}\mathbb{1}_{\{q_t>\hat{\rho}\}} = \hat{A}_2 x_t(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}})$. Thus, (B.44) can be stated as

$$\begin{aligned} T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} &= \hat{A}_1\underbrace{\Big(T^{-1}\sum_t x_t x_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\hat{\rho}\}}\Big)}_{=\Xi_1}\hat{A}_1 \\ &\quad + \hat{A}_2\underbrace{\Big(T^{-1}\sum_t x_t x_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}})\Big)}_{=\Xi_2}\hat{A}_2. \end{aligned} \tag{B.45}$$

Next, we show the asymptotic behavior of the two terms Ξ_1 and Ξ_2. Following the reasoning from case (a), we find for Ξ_1 that, uniformly in γ,

$$\begin{aligned} \Xi_1 &= T^{-1}\sum_t x_t x_t^\top [\hat{w}_t^\top(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_t]^2 \mathbb{1}_{\{q_t\le\hat{\rho}\}} \\ &= T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 \mathbb{1}_{\{q_t\le\hat{\rho}\}} + 2T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t x_t^\top \hat{A}_1^\top(\theta^0 - \hat{\theta}_1^\gamma)\mathbb{1}_{\{q_t\le\gamma\}} \\ &\quad + T^{-1}\sum_t x_t x_t^\top (\theta^0 - \hat{\theta}_1^\gamma)^\top \hat{A}_1 x_t x_t^\top \hat{A}_1^\top(\theta^0 - \hat{\theta}_1^\gamma)\mathbb{1}_{\{q_t\le\hat{\rho}\}} \end{aligned}$$

$$
\begin{aligned}
&= T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 \mathbb{1}_{\{q_t \le \hat{\rho}\}} + o_p(1) \\
&= T^{-1}\sum_t x_t x_t s_t^2 \mathbb{1}_{\{q_t \le \hat{\rho}\}} + 2T^{-1}\sum_t x_t x_t^\top s_t x_t^\top (\Pi_1^0 - \hat{\Pi}_1)\theta_z^0 \mathbb{1}_{\{q_t \le \hat{\rho}\}} \\
&+ T^{-1}\sum_t x_t x_t^\top \theta_z^{0\top}(\Pi_1^0 - \hat{\Pi}_1)^\top x_t x_t^\top (\Pi_1^0 - \hat{\Pi}_1)\theta_z^0 \mathbb{1}_{\{q_t \le \hat{\rho}\}} + o_p(1) \\
&= T^{-1}\sum_t x_t x_t s_t^2 \mathbb{1}_{\{q_t \le \hat{\rho}\}} + o_p(1).
\end{aligned} \tag{B.46}
$$

Together with the facts that $\hat{A}_1 = A_1^0 + o_p(1) = \mathcal{O}_p(1)$, that the data has bounded fourth moments by Assumption A.1.2 and that $\hat{\rho} \xrightarrow{p} \rho^0$ (cf. Lemma 2) it thus follows that

$$
\hat{A}_1 \Xi_1 \hat{A}_1^\top \xrightarrow{p} A_1^0 \mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t \le \rho^0\}}] A_1^{0\top}. \tag{B.47}
$$

For the term Ξ_2, again by similar reasoning as in case (a) and uniformly in γ, we find that

$$
\begin{aligned}
\Xi_2 &= T^{-1}\sum_t x_t x_t^\top [\hat{w}_t(\theta^0 - \hat{\theta}_1^\gamma) + \tilde{\epsilon}_t]^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&= T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&+ 2T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t x_t^\top \hat{A}_2^\top (\theta^0 - \hat{\theta}_1^\gamma)(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&+ T^{-1}\sum_t x_t x_t^\top (\theta^0 - \hat{\theta}_1^\gamma)^\top \hat{A}_2 x_t x_t^\top \hat{A}_2^\top (\theta^0 - \hat{\theta}_1^\gamma)(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&= T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) + o_p(1) \\
&= T^{-1}\sum_t x_t x_t^\top [s_t + x_t^\top (\Pi_1^0 - \hat{\Pi}_2)\theta_z^0]^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) + o_p(1) \\
&= T^{-1}\sum_t x_t x_t^\top s_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&+ 2T^{-1}\sum_t x_t x_t^\top s_t x_t^\top (\Pi_1^0 - \hat{\Pi}_2)\theta_z^0 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&+ T^{-1}\sum_t x_t x_t^\top \theta_z^{0\top}(\Pi_1^0 - \hat{\Pi}_2)^\top x_t x_t^\top (\Pi_1^0 - \hat{\Pi}_2)\theta_z^0 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) + o_p(1)
\end{aligned} \tag{B.48}
$$

where the second equality used the fact that $\hat{w}_t(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) = \hat{w}_t \mathbb{1}_{\{q_t \le \gamma\}} \mathbb{1}_{\{q_t > \hat{\rho}\}} = \hat{A}_2 x_t \mathbb{1}_{\{q_t \le \gamma\}} \mathbb{1}_{\{q_t > \hat{\rho}\}} = \hat{A}_2 x_t (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}})$. The third equality holds by the same reasoning as in the LRF case (cf. Equation (A.32) and the subsequent arguments). The fourth and fifth equalites utilize the definition of $\tilde{\epsilon}_t = \epsilon_t + (z_t - \hat{z}_t)^\top \theta_z^0$ and the facts that $z_t(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) = z_t \mathbb{1}_{\{q_t \le \gamma\}} \mathbb{1}_{\{q_t > \hat{\rho}\}} = (\Pi_1^{0\top} x_t + u_t)(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}})$ and

$\hat{z}_t(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) = z_t \mathbb{1}_{\{q_t \le \gamma\}} \mathbb{1}_{\{q_t > \hat{\rho}\}} = \hat{\Pi}_2^\top x_t (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}})$.
Next, it holds that $\hat{\rho} = \rho^0 + \mathcal{O}_p(T^{-1})$ by Lemma 2 and thus that, uniformly in γ,

$$\Xi_2 = \mathcal{O}_p(T^{-1}) \tag{B.49}$$

implying that

$$\hat{A}_2 \Xi_2 \hat{A}_2^\top \xrightarrow{p} 0 \tag{B.50}$$

because $\hat{A}_2 = A_2^0 + o_p(1) = \mathcal{O}_p(1)$ by Lemma 2. Thus, combining (B.47) and (B.50) yields that, uniformly in γ,

$$T^{-1} \sum_t \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t \le \gamma\}} \xrightarrow{p} A_1^0 \mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t \le \gamma\}}] A_1^{0\top}. \tag{B.51}$$

Case (c). For both sub-cases of case c it holds that the structural form partition based on $\mathbb{1}_{\{q_t \le \gamma\}}$ is not a further restriction on the (estimated) reduced form partition based on $\mathbb{1}_{\{q_t \le \hat{\rho}\}}$, as in sub-case $(b.2)$. To circumvent this problem we will use a decomposition of the involved summation. That is, in sub-case $(c.1)$ we will first sum over all observations satisfying $\mathbb{1}_{\{q_t \le \hat{\rho}\}} = 1$, then all observations satisfying $\mathbb{1}_{\{q_t \le \rho^0\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}} = 1$ and finally all observations with $\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}} = 1$. Similarly, in sub-case $(c.2)$ the decomposition is based on $\mathbb{1}_{\{q_t \le \rho^0\}} = 1$, $\mathbb{1}_{\{q_t \le \hat{\rho}\}} - \mathbb{1}_{\{q_t \le \rho^0\}} = 1$ and $\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}} = 1$.

Therefore, in sub-case $(c.1)$ it holds that

$$\begin{aligned}
T^{-1} \sum_t \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t \le \gamma\}} &= T^{-1} \sum_t \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t \le \hat{\rho}\}} \\
&\quad + T^{-1} \sum_t \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t \le \rho^0\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \\
&\quad + T^{-1} \sum_t \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \\
&= \hat{A}_1 \underbrace{\Big(T^{-1} \sum_t x_t x_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t \le \hat{\rho}\}} \Big)}_{=D_1} \hat{A}_1^\top \\
&\quad + \hat{A}_2 \underbrace{\Big(T^{-1} \sum_t x_t x_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t \le \rho^0\}} - \mathbb{1}_{\{q_t \le \hat{\rho}\}}) \Big)}_{=D_2} \hat{A}_2^\top \\
&\quad + \hat{A}_2 \underbrace{\Big(T^{-1} \sum_t x_t x_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \Big)}_{=D_3} \hat{A}_2^\top
\end{aligned} \tag{B.52}$$

As already shown (cf. Equation (B.47)) it holds uniformly in γ that

$$\hat{A}_1 D_1 \hat{A}_1^\top \xrightarrow{p} A_1^0 \mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t \le \rho^0\}}] A_1^{0\top}. \tag{B.53}$$

Moreover, as argued for (B.49) it also holds that

$$\hat{A}_2 D_2 \hat{A}_2^\top = o_p(1) \tag{B.54}$$

because $\hat{\rho} = \rho^0 + \mathcal{O}_p(T^{-1})$ by Lemma 2 implying $D_2 = \mathcal{O}_p(T^{-1})$ and $\hat{A}_2 = A_2^0 + o_p(1)$ by Lemma 2.
For the remaining quantity D_3 in (B.52) we find that

$$\begin{aligned} D_3 &= T^{-1} \sum_t x_t x_t^\top \tilde{\epsilon}_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \\ &\quad + 2T^{-1} \sum_t x_t x_t^\top \tilde{\epsilon}_t x_t^\top \hat{A}_2^\top (\theta^0 - \hat{\theta}_1^\gamma)(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \\ &\quad + T^{-1} \sum_t x_t x_t^\top (\theta^0 - \hat{\theta}_1^\gamma)^\top \hat{A}_2 x_t x_t^\top \hat{A}_2^\top (\theta^0 - \hat{\theta}_1^\gamma)(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \end{aligned} \tag{B.55}$$

where the last two quantities on the right hand side are, as already shown, of order $o_p(1)$ because $\theta^0 - \hat{\theta}_1^\gamma = o_p(1)$ by Lemma 3 and the (scaled) sums are of order $\mathcal{O}_p(1)$. Thus,

$$D_3 = T^{-1} \sum_t x_t x_t^\top \tilde{\epsilon}_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) + o_p(1). \tag{B.56}$$

Last, because $\tilde{\epsilon}_t(\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) = [s_t + x_t^\top (\Pi_2^0 - \hat{\Pi}_2)\theta_z^0](\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}})$ it follows, as before, that

$$\begin{aligned} D_3 &= T^{-1} \sum_t x_t x_t^\top s_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \\ &\quad + 2T^{-1} \sum_t x_t x_t^\top s_t x_t^\top (\Pi_2^0 - \hat{\Pi}_2)\theta_z^0 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \\ &\quad + T^{-1} \sum_t x_t x_t^\top \theta_z^{0\top} (\Pi_2^0 - \hat{\Pi}_2)^\top x_t x_t^\top (\Pi_2^0 - \hat{\Pi}_2)\theta_z^0 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) \\ &= T^{-1} \sum_t x_t x_t^\top s_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}}) + o_p(1), \end{aligned} \tag{B.57}$$

uniformly in γ.
(B.57) together with (B.52)–(B.54) and that $\hat{A}_2 = A_2^0 + o_p(1)$ imply that

$$\begin{aligned} T^{-1} \sum_t \hat{w}_t \hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t \le \gamma\}} &\xrightarrow{p} A_1^0 \mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t \le \rho^0\}}] A_1^{0\top} \\ &\quad + A_2^0 \mathbb{E}[x_t x_t^\top s_t^2 (\mathbb{1}_{\{q_t \le \gamma\}} - \mathbb{1}_{\{q_t \le \rho^0\}})] A_2^{0\top}. \end{aligned} \tag{B.58}$$

In sub-case $(c.2)$ we have that

$$
\begin{aligned}
T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\gamma\}} &= T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\rho^0\}} \\
&+ T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t\le\hat{\rho}\}} - \mathbb{1}_{\{q_t\le\rho^0\}}) \\
&+ T^{-1}\sum_t \hat{w}_t\hat{w}_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) \\
&= \hat{A}_1 \underbrace{\Big(T^{-1}\sum_t x_t x_t^\top \hat{e}_t^2 \mathbb{1}_{\{q_t\le\rho^0\}}\Big)}_{=E_1} \hat{A}_1^\top \\
&+ \hat{A}_1 \underbrace{\Big(T^{-1}\sum_t x_t x_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t\le\hat{\rho}\}} - \mathbb{1}_{\{q_t\le\rho^0\}})\Big)}_{=E_2} \hat{A}_1^\top \\
&+ \hat{A}_2 \underbrace{\Big(T^{-1}\sum_t x_t x_t^\top \hat{e}_t^2 (\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}})\Big)}_{=E_3} \hat{A}_2^\top . \qquad \text{(B.59)}
\end{aligned}
$$

Again, as already shown it holds uniformly in γ that

$$
\hat{A}_1 E_1 \hat{A}_1^\top \xrightarrow{p} A_1^0 \mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\rho^0\}}] A_1^{0\top} \qquad \text{(B.60)}
$$

and that

$$
\hat{A}_1 E_2 \hat{A}_1^\top = o_p(1). \qquad \text{(B.61)}
$$

For the remaining quantity E_3 in (B.57) it holds uniformly in γ that

$$
\begin{aligned}
E_3 &= T^{-2}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 (\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) \\
&+ 2T^{-1}\sum_t x_t x_t^\top \tilde{\epsilon}_t x_t^\top \hat{A}_2^\top (\theta^0 - \hat{\theta}_1^\gamma)(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) \\
&+ T^{-1}\sum_t x_t x_t^\top (\theta^0 - \hat{\theta}_1^\gamma)^\top \hat{A}_2 x_t x_t^\top \hat{A}_2^\top (\theta^0 - \hat{\theta}_1^\gamma)(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) \\
&= T^{-2}\sum_t x_t x_t^\top \tilde{\epsilon}_t^2 (\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) + o_p(1) \qquad \text{(B.62)}
\end{aligned}
$$

as before.
Next, because $\tilde{\epsilon}_t(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}}) = [s_t + x_t^\top(\Pi_2^0 - \hat{\Pi}_2)\theta_z^0](\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat{\rho}\}})$ it follows

that

$$
\begin{aligned}
E_3 &= T^{-1}\sum_t x_t x_t^\top s_t^2(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat\rho\}}) \\
&+ 2T^{-1}\sum_t x_t x_t^\top s_t x_t^\top(\Pi_2^0 - \hat\Pi_2)\theta_z^0(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat\rho\}}) \\
&+ 2T^{-1}\sum_t x_t x_t^\top \theta_z^{0\top}(\Pi_2^0 - \hat\Pi_2)^\top x_t x_t^\top(\Pi_2^0 - \hat\Pi_2)\theta_z^0(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat\rho\}}) + o_p(1) \\
&= T^{-1}\sum_t x_t x_t^\top s_t^2(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\hat\rho\}}) + o_p(1)
\end{aligned}
\tag{B.63}
$$

uniformly in γ.
(B.63) together with (B.57)–(B.61), (B.63) and that $\hat A_i = A_i^0 + o_p(1)$ for $i = 1, 2$ thus implies that

$$
\begin{aligned}
T^{-1}\sum_t \hat w_t \hat w_t^\top \hat e_t^2 \mathbb{1}_{\{q_t\le\gamma\}} &\xrightarrow{p} A_1^0\mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\rho^0\}}]A_1^{0\top} \\
&+ A_2^0\mathbb{E}[x_t x_t^\top s_t^2(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\rho^0\}})]A_2^{0\top},
\end{aligned}
\tag{B.64}
$$

uniformly in γ.
Thus, combining results (B.42), (B.43), (B.51), (B.58) and (B.64) yield

$$
\begin{aligned}
T^{-1}\sum_t \hat w_t \hat w_t^\top \hat e_t^2 \mathbb{1}_{\{q_t\le\gamma\}} &\xrightarrow{p} A_1^0\mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\le\gamma\wedge\rho^0\}}]A_1^{0\top} \\
&+ A_2^0\mathbb{E}[x_t x_t^\top s_t^2(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{\gamma\wedge\rho^0\}})]A_2^{0\top} \\
&\equiv H_{A,1}(\gamma)
\end{aligned}
\tag{B.65}
$$

uniformly in $\gamma \in \Gamma$.
With similar reasoning we obtain

$$
\begin{aligned}
T^{-1}\sum_t \hat w_t \hat w_t^\top \hat e_t^2 \mathbb{1}_{\{q_t>\gamma\}} &\xrightarrow{p} A_2^0\mathbb{E}[x_t x_t^\top s_t^2 \mathbb{1}_{\{q_t\ge\gamma\vee\rho^0\}}]A_2^{0\top} \\
&+ A_1^0\mathbb{E}[x_t x_t^\top s_t^2(\mathbb{1}_{\{q_t>\gamma\}} - \mathbb{1}_{\{q_t\ge\gamma\vee\rho^0\}})]A_1^{0\top} \\
&\equiv H_{A,2}(\gamma)
\end{aligned}
\tag{B.66}
$$

uniformly in $\gamma \in \Gamma$. In fact these derivations follows the same lines as for $T^{-1}\sum_t \hat w_t \hat w_t^\top \hat e_t^2 \mathbb{1}_{\{q_t\le\gamma\}}$ with the cases (a) (b) and (c) and their respective sub-cases inverting their roles.
Those two equation together with (B.34), the continuous mapping theorem and weak convergence uniformly in γ thus yields

$$
\sup_{\gamma\in\Gamma} W_{T,\mathrm{TRF}}^{2\mathrm{SLS}}(\gamma) \Longrightarrow \mathcal{E}_A^\top(\gamma)[C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma) + C_{A,2}^{-1}(\gamma)H_{A,2}(\gamma)C_{A,2}^{-1}(\gamma)]^{-1}\mathcal{E}_A(\gamma)
\tag{B.67}
$$

proving the claim. □

Proof of Corollary 3. Again, we will prove th claim in two steps. First, we show the desired result for the sup LR-test and afterwards, for the sup Wald-test.

sup LR-test. As in the case of a linear reduced form, the only quantities affected by Assumption A.2 are $\mathcal{GP}_{\text{mat},1}(\gamma)$ (and therefore, $\mathcal{B}_{A,1}(\gamma)$, $\mathcal{B}_{A,2}(\gamma)$ and $\mathcal{E}(\gamma)$), $H_{A,1}(\gamma)$ and $H_{A,2}(\gamma)$. Moreover, by the same arguments as in the proof of Corollary 1 it holds that $\mathcal{GP}_{\text{mat},1}(\gamma) = \tilde{\mathcal{GP}}_{\text{mat},1}(\gamma)Q$. Therefore, it immediately follows that $\mathcal{B}_{A,1}(\gamma) = \tilde{\mathcal{B}}_{A,1}(\gamma)$ and hence, that $\mathcal{E}(\gamma) = \tilde{\mathcal{E}}(\gamma)$. Thus, the claim immediately follows.

sup Wald-test. As argued above, $\mathcal{E}(\gamma) = \tilde{\mathcal{E}}(\gamma)$. Hence, as in the linear reduced form case, we are left to show that $[C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma)+C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma)]^{-1} = C_{A,2}(\gamma)C^{-1}C_{A,1}(\gamma)/\sigma^2$. We will do this in the same manner as in the proof of Corollary 1. I.e. we will show that $H_{A,i}(\gamma) = \sigma^2 C_{A,i}(\gamma)$, $i = 1, 2$. The claim then follows as in the proof of Corollary 1 by showing that $[C_{A,1}^{-1}(\gamma) + C_{A,2}^{-1}(\gamma)]^{-1} = C_{A,2}(\gamma)C_A^{-1}C_{A,1}(\gamma)$ via the Kailath identity.
Under Assumption A.2 it follows for $H_{A,1}(\gamma)$ that

$$\begin{aligned}
H_{A,1}(\gamma) &= A_1^0\mathbb{E}[x_t x_t^\top(\epsilon_t + u_t^\top\theta_z^0)^2\mathbb{1}_{\{q_t\le\gamma\wedge\rho^0\}}]A_1^{0\top}\\
&= A_2^0\mathbb{E}[x_t x_t^\top(\epsilon_t + u_t^\top\theta_z^0)^2(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\gamma\wedge\rho^0\}})]A_2^{0\top}\\
&= A_1^0\mathbb{E}\left[x_t x_t^\top\mathbb{1}_{\{q_t\le\gamma\wedge\rho^0\}}\mathbb{E}[(\epsilon_t + u_t^\top\theta_z^0)^2|x_t, q_t]\right]A_1^{0\top}\\
&= A_2^0\mathbb{E}\left[x_t x_t^\top(\mathbb{1}_{\{q_t\le\gamma\}} - \mathbb{1}_{\{q_t\le\gamma\wedge\rho^0\}})\mathbb{E}[(\epsilon_t + u_t^\top\theta_z^0)^2|x_t, q_t]\right]A_2^{0\top}\\
&= \sigma^2[A_1^0M_1(\gamma\wedge\rho^0)A_1^{0\top} + A_2^0[M_1(\gamma) - M_1(\gamma\wedge\rho^0)]A_2^{0\top}]\\
&= \sigma^2 C_{A,1}(\gamma),
\end{aligned} \tag{B.68}$$

where the last equality holds by definition of $C_{A,1}(\gamma)$ (cf. equation (4.5)).
Similarly, for $H_{A,2}(\gamma)$ we have that

$$\begin{aligned}
H_{A,2}(\gamma) &= A_1^0\mathbb{E}[x_t x_t^\top(\epsilon_t + u_t^\top\theta_z^0)^2(\mathbb{1}_{\{q_t>\gamma\}} - \mathbb{1}_{\{q_t>\gamma\vee\rho^0\}})]A_1^{0\top}\\
&+ A_2^0\mathbb{E}[x_t x_t^\top(\epsilon_t^2 + u_t^\top\theta_z^0)^2\mathbb{1}_{\{q_t>\gamma\vee\rho^0\}}]A_2^{0\top}\\
&= \sigma^2[A_1^0[M_2(\gamma) - M_2(\gamma\vee\rho^0)]A_1^{0\top} + A_2^0M_2(\gamma\vee\rho^0)A_2^{0\top}].
\end{aligned} \tag{B.69}$$

Next, note that

$$\begin{aligned}
C_{A,2}(\gamma) &= C_A - C_{A,1}(\gamma)\\
&= A_1^0M_1(\rho^0)A_1^{0\top} + A_2^0M_2(\rho^0)A_2^{0\top}\\
&- A_1^0M_1(\gamma\wedge\rho^0)A_1^{0\top} - A_2^0M_1(\gamma)A_2^{0\top} + A_2^0M_1(\gamma\wedge\rho^0)A_2^{0\top}\\
&= A_1^0[M_1(\rho^0) - M_1(\gamma\wedge\rho^0)]A_1^{0\top}\\
&+ A_2^0[M_2(\rho^0) - M_1(\gamma) + M_1(\gamma\wedge\rho^0)]A_2^{0\top}.
\end{aligned} \tag{B.70}$$

So, in a last step of proving that $H_{A,2}(\gamma) = \sigma^2 C_{A,2}(\gamma)$ we need to show that

$$\begin{aligned}
&A_1^0[M_2(\gamma) - M_2(\gamma\vee\rho^0)]A_1^{0\top} + A_2^0M_2(\gamma\vee\rho^0)A_2^{0\top}\\
=&A_1^0[M_1(\rho^0) - M_1(\gamma\wedge\rho^0)]A_1^{0\top} + A_2^0[M_2(\rho^0) - M_1(\gamma) + M_1(\gamma\wedge\rho^0)]A_2^{0\top}.
\end{aligned} \tag{B.71}$$

Therefore, note that

$$M_1(\rho^0) = M - M_2(\rho^0) \tag{B.72a}$$
$$M_1(\gamma \wedge \rho^0) = M - M_2(\gamma) - M_2(\rho^0) + M_2(\gamma \vee \rho^0) \tag{B.72b}$$
$$M_1(\gamma) = M - M_2(\gamma). \tag{B.72c}$$

Finally, plugging (B.72a)–(B.72c) into (B.71) yields the desired result that $H_{A,2}(\gamma) = \sigma^2 C_{A,2}(\gamma)$ and together with (B.68) that

$$[C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma) + C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma)]^{-1} = \frac{[C_{A,1}^{-1}(\gamma) + C_{A,2}^{-1}(\gamma)]^{-1}}{\sigma^2}.$$

Finally, applying the Kailath identity to the numerator of this last expression, as in the linear reduced form case, yields the claim that

$$\begin{aligned}&\mathcal{E}_A^\top(\gamma)[C_{A,1}^{-1}(\gamma)H_{A,1}(\gamma)C_{A,1}^{-1}(\gamma) + C_{A,2}^{-1}(\gamma)H_{A,2}(\gamma)C_{A,2}^{-1}(\gamma)]^{-1}\mathcal{E}_A(\gamma)\\ =&\frac{\tilde{\mathcal{E}}_A^\top(\gamma)C_{A,2}(\gamma)C_A^{-1}C_{A,1}(\gamma)\tilde{\mathcal{E}}_A(\gamma)}{\sigma^2}\end{aligned}$$

□

C. GMM Results

Proof of Corollary 4. First we note that by Assumption A.1.4 it follows that $\text{Prob}(q_t \leq \gamma) = G(\gamma)$ is continuous. Further, we will replace the threshold parameter γ by an equivalent value, say λ, defined on the open unit interval $(0; 1)$. To see how this works, note first that $\Gamma \subset (\gamma_{\min}, \gamma_{\max})$. Then, $\text{Prob}(q_t \leq \gamma_{\min}) = 0$ and $\text{Prob}(q_t \leq \gamma_{\max}) = 1$. 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

$$\Omega_1(\gamma) = \lambda\Omega, \quad \Omega_2(\gamma) = (1-\lambda)\Omega \tag{C.1}$$
$$N_1(\gamma) = \lambda N, \quad N_2(\gamma) = (1-\lambda)N \tag{C.2}$$
$$V_1(\gamma) = \lambda^{-1}\left[N\Omega^{-1}N^\top\right]^{-1} = \lambda^{-1}V \tag{C.3}$$
$$V_2(\gamma) = (1-\lambda)^{-1}\left[N\Omega^{-1}N^\top\right]^{-1} = (1-\lambda)^{-1}V \tag{C.4}$$
$$V_1(\gamma) + V_2(\gamma) = \frac{V}{\lambda(1-\lambda)} \tag{C.5}$$
$$V_1(\gamma)N_1(\gamma)\Omega_1^{-1}(\gamma) = \lambda^{-1}VN\Omega^{-1} \tag{C.6}$$
$$V_2(\gamma)N_2(\gamma)\Omega_2^{-1}(\gamma) = (1-\lambda)^{-1}VN\Omega^{-1} \tag{C.7}$$

Moreover, (C.1) implies that –under Assumptions A.2 and A.3– the Gaussian process $\mathcal{GP}_1(\gamma)$ can be restated as

$$\mathcal{GP}_1(\gamma) = \Omega^{1/2}\overline{\mathcal{BM}}(\lambda) \tag{C.8}$$
$$\mathcal{GP} = \Omega^{1/2}\overline{\mathcal{BM}}(1) \tag{C.9}$$

where $\overline{\mathcal{BM}}(\cdot)$ is a $q \times 1$-vector of independent Brownian motions on the unit interval. Thus, the term $V_1(\gamma)N_1(\gamma)\Omega_1^{-1}(\gamma)\mathcal{GP}_1(\gamma) - V_2(\gamma)N_2(\gamma)\Omega_2^{-1}(\gamma)\mathcal{GP}_2(\gamma))$ can be restated in terms of λ as

$$\begin{aligned}
&V_1(\gamma)N_1(\gamma)\Omega_1^{-1}(\gamma)\mathcal{GP}_1(\gamma) - V_2(\gamma)N_2(\gamma)\Omega_2^{-1}(\gamma)\mathcal{GP}_2(\gamma) \\
&= \lambda^{-1}VN\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) - (1-\lambda)^{-1}VN\Omega^{-1/2}(\overline{\mathcal{BM}}(1) - \overline{\mathcal{BM}}(\lambda)) \\
&= \frac{VN\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) - \lambda VN\Omega^{-1/2}\overline{\mathcal{BM}}(1)}{\lambda(1-\lambda)}. \qquad \text{(C.10)}
\end{aligned}$$

Hence, using results (C.5) and (C.10) the asymptotic distribution simplifies, again using similar arguments as in the proof of Corollary 2, to

$$\begin{aligned}
&\frac{\left[VN\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) - \lambda VN\Omega^{-1/2}\overline{\mathcal{BM}}(1)\right]^\top}{\lambda(1-\lambda)} \\
&\times V^{-1}\lambda(1-\lambda) \\
&\times \frac{\left[VN\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) - \lambda VN\Omega^{-1/2}\overline{\mathcal{BM}}(1)\right]}{\lambda(1-\lambda)} \\
&= \left[V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) - \lambda V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(1)\right]^\top \\
&\times \frac{1}{\lambda(1-\lambda)} \\
&\times \left[V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) - \lambda V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(1)\right]. \qquad \text{(C.11)}
\end{aligned}$$

Now, for the expression $V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda)$ in (C.11) we find

$$\begin{aligned}
V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) &\sim \mathcal{N}\left(0_p, V^{1/2}N\Omega^{-1}N^\top V^{1/2}\right) \\
&= \mathcal{N}(0_p, I_p) \qquad \text{(C.12)}
\end{aligned}$$

where the equality uses $V = (N\Omega^{-1}N^\top)^{-1}$. Thus, we have

$$V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(\lambda) \stackrel{\mathcal{D}}{=} \mathcal{BM}(\lambda) \qquad \text{(C.13)}$$

where $\mathcal{BM}(\lambda)$ is a $p \times 1$-vector of independent Brownian motions on the unit interval. Similarly,

$$V^{1/2}N\Omega^{-1/2}\overline{\mathcal{BM}}(1) \stackrel{\mathcal{D}}{=} \mathcal{BM}(1). \qquad \text{(C.14)}$$

Combining results (C.11)–(C.14) then immediately yields

$$\begin{aligned}
&\left[V_1(\gamma)N_1(\gamma)\Omega_1^{-1}(\gamma)\mathcal{GP}_1^U(\gamma) - V_2(\gamma)N_2(\gamma)\Omega_2^{-1}(\gamma)\mathcal{GP}_2^U(\gamma))\right]^\top \\
&\times \left[V_1(\gamma) + V_2(\gamma)\right]^{-1} \\
&\times \left[V_1(\gamma)N_1(\gamma)\Omega_1^{-1}(\gamma)\mathcal{GP}_1^U(\gamma) - V_2(\gamma)N_2(\gamma)\Omega_2^{-1}(\gamma)\mathcal{GP}_2^U(\gamma))\right] \\
&\stackrel{\mathcal{D}}{=} \frac{\left[\mathcal{BM}(\lambda) - \lambda\mathcal{BM}(1)\right]^\top\left[\mathcal{BM}(\lambda) - \lambda\mathcal{BM}(1)\right]}{\lambda(1-\lambda)}. \qquad \text{(C.15)}
\end{aligned}$$

The claim then follows by continuity of the process $[\mathcal{BM}(\lambda) - \lambda\mathcal{BM}(1)]$, the continuous mapping theorem and weak convergence (uniform in λ). □