

Tilburg University

Testing for a Threshold in Models with Endogenous Regressors

Rothfelder, Mario P.; Boldea, Otilia

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Rothfelder, M. P., & Boldea, O. (2022). *Testing for a Threshold in Models with Endogenous Regressors*. (Arxiv). Cornell University Library.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Testing for a Threshold in Models with Endogenous Regressors

Mario P. Rothfelder* and Otilia Boldea[†]

July 21, 2022

Abstract

We show by simulation that the test for an unknown threshold in models with endogenous regressors - proposed in Caner and Hansen (2004) - can exhibit severe size distortions both in small and in moderately large samples, pertinent to empirical applications. We propose three new tests that rectify these size distortions. The first test is based on GMM estimators. The other two are based on unconventional 2SLS estimators, that use additional information about the linearity (or lack of linearity) of the first stage. Just like the test in Caner and Hansen (2004), our tests are non-pivotal, and we prove their bootstrap validity. The empirical application revisits the question in Ramey and Zubairy (2018) whether government spending multipliers are larger in recessions, but using tests for an unknown threshold. Consistent with Ramey and Zubairy (2018), we do not find strong evidence that these multipliers are larger in recessions.

Keywords: 2SLS, GMM, instrumental variables, government spending, unknown threshold, wild bootstrap

*Amsterdam School of Economics, University of Amsterdam, and Tinbergen Institute. Address: University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. Email: m.p.rothfelder@uva.nl

[†]Corresponding Author. Department of Econometrics and Operations Research, Tilburg School of Economics and Management, Tilburg University. Address: Tilburg University, Warandelaan 2, 5037AB, Tilburg, The Netherlands. Email: o.boldea@tilburguniversity.edu

[‡]We would like to thank for very valuable comments and suggestions on this work: Jeffrey Campbell, Mehmet Caner, Pavel Čížek, Bruce Hansen, Frank Kleibergen, André Lucas, Bertrand Melenberg, David Pacini, Valerie Ramey, Nikolaus Schweizer, Bas Werker, and the participants at the Netherlands Econometrics Study Group Conference in Tilburg (2014) and Maastricht (2015), the Bristol Econometric Study Group (2015), the IAAE Conference in Thessaloniki (2015), the NBER-NSF Time Series Conference in Vienna (2015), the ESEM Conference in Geneva (2016) and the NASM Conference in Philadelphia (2016). Otilia Boldea gratefully acknowledges the support of the NWO VENI Grant 451-11-001.

1 Introduction

In the aftermath of the 2008 financial crisis, there has been a surge in the macroeconomic literature investigating whether the response of many key macroeconomic variables to monetary and fiscal policies depends on the state of the economy - see, among others, Auerbach and Gorodnichenko (2013), Owyang et al. (2013), Caggiano et al. (2015), Cugnasca and Rother (2015), Ramey and Zubairy (2018), Alloza (2022) and Jo and Zubairy (2022) for fiscal policy examples, and Santoro et al. (2014), Barnichon and Matthes (2018), Jordá et al. (2020), Alpanda et al. (2021), Bruns and Piffer (2021) and Klepacz (2021) for monetary policy examples. These papers model state dependence in various ways, including via threshold models, in which case the state dependence is typically driven by a particular variable such as the unemployment rate, interest rates, or credit conditions.

Threshold models were also widely used in economics to model unemployment, growth, bank profits, asset prices, exchange rates, and interest rates; see Hansen (2011) for a survey of economic applications. While threshold models with exogenous regressors have been widely studied and their asymptotic properties are well known¹, the literature on threshold models with endogenous regressors remains relatively scarce.² Nevertheless, in many applications, the regressors are *endogenous* and the existence of a threshold has important policy implications. For example, among the empirical papers cited above, Owyang et al. (2013), Cugnasca and Rother (2015), Ramey and Zubairy (2018) and Jo and Zubairy (2022) use a threshold model with endogenous regressors, where the state dependence of the macroeconomic response is driven by a threshold variable being above or below a certain a-priori fixed value. Ramey and Zubairy (2018) (RZ henceforth) used a threshold model with endogenous regressors to investigate whether the government spending multiplier is larger in recessions, where recessions were defined by the unemployment rate being below or above a threshold parameter. This has important policy implications, because if the government spending multiplier is larger (above one) in recessions, it implies that governments should spend more in recessions to boost the economy.

In their analysis, RZ fix this threshold parameter at an unemployment rate of 6.5%.³ As the threshold parameter is typically *unknown*, we revisit their question and test for an unknown threshold, using - to our knowledge - the only parametric test available for linear *time series* models with endogenous regressors that directly applies to the RZ model. This test was proposed in Caner and Hansen (2004) (CH henceforth). CH first compute a Wald test statistic for all candidate threshold values between the ϵ and $(1 - \epsilon)$ quantiles of the threshold variable, then take the maximum over this sequence to obtain a test for the null hypothesis of no threshold against the alternative hypothesis of an unknown threshold in (otherwise) linear models with

¹See *inter alia* Tong (1990), Hansen (1996, 1999, 2000) and Gonzalo and Wolf (2005) for inference, Gonzalo and Pitarakis (2002) for multiple threshold regression and model selection, Caner and Hansen (2001) and Gonzalo and Pitarakis (2006) for threshold regression with unit roots, Seo and Linton (2007) for smoothed estimators of threshold models, Lee et al. (2011) for testing for thresholds, and Hansen (2017) for threshold regressions with a kink.

²For some contributions with endogenous regressors, see *inter alia*: for time-series, Caner and Hansen (2004), who consider exogenous threshold variables and Kourtellis et al. (2015) who consider endogenous threshold variables; for cross-sections and (short) panels, Seo and Shin (2016) (and references therein), Yu and Phillips (2018) and Christopoulos et al. (2021), who consider endogenous threshold variables.

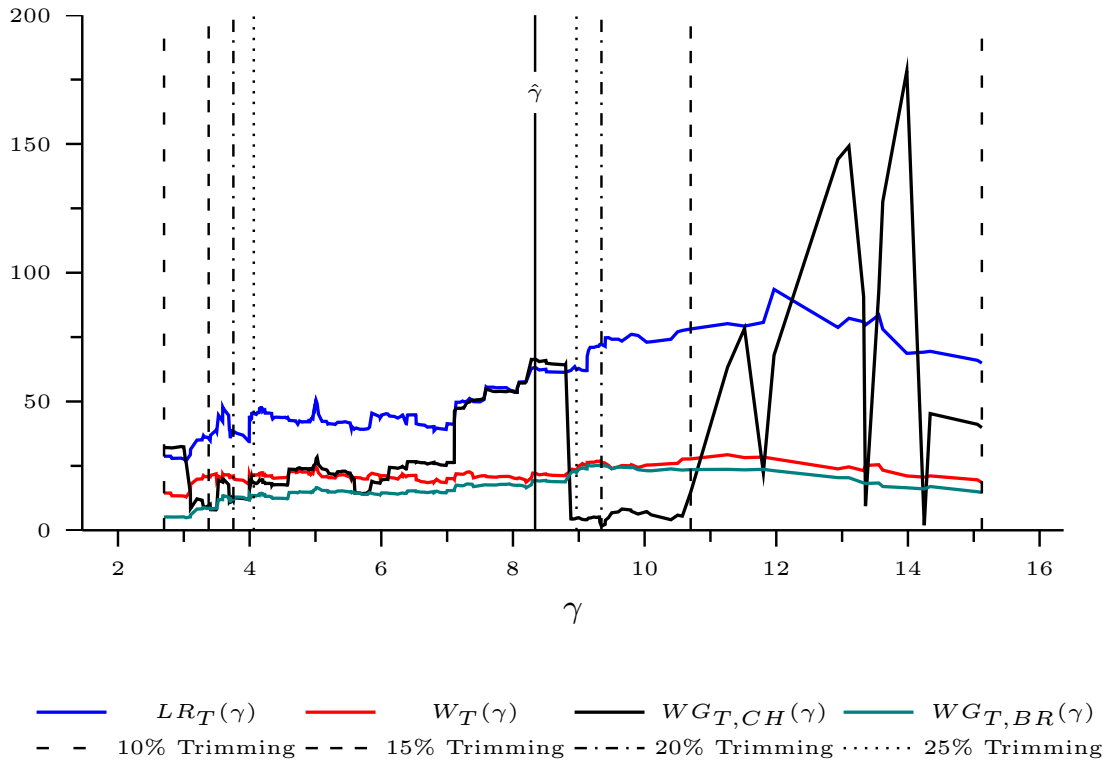
³This is based on the Federal Reserve's use of this threshold in a policy announcement. They later do robustness checks with a larger threshold, and modelled time-varying thresholds.

endogenous regressors and exogenous threshold variables.

Our simulations show that this test has serious size distortions, with rejection frequencies up to three times the nominal size in small samples (see Tables 1 and 2), accompanied by a reversal to severe under-rejections for larger samples of 1000 observations. Tables 1 and 2 show that these size distortions are already present in just-identified models with strong instruments and homoskedastic data. We identify two problems with the CH test that lead to these size distortions, and proceed to correct them.

The first problem is illustrated in Figure 1, where we test for an unknown threshold in the RZ model, and plot the sequence of the CH test statistics over the candidate threshold values, along with the same sequence for three tests we propose.⁴ The plot shows erratic behavior of the CH test sequence, switching frequently between low and high values, especially around the sample edges, but starting already at the 25% and the 75% sample quantiles of the threshold variable. Therefore, the CH test, the maximum of the plotted sequence of tests, can change by a large amount when slightly changing the trimming.⁵ This is problematic for its application in practice, as in general, it may lead to both over- or under-rejection of the null hypothesis, especially since this non-monotonic behavior is not well replicated by the bootstrap critical values even for samples of 1000 observations, as shown in our simulations.

Figure 1: Plot of the CH test sequence (black) versus our proposed test sequences (blue, red, green) for the model in RZ



Note: Above, $WG_{T,CH}(\gamma)$ refers to the CH test sequence, and $WG_{T,BR}(\gamma)$, $W_T(\gamma)$ and $LR_T(\gamma)$ to our three test sequences: the first one is based on GMM estimators, and the last two on 2SLS estimators. The sample size is 500.

⁴Section 2.1 explains how these tests are calculated. Section 4 describes the threshold estimator, the model and the data.

⁵Note that this is not due to the actual threshold estimate being between cut-off points: if there was a threshold, its consistent estimate, based on 2SLS or in CH, with 25% cut-off, is 8.33; however, in our application in Section 4, and in line with RZ, we do not find evidence of such a threshold.

We identify the source of this problem to lie in the computation of the variance estimator in the middle of each Wald test for a candidate threshold. The residuals in the variance estimator are obtained with sub-sample parameter estimators, using observations only below or above each candidate threshold value. When the threshold value is close to the sample edges, these residuals can be very inaccurate approximations of the true underlying errors, because of the slow convergence rate of the sub-sample estimators employed to obtain them. We correct this by obtaining the residuals with full-sample estimators instead. Figure 1 shows that all three test statistics we propose no longer display this non-monotonic behavior, whether computed with generalized method of moment estimators (GMM) estimators as in CH, or with two-stage least squares estimators (2SLS).

A second, yet related issue arises in the construction of the critical values of the CH test. The critical values of unknown threshold tests typically depend on the data, and therefore need to be simulated or bootstrapped. CH propose to bootstrap the critical values via a wild fixed regressor bootstrap and prove the bootstrap validity of their test. However, just like the variance estimator, the bootstrap residuals (and therefore the bootstrap samples) are computed with estimators under the alternative of each candidate threshold value.⁶

While bootstrapping under the alternative does not affect the asymptotic validity of the CH test, it is problematic for two reasons. First, it is computationally much more intensive than computing the bootstrap samples just once, under the null hypothesis, using full-sample estimators. This is because for each bootstrapped test, one needs to compute many bootstrap samples corresponding to each candidate threshold value. Second, just like their sample equivalents, the bootstrap residuals will be inaccurate at the sample edges due to slow convergence of the sub-sample estimators used to employ them. When taking the maximum over the sequence of bootstrapped tests, then doing so for many bootstrap samples, the bootstrapped critical values can become highly unreliable for the original test statistic, even for sample sizes up to 1000 observations. Tables 1 and 2 in the simulation section show severe under-rejection of the null hypothesis for sample sizes of 1000 observations. They also show that bootstrapping under the null hypothesis fixes this issue, leading to correctly sized tests, but only if the variance correction discussed earlier is also employed.⁷

In this paper, we propose three test statistics for testing the null hypothesis of an unknown threshold in threshold models with endogenous regressors and exogenous regressors, and because both their computation and the bootstrap is different than CH, we derive for all three tests their asymptotic distribution and bootstrap validity. The first test we propose is similar to the CH test and uses sub-sample GMM estimators, but, unlike CH, employs a different variance estimator and a null bootstrap. The other two tests are a likelihood ratio (LR) test and a Wald test, both based on 2SLS estimators. The 2SLS estimators are not conventional

⁶Note that bootstrapping under the alternative is not necessary even when the variance estimator is computed with residuals under each alternative hypothesis of a candidate threshold value.

⁷Note that all test statistics for an unknown threshold we consider are non-pivotal, so one cannot expect any bootstrap to provide asymptotic refinements. While for the (trimmed) edges of the sample, the residuals computed with sub-sample estimators and their bootstrapped version are clearly inaccurately estimating the true underlying errors, because of slow convergence of the sub-sample estimator employed to construct them, this is not the case for the middle of the sample. Because both our tests and the CH test take the maximum over all candidate threshold values, around the (trimmed) sample edges or not, it is not possible to derive uniform asymptotic refinements of our tests over the CH test; these refinements will only hold for candidate threshold values around the sample edges. We would like to thank a referee for raising this issue.

and therefore not a special case of the sub-sample GMM estimators used in the CH test, because they use additional information about the first stage being either linear or having itself a threshold, while the GMM estimators do not use this information by construction. Therefore, the resulting test statistics can be equally reliable to the test based on GMM estimators, as shown in our simulations. Because the 2SLS with a first stage threshold require consistent estimators of the first stage threshold parameter, as a by-product of our analysis, we also prove the consistency of ordinary least-squares threshold estimators with a fixed threshold, a result we could not find in the extant literature, only for very specific regression models.⁸

Our paper is closely related to several papers in the change-point literature. Boldea et al. (2019) study the same 2SLS-based test statistics as this paper but for change-points. They also prove bootstrap validity of their tests, however we employ different proof techniques in this paper because the threshold variable is typically correlated with regressors, while the change-points are not, and the asymptotic distributions will also be different. Magnusson and Mavroeidis (2014) use information about change-points in the first stage to improve the power of tests for moment conditions, while we use similar information to improve the size of our tests. Antoine and Boldea (2015) and Antoine and Boldea (2018) also use a full sample first stage or change points in the first stage for more efficient estimation, while we focus on testing.

It should be noted that we allow for endogenous regressors, but not for endogenous threshold variables. For the latter, see *inter alia* Kourtellos et al. (2015), Yu and Phillips (2018), Christopoulos et al. (2021) and Liao et al. (2019). To account for regressor endogeneity, we use instruments for constructing parametric test statistics for thresholds. As a result, our tests have nontrivial local power for $O(T^{-1/2})$ threshold shifts, where T is the sample size. This is in contrast to Yu and Phillips (2018), who do not use instruments, but rather local shifts around the threshold to construct a nonparametric threshold test. As a result, their test covers more general functional forms, at the cost of losing power in $O(T^{-1/2})$ neighborhoods. Additionally, the later paper focuses on cross-sectional models, while our tests are applicable to both cross-sectional models and time series models.

In the empirical application, using the same data and model specification as in RZ, we revisit the question of whether the government spending multipliers are larger in recessions. As in RZ, we cannot rule out that the cumulative government spending multipliers are the same in recessions and expansions. However, we estimate the threshold unemployment rate to be 8.3%, rather than 6.5% as imposed in RZ. This new threshold causes the military spending instrument constructed in RZ to become weaker for deep recessions, suggesting that this instrument is probably most informative at moderate unemployment rates somewhere between 6.5% and 8.3%.

The paper is organized as follows. Section 2 describes the model, the CH test and our test statistics, the proposed bootstrap, as well as the assumptions and all the bootstrap validity results. Section 3 contains simulations and Section 4 contains the empirical application. Section 5 concludes. The Online Supplement, at the end of this document, contains all the proofs.

⁸See Theorem A.2 in the Online Supplement.

2 Threshold Model, Test Statistics and Bootstrap Validity

2.1 Threshold Model and Test Statistics

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

$$\begin{aligned} y_t &= (x_t^\top \theta_{1x}^0 + z_{1t}^\top \theta_{1z}^0) \mathbf{1}[q_t \leq \gamma^0] + (x_t^\top \theta_{2x}^0 + z_{1t}^\top \theta_{2z}^0) \mathbf{1}[q_t > \gamma^0] + \epsilon_t \\ &= w_t^\top \theta_1^0 \mathbf{1}[q_t \leq \gamma^0] + w_t^\top \theta_2^0 \mathbf{1}[q_t > \gamma^0] + \epsilon_t, \end{aligned} \quad (1)$$

where y_t is the scalar dependent variable, x_t is a $p_1 \times 1$ vector of endogenous variables, z_{1t} a $p_2 \times 1$ vector of exogenous variables including the intercept and possibly lags of y_t , q_t is the scalar exogenous threshold variable, $\mathbf{1}[\cdot]$ is the indicator function, $w_t = (x_t^\top, z_{1t}^\top)^\top$ and $\theta_i^0 = (\theta_{ix}^{0\top}, \theta_{iz}^{0\top})^\top$. Let $\gamma^0 \in \Gamma$, a strict subset of the support of q_t , and let $p = p_1 + p_2$. The threshold variable is assumed exogenous and it can be a function of the exogenous regressors. As in CH, the first stage can be a linear model:

$$x_t = \Pi^{0\top} z_t + u_t, \quad (2)$$

or a threshold model:

$$x_t = \Pi_1^{0\top} z_t \mathbf{1}[q_t \leq \rho^0] + \Pi_2^{0\top} z_t \mathbf{1}[q_t > \rho^0] + u_t, \quad (3)$$

where $\rho^0 \in \Gamma$ is a threshold not necessarily coinciding with γ^0 , and z_t are $q \times 1$ strong and valid instruments, including z_{1t} , with $q - p_2 \geq p_1$. We assume that $\mathbb{E}[(\epsilon_t, u_t^\top) | \mathfrak{F}_t] = 0$, where $\mathfrak{F}_t = \sigma\{z_{t-s}, v_{t-s-1}, q_{t-s} | s \geq 0\}$, so that equation (1) can be estimated by either 2SLS or by GMM.

We are interested in testing for an unknown threshold, i.e. the null hypothesis $\mathbb{H}_0 : \theta_1^0 = \theta_2^0 = \theta^0$. CH proposed a test based on GMM estimators of θ_i^0 , ($i = 1, 2$) for each $\gamma \in \Gamma$. Because z_t and q_t are exogenous, the moment conditions

$$E[z_t \epsilon_t \mathbf{1}[q_t \leq \gamma]] = 0, \quad E[z_t \epsilon_t \mathbf{1}[q_t > \gamma]] = 0 \quad (4)$$

hold for all $\gamma \in \Gamma$. Based on these moment conditions, they construct the two-step GMM estimators:

$$\hat{\theta}_{i\gamma, (2)} = \left(\hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{-1} \hat{N}_{i\gamma}^\top \right)^{-1} \hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{-1} \left(T^{-1} \sum_{i\gamma} z_t y_t \right), \quad (5)$$

with $\sum_{1\gamma}(\cdot) = \sum_{t=1}^T (\cdot) \mathbf{1}[q_t \leq \gamma]$, $\sum_{2\gamma}(\cdot) = \sum_{t=1}^T (\cdot) \mathbf{1}[q_t > \gamma]$, $\hat{N}_{i\gamma} = T^{-1} \sum_{i\gamma} w_t z_t^\top$ and with $\hat{H}_{\epsilon, i\gamma} = T^{-1} \sum_{i\gamma} \hat{\epsilon}_{t, \gamma, (1)}^2 z_t z_t^\top$. Here, $\hat{\epsilon}_{t, \gamma, (1)} = y_t - w_t^\top \hat{\theta}_{1\gamma, (1)} \mathbf{1}[q_t \leq \gamma] - w_t^\top \hat{\theta}_{2\gamma, (1)} \mathbf{1}[q_t > \gamma]$ are the first step GMM residuals for each γ , and $\hat{\theta}_{i\gamma, (1)}$ are consistent first-step versions of $\hat{\theta}_{i\gamma, (2)}$, for example by replacing $\hat{H}_{\epsilon, i\gamma}$ with $\hat{M}_{i\gamma} = T^{-1} \sum_{i\gamma} z_t z_t^\top$. These estimators can be used to construct a Wald test for each γ , and taking the supremum of this sequence of Wald tests over $\gamma \in \Gamma$ yields the test in CH:

$$WG_{T, CH} = \sup_{\gamma \in \Gamma} WG_{T, CH}(\gamma) = \sup_{\gamma \in \Gamma} T \left(\hat{\theta}_{1\gamma, (2)} - \hat{\theta}_{2\gamma, (2)} \right)^\top \hat{V}_{\gamma, (1)}^{-1} \left(\hat{\theta}_{1\gamma, (2)} - \hat{\theta}_{2\gamma, (2)} \right), \quad (6)$$

where $\hat{V}_{\gamma,(1)} = \sum_{i=1}^2 \left(\hat{N}_{i\gamma} \hat{H}_{\epsilon,i\gamma}^{-1} \hat{N}_{i\gamma}^\top \right)^{-1}$.

As shown in CH, the asymptotic distribution of the test statistic (6) is non-pivotal and therefore needs to be simulated/bootstrapped for a given application. CH propose to generate new pseudo-dependent variables $y_{t,\gamma}^b = \hat{\epsilon}_{t,\gamma,(2)} \eta_t$, where $\eta_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ and $\hat{\epsilon}_{t,\gamma,(2)}$ denote the second step GMM residuals for each value of γ , and recalculate the test statistic (6) for each γ using $y_{t,\gamma}^b$ instead of y_t , and then for many bootstrap samples.⁹ Even though CH prove validity of their bootstrap procedure in large samples, Tables 1 and 2 show that their bootstrap does not replicate well the empirical distribution of the test statistic in finite samples, being severely over-/undersized for small/large samples.

Tables 1 and 2 in the simulation section show that these size distortions are due to two interacting phenomena: the type of bootstrap employed, and the way the heteroskedasticity-robust variance estimators are computed. We therefore employ two corrections. First, we adjust the bootstrap such that the pseudo-dependent variable y_t^b is constructed using full-sample residuals. That is, we replace $y_t^b = \hat{\epsilon}_{t,\gamma,(2)} \eta_t$ by $y_t^b = \hat{\epsilon}_{t,(2)} \eta_t$ where $\hat{\epsilon}_{t,(2)} = y_t - w_t^\top \hat{\theta}_{(2)}$ and $\hat{\theta}_{(2)}$ is the second step GMM estimate under \mathbb{H}_0 . This gets rid of the undersizing of the CH test statistic documented in the simulation section: the residuals become more accurate around the sample edges as they are not constructed with sub-sample estimators. However, the simulations now indicate that the test is oversized (see Tables 1 and 2, column ‘‘Mix’’).

Therefore, we employ a second correction, where the heteroskedasticity-robust variance estimators are also computed with full-sample parameter estimates. More exactly, rather than using $\hat{\epsilon}_{t,\gamma,(1)}$ and $\hat{\epsilon}_{t,\gamma,(2)}$ in the expression for $\hat{H}_{\epsilon,i\gamma}$, we use $\hat{\epsilon}_{t,(1)} = y_t - w_t^\top \hat{\theta}_{(1)}$ instead, where $\hat{\theta}_{(1)}$ is the first step full-sample GMM estimate (so we redefine $\hat{H}_{\epsilon,i\gamma} = \sum_{i\gamma} \hat{\epsilon}_{t,(1)}^2 z_t z_t^\top$). As Tables 1 and 2 show (column ‘‘BR’’, ‘‘bootstrap/rectification’’), this yields correctly sized sample test statistics in all samples considered.

Note that both effects that we correct for are due to unstable estimates of the residuals at the sample edges below/above the 15%/85%-quantiles of the empirical distribution of q_t . The test employing these two corrections is denoted by $WG_{T,BR} = \sup_{\gamma \in \Gamma} WG_{T,BR}(\gamma)$.

We also consider two 2SLS-based test statistics, because the GMM estimators involved in the computation of the tests above do not use information about the linearity or lack of linearity of the first stage. Therefore, they are not more efficient than the 2SLS estimators that use this information (see Antoine and Boldea (2015) for a formal proof of this statement for change-point models), so there is no reason to expect that 2SLS-based tests will be inferior to the GMM-based tests.

⁹Note that the pseudo-dependent variables are generated without adding back the estimated mean to the bootstrap residuals. This is inconsequential to the analysis because the test statistic is based on mean differences across regimes of low or high q_t , and these are zero under the null of no threshold.

The likelihood-ratio type and a Wald-type test statistic for $\mathbb{H}_0 : \theta_1^0 = \theta_2^0$ based on 2SLS estimators are:

$$LR_T = \sup_{\gamma \in \Gamma} \frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/(T - 2p)} \quad (7a)$$

$$W_T = \sup_{\gamma \in \Gamma} T \left(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma} \right)^\top \hat{V}_\gamma^{-1} \left(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma} \right) \quad (7b)$$

where $SSR_0 = \sum_{t=1}^T (y_t - \hat{w}_t^\top \hat{\theta})^2$, with $\hat{\theta} = (\sum_{t=1}^T \hat{w}_t \hat{w}_t^\top)^{-1} (\sum_{t=1}^T \hat{w}_t y_t)$ the full-sample 2SLS estimator, $SSR_1(\gamma) = \sum_{i=1}^2 \sum_{i\gamma} (y_t - \hat{w}_t^\top \hat{\theta}_{i\gamma})^2$, with $\hat{\theta}_{i\gamma} = (\sum_{i\gamma} \hat{w}_t \hat{w}_t^\top)^{-1} (\sum_{i\gamma} \hat{w}_t y_t)$ the split-sample 2SLS estimators. Here, $\hat{w}_t = (\hat{x}_t^\top, z_{1t}^\top)^\top$ stacks the predicted endogenous variables \hat{x}_t and the exogenous variables z_{1t} . The predicted endogenous variables are obtained either via estimating the linear first stage equation (2):

$$\hat{\Pi} = \left(\sum_{t=1}^T z_t z_t^\top \right)^{-1} \left(\sum_{t=1}^T z_t x_t^\top \right), \quad \hat{x}_t = \hat{\Pi}^\top z_t \quad (8)$$

or via estimating the threshold first-stage equation (3):

$$\hat{\Pi}_{i\rho} = \left(\sum_{i\rho} z_t z_t^\top \right)^{-1} \left(\sum_{i\rho} z_t x_t^\top \right), \quad i = 1, 2, \quad \hat{\Pi}_{t\rho} = \hat{\Pi}_{1\rho} \mathbf{1}[q_t \leq \rho] + \hat{\Pi}_{2\rho} \mathbf{1}[q_t > \rho] \quad (9a)$$

$$\hat{\rho} = \underset{\rho \in \Gamma}{\operatorname{argmin}} \left[\operatorname{tr} \sum_{t=1}^T (x_t - \hat{\Pi}_{t\rho}^\top z_t)(x_t - \hat{\Pi}_{t\rho}^\top z_t)^\top \right] \quad \hat{\Pi}_i = \left(\sum_{i\hat{\rho}} z_t z_t^\top \right)^{-1} \left(\sum_{i\hat{\rho}} z_t x_t^\top \right), \quad i = 1, 2 \quad (9b)$$

$$\hat{x}_t = \hat{\Pi}_1^\top z_t \mathbf{1}[q_t \leq \hat{\rho}] + \hat{\Pi}_2^\top z_t \mathbf{1}[q_t > \hat{\rho}] \quad (9c)$$

Lastly, $\hat{V}_\gamma \xrightarrow{p} V_\gamma = \lim \operatorname{Var}[T^{1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma})]$.¹⁰ Unlike the sup Wald test in Hall et al. (2012), which is the change-point counterpart of the test here, our test – through the way \hat{V}_γ is defined – takes into account that the 2SLS estimators $\hat{\theta}_{1\gamma}$ and $\hat{\theta}_{2\gamma}$ are correlated through either a full-sample first-stage or through misalignment of ρ^0 and γ . Moreover, as in the case of CH’s GMM-test, the 2SLS test-statistics are non-pivotal and, therefore, need to be simulated/bootstrapped. The next subsection describes the bootstrap we propose and contains results for asymptotic validity of this bootstrap for all three tests proposed.

2.2 Bootstraps and their Validity

The bootstrap employed for both CH GMM test and our GMM test is a wild bootstrap with fixed regressors because it does not bootstrap the regressors w_t, x_t and the instruments $z_{1,t}$. We already alluded to the proposed change in the bootstrap procedure for the CH test in the previous section. These changes are summarized in the Algorithms 1 and 2 below. The difference between the CH test and our test are highlighted in lines 3–6 of the below algorithms. Since CH construct their pseudo-dependent variable for *each* γ separately, the for-loop over γ starts already in line 3 of Algorithm 1, as opposed to line 5 in Algorithm 2 when the *same*

¹⁰The explicit expressions for \hat{V}_γ and V_γ are given in Online Supplement Section A.3.1, and Definition 1 for a linear first stage, and in Online Supplement Section A.4.1, Definition 2 for a threshold first stage, together with the expressions in the asymptotic distributions of the 2SLS test-statistics.

pseudo-dependent variable is used for all values of γ . Line 6 in both algorithms indicates the difference in constructing the heteroskedasticity-robust variance estimators $\hat{H}_{\epsilon, i\gamma}^b$ and in $\hat{H}_{\epsilon, i\gamma}$.

Algorithm 1: CH Bootstrap	Algorithm 2: Modified CH Bootstrap
<pre> 1: Input: $\{\hat{\epsilon}_{t,\gamma,(2)}\}_t, \gamma, \{w_t\}_t, \Gamma, B$ 2: for $b \in \{1, \dots, B\}$ do 3: for $\gamma \in \Gamma$ do 4: $\eta_t \sim i.i.d. \mathcal{N}(0, 1)$ 5: $y_{t,\gamma}^b \leftarrow \hat{\epsilon}_{t,\gamma,(2)} \eta_t$ 6: $WG_{T,CH}^b(\gamma) \leftarrow$ use $\{(y_{t,\gamma}^b, w_t)\}_t$ to compute RHS in (6) using $\hat{\epsilon}_{t,\gamma,(1)}$ in $\hat{H}_{\epsilon, i\gamma}^b$ 7: end for 8: $WG_{T,CH}^b \leftarrow \sup_{\gamma \in \Gamma} WG_{T,CH}^b(\gamma)$ 9: end for 10: Return $\{WG_{T,CH}^b\}_b$ </pre>	<pre> 1: Input: $\{\hat{\epsilon}_{t,(2)}\}_t, \{w_t\}_t, \Gamma, B$ 2: for $b \in \{1, \dots, B\}$ do 3: $\eta_t \sim i.i.d. (0, 1)$ 4: $y_t^b \leftarrow \hat{\epsilon}_{t,(2)} \eta_t$ 5: for $\gamma \in \Gamma$ do 6: $WG_{T,BR}^b(\gamma) \leftarrow$ use $\{(y_t^b, w_t)\}_t$ to compute RHS in (6) using $\hat{\epsilon}_{t,(1)}$ in $\hat{H}_{\epsilon, i\gamma}^b$ 7: end for 8: $WG_{T,BR}^b \leftarrow \sup_{\gamma \in \Gamma} WG_{T,BR}^b(\gamma)$ 9: end for 10: Return $\{WG_{T,BR}^b\}_b$ </pre>

Algorithm 3 below describes the wild fixed regressor bootstrap for our proposed 2SLS test-statistics, where regressors w_t and instruments $z_{1,t}$ are kept fixed in the bootstrap. The first stage linearity or lack thereof is taken into account in computing \hat{x}_t - equation (8) or (9c) respectively. For these tests, we need to know whether the first stage is linear or not; however, this is not necessarily a drawback in empirical work, because such knowledge is required for estimating the threshold parameter γ^0 consistently (see CH).

Algorithm 3: 2SLS Bootstrap
<pre> 1: Input: $\{y_t, x_t, z_t, \hat{x}_t\}_t, \hat{\theta}, \Gamma, B$ 2: $w_t \leftarrow (x_t^\top, z_{1t}^\top)^\top$ 3: $\hat{\epsilon}_t \leftarrow y_t - w_t^\top \hat{\theta}$ 4: $\hat{u}_t \leftarrow x_t - \hat{x}_t$ 5: for $b \in \{1, \dots, B\}$ do 6: $\eta_t \sim (0, 1)$ 7: $u_t^b \leftarrow \hat{u}_t \eta_t$ 8: $x_t^b \leftarrow \hat{x}_t + u_t^b$ 9: $w_t^b \leftarrow (x_t^{b\top}, z_{1t}^\top)^\top$ 10: $\epsilon_t^b \leftarrow \hat{\epsilon}_t \eta_t$ 11: $y_t^b \leftarrow w_t^{b\top} \hat{\theta} + \epsilon_t^b$ 12: for $\gamma \in \Gamma$ do 13: $LR_T^b(\gamma) \leftarrow$ use (y_t^b, w_t^b, z_t) to compute RHS in (7a) using either (8) or (9a)–(9c) 14: $W_T^b(\gamma) \leftarrow$ use (y_t^b, w_t^b, z_t) to compute RHS in (7b) using either (8) or (9a)–(9c) 15: end for 16: $LR_T^b \leftarrow \sup_{\gamma \in \Gamma} LR_T^b(\gamma)$ 17: $W_T^b \leftarrow \sup_{\gamma \in \Gamma} W_T^b(\gamma)$ 18: end for 19: Return: $\{LR_T^b, W_T^b\}_b$ </pre>

We now derive the asymptotic properties of our tests¹¹ and show their bootstrap validity. First define $g_t = x_t - u_t$, $h_t = y_t - \epsilon_t$, $M_{1\gamma} = E[z_t z_t^\top \mathbf{1}[q_t \leq \gamma]]$, $M = \text{plim}_{\gamma \rightarrow \infty} M_{1\gamma} = E[z_t z_t^\top]$, $M_{2\gamma} = M - M_{1\gamma}$, and

¹¹We focus on the GMM-based test; the asymptotic distributions of the 2SLS tests are given in the Online Supplement, Sections A.3.1 and A.4.1.

$v_t = (\epsilon_t, u_t^\top)^\top$. Let $\|\cdot\|$ be the Euclidean norm. The following assumptions are similar to CH.

Assumption 1.

- (a) $E[v_t | \mathfrak{F}_t] = 0$ with $\mathfrak{F}_t = \sigma\{z_{t-s}, v_{t-s-1}, q_{t-s} | s \geq 0\}$;
- (b) The series $(v_t, g_t, h_t, q_t, z_t)$ is strictly stationary with ρ -mixing coefficient $\rho(m) = \mathcal{O}(m^{-A})$ for some $A > \frac{a}{a-1}$ and $1 < a \leq r$;
- (c) $E[\|z_t\|^{4r}] < \infty$, $E[\|v_t\|^{4r}] < \infty$ for some $r > 1$;
- (d) $\inf_{\gamma \in \Gamma^0} \det M_{1\gamma} > 0$, and if (3) holds, then $M_{1\gamma_1} - M_{1\gamma_2}$ has eigenvalues bounded away from zero for any $\gamma_1 > \gamma_2$. Moreover, $M_{1\gamma}$ is continuous at $\gamma = \rho^0$;
- (e) The threshold variable q_t has a continuous pdf $f(q_t)$ with $\sup_{q_t \in \Gamma} |f(q_t)| < \infty$;
- (f) $E[v_t v_t^\top]$ and $E[(v_t v_t^\top) \otimes (z_t z_t^\top)]$ are two p.d. matrices of constants;
- (g) The coefficient matrices Π^0 (for the linear first stage (2)) or Π_1^0, Π_2^0 (for the threshold first stage (3)) are full rank, and $\Pi_1^0 - \Pi_2^0 \neq 0$.

Most of these assumptions are also used in CH. Assumption 1 (a) is typically needed for nonlinear models. Assumption 1 (b) is also needed, as the only uniform law of large numbers and functional central limit theorem for partial sums in $\mathbf{1}[q_t \leq \gamma]$ that we are aware of derives from Hansen (1996) and require strict stationarity (see Lemma A.1-A.2 in the Online Supplement). Assumption 1 (c) is a typical moment condition. Assumption 1 (d) is slightly different than CH: they also impose that $M_{1\gamma}$ is p.d. for all γ , but we require that the increments in $M_{1\gamma}$ are p.d. in the limit with eigenvalues bounded away from zero. The latter is technical in nature and required to obtain quantities bounded in probability in order to provide a self-contained proof of super-consistency of $\hat{\rho}$ in a threshold first stage model. Assumptions 1 (e) is standard in the threshold literature, and 1 (g) is an identification condition for a possible threshold in the first stage. Assumption 1 (f) is needed for uniqueness of the asymptotic distributions of the test statistics proposed.

With this assumption, we first show that employing the new heteroskedasticity-robust estimators do not alter the distribution of the CH test.

Theorem 1 (Asymptotics Modified GMM Wald-Test). *Let y_t be generated by (1), x_t be generated by either a linear first stage (2) or a threshold first stage (3). Moreover, let the GMM test be computed as in (6) but with $\hat{H}_{\epsilon, i\gamma}$ constructed using $\hat{\epsilon}_{t,(1)}$. Then, under \mathbb{H}_0 and Assumption 1, $WG_{T, BR}$ has the same limiting distribution as $WG_{T, CH}$.*

To show the validity of the null bootstrap, we require the following additional assumption:

Assumption 2.

- (a) $\eta_t \stackrel{iid}{\sim} (0, 1)$ with $E^b[\eta_t^4] < \infty$, where $E^b[\cdot]$ is the expectation with respect to the bootstrap probability measure;

(b) $E\left[\|z_t\|^{8r}\right] < \infty$ and $E\left[\|v_t\|^{8r}\right] < \infty$ for some $r > 1$.

Assumption 2 (a) is common for the wild bootstrap (also see Boldea et al., 2019), and typical choices for η_t are the normal distribution, the Rademacher distribution, and the asymmetric two-point distribution in Mammen (1993). CH propose using the normal distribution, but we use both the normal distribution and the Mammen (1993) distribution, as the latter yields better results for the GMM-Wald test, see Tables 1–4. Assumption 2 (b) is only needed for $\hat{H}_{\epsilon, i\gamma}^b$ to weakly converge to $H_{\epsilon, i\gamma}^b$ in probability under the bootstrap measure. Theorem 2 proves the asymptotic validity of the null bootstrap for $WG_{T, BR}$.

Theorem 2 (Bootstrap Validity Modified GMM Wald-Test). *Let y_t be generated by (1), x_t be generated by either a linear first stage (2) or a threshold first stage (3). Moreover, let the GMM Wald-test be computed as in (6) but with $\hat{H}_{\epsilon, i\gamma}$ constructed using $\hat{\epsilon}_{t,(1)}$ and using the bootstrap in Algorithm 2. Then, under \mathbb{H}_0 and Assumptions 1 and 2,*

$$\sup_{c \in \mathbb{R}} |P^b(WG_{T, BR}^b \leq c) - P(WG_{T, BR} \leq c)| \xrightarrow{P} 0.$$

For the 2SLS-based test-statistics, the asymptotic distributions are cumbersome and not of main interest. Therefore, we relegate these results to the Online Supplement, Sections A.3.1 and A.4.1. However, in order to derive these asymptotic distributions, we also provide in the Online Supplement, Theorem 5, a self-contained proof of super-consistency of the first stage (ordinary least-squares) threshold parameter estimate $\hat{\rho}$. This was also shown in Chan (1993), but for a threshold autoregressive model where z_t and q_t are lags of x_t . This proof may be of interest in its own right, as it extends proof techniques from change point analysis to threshold models.

The asymptotic distributions of the 2SLS based tests are also non-pivotal, and we conclude this section by stating the asymptotic validity of the bootstrap for these tests.

Theorem 3 (Bootstrap Validity – 2SLS Tests). *Let y_t be generated by (1) and x_t be generated by the linear first stage (2) or by the threshold first stage (3). Moreover, let the 2SLS-based test-statistics be computed as described in (7a) and (7b), and using the bootstrap in Algorithm 3. Then, under \mathbb{H}_0 and Assumption 1,*

(i) $\sup_{c \in \mathbb{R}} |P^b(LR_T^b \leq c) - P(LR_T \leq c)| \xrightarrow{P} 0$ if Assumption 2 (a) holds in addition;

(ii) $\sup_{c \in \mathbb{R}} |P^b(W_T^b \leq c) - P(W_T \leq c)| \xrightarrow{P} 0$ if Assumption 2 holds in addition.

3 Simulations

Consider the following data generating process (DGP) for $t = 1, \dots, T$:

$$y_t = 1 + x_t + \delta_x x_t \mathbf{1}[q_t > \gamma^0] + \epsilon_t, \quad x_t = 1 + z_t + \delta_\Pi z_t \mathbf{1}[q_t > \rho^0] + u_t,$$

where $z_t \stackrel{iid}{\sim} \mathcal{N}(1, 1)$, $q_t = z_t + 1$, and z_t , x_t , and q_t are scalars. We set $\delta_\Pi = 0$ for a linear first stage (LFS) and $\delta_\Pi \in \{-0.5, 0.5, 1\}$ for a threshold first stage (TFS) with $\rho^0 = 1.75$. Under the null hypothesis, $\delta_x = 0$,

and under the alternative hypothesis, $\delta_x = 0.25$ with $\gamma^0 = 2.25$.¹² To generate ϵ_t , we define $e_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ and consider the following three cases.

In case (a), the errors are homoskedastic, i.e. $\epsilon_t = e_t$, and the econometrician knows this. Therefore, we use the i.i.d. bootstrap instead of the wild bootstrap, and make two adjustments to the computation of the test statistics. First, $v_t^b \stackrel{iid}{\sim} \mathcal{N}(0, \hat{\Sigma}_v)$ with $\hat{\Sigma}_v = T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top$ for 2SLS. For GMM, $\epsilon_t^b \stackrel{iid}{\sim} \mathcal{N}(0, \hat{\sigma}_\epsilon^2)$ given the data, with $\hat{\sigma}_\epsilon^2 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{t,(1)}^2$, and $\hat{\epsilon}_{t,(1)} = y_t - w_t^\top \hat{\theta}$, with $\hat{\theta}$ the full-sample 2SLS estimator. Second, all heteroskedasticity-robust estimators are replaced by their homoskedastic analogs. For example, $E[z_t z_t^\top \epsilon_t^2 \mathbf{1}[q_t \leq \gamma]]$ is no longer estimated by $T^{-1} \sum_{1\gamma} z_t z_t^\top \hat{\epsilon}_{t,(1)}^2 \mathbf{1}[q_t \leq \gamma]$, but by $\hat{\sigma}_\epsilon^2 (T^{-1} \sum_{1\gamma} z_t z_t^\top \mathbf{1}[q_t \leq \gamma])$. The same applies to the case of CH’s original bootstrap, except that the residuals are computed for each value of γ rather than under \mathbb{H}_0 .

In case (b), the errors are still homoskedastic, i.e. $\epsilon_t = e_t$, but this is not known to the econometrician. Therefore, the heteroskedasticity-robust variance estimators described in Sections 2.1 and 2.2 are employed.

In case (c), the errors are conditional heteroskedastic, i.e. $\epsilon_t = e_t \cdot z_t / \sqrt{2}$ with $Var(e_t) = Var(u_t) = 1$ and $Cov(u_t, e_t) = 0.5$, and heteroskedasticity-robust variance estimators are employed.

In Tables 1 and 3 cases (b)-(c), the bootstrap is performed using $\eta_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$. For all other results, $\eta_t \stackrel{iid}{\sim} (0, 1)$ with draws from the asymmetric two-point distribution proposed by Mammen (1993). In all cases besides (a), we use the wild bootstrap as described in Section 2.2.

There are 500 bootstrap samples. For each simulation, we compute the 95% quantile of the bootstrap distribution of the test statistic, and if the test in the original sample is above this quantile, we reject, else we do not reject. γ is varied between all sample realizations of q_t from its 15% quantile to its 85% quantile. We report the rejection frequency of each test statistic in 1000 simulations under the null and at 5% nominal size (Tables 1–4), and under the alternative we plot the size-adjusted power, where the size-adjustment is made relative to the null DGPs described above (Figure 2).

Tables 1 and 2 show that the bootstrap procedure originally proposed by CH has heavy size distortions in both directions. In particular, the test moves from being heavily oversized in small samples to severely undersized in large samples (columns “CH”). This originates from imprecise residual estimates and imprecise $\hat{H}_{\epsilon, i\gamma}$ for small and moderate sample sizes pertinent to applications. In particular, when γ is close to the 15% or 85% quantiles of q_t , there is not enough data to obtain precise residuals and precisely estimate $H_{\epsilon, i\gamma}$. Moreover, changing the CH bootstrap to a null bootstrap results in oversized tests for all considered sample sizes (columns “Mix”). This problem is rectified by modifying $\hat{H}_{\epsilon, i\gamma}$ in the original test statistic, as evident from columns “BR” in Tables 1 and 2, where the empirical sizes are much closer to the nominal size. Tables 3 and 4 show that the 2SLS tests are in almost all cases close to nominal sizes, even in small samples, and that there is no clear ranking among the three proposed tests.

¹²Note that because of just-identification, there is no difference between the first and the second-step GMM estimators, therefore $\hat{\theta}_{(2)} = \hat{\theta}_{(1)} = (\sum_{t=1}^T w_t z_t^\top)^{-1} (T^{-1} \sum_{t=1}^T z_t y_t)$, and $\hat{\theta}_{i\gamma, (2)} = \hat{\theta}_{i\gamma, (1)} = (\hat{N}_{i\gamma}^\top)^{-1} (T^{-1} \sum_{i\gamma} z_t y_t)$.

Table 1: Rejection frequencies under the null DGP, 5% nominal size and $\eta_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ – GMM Tests

	T	LFS			TFS, $\delta_{\Pi} = -0.5$			TFS, $\delta_{\Pi} = 0.5$			TFS, $\delta_{\Pi} = 1$		
		CH	Mix	BR	CH	Mix	BR	CH	Mix	BR	CH	Mix	BR
(a)	100	11.5%	17.9%	4.8%	10.8%	17.0%	3.9%	10.0%	15.0%	4.5%	8.0%	13.4%	4.5%
	250	8.6%	12.9%	5.8%	7.8%	13.6%	4.8%	6.1%	10.1%	5.7%	5.1%	8.9%	5.7%
	500	4.4%	11.2%	4.3%	4.6%	14.2%	5.5%	2.9%	9.6%	4.6%	2.5%	8.8%	4.6%
	1000	1.8%	9.5%	4.9%	2.9%	11.3%	4.9%	1.7%	9.1%	4.9%	1.2%	8.8%	4.5%
(b)	100	12.0%	17.4%	7.5%	11.1%	17.3%	6.2%	10.5%	15.6%	7.9%	9.6%	15.0%	8.0%
	250	4.9%	11.8%	7.5%	6.0%	14.1%	7.2%	4.8%	10.9%	7.2%	4.3%	10.2%	7.1%
	500	2.4%	10.7%	6.4%	3.5%	12.3%	6.3%	2.3%	9.4%	6.9%	2.4%	9.1%	6.8%
	1000	0.7%	8.7%	4.6%	1.3%	10.6%	4.6%	0.7%	8.2%	4.3%	0.6%	7.8%	4.4%
(c)	100	6.9%	13.8%	8.1%	7.2%	12.9%	6.8%	7.8%	12.6%	8.6%	8.0%	12.6%	8.9%
	250	2.7%	11.2%	6.3%	3.1%	11.9%	5.6%	2.8%	9.6%	5.9%	2.2%	8.8%	6.0%
	500	2.0%	8.6%	6.6%	2.3%	11.0%	7.6%	1.8%	7.8%	6.7%	1.7%	7.1%	6.6%
	1000	0.4%	9.7%	4.7%	0.8%	11.2%	4.6%	0.5%	8.8%	4.4%	0.5%	8.4%	3.8%

CH refers to the $WG_{T,CH}$ test in CH (Algorithm 1) and BR refers to our modified $WG_{T,BR}$ test (Algorithm 2). Mix refers to the mixture between the two, where the bootstrap is conducted using \mathbb{H}_0 residuals, whereas $\hat{H}_{\epsilon, i\gamma}$ are constructed using \mathbb{H}_A -residuals. Panel (a) corresponds to known homoskedasticity, panel (b) to unknown homoskedasticity and panel (c) to known heteroskedasticity.

Table 2: Rejection frequencies under the null DGP, 5% nominal size and $\eta_t \stackrel{iid}{\sim} Mammen$ – GMM Tests

	T	LFS			TFS, $\delta_{\Pi} = -0.5$			TFS, $\delta_{\Pi} = 0.5$			TFS, $\delta_{\Pi} = 1$		
		CH	Mix	BR	CH	Mix	BR	CH	Mix	BR	CH	Mix	BR
(b)	100	12.0%	13.1%	5.5%	11.7%	14.1%	4.9%	10.6%	11.6%	6.1%	9.6%	11.0%	5.9%
	250	4.6%	11.9%	5.5%	5.5%	13.7%	4.7%	4.0%	10.5%	5.6%	3.9%	9.5%	5.6%
	500	2.2%	8.7%	5.3%	2.7%	11.3%	4.8%	1.8%	8.6%	5.0%	1.8%	8.5%	5.2%
	1000	0.9%	9.3%	5.6%	1.7%	9.4%	5.5%	0.7%	8.8%	5.7%	0.6%	8.0%	5.4%
(c)	100	7.9%	10.5%	6.1%	7.3%	10.6%	4.8%	7.9%	9.4%	6.3%	8.3%	8.9%	6.3%
	250	3.3%	9.9%	5.0%	3.7%	11.4%	4.5%	2.9%	7.7%	4.5%	2.8%	7.3%	4.1%
	500	1.2%	6.6%	4.8%	1.5%	7.8%	4.6%	1.2%	5.0%	4.9%	1.0%	4.6%	4.9%
	1000	0.8%	6.9%	5.3%	1.3%	9.3%	5.1%	0.6%	6.4%	5.4%	0.3%	5.9%	5.7%

See Table 1 Notes.

Table 3: Rejection frequencies under the null DGP, 5% nominal size and $\eta_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ – 2SLS Tests

	T	LFS		TFS, $\delta_{\Pi} = -0.5$		TFS, $\delta_{\Pi} = 0.5$		TFS, $\delta_{\Pi} = 1$	
		LR	W	LR	W	LR	W	LR	W
(a)	100	6.0%	6.2%	2.9%	4.5%	2.1%	2.4%	2.9%	2.6%
	250	5.0%	5.0%	2.0%	4.0%	2.6%	3.2%	3.4%	3.8%
	500	5.3%	5.0%	2.8%	3.4%	2.8%	2.8%	4.5%	4.3%
	1000	4.2%	4.4%	3.9%	3.4%	4.2%	3.3%	4.9%	3.7%
(b)	100	4.6%	8.4%	2.3%	8.2%	1.5%	5.5%	2.1%	7.2%
	250	5.3%	8.8%	2.3%	6.0%	2.3%	5.4%	4.2%	7.0%
	500	5.7%	6.7%	4.0%	6.0%	3.3%	5.6%	5.5%	6.7%
	1000	6.5%	6.9%	3.0%	4.8%	3.7%	5.5%	4.8%	6.3%
(c)	100	5.1%	8.2%	3.7%	7.8%	2.7%	5.1%	3.2%	6.3%
	250	5.3%	6.9%	3.9%	5.4%	3.9%	5.4%	5.7%	6.7%
	500	7.0%	7.6%	5.9%	6.6%	5.9%	5.2%	6.9%	7.2%
	1000	6.6%	5.3%	5.8%	4.1%	5.2%	3.5%	5.5%	4.0%

LR and W refer to the LR_T and W_T tests. Panel (a) corresponds to known homoskedasticity, panel (b) to unknown homoskedasticity and panel (c) to known heteroskedasticity.

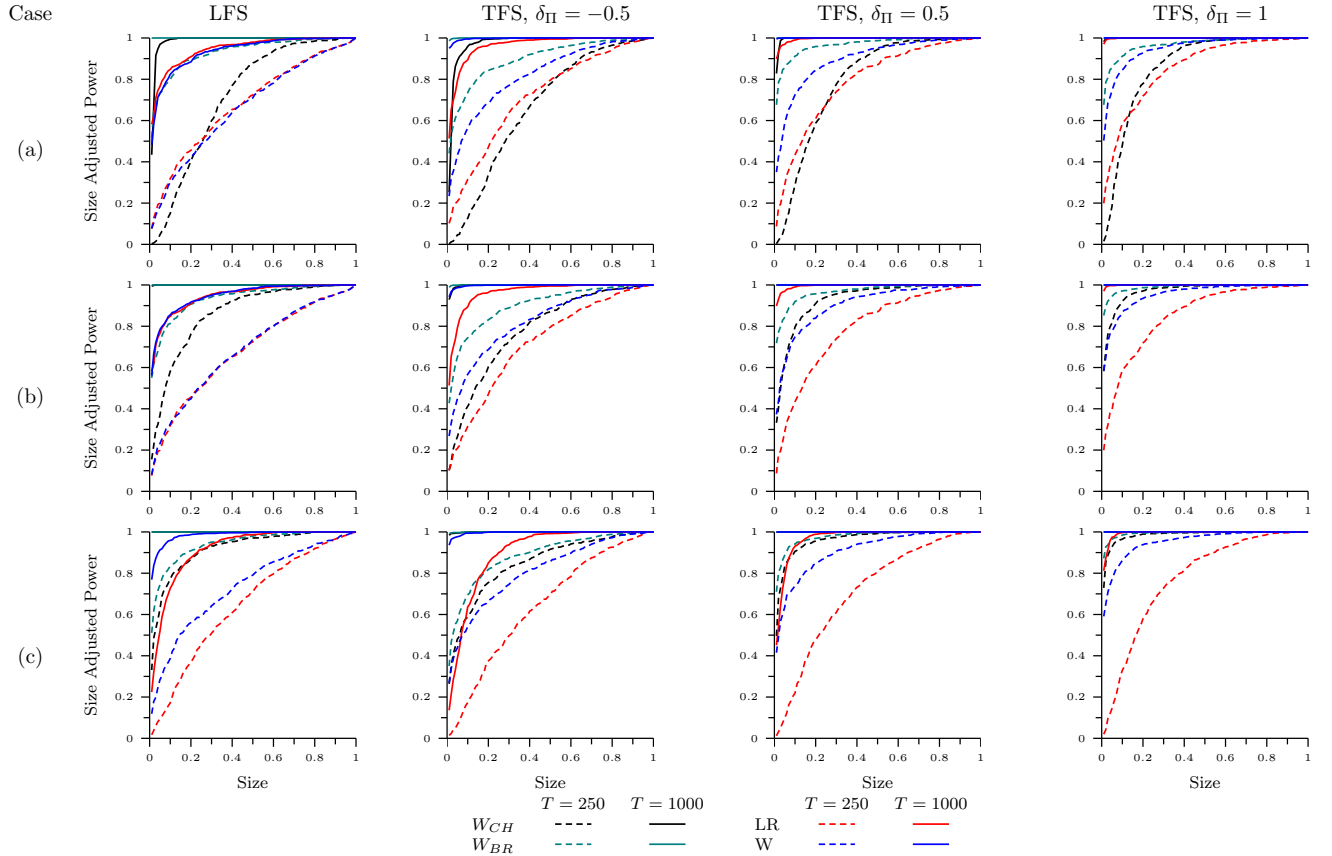
Table 4: Rejection frequencies under the null DGP, 5% nominal size and $\eta_t \stackrel{iid}{\sim} Mammen$ – 2SLS Tests

	T	LFS		TFS, $\delta_{\Pi} = -0.5$		TFS, $\delta_{\Pi} = 0.5$		TFS, $\delta_{\Pi} = 1$	
		LR	W	LR	W	LR	W	LR	W
(b)	100	5.0%	5.9%	2.1%	6.3%	2.0%	4.1%	2.2%	5.1%
	250	5.6%	6.8%	2.5%	5.2%	1.9%	4.2%	3.3%	5.1%
	500	4.4%	4.6%	3.2%	4.0%	3.1%	4.1%	4.7%	4.3%
	1000	5.4%	6.0%	4.6%	4.7%	3.0%	4.5%	4.4%	5.2%
(c)	100	5.3%	5.2%	3.8%	6.4%	3.2%	3.9%	3.7%	4.0%
	250	5.3%	5.6%	3.4%	4.8%	3.3%	3.8%	4.4%	3.9%
	500	5.3%	5.9%	4.6%	2.8%	4.7%	3.2%	5.4%	4.1%
	1000	5.1%	6.6%	4.0%	4.8%	4.4%	4.3%	4.4%	5.3%

See Table 3 Notes

We also assess the power of all tests. For a large threshold $\delta_x = 1$, all tests have power virtually equal to one even for sample sizes of $T = 250$ and therefore we do not report these results. Figure 2 shows the power properties for a small threshold of $\delta_x = 0.25$. In small samples, the Wald tests dominate the LR test for all cases (a)-(c). Note that this is not necessarily for classical reasons of correcting for heteroskedasticity, as all tests are non-pivotal and bootstrapped. The power differences among all three tests vanish as the sample size grows. Therefore, we argue that all the tests proposed provide reliable alternatives in moderate samples pertinent to macroeconomic applications.

Figure 2: Size-Adjusted Power Plots for $\delta_x = 0.25$



LR and W refer to our LR_T and W_T tests. CH refers to the $WG_{T,CH}$ test in CH (Algorithm 1) and BR refers to our modified $WG_{T,BR}$ test (Algorithm 2).

4 Application to government spending multipliers

In this section, we revisit the question whether government spending is more effective in recessions, and address it as in RZ, using exactly the same data and model specifications, except that we test and estimate an unknown threshold rather than imposing it. For simplicity, we first focus on the instantaneous government

spending multiplier $\theta_{g,i}$ ($i = 1, 2$), estimated similarly to RZ from:

$$y_t = (\theta_{g,1} g_t + z_{1,t}^\top \theta_{z,1}) \mathbf{1}[q_t \leq \gamma^0] + (\theta_{g,2} g_t + z_{1,t}^\top \theta_{z,2}) \mathbf{1}[q_t > \gamma^0] + \epsilon_t \quad (10)$$

$$g_t = \Pi_1^\top z_t \mathbf{1}[q_t \leq \rho^0] + \Pi_2^\top z_t \mathbf{1}[q_t > \rho^0] + v_t \quad (11)$$

where y_t is real GDP divided by trend GDP, g_t is real government spending divided by trend GDP – which is endogenous and instrumented by military spending news m_t – and the threshold variable is q_t , the first lag of the unemployment rate. The exogenous regressors z_{1t} are also included in z_t and contain an intercept and four lags of g_t , y_t , m_t . Thus, $z_t = [z_{1t}^\top, m_t^\top]^\top$.

The data is from the RZ replication package.¹³ For details on the data construction, instrument validity, or interpretation of $\theta_{g,i}$ ($i = 1, 2$) as cumulative spending multipliers, we refer the interested reader to RZ.

Letting $\theta_i^0 = (\theta_{g,i}, \theta_{z,i}^\top)^\top$ and $w_t = (g_t, z_{1,t}^\top)^\top$, the RZ estimators of θ_i^0 are exactly the just-identified GMM (or instrumental variables, IV henceforth) estimators $\hat{\theta}_{i\gamma,(1)}$ defined in Section 2.1, but evaluated in RZ at $\gamma = 6.5$ (and ignoring the first stage which is irrelevant for conventional IV estimators).¹⁴ The threshold $\gamma = 6.5$ is chosen by RZ as in Owyang et al. (2013), based on the US Federal Reserve use of this threshold in its policy announcement; RZ also do a robustness check with a threshold of 8.0. Since it is unclear why 6.5 or 8.0 would be the threshold that defines recessions versus expansions, we do not assume that the threshold γ^0 is known or even that there is a threshold γ^0 ; we instead test for the presence of γ^0 first.

The 2SLS tests require first estimating ρ^0 in equation (11). Table 5 reports the multivariate threshold estimates $\hat{\rho}$ described in Section 2.1, along with the decisions of a LFS or a TFS based on the BIC3 criterion proposed in Gonzalo and Pitarakis (2002) and on the ordinary least-squares (OLS) versions of LR_T^b and W_T^b tests described in Section 2.2, which were proposed in Hansen (1996). The estimate of ρ^0 change with the cut-off considered, but there is considerable evidence of a threshold in the first stage. The maximizer of the OLS version of $LR_T(\gamma)$ is exactly $\hat{\rho}$, a consistent estimator of ρ^0 as shown in Theorem A.2. Therefore, we use a TFS with $\hat{\rho}$ in Table 5.

Table 5: Presence of Thresholds in the First Stage

Trim	$\hat{\rho}$	BIC3	LR	W
10%	3.5264	TFS	TFS	LFS
15%	3.5264	TFS	TFS	LFS
20%	3.7530	LFS	TFS	LFS
25%	4.0636	LFS	TFS	LFS

BIC3 is the BIC3 criterion in Gonzalo and Pitarakis (2002), and LR and W are the OLS bootstrap equivalents of our tests LR_T and W_T

Given $\hat{\rho}$ obtained for each cut-off, we test for an unknown threshold in equation (10). Table 6 shows that

¹³http://econweb.ucsd.edu/~vramey/research/Ramey_Zubairy_replication_codes.zip

¹⁴All numbers referring to unemployment rates, such as 6.5, should be interpreted as percentages: 6.5%.

Table 6: Presence of Thresholds in the Equation of Interest

Trim	TFS $\hat{\rho}$	$\hat{\gamma}$	GMM Tests					
			$WG_{T,CH}$	5%CV	Reject	$WG_{T,BR}$	5%CV	Reject
10%	3.5264	11.9660	178.296	1213.488	No	25.258	27.007	No
15%	3.5264	10.7000	66.523	1097.046	No	25.258	27.327	No
20%	3.7530	9.3443	66.523	809.764	No	25.258	26.631	No
25%	4.0636	8.3363	66.523	250.830	No	22.787	27.133	No

Trim	TFS $\hat{\rho}$	$\hat{\gamma}$	2SLS Tests					
			LR_T	5%CV	Reject	W_T	5%CV	Reject
10%	3.5264	11.9660	93.526	74.011	Yes	29.335	29.348	No
15%	3.5264	10.7000	78.158	56.963	Yes	27.698	29.348	No
20%	3.7530	9.3443	75.332	52.066	Yes	27.505	29.197	No
25%	4.0636	8.3363	65.719	50.498	Yes	24.365	29.137	No

“5%CV” display the bootstrap 5% critical values, “Reject” indicates whether the null of no threshold in (8) is rejected. For all specifications, a TFS is used with $\hat{\rho}$ obtained with the same cut-offs in the first stage as the column “Trim” indicates.

the LR test rejects the null. The 2SLS Wald test and our modified GMM Wald-test do not reject (but their values are relatively close to the critical values at certain cut-offs). From Figure 1, it is evident that the sequence of all our test statistics are relatively flat for all values of γ . The CH test also never rejects the null, but its sequence is not flat: its value is relatively large at 10% trimming, and its critical values are very large at all trimming levels. This is in line with our simulations, which indicated that the tests are undersized at 500 observations, the number of observations in our sample. Its erratic behavior near the sample edges was further illustrated in Figure 1.

Because Equations (10)–(11) control for several lags – in line with the RZ specification – we choose the 25% cut-off results with $\hat{\rho} = 4.0636$ and $\hat{\gamma} = 8.3363$, where the latter is the 2SLS threshold estimate proposed in CH (or, equivalently, the implicit maximizer of the $LR_T(\gamma)$ quantity in this paper).¹⁵

We could conclude based on Figure 1 and Table 6 that there is little evidence that the instantaneous multipliers are different in recessions and expansions. In what follows, we also show that there is little evidence that the multipliers at other horizons than zero are different. To that end, as in RZ, we compute the cumulative government spending multipliers $\theta_{g,i}^h (i = 1, 2)$ at horizon $h = 1, \dots, H$ from the IV regression:

$$\begin{aligned} \sum_{h=0}^H y_{t+h}^h &= (\theta_{g,1}^h \sum_{h=0}^H g_{t+h} + z_{1,t}^\top \theta_{z,1}^h) \mathbf{1}[q_t \leq \hat{\gamma}] \\ &+ (\theta_{g,2}^h \sum_{h=0}^H g_{t+h} + z_{1,t}^\top \theta_{z,2}^h) \mathbf{1}[q_t > \hat{\gamma}] + \epsilon_t, \end{aligned}$$

where $\sum_{h=0}^H g_{t+h}$ is instrumented by m_t .¹⁶

Tables 7-9 show the RZ multipliers (using 6.5 and 8 - robustness check in RZ - as thresholds), and our

¹⁵The confidence sets for both these thresholds obtained by inverting the likelihood ratio tests in Hansen (2000) and CH, or by simulating the asymptotic distribution in CH, are very tight when using the default nonparametric kernel. However, since both estimators are close to the 25% cut-off, and increase ($\hat{\gamma}$) or decrease ($\hat{\rho}$) when decreasing the cut-offs used, we can only interpret these estimators as close to the lower bounds of the true threshold values that are identified in the sample.

¹⁶It is unclear how to use the TFS specification (11) to obtain cumulative government spending multipliers at $h > 0$, because of the misalignment between the first and the second stage threshold, and we leave this to future research.

multipliers for fifteen quarters ahead, calculated exactly as in RZ but with $\hat{\gamma} = 8.3383$. We also report classical heteroskedasticity and autocorrelation (HAC) robust standard errors, weak instrument HAC robust confidence sets, and classical and weak-instrument HAC-robust tests for the difference in multipliers at the imposed thresholds. These tables show that in all cases, there is no evidence that government spending multipliers are different in recessions, once the possibility of weak instruments is taken into account at all horizons.

Table 7: IV Multipliers with RZ threshold

h	State 1, $q_t \leq 6.5$: 319 obs.				State 2, $q_t > 6.5$: 181 obs.				p -val.	AR p -val.
	Mult.	s.e.	AR LB	AR UB	Mult.	s.e.	AR LB	AR UB		
0	1.24	0.45	-0.51	2.99	-0.61	0.98	-4.43	3.22	0.04	0.22
1	1.11	0.29	-0.02	2.24	-1.92	1.54	-7.95	4.10	0.04	0.24
2	0.89	0.19	0.13	1.64	-0.17	0.25	-1.16	0.81	0.00	0.24
3	0.71	0.14	0.15	1.28	0.22	0.16	-0.42	0.87	0.01	0.25
4	0.64	0.12	0.17	1.12	0.46	0.14	-0.09	1.01	0.26	0.39
5	0.63	0.10	0.24	1.03	0.54	0.12	0.08	1.00	0.52	0.57
6	0.62	0.09	0.26	0.99	0.59	0.11	0.17	1.01	0.81	0.82
7	0.59	0.09	0.24	0.95	0.60	0.10	0.23	0.97	0.95	0.95
8	0.59	0.09	0.23	0.95	0.62	0.09	0.29	0.95	0.82	0.82
9	0.62	0.10	0.25	1.00	0.63	0.08	0.33	0.92	0.97	0.97
10	0.66	0.10	0.27	1.05	0.64	0.07	0.37	0.91	0.87	0.87
11	0.68	0.10	0.28	1.08	0.64	0.07	0.39	0.90	0.79	0.80
12	0.68	0.11	0.27	1.10	0.65	0.06	0.41	0.90	0.81	0.82
13	0.68	0.11	0.26	1.11	0.67	0.06	0.44	0.89	0.89	0.90
14	0.68	0.11	0.24	1.13	0.68	0.05	0.47	0.89	0.99	0.99
15	0.67	0.12	0.19	1.15	0.68	0.05	0.48	0.88	0.92	0.92

"Mult." indicates the IV estimates at each horizon, "obs." the number of observations, and "s.e." the Newey-West HAC standard errors using the Bartlett kernel and the data-dependent bandwidth. "AR LB (AR UB)" refer to 95% Anderson-Rubin confidence lower (upper) bounds. " p -val." indicate classical p -values for the t -test of no difference between the multipliers, and "AR p -val." indicate Anderson-Rubin p -values for the same test. All the results are computed with RZ's replication package code.

Table 8: IV Multipliers with a threshold equal to 8 (robustness check in RZ)

h	State 1, $q_t \leq 8$: 396 obs.				State 2, $q_t > 8$: 104 obs.				p -val.	AR p -val.
	Mult.	s.e.	AR LB	AR UB	Mult.	s.e.	AR CI LB	AR CI UB		
0	1.29	0.37	-0.17	2.74	-0.44	1.11	-4.79	3.92	0.10	0.26
1	1.13	0.24	0.20	2.05	-1.05	1.25	-5.94	3.83	0.08	0.30
2	0.92	0.15	0.31	1.52	-0.29	0.41	-1.89	1.31	0.01	0.31
3	0.73	0.12	0.27	1.19	0.16	0.25	-0.82	1.13	0.05	0.36
4	0.66	0.10	0.26	1.07	0.48	0.23	-0.42	1.38	0.47	0.59
5	0.65	0.09	0.28	1.02	0.65	0.21	-0.17	1.47	0.99	0.99
6	0.63	0.09	0.26	1.00	0.76	0.21	-0.06	1.57	0.60	0.58
7	0.60	0.09	0.24	0.97	0.80	0.21	-0.01	1.60	0.40	0.42
8	0.60	0.09	0.24	0.95	0.79	0.18	0.08	1.50	0.35	0.38
9	0.63	0.09	0.28	0.98	0.76	0.15	0.18	1.34	0.45	0.48
10	0.66	0.09	0.30	1.01	0.75	0.13	0.25	1.24	0.58	0.58
11	0.68	0.09	0.31	1.04	0.72	0.11	0.28	1.16	0.77	0.77
12	0.68	0.09	0.31	1.05	0.71	0.11	0.29	1.12	0.85	0.84
13	0.67	0.10	0.29	1.05	0.72	0.10	0.33	1.12	0.72	0.72
14	0.66	0.10	0.26	1.07	0.75	0.10	0.36	1.13	0.55	0.55
15	0.65	0.12	0.20	1.10	0.76	0.10	0.37	1.15	0.45	0.43

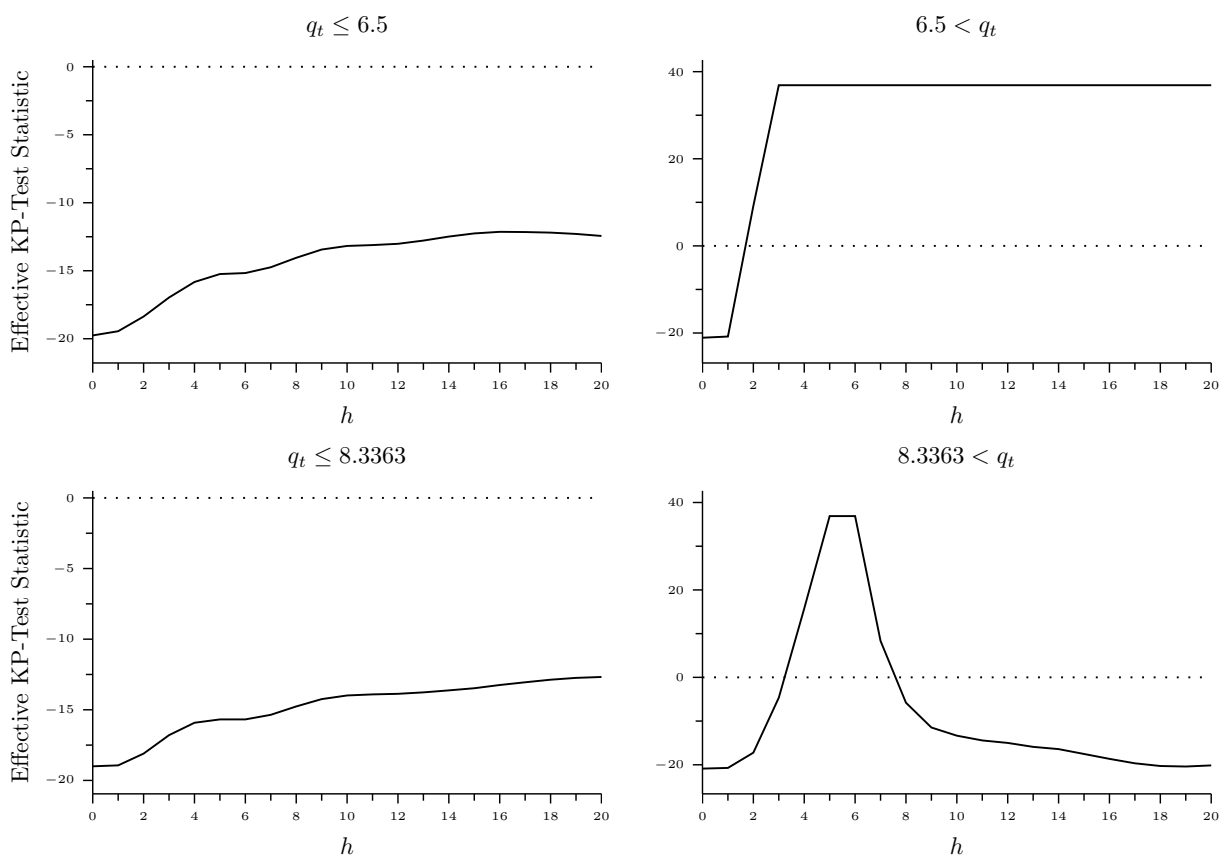
See Table 7 notes.

Table 9: IV Multipliers with our threshold

h	State 1, $q_t \leq 8.3363$: 410 obs.				State 2, $q_t > 8.3363$: 90 obs.				p -val.	AR p -val.
	Mult.	s.e.	AR LB	AR UB	Mult.	s.e.	AR LB	AR UB		
0	1.30	0.38	-0.20	2.80	-0.93	1.36	-6.27	4.41	0.05	0.19
1	1.14	0.25	0.17	2.11	-1.68	1.37	-7.05	3.68	0.03	0.23
2	0.93	0.16	0.28	1.57	-0.55	0.48	-2.42	1.33	0.00	0.25
3	0.74	0.13	0.25	1.23	-0.02	0.20	-0.80	0.76	0.00	0.26
4	0.67	0.11	0.24	1.09	0.32	0.17	-0.36	1.00	0.07	0.31
5	0.65	0.10	0.28	1.03	0.52	0.18	-0.19	1.22	0.45	0.52
6	0.63	0.10	0.26	1.01	0.62	0.19	-0.14	1.38	0.94	0.94
7	0.60	0.09	0.24	0.96	0.66	0.20	-0.13	1.45	0.77	0.78
8	0.60	0.09	0.25	0.94	0.66	0.18	-0.04	1.37	0.71	0.73
9	0.62	0.09	0.29	0.96	0.66	0.15	0.07	1.25	0.84	0.85
10	0.65	0.09	0.31	1.01	0.66	0.13	0.14	1.18	0.96	0.96
11	0.67	0.09	0.32	1.02	0.65	0.12	0.17	1.13	0.87	0.87
12	0.67	0.09	0.32	1.03	0.64	0.12	0.17	1.11	0.85	0.85
13	0.66	0.09	0.30	1.03	0.67	0.11	0.22	1.11	0.98	0.98
14	0.65	0.10	0.27	1.03	0.70	0.11	0.28	1.11	0.75	0.76
15	0.64	0.11	0.22	1.05	0.71	0.10	0.31	1.18	0.58	0.60

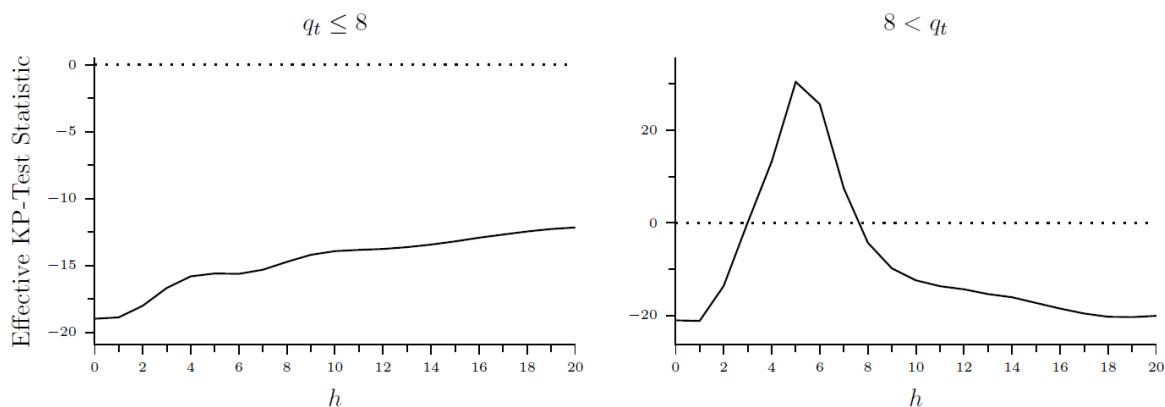
See Table 7 notes.

Figure 3: Effective F-Statistic



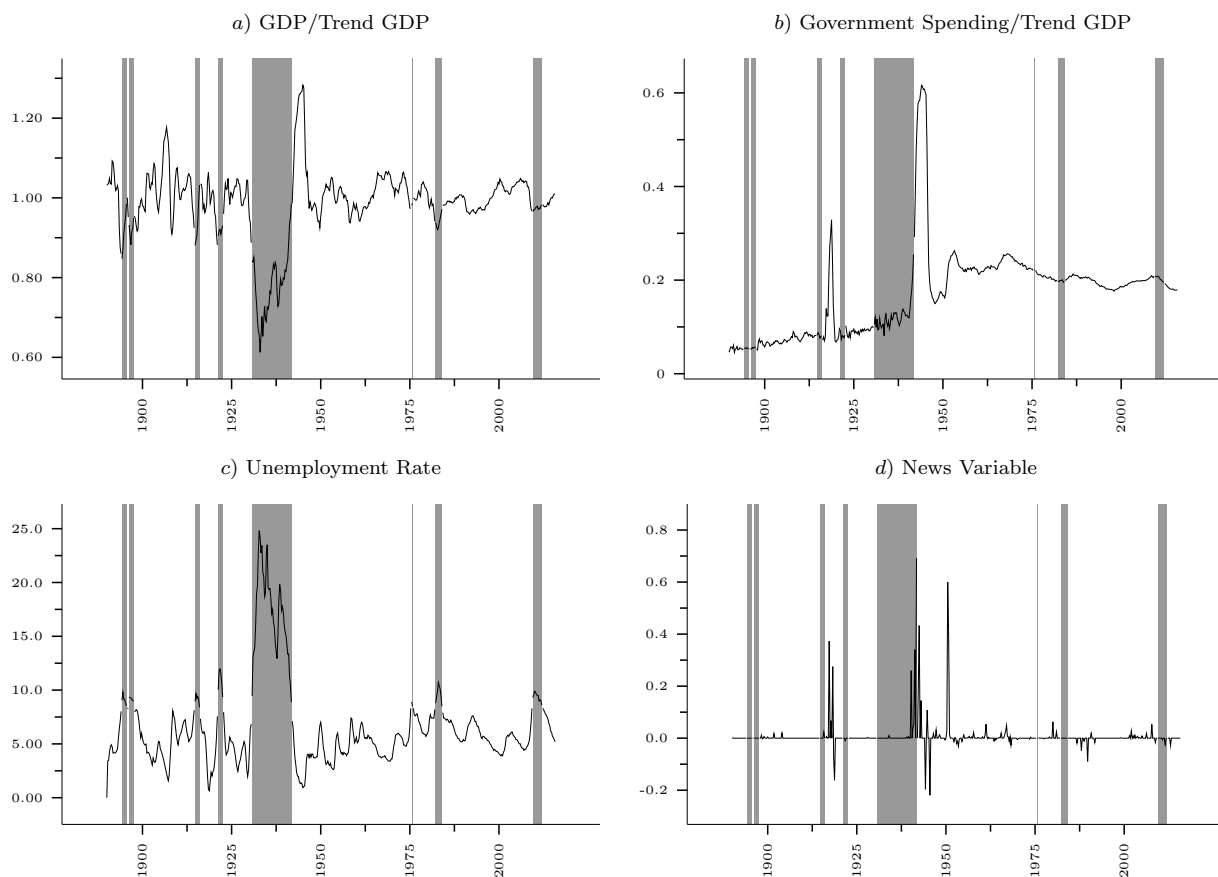
The effective F- or KP-statistic is calculated as the Kleibergen and Paap (2006) Wald rank test in Stata, equal to the (HAC-robust) F-test on the first-stage coefficient on m_t in each regime, minus the Montiel-Olea and Pflueger (2013) critical value 23.11, the 5% critical value that tolerates 10% relative bias of 2SLS compared to OLS. The values are capped just below 40. The first two plots reproduce Figure 10 in RZ.

Figure 4: Effective F-Statistic



See Figure 3 notes.

Figure 5: Data plots with with shaded areas $q_t > 8.3363$



We therefore assess the possibility of weak instruments at various horizons in Figures 3 and 4, plotting the effective F-statistic for the null hypothesis of weak instruments in each regime across horizons. These figures show evidence of weak instruments in both regimes at short horizons, for all thresholds. This also holds for the effective F-statistics minus their critical value for our TFS specification with $\hat{\rho} = 4.0636$: they are equal to approximately -19 for $q_t \leq 4.0636$ (101 observations), and -17.5 for $q_t > 4.0636$ (399 observations), so

well below zero. Therefore, the weak instrument robust p-values should be used, even for Table 9, at shorter horizons. Hence, once weak instruments are accounted for, there is no evidence that government spending multipliers are different in recessions, both in our paper and in RZ.

What we do learn from the analysis is that military spending news becomes a weaker instrument for longer horizons when the threshold increases from 6.5 to 8.0 or to 8.3363, and therefore that the instrument relevance is not robust to the threshold used. This is also indicated in Figure 5, which shows that, except for the World War II period, the news variable does not exhibit much variation when the unemployment rate is above 8.3363. This suggests that the RZ military news instrument is more informative for intermediate values of unemployment, so for "normal" recessions rather than "deep" recessions.

5 Conclusions

In this paper we proposed two adjustments to the GMM Wald test of Caner and Hansen (2004), and two new 2SLS test statistics for threshold detection in linear models with endogenous regressors and exogenous thresholds. We derived the asymptotic validity of their null bootstrap equivalents, and showed through simulations and an application that these tests have better finite sample properties than the test proposed in Caner and Hansen (2004).

Rothfelder and Boldea (2016) show in their Theorem 1 that under conditional homoskedasticity and one endogenous regressor, the 2SLS estimators with a linear first stage or a threshold first stage can be more efficient than the GMM estimators that ignore this information. It would be interesting to assess when this efficiency carries over to more general settings, and whether there exists an optimal GMM estimator that uses similar information from the first stage as the 2SLS estimators.

References

- ALLOZA, M. (2022): "Is fiscal policy more effective during recessions?" *International Economic Review*, forthcoming.
- ALPANDA, S., E. GRANZIERA, AND S. ZUBAIRY. (2021): "State dependence of monetary policy across business, credit and interest rate cycles," *European Economic Review*, 140, 103936.
- ANTOINE, B. AND O. BOLDEA (2015): "Inference in linear models with structural changes and mixed identification strength," *SFU Working Paper 15-5*, <https://ideas.repec.org/p/sfu/sfudps/dp15-05.html>.
- (2018): "Efficient estimation with time-varying information and the New Keynesian Phillips curve," *Journal of Econometrics*, 204, 268–300.
- AUERBACH, A. AND Y. GORODNICHENKO (2013): "Fiscal multipliers in recession and expansion," in *Fiscal Policy After the Financial Crisis*, ed. by A. Alesian and F. Giavazzi, University of Chicago Press, 63–98.

- BAI, J. AND P. PERRON (1998): “Estimating and testing linear models with multiple structural changes,” *Econometrica*, 66, 47–78.
- BARNICHON, R. AND C. MATTHES (2018): “Functional approximation of impulse responses,” *Journal of Monetary Economics*, 99, 41–55.
- BOLDEA, O., A. CORNEA-MADEIRA, AND A. R. HALL (2019): “Bootstrapping structural change tests,” *Journal of Econometrics*, 213, 357–397.
- BRUNS, M. AND M. PIFFER (2021): “Monetary policy shocks over the business cycle: Extending the smooth transition framework,” *Working Paper No. 2021-07 School of Economics, University of East Anglia*, <https://ueaeco.github.io/working-papers/papers/ueaeco/UEA-ECO-21-07.pdf>.
- CAGGIANO, G., E. CASTELNUOVO, AND V. COLOMBO (2015): “Estimating fiscal multipliers: News from a nonlinear world,” *The Economic Journal*, 125, 746–776.
- CANER, M. AND B. E. HANSEN (2001): “Threshold autoregression with a unit root,” *Econometrica*, 69, 1555–1596.
- (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.
- CHRISTOPOULOS, D., P. MCADAM, AND E. TZAVALIS (2021): “Dealing with endogeneity in threshold models using copulas,” *Journal of Business and Economic Statistics*, 39, 166–178.
- CUGNASCA, A. AND P. ROTHER (2015): “Fiscal multipliers during consolidation: evidence from the European Union,” *ECB Working Paper No. 1863*.
- 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.
- 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.
- (2017): “Regression kink with an unknown threshold,” *Journal of Business and Economic Statistics*, 35, 228–240.
- JO, Y. AND S. ZUBAIRY (2022): “State dependent government spending multipliers: Downward nominal wage rigidity and sources of business cycle fluctuations,” *Working Paper*, https://drive.google.com/file/d/1BIh-QGCIvR7voL_HVUAERXrN8vZ1kGe9/view.
- JORDÁ, O., M. SCHULARICK, AND A. M. TAYLOR (2020): “The effects of quasi-random monetary experiments,” *Journal of Monetary Economics*, 112, 22–40.
- KLEPACZ, M. (2021): “Price setting and volatility: Evidence from oil price volatility shocks,” *International Finance Discussion Paper 1316*, <https://www.federalreserve.gov/econres/ifdp/files/ifdp1316.pdf>.
- KOURTELLOS, A., T. STENGOS, AND C. TAN (2015): “Structural threshold regression,” *Econometric Theory*, 1–34.
- 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.
- LIAO, Q., P. C. B. PHILLIPS, AND P. YU (2019): “Inference and specification testing in threshold regression with endogeneity,” *Cowles Foundation Discussion Paper No. 2209*, http://web.hku.hk/~pingyu/WorkingPapers/EndoTR_Infer.pdf.
- MAGNUSSON, L. AND S. MAVROEIDIS (2014): “Identification using stability restrictions,” *Econometrica*, 82, 1799–1851.
- MAMMEN, E. (1993): “Bootstrap and wild bootstrap for high-dimensional linear models,” *Annals of Statistics*, 21, 255–285.
- OWYANG, M., V. RAMEY, AND S. ZUBAIRY (2013): “Are government spending multipliers greater during periods of slack? Evidence from twentieth-century historical data,” *American Economic Review: Papers & Proceedings*, 103, 129–134.
- RAMEY, V. A. AND S. ZUBAIRY (2018): “Government spending multipliers in good times and in bad: Evidence from U.S. historical data,” *Journal of Political Economy*, 126, 850–901.

- ROTHFELDER, M. AND O. BOLDEA (2016): “Testing for a threshold in models with endogenous regressors,” *CentER Working Paper 2016-029*, https://pure.uvt.nl/ws/portalfiles/portal/12620378/2016_029.pdf.
- SANTORO, E., I. PETRELLA, D. PFAJFAR, AND E. GAFFEO (2014): “Loss aversion and the asymmetric transmission of monetary policy,” *Journal of Monetary Economics*, 68, 19–36.
- SEO, M. AND O. LINTON (2007): “A smoothed least squares estimator for threshold regression models,” *Journal of Econometrics*, 141, 704–735.
- SEO, M. H. AND C. SHIN (2016): “Dynamic panels with threshold effect and endogeneity,” *Journal of Econometrics*, 195, 169–186.
- TONG, H. (1990): *Nonlinear Time Series - A Dynamical System Approach*, Oxford: Clarendon Press.
- YU, P. AND P. C. B. PHILLIPS (2018): “Threshold Regression with Endogeneity,” *Journal of Econometrics*, 203, 50–68.

A Online Supplement: Proofs

A.1 General Notation and Results

Before proving our results, we introduce some more required notation and provide additional Lemmas with helpful results.

General notation. Let $\|\cdot\|$ be the Euclidean norm for vectors, respectively the Frobenius norm for matrices: $\|P\| = \sqrt{\text{tr}(P^\top P)}$. Also, for a (matrix valued) random variable P , let $\|P\|_\alpha = (E\|P\|^\alpha)^{1/\alpha}$, for any $\alpha > 0$. Let I_m the $m \times m$ -identity matrix, $0_{a \times b}$ a $a \times b$ vector of zeros (we use this notation only when the dimension is not obvious from the derivations, else we use 0), and let K denote a generic constant.

Let P_t be a matrix of random variables. We define $P_{1\gamma} = E[P_t \mathbf{1}[q_t \leq \gamma]]$, $P = \lim_{\gamma \rightarrow \infty} P_{1\gamma} = E[P_t]$, and $P_{2\gamma} = P - P_{1\gamma}$. For example, since $M_{1\gamma} = E[z_t z_t^\top \mathbf{1}[q_t \leq \gamma]]$, we have $M = E[z_t z_t^\top]$ and $M_{2\gamma} = E[z_t z_t^\top \mathbf{1}[q_t > \gamma]]$. Let their sample equivalents (replacing expectations by averages and unobserved quantities with estimates) be denoted by hats, for example, for $M_{1\gamma}$, its sample equivalent is $\hat{M}_{1\gamma} = T^{-1} \sum_{1\gamma} z_t z_t^\top$, for M it is $\hat{M} = T^{-1} \sum_{t=1}^T z_t z_t^\top$, for $H_{\epsilon,1\gamma} = E[z_t z_t^\top \epsilon_t^2 \mathbf{1}[q_t \leq \gamma]]$, it is $\hat{H}_{\epsilon,1\gamma} = T^{-1} \sum_{1\gamma} z_t z_t^\top \hat{\epsilon}_t^2$, where $\hat{\epsilon}_t$ is an estimate of the residual ϵ_t , and so on. When the notation $\hat{P}_{i\gamma}$ does not conform with this definition, it is specifically indicated in the text.

Throughout the text, quantities of the form $v_t v_t^\top \otimes z_t z_t^\top$ should be read as $(v_t v_t^\top) \otimes (z_t z_t^\top)$. Let $\mathcal{G}_{1\gamma} = (\mathcal{G}_{\epsilon,1\gamma}^\top, \mathcal{G}_{u,1\gamma}^\top)^\top$ be a $q(p_1 + 1)$ vector of zero mean Gaussian processes with covariance matrix

$$H_{1\gamma} = E[(v_t v_t^\top \otimes z_t z_t^\top) \mathbf{1}[q_t \leq \gamma]] = \begin{pmatrix} E[\epsilon_t^2 z_t z_t^\top \mathbf{1}[q_t \leq \gamma]] & E[\epsilon_t u_t^\top \otimes z_t z_t^\top \mathbf{1}[q_t \leq \gamma]] \\ E[\epsilon_t u_t \otimes z_t z_t^\top \mathbf{1}[q_t \leq \gamma]] & E[u_t u_t^\top \otimes z_t z_t^\top \mathbf{1}[q_t \leq \gamma]] \end{pmatrix} = \begin{pmatrix} H_{\epsilon,1\gamma} & H_{\epsilon u,1\gamma}^\top \\ H_{\epsilon u,1\gamma} & H_{u,1\gamma} \end{pmatrix}$$

and covariance function $E[\mathcal{G}_{1\gamma_1} \mathcal{G}_{1\gamma_2}^\top] = E[(v_t v_t^\top \otimes z_t z_t^\top) \mathbf{1}[q_t \leq (\gamma_1 \wedge \gamma_2)]]$. Hence, $\mathcal{G}_{\epsilon,1\gamma}$, respectively $\mathcal{G}_{u,1\gamma}$ corresponds to the parts of $\mathcal{G}_{1\gamma}$ induced by ϵ_t , respectively u_t . Moreover, $\mathcal{G} = \lim_{\gamma \rightarrow \infty} \mathcal{G}_{1\gamma}$ and $\mathcal{G}_{2\gamma} = \mathcal{G} - \mathcal{G}_{1\gamma}$.

Let $\sigma^2 = E[(\epsilon_t + u_t^\top \theta_x^0)^2]$, $\tilde{\theta}^0 = \text{vec}(1, \theta_x^0)$ and $\check{\theta}^0 = \text{vec}(0, \theta_x^0)$. Define $\tilde{\epsilon}_t = \epsilon_t + (x_t - \hat{x}_t)^\top \theta_x^0$, where \hat{x}_t is obtained either with a LFS (linear first stage) or a TFS (threshold first stage) specification, depending on the context. Define $\hat{C}_{i\gamma} = T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top$, and $\hat{C} = \hat{C}_{1\gamma} + \hat{C}_{2\gamma}$, both for a LFS and a TFS.

Let $A^0 = [\Pi^0, S^\top]^\top$ be the augmented matrix of the LFS slope parameters, where $S = [I_{p_2}, 0_{p_2 \times (q-p_2)}]$. Hence, $z_{1t} = S z_t$ and $w_t = A^0 z_t + (u_t^\top, 0_{1 \times q_1})^\top = A^0 z_t + \bar{u}_t$. Similarly, let $A_1^0 = [\Pi_1^0, S^\top]^\top$ and $A_2^0 = [\Pi_2^0, S^\top]^\top$ be the augmented matrix of TFS parameters such that $w_t = (A_1^0 \mathbf{1}[q_t \leq \rho^0] + A_2^0 \mathbf{1}[q_t > \rho^0]) z_t + \bar{u}_t$.

All convergence results, if not stated otherwise, are uniformly in $\gamma \in \Gamma$, and all $o_p(1)$ terms are uniform in γ . “ \Rightarrow ” stands for weak convergence in Skorokhod metric, “ $\xrightarrow{\text{d}^b}$ ” for weak convergence in Skorokhod metric under the bootstrap measure, and “ $\xrightarrow{\text{p}^b}$ ” for weak convergence in probability under the bootstrap measure.

Lemma A.1. [ULLN] If (i) $\{a_t\}$ and $\{q_t\}$ are scalar strictly stationary and ρ -mixing series, with mixing coefficient $\rho(m) = \mathcal{O}(m^{-A})$ for some $A > \frac{a}{a-1}$ and $1 < a \leq r$; (ii) $\|a_t\|_r < \infty$ for some $r > 1$; (iii) q_t has a contin-

uous distribution, with pdf $f(\cdot)$ bounded: $\sup_{x \in \Gamma} |f(x)| < \infty$, then $\sup_{\gamma \in \Gamma} \left| T^{-1} \sum_{1\gamma} a_t - E[a_t \mathbf{1}[q_t \leq \gamma]] \right| \xrightarrow{P} 0$.

Proof of Lemma A.1. This uniform law of large numbers (ULLN) can be proven using the same steps as the proof of Lemma 1 in Hansen (1996), with a slight modification as we do not assume that a_t has a continuous and bounded pdf. First, note that ρ -mixing implies ergodicity. Second, set in the proof of their Lemma 1 $w_t = (a_t, q_t)$, $\phi(w_t) = a_t$, and $\{w_t \leq \gamma\} = \mathbf{1}[q_t \leq \gamma]$. Follow the steps in Hansen (1996), until their equation (15). Then note that $\|a_t\|_r = K < \infty$ by Assumption 1(c), and for some $\epsilon > 0$, set $K_\epsilon = (2K/\epsilon)^{r/(r-1)}$. Since q_t is assumed to have a continuous and bounded pdf, there exists an ϵ such that $\int_{\gamma_k}^{\gamma_{k+1}} f(x) dx \leq 1/K_\epsilon = (\epsilon/(2K))^{r/(1-r)}$. Therefore, replace equation (15) by the inequality below (derived using Hölder's inequality with $p = r$ and $q = r/(r-1)$):

$$\begin{aligned} E[a_t \mathbf{1}[\gamma_k < q_t \leq \gamma_{k+1}]] &\leq \|a_t\|_r \|\mathbf{1}[\gamma_k < q_t \leq \gamma_{k+1}]\|_{r/(r-1)} \\ &= K \left(\int_{\gamma_k}^{\gamma_{k+1}} f(x) dx \right)^{(r-1)/r} \leq K(\epsilon/(2K)) = \epsilon/2. \end{aligned}$$

The rest of the proof is as in Hansen (1996), where only the last equation in their proof should be replaced by:

$$E|f_{\epsilon,k}^u(w_t) - f_{\epsilon,k}^l(w_t)| \leq 2E[a_t \mathbf{1}[\gamma_k < q_t \leq \gamma_{k+1}]] \leq \epsilon.$$

■

Lemma A.2. [FCLT] If the assumptions in Lemma A.1 hold but with a_t being a vector of m.d.s, and additionally (i) $\|a_t\|_{2r} < \infty$ for some $r > 1$; (ii) $E[a_t a_t^\top \mathbf{1}[q_t \leq \gamma]] = F_\gamma$, a p.d. matrix of constants, (iii) $\inf_{\gamma \in \Gamma} \det F_\gamma > 0$, then: $T^{-1/2} \sum_{1\gamma} a_t \Rightarrow \mathcal{J}_{1\gamma}$, a vector of Gaussian processes with covariance function $E[a_t a_t^\top \mathbf{1}[q_t \leq (\gamma_1 \wedge \gamma_2)]]$.

Proof of Lemma A.2. This functional central limit theorem (FCLT) follows directly from Theorem 3 and then Theorem 1 in Hansen (1996). Note that only $\|a_t\|_{2r} < \infty$ is needed, as evident from replacing $x_t \epsilon_t$ with a_t in the first two equations of the proof of Theorem 3 in Hansen (1996). ■

Note that, Lemmas A.1 and A.2 imply that $\sup_{\gamma \in \Gamma} |T^{-1} \sum_{2\gamma} a_t - E[a_t \mathbf{1}[q_t > \gamma]]| \xrightarrow{P} 0$, and $T^{-1/2} \sum_{2\gamma} a_t \Rightarrow \mathcal{J}_{2\gamma}$ since $\sum_{2\gamma}(\cdot) = \sum_{t=1}^T(\cdot) - \sum_{1\gamma}(\cdot)$ by definition.

Lemma A.3. Suppose Assumption 1 holds. Then: (i) $T^{-1} \sum_{1\gamma} z_t z_t^\top \xrightarrow{P} M_{1\gamma}$ and $T^{-1} \sum_{2\gamma} z_t z_t^\top \xrightarrow{P} M_{2\gamma}$, and (ii) $T^{-1/2} \sum_{i\gamma} v_t \otimes z_t \Rightarrow \mathcal{G}_{i\gamma}$.

If, additionally, Assumption 2 holds, then, under the bootstrap measure: (iii) $T^{-1} \sum_{i\gamma} z_t z_t^\top \eta_t = o_p^b(1)$ and (iv) $T^{-1/2} \sum_{i\gamma} v_t \eta_t \otimes z_t \xrightarrow{q_p^b} \mathcal{G}_{i\gamma}$.

Proof of Lemma A.3. Part (i) follows from Assumptions 1 (b), (c) and (e), ensuring that the assumptions of Lemma A.1 are satisfied for elements of $z_t z_t^\top$. Hence, $T^{-1} \sum_{i\gamma} z_t z_t^\top \xrightarrow{P} M_{i\gamma}$.

Part (ii) follows from Assumptions 1 (a)–(f), ensuring that the conditions of Lemma A.2 are satisfied since $\|v_t \otimes z_t\|_{2r} \leq \|v_t\|_{4r} \|z_t\|_{4r} < \infty$. So, $T^{-1/2} \sum_{1\gamma} v_t \otimes z_t \Rightarrow \mathcal{G}_{1\gamma}$.

We will show Parts (iii) and (iv) only for $i = 1$; for $i = 2$, the proofs follow the same arguments by noting that, for example, $\sum_{2\gamma} z_t z_t^\top \eta_t = \sum_{t=1}^T z_t z_t^\top \eta_t - \sum_{1\gamma} z_t z_t^\top \eta_t$. To prove Part (iii), note that $E^b[T^{-1} \sum_{1\gamma} z_t z_t^\top \eta_t] = 0$. Hence, Chebyshev's and Minkowski's inequalities imply $P^b(\|T^{-1} \sum_{1\gamma} z_t z_t^\top \eta_t\| > K) \leq K^{-2} T^{-1} (T^{-1} \sum_{1\gamma} \|z_t\|^4 + (E^b|\eta_t|)^2 T^{-1} \sum_{1\gamma, t \neq s} \|z_t\|^2 \|z_s\|^2) = K^{-2} T^{-1} (O_p(1) + O(1)O_p(1)) = o_p(1)$, where the second to last equality follows from Assumptions 1 (b), (c), (e), 2 (a) and Lemma A.1. So, $T^{-1} \sum_{1\gamma} z_t z_t^\top \eta_t = o_p^b(1)$.

Finally, to prove Part (iv) for $i = 1$, we apply Lemma A.2 and verify that $\text{Var}^b(T^{-1/2} \sum_{1\gamma} v_t \eta_t \otimes z_t) \xrightarrow{P} H_{1\gamma}$. First, $E^b(v_t \eta_t \otimes z_t) = 0$. Conditions (ii) and (iii) in Lemma A.2 are satisfied by Assumption 1 (b), (c), (e) and Assumption 2 (a). Condition (i) is satisfied because $(E^b\|v_t \eta_t \otimes z_t\|^r)^{1/(r)} = (E^b|\eta_t|^r)^{1/(r)} \|v_t\| \|z_t\| = O_p(1)$ by Assumptions 1 (c) and 2. Finally, $\text{Var}^b(T^{-1/2} \sum_{1\gamma} v_t \eta_t \otimes z_t) = T^{-1} \sum_{1\gamma} v_t v_t^\top \otimes z_t z_t^\top \xrightarrow{P} H_{1\gamma}$, where the last statement follows by applying Lemma A.1. So, by Lemma A.2, $T^{-1/2} \sum_{1\gamma} v_t \eta_t \otimes z_t \xrightarrow{d^b} \mathcal{G}_{1\gamma}$. \blacksquare

A.2 Proofs of GMM Results

A.2.1 Asymptotic Distribution of GMM

In order to simplify exposition for these proofs, define the quantity

$$\tilde{A} = \begin{cases} A^0 & \text{if LFS} \\ A_1^0 \mathbf{1}[q_t \leq \rho^0] + A_2^0 \mathbf{1}[q_t > \rho^0] & \text{if TFS} \end{cases}.$$

Hence, $w_t = \tilde{A} z_t + \bar{u}_t$.

Proof of Theorem 1. First, we show that $\hat{N}_{i\gamma} \xrightarrow{P} N_{i\gamma}$. Note, $\hat{N}_{i\gamma} = T^{-1} \sum_{i\gamma} w_t z_t^\top = \tilde{A} (T^{-1} \sum_{i\gamma} z_t z_t^\top) + T^{-1} \sum_{i\gamma} \bar{u}_t z_t^\top = \tilde{A} M_{i\gamma} + o_p(1) = N_{i\gamma}$, where the second to last equality follows from Lemma A.3 (i)–(ii) and the last equality from Assumption 1 (a) implying that $N_{1\gamma} = \mathbb{E}[w_t z_t^\top \mathbf{1}\{q_t \leq \gamma\}] = \mathbb{E}[(A^0 z_t + \bar{u}_t) z_t^\top \mathbf{1}\{q_t \leq \gamma\}] = A^0 \mathbb{E}[z_t z_t^\top \mathbf{1}\{q_t \leq \gamma\}] = A^0 M_{1\gamma}$ and, similarly, $N_{2\gamma} = A^0 M_{2\gamma}$ in case of a LFS. In case of a TFS, some more algebra provides the appropriate result.

Next, we show that $\hat{H}_{\epsilon, i\gamma} \xrightarrow{P} H_{\epsilon, i\gamma}$. Note that $\hat{H}_{\epsilon, i\gamma} = T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t^2$ where $\hat{\epsilon}_t = y_t - w_t^\top \hat{\theta}_{(1)}$. Hence, $\hat{H}_{\epsilon, i\gamma} = T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 - 2[T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes \epsilon_t w_t^\top][I_q \otimes (\hat{\theta}_{(1)} - \theta^0)] + [I_q \otimes (\hat{\theta}_{(1)} - \theta^0)^\top][T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top][I_q \otimes (\hat{\theta}_{(1)} - \theta^0)]$. By Assumptions 4.1 (b), (c), (e) and Lemma A.1 it holds that $T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 = H_{\epsilon, i\gamma} + o_p(1)$. Moreover, for any $r > 1$, $\|z_t z_t^\top \otimes w_t w_t^\top\|_r \leq \|I_q \otimes \tilde{A}\|^2 \cdot \|z_t z_t^\top \otimes z_t z_t^\top\|_r + 2\|I_q \otimes \tilde{A}\| \cdot \|z_t z_t^\top \otimes z_t \bar{u}_t^\top\|_r + \|z_t z_t^\top \otimes \bar{u}_t \bar{u}_t^\top\|_r$ by Minkowski's inequality and sub-multiplicativity of the Frobenius norm. Using Hölder's inequality and Assumption 4.1 (c) it follows that $\|z_t z_t^\top \otimes z_t z_t^\top\|_r \leq \|z_t\|_{4r}^4 < \infty$, $\|z_t z_t^\top \otimes z_t \bar{u}_t^\top\|_r \leq \|z_t\|_{4r}^3 \|\bar{u}_t\|_{4r} < \infty$ and $\|z_t z_t^\top \otimes \bar{u}_t \bar{u}_t^\top\|_r \leq \|z_t\|_{4r}^2 \|\bar{u}_t\|_{4r}^2 < \infty$. Additionally, $\|I_q \otimes \tilde{A}\| = O(1)$. Hence, $T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top = O_p(1)$ by Lemma A.2. By standard arguments for GMM estimators of linear models $\hat{\theta}_{(1)} - \theta^0 = O_p(T^{-1/2})$. Therefore,

$[I_q \otimes (\hat{\theta}_{(1)} - \theta)^\top][T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top][I_q \otimes (\hat{\theta}_{(1)} - \theta^0)] = O_p(T^{-1/2})O_p(1)O_p(T^{-1/2}) = o_p(1)$. Similarly, $[T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes \epsilon_t w_t^\top][I_q \otimes (\hat{\theta}_{(1)} - \theta^0)] = o_p(1)$. Hence, $\hat{H}_{\epsilon, i\gamma} = H_{\epsilon, i\gamma} + o_p(1)$.

Last, we show that $T^{1/2}(\hat{\theta}_{i\gamma} - \theta^0) \Rightarrow (N_{i\gamma} H_{\epsilon, i\gamma}^{-1} N_{i\gamma}^\top)^{-1} N_{i\gamma} H_{\epsilon, i\gamma}^{-1} \mathcal{G}_{\epsilon, i\gamma}$. Since $\hat{\theta}_{i\gamma} = \theta^0 + (\hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{-1} \hat{N}_{i\gamma}^\top)^{-1} (\hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{-1} [T^{-1} \sum_{i\gamma} z_t \epsilon_t])$, $T^{-1/2} \sum_{i\gamma} z_t \epsilon_t \Rightarrow \mathcal{G}_{\epsilon, i\gamma}$ by Lemma A.3 (ii), parts (i) and (ii) above, the continuous mapping theorem and Slutsky's theorem it follows that $T^{1/2}(\hat{\theta}_{i\gamma} - \theta^0) \Rightarrow (N_{i\gamma} H_{\epsilon, i\gamma}^{-1} N_{i\gamma}^\top)^{-1} N_{i\gamma} H_{\epsilon, i\gamma}^{-1} \mathcal{G}_{\epsilon, i\gamma}$.

Using parts (i)–(iii) and continuous mapping theorem and Slutsky's theorem concludes the proof. \blacksquare

A.2.2 Bootstrap Validity for GMM Tests

Lemma A.4. *Suppose Assumptions 1 and 2 hold. Then, under \mathbb{H}_0 and for $i = 1, 2$, (i) $T^{-1/2} \sum_{i\gamma} z_t y_t^b \xrightarrow{d_p^b} \mathcal{G}_{\epsilon, i\gamma}$, (ii) $T^{1/2} \hat{\theta}_{(1)}^b = O_p^b(1)$, and (iii) $\hat{H}_{\epsilon, i\gamma}^b \xrightarrow{p^b} H_{\epsilon, i\gamma}$*

Proof of Lemma A.4. *Part (i).* Note $T^{-1/2} \sum_{i\gamma} z_t y_t^b = T^{-1/2} \sum_{i\gamma} z_t \eta_t \epsilon_t - \left(T^{-1} \sum_{i\gamma} z_t w_t^\top \eta_t \right) \left(T^{1/2} [\hat{\theta}_{(2)} - \theta^0] \right)$. From Lemma A.3 (iii), respectively standard arguments for GMM estimators for linear models it follows that $T^{-1/2} \sum_{i\gamma} z_t \eta_t \epsilon_t \xrightarrow{d_p^b} \mathcal{G}_{\epsilon, i\gamma}$, respectively $T^{1/2} [\hat{\theta}_{(2)} - \theta^0] = O_p(1)$. Moreover, $T^{-1} \sum_{i\gamma} z_t w_t^\top \eta_t = \left(T^{-1} \sum_{i\gamma} z_t z_t^\top \eta_t \right) \tilde{A}^\top + T^{-1} \sum_{i\gamma} z_t \tilde{u}_t^\top \eta_t = o_p^b(1)O(1) + o_p^b(1) = o_p^b(1)$ by Lemma A.3 (iii)–(iv) and since $\tilde{A} = O(1)$. Thus, $T^{-1/2} \sum_{i\gamma} z_t y_t^b \xrightarrow{d_p^b} \mathcal{G}_{\epsilon, i\gamma}$ under the null hypothesis.

Part (ii). Recall that $\hat{\theta}_{(1)} = (\hat{N} \hat{M}^{-1} \hat{N}^\top)^{-1} (\hat{N} \hat{M}^{-1} \sum_{t=1}^T z_t y_t^b)$ where $\hat{M} \xrightarrow{p} M$ by Lemma A.3, $\hat{N} \xrightarrow{p} N$ as shown in the proof of Theorem 1 and $T^{-1/2} \sum_{t=1}^T z_t y_t^b \xrightarrow{d_p^b} \mathcal{G}_\epsilon$ by part (i). Hence, $T^{1/2} \hat{\theta}_{(1)}^b = O_p^b(1)$.

Part (iii). Note that $\hat{\epsilon}_t^b = \epsilon_t \eta_t - w_t^\top (\hat{\theta}_{(2)} - \theta^0) \eta_t - w_t^\top \hat{\theta}_{(1)}^b$. Hence,

$$\begin{aligned} \hat{H}_{\epsilon, i\gamma}^b &= T^{-1} \sum_{i\gamma} z_t z_t^\top \left(\epsilon_t^2 \eta_t^2 + w_t^\top (\hat{\theta}_{(2)} - \theta^0) (\hat{\theta}_{(2)} - \theta^0)^\top w_t \eta_t^2 + w_t^\top \hat{\theta}_{(1)}^b \hat{\theta}_{(1)}^{b\top} w_t \right. \\ &\quad \left. - 2\epsilon_t \eta_t^2 w_t^\top (\hat{\theta}_{(2)} - \theta^0) - 2\epsilon_t \eta_t w_t^\top \hat{\theta}_{(1)}^b + 2\eta_t w_t^\top (\hat{\theta}_{(2)} - \theta^0) \hat{\theta}_{(1)}^{b\top} w_t \right) \\ &= I + II + III - 2IV - 2V + 2VI. \end{aligned} \tag{A.12}$$

We proceed by first showing that $II, \dots, VI = o_p^b(1)$ and then that $I = H_{\epsilon, i\gamma} + o_p^b(1)$.

Note that $II = \left(I_q \otimes (\hat{\theta}_{(2)} - \theta^0)^\top \right) \left[T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top \eta_t^2 \right] \left(I_q \otimes (\hat{\theta}_{(2)} - \theta^0) \right)$. Using Markov's inequality, the term in brackets satisfies $P^b(\|T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top \eta_t^2\| \geq K) \leq K^{-1} T^{-1} \sum_{i\gamma} \|z_t z_t^\top \otimes w_t w_t^\top\| = O_p(1)$ where the last equality was shown in the proof of Theorem 1. Hence, $T^{-\alpha} [T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top \eta_t^2] = o_p^b(1)$ for any $\alpha > 0$. By standard arguments for GMM-estimators in linear models, $\hat{\theta}_{(2)} - \theta^0 = O_p(T^{-1/2})$ under Assumption 4.1. Hence, $II = O_p(T^{-1/2}) o_p^b(T^\alpha) O_p(T^{-1/2}) = T^{\alpha-1} O_p(1) o_p^b(1) O_p(1) = o_p^b(1)$ for $0 < \alpha \leq 1$.

Moreover, $III = \left(I_q \otimes \hat{\theta}_{(1)}^{b\top} \right) \left[T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top \right] \left(I_q \otimes \hat{\theta}_{(1)}^b \right)$. By part (ii), $T^{1/2} \hat{\theta}_{(1)}^b = O_p^b(1)$. As shown above, $T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top = O_p(1)$. Hence, $III = O_p^b(T^{-1/2}) O_p(1) O_p^b(T^{-1/2}) = o_p^b(1)$.

By the same arguments as for terms II and III above, $IV = O_p(T^{-1/2}) o_p^b(T^\alpha) = o_p^b(1)$ and $V = O_p^b(T^{-1/2}) o_p^b(T^\alpha) = o_p^b(1)$ for any $0 < \alpha \leq \frac{1}{2}$, respectively $VI = O_p(T^{-1/2}) o_p^b(T^\alpha) O_p^b(T^{-1/2}) = o_p^b(1)$ for any $0 < \alpha \leq 1$. Hence, $\hat{H}_{\epsilon, i\gamma}^b = T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 \eta_t^2 + o_p^b(1)$, for any $0 < \alpha \leq \frac{1}{2}$.

Thus, it is left to show that $T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 \eta_t^2 - T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 = o_p^b(1)$ since $T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 = H_{\epsilon, i\gamma} +$

$o_p(1)$ by Assumptions 4.1 (b), (c), (e) and Lemma A.1. Since $E^b[T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 \eta_t^2] = T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2$, we verify the claim by Chebychev's inequality: $P^b(\|T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 (\eta_t^2 - 1)\| > K) \leq K^{-2} E^b \|T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 (\eta_t^2 - 1)\|^2 \leq K^{-2} T^{-1} [T^{-1} \sum_{i\gamma, t \neq s} \|z_t z_t^\top \epsilon_t^2 z_s z_s^\top \epsilon_s^2\| E^b |(\eta_t^2 - 1)(\eta_s^2 - 1)|]$. By Assumption 2 and applying Hölder's inequality with $p = q = 2r$ iteratively, $E^b |(\eta_t^2 - 1)(\eta_s^2 - 1)| < \infty$ and $\|z_t z_t^\top \epsilon_t^2 z_s z_s^\top \epsilon_s^2\|_r \leq \|z_t\|_{8r}^4 \|\epsilon_t\|_{8r}^4 < \infty$ for some $r > 1$. Hence, $P^b(\|T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 (\eta_t^2 - 1)\| > K) \leq T^{-1} O_p(1) = o_p(1)$ by Lemma A.1, concluding the proof. \blacksquare

Proof of Theorem 2. First, by Lemma A.4 (iii), the first result in the proof of Theorem 1, Slutsky's theorem and the continuous mapping theorem it follows that $\hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{b-1} \hat{N}_{i\gamma}^\top \xrightarrow{p^b} N_{i\gamma} H_{\epsilon, i\gamma}^{-1} N_{i\gamma}^\top \equiv V_{i\gamma}$ for $i = 1, 2$. Next, recall that $\hat{\theta}_{i\gamma, (2)}^b = (\hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{b-1} \hat{N}_{i\gamma}^\top)^{-1} (\hat{N}_{i\gamma} \hat{H}_{\epsilon, i\gamma}^{b-1} T^{-1} \sum_{i\gamma} z_t y_t^b)$ for $i = 1, 2$. Hence, $T^{1/2} (\hat{\theta}_{1\gamma, (2)}^b - \hat{\theta}_{2\gamma, (2)}^b) \xrightarrow{d_p^b} V_{1\gamma}^{-1} N_{1\gamma} H_{\epsilon, 1\gamma} G_{\epsilon, 1\gamma} - V_{2\gamma}^{-1} N_{2\gamma} H_{\epsilon, 2\gamma} G_{\epsilon, 2\gamma}$, by Lemma A.4, Slutsky's theorem and the continuous mapping theorem. Putting these results together yields the claim. \blacksquare

A.3 Proofs of 2SLS Results with a LFS

A.3.1 Asymptotic Distribution of 2SLS Test Statistics with a LFS

Definition 1. Let $C_{i\gamma} = A^0 M_{i\gamma} A^{0\top}$, $R_{i\gamma} = M_{i\gamma} M^{-1}$, $C_\gamma = [C_{1\gamma}^{-1}, -C_{2\gamma}^{-1}]$, and $Q_\gamma = C_{1\gamma} C^{-1} C_{2\gamma}$. Also, define the Gaussian processes $\mathcal{B}_{i\gamma} = A^0 [(\tilde{\theta}^{0\top} \otimes I_q) \mathcal{G}_{i\gamma} - R_{i\gamma} (\tilde{\theta}^{0\top} \otimes I_q) \mathcal{G}]$ for $i = 1, 2$, and $\mathcal{B} = \mathcal{B}_{1\gamma} + \mathcal{B}_{2\gamma}$, as well as the processes $\mathcal{B}_\gamma = \text{vec}(\mathcal{B}_{1\gamma}, \mathcal{B}_{2\gamma})$ and $\mathcal{E}_\gamma = C_\gamma \mathcal{B}_\gamma$. Let $V_{\mathcal{B}_{i\gamma}} = A^0 D_{i\gamma} H_{i\gamma} D_{i\gamma}^\top A^0$ for $i = 1, 2$, and $V_{\mathcal{B}} = A^0 D H D^\top A^0$ and $V_{\mathcal{B}_{12}, \gamma} = A^0 D_{1\gamma} H_{1\gamma} D_{1\gamma}^\top A^{0\top}$, $V_{\mathcal{B}_{2\gamma}}$ denote the covariances of $\mathcal{B}_{i\gamma}$, \mathcal{B} , and the covariance between $\mathcal{B}_{1\gamma}$ and $\mathcal{B}_{2\gamma}$. Then, $V_{\mathcal{B}_\gamma} = \begin{pmatrix} V_{\mathcal{B}_{1\gamma}} & V_{\mathcal{B}_{12}, \gamma} \\ V_{\mathcal{B}_{12}, \gamma}^\top & V_{\mathcal{B}_{2\gamma}} \end{pmatrix}$, and $V_\gamma = C_\gamma V_{\mathcal{B}_\gamma} C_\gamma^\top$, respectively. Define $\bar{D}_{i\gamma} = [\tilde{\theta}^\top \otimes I_q] - [\tilde{\theta}^\top \otimes \hat{R}_{i\gamma}]$, where $\tilde{\theta} = [1, \hat{\theta}_x]$ and $\tilde{\theta} = [0, \hat{\theta}_x]$. Then \hat{V}_γ is defined as V_γ , but replacing $C_{i\gamma}$ with $\hat{C}_{i\gamma}$, A^0 with \hat{A} , $D_{i\gamma}$ by $\bar{D}_{i\gamma}$, and $H_{1\gamma}$ by $\hat{H}_{1\gamma} = T^{-1} \sum_\gamma \hat{v}_t \hat{v}_t^\top \otimes z_t z_t^\top$, and H by $\hat{H} = T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top \otimes z_t z_t^\top$.

Lemma A.5. Suppose Assumption 1 holds, y_t is generated by (1) and x_t is generated by the LFS (2). Then, under \mathbb{H}_0 and for $i = 1, 2$, (i) $T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top \xrightarrow{p} C_{i\gamma}$; (ii) $T^{-1/2} \sum_{i\gamma} \hat{w}_t \tilde{\epsilon}_t \Rightarrow \mathcal{B}_{i\gamma}$, and $T^{-1/2} \text{vec}(\sum_{1\gamma} \hat{w}_t \tilde{\epsilon}_t, \sum_{2\gamma} \hat{w}_t \tilde{\epsilon}_t) \Rightarrow \mathcal{B}_\gamma$.

Proof of Lemma A.5. Part (i). $T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top = T^{-1} \sum_{i\gamma} \hat{A} z_t z_t^\top \hat{A}^\top$. By Assumption 1 and standard arguments, we have that $\hat{A} = A^0 + o_p(1)$, so $T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top = (A^0 + o_p(1)) T^{-1} \sum_{i\gamma} z_t z_t^\top (A^0 + o_p(1))^\top$. By Lemma A.3, $T^{-1} \sum_{i\gamma} z_t z_t^\top \xrightarrow{p} M_{i\gamma}$, for $i = 1, 2$. Hence, $T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top \xrightarrow{p} A^0 M_{i\gamma} A^{0\top} = C_{i\gamma}$.

Part (ii). Note,

$$\begin{aligned} T^{-1/2} \sum_{i\gamma} \hat{w}_t \tilde{\epsilon}_t &= \hat{A} \left(T^{-1/2} \sum_{i\gamma} z_t (\epsilon_t + u_t^\top \theta_x^0) - \hat{R}_{i\gamma} T^{-1/2} \sum_{t=1}^T z_t u_t^\top \theta_x^0 \right) \\ &= \hat{A} \left([\tilde{\theta}^{0\top} \otimes I_q] T^{-1/2} \sum_{i\gamma} v_t \otimes z_t - \hat{R}_{i\gamma} [\tilde{\theta}^{0\top} \otimes I_q] T^{-1/2} \sum_{t=1}^T v_t \otimes z_t \right). \end{aligned}$$

By Lemma A.3, $\hat{R}_{i\gamma} \xrightarrow{p} R_{i\gamma}$, $T^{-1/2} \sum_{i\gamma} v_t \otimes z_t \Rightarrow \mathcal{G}_{i\gamma}$ and $T^{-1/2} \sum_{t=1}^T v_t \otimes z_t \Rightarrow \mathcal{G}$. Hence, $T^{-1/2} \sum_{i\gamma} \hat{w}_t \tilde{\epsilon}_t \Rightarrow \mathcal{B}_{i\gamma}$

and $T^{-1/2} \text{vec}(\sum_{1\gamma} \hat{w}_t \tilde{\epsilon}_t, \sum_{2\gamma} \hat{w}_t \tilde{\epsilon}_t) \Rightarrow \mathcal{B}_\gamma$. ■

Theorem A.1 (Asymptotic Distribution LFS). *Let y_t be generated by (1) and x_t be generated by the LFS (2). Then, under \mathbb{H}_0 and Assumption 1,*

$$(i) \quad T^{1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) \Rightarrow \mathcal{E}_\gamma,$$

$$(ii) \quad \sup_{\gamma \in \Gamma} LR_T(\gamma) \Rightarrow \sup_{\gamma \in \Gamma} [\mathcal{E}_\gamma^\top Q_\gamma \mathcal{E}_\gamma / \sigma^2],$$

$$(iii) \quad \sup_{\gamma \in \Gamma} W_T(\gamma) \Rightarrow \sup_{\gamma \in \Gamma} \mathcal{E}_\gamma^\top V_\gamma^{-1} \mathcal{E}_\gamma, \text{ where } V_\gamma \text{ was defined in Definition 1.}$$

Proof of Theorem A.1. *Part (i).* $T^{1/2}(\hat{\theta}_{i\gamma} - \theta^0) = \hat{C}_{i\gamma}^{-1}(T^{-1/2} \sum_{i\gamma} \hat{w}_t \tilde{\epsilon}_t) \Rightarrow C_{i\gamma}^{-1} \mathcal{B}_{i\gamma}$ by Lemma A.5. So, $T^{1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) \Rightarrow [C_{1\gamma}^{-1}, -C_{2\gamma}^{-1}] \text{vec}(\mathcal{B}_{1\gamma}, \mathcal{B}_{2\gamma}) = C_\gamma \mathcal{B}_\gamma = \mathcal{E}_\gamma$.

Part (ii). Since $\hat{\theta} = \hat{C}^{-1}(T^{-1} \sum_{t=1}^T \hat{w}_t y_t) = \hat{C}^{-1}(\sum_{i=1}^2 T^{-1} \sum_{i\gamma} \hat{w}_t y_t)$ and $\hat{\theta}_{i\gamma} = \hat{C}_{i\gamma}^{-1}(T^{-1} \sum_{i\gamma} \hat{w}_t y_t)$, it follows that $\hat{\theta} = \sum_{i=1}^2 \hat{C}^{-1} \hat{C}_{i\gamma} \hat{\theta}_{i\gamma}$. So $\hat{\theta} = \sum_{i=1}^2 C^{-1} C_{i\gamma} \hat{\theta}_{i\gamma} + o_p(1)$, and therefore $\hat{\theta}_{1\gamma} - \hat{\theta} = C^{-1} C_{2\gamma} (\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) + o_p(1)$ and $\hat{\theta}_{2\gamma} - \hat{\theta} = C^{-1} C_{1\gamma} (\hat{\theta}_{2\gamma} - \hat{\theta}_{1\gamma}) + o_p(1)$. Hence,

$$\begin{aligned} SSR_0 - SSR_1(\gamma) &= \sum_{i=1}^2 \left[\sum_{i\gamma} (y_t - \hat{w}_t^\top \hat{\theta})^2 - (y_t - \hat{w}_t^\top \hat{\theta}_{i\gamma})^2 \right] \\ &= \sum_{i=1}^2 (\hat{\theta}_{i\gamma} - \hat{\theta})^\top \left[2 \sum_{i\gamma} \hat{w}_t \tilde{\epsilon}_t - \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top (\hat{\theta} - \theta^0) - \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top (\hat{\theta}_{i\gamma} - \theta^0) \right] \\ &= \sum_{i=1}^2 T^{1/2} (\hat{\theta}_{i\gamma} - \hat{\theta})^\top \left(T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top \right) T^{1/2} (\hat{\theta}_{i\gamma} - \hat{\theta}) \\ &= T^{1/2} (\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) [C_{2\gamma} C^{-1} C_{1\gamma} C^{-1} C_{2\gamma} + C_{1\gamma} C^{-1} C_{2\gamma} C^{-1} C_{1\gamma}] T^{1/2} (\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) + o_p(1) \\ &= T^{1/2} (\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma})^\top Q_\gamma T^{1/2} (\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) + o_p(1), \end{aligned}$$

where the last line follows because $C = \sum_{i=1}^2 C_{i\gamma}$, therefore $C^{-1} C_{1\gamma} = I_p - C^{-1} C_{2\gamma}$, $C_{1\gamma} C^{-1} C_{2\gamma} = (C - C_{2\gamma}) C^{-1} (C - C_{1\gamma}) = C - C_{1\gamma} - C_{2\gamma} + C_{2\gamma} C^{-1} C_{1\gamma} = C_{2\gamma} C^{-1} C_{1\gamma}$, so $C_{2\gamma} C^{-1} C_{1\gamma} C^{-1} C_{2\gamma} + C_{1\gamma} C^{-1} C_{2\gamma} C^{-1} C_{1\gamma} = C_{2\gamma} C^{-1} C_{1\gamma} C^{-1} C_{2\gamma} + C_{1\gamma} C^{-1} C_{2\gamma} (I_p - C^{-1} C_{2\gamma}) = Q_\gamma + (C_{2\gamma} C^{-1} C_{1\gamma} - C_{1\gamma} C^{-1} C_{2\gamma}) C^{-1} C_{2\gamma} = Q_\gamma$. Since $T^{-1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) \Rightarrow \mathcal{E}_\gamma$, $SSR_0 - SSR_1(\gamma) \Rightarrow \mathcal{E}_\gamma^\top Q_\gamma \mathcal{E}_\gamma$.

Next, $SSR_1(\gamma)/(T - 2p) = \sigma^2 + o_p(1)$, since, as shown below, $T^{-1} SSR_1(\gamma) \xrightarrow{P} \sigma^2$:

$$\begin{aligned} T^{-1} SSR_1(\gamma) &= \sum_{i=1}^2 T^{-1} \sum_{i\gamma} (y_t - \hat{w}_t^\top \hat{\theta}_{i\gamma})^2 \\ &= \sum_{i=1}^2 T^{-1} \sum_{i\gamma} (\tilde{\epsilon}_t - \hat{w}_t^\top (\hat{\theta}_{i\gamma} - \theta^0))^2 \\ &= T^{-1} \sum_{t=1}^T \tilde{\epsilon}_t^2 - 2 \sum_{i=1}^2 T^{-1} \sum_{i\gamma} \tilde{\epsilon}_t \hat{w}_t^\top (\hat{\theta}_{i\gamma} - \theta^0) + \sum_{i=1}^2 (\hat{\theta}_{i\gamma} - \theta^0)^\top (T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top) (\hat{\theta}_{i\gamma} - \theta^0). \end{aligned}$$

By Lemma A.5, $T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^\top \xrightarrow{P} C_{i\gamma}$, and $T^{-1} \sum_{i\gamma} \tilde{\epsilon}_t \hat{w}_t^\top = o_p(1)$ and, as shown in *Part (ii)* above, $\hat{\theta}_{i\gamma} - \theta^0 =$

$o_p(1)$. Hence:

$$\begin{aligned}
T^{-1}SSR_1(\gamma) &= T^{-1} \sum_{t=1}^T (\epsilon_t + u_t^\top \theta_x^0 - z_t^\top (\hat{\Pi} - \Pi^0) \theta_x^0)^2 + o_p(1) \\
&= T^{-1} \sum_{t=1}^T (\epsilon_t + u_t^\top \theta_x^0)^2 + \theta_x^{0\top} (\hat{\Pi} - \Pi^0) T^{-1} \sum_{t=1}^T z_t z_t^\top (\hat{\Pi} - \Pi^0) \theta_x^0 \\
&\quad - 2T^{-1} \sum_{t=1}^T z_t (\epsilon_t + u_t^\top \theta_x^0) (\hat{\Pi} - \Pi^0) \theta_x^0 + o_p(1) \\
&= T^{-1} \sum_{t=1}^T (\epsilon_t + u_t^\top \theta_x^0)^2 + o_p(1), \tag{A.13}
\end{aligned}$$

where the last equality used Lemma A.3 (ii) and the fact that $\hat{\Pi} - \Pi^0 = o_p(1)$. We now apply Lemma A.1 to $(\epsilon_t + u_t^\top \theta_x^0)^2$. First, $E[(\epsilon_t + u_t^\top \theta_x^0)^2] = \sigma^2$. Second, by Assumption 1 (b), $(\epsilon_t + u_t^\top \theta_x^0)^2$ is strictly stationary with ρ -mixing coefficients satisfying condition (i) in Lemma A.1. Third, by Minkowski's inequality, $\|(\epsilon_t + u_t^\top \theta_x^0)^2\|_2 \leq \|\epsilon_t^2\|_2 + \|(\theta_x^{0\top} u_t)^2\|_2 + 2\|\epsilon_t u_t^\top \theta_x^0\|_2$. Note that $\|\epsilon_t^2\|_2 = \|\epsilon_t\|_4^2 < K$, $\|(\theta_x^{0\top} u_t)^2\|_2 \leq \|\theta_z^0\|^2 \|u_t\|_4^2 < K$ and $\|\epsilon_t u_t^\top \theta_x^0\|_2 \leq \|\theta_x^0\| \|\epsilon_t\|_4 \|u_t\|_4 < K$ by Assumption 1 (c). Therefore, by Lemma A.1, $T^{-1} \sum_{t=1}^T (\epsilon_t + u_t^\top \theta_x^0)^2 \xrightarrow{P} \sigma^2$, completing the proof of Part (ii).

Part (iii). We are left to show $\hat{V}_\gamma \xrightarrow{P} V_\gamma$. Since $\hat{C}_{i\gamma} \xrightarrow{P} C_{i\gamma}$, $\hat{A} \xrightarrow{P} A^0$, $\hat{\theta}_x \xrightarrow{P} \theta_x^0$, to show that $\hat{V}_\gamma \xrightarrow{P} V_\gamma$, it suffices to show that $\hat{H}_{i\gamma} = T^{-1} \sum_{i\gamma} \hat{v}_t \hat{v}_t^\top \otimes z_t z_t^\top \xrightarrow{P} H_{i\gamma}$ for $i = 1, 2$. We proceed to show that $\hat{H}_{\epsilon, i\gamma} = T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t^2 \xrightarrow{P} H_{\epsilon, i\gamma}$:

$$\hat{H}_{\epsilon, i\gamma} = T^{-1} \sum_{1\gamma} z_t z_t^\top \hat{\epsilon}_t^2 = T^{-1} \sum_{1\gamma} z_t z_t^\top [\epsilon_t + w_t^\top (\theta^0 - \hat{\theta})]^2 \tag{A.14}$$

$$\begin{aligned}
&= T^{-1} \sum_{1\gamma} z_t z_t^\top \epsilon_t^2 + T^{-1} \sum_{1\gamma} z_t z_t^\top [(\theta^0 - \hat{\theta})^\top w_t w_t^\top (\theta^0 - \hat{\theta})] \\
&\quad + 2T^{-1} \sum_{1\gamma} z_t z_t^\top [(\theta^0 - \hat{\theta})^\top w_t \epsilon_t]. \tag{A.15}
\end{aligned}$$

First, by Assumption 1 (b), (c) and (e), and Lemma A.1, $T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t^2 \xrightarrow{P} H_{\epsilon, i\gamma}$. Second, $T^{-1} \sum_{i\gamma} z_t z_t^\top [(\theta^0 - \hat{\theta})^\top w_t w_t^\top (\theta^0 - \hat{\theta})] = [I_q \otimes (\theta^0 - \hat{\theta})]^\top [T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top] [I_q \otimes (\theta^0 - \hat{\theta})]$. Note that $\theta^0 - \hat{\theta} = O_p(T^{-1/2})$. Moreover, the asymptotic behavior of the terms *II* and *III* in (A.12), implies that $T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top = O_p(1)$. Hence, $T^{-1} \sum_{i\gamma} z_t z_t^\top [(\theta^0 - \hat{\theta})^\top w_t w_t^\top (\theta^0 - \hat{\theta})] = o_p(1)$. Similarly, $T^{-1} \sum_{i\gamma} z_t z_t^\top (\theta^0 - \hat{\theta})^\top w_t \epsilon_t = o_p(1)$. Therefore, $\hat{H}_{\epsilon, i\gamma} \xrightarrow{P} H_{\epsilon, i\gamma}$.

By similar arguments, $\hat{H}_{u, i\gamma} \xrightarrow{P} H_{u, i\gamma}$ and $\hat{H}_{\epsilon u, i\gamma} \xrightarrow{P} H_{\epsilon u, i\gamma}$, completing the proof of Part (iii). \blacksquare

A.3.2 Bootstrap Validity for 2SLS Test Statistics with LFS

Lemma A.6. *Let Assumptions 1-2 hold, y_t be generated by (1), and x_t be generated by (2). Then, under \mathbb{H}_0 and for $i = 1, 2$, (i) $T^{-1/2} \sum_{i\gamma} v_t^b \otimes z_t \xrightarrow{d^b} \mathcal{G}_{i\gamma}$; (ii) $T^{1/2}(\hat{\Pi}^b - \hat{\Pi}) = T^{1/2}(\hat{\Pi} - \Pi^0) + o_p^b(1)$; (iii) $T^{-1} \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \xrightarrow{P^b} C_{i\gamma}$; (iv) $T^{-1/2} \sum_{i\gamma} \hat{w}_t^b \hat{\epsilon}_t^b \xrightarrow{d^b} \mathcal{B}_{i\gamma}$ and $\text{vec}(T^{-1/2} \sum_{1\gamma} \hat{w}_t^b \hat{\epsilon}_t^b, T^{-1/2} \sum_{2\gamma} \hat{w}_t^b \hat{\epsilon}_t^b) \xrightarrow{d^b} \mathcal{B}_\gamma$, where $\hat{\epsilon}_t^b = y_t^b - \hat{w}_t^{b\top} \hat{\theta}$.*

Proof of Lemma A.6. *Part (i).* Since $\hat{u}_t = x_t - \hat{x}_t = u_t - (\hat{\Pi} - \Pi^0)^\top z_t$,

$$\begin{aligned}
T^{-1/2} \sum_{i\gamma} u_t^b \otimes z_t &= T^{-1/2} \sum_{i\gamma} \hat{u}_t \eta_t \otimes z_t \\
&= T^{-1/2} \sum_{i\gamma} u_t \eta_t \otimes z_t - (T^{1/2}(\hat{\Pi} - \Pi^0)^\top \otimes I_q) (T^{-1} \sum_{i\gamma} z_t \eta_t \otimes z_t) \\
&= T^{-1/2} \sum_{i\gamma} u_t \eta_t \otimes z_t + O_p(1) \times \text{vec}(T^{-1} \sum_{i\gamma} z_t z_t^\top \eta_t) \\
&= T^{-1/2} \sum_{i\gamma} u_t \eta_t \otimes z_t + o_p^b(1).
\end{aligned}$$

where the last equality follows from Lemma A.3 (iii). Similarly, $T^{-1/2} \sum_{i\gamma} \epsilon_t^b z_t = T^{-1/2} \sum_{i\gamma} \epsilon_t \eta_t z_t + o_p^b(1)$, therefore $T^{-1/2} \sum_{i\gamma} v_t^b \otimes z_t \xrightarrow{d_p^b} \mathcal{G}_{i\gamma}$ by Lemma A.3 (iv).

Part (ii). Since $\text{vec}(z_t u_t^{b\top}) = u_t^b \otimes z_t$, from *Part (i)*, and Lemma A.3 (ii) and (iv), $T^{1/2}(\hat{\Pi}^b - \hat{\Pi}) = \hat{M}^{-1}(T^{-1/2} \sum_{t=1}^T z_t u_t^{b\top}) = \hat{M}^{-1}(T^{-1/2} \sum_{t=1}^T z_t u_t^\top + o_p^b(1)) = T^{1/2}(\hat{\Pi} - \Pi^0) + o_p^b(1)$.

Part (iii). $T^{-1} \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} = \hat{A}^b (T^{-1} \sum_{i\gamma} z_t z_t^\top) \hat{A}^{b\top}$, where $\hat{A}^b = [\hat{\Pi}^b, S^\top]^\top$. By *Part (ii)*, $T^{1/2}(\hat{\Pi}^b - \Pi^0) = O_p^b(1)$, so $\hat{\Pi}^b \xrightarrow{p^b} \Pi^0$, therefore $\hat{A}^b \xrightarrow{p^b} A^0$, and $T^{-1} \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \xrightarrow{p^b} C_{i\gamma}$.

Part (iv). By *Parts (i) and (ii)* and Lemma A.5 (i)-(ii), and recalling that $[\tilde{\theta}^\top \otimes I_q] - [\check{\theta}^\top \otimes \hat{R}_{i\gamma}] \xrightarrow{p^b} [\tilde{\theta}^{0\top} \otimes I_q] - [\check{\theta}^{0\top} \otimes R_{i\gamma}]$, we have:

$$\begin{aligned}
T^{-1/2} \sum_{i\gamma} \hat{w}_t^b \check{\epsilon}_t^b &= \hat{A}^b \left(T^{-1/2} \sum_{i\gamma} z_t (\epsilon_t^b + (w_t^b - \hat{w}_t^b)^\top \hat{\theta}) \right) \\
&= \hat{A}^b \left(T^{-1/2} \sum_{i\gamma} z_t (\epsilon_t^b + u_t^{b\top} \hat{\theta}_x) - [T^{-1} \sum_{i\gamma} z_t z_t^\top] T^{1/2} (\hat{\Pi}^b - \hat{\Pi}) \hat{\theta}_x \right) \\
&= \hat{A}^b \left(T^{-1/2} \sum_{i\gamma} z_t (\epsilon_t^b + u_t^{b\top} \hat{\theta}_x) - \hat{M}_{i\gamma} \hat{M}^{-1} T^{-1/2} \sum_{t=1}^T z_t u_t^\top \hat{\theta}_x \right) + o_p^b(1) \\
&= A^0 \left([\tilde{\theta}^\top \otimes I_q] T^{-1/2} \sum_{i\gamma} v_t^b \otimes z_t - \hat{R}_{i\gamma} [\check{\theta}^\top \otimes I_q] \left(T^{-1/2} \sum_{t=1}^T v_t^b \otimes z_t \right) \right) + o_p^b(1) \\
&= A^0 \left([\tilde{\theta}^{0\top} \otimes I_q] T^{-1/2} \sum_{i\gamma} v_t^b \otimes z_t - R_{i\gamma} [\check{\theta}^{0\top} \otimes I_q] \left(T^{-1/2} \sum_{t=1}^T v_t^b \otimes z_t \right) \right) + o_p^b(1) \\
&\xrightarrow{d_p^b} \mathcal{B}_{1\gamma}.
\end{aligned}$$

Hence, also $\text{vec}(T^{-1/2} \sum_{1\gamma} \hat{w}_t^b \check{\epsilon}_t^b, T^{-1/2} \sum_{2\gamma} \hat{w}_t^b \check{\epsilon}_t^b) \xrightarrow{d_p^b} \mathcal{B}_\gamma$. ■

Proof of Theorem 3. *Part (i).* We have

$$\begin{aligned}
SSR_0^b - SSR_1^b(\gamma) &= \sum_{i=1}^2 \left[\sum_{i\gamma} (y_t^b - \hat{w}_t^{b\top} \hat{\theta}^b)^2 - (y_t^b - \hat{w}_t^{b\top} \hat{\theta}_{i\gamma}^b)^2 \right] \\
&= \sum_{i=1}^2 (\hat{\theta}_{i\gamma}^b - \hat{\theta}^b)^\top \left[2 \sum_{i\gamma} \hat{w}_t^b \check{\epsilon}_t^b - \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} (\hat{\theta}^b - \hat{\theta}) - \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} (\hat{\theta}_{i\gamma}^b - \hat{\theta}) \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^2 T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b)^\top \left(T^{-1} \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \right) T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b) \\
&= \sum_{i=1}^2 T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b)^\top C_{i\gamma} T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b) + o_p^b(1) \\
&= T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b)^\top Q_\gamma T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b) + o_p^b(1),
\end{aligned} \tag{A.16}$$

where the second to last line follows by Lemma A.6 (iii). By Lemma A.6 (iii) and (iv),

$$\begin{aligned}
T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b) &= T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}) - T^{1/2}(\hat{\theta}_{2\gamma}^b - \hat{\theta}) \\
&= (T^{-1} \sum_{1\gamma} \hat{w}_t^b \hat{w}_t^{b\top})^{-1} T^{-1/2} \sum_{1\gamma} \hat{w}_t^b \tilde{\epsilon}_t^{b\top} - (T^{-1} \sum_{2\gamma} \hat{w}_t^b \hat{w}_t^{b\top})^{-1} T^{-1/2} \sum_{2\gamma} \hat{w}_t^b \tilde{\epsilon}_t^{b\top} \\
&\stackrel{\text{d.b.}}{\Rightarrow} C_{1\gamma}^{-1} \mathcal{B}_{1\gamma} - C_{2\gamma}^{-1} \mathcal{B}_{2\gamma} = \mathcal{E}_\gamma.
\end{aligned} \tag{A.17}$$

Using (A.17) in (A.16), we have: $SSR_0^b - SSR_1^b(\gamma) \stackrel{\text{d.b.}}{\Rightarrow} \mathcal{E}_\gamma^\top Q_\gamma \mathcal{E}_\gamma$.

It remains to show that $SSR_1^b(\gamma)/(T-2p) \xrightarrow{\text{p.b.}} \sigma^2$, or, equivalently, that $T^{-1}SSR_1^b(\gamma) \xrightarrow{\text{p.b.}} \sigma^2$:

$$\begin{aligned}
T^{-1}SSR_1^b(\gamma) &= \sum_{i=1}^2 T^{-1} \sum_{i\gamma} (y_t^b - \hat{w}_t^{b\top} \hat{\theta}_{i\gamma}^b)^2 \\
&= \sum_{i=1}^2 T^{-1} \sum_{i\gamma} (\tilde{\epsilon}_t^b - \hat{w}_t^{b\top} (\hat{\theta}_{i\gamma}^b - \hat{\theta}))^2 \\
&= T^{-1} \sum_{t=1}^T (\tilde{\epsilon}_t^b)^2 - 2 \sum_{i=1}^2 T^{-1} \sum_{i\gamma} \tilde{\epsilon}_t^b \hat{w}_t^{b\top} (\hat{\theta}_{i\gamma}^b - \hat{\theta}) \\
&\quad + \sum_{i=1}^2 (\hat{\theta}_{i\gamma}^b - \hat{\theta})^\top \left(T^{-1} \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \right) (\hat{\theta}_{i\gamma}^b - \hat{\theta}).
\end{aligned}$$

By Lemma A.6 (iii) and (iv), $T^{-1} \sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \xrightarrow{\text{p.b.}} C_{i\gamma}$, and $T^{-1} \sum_{i\gamma} \tilde{\epsilon}_t^b \hat{w}_t^{b\top} = O_p^b(T^{-1/2})$. From (A.17), $\hat{\theta}_{i\gamma}^b - \hat{\theta} = O_p^b(T^{-1/2})$. Therefore, $T^{-1}SSR_1^b(\gamma) = T^{-1} \sum_{t=1}^T (\tilde{\epsilon}_t^b)^2 + o_p^b(1) = T^{-1} \sum_{t=1}^T (\epsilon_t^b + u_t^{b\top} \hat{\theta}_x - z_t^\top (\hat{\Pi}^b - \hat{\Pi})^\top \hat{\theta}_x)^2 + o_p^b(1) = T^{-1} \sum_{t=1}^T (\epsilon_t^b + u_t^{b\top} \theta_x^0)^2 + o_p^b(1)$, where the last equality used Lemma A.6 (ii), which implies $\hat{\Pi}^b - \hat{\Pi} = o_p(1)$, and the fact that $\hat{\theta}_x \xrightarrow{\text{p.b.}} \theta_x^0$. We now show that $T^{-1} \sum_{t=1}^T (\epsilon_t^b + u_t^{b\top} \theta_x^0)^2 \xrightarrow{\text{p.b.}} \sigma^2$, which then completes the proof of Part (i).

Since $T^{-1} \sum_{t=1}^T (\epsilon_t^b + u_t^{b\top} \theta_x^0)^2 = \tilde{\theta}^{0\top} (T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top \eta_t^2) \tilde{\theta}^0$, we analyze $T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top \eta_t^2$. First consider $T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^2 \eta_t^2$, the first element of this matrix. We have:

$$\begin{aligned}
T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^2 \eta_t^2 &= T^{-1} \sum_{t=1}^T (\epsilon_t - w_t^\top (\hat{\theta} - \theta^0))^2 \eta_t^2 \\
&= T^{-1} \sum_{t=1}^T \epsilon_t^2 \eta_t^2 + (\hat{\theta} - \theta^0)^\top A^0 \left(T^{-1} \sum_{t=1}^T z_t z_t^\top \eta_t^2 \right) A^{0\top} (\hat{\theta} - \theta^0) \\
&\quad - 2 \left(T^{-1} \sum_{t=1}^T \epsilon_t z_t^\top \eta_t^2 \right) A^{0\top} (\hat{\theta} - \theta^0)
\end{aligned}$$

$$= T^{-1} \sum_{t=1}^T \epsilon_t^2 \eta_t^2 + o_p(1) \left(T^{-1} \sum_{t=1}^T z_t z_t^\top \eta_t^2 \right) o_p(1) - \left(T^{-1} \sum_{t=1}^T \epsilon_t z_t^\top \eta_t^2 \right) o_p(1). \quad (\text{A.18})$$

First, we show that $T^{-1} \sum_{t=1}^T \epsilon_t^2 \eta_t^2 \xrightarrow{p^b} E[\epsilon_t^2]$. Note that $E^b[T^{-1} \sum_{t=1}^T \epsilon_t^2 (\eta_t^2 - 1)] = 0$. Hence, we have $\text{Var}^b(T^{-1} \sum_{t=1}^T \epsilon_t^2 (\eta_t^2 - 1)) = E^b[(T^{-1} \sum_{t=1}^T \epsilon_t^2 (\eta_t^2 - 1))^2] = E^b[(\eta_t^2 - 1)^2] T^{-2} \sum_{t=1}^T \epsilon_t^4 = o_p(1)$, since ϵ_t^4 satisfies the assumptions of Lemma A.1 and $E^b[(\eta_t^2 - 1)^2] < \infty$ by Assumption 2 (i). Thus, $P^b(|T^{-1} \sum_{t=1}^T \epsilon_t^2 (\eta_t^2 - 1)| > K) \leq K^{-2} E^b[(\eta_t^2 - 1)^2] T^{-2} \sum_{t=1}^T \epsilon_t^4 = o_p(1)$. Hence, $T^{-1} \sum_{t=1}^T \epsilon_t^2 \eta_t^2 \xrightarrow{p^b} E[\epsilon_t^2]$.

Second, we show that $T^{-1} \sum_{t=1}^T z_t z_t^\top \eta_t^2 \xrightarrow{p^b} M$. Note $E^b[T^{-1} \sum_{t=1}^T z_t z_t^\top \eta_t^2] = T^{-1} \sum_{t=1}^T z_t z_t^\top \xrightarrow{p} E[z_t z_t^\top]$ by Lemma A.1. Hence, $P^b(\|T^{-1} \sum_{t=1}^T z_t z_t^\top (\eta_t^2 - 1)\| > K) \leq K^{-2} T^{-1} (T^{-1} \sum_{t,s} \|z_t\|^2 \|z_s\|^2 E^b[(\eta_t^2 - 1)(\eta_s^2 - 1)]) = o_p(1)$ where the last equality follows from Assumption 2 (i), implying $E^b[(\eta_t^2 - 1)(\eta_s^2 - 1)] \leq \tilde{K} < \infty$ for some $\tilde{K} > 0$ and for all s, t , and Lemma A.1 applied to $\|z_t\|^2 \|z_s\|^2$, which ensures $T^{-1} \sum_{t,s} \|z_t\|^2 \|z_s\|^2 = O_p(1)$. Thus, $T^{-1} \sum_{t=1}^T z_t z_t^\top \eta_t^2 \xrightarrow{p^b} E[z_t z_t^\top]$.

Third, we show that $T^{-1} \sum_{t=1}^T \epsilon_t z_t \eta_t^2 = o_p^b(1)$. Note that $E^b[T^{-1} \sum_{t=1}^T \epsilon_t z_t (\eta_t^2 - 1)] = 0$. Hence, $P^b(\|T^{-1} \sum_{t=1}^T \epsilon_t z_t (\eta_t^2 - 1)\| > K) \leq K^{-2} T^{-1} (T^{-1} \sum_{t=1}^T \|z_t\| \|z_s\| |\epsilon_t \epsilon_s|) = o_p^b(1)$, by the same arguments as before. Thus, $T^{-1} \sum_{t=1}^T \epsilon_t z_t \eta_t^2 \xrightarrow{p^b} 0$. Substituting these results into (A.18), $T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^2 \eta_t^2 \xrightarrow{p^b} E[\epsilon_t^2]$. Next,

$$\begin{aligned} T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t^\top \eta_t^2 &= T^{-1} \sum_{t=1}^T (u_t + (\hat{\Pi} - \Pi^0)^\top z_t) (u_t + (\hat{\Pi} - \Pi^0)^\top z_t)^\top \eta_t^2 \\ &= T^{-1} \sum_{t=1}^T u_t u_t^\top \eta_t^2 + (\hat{\Pi} - \Pi^0)^\top T^{-1} \sum_{t=1}^T z_t z_t^\top \eta_t^2 (\hat{\Pi} - \Pi^0) \\ &\quad + [T^{-1} \sum_{t=1}^T u_t z_t^\top \eta_t^2 (\hat{\Pi} - \Pi^0)] + (\hat{\Pi} - \Pi^0)^\top [T^{-1} \sum_{t=1}^T u_t z_t \eta_t^2] \\ &\xrightarrow{p^b} E[u_t u_t^\top], \end{aligned}$$

by similar arguments as for $T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^2 \eta_t^2 \xrightarrow{p^b} E[\epsilon_t^2]$. Similarly, $T^{-1} \sum_{t=1}^T \hat{u}_t \hat{\epsilon}_t^\top \eta_t^2 \xrightarrow{p^b} E[u_t \epsilon_t]$. Therefore, $T^{-1} \sum_{t=1}^T (\epsilon_t^b + u_t^b \theta_x^0)^2 = \tilde{\theta}^{0\top} (T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top \eta_t^2) \tilde{\theta}^0 = \tilde{\theta}^{0\top} E[v_t v_t^\top] \tilde{\theta}^0 + o_p^b(1) = \sigma^2 + o_p^b(1)$, completing the proof of Part (i).

Part (ii). From (A.17), $T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b) \xrightarrow{d^b} \mathcal{E}_\gamma$, so it remains to show that $\hat{V}_\gamma^b \xrightarrow{p^b} V_\gamma$. We will only show that $\hat{V}_{\mathcal{B}, i\gamma}^b = \hat{A}^b \{(\tilde{\theta}^{b\top} \otimes I_q) \hat{H}_{i\gamma}^b (\tilde{\theta}^b \otimes I_q) + (\tilde{\theta}^{b\top} \otimes I_q) \hat{H}^b (\tilde{\theta}^b \otimes I_q) - (\tilde{\theta}^{b\top} \otimes I_q) \hat{H}_{i\gamma}^b (\tilde{\theta}^b \otimes I_q) - (\tilde{\theta}^{b\top} \otimes I_q) \hat{H}_{i\gamma}^b (\tilde{\theta}^b \otimes I_q)\} \hat{A}^{b\top} \xrightarrow{p^b} A^0 \{(\tilde{\theta}^{0\top} \otimes I_q) H_{i\gamma} (\tilde{\theta}^0 \otimes I_q) + (\tilde{\theta}^{0\top} \otimes I_q) H (\tilde{\theta}^0 \otimes I_q) - (\tilde{\theta}^{0\top} \otimes I_q) H_{i\gamma} (\tilde{\theta}^0 \otimes I_q) - (\tilde{\theta}^{0\top} \otimes I_q) H_{i\gamma} (\tilde{\theta}^0 \otimes I_q)\} A^{0\top} = V_{\mathcal{B}, i\gamma}$, where $\hat{H}_{i\gamma}^b = T^{-1} \sum_{i\gamma} \hat{v}_t^b \hat{v}_t^{b\top} \otimes z_t z_t^\top$; the rest follows by similar arguments. We have already shown that $\hat{A}^b \xrightarrow{p^b} A^0$, $\tilde{\theta}^b \xrightarrow{p^b} \tilde{\theta}^0$, and $\tilde{\theta}^b \xrightarrow{p^b} \tilde{\theta}^0$. Since the proof for $\hat{H}_{i\gamma}^b \xrightarrow{p^b} H_{i\gamma}$ is similar to $\hat{H}_{\epsilon, i\gamma}^b \xrightarrow{p^b} H_{\epsilon, i\gamma}$, where $\hat{H}_{\epsilon, i\gamma}^b = T^{-1} \sum_{i\gamma} (\hat{\epsilon}_t^b)^2 z_t z_t^\top$, we only show $\hat{H}_{\epsilon, i\gamma}^b \xrightarrow{p^b} H_{\epsilon, i\gamma}$. As in Equation (A.15) in the proof of Theorem A.1,

replacing ϵ_t , w_t and θ^0 by ϵ_t^b , w_t^b and $\hat{\theta}$:

$$\begin{aligned}\hat{H}_{\epsilon, i\gamma}^b &= T^{-1} \sum_{i\gamma} z_t z_t^\top (\epsilon_t^b)^2 + T^{-1} \sum_{i\gamma} z_t z_t^\top [w_t^{b\top} (\hat{\theta}_{i\gamma}^b - \hat{\theta})]^2 + 2T^{-1} \sum_{i\gamma} z_t z_t^\top [\epsilon_t^b w_t^{b\top} (\hat{\theta}_{i\gamma}^b - \hat{\theta})] \\ &= T^{-1} \sum_{i\gamma} z_t z_t^\top (\epsilon_t^b)^2 + [I_q \otimes ((\hat{\theta}_{i\gamma}^b - \hat{\theta})^\top \hat{A}^b)] \left(T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes z_t z_t^\top \right) [I_q \otimes (\hat{A}^{b\top} (\hat{\theta}_{i\gamma}^b - \hat{\theta}))] \\ &\quad + 2[I_q \otimes ((\hat{\theta}_{i\gamma}^b - \hat{\theta})^\top \hat{A}^b)] \left(T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^b \otimes z_t \right).\end{aligned}$$

We have already shown that $T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top = O_p(1)$ (in the proof of Theorem 1), $\hat{\theta}_{i\gamma}^b - \hat{\theta} = O_p(T^{-1/2})$ (from equation (A.17)), and $\hat{A}^b = A^0 + o_p^b(1)$ (proof of Part (ii) in Lemma A.3). Recall that $\epsilon_t^b = \hat{\epsilon}_t \eta_t$, so

$$\hat{H}_{\epsilon, i\gamma}^b = T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t^2 \eta_t^2 + O_p^b(T^{-1/2}) \left(T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t \eta_t \otimes z_t \right) + o_p^b(1)$$

As before,

$$\begin{aligned}T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t^2 \eta_t^2 &= T^{-1} \sum_{i\gamma} z_t z_t^\top \epsilon_t^2 \eta_t^2 \\ &\quad + [I_q \otimes ((\hat{\theta} - \theta^0)^\top A^0)] \left(T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t^2 \right) [I_q \otimes (A^{0\top} (\hat{\theta} - \theta^0))] \\ &\quad - 2[I_q \otimes ((\hat{\theta} - \theta^0)^\top A^0)] \left(T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes \epsilon_t z_t \eta_t^2 \right) \\ &\xrightarrow{p^b} H_{\epsilon, i\gamma}\end{aligned}$$

where the convergence was already shown in the Proof of Theorem 2 (cf. the asymptotic behaviour of the terms I , II and IV in Equation (A.12)). By similar arguments,

$$\begin{aligned}T^{-1} \sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t \eta_t \otimes z_t &= T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes z_t \epsilon_t \eta_t - (T^{-1} \sum_{i\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t) [I_q \otimes A^{0\top} (\hat{\theta} - \theta^0)] \\ &= o_p^b(T^\alpha) + o_p^b(T^\alpha) O_p(T^{-1/2}).\end{aligned}$$

Hence, $O_p(T^{-1/2}) \left(T^{-1} \sum_{1\gamma} z_t z_t^\top \hat{\epsilon}_t \eta_t \otimes z_t \right) = O_p^b(T^{-1/2}) (o_p^b(T^\alpha) + o_p^b(T^\alpha) O_p(T^{-1/2})) = o_p^b(1)$ if $\alpha \leq \frac{1}{2}$. Hence, we conclude that $\hat{H}_{\epsilon, i\gamma}^b = \mathbb{E}[z_t z_t^\top \epsilon_t^2 \mathbf{1}[q_t \leq \gamma]] + o_p^b(1)$ for any $\alpha \leq \frac{1}{2}$. \blacksquare

A.4 Proofs of 2SLS Results with a TFS

Theorem A.2. Under Assumption 1 and the TFS in (3), (i) $\hat{\rho} - \rho^0 \xrightarrow{p} 0$; (ii) $T(\hat{\rho} - \rho^0) = O_p(1)$; (iii) $T^{1/2} \text{vec}(\hat{\Pi}_i - \Pi_i^0) = T^{1/2} \text{vec}(\hat{\Pi}_{i\rho^0} - \Pi_i^0) + o_p(1)$.

Proof of Theorem A.2. Wlog, assume that $\hat{\rho} \leq \rho^0$ (the proofs for $\hat{\rho} > \rho^0$ are similar and omitted for simplicity).

Part (i). Let $Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) = \text{tr} \left(T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t^\top \right) = T^{-1} \sum_{t=1}^T \hat{u}_t^\top \hat{u}_t = T^{-1} \sum_{i=1}^2 \sum_{i\hat{\rho}} (x_t^\top - z_t^\top \hat{\Pi}_{i\hat{\rho}}) (x_t - \hat{\Pi}_{i\hat{\rho}}^\top z_t)$, and $d_t = \hat{u}_t - u_t$. Then, by definition, $Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) = T^{-1} \sum_{t=1}^T (u_t + d_t)^\top (u_t + d_t) \leq Q(\Pi_1^0, \Pi_2^0, \rho^0) = T^{-1} \sum_{t=1}^T u_t^\top u_t$ with probability one. This implies that $2T^{-1} \sum_{t=1}^T u_t^\top d_t + T^{-1} \sum_{t=1}^T d_t^\top d_t \leq 0$ with probability one. We now prove consistency in two steps. In *part (i1)*, we show that $T^{-1} \sum_{t=1}^T u_t^\top d_t = o_p(1)$, implying $T^{-1} \sum_{t=1}^T d_t^\top d_t = o_p(1)$, since $T^{-1} \sum_{t=1}^T d_t^\top d_t \geq 0$. In *part (i2)*, we show that if $\hat{\rho} \xrightarrow{P} \rho^0$, then, with strictly positive probability, $T^{-1} \sum_{t=1}^T d_t^\top d_t > K$ for some $K > 0$, contradicting $T^{-1} \sum_{t=1}^T d_t^\top d_t = o_p(1)$, and therefore yielding $\hat{\rho} \xrightarrow{P} \rho^0$.

Part (i1). Define $\sum_{\Delta}(\cdot) = \sum_{t=1}^T (\cdot) \mathbf{1}[\rho < q_t \leq \rho^0]$. Then, for any $\rho \leq \rho^0$ instead of just $\hat{\rho}^{17}$, and using $\text{vec}(ABC) = (C^\top \otimes A) \text{vec}(B)$, we have:

$$\begin{aligned} T^{-1} \sum_{t=1}^T u_t^\top d_t &= T^{-1} \sum_{1\rho} u_t^\top (\Pi_1^0 - \hat{\Pi}_{1\rho})^\top z_t + T^{-1} \sum_{\Delta} u_t^\top (\Pi_1^0 - \hat{\Pi}_{2\rho})^\top z_t + T^{-1} \sum_{2\rho^0} u_t^\top (\Pi_2^0 - \hat{\Pi}_{2\rho})^\top z_t \\ &= (T^{-1} \sum_{1\rho} z_t \otimes u_t)^\top \text{vec}(\Pi_1^0 - \hat{\Pi}_{1\rho}) + (T^{-1} \sum_{\Delta} z_t \otimes u_t)^\top \text{vec}(\Pi_1^0 - \hat{\Pi}_{2\rho}) \\ &\quad + (T^{-1} \sum_{2\rho^0} z_t \otimes u_t)^\top \text{vec}(\Pi_2^0 - \hat{\Pi}_{2\rho}). \end{aligned}$$

Since $\rho \leq \rho^0$, it follows that $\hat{\Pi}_{1\rho}$ is computed based on observations corresponding to subsamples $\mathbf{1}[q_t \leq \rho^0]$ only and, therefore, by standard arguments $\hat{\Pi}_{1\rho} - \Pi_1^0 = o_p(1)$. On the other hand, $\hat{\Pi}_{2\rho}$ is constructed using observations from both subsamples satisfying $\mathbf{1}[q_t \leq \rho^0]$, respectively $\mathbf{1}[q_t > \rho]$. Straightforward calculations give $\hat{\Pi}_2 = (T^{-1} \sum_{2\rho} z_t z_t^\top)^{-1} [(T^{-1} \sum_{\Delta} z_t z_t^\top) \Pi_1^0 + (T^{-1} \sum_{2\rho^0} z_t z_t^\top) \Pi_2^0 + (T^{-1} \sum_{2\rho} z_t u_t^\top)] = O_p(1)[O_p(1) + O_p(1) + o_p(1)] = O_p(1)$, by Lemma A.3 (i) and (ii). Hence, $\Pi_1^0 - \hat{\Pi}_{2\rho} = O_p(1)$, and $\Pi_2^0 - \hat{\Pi}_{2\rho} = O_p(1)$. Also by Lemma A.3 (ii), $T^{-1} \sum_{1\rho} z_t \otimes u_t$, $T^{-1} \sum_{2\rho^0} z_t \otimes u_t$, $T^{-1} \sum_{\Delta} z_t \otimes u_t = o_p(1)$. Therefore, $T^{-1} \sum_{t=1}^T u_t^\top d_t = o_p(1)$ (uniformly in ρ). Because these results hold uniformly over ρ , we have $o_p(1) + T^{-1} \sum_{t=1}^T d_t^\top d_t \leq 0$ uniformly over ρ , and therefore also at $\hat{\rho}$, so $T^{-1} \sum_{t=1}^T d_t^\top d_t = o_p(1)$.

Part (i2). By the continuity assumption 1 (e), there exists an $\epsilon > 0$ such that with positive probability, $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$. If $\hat{\rho} \xrightarrow{P} \rho^0$, then $\hat{\rho} < \rho^0 - \epsilon$ because $\hat{\rho} \leq \rho^0$. Consequently, the residuals evaluated over the sub-sample $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$ will also be evaluated at $\hat{\Pi}_{2\hat{\rho}} = \hat{\Pi}_2$ since $\hat{\Pi}_2$ is the multivariate LS estimator in the sample $q_t > \hat{\rho}$. However, the true parameter values are Π_1^0 for $q_t \in [\rho^0 - \epsilon, \rho^0]$, and Π_2^0 for $q_t \in (\rho^0, \rho^0 + \epsilon]$. Let $\sum_A = \sum \mathbf{1}[q_t \in [\rho^0 - \epsilon, \rho^0]]$, $\sum_B = \sum \mathbf{1}[q_t \in (\rho^0, \rho^0 + \epsilon]]$, and $\sum_{AB} = \sum \mathbf{1}[q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]]$. Denote by $\pi_{1,i}^0$, $\pi_{2,i}^0$ and $\hat{\pi}_{2,i}$ the i -th columns of Π_1^0 , Π_2^0 , respectively $\hat{\Pi}_2$. Further, let η_1 and η_2 be the minimum eigenvalues of $M_{1\rho^0} - M_{1\rho^0 - \epsilon}$, respectively $M_{1\rho^0 + \epsilon} - M_{1\rho^0}$. Recall that $\text{tr}(ABC) = \text{vec}(A^\top)^\top (I \otimes B) \text{vec}(C)$ for conformable matrices A , B and C . Then:

$$\begin{aligned} T^{-1} \sum_{AB} d_t^\top d_t &= T^{-1} \sum_A \text{tr} \left([\Pi_1^0 - \hat{\Pi}_2]^\top z_t z_t^\top [\Pi_1^0 - \hat{\Pi}_2] \right) + T^{-1} \sum_B \text{tr} \left([\Pi_2^0 - \hat{\Pi}_2]^\top z_t z_t^\top [\Pi_2^0 - \hat{\Pi}_2] \right) \\ &= \text{vec}(\Pi_1^0 - \hat{\Pi}_2)^\top \left(T^{-1} \sum_A I_{p_1} \otimes z_t z_t^\top \right) \text{vec}(\Pi_1^0 - \hat{\Pi}_2) \\ &\quad + \text{vec}(\Pi_2^0 - \hat{\Pi}_2)^\top \left(T^{-1} \sum_B I_{p_1} \otimes z_t z_t^\top \right) \text{vec}(\Pi_2^0 - \hat{\Pi}_2) \end{aligned}$$

¹⁷Note that $\sum_{\Delta}(\cdot) = 0$ in case of $\rho = \rho^0$.

$$\begin{aligned}
&= \sum_{i=1}^{p_1} (\pi_{1,i}^0 - \hat{\pi}_{2,i})^\top (M_{1\rho^0} - M_{1\rho^0-\epsilon})(\pi_{1,i}^0 - \hat{\pi}_{2,i}) \\
&\quad + \sum_{i=1}^{p_1} (\pi_{2,i}^0 - \hat{\pi}_{2,i})^\top (M_{1\rho^0+\epsilon} - M_{1\rho^0})(\pi_{2,i}^0 - \hat{\pi}_{2,i}) + o_p(1) \\
&\geq \eta_1 \sum_{i=1}^{p_1} \|\pi_{1,i}^0 - \hat{\pi}_{2,i}\|^2 + \eta_2 \sum_{i=1}^{p_1} \|\pi_{2,i}^0 - \hat{\pi}_{2,i}\|^2 + o_p(1) \\
&\geq \min\{\eta_1, \eta_2\} \sum_{i=1}^{p_1} \frac{\|\pi_{1,i}^0 - \pi_{2,i}^0\|^2}{2} + o_p(1)
\end{aligned}$$

where the last equality used continuity of $M_{1\rho}$ at ρ^0 (Assumption 1 (d)) and the last inequality used the fact that for any vectors a, b, c of the same length, $(a - b)^\top (a - b) + (c - b)^\top (c - b) \geq (a - c)^\top (a - c)/2$. Next, $\min\{\eta_1, \eta_2\} > K > 0$ by Assumption 1 (d) and there exists at least one $i = 1, \dots, p_1$ such that $\|\Pi_{1,i}^0 - \Pi_{2,i}^0\| > 0$ by Assumption 1 (g). Hence, $\text{plim } T^{-1} \sum_{t=1}^T d_t^\top d_t > K > 0$ with strictly positive probability, reaching a contradiction.

Part (ii). Let $\zeta_0 = T^{-1} \sum_{t=1}^T \mathbf{1}[q_t \leq \rho^0]$, $\zeta = T^{-1} \sum_{t=1}^T \mathbf{1}[q_t \leq \rho]$ and $\hat{\zeta} = T^{-1} \sum_{t=1}^T \mathbf{1}[q_t \leq \hat{\rho}]$. Then $\rho_0 - \rho < \epsilon$, for some small ϵ , can equivalently be written as $\zeta_0 - \zeta < \omega$, for some small ω (where there is a one-to-one correspondence between ω and ϵ), or in other words, the difference between the *fraction* of observations below the ρ^0 and below the ρ quantiles is smaller than ω , as the distribution of q_t is continuous. Define the set $V_\omega = \{\zeta : \zeta_0 - \zeta < \omega\}$ for some $\omega > 0$. Since $\hat{\rho} \xrightarrow{P} \rho^0$ by Part (i), it holds that $\hat{\zeta} \in V_\omega$, for large enough T and small enough ω . Thus, we only need to consider the case where $\zeta_0 - \zeta < \omega$, for some small $\omega > 0$. For $C > 0$, define the set $V_\omega(C) = \{\zeta : \zeta^0 - \zeta \leq \omega, T(\zeta^0 - \zeta) > C\}$, i.e. the set of all values such that the *number* of observations between the two quantiles is larger C . By construction $V_\omega(C) \subset V_\omega$. Thus, if we can show that $\hat{\zeta} \notin V_\omega(C)$, it follows that the number of observations $T\zeta^0 - T\zeta \leq C$, which means in turn that $T(\hat{\rho} - \rho^0) = O_p(1)$. Hence, the quantiles $\hat{\rho}$ and ρ cannot be more than a fixed number apart with large probability. To this extent, recall the definition $Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_{2\rho}, \rho)$ of the multivariate SSR in Part (i) and that $Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) \leq Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0)$ with probability one. To verify the claim, we show that for each $K > 0$ there exist $C > 0$ and $\epsilon > 0$ such that $P(\min_{\zeta \in V_\omega(C)} \{Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_{2\rho}, \rho) - Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0)\} \leq 0) < K$ for large T . That is, we are going to show that $\hat{\zeta} \notin V_\omega(C)$. Thus, it must follow that $T(\hat{\zeta} - \zeta^0) \leq C$ with large probability.

To this extent, define $Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0)$ where $\hat{\Pi}_\Delta$ is constructed using observations $\rho < q_t \leq \rho^0$; i.e., next to ρ we introduce a second threshold at ρ^0 such that $\hat{\Pi}_\Delta$ is evaluated over samples associated with Π_1^0 only. Note that $Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) - Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0) = (Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) - Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0)) - (Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0) - Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0))$. Define $\hat{Q}_\Delta = \frac{1}{T(\zeta^0 - \zeta)} \sum_\Delta z_t z_t^\top$ and $\hat{M}_\Delta = (\zeta^0 - \zeta) \hat{Q}_\Delta$. Then, by Bai and Perron (1998), pp. 70,

$$\begin{aligned}
\frac{Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) - Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0)}{T(\zeta^0 - \zeta)} &= \text{tr}[(\hat{\Pi}_{2\rho^0} - \hat{\Pi}_\Delta)^\top (\hat{Q}_\Delta - \hat{Q}_\Delta \hat{M}_{2\rho}^{-1} \hat{M}_\Delta) (\hat{\Pi}_{2\rho^0} - \hat{\Pi}_\Delta)] \\
\frac{Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0) - Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0)}{T(\zeta^0 - \zeta)} &= \text{tr}[(\hat{\Pi}_{1\rho} - \hat{\Pi}_\Delta)^\top (\hat{Q}_\Delta - \hat{Q}_\Delta \hat{M}_{1\rho^0}^{-1} \hat{M}_\Delta) (\hat{\Pi}_{1\rho} - \hat{\Pi}_\Delta)].
\end{aligned}$$

By Lemma A.3 (i) and (ii), $\hat{M}_{1\rho^0} = M_{1\rho^0} + o_p(1)$ and $\hat{M}_{2\rho} = M_{2\rho} + o_p(1)$ so that their inverses are $O_p(1)$. Thus, also on $V_\omega(C)$. Similarly, $\hat{\Pi}_{1\rho} = \Pi_1^0 + o_p(1)$ (since it only uses observations satisfying $q_t \leq \rho < \rho^0$) and $\hat{\Pi}_{2\rho^0} = \Pi_2^0 + o_p(1)$. Straightforward calculations show that $\hat{\Pi}_\Delta = \Pi_1^0 + (T^{-1} \sum_\Delta z_t z_t^\top)^{-1} (T^{-1} \sum_\Delta z_t u_t^\top)$ if $\rho < \rho^0$. By Lemma 3, we have $\hat{M}_\Delta = T^{-1} \sum_\Delta z_t z_t^\top = T^{-1} \sum_{1\rho^0} z_t z_t^\top - T^{-1} \sum_{1\rho} z_t z_t^\top = M_{1\rho^0} - M_{1\rho} + o_p(1)$ and, similarly $T^{-1} \sum_\Delta z_t u_t^\top = o_p(1)$. For the former term, we have that $M_{1\rho^0} - M_{1\rho}$ has eigenvalues bounded away from 0 on $V_\omega(C)$ by Assumption 1 (d) and its inverse is therefore bounded (in the sense that its largest eigenvalue is smaller/equal than some finite constant). Hence, it follows that $\hat{\Pi}_\Delta = \Pi_1^0 + o_p(1)$ on $V_\omega(C)$. Moreover, on $V_\omega(C)$, we have

$$\begin{aligned} \|\hat{M}_\Delta\| &= \|E[z_t z_t^\top (\mathbf{1}[q_t \leq \rho^0] - \mathbf{1}[q_t \leq \rho])] + o_p(1)\| \leq (E\|z_t\|^4)^{1/2} \left(\int_\rho^{\rho^0} f(x) dx \right)^{1/2} + o_p(1) \\ &\leq M \left(\int_{\rho^0 - \epsilon}^{\rho^0} f(x) dx \right)^{1/2} + o_p(1) = M \sqrt{f(b)\epsilon} + o_p(1) = \sqrt{\epsilon} O_p(1), \end{aligned}$$

where the second-to-last equality holds by continuity and boundedness of $f(x)$ (Assumption 1 (e)) for some $b \in [\rho^0 - \epsilon, \rho^0]$. Moreover, $\hat{Q}_\Delta = \frac{1}{T(\zeta^0 - \zeta)} \sum_\Delta z_t z_t^\top = O_p(1)$ on $V_\omega(C)$ for large enough C by Lemma A.3 since $T(\zeta^0 - \zeta) > C > 0$ such that $[T(\zeta^0 - \zeta)]^{-1}$ is bounded. Thus, on $V_\omega(C)$ it follows that

$$\begin{aligned} \frac{Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) - Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0)}{T(\zeta^0 - \zeta)} &= \text{tr}[(\Pi_2^0 - \Pi_1^0)^\top \hat{Q}_\Delta (\Pi_2^0 - \Pi_1^0)^\top] + \sqrt{\epsilon} O_p(1) + o_p(1) \\ \frac{Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0) - Q(\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}, \rho, \rho^0)}{T(\zeta^0 - \zeta)} &= o_p(1) \end{aligned}$$

and consequently that

$$\frac{Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) - Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0)}{T(\zeta^0 - \zeta)} = \text{tr}[(\Pi_2^0 - \Pi_1^0)^\top \hat{Q}_\Delta (\Pi_2^0 - \Pi_1^0)^\top] + \sqrt{\epsilon} O_p(1) + o_p(1).$$

Finally, $\hat{Q}_\Delta = \frac{M_{1\rho^0} - M_{1\rho}}{\zeta^0 - \zeta} + o_p(1)$ on $V_\omega(C)$ for large enough C . Since $M_{1\rho^0} - M_{1\rho}$ has smallest eigenvalue bounded away from zero (Assumption 1 (d)) and since $\rho^0 - \rho > 0$ it follows that $\text{tr}[(\Pi_2^0 - \Pi_1^0)^\top \frac{M_{1\rho^0} - M_{1\rho}}{\zeta^0 - \zeta} (\Pi_2^0 - \Pi_1^0)^\top] > K > 0$. Hence, $\frac{Q(\hat{\Pi}_{1\hat{\rho}}, \hat{\Pi}_{2\hat{\rho}}, \hat{\rho}) - Q(\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}, \rho^0)}{T(\zeta^0 - \zeta)} > K > 0$. Hence, $\hat{\zeta} \notin V_\omega(C)$ and therefore, $T(\zeta^0 - \hat{\zeta}) \leq C$ with large probability, hence $T(\rho^0 - \rho) = O_p(1)$.

Part (iii). Since any partial sum in the expression of $\hat{\Pi}_i$ differs by the partial sum in the expression of $\hat{\Pi}_{i\rho^0}$ by $|T(\rho^0 - \hat{\rho})| < D$ terms, for some $D > 0$, which are uniformly bounded by Assumption 1(c), it follows that $T^{1/2} \text{vec}(\hat{\Pi}_i - \Pi_i^0) = T^{1/2} \text{vec}(\hat{\Pi}_{i\rho^0} - \Pi_i^0) + o_p(1)$, for $i = 1, 2$. The rest of the proof follows standard arguments. \blacksquare

By Theorem A.2 and its proof, wlog, we treat in what follows $\hat{\rho}$ as if it was equal to ρ^0 .

A.4.1 Asymptotic Distribution of 2SLS Test Statistics with a TFS

Definition 2. Let $A_i^0 = [\Pi_i^0, S^\top]^\top$ be the augmented matrices of the FS slope parameters, where $S = [I_{p_2}, \mathbf{0}_{p_2 \times q_1}]$, $q_1 = q - p_2$, and $\hat{A}_i = [\hat{\Pi}_i, S^\top]^\top$. Hence, $z_{1t} = Sz_t$ and $w_t = A_1^0 z_t \mathbf{1}[q_t \leq \rho^0] + A_2^0 z_t \mathbf{1}[q_t > \rho^0] + \bar{u}_t$. Let $A_t^0 = A_1^0 \mathbf{1}[q_t \leq \rho^0] + A_2^0 \mathbf{1}[q_t > \rho^0]$, $\Pi_t = \Pi_1^0 \mathbf{1}[q_t \leq \rho^0] + \Pi_2^0 \mathbf{1}[q_t > \rho^0]$.

Let \wedge and \vee define the minimum and maximum operators. Let $C_{1\gamma} = A_1^0 M_{1, \rho^0 \wedge \gamma} A_1^{0\top} + A_2^0 (M_{1\gamma} - M_{1, \rho^0 \wedge \gamma}) A_2^{0\top}$ and $C_{2\gamma} = A_1^0 (M_{1, \rho^0 \vee \gamma} - M_{1\gamma}) A_1^{0\top} + A_2^0 M_{2, \rho^0 \vee \gamma} A_2^{0\top}$. Also, $C_\gamma = [C_{1\gamma}^{-1}, -C_{2\gamma}^{-1}]$, $C = C_{1\gamma} + C_{2\gamma}$, $R_{i\gamma} = M_{i\gamma} M_{i\rho^0}^{-1}$ for $i = 1, 2$, $D = [1, 0_{1 \times p_1}] \otimes I_q$, and $Q_\gamma = C_{1\gamma} C^{-1} C_{2\gamma}$. Let $F_{i\gamma} = [\check{\theta}^{0\top} \otimes R_{i\gamma}]$ and $D_{i\gamma} = [\tilde{\theta}^{0\top} \otimes I_q] - F_{i\gamma}$. Also define the Gaussian processes:

$$\mathcal{B}_1(\gamma) = \begin{cases} A_1^0 (D_{1\gamma} \mathcal{G}_1(\gamma) - F_{1\gamma} (\mathcal{G}_1(\rho^0) - \mathcal{G}_1(\gamma))), & \gamma \leq \rho^0 \\ \mathcal{B} - A_2^0 (D_{2\gamma} \mathcal{G}_2(\gamma) - F_{2\gamma} (\mathcal{G}_1(\gamma) - \mathcal{G}_1(\rho^0))) & \gamma > \rho^0, \end{cases}$$

where $\mathcal{B} = \sum_{i=1}^2 A_i^0 D \mathcal{G}_i(\rho^0)$. Let $V_{\mathcal{B}} = \sum_{i=1}^2 A_i^0 D H_{i, \rho^0} D^\top A_i^{0\top}$,

$$V_{\mathcal{B}, 1\gamma} = \begin{cases} A_1^0 [D_{1\gamma} H_{1\gamma} D_{1\gamma} + F_{1\gamma} (H_{1, \rho^0} - H_{1\gamma}) F_{1\gamma}] A_1^{0\top}, & \gamma \leq \rho^0, \\ A_1^0 D H_{1, \rho^0} D A_1^0 + A_2^0 [(D + F_{2\gamma})(H_{1\gamma} - H_{1, \rho^0})(D + F_{2\gamma})^\top \\ + (D - D_{2\gamma}) H_{2\gamma} (D - D_{2\gamma})^\top] A_2^{0\top}, & \gamma > \rho^0 \end{cases},$$

$$V_{\mathcal{B}, 12\gamma} = \begin{cases} A_1^0 [D_{1\gamma} H_{1\gamma} - F_{1\gamma} (H_{1, \rho^0} - H_{1\gamma})] D^\top A_1^{0\top} - V_{\mathcal{B}, 1\gamma}, & \gamma \leq \rho^0 \\ A_2^0 [(D - D_{2\gamma}) H_{2\gamma} D_{2\gamma}^\top - (F_{2\gamma} + D)(H_{1\gamma} - H_{1, \rho^0}) F_{2\gamma}^\top] A_2^{0\top}, & \gamma > \rho^0 \end{cases},$$

and $V_{\mathcal{B}, 2\gamma} = V_{\mathcal{B}} - V_{\mathcal{B}, 1\gamma} - V_{\mathcal{B}, 12, \gamma} - V_{\mathcal{B}, 12, \gamma}^\top$ denote the covariances of \mathcal{B} , $\mathcal{B}_{1\gamma}$, the covariance between $\mathcal{B}_{1\gamma}$ and

$\mathcal{B}_{2\gamma}$, as well as the covariance of $\mathcal{B}_{2\gamma}$. Then $V_{\mathcal{B}, \gamma} = \begin{bmatrix} V_{\mathcal{B}, 1\gamma} & V_{\mathcal{B}, 12, \gamma} \\ V_{\mathcal{B}, 12, \gamma}^\top & V_{\mathcal{B}, 2\gamma} \end{bmatrix}$, and $V_\gamma = C_\gamma V_{\mathcal{B}, \gamma} C_\gamma^\top$.

Let $\bar{F}_{i\gamma} = [\check{\theta}^\top \otimes \hat{R}_{i\gamma}]$ and $\bar{D}_{i\gamma} = [\tilde{\theta}^\top \otimes I_q] - \bar{F}_{i\gamma}$, where $\tilde{\theta}^\top = [1, \hat{\theta}_x]$ and $\check{\theta} = [0, \hat{\theta}^\top]$. where $\tilde{\theta} = [1, \hat{\theta}_x]$ and $\check{\theta} = [0, \hat{\theta}_x]$. Then \hat{V}_γ is defined as V_γ , but replacing $C_{i\gamma}$ with $\hat{C}_{i\gamma}$, A_i^0 with \hat{A}_i , $D_{i\gamma}$ by $\bar{D}_{i\gamma}$, $F_{i\gamma}$ by $\bar{F}_{i\gamma}$, $H_{i\gamma}$ by $\hat{H}_{i\gamma} = T^{-1} \sum_{i\gamma} \hat{v}_t \hat{v}_t^\top \otimes z_t z_t^\top$.

With this new notation, we now reprove Lemmas A.5 and A.6, and Theorems A.1 and 3, for x_t generated by the TFS (3) instead of the LFS (2).

Proof of Lemma A.5. Part (i). $T^{-1} \sum_{i\gamma} z_t z_t^\top \xrightarrow{P} M_{i\gamma}$ still holds, as the result is not specific to a LFS or

TFS. So,

$$\begin{aligned}
T^{-1} \sum_{1\gamma} \hat{w}_t \hat{w}_t^\top &= \hat{A}_1 T^{-1} \sum_{1, \rho^0 \wedge \gamma} z_t z_t^\top \hat{A}_1^\top + \hat{A}_2 \left(T^{-1} \sum_{1\gamma} z_t z_t^\top - T^{-1} \sum_{1, \rho^0 \wedge \gamma} z_t z_t^\top \right) \hat{A}_2^\top \\
&\stackrel{P}{\rightarrow} A_1^0 M_{1, \rho^0 \wedge \gamma} A_1^{0\top} + A_2^0 (M_{1\gamma} - M_{1, \rho^0 \wedge \gamma}) A_2^{0\top} = C_{1\gamma} \\
T^{-1} \sum_{2\gamma} \hat{w}_t \hat{w}_t^\top &= \hat{A}_1 \left(T^{-1} \sum_{1, \rho^0 \vee \gamma} z_t z_t^\top - T^{-1} \sum_{1\gamma} z_t z_t^\top \right) \hat{A}_1^\top + \hat{A}_2 T^{-1} \sum_{2, \rho^0 \vee \gamma} z_t z_t^\top \hat{A}_2^\top \\
&\stackrel{P}{\rightarrow} A_1^0 (M_{1, \rho^0 \vee \gamma} - M_{1\gamma}) A_1^{0\top} + A_2^0 M_{2, \rho^0 \vee \gamma} A_2^{0\top} = C_{2\gamma}.
\end{aligned}$$

Part (ii). The result $T^{-1/2} \sum_{i\gamma} v_t \otimes z_t \Rightarrow \mathcal{G}_i(\gamma)$ still holds. But now,

$$\tilde{\epsilon}_t = \epsilon_t + (\hat{x}_t - x_t)^\top \theta_x^0 = \epsilon_t + u_t^\top \theta_x^0 - \mathbf{1}[q_t \leq \rho^0] [z_t^\top (\hat{\Pi}_1 - \Pi_1^0) \theta_x^0] - \mathbf{1}[q_t > \rho^0] [z_t^\top (\hat{\Pi}_2 - \Pi_2^0) \theta_x^0].$$

Therefore, for $\gamma \leq \rho^0$,

$$\begin{aligned}
T^{-1/2} \sum_{1\gamma} \hat{w}_t \tilde{\epsilon}_t &= A_1^0 \left(T^{-1/2} \sum_{1\gamma} z_t (\epsilon_t + u_t^\top \theta_x^0) - M_{1\gamma} M_{1\rho^0}^{-1} T^{-1/2} \sum_{1\rho^0} z_t u_t^\top \theta_x^0 \right) \\
&\Rightarrow A_1^0 ([\tilde{\theta}^0 \otimes I_q] \mathcal{G}_1(\gamma) - [\tilde{\theta}^0 \otimes R_{1\gamma}] \mathcal{G}_1(\rho^0)) = A_1^0 (D_{1\gamma} \mathcal{G}_1(\gamma) - F_{1\gamma}(\mathcal{G}_1(\rho^0) - \mathcal{G}_1(\gamma))) \\
&= \mathcal{B}_1(\gamma).
\end{aligned}$$

For $\gamma > \rho^0$,

$$\begin{aligned}
T^{-1/2} \sum_{1\gamma} \hat{w}_t \tilde{\epsilon}_t &= T^{-1/2} \sum_{t=1}^T \hat{w}_t \tilde{\epsilon}_t - T^{-1/2} \sum_{2\gamma} \hat{w}_t \tilde{\epsilon}_t \\
&\Rightarrow A_1^0 D \mathcal{G}_1(\rho^0) + A_2^0 D \mathcal{G}_2(\rho^0) - A_2^0 ([\tilde{\theta}^0 \otimes I_q] \mathcal{G}_2(\gamma) - F_{2\gamma} \mathcal{G}_2(\rho^0)) \\
&= \mathcal{B} - A_2^0 (D_{2\gamma} \mathcal{G}_2(\gamma) - F_{2\gamma}(\mathcal{G}_1(\gamma) - \mathcal{G}_1(\rho^0))) = \mathcal{B}_1(\gamma).
\end{aligned}$$

Because $T^{-1/2} \sum_{t=1}^T \hat{w}_t \tilde{\epsilon}_t \Rightarrow A_1^0 D \mathcal{G}_1(\rho^0) + A_2^0 D \mathcal{G}_2(\rho^0)$, $T^{-1/2} \sum_{2\gamma} \hat{w}_t \tilde{\epsilon}_t \Rightarrow A_1^0 D \mathcal{G}_1(\rho^0) + A_2^0 D \mathcal{G}_2(\rho^0) - \mathcal{B}_1(\gamma) = \mathcal{B} - \mathcal{B}_1(\gamma) = \mathcal{B}_2(\gamma)$, and $\text{vec}(T^{-1/2} \sum_{1\gamma} \hat{w}_t \tilde{\epsilon}_t, T^{-1/2} \sum_{2\gamma} \hat{w}_t \tilde{\epsilon}_t) \Rightarrow \mathcal{B}(\gamma)$. \blacksquare

Proof of Theorem A.1. *Part (i).* Because $T^{-1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) = \hat{C}_{1\gamma}^{-1} T^{-1/2} \sum_{1\gamma} \hat{w}_t \tilde{\epsilon}_t - \hat{C}_{2\gamma}^{-1} T^{-1/2} \sum_{2\gamma} \hat{w}_t \tilde{\epsilon}_t$, the desired result follows directly from Lemma A.3.

Part (ii). Follows the same steps as for the LFS proof until equation (A.13). Then note that because $\hat{\Pi}_i - \Pi_i^0 = o_p(1)$,

$$\begin{aligned}
T^{-1} SSR_1(\gamma) &= T^{-1} \sum_{i=1}^2 \left(\sum_{i\rho^0} (\epsilon_t + u_t^\top \theta_x^0)^2 - 2 \sum_{i\rho^0} (\epsilon_t + u_t^\top \theta_x^0) z_t^\top (\hat{\Pi}_i - \Pi_i^0) \theta_x^0 \right. \\
&\quad \left. + \sum_{i\rho^0} \theta_x^{0\top} (\hat{\Pi}_i - \Pi_i^0)^\top T^{-1} \sum_{i\rho^0} z_t z_t^\top (\hat{\Pi}_i - \Pi_i^0) \theta_x^0 \right) \\
&= \sigma^2 + o_p(1),
\end{aligned}$$

following the same arguments as in the LFS proof.

Part (iii). It can be shown by similar arguments to the LFS, but now separately for cases $\gamma \leq \rho^0$ and $\gamma \geq \rho^0$, and taking to account the different parameter estimates in different regimes, that $\hat{V}_\gamma \xrightarrow{P} V_\gamma$. Because of part (i) of this theorem, the desired result follows. ■

Proof of Lemma A.5 and Theorem 3. As evident from the proof of Theorem A.1 for a TFS, besides replacing $\hat{\Pi}$ with $\hat{\Pi}_i$, and $\hat{\Pi}^b$ with $\hat{\Pi}_i^b$, and re-deriving the terms involving these, there are no essential differences between the proofs for a LFS and a TFS, and for brevity we omit these proofs. ■