Testing for a Threshold in Models with Endogenous Regressors

Rothfelder, Mario; Boldea, Otilia

Publication date:
2019

Document Version
Early version, also known as pre-print

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Rothfelder, M., & Boldea, O. (2019). *Testing for a Threshold in Models with Endogenous Regressors*. (Center Discussion Paper; Vol. 2019-030). CentER, Center for Economic Research.

CentER

Discussion Paper

No. 2019-030

TESTING FOR A THRESHOLD IN MODELS WITH ENDOGENOUS REGRESSORS

By

Mario P. Rothfelder, Otilia Boldea

30 October 2019

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

Testing for a Threshold in Models with Endogenous Regressors

Mario P. Rothfelder[†] and Otilia Boldea[‡]

[†] *Tilburg University, Dept. of Econometrics & OR and CentER*
E-mail: m.p.rothfelder@tilburguniversity.edu
[‡] *Tilburg University, Dept. of Econometrics & OR and CentER*
E-mail: o.boldea@tilburguniversity.edu

Summary We propose two new parametric tests for an unknown threshold in models with endogenous regressors. They are both based on unconventional 2SLS estimators that use additional information about the linearity of the first stage. This information leads to more accurate residuals and therefore tests with better size properties than the Wald GMM test in Caner and Hansen (2004), which we show exhibits severe size distortions in small samples pertinent to empirical applications.

We prove the bootstrap validity of our tests and evaluate their empirical relevance by revisiting the question whether government spending multipliers are larger in recessions. As Ramey and Zubairy (2018), we cannot rule out that they are the same in recessions or expansions.

Keywords: *2SLS, GMM, threshold, bootstrap, government spending*

1. INTRODUCTION

Threshold models are widely used in economics to model unemployment, output, growth, bank profits, asset prices, exchange rates, and interest rates; see Hansen (2011) for a survey of economic applications. While threshold models with exogenous regressors have been widely studied and their asymptotic properties are well known[1], the literature on threshold models with endogenous regressors remains relatively scarce.[2] Nevertheless, in many applications, the regressors are *endogenous*. For example, Ramey and Zubairy (2018) (RZ henceforth) recently used a threshold model with endogenous regressors to show that the US cumulative government spending multiplier is not necessarily larger in recession regimes defined by high unemployment. One of the main questions of interest in their paper is whether there are two government spending multiplier regimes defined by the unemployment rate being below or above a threshold. This question can be an-

[1]See *interalia* 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 (2016) for threshold regressions with a kink.

[2]For some contributions with endogenous regressors, see *interalia*: for time-series, Caner and Hansen (2004), who consider exogenous threshold variables and Kourtellos 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. (2019), who consider endogenous threshold variables.

swered by testing the null hypothesis of *an unknown* threshold in models with endogenous regressors, as in general, the threshold is not known to the researcher *apriori*.[3]

Caner and Hansen (2004) (CH henceforth) proposed a Wald test for exactly this hypothesis based on GMM (generalized method of moments) estimation.[4] This test computes the GMM estimators before and after each candidate threshold values between, say, the 15% and 85% quantiles of the threshold variable that may drive the two regimes, and compares the two estimators by computing the maximum over the Wald tests for each of these candidate threshold values. Unfortunately, we find in simulations that this test has serious size distortions, with rejection frequencies up to three times the nominal size in small samples pertinent to applications (see tables 1 and 2), and erratic behavior in our empirical application (see figure 2). Tables 1 and 2 show that these size distortions are already present in just-identified models with strong instruments, even with homoskedastic data, and even though the test is bootstrapped under the null hypothesis. Because they do not occur with homoskedastic variance estimators, we find that the source of size distortions is inaccurate heteroskedasticity-robust variance estimators when using data only at the sample edges (e.g. below the 15% quantile or above the 85% quantiles). The Online Supplement to this paper shows that the entries of these heteroskedasticity-robust estimators can be more than a hundred times larger that their respective limits and five times smaller than their bootstrap equivalents in small samples.

Since the use of heteroskedasticity-robust variance estimators cannot be avoided for optimal GMM estimation, in this paper, we focus instead on 2SLS (two-stage least squares) estimation. The 2SLS estimators we compute are not conventional and therefore not a special case of the GMM estimators used in the CH Wald 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. Using this information results in more accurate residuals and sample variances for data at the sample edges. Therefore, the two test statistics we propose - a 2SLS likelihood ratio test and a 2SLS Wald test - fix the small sample issues associated with the GMM test.

As for the GMM test, the null asymptotic distributions of our tests depend on the data generating process. We propose obtaining critical values via the wild bootstrap to mimic potential heteroskedasticity in the initial sample. We prove the asymptotic validity of the two bootstrap test statistics proposed and show via simulations that they do not have similar size distortions to the GMM test in small samples. Since we find no systematic difference between the two 2SLS tests, we conclude that both are valuable diagnostics for the existence of an unknown threshold.

Our paper is closely related to several papers in the change-point literature. Boldea et al. (2019) study the equivalent of the two test statistics in 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. 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

[3]RZ fix this threshold throughout most of their analysis at an unemployment rate of 6.5%, based on the Federal Reserve's use of this threshold in a policy announcement, and later do robust checks with different fixed or time-varying thresholds.

[4]This is the only test that we are aware of that is specific to parametric time series threshold models with endogenous regressors and an exogenous threshold variable.

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 *interalia* Kourtellos et al. (2015), Yu and Phillips (2018) and Christopoulos 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. Based on the tests proposed, we find evidence that the instantaneous government spending multiplier is different in recessions, but as RZ, we cannot rule out that the cumulative government spending multipliers are the same in recessions and expansions due to relatively weaker instruments in the recession regime. However, we do estimate the threshold unemployment rate to be 8.3%, rather than 6.5% as imposed in RZ (or than the 8% which RZ employ as a robustness check.) This new threshold (but also 8%) causes the military spending instrument constructed in RZ to become weaker for deep recessions, suggesting that it is probably most informative at moderate unemployment rates somewhere between 6.5% and 8.3%.

The paper is organized as follows. Section 2 describe the model and test statistics. Section 3 describes the bootstrap. Section 4 contains the assumptions and all the bootstrap validity results. Section 5 contains simulations and section 6 contains the empirical application. Section 7 concludes. Appendix A contains all the tables and figures, and Appendix B all the proofs. An Online Supplement contains additional results in support of sections 6 and 7.

2. THRESHOLD MODEL AND TEST STATISTICS

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

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

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

$$x_t = \Pi^{0\top} z_t + u_t, \tag{2}$$

or a threshold model:

$$x_t = \Pi_1^{0\top} z_t \mathbf{1}[q_t \le \rho^0] + \Pi_2^{0\top} z_t \mathbf{1}[q_t > \rho^0] + u_t, \tag{3}$$

where $\rho^0 \in \Gamma$, and z_t are $q \times 1$ strong and valid instruments, including z_{1t}, with $q \geq p_1$. We assume that $E((\epsilon_t, u_t)|\mathcal{F}_t) = 0$, where $\mathcal{F}_t = \{q_t, q_{t-1}, \ldots, z_t, z_{t-1}, \ldots, x_{t-1}, \ldots\}$, so that equation (1) can be estimated by either 2SLS or by GMM.

For the theory contribution, we assume that the researcher knows whether there is a linear first stage (LFS) or a threshold first stage (TFS), but in general this is not known and can be determined by testing for a threshold in the first stage (see Hansen (1996)) or by using the BIC criteria in Gonzalo and Pitarakis (2002).

We are interested in testing for an unknown threshold, i.e. $\mathbb{H}_0 : \theta_1^0 = \theta_2^0 \equiv \theta^0$. Such a test is already available from CH.[5] They proposed a test based on GMM estimators of $\theta_i^0, (i = 1, 2)$ for each $\gamma \in \Gamma$. Because z_t and q_t are exogenous, the moment conditions

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

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

$$\hat{\theta}_{i\gamma,(2)} = (\hat{N}_{i\gamma}\hat{\Omega}_{i\gamma}^{-1}\hat{N}_{i\gamma}^\top)^{-1}\hat{N}_{i\gamma}\hat{\Omega}_{i\gamma}^{-1}(T^{-1}\textstyle\sum_{i\gamma} z_t y_t),$$

where $\hat{N}_{i\gamma} = T^{-1}\sum_{i\gamma} w_t z_t^\top$, $\hat{\Omega}_{i\gamma} = T^{-1}\sum_{i\gamma}\hat{\epsilon}_{t,(1)}^2 z_t z_t^\top$, $\sum_{1\gamma}(\cdot) = \sum_{t=1}^T(\cdot)\mathbf{1}[q_t \leq \gamma]$, $\sum_{2\gamma}(\cdot) = \sum_{t=1}^T(\cdot)\mathbf{1}[q_t > \gamma]$, $\hat{\epsilon}_{t,(1)} = y_t - w_t^\top\hat{\theta}_{1\gamma,(1)}\mathbf{1}[q_t \leq \gamma] - w_t^\top\hat{\theta}_{2\gamma,(1)}\mathbf{1}[q_t > \gamma]$ are the first step GMM residuals, and $\hat{\theta}_{i\gamma,(1)}$ are consistent first-step versions of $\hat{\theta}_{i\gamma,(2)}$, for example by replacing $\hat{\Omega}_{i\gamma}$ 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 maximum of this sequence of Wald tests over $\gamma \in \Gamma$ yields the test in CH:

$$WG_T = \sup_{\gamma\in\Gamma} WG_T(\gamma) = \sup_{\gamma\in\Gamma} T\left(\hat{\theta}_{1\gamma,(2)} - \hat{\theta}_{2\gamma,(2)}\right)^\top \hat{V}_{\gamma,(1)}^{-1}\left(\hat{\theta}_{1\gamma,(2)} - \hat{\theta}_{2\gamma,(2)}\right), \tag{5}$$

where $\hat{V}_{\gamma,(1)} = \sum_{i=1}^2 \hat{V}_{i\gamma,(1)}$ and $\hat{V}_{i\gamma,(1)} = (\hat{N}_{i\gamma}\hat{\Omega}_{i\gamma}^{-1}\hat{N}_{i\gamma}^\top)^{-1}$.

As shown in the simulation section 5, this test has serious size distortions for sample sizes often encountered in macroeconomic data. Even though CH prove bootstrap validity of their test in large samples, Tables 1-2 shows that the bootstrap does not replicate well the empirical distribution of the test statistic in finite samples.

Note that both for a LFS and a TFS, the moment conditions (4) are still valid, but they do not reflect the additional information that the instruments are equally strong over the entire sample (for a LFS), or they change strength over the sample (for a TFS). In both cases, it may be possible to construct more efficient GMM estimators which use this information, however the optimal GMM estimation require the use of heteroskedasticity-robust variance estimators before and after each γ, leading precisely to the sample edge inaccuracies in these estimators that we want to avoid.

For this reason and also because the 2SLS estimation we consider is a necessary step if one also wants a consistent estimate of the threshold γ^0 under the alternative hypothesis $\mathbb{H}_A : \theta_1^0 \neq \theta_2^0$, the paper focuses on 2SLS estimation of equations (1) and (2) or (1) and (3).[6]

[5] They only considered a model with endogenous regressors, but their results straightforwardly generalize to the case of both endogenous and exogenous regressors.

[6] The only available parametric method for consistently estimating γ^0 in (1) we are aware of was proposed in CH and is based on 2SLS. Note that it is not known whether the implicit maximizer of $WG_T(\gamma)$ is a consistent estimator of γ^0.

The LFS (2) is estimated over the full sample:

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

For the TFS (3), we first estimate the threshold ρ^0 as in CH, by multivariate least-squares:

$$\hat{\rho} = \arg\min_{\rho\in\Gamma} \operatorname{trace} \sum_{t=1}^{T} \left[(x_t - \hat{\Pi}_{t\rho}^\top z_t)(x_t - \hat{\Pi}_{t\rho}^\top z_t)^\top\right],$$

where $\hat{\Pi}_{i\rho} = (\sum_{i\rho} z_t z_t')^{-1} \sum_{i\rho} z_t x_t^\top$ for $i = 1, 2$, and $\hat{\Pi}_{\rho t} = \hat{\Pi}_{1\rho}\mathbf{1}[q_t \le \rho] + \hat{\Pi}_{2\rho}\mathbf{1}[q_t > \rho]$. At this threshold estimator,

$$\hat{\Pi}_i = \hat{\Pi}_{i\hat{\rho}}, \qquad \hat{x}_t = \hat{\Pi}_{\hat{\rho}t}^\top z_t.$$

With this notation, in both cases, the 2SLS estimators $\hat{\theta}_{i\gamma}$ of equation (1) - over all $\gamma \in \Gamma$ - are[7]

$$\hat{w}_t = (\hat{x}_t^\top, \hat{z}_{1t}^\top)^\top, \qquad \hat{\theta}_{i\gamma} = \left(\sum_{i\gamma} \hat{w}_t \hat{w}_t^\top\right)^{-1} \sum_{i\gamma} \hat{w}_t y_t.$$

We now propose two tests for $\mathbb{H}_0 : \theta_1^0 = \theta_2^0$ and an unknown threshold. The first one is a likelihood-ratio type test:

$$LR_T = \sup_{\gamma\in\Gamma} LR_T(\gamma) = \sup_{\gamma\in\Gamma} \frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/(T-2p)} \tag{6}$$

Here, $SSR_0 = \sum_{t=1}^{T}(y_t - \hat{w}_t^\top \hat{\theta})$, 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, and $SSR_1(\gamma) = \sum_{i=1}^{2} \sum_{i\gamma}(y_t - \hat{w}_t^\top \hat{\theta}_{i\gamma})^2$. This test has the advantage that the implicit maximum is achieved at $\hat{\gamma}$, which is a consistent estimator of γ^0 in equation (1) under the alternative $\mathbb{H}_A : \hat{\theta}_1 = \hat{\theta}_2$, as shown in CH.

The second one is a Wald-type test:

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

where $\hat{V}_\gamma \xrightarrow{\text{p}} V_\gamma = \lim \operatorname{Var}\left[T^{1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma})\right]$, and the explicit expression for $\hat{V}_\gamma$ and V_γ can be found in Appendix B, Definition B.1 for a LFS and Appendix B, Definition B.2 for a TFS. 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 $\hat{V}_\gamma$ 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.

Since the asymptotic distribution of both 2SLS tests is shown in Appendix B, Theorem B.1 to depend on moments and parameters of the data generating process, we propose bootstrapping these tests and prove the asymptotic validity of the bootstrap. The GMM test was also shown to be non-pivotal in CH, and they also proposed a (variant of) bootstrap which they showed to be asymptotically valid. All these bootstraps are described in the next section.

[7] In the computation, one does not search over all possible values in Γ, but only consider unique values of q_t taken in the sample, from the η to the $(1-\eta)$ quantile, where typically $\eta \in [0.10, 0.25]$.

3. BOOTSTRAP

Motivated by the good performance of the wild bootstrap for 2SLS change-point tests in Boldea et al. (2019), for our 2SLS tests, we propose using the fixed regressor wild bootstrap, where the instruments and the exogenous regressors are kept fixed, and the residuals are multiplied with the i.i.d. variable η_t described in Assumption 4.2 to preserve the potential conditional heteroskedasticity in the data.[8] A bootstrap sample is computed always under $\mathbb{H}_0 : \theta_1^0 = \theta_2^0$, as follows:

- estimate a LFS or TFS (with $\hat{\rho}$ fixed), and compute the full-sample 2SLS estimates $\hat{\theta} = (T^{-1} \sum_{t=1} \hat{w}_t \hat{w}_t^\top)^{-1} T^{-1} \sum_{t=1} \hat{w}_t y_t$;
- obtain the residuals $\hat{u}_t = x_t - \hat{x}_t$ and $\hat{\epsilon}_t = y_t - w_t^\top \hat{\theta}$;
- let $u_t^b = \hat{u}_t \eta_t$ and $\epsilon_t^b = \hat{\epsilon}_t \eta_t$, where $\eta_t \sim i.i.d.(0, 1)$
- let $x_t^b = \hat{\Pi}_t^\top z_t + v_t^b$, $w_t^b = (x_t^{b\top}, z_{1t}^\top)^\top$ and $y_t^b = w_t^b \hat{\theta} + u_t^b$.

Let $\hat{\Pi}_t^b$ be computed as $\hat{\Pi}_{\hat{\rho}t}$ but with the bootstrap data, so either under a LFS, in which case $\hat{\Pi}_t^b = \hat{\Pi}^b = (\sum_{t=1}^T z_t z_t^\top)^{-1} \sum_{t=1}^t z_t x_t^{b\top}$, or under a TFS, in which case $\hat{\Pi}_t^b = \mathbf{1}[q_t \leq \hat{\rho}]\Pi_1^b + \mathbf{1}[q_t > \hat{\rho}]\Pi_2^b$, and $\hat{\Pi}_i^b = (\sum_{i\hat{\rho}} z_t z_t^\top)^{-1} \sum_{i\hat{\rho}} z_t x_t^{b\top}$ for $i = 1, 2$.

Let $\hat{\theta}_{i\gamma}^b = (\sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b^\top})^{-1} \sum_{i\gamma} \hat{w}_t^b y_t^b$, where $\hat{x}_t^b = \hat{\Pi}_t^{b\top} z_t$ and $\hat{w}_t^b = (\hat{x}_t^{b\top}, z_{1t}^\top)^\top$, and let $\hat{\theta}^b = (\sum_{t=1}^T \hat{w}_t^b \hat{w}_t^{b^\top})^{-1} \sum_{t=1}^T \hat{w}_t^b y_t^b$.

Then the bootstrap equivalent of $LR_T = \sup_{\gamma\in\Gamma} LR_T(\gamma)$ is $LR_T^b = \sup_{\gamma\in\Gamma} LR_T^b(\gamma)$, where $LR_T^b(\gamma)$ is computed in the same way as $LR_T(\gamma)$, but with SSR_0 replaced by $SSR_0^b = \sum_{t=1}^T (y_t^b - \hat{w}_t^{b\top} \hat{\theta}^b)^2$ and $SSR_1(\gamma)$ by $SSR_1^b(\gamma) = \sum_{i=1}^2 \sum_{i\gamma} (y_t^b - \hat{w}_t^{b\top} \hat{\theta}_{i\gamma}^b)^2$.

The bootstrap equivalent of $W_T = \sup_{\gamma\in\Gamma} W_T(\gamma)$ is $W_T^b = \sup_{\gamma\in\Gamma} W_T^b(\gamma)$, where $W_T^b(\gamma) = T(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b)^\top (\hat{V}_\gamma^b)^{-1} (\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b)$, with $\hat{V}_\gamma^b$ obtained as $\hat{V}_\gamma$ in Appendix B, Definition B.1 for a LFS or Definition B.2 for a TFS, but replacing all sample estimators by their bootstrap equivalents, except for $\hat{\rho}$, which is kept fixed throughout the bootstrap samples for a TFS.

We now describe the bootstrap in CH, to which we make two slight modifications which have no asymptotic consequence for their test, but improve its finite sample performance under $\mathbb{H}_0$. A bootstrap sample is computed as follows:

- obtain the full-sample two-step GMM estimator
$$\hat{\theta}_{(2)} = (\hat{N}\hat{\Omega}^{-1}\hat{N}^\top)^{-1}\hat{N}\hat{\Omega}^{-1}(T^{-1}\sum_{t=1}^T z_t y_t),$$
with $\hat{\Omega} = T^{-1} \sum_{t=1}^T z_t z_t^\top \check{\epsilon}_{t,(1)}^2$, $\check{\epsilon}_{t,(1)} = y_t - w_t^\top \hat{\theta}_{(1)}$, where $\hat{\theta}_{(1)}$ is obtained as $\hat{\theta}_{(2)}$ but replacing $\hat{\Omega}$ with $T^{-1} \sum_{t=1}^T z_t z_t^\top$, and $\hat{N} = T^{-1} \sum_{t=1}^T w_t z_t^\top$;
- obtain the two-step residuals $\check{\epsilon}_{t,(2)} = y_t - w_t^\top \hat{\theta}_{(2)}$;
- let $\epsilon_t^b = \check{\epsilon}_{t,(2)} \eta_t$, where $\eta_t \sim i.i.d.(0, 1)$;
- let $y_t^b = w_t^\top \hat{\theta}_{(2)} + \epsilon_t^b$.

[8] It is perhaps possible to design a recursive wild bootstrap method for threshold models, but as shown in Boldea et al. (2019), it would come at the cost of stronger moment assumptions. One could also consider a pairs bootstrap, but as shown in MacKinnon (2009), it does not necessarily outperform the wild bootstrap in small samples.

Note that the two changes to CH are: (1) that we bootstrap under the null rather than the alternative, and therefore use $\check{\epsilon}_{t,(1)}$ instead of $\hat{\epsilon}_{t,(1)}$ and $\check{\epsilon}_{t,(2)}$ instead of $\hat{\epsilon}_{t,(2)} = y_t - w_t^\top(\hat{\theta}_{1\gamma}\mathbf{1}[q_t \leq \gamma] + \hat{\theta}_{2\gamma}\mathbf{1}[q_t > \gamma])$; (2) that we add back the conditional mean for y_t^b instead of writing $y_t^b = \epsilon_t^b$, for the purposes of better replicating the variation in y_t. Then the bootstrap equivalent of the GMM test is $WG_T^b = \sup_{\gamma\in\Gamma} WG_T^b$, with

$$WG_T^b = T\left(\hat{\theta}^b_{1\gamma,(2)} - \hat{\theta}^b_{2\gamma,(2)}\right)^\top (\hat{V}^b_{\gamma,(1)})^{-1}\left(\hat{\theta}^b_{1\gamma,(2)} - \hat{\theta}^b_{2\gamma,(2)}\right),$$

where $\hat{\theta}^b_{i\gamma,(2)} = (\hat{N}_{i\gamma}\hat{\Omega}^{b-1}_{i\gamma}\hat{N}^\top_{i\gamma})^{-1}\hat{N}_{i\gamma}\hat{\Omega}^{b-1}_{i\gamma}(T^{-1}\sum_{i\gamma} w_t y_t^b)$, $\hat{V}^b_{\gamma,(1)} = \sum_{i=1}^2 \hat{V}^b_{i\gamma,(1)}$, $\hat{V}^b_{i\gamma,(1)} = (\hat{N}_{i\gamma}(\hat{\Omega}^b_{i\gamma})^{-1}\hat{N}^\top_{i\gamma})^{-1}$, $\hat{\Omega}^b_{i\gamma} = T^{-1}\sum_{i\gamma}(\hat{\epsilon}^b_{t,(1)})^2 z_t z_t^\top$ and $\hat{\epsilon}^b_{t,(1)} = y_t^b - w_t^\top\hat{\theta}^b_{1\gamma,(1)}\mathbf{1}[q_t \leq \gamma] - w_t^\top\hat{\theta}^b_{2\gamma,(1)}\mathbf{1}[q_t > \gamma]$.

Note that this bootstrap keeps all regressors x_t, z_t, q_t fixed, and only bootstraps y_t, while our test also bootstraps x_t; the latter cannot be avoided for 2SLS.

4. BOOTSTRAP VALIDITY

Define $g_t = x_t - u_t$, $h_t = y_t - \epsilon_t$, $M_1(\gamma) = E[z_t z_t^\top \mathbf{1}[q_t \leq \gamma]]$, $M = M(\gamma_{\max}) = E[z_t z_t^\top]$, where $\Gamma^0 = [\gamma_{min}, \gamma_{max}]$ is the support of q_t, $M_2(\gamma) = M - M_1(\gamma)$, and $v_t = (\epsilon_t, u_t^\top)^\top$. Let $\|\cdot\|$ be the Euclidean norm. The following assumptions are similar to CH.

Assumption 4.1.

(a) $\mathbb{E}[v_t|\mathfrak{F}_t] = 0$ *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) $\sup_t \mathbb{E}\|z_t\|^{4r} < \infty, \sup_t \mathbb{E}\|v_t\|^{4r} < \infty$;
(d) $\inf_{\gamma\in\Gamma^0} \det M_{1\gamma} > 0$, *and if* (3) *holds, then* $M_{1\gamma}$ *is strictly increasing in* γ *(strictly increasing here means that if* $\gamma_1 > \gamma_2 > 0$*, then* $M_{1\gamma_1} - M_{1\gamma_2}$ *is p.d.);*
(e) The threshold variable q_t *has a continuous pdf* $f(q_t)$ *with* $\sup_{q_t\in\Gamma^0} |f(q_t)| < \infty$;
(f) $\mathbb{E}[v_t v_t^\top]$ *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 LFS* (2)*) or* Π_1^0, Π_2^0 *(for the TFS* (3)*) are full rank, and* $\Pi_1^0 - \Pi_2^0 \neq 0$.

Assumption 4.1(a) is typically needed for nonlinear models, as discussed in CH. Assumption 4.1(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 B.1-B.2 in the Appendix). This assumption is also in CH. Assumption 4.1(c) is a typical moment condition similar to CH. Assumption 4.1(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. This assumption is used to provide a self-contained proof of super-consistency of $\hat{\rho}$ in a TFS, and is further discussed after Theorem 4.1. Assumptions 4.1(e) and 4.1(g) are also imposed in CH, and Assumption 4.1(f) is needed for uniqueness of the asymptotic distributions of the test statistics proposed.

With this assumption, we first show super-consistency of $\hat{\rho}$, which allows one to treat the first stage threshold as if it was known in subsequent analysis. This result is also needed for deriving the asymptotic distribution of the proposed tests and their bootstrap versions.

THEOREM 4.1. *Under Assumption 4.1 and the TFS in* (3), *(i)* $\hat{\rho}-\rho^0 \xrightarrow{\text{P}} 0$*; (ii)* $T(\hat{\rho}-\rho^0) = \mathcal{O}_p(1)$*; (iii)* $T^{1/2}\,\text{vec}(\hat{\Pi}_i - \Pi_i^0) = T^{1/2}\,\text{vec}(\hat{\Pi}_{i\rho^0} - \Pi_i^0) + o_p(1)$.

This theorem was also shown in Chan (1993), under different assumptions that pertain directly to a threshold autoregressive model where z_t, q_t are lags of x_t. CH state that under their assumptions, the result in Theorem 4.1 holds, but they do not provide an explicit proof. We therefore provide a self-contained proof of this theorem in Appendix B. This proof may be of interest in its own right, as it extends proof techniques from change point analysis to threshold models.

We require the following assumption for proving bootstrap validity.

ASSUMPTION 4.2. *(a)* $\eta_t \overset{iid}{\sim} (0,1)$ *with* $\sup_t E^b(\eta_t^4) < \infty$, *where* $E^b(\cdot)$ *is the expectation with respect to the bootstrap probability measure; (b)* $\sup_t E||z_t||^{4r/(r-1)} = o(T)$.

Assumption 4.2(a) is common for wild bootstraps (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 Wald GMM test. Assumption 4.2(b) is only needed for the W_T^b test, in particular for $\hat{V}_\gamma^b$ to weakly converge to V_γ in probability under the bootstrap measure.

THEOREM 4.2. (BOOTSTRAP VALIDITY) *Let* y_t *be generated by* (1) *and* x_t *be generated by the LFS* (2) *or by the TFS* (3). *Then, under* $\mathbb{H}_0$,
(i) $\sup_{c\in\mathbb{R}} \left|P^b\left(|LR_T^b - LR_T| \le c\right)\right| \xrightarrow{\text{P}} 0$ *under Assumptions 4.1-4.2(a);*
(ii) $\sup_{c\in\mathbb{R}} \left|P^b\left(|W_T^b - W_T| \le c\right)\right| \xrightarrow{\text{P}} 0$, *under Assumptions 4.1-4.2.*

This theorem proves bootstrap validity of the proposed test statistics. The bootstrap validity of WG_T^b was also explained in CH under their Assumption 1, Assumption 2 in Hansen (1996) and $\eta_t \overset{iid}{\sim} \mathcal{N}(0,1)$, although from Hansen (1996) it is clear that the normality assumption is not necessary for their results to go through. All of these are comparable to our Assumptions 4.1-4.2. In the next section, we compare the three test statistics through simulations for which both our assumptions and their assumptions are satisfied.

5. SIMULATIONS

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

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

where $z_t \stackrel{iid}{\sim} \mathcal{N}(1,1)$, $q_t = z_t + 1$, and z_t, x_t, q_t are scalars. We set $\delta_\Pi = 0$ for a LFS and $\delta_\Pi \in \{-0.5, 0.5, 1\}$ for a TFS. For a TFS, $\rho^0 = 1.75$. Under the null, $\delta_x = 0$, and under the alternative, $\delta_x = 0.25$ with $\gamma^0 = 2.25$.[9]

Let $\nu_t \stackrel{iid}{\sim} \mathcal{N}(0,1)$. We consider three cases. In case (a), the errors are homoskedastic - $\epsilon_t = \nu_t$, we use the i.i.d. bootstrap instead of the wild bootstrap, and make two adjustments to the computation of the test statistics. First, $v_t^b \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \hat{\Sigma}_v)$ with $\hat{\Sigma}_v = T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top$ for 2SLS. For GMM, $\epsilon_t^b \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \hat{\sigma}_\epsilon^2)$ with $\hat{\sigma}_\epsilon^2 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{t,(1)}^2$, and $\hat{\epsilon}_{t,(1)} = y_t - w_t^\top \hat{\theta}$, where recall that $\hat{\theta}$ is the full-sample 2SLS estimator. Second, all heteroskedasticity-robust estimators are replaced by their homoskedastic analogs. For example, $E[z_t z_t^\top \epsilon_t^2 \mathbb{1}\{q_t \leq \gamma\}]$ is no longer estimated by $T^{-1} \sum_{1\gamma} z_t z_t^\top \hat{\epsilon}_t^2 \mathbf{1}[q_t \leq \gamma]$, but by $\hat{\sigma}_\epsilon^2 (T^{-1} \sum_{1\gamma} z_t z_t^\top \mathbf{1}[q_t \leq \gamma])$.

In case (b), the errors are still homoskedastic - $\epsilon_t = \nu_t$, but we assume we do not know this, and therefore use the heteroskedasticity-robust variance estimators described in sections 2 and 3. In case (c), the errors are conditional heteroskedastic - $\epsilon_t = \nu_t \cdot z_t/\sqrt{2}$ with $Var(\nu_t) = Var(u_t) = 1$ and $Cov(u_t, \nu_t) = 0.5$, and we use the same heteroskedasticity-robust variance estimators. In table 1 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 3.

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 and 2), and under the alternative we plot the size-adjusted power, where the size-adjustment is made relative to the null DGPs described above (figure 1).

Tables 1 and 2 show that our tests have in all cases close to nominal sizes, even in small samples. Under known conditional homoskedasticity, the GMM test is also close to its nominal size. However, as soon as we correct for (potential) heteroskedasticity, it become oversized in small samples, with empirical sizes more than twice the nominal size. This means that it will overreject the null of no threshold and cannot be trusted even for sample sizes around $T = 500$, unless one truly believes the data is conditional homoskedastic. This problem is mitigated in some cases by the use of the asymmetric distribution in Mammen (1993) instead of the normal distribution, but the test still displays sizes of up to 15.5% at a 5% nominal size, as shown in table 2.

The size distortions of the GMM test originate from imprecise heteroskedasticity-robust estimates $\hat{\Omega}_{i\gamma}$ 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 precisely estimate $\Omega_{i\gamma}$. In the Online supplement, Section S1, we compute $\Omega_{i\gamma}$, $\hat{\Omega}_{i\gamma}$ and $\hat{\Omega}_{i\gamma}^b$ for a sample size of $T = 100$ and case (c). We show that the entries of $\hat{\Omega}_{i\gamma}$ can be 100 times larger than their limits in $\Omega_{i\gamma}$, and also twice as large or five times smaller than

[9] Note 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 $\hat{\theta}_{i\gamma,(2)} = \hat{\theta}_{i\gamma,(1)} = (\hat{N}_{i\gamma}^\top)^{-1} (T^{-1} \sum_{i\gamma} z_t y_t)$. Also, $\hat{V}_{\gamma,(1)} = \sum_{i=1}^2 \hat{V}_{i\gamma,(1)}$, $\hat{V}_{i\gamma,(1)} = (\hat{N}_{i\gamma} \hat{\Omega}_{i\gamma}^{-1} \hat{N}_{i\gamma}^\top)^{-1}$ and $\hat{\epsilon}_{t,(2)} = \hat{\epsilon}_{t,(1)} = y_t - w_t^\top \hat{\theta}_{1\gamma,(1)} \mathbf{1}[q_t \leq \gamma] - w_t^\top \hat{\theta}_{2\gamma,(1)} \mathbf{1}[q_t > \gamma]$.

their bootstrap equivalents in $\hat{\Omega}^b_{i\gamma}$ when γ is at the edges of the sample, but that the same entries are approximately equal to their limits and close to their bootstrap equivalents if γ is equal to the median.[10] This is precisely why the bootstrap does not replicate well the empirical distribution of the GMM test in small samples.

Note that this is exactly the problem that our tests mitigate, because for γ close the edges of the sample, the 2SLS residuals are estimated using additional data from the first stage, either from the full-sample in the case of a LFS, or from a larger sub-sample dictated by the TFS, improving their finite sample properties.

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 1 shows the power properties for a small threshold of $\delta_x = 0.25$. In small samples, the Wald tests dominates 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. For the GMM test, the power is suspiciously large for a small threshold, and it is probably due to the same inaccuracies of $\hat{\Omega}_{i\gamma}$ at the edges of the sample mentioned earlier. However, the power differences among all three tests vanish as the sample size grows.

Therefore, we argue that our tests provide a reliable alternative to the GMM test in smaller samples pertinent to macroeconomic applications.

6. 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.[11] For simplicity, we first focus on the instantaneous government spending multiplier $\theta_{g,i}(i = 1, 2)$, estimated similarly to RZ from:

$$y_t = (\theta_{g,1}\, g_t + z_{1,t}^\top \theta_{z,1})\mathbf{1}[q_t \leq \gamma^0] + (\theta_{g,2}\, g_t + z_{1,t}^\top \theta_{z,2})\mathbf{1}[q_t > \gamma^0]) + \epsilon_t \tag{8}$$

$$g_t = \Pi_1^\top z_t \mathbf{1}[q_t \leq \rho^0] + \Pi_2^\top z_t \mathbf{1}[q_t > \rho^0] + v_t \tag{9}$$

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 which can be found at `http://econweb.ucsd.edu/~vramey/research/Ramey_Zubairy_replication_codes.zip`. For details on the data construction, the validity of instruments, or the interpretation of $\theta_{g,i}(i = 1, 2)$ as cumulative spending multipliers, we refer the interested reader to RZ.

[10] These size distortions persist when using other heteroskedasticity-robust variance estimators such as the HC1-HC3 described in Davidson and MacKinnon (1993). HC1 is just a degree of freedom adjustment that leads to exactly the same results, because the test statistics and the bootstrap critical value are multiplied by the same constant. With the same DGP as in table 2, we obtain sizes up to 13.4% for $T = 100$ and 12.5% for $T = 250$ using HC2, and sizes up to 13.7% for $T = 100$ and 12.5% for $T = 250$ using HC3.

[11] An earlier draft of this paper, Rothfelder and Boldea (2016), answered the same question but only with more recent data. RZ explained that this data does not vary enough with the business cycles to identify changes in government spending multipliers. We are grateful for the rich dataset they constructed and we redid the empirical analysis in Rothfelder and Boldea (2016) using their data and approach.

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, but evaluated in RZ at $\gamma = 6.5$ (and ignoring the first stage which is irrelevant for conventional IV estimators).[12] 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.

Our methods require first estimating ρ^0 in equation (9).[13] Table 3 reports the multivariate threshold estimates $\hat{\rho}$ described in Section 2, 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 3, which were proposed in Hansen (1996). The estimate of ρ^0 and the decisions change with the cut-off considered. But if there is a threshold, the maximizer of the OLS version of $LR_T(\gamma)$ is a consistent estimator of ρ^0. This is because the likelihood ratio test is maximized exactly at $\hat{\rho}$, the minimizer of the multivariate sum of squared residuals described in section 2, which is consistent as shown in Theorem 4.1. Therefore, we use a TFS with $\hat{\rho}$ in table 3.

Given $\hat{\rho}$ obtained for each cut-off, we test for an unknown threshold in equation (8). Table 4 shows that the LR test clearly rejects the null. The 2SLS Wald test does not always reject but its values are relatively close to the critical values at certain cut-offs. The GMM Wald test only rejects with 25% trimming, but both the test and its critical values are very large at 10% trimming, and these may be due to its distortions around cut-offs discussed in the simulation section and further illustrated in figure 2, where the sample-edge erratic behavior of the test is apparent.

Because equations (8)-(9) 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 (equivalently, the implicit maximizer of the $LR_T(\gamma)$ quantity in this paper).[14]

Our empirical illustration could conclude that the instantaneous government spending multiplier is different in recessions, and that the threshold defining recessions is larger than 6.5. However, we now show results that cast doubt on this conclusion and further explain why both RZ and this paper cannot rule out that multipliers are the same in recessions and expansions. As in RZ, we compute the cumulative government spending

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

[13] RZ do not further investigate the regimes in equation (9) because they are irrelevant for the IV estimation. For IV estimation, one can treat equation (9) as a mere projection that helps estimate $\theta_{g,i}$. We treat equation (9) as the "true specification" for testing for a threshold in instantaneous government multipliers, although it could also be viewed as a projection, as long as it satisfies Assumption 4.1.

[14] The 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.

multipliers $\theta^h_{g,i}(i = 1, 2)$ at horizon $h = 0, 1, \ldots, H$ from the IV regression:

$$\sum_{h=0}^{H} y^h_{t+h} = (\theta^h_{g,1} \sum_{h=0}^{H} g_{t+h} + z_{1,t}^\top \theta^h_{z,1})\mathbf{1}[q_t \le \hat{\gamma}] + (\theta^h_{g,2} \sum_{h=0}^{H} g_{t+h} + z_{1,t}^\top \theta^h_{z,2})\mathbf{1}[q_t > \hat{\gamma}] + \epsilon_t,$$

where $\sum_{h=0}^{H} g_{t+h}$ is instrumented by m_t.[15]

Tables 5 and 6 show the RZ multipliers and our multipliers for two years 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.[16] These tables show that in both cases, there is no evidence that government spending multipliers are different in recessions, once the possibility of weak instruments is taken into account.

We therefore plot in figure 3 a weak instrument test across horizons (the effective F-statistic for the null hypothesis of weak instruments in each regime) for both thresholds.[17] These figures show evidence of weak instruments in both regimes at horizon $h = 0$, and for both our threshold and the RZ threshold. This also holds for the effective F-statistics for our TFS specification with $\hat{\rho} = 4.0636$: they are equal to approximately -19 for $q_t \le 4.0636$ (101 observations), and -17.5 for $q_t > 4.0636$ (399 observations), so well below zero. Therefore, the weak instrument problem hampers further conclusions about the difference in government spending multipliers 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 the usefulness of this variable as an instrument for government spending is not robust to the threshold used.[18] This is also indicated in figure 4, 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. Apart from small sample and specification issues, the results suggest that the RZ military news instrument is more informative for intermediate values of unemployment, so for "normal" recessions rather than "deep" recessions.

7. CONCLUSIONS

In this paper we proposed two new test statistics for threshold detection in linear models with endogenous regressors and exogenous thresholds. Even though the paper is derived for time series models, our assumptions indicate that it can also be applied to cross-sectional models. We showed that our test statistics works better in finite samples than an existing GMM test because they use more information available from the first stage.

[15]It is unclear how to use the TFS specification (9) 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.

[16]Table 1 in the Online Supplement, Section S2, also shows these results with a 8.0 threshold; the results are similar to the $\hat{\gamma} = 8.3363$ threshold, and it makes sense as the difference in the regimes amounts to 14 more observations.

[17]Figure 1 in the Online Appendix shows the same plot with a 8.0 threshold; the results are similar to the $\hat{\gamma} = 8.3363$ threshold.

[18]The Online Supplement, figure 1, shows that this problem already occurs for a threshold of 8.0.

Rothfelder and Boldea (2016) show in their Theorem 1 that under conditional homoskedasticity and one endogenous regressor, the 2SLS estimators with a LRF or a TRF 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.

ACKNOWLEDGEMENTS

We would like to thank for very valuable comments and suggestions on this work: Jeffrey Campbell, 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.

REFERENCES

Antoine, B. and O. Boldea (2015). Inference in linear models with structural changes and mixed identification strength. SFU Working Paper 15-5.

Antoine, B. and O. Boldea (2018). Efficient estimation with time-varying information and the New Keynesian Phillips curve. *Journal of Econometrics 204*, 268–300.

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

Boldea, O., A. Cornea-Madeira, and A. R. Hall (2019). Bootstrapping structural change tests. *Journal of Econometrics 213*, 357–397.

Caner, M. and B. E. Hansen (2001). Threshold autoregression with a unit root. *Econometrica 69*, 1555–1596.

Caner, M. and B. E. Hansen (2004). Instrumental variable estimation of a threshold model. *Econometric Theory 20*, 813–843.

Chan, K. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics 21*, 520–533.

Christopoulos, D., P. McAdam, and E. Tzavalis (2019). Dealing with endogeneity in threshold models using copulas. *Journal of Business and Economic Statistics (forthcoming)*.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Gonzalo, J. and J.-Y. Pitarakis (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics 110*, 319–352.

Gonzalo, J. and J.-Y. Pitarakis (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.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics 93*, 345–368.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*, 575–603.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and Its Interface 4*, 123–127.

Hansen, B. E. (2016). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*. forthcoming.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*, 97–126.

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.

MacKinnon, J. (2009). *Bootstrap Hypothesis Testing*, Chapter 6, pp. 183–213. Wiley.

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.

Montiel-Olea, J. L. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics 31*, 358–369.

Owyang, M. T., V. A. Ramey, and S. Zubairy (2013). Are government spending multipliers greater during times of slack? Evidence from from twentieth-century historical data. *American Economic Review 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*.

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. APPENDIX A

Table 1. Rejection frequencies under the null DGP and 5% nominal size

Case	T	LFS			TFS, $\delta_\Pi = -0.5$			TFS, $\delta_\Pi = 0.5$			TFS, $\delta_\Pi = 1$		
		LR	W	WG	LR	W	WG	LR	W	WG	LR	W	WG
(a)	100	6.1%	5.9%	4.3%	2.8%	5.6%	2.7%	2.5%	3.4%	5.0%	3.8%	4.1%	5.4%
	250	5.3%	5.5%	2.6%	2.7%	2.8%	2.7%	1.0%	2.4%	3.4%	3.8%	3.7%	3.5%
	500	5.1%	5.3%	3.5%	2.7%	3.8%	3.1%	3.6%	4.1%	4.2%	5.4%	4.6%	4.4%
	1000	5.5%	5.8%	4.0%	5.2%	4.5%	3.2%	4.4%	4.3%	4.4%	5.4%	5.1%	4.4%
(b)	100	4.6%	7.3%	11.9%	6.1%	10.0%	12.7%	3.1%	5.9%	12.7%	3.5%	7.7%	12.0%
	250	4.2%	5.8%	12.2%	5.9%	7.9%	11.0%	3.6%	4.6%	9.8%	5.2%	5.9%	9.3%
	500	4.9%	4.6%	11.1%	5.3%	6.3%	8.8%	5.3%	3.9%	8.4%	6.2%	5.4%	8.0%
	1000	5.1%	4.8%	9.2%	5.9%	5.9%	7.4%	5.8%	5.9%	6.2%	5.8%	6.1%	5.9%
(c)	100	4.6%	7.3%	11.9%	6.1%	10.0%	12.7%	3.1%	5.9%	12.7%	3.5%	7.7%	12.0%
	250	4.2%	5.8%	12.2%	5.9%	7.9%	11.0%	3.6%	4.6%	9.8%	5.2%	5.9%	9.3%
	500	4.9%	4.6%	11.1%	5.3%	6.3%	8.8%	5.3%	3.9%	8.4%	6.2%	5.4%	8.0%
	1000	5.1%	4.8%	9.2%	5.9%	5.9%	7.4%	5.8%	5.9%	6.2%	5.8%	6.1%	5.9%

Here, $\eta_t \overset{iid}{\sim} \mathcal{N}(0,1)$. LR and W refer to our tests LR_T and W_T, and WG refers to the WG_T test in CH

Table 2. Rejection frequencies under the null DGP and 5% nominal size

Case	T	LFS			TFS, $\delta_\Pi = -0.5$			TFS, $\delta_\Pi = 0.5$			TFS, $\delta_\Pi = 1$		
		LR	W	WG	LR	W	WG	LR	W	WG	LR	W	WG
(b)	100	5.2%	6.8%	12.2%	2.4%	6.1%	12.7%	2.3%	5.1%	10.1%	3.7%	5.9%	9.9%
	250	5.1%	5.8%	12.4%	1.7%	4.5%	15.5%	2.7%	4.6%	11.0%	3.6%	6.2%	10.8%
	500	4.8%	5.4%	8.9%	3.5%	3.9%	10.0%	3.1%	4.0%	8.0%	3.9%	4.3%	7.0%
	1000	4.3%	4.9%	9.5%	3.3%	4.0%	12.3%	3.6%	3.8%	8.7%	4.7%	5.0%	8.7%
(c)	100	6.2%	6.7%	10.0%	4.5%	6.1%	10.3%	3.5%	3.9%	8.8%	4.1%	5.1%	8.8%
	250	4.6%	6.2%	10.4%	2.9%	4.9%	12.2%	4.1%	4.8%	7.9%	4.8%	5.4%	7.0%
	500	4.9%	6.4%	7.8%	4.1%	4.1%	9.8%	4.2%	4.9%	6.5%	4.6%	4.9%	5.7%
	1000	5.4%	4.6%	6.6%	5.1%	4.9%	9.8%	5.0%	5.5%	6.7%	4.9%	5.8%	6.3%

See table 1: test statistics are the same but $\eta_t \overset{iid}{\sim} (0,1)$ follows the Mammen (1993) distribution.

Table 3. Presence of Thresholds in the First Stage

	$\hat{\rho}$	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 LR_T and W_T.

Table 4. Presence of Thresholds in model (8)

Trim	TFS $\hat{\rho}$	$\hat{\gamma}$	LR	5%CV	Reject	W	5%CV	Reject	WG	5%CV	Reject
10%	3.5264	11.9660	93.526	77.308	Yes	29.335	28.775	Yes	178.296	198.705	No
15%	3.5264	10.7000	78.158	60.051	Yes	27.698	28.775	No	66.523	83.339	No
20%	3.7530	9.3443	75.332	51.311	Yes	27.505	29.772	No	66.523	67.292	No
25%	4.0636	8.3363	65.719	48.726	Yes	24.365	28.352	No	66.523	66.433	Yes

"5%$C.V.$" display the bootstrap 5% critical values. "Reject" indicates whether the null of no threshold in (8) is rejected. For all cases, we use a TFS with $\hat{\rho}$ obtained with the same cut-offs as the column "Trim" indicates.

Table 5. IV Multipliers with RZ threshold

	State 1, $q_t \leq 6.5$: 319 obs.				State 2, $q_t > 6.5$: 181 obs.					
h	Mult.	s.e.	AR LB	AR UB	Mult.	s.e.	AR LB	AR UB	p-val.	AR p-val.
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 6. IV Multipliers with our threshold

	State 1, $q_t \leq 8.3363$: 410 obs.				State 2, $q_t > 8.3363$: 90 obs.					
h	Mult.	s.e.	AR LB	AR UB	Mult.	s.e.	AR LB	AR UB	p-val.	AR p-val.
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	0.10	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 5 notes.

Figure 1. Size-adjusted power plots for $\delta_x = 0.25$

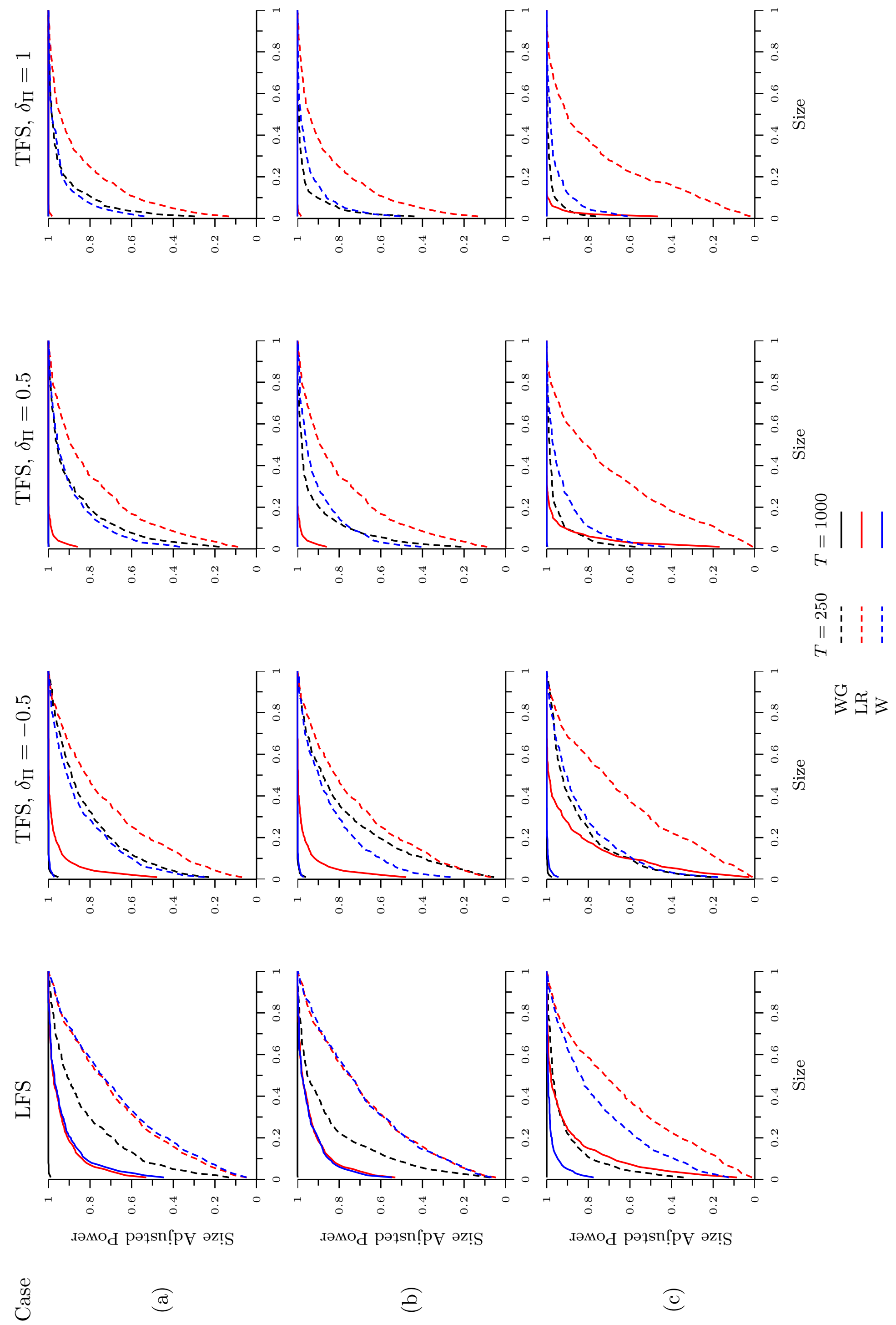


Figure 2. Plot of $LR_T(\gamma)$, $W_T(\gamma)$ and $WG_T(\gamma)$

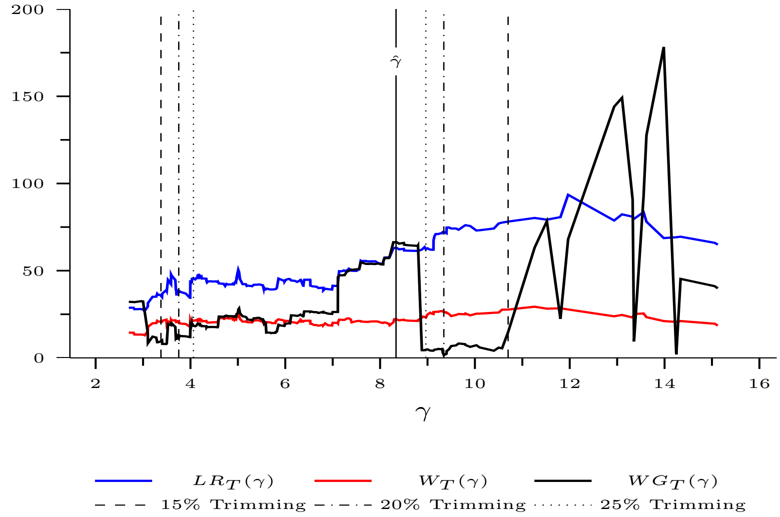
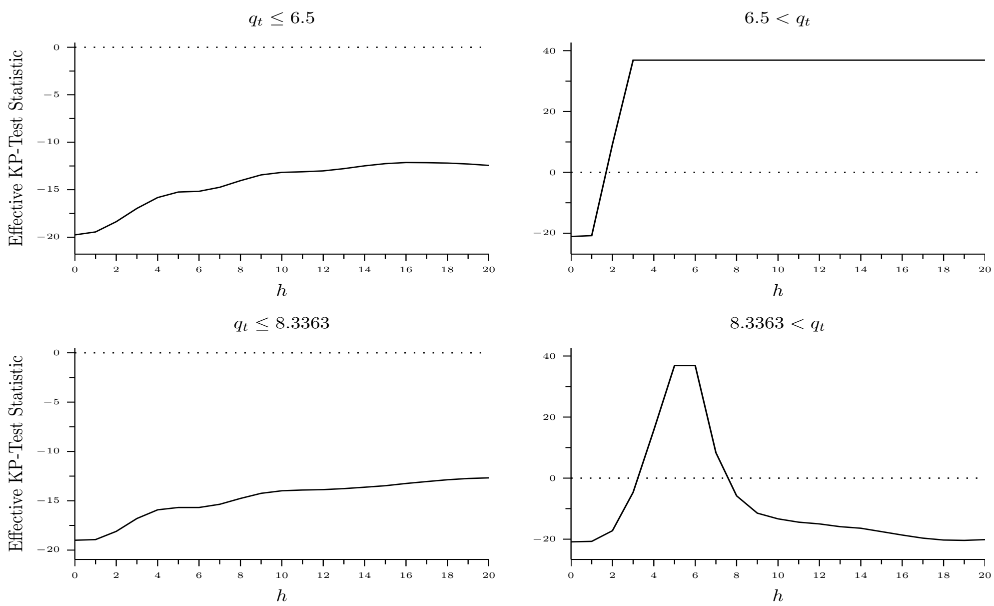
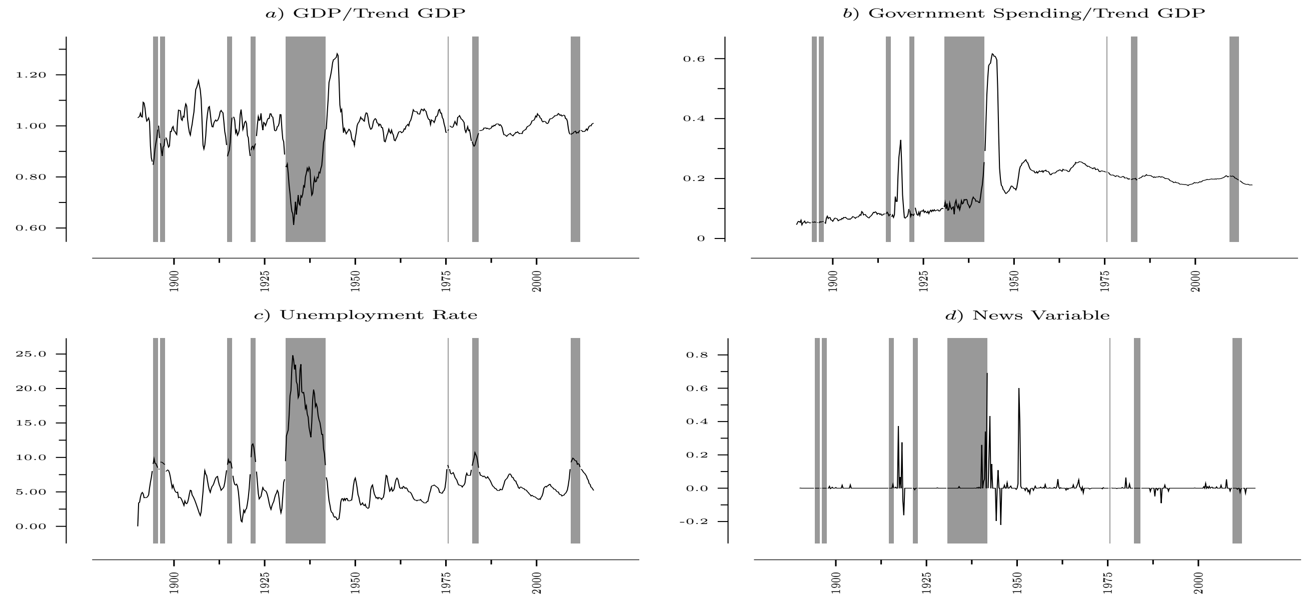


Figure 3. Effective F-statistic



The effective F- or KP-Test statistic is the Kleibergen and Paap (2006) Wald rank test in Stata, equal to the HAC-robust Wald significance test of 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. Data plots with with shaded areas $q_t > 8.3363$



B. APPENDIX B

General notation. Let $\|\cdot\|$ be the Euclidean norm for vectors, and the Frobenius norm for matrices: $\|P\| = \sqrt{\mathrm{tr}(P^\top P)}$. Also, for (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. 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, "$\overset{\mathrm{d^b}}{\Rightarrow}$" for weak convergence in Skorokhod metric under the bootstrap measure, and "$\overset{\mathrm{p^b}}{\longrightarrow}$" for weak convergence in probability under the bootstrap measure. Let K be an universal constant.

Lemma B.1. *[ULLN] If (i) $\{a_t\}$ and $\{q_t\}$ are two scalar strictly stationary and ρ-mixing series, with ρ-mixing coefficient $\rho(m) = \mathcal{O}(m^{-A})$ for some $A > \frac{a}{a-1}$ and $1 < a < r$;(ii) $\sup_t \|a_t\|_{1+\delta} < \infty$ for some $\delta > 0$; (iii) q_t has a continuous distribution, with its pdf $f(\cdot)$ bounded: $\sup_{x\in\Gamma} |f(x)| < \infty$, then $\sup_{\gamma\in\Gamma} \left|T^{-1}\sum_{1\gamma} a_t - E[a_t\mathbf{1}[q_t \le \gamma]]\right| \overset{p}{\to} 0$.*

Proof of Lemma B.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 $\{w_t \le \gamma\} = \mathbf{1}[q_t \le \gamma]$. Follow the steps in Hansen (1996), until their equation (15). Then note that $\sup_t ||a_t||_{1+\delta} \le K$ by Assumption 4.1(c), and for some $\epsilon > 0$, set $K_\epsilon = (2K/\epsilon)^{(1+\delta)/\delta}$. Since q_t is assumed to have a continuous and bounded pdf, there exists an ϵ such that $\int_{\gamma_k}^{\gamma_{k+1}} f(x)dx \leq 1/K_\epsilon = (\epsilon/(2K))^{(1+\delta)/\delta}$. Therefore, replace equation (15) by the inequality below (derived using Hölder's inequality with $p = 1/(1+\delta)$ and $q = \delta/(1+\delta)$):

$$E[|a_t|\mathbf{1}[\gamma_k < q_t \leq \gamma_{k+1}]] \leq \|a_t\|_{1+\delta}\|\mathbf{1}[\gamma_k < q_t \leq \gamma_{k+1}]]\|_{(1+\delta)/\delta}$$

$$= \sup_t \|a_t\|_{1+\delta}\left(\int_{\gamma_k}^{\gamma_{k+1}} f(x)dx\right)^{\delta/(1+\delta)} \leq K\left(\int_{\gamma_k}^{\gamma_{k+1}} f(x)dx\right)^{\delta/(1+\delta)} \leq K(\epsilon/(2K)) = \epsilon/2.$$

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

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

Lemma B.2. *[FCLT] If the assumptions in Lemma B.1 hold but with a_t being a vector of m.d.s, and additionally (i) $\sup_t \|a_t\|_{2r} < \infty$; (ii) $E[a_t a_t' \mathbf{1}[q_t \leq \gamma]] = F_\gamma$, a p.d. matrix of constants, (iii) $\inf_{\gamma\in\Gamma} \det F_\gamma > 0$, then: $T^{-1/2}\sum_{1\gamma} a_t \Rightarrow \mathcal{J}(\gamma)$, a vector of Gaussian processes with covariance function $E[a_t a_t' \mathbf{1}[q_t \leq (\gamma_1 \wedge \gamma_2)]]$.*

Proof of Lemma B.2. This functional central limit theorem (FCLT) follows directly from Theorem 3 and then Theorem 1 in Hansen (1996). Note that only $\sup_t \|a_t\|_{2r} < \infty$ 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).

General notation continued. For a generic matrix P, with $P_{1\gamma} = E[P_t\mathbf{1}[q_t \leq \gamma]]$, and $\{P_t\}_{t=1}^T$ a matrix of random variables, we let $P = P_{1\gamma_{max}} = E[P_t]$, 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 $M_{2\gamma} = E[z_t z_t' \mathbf{1}[q_t > \gamma]]$. 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 $H_{a,1\gamma} = E[z_t z_t^\top \epsilon_t^2 \mathbf{1}[q_t \leq \gamma]]$, it is $\hat{H}_{a,1\gamma} = T^{-1}\sum_{1\gamma} z_t z_t^\top \hat{\epsilon}_t^2$, 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. Let $\mathcal{G}_1(\gamma)$ be a $q(p_1+1)$ vector of independent zero mean Gaussian processes with covariance matrix $H_{1\gamma} = E[(v_t v_t^\top \otimes z_t z_t^\top)\mathbf{1}[q_t \leq \gamma]]$, and covariance function $E[\mathcal{G}_1(\gamma_1)\mathcal{G}_1^\top(\gamma_2)] = E[(v_t v_t^\top \otimes z_t z_t^\top)\mathbf{1}[q_t \leq (\gamma_1 \wedge \gamma_2)]$. We denote $\mathcal{G} = \mathcal{G}(\gamma_{max})$, where recall that $[\gamma_{min}, \gamma_{max}]$ is the support of q_t, and $\mathcal{G}_2(\gamma) = G - \mathcal{G}_1(\gamma)$. Let $\sigma^2 = E(\epsilon_t + u_t^\top \theta_x^0)^2$, $\tilde{\theta}^0 = \text{vec}(1, \theta_x^0)$ and $\check{\theta}^0 = \text{vec}(0, \theta_x^0)$.

Throughout the text, quantities of the form $v_t v_t^\top \otimes z_t z_t^\top$ should be read as $(v_t v_t^\top) \otimes (z_t z_t^\top)$. 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 or a TFS 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.

LINEAR FIRST STAGE

Notation specific for a LFS: Let $A^0 = [\Pi^0, S^\top]^\top$ be the augmented matrix of the FS slope parameters, where $S = [I_{p_2}, \mathbf{0}_{p_2\times(q-p_2)}]$ and let $\hat{A} = [\hat{\Pi}, S^\top]^\top$. Hence, $z_{1t} = Sz_t$ and $w_t = A^0 z_t + (u_t^\top, 0_{1\times q_1})^\top$. Define $C_{i\gamma} = A^0 M_{i\gamma} A^{0\top}$, $R_{i\gamma} = M_{i\gamma} M^{-1}$, $C_\gamma = [C_{1\gamma}^{-1}, -C_{2\gamma}^{-1}]$, $D_{i\gamma} = [\tilde{\theta}^{0\top} \otimes I_q] - [\check{\theta}^{0\top} \otimes R_{i\gamma}]$, $\hat{D}_{i\gamma} = [\tilde{\theta}^{0\top} \otimes I_q] - [\check{\theta}^{0\top} \otimes \hat{R}_{i\gamma}]$, $\hat{R}_{i\gamma} = \hat{M}_{i\gamma}\hat{M}^{-1}$, $D = D_{1\gamma} + D_{2\gamma} = [1, 0_{1\times p_2}] \otimes I_q$ and $Q_\gamma = C_{1\gamma} C^{-1} C_{2\gamma}$. Also, define the Gaussian processes $\mathcal{B}_i(\gamma) = A^0 D_{i\gamma}\,\mathcal{G}_i(\gamma)$for $i = 1, 2$, $\mathcal{B} = \mathcal{B}_1(\gamma) + \mathcal{B}_2(\gamma) = A^0 D\,\mathcal{G}$, as well as the processes $\mathcal{B}(\gamma) = \text{vec}(\mathcal{B}_1(\gamma), \mathcal{B}_2(\gamma))$ and $\mathcal{E}(\gamma) = C_\gamma \mathcal{B}(\gamma)$, whose covariance matrices $V_{\mathcal{B},\gamma}$ and respectively V_γ are defined below, along with $\hat{V}_\gamma$ used in the construction of the $\sup_{\gamma\in\Gamma} W_T(\gamma)$ test.

DEFINITION B.1. *Denote* $V_{\mathcal{B},1\gamma} = A^0 D_{1\gamma} H_{1\gamma} D_{1\gamma}^\top A^0$, $V_{\mathcal{B}} = A^0 D\, H\, D^\top A^0$, $V_{\mathcal{B},12,\gamma} = A^0 D_{1\gamma} H_{1\gamma} D^\top A^{0\top}$, $V_{\mathcal{B},2\gamma} = V_{\mathcal{B}} + V_{\mathcal{B},1\gamma} - V_{\mathcal{B},12,\gamma} - V_{\mathcal{B},12,\gamma}^\top$. *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* $V_\gamma = C_\gamma V_{\mathcal{B},\gamma} C_\gamma^\top$. *Define* $\bar{D}_{i\gamma} = [\tilde{\theta}^\top \otimes I_q] - [\check{\theta}^\top \otimes \hat{R}_{i\gamma}]$, *where* $\tilde{\theta} = [1, \hat{\theta}_x]$ *and* $\check{\theta} = [0, \hat{\theta}_x]$. *Then* $\hat{V}_\gamma$ *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 B.3. *Suppose Assumption 4.1 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)* $T^{-1}\sum_{i\gamma} z_t z_t^\top \xrightarrow{p} M_{i\gamma}$ *and* $T^{-1}\sum_{i\gamma} \hat{w}_t\hat{w}_t^\top \xrightarrow{p} C_{i\gamma}$; *(ii)* $T^{-1/2}\sum_{i\gamma} v_t \otimes z_t \Rightarrow \mathcal{G}_i(\gamma)$, $T^{-1/2}\sum_{i\gamma}\hat{w}_t\tilde{\epsilon}_t \Rightarrow \mathcal{B}_i(\gamma)$, *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 B.3. 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 4.1 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 Assumption 4.1(b),(c) and (e), the assumptions of Lemma B.1 are satisfied for elements of $a_t = z_t z_t^\top$, so $T^{-1}\sum_{1\gamma} z_t z_t^\top \xrightarrow{p} M_{1\gamma}$, $T^{-1}\sum_{2\gamma} z_t z_t^\top \xrightarrow{p} M_{2\gamma}$, and $T^{-1}\sum_{i\gamma}\hat{w}_t\hat{w}_t^\top \xrightarrow{p} A^0 M_{i\gamma} A^{0\top} = C_{i\gamma}$.
Part (ii).

$$
\begin{aligned}
T^{-1/2}\sum_{1\gamma}\hat{w}_t\tilde{\epsilon}_t &= \hat{A}\left(T^{-1/2}\sum\nolimits_{1\gamma} z_t(\epsilon_t + u_t^\top\theta_x^0) - \hat{R}_{1\gamma}T^{-1/2}\sum\nolimits_{1\gamma} z_t u_t^\top\theta_x^0\right) \\
&= \hat{A}\,\hat{D}_{1\gamma}\left(T^{-1/2}\sum\nolimits_{1\gamma} z_t \otimes v_t\right),
\end{aligned}
$$

where by Lemma B.1(i), $\hat{M}_{1\gamma} \xrightarrow{p} M_{1\gamma}$ and $\hat{M} \xrightarrow{p} M$, therefore $\hat{D}_{1\gamma} \xrightarrow{p} D_{1\gamma}$. By Assumption 4.1(a)-(f), the conditions of Lemma B.2 are satisfied (because $\|v_t \otimes z_t\|_{2r} \le \|v_t\|_{4r}\|z_t\|_{4r} < K$), so $T^{-1/2}\sum_{1\gamma} v_t \otimes z_t \Rightarrow \mathcal{G}_1(\gamma)$. Because $\hat{A} \xrightarrow{p} A^0$, $T^{-1/2}\sum_{1\gamma}\hat{w}_t\tilde{\epsilon}_t \Rightarrow A^0\,D_{1\gamma}\mathcal{G}_1(\gamma) = \mathcal{B}_1(\gamma)$. Because

$$
T^{-1/2}\sum\nolimits_{t=1}^T \hat{w}_t\tilde{\epsilon}_t = \hat{A}T^{-1/2}\sum\nolimits_{t=1}^T z_t\epsilon_t = \hat{A}\,D(T^{-1/2}\sum\nolimits_{t=1}^T z_t \otimes v_t) \Rightarrow A^0\,D\,\mathcal{G} = \mathcal{B},
$$

$T^{-1/2}\sum_{2\gamma}\hat{w}_t\tilde{\epsilon}_t = T^{-1/2}\sum_{t=1}^T \hat{w}_t\tilde{\epsilon}_t - T^{-1/2}\sum_{1\gamma}\hat{w}_t\tilde{\epsilon}_t \Rightarrow \mathcal{B} - \mathcal{B}_1(\gamma) = \mathcal{B}_2(\gamma)$, and therefore $T^{-1/2}\,\text{vec}(\sum_{1\gamma}\hat{w}_t\tilde{\epsilon}_t, \sum_{2\gamma}\hat{w}_t\tilde{\epsilon}_t) \Rightarrow \mathcal{B}(\gamma)$.

THEOREM B.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 4.1,*
(i) $T^{1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma}) \Rightarrow \mathcal{E}(\gamma)$,
(ii) $\sup_{\gamma\in\Gamma} LR_T(\gamma) \Rightarrow \sup_{\gamma\in\Gamma}[\mathcal{E}^\top(\gamma)\ Q_\gamma\ \mathcal{E}(\gamma)/\sigma^2]$,
(iii) $\sup_{\gamma\in\Gamma} W_T(\gamma) \Rightarrow [\sup_{\gamma\in\Gamma}\mathcal{E}^\top(\gamma)V_\gamma^{-1}\mathcal{E}(\gamma)]$, *where* V_γ *is in Definition B.1.*

Proof of Theorem B.1. Part (i). By Lemma B.3, we have that $T^{-1/2}(\hat{\theta}_{i\gamma} - \theta^0) =$

$\hat{C}_{i\gamma}^{-1}(T^{-1/2}\sum_{i\gamma}\hat{w}_t\tilde{\epsilon}_t)\Rightarrow C_{i\gamma}^{-1}\mathcal{B}_i(\gamma)$, so

$$T^{1/2}(\hat{\theta}_{1\gamma}-\hat{\theta}_{2\gamma})\Rightarrow[C_{1\gamma}^{-1},-C_{2\gamma}^{-1}]\,\mathrm{vec}(\mathcal{B}_{1\gamma},\mathcal{B}_{2\gamma})=C_\gamma\,\mathcal{B}(\gamma)=\mathcal{E}(\gamma).$$

Part (ii). Since $\hat{\theta}=\hat{C}^{-1}(T^{-1}\sum_{t=1}^T\hat{w}_ty_t)$, $\hat{C}\hat{\theta}=\sum_{i=1}^2\hat{C}_{i\gamma}\hat{\theta}_{i\gamma}$, so $\hat{\theta}=\sum_{i=1}^2C^{-1}C_{i\gamma}\hat{\theta}_{i\gamma}+o_p(1)$, 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,

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

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

Now $SSR_1(\gamma)/(T-2p)=T^{-1}SSR_1(\gamma)+o_p(1)$, since show below $T^{-1}SSR_1(\gamma)\xrightarrow{\text{p}}\sigma^2$:

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

By Lemma B.3, $T^{-1}\sum_{i\gamma}\hat{w}_t\hat{w}_t^\top\xrightarrow{\text{p}}C_{i\gamma}$, and $T^{-1}\sum_{i\gamma}\tilde{\epsilon}_t\hat{w}_t^\top=o_p(1)$. Since $\hat{\theta}_{i\gamma}-\theta^0=o_p(1)$,

$$\begin{aligned}
&T^{-1}SSR_1(\gamma)=T^{-1}\sum_{t=1}^T(\epsilon_t+u_t^\top\theta_x^0-z_t^\top(\hat{\Pi}-\Pi^0)\theta_x^0)^2+o_p(1)\\
&=T^{-1}\sum_{t=1}^T(\epsilon_t+u_t^\top\theta_x^0)^2+\theta_x^{0\top}(\hat{\Pi}-\Pi^0)T^{-1}\sum_{t=1}^Tz_tz_t^\top(\hat{\Pi}-\Pi^0)\theta_x^0\\
&-2T^{-1}\sum_{t=1}^Tz_t(\epsilon_t+u_t^\top\theta_x^0)(\hat{\Pi}-\Pi^0)\theta_x^0+o_p(1)\\
&=T^{-1}\sum_{t=1}^T(\epsilon_t+u_t^\top\theta_x^0)^2+o_p(1),
\end{aligned}\tag{10}$$

where the last equality used Lemma B.3(ii) and $\hat{\Pi}-\Pi^0=o_p(1)$. We now apply Lemma B.1 to $a_t=(\epsilon_t+u_t^\top\theta_x^0)^2$. First, $E(\epsilon_t+u_t^\top\theta_x^0)^2=\sigma^2$. Second, by Assumption 4.1(b), a_t is strictly stationary with ρ-mixing coefficients satisfying condition (i) in Lemma B.1. Third, by Minkowski's inequality, $\|(\epsilon_t+u_t^\top\theta_x^0)^2\|_2\le\|\epsilon_t^2\|_2+\|\theta_x^{0\top}u_tu_t^\top\theta_x^0\|_2+2\|\epsilon_tu_t^\top\theta_x^0\|_2$. Note that $\|\epsilon_t^2\|_2=(E|\epsilon_t^4|)^{1/2}<K$ by Assumption 4.1(c). By the same assumption and Hölder and Minkowksi inequalities, letting $\theta_{x,j}^0$ be the j^{th} element of θ_x^0, we have:

$$\|\theta_x^{0\top}u_tu_t^\top\theta_x^0\|_2\le\|\sum_{j,k=1}^{p_1}\theta_{x,j}^0\theta_{x,k}^0u_{t,j}u_{t,k}\|_2\le\max_{j,k}|\theta_{x,j}^0\theta_{x,k}^0|\,p_1^2\|u_t\|_2^2<K,$$

and similarly, $\|\epsilon_t u_t^\top \theta_x^0\|_2 < K$. Therefore, by Lemma B.1, $T^{-1}\sum_{t=1}^T(\epsilon_t + u_t^\top\theta_x^0)^2 \xrightarrow{\text{p}} \sigma^2$, completing the proof of part (ii).

Part (iii). We are left to show $\hat{V}_\gamma \xrightarrow{\text{p}} V_\gamma$. Since $\hat{C}_{i\gamma} \xrightarrow{\text{p}} C_{i\gamma}$, $\hat{A} \xrightarrow{\text{p}} A^0$, $\hat{\theta}_x \xrightarrow{\text{p}} \theta_x^0$, and therefore $\bar{D}_{i\gamma} \xrightarrow{\text{p}} D_{i\gamma}$, to show that $\hat{V}_\gamma \xrightarrow{\text{p}} V_\gamma$, it suffices to show that $\hat{H}_{i\gamma} = \text{Var}(T^{-1/2}\sum_{i\gamma}\hat{v}_t \otimes z_t) \xrightarrow{\text{p}} H_{i\gamma}$ for $i = 1,2$. We show $\hat{H}_{a,1\gamma} = T^{-1}\sum_{i\gamma} z_t z_t^\top \hat{\epsilon}_t^2 \xrightarrow{\text{p}} H_{a,1\gamma} = E[z_t z_t^\top \epsilon_t^2 \mathbf{1}[q_t \leq \gamma]]$; the rest of the proof is similar.

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

First, by Assumption 4.1(b),(c) and (e), and Lemma B.1, $T^{-1}\sum_{1\gamma} z_t z_t^\top \epsilon_t^2 \xrightarrow{\text{p}} H_{a,1\gamma}$. Second, $T^{-1}\sum_{1\gamma} z_t z_t^\top[(\theta^0 - \hat{\theta})^\top w_t w_t^\top(\theta^0 - \hat{\theta})] = [I_q \otimes (\theta^0 - \hat{\theta})]^\top T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes w_t w_t^\top [I_q \otimes (\theta^0 - \hat{\theta})]$. We know that $\theta^0 - \hat{\theta} = O_p(T^{-1/2})$. Also, it can be shown that for typical element $\xi_t = z_{t,i} z_{t,j} w_{t,k} w_{t,l}$, with $i,j = 1,\ldots,q$ and $k,l = 1,\ldots p$, we have: $\|\xi_t\|_r \leq K \max_i \|z_{t,i}\|_{4r} < \infty$, by Assumption 4.1(c), where $r > 1$. Therefore, by Assumption 4.1(b)-(c) and Lemma B.1, $T^{-1}\sum_{i\gamma} z_t z_t^\top \otimes w_t w_t^\top = O_p(1)$ uniformly in γ. It follows that $T^{-1}\sum_{1\gamma} z_t z_t^\top[(\theta^0 - \hat{\theta})^\top w_t w_t^\top(\theta^0 - \hat{\theta})] = o_p(1)$. Similarly, $T^{-1}\sum_{1\gamma} z_t z_t^\top(\theta^0 - \hat{\theta})^\top w_t \epsilon_t = o_p(1)$, therefore $\hat{H}_{a,1\gamma} \xrightarrow{\text{p}} H_{a,1\gamma}$, so by similar arguments, $\hat{H}_{a,2\gamma} \xrightarrow{\text{p}} H_{a,2\gamma}$, completing the proof of part (ii).

BOOTSTRAP VALIDITY LFS

LEMMA B.4. *Let Assumptions 4.1-4.2 hold, y_t is generated by* (1)*, and x_t is generated by the LFS* (2)*. Then, under $\mathbb{H}_0$, (i) $T^{-1}\sum_{i\gamma} v_t \eta_t \otimes z_t = o_p^b(1)$; (ii) $T^{-1}\sum_{i\gamma} z_t z_t^\top \eta_t = o_p^b(1)$; (iii) $T^{-1/2}\sum_{i\gamma} v_t \eta_t \otimes z_t \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{G}_i(\gamma)$; (iv) $T^{-1/2}\sum_{i\gamma} v_t^b \otimes z_t \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{G}_i(\gamma)$; (v) $T^{1/2}(\hat{\Pi}^b - \hat{\Pi}) = T^{1/2}(\hat{\Pi} - \Pi^0) + o_p^b(1)$; (vi) $T^{-1}\sum_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \xrightarrow{\text{p}^\text{b}} C_{i\gamma}$; (vii) $T^{-1/2}\sum_{i\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{B}_i(\gamma)$ and $\text{vec}(T^{-1/2}\sum_{1\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b, T^{-1/2}\sum_{2\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b) \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{B}(\gamma)$, where $\tilde{\epsilon}_t^b = y_t^b - \hat{w}_t^{b\top}\hat{\theta}$.*

Proof of Lemma B.4. We will show Lemma B.4 for $i = 1$; for $i = 2$, the proof is similar and omitted for brevity.

Part (i). By Markov's inequality and Assumptions 4.1(c) and 4.2(a), $P^b(\|T^{-1}\sum_{i\gamma} v_t \eta_t \otimes z_t\| > c) \leq c^{-1}E^b\|T^{-1}\sum_{i\gamma} v_t \eta_t \otimes z_t\| = c^{-1}\|T^{-1}\sum_{i\gamma} v_t \otimes z_t\| E^b|\eta_t| \leq c^{-1}K\|T^{-1}\sum_{i\gamma} v_t \otimes z_t\| = o_p(1)$, where the last equality follows from Lemma B.3(ii). Therefore, $T^{-1}\sum_{i\gamma} v_t \eta_t \otimes z_t = o_p^b(1)$.

Part (ii). $\|T^{-1}\sum_{1\gamma} z_t z_t^\top \eta_t\| \leq \|T^{-1}\sum_{1\gamma}(z_t z_t^\top - M_{1\gamma})\eta_t\| + \|M_{1\gamma}T^{-1}\sum_{1\gamma}\eta_t\|$. First, $P^b(\|T^{-1}\sum_{1\gamma}(z_t z_t^\top - M_{1\gamma})\eta_t\| > c) \leq \eta_t K E^b|\eta_t|\ \|T^{-1}\sum_{1\gamma}(z_t z_t^\top - M_{1\gamma})\| = o_p^b(1)$, where the last equality follows from Lemma B.3(i). By Assumptions 4.1(e), 4.2 and Lemma B.1, it can be shown that $T^{-1}\sum_{1\gamma}\eta_t = o_p^b(1)$, therefore $\|M_{1\gamma}T^{-1}\sum_{1\gamma}\eta_t\| \leq \|M_{1\gamma}\|\|T^{-1}\sum_{i\gamma}\eta_t| = o_p^b(1)$, so $T^{-1}\sum_{1\gamma} z_t z_t^\top \eta_t = o_p^b(1)$.

Part (iii). To show $T^{-1/2}\sum_{1\gamma} v_t \eta_t \otimes z_t \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{G}P_1(\gamma)$, we apply Lemma B.2 (FCLT) and verify that $Var^b(T^{-1/2}\sum_{1\gamma} v_t \eta_t \otimes z_t) \xrightarrow{\text{p}} H_{1\gamma}$. First, $E^b(v_t \eta_t \otimes z_t) = 0$. Conditions (ii) and (iii) in Lemma B.2 are satisfied by Assumption 4.1(b),(c), (e) and Assumption 4.2(a).

Condition (i) is satisfied because $E^b(\|v_t\eta_t \otimes z_t\|^{2r})^{1/(2r)} \leq (E^b|\eta_t|^{2r})^{1/(2r)}\|v_t z_t\| = O_p(1)$ by Assumptions 4.1(c) and 4.2. Finally, $Var^b(T^{-1/2}\sum_{1\gamma} v_t\eta_t \otimes z_t) = T^{-1}\sum_{1\gamma} v_t v_t^\top \otimes z_t z_t^\top \xrightarrow{\text{p}} H_{1\gamma}$, where the last statement follows by applying Lemma B.1(ULLN). Therefore, by Lemma B.2, $T^{-1/2}\sum_{1\gamma} v_t\eta_t \otimes z_t \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{G}_1(\gamma)$.

Part (iv). Since $\hat{u}_t = x_t^\top - \hat{x}_t^\top = u_t - (\hat{\Pi} - \Pi^0)^\top z_t$,

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

Similarly, $T^{-1/2}\sum_{1\gamma} \epsilon_t^b z_t = T^{-1/2}\sum_{1\gamma} \epsilon_t\eta_t \otimes z_t + o_p^b(1)$, therefore $T^{-1/2}\sum_{1\gamma} v_t^b \otimes z_t \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{G}_1(\gamma)$.

Part (v). Since $v_t = (\epsilon_t, u_t^\top)^\top$, from part (iv) and Lemma B.3(ii), $T^{1/2}(\hat{\Pi}^b - \hat{\Pi}) = \hat{M}(T^{-1/2}\sum_{t=1}^T z_t u_t^{b\top}) = \hat{M}(T^{-1/2}\sum_{t=1}^T z_t u_t^\top + o_p^b(1)) = T^{1/2}(\hat{\Pi} - \Pi^0) + o_p^b(1)$.

Part (vi). $T^{-1}\sum_{1\gamma} \hat{w}_t^b \hat{w}_t^b = \hat{A}^b(T^{-1}\sum_{1\gamma} z_t z_t^\top)\hat{A}^{b\top}$, where $\hat{A}^b = [\hat{\Pi}^b, S^\top]^\top$. From part (v), $T^{1/2}(\hat{\Pi}^b - \Pi^0) = O_p^b(1)$, so $\hat{\Pi}^b \xrightarrow{\text{p}^\text{b}} \Pi^0$, therefore $\hat{A}^b \xrightarrow{\text{p}^\text{b}} A^0$, and $T^{-1}\sum_{1\gamma} \hat{w}_t^b \hat{w}_t^{b\top} \xrightarrow{\text{p}^\text{b}} C_{1\gamma}$.

Part (vii). By part (iv) and (v) and Lemma B.3(i)-(ii), and recalling that $\bar{D}_{1\gamma} = [\tilde{\theta} \otimes I_q] - [\check{\theta} \otimes \hat{R}_{1\gamma}] \xrightarrow{\text{p}} D_{1\gamma}$, we have:

$$
\begin{aligned}
&T^{-1/2}\sum\nolimits_{1\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b = \hat{A}^b\left(T^{-1/2}\sum\nolimits_{1\gamma} z_t(\epsilon_t^b + (w_t^b - \hat{w}_t^b)^\top \hat{\theta})\right) \\
&= \hat{A}^b\left(T^{-1/2}\sum\nolimits_{1\gamma} z_t(\epsilon_t^b + u_t^{b\top}\hat{\theta}_x) - T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top T^{1/2}(\hat{\Pi}^b - \hat{\Pi})\hat{\theta}_x\right) \\
&= A^0\bar{D}_{1\gamma}\big(T^{-1/2}\sum\nolimits_{1\gamma} v_t^b \otimes z_t\big) + o_p^b(1) = A^0 D_{1\gamma}\big(T^{-1/2}\sum\nolimits_{1\gamma} v_t^b \otimes z_t\big) + o_p^b(1) \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{B}_1(\gamma).
\end{aligned}
$$

Similarly, $T^{-1/2}\sum_{2\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{B}_2(\gamma)$, so $\text{vec}(T^{-1/2}\sum_{1\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b, T^{-1/2}\sum_{2\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b) \overset{\text{d}^\text{b}}{\Rightarrow} \mathcal{B}(\gamma)$.

Proof of Theorem 4.2. Part (i).

$$
\begin{aligned}
&SSR_0^b - SSR_1^b(\gamma) = \sum\nolimits_{i=1}^2 \left[\sum\nolimits_{i\gamma} (y_t^b - \hat{w}_t^{b\top}\hat{\theta}^b)^2 - (y_t^b - \hat{w}_t^{b\top}\hat{\theta}_{i\gamma}^b)^2\right] \\
&= \sum\nolimits_{i=1}^2 (\hat{\theta}_{i\gamma}^b - \hat{\theta}^b)^\top \left[2\sum\nolimits_{i\gamma} \hat{w}_t^b \tilde{\epsilon}_t^b - \sum\nolimits_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top}(\hat{\theta}^b - \hat{\theta}) - \sum\nolimits_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top}(\hat{\theta}_{i\gamma}^b - \hat{\theta})\right] \\
&= \sum\nolimits_{i=1}^2 T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b)^\top \left(T^{-1}\sum\nolimits_{i\gamma} \hat{w}_t^b \hat{w}_t^{b\top}\right)(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b) \\
&= \sum\nolimits_{i=1}^2 T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b)^\top C_{i\gamma} \sum\nolimits_{i=1}^2 T^{1/2}(\hat{\theta}_{i\gamma}^b - \hat{\theta}^b) + o_p^b(1) \\
&= T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b)^\top Q_\gamma T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b) + o_p^b(1),
\end{aligned}
\tag{11}
$$

where the second to last line follows by Lemma B.4(vi), and the last line by similar calculations as in the first two paragraphs of the Proof of Theorem B.1(ii). By Lemma

B.4(vi) and (vii),

$$\begin{aligned}
&T^{1/2}(\hat{\theta}^b_{1\gamma} - \hat{\theta}^b_{2\gamma}) = T^{1/2}(\hat{\theta}^b_{1\gamma} - \hat{\theta}) - T^{1/2}(\hat{\theta}^b_{1\gamma} - \hat{\theta}) \\
&= (T^{-1}\sum\nolimits_{1\gamma} \hat{w}^b_t \hat{w}^{b\top}_t)^{-1} T^{-1/2} \sum\nolimits_{1\gamma} \hat{w}^b_t \tilde{\epsilon}^{b\top}_t - (T^{-1}\sum\nolimits_{2\gamma} \hat{w}^b_t \hat{w}^{b\top}_t)^{-1} T^{-1/2} \sum\nolimits_{2\gamma} \hat{w}^b_t \tilde{\epsilon}^{b\top}_t \\
&\overset{\mathrm{d^b}}{\Rightarrow} C^{-1}_{1\gamma}\mathcal{B}_1(\gamma) - C^{-1}_{2\gamma}\mathcal{B}_2(\gamma) = \mathcal{E}(\gamma). \qquad (12)
\end{aligned}$$

Using (12) into (11), we have: $SSR^b_0 - SSR^b_1(\gamma) \overset{\mathrm{d^b}}{\Rightarrow} \mathcal{E}^\top(\gamma) Q_\gamma\, \mathcal{E}(\gamma)$. It remains to show that $SSR^b_1(\gamma)/(T-2p) \overset{\mathrm{p^b}}{\longrightarrow} \sigma^2$, or, equivalently, that $T^{-1}SSR^b_1(\gamma) \overset{\mathrm{p^b}}{\longrightarrow} \sigma^2$.

$$\begin{aligned}
&T^{-1}SSR^b_1(\gamma) = \sum\nolimits_{i=1}^2 T^{-1}\sum\nolimits_{i\gamma}(y^b_t - \hat{w}^{b\top}_t\hat{\theta}^b_{i\gamma})^2 = \sum\nolimits_{i=1}^2 T^{-1}\sum\nolimits_{i\gamma}(\tilde{\epsilon}^b_t - \hat{w}^{b\top}_t(\hat{\theta}^b_{i\gamma} - \hat{\theta}))^2 \\
&= T^{-1}\sum_{t=1}^T (\tilde{\epsilon}^b_t)^2 - 2\sum\nolimits_{i=1}^2 T^{-1}\sum\nolimits_{i\gamma} \tilde{\epsilon}^b_t \hat{w}^{b\top}_t(\hat{\theta}^b_{i\gamma} - \hat{\theta}) \\
&+ \sum\nolimits_{i=1}^2 (\hat{\theta}^b_{i\gamma} - \hat{\theta})^\top (T^{-1}\sum\nolimits_{i\gamma}\hat{w}^b_t\hat{w}^{b\top}_t)(\hat{\theta}^b_{i\gamma} - \hat{\theta}).
\end{aligned}$$

By Lemma B.4(vi) and (vii), $T^{-1}\sum_{i\gamma}\hat{w}^b_t\hat{w}^{b\top}_t \overset{\mathrm{p^b}}{\longrightarrow} C_{i\gamma}$, and $T^{-1}\sum_{i\gamma}\tilde{\epsilon}^b_t\hat{w}^{b\top}_t = O^b_p(T^{-1/2})$. From above,

$$\hat{\theta}^b_{i\gamma} - \hat{\theta} = O^b_p(T^{-1/2}) = o^b_p(1). \qquad (13)$$

Therefore, $T^{-1}SSR^b_1(\gamma) = T^{-1}\sum_{t=1}^T(\tilde{\epsilon}^b_t)^2 + o^b_p(1) = T^{-1}\sum_{t=1}^T(\epsilon^b_t + u^{b\top}_t\hat{\theta}_x - z^\top_t(\hat{\Pi}^b - \hat{\Pi})^\top\hat{\theta}_x)^2 + o^b_p(1) = T^{-1}\sum_{t=1}^T(\epsilon^b_t + u^{b\top}_t\theta^0_x)^2 + o^b_p(1)$, where the last equality used Lemma B.4(v), which implies $\hat{\Pi}^b - \hat{\Pi} = o_p(1)$. We now show that $T^{-1}\sum_{t=1}^T(\epsilon^b_t + u^{b\top}_t\theta^0_x)^2 \overset{\mathrm{p^b}}{\longrightarrow} \sigma^2$, which then would complete the proof of part (i).

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

$$\begin{aligned}
&T^{-1}\sum\nolimits_{t=1}^T \hat{\epsilon}^2_t\eta^2_t = T^{-1}\sum\nolimits_{t=1}^T(\epsilon_t - w^\top_t(\hat{\theta} - \theta^0))^2\eta^2_t = T^{-1}\sum\nolimits_{t=1}^T \epsilon^2_t\eta^2_t \\
&+ (\hat{\theta} - \theta^0)^\top A^0(T^{-1}\sum\nolimits_{t=1}^T z_t z^\top_t\eta^2_t)A^{0\top}(\hat{\theta} - \theta^0) - 2(T^{-1}\sum\nolimits_{t=1}^T \epsilon_t z^\top_t\eta^2_t)A^{0\top}(\hat{\theta} - \theta^0) \\
&= T^{-1}\sum\nolimits_{t=1}^T \epsilon^2_t\eta^2_t + o_p(1)(T^{-1}\sum\nolimits_{t=1}^T z_t z^\top_t\eta^2_t)o_p(1) - (T^{-1}\sum_{t=1}^T \epsilon_t z^\top_t\eta^2_t)o_p(1). \qquad (14)
\end{aligned}$$

First, we show that $T^{-1}\sum_{t=1}^T\epsilon^2_t\eta^2_t \overset{\mathrm{p^b}}{\longrightarrow} E(\epsilon^2_t)$. Note that $E^b(T^{-1}\sum_{t=1}^T\epsilon^2_t\eta^2_t) = T^{-1}\sum_{t=1}^T\epsilon^2_t$. Also, letting $K^* = E^b(\eta^4_t)$, $E^b(T^{-1}\sum_{t=1}^T\epsilon^2_t\eta^2_t)^2 = T^{-2}\sum_{t=1}^T\epsilon^4_t K^* + T^{-2}\sum_{t,s=1,t\neq s}^T\epsilon^2_t\epsilon^2_s$. By Assumption 4.1(b)-(c), ϵ^4_t satisfies Lemma B.1, so $T^{-1}\sum_{t=1}^T\epsilon^4_t \overset{\mathrm{p}}{\rightarrow} E(\epsilon^4_t)$, and therefore $T^{-2}\sum_{t=1}^T\epsilon^4_t K^* = o_p(1)$. Also, note that $T^{-2}\sum_{t,s=1,t\neq s}^T\epsilon^2_t\epsilon^2_s = (T^{-1}\sum_{t=1}^T\epsilon^2_t)^2 - T^{-2}\sum_{t=1}^T\epsilon^4_t \overset{\mathrm{p}}{\rightarrow}[E(\epsilon^2_t)]^2$. Hence, $E^b(T^{-1}\sum_{t=1}^T\epsilon^2_t(\eta^2_t - 1)) = 0$, and $E^b(T^{-1}\sum_{t=1}^T\epsilon^2_t(\eta^2_t - 1))^2 = o_p(1)$, so, by Markov's inequality, $P^b(|T^{-1}\sum_{t=1}^T\epsilon^2_t\eta^2_t - \epsilon^2_t| > \eta) \overset{\mathrm{p}}{\rightarrow} 0$, therefore $T^{-1}\sum_{t=1}^T\epsilon^2_t\eta^2_t \overset{\mathrm{p^b}}{\longrightarrow} E(\epsilon^2_t)$.

Second, we show that $T^{-1}\sum_{t=1}^T z_t z_t^\top \eta_t^2 = O_p^b(1)$. We have $E^b(z_t z_t^\top \eta_t^2) = z_t z_t^\top$. Also,

$$\begin{aligned}
&E^b[(\mathrm{vec}(T^{-1}\sum\nolimits_{t=1}^T z_t z_t^\top \eta_t^2))\,\mathrm{vec}(T^{-1}\sum\nolimits_{t=1}^T z_t z_t^\top \eta_t^2)] \\
&= T^{-2}\sum\nolimits_{t=1}^T (z_t z_t^\top \otimes z_t z_t^\top)K + (T^{-1}\sum\nolimits_{t=1}^T z_t \otimes z_t)(T^{-1}\sum\nolimits_{t=1}^T z_t \otimes z_t^\top) \\
&- T^{-2}\sum\nolimits_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top,
\end{aligned}$$

By Assumption 4.1(b)-(c) and Lemma B.1, $T^{-2}\sum_{t=1}^T (z_t z_t^\top \otimes z_t z_t^\top) = o_p(1)$. Also, by Lemma B.3(a), $T^{-1}\sum_{t=1}^T z_t \otimes z_t = \mathrm{vec}\,\hat{M} \xrightarrow{\mathrm{p}} \mathrm{vec}\,M$. So, $E^b\,\mathrm{vec}(T^{-1}\sum_{t=1}^T z_t z_t^\top \eta_t^2) = z_t \otimes z_t$, and $Var^b(\mathrm{vec}(T^{-1}\sum_{t=1}^T z_t z_t^\top \eta_t^2 - z_t z_t^\top)) = 0$. Therefore, by Markov's inequality, $T^{-1}\sum_{t=1}^T z_t z_t^\top \eta_t^2 - z_t z_t^\top = o_p^b(1)$, hence $T^{-1}\sum_{t=1}^T z_t z_t^\top \eta_t^2 = T^{-1}\sum_{t=1}^T z_t z_t^\top + o_p^b(1) = M + o_p^b(1) = O_p^b(1)$.

Third, we show that $\xi_t = T^{-1}\sum_{t=1}^T \epsilon_t z_t \eta_t^2 = o_p^b(1)$. Note that $E^b(T^{-1}\sum_{t=1}^T \epsilon_t z_t \eta_t^2) = \epsilon_t z_t$, and, by similar arguments as before, $E^b(\xi_t \xi_t^\top) \xrightarrow{\mathrm{p}} 0$, so, by Markov's inequality, $\xi_t \xrightarrow{\mathrm{p^b}} 0$. Substituting these results into (14), it follows that $T^{-1}\sum_{t=1}^T \hat{\epsilon}_t^2 \eta_t^2 \xrightarrow{\mathrm{p^b}} E(\epsilon_t^2)$. Next,

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

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

Part (ii). Let $\bar{D}_{i\gamma}^b = [\tilde{\theta}^{b\top} \otimes I_q] - [\check{\theta}^{b\top} \otimes \hat{R}_{i\gamma}]$, where $\tilde{\theta}^b = [1, \hat{\theta}_x^b]$ and $\check{\theta}^b = [0, \hat{\theta}_x^b]$. From (12), $T^{1/2}(\hat{\theta}_{1\gamma}^b - \hat{\theta}_{2\gamma}^b) \overset{\mathrm{d^b}}{\Rightarrow} \mathcal{E}(\gamma)$, so it remains to show that $\hat{V}_\gamma^b \xrightarrow{\mathrm{p^b}} V_\gamma$. We will just show that $\hat{V}_{\mathcal{B},1\gamma}^b = \hat{A}^b \bar{D}_{1\gamma}^b \hat{H}_{1\gamma}^b \bar{D}_{1\gamma}^{b\top} \hat{A}^{b\top} \xrightarrow{\mathrm{p}} A^0\, D_{1\gamma}\, H_{1\gamma} D_{1\gamma}^\top A^{0\top} = V_{\mathcal{B},1\gamma}$, where $\hat{H}_{1\gamma}^b = T^{-1}\sum_{1\gamma} \hat{v}_t^b \hat{v}_t^b \otimes z_t z_t^\top$; the rest follows by similar arguments. Since $\hat{A}^b \xrightarrow{\mathrm{p^b}} A^0$, $\bar{D}_{i\gamma}^b \xrightarrow{\mathrm{p}} D_{i\gamma}$, and the proof for $\hat{H}_{1\gamma}^b \xrightarrow{\mathrm{p^b}} H_{1\gamma}$ is similar to $\hat{H}_{a,1\gamma}^b \xrightarrow{\mathrm{p^b}} H_{a,1\gamma}$, where $\hat{H}_{a,1\gamma}^b = T^{-1}\sum_{1\gamma}(\hat{\epsilon}_t^b)^2 z_t z_t^\top$, we only show $\hat{H}_{a,1\gamma}^b \xrightarrow{\mathrm{p^b}} H_{a,1\gamma}$.

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

We already showed that $T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top = O_p(1)$, $\hat\theta - \hat\theta^b_{i\gamma} = O^b_p(T^{-1/2})$, and $\hat A^b = A^0 + o^b_p(1)$. Therefore,

$$\begin{aligned}\hat H^b_{a,i\gamma} &= T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \hat\epsilon_t^2 \eta_t^2 + O^b_p(T^{-1/2})(T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \epsilon^b_t \otimes z_t) + o^b_p(1)\\ &\equiv L^b_1 + L^b_2 + o^b_p(1). \qquad (15)\end{aligned}$$

$$\begin{aligned}L^b_1 = {} & T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \hat\epsilon_t^2\eta_t^2 = T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top (\epsilon_t - w_t^\top(\hat\theta-\theta^0))^2\eta_t^2 = T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \epsilon_t^2\eta_t^2\\ & +[I_q\otimes((\hat\theta-\theta^0)^\top A^0)]\,(T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t^2)\,[I_q\otimes(A^{0\top}(\hat\theta-\theta^0))]\\ & -2[I_q\otimes((\hat\theta-\theta^0)^\top A^0)]\,(T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \otimes \epsilon_t z_t \eta_t^2\,)\\ & = T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \epsilon_t^2\eta_t^2 + O_p(T^{-1/2})\,(T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t^2\,)O_p(T^{-1/2})\\ & +O_p(T^{1/2})\,(T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top \otimes \epsilon_t z_t \eta_t^2\,) \equiv \sum_{i=1}^{3} L^b_{1,i}. \qquad (16)\end{aligned}$$

For any element ξ_t of $z_t z_t^\top \otimes z_t z_t^\top$, $E^b|\xi_t\eta_t^2| = |\xi_t|E^b(\eta_t^2) = |\xi_t|$. So, $T^{-1-\alpha}\sum_{1\gamma} E^b|\xi_t\eta_t^2| \le T^{-\alpha}(T^{-1}\sum_{1\gamma}|\xi_t|) = T^{-\alpha}O_p(1) = o_p(1)$. Therefore, by Markov's inequality (for the bootstrap probability measure), $T^{-1-\alpha}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t^2 = o^b_p(1)$, for any $\alpha > 0$, so $T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t^2 = O^b_p(T^\alpha)$. Therefore, $L^b_{1,2} = O_p(T^{-1/2})\,(T^{-1}\sum_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top \eta_t^2\,)O_p(T^{-1/2}) = O^b_p(T^{\alpha-1}) = o^b_p(1)$, choosing $\alpha < 1$. Similarly, it can be shown that $L^b_{1,3} = o^b_p(1)$. Also, $E^b(L^b_{1,1}) = T^{-1}\sum_{1\gamma} z_t z_t^\top \epsilon_t^2$. Recalling that $E^b(\eta_t^4) = K^*$, noting that $\mathrm{vec}(z_t z_t^\top) = z_t \otimes z_t$, and using $Var(a) = E(aa^\top) - E(a)E(a^\top)$ for a vector a,

$$\begin{aligned}& Var^b(\mathrm{vec}\,L^b_{1,1}\,\mathrm{vec}\,L^{b\top}_{1,1}) = T^{-2}\sum\nolimits_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^4 K^*\\ & + T^{-2}\sum\nolimits_{t=1}^T\sum\nolimits_{s=1,s\neq t}^T z_t z_t^\top \otimes z_s z_s^\top \epsilon_t^2\epsilon_s^2 \mathbf{1}[q_t\le\gamma]\mathbf{1}[q_s\le\gamma]\\ & -(T^{-1}\sum\nolimits_{1\gamma} z_t\otimes z_t\epsilon_t^2)(T^{-1}\sum\nolimits_{1\gamma} z_t^\top \otimes z_t^\top\epsilon_t^2) = T^{-2}\sum\nolimits_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^4(K^*-1).\end{aligned}$$

By Hölder's inequality applied for $p = r/r-1$ and $q = r$, and Assumption 4.2(b),

$$\begin{aligned}E\|z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^4\| &\le (E\|z_t z_t^\top \otimes z_t z_t^\top\|^{r/(r-1)})^{(r-1)/r}(E\|\epsilon_t\|^{4r})^{1/4r}\\ &\le \sup\nolimits_t(E\|z_t\|^{4r/(r-1)})^{(r-1)/(4r)}K = o(T).\end{aligned}$$

Therefore, by Markov's inequality, $T^{-2}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^4(K^*-1) = o_p(1)$. From Chebyshev's inequality, it follows that $L^b_{1,1} - T^{-1}\sum_{1\gamma} z_t z_t^\top \epsilon_t^2 = o_p(1)$. By Assumption 4.1(b),(c) and (e) and Lemma B.1, $T^{-1}\sum_{1\gamma} z_t z_t^\top \epsilon_t^2 \xrightarrow{\mathrm{p}} H_{a,1\gamma}$, therefore $L^b_{1,1} \xrightarrow{\mathrm{p^b}} H_{a,1\gamma}$, so $L^b_1 \xrightarrow{\mathrm{p^b}} H_{a,1\gamma}$.

We now analyze $L^b_2 = O^b_p(T^{-1/2})(T^{-1}\sum_{1\gamma} z_t z_t^\top \epsilon^b_t \otimes z_t)$.

$$\begin{aligned}T^{-1}\sum_{1\gamma} z_t z_t^\top \epsilon^b_t\otimes z_t &= T^{-1}\sum_{1\gamma} z_t z_t^\top \hat\epsilon_t\eta_t\otimes z_t = T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t(\epsilon_t - w_t^\top(\hat\theta-\theta^0))\eta_t\\ &= T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t\epsilon_t\eta_t - (T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top\eta_t)\,[I_q\otimes A^{0\top}(\hat\theta-\theta^0)]\\ &= T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t\epsilon_t\eta_t - (T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top\eta_t)O_p(T^{-1/2}).\end{aligned}$$

For any element ξ_t of $z_t z_t^\top \otimes z_t \epsilon_t$, $E^b|T^{-1-\alpha}\sum_{1\gamma}\xi_t\eta_t| \leq T^{-\alpha}(T^{-1}\sum_{1\gamma}\xi_t)\sup_t E^b|\eta_t| = T^{-\alpha}O_p(1)O(1) = o_p(1)$, give any $\alpha > 0$. Therefore, $T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t\epsilon_t\eta_t = o_p^b(T^\alpha)$. Similarly, $T^{-1}\sum_{1\gamma} z_t z_t^\top \otimes z_t z_t^\top \eta_t = o_p^b(T^\alpha)$. Choose $\alpha < 1/2$; then $L_2^b = O_p^b(T^{-1/2})o_p^b(T^\alpha) = o_p^b(1)$. Substituting this and $L_1^b \xrightarrow{\text{p}^\text{b}} H_{a,1\gamma}$ into (15), it follows that $\hat{H}^b_{a,1\gamma} \xrightarrow{\text{p}^\text{b}} H_{a,1\gamma}$, completing the proof.

THRESHOLD FIRST STAGE

Wlog, assume that $\rho \leq \rho^0$ (the proofs for $\rho > \rho^0$ are similar and omitted for simplicity).
Proof of Theorem 4.1. Part (i).
Let $Q(\hat{\Pi}_{i\hat{\rho}}, \hat{\rho}) \equiv T^{-1}\sum_{t=1}^T \hat{u}_t^\top \hat{u}_t = T^{-1}\sum_{i=1}^2\sum_{1\hat{\rho}}(x_t^\top - z_t^\top\hat{\Pi}_{i\hat{\rho}})(x_t - \hat{\Pi}_{i\hat{\rho}}^\top z_t)$, and $d_t = \hat{u}_t - u_t$. Then, by definition, $Q(\hat{\Pi}_{i\hat{\rho}}, \hat{\rho}) = T^{-1}\sum_{t=1}^T(u_t + d_t)^\top(u_t + d_t) \leq Q(\Pi_i^0, \rho^0) = T^{-1}\sum_{t=1}^T u_t^\top u_t$. This implies that $2T^{-1}\sum_{t=1}^T u_t^\top d_t + T^{-1}\sum_{t=1}^T d_t^\top d_t \leq 0$. We now prove consistency in two steps. In part (i1), we show that $T^{-1}\sum_{t=1}^T d_t^\top d_t = O_p(1)$ and that $T^{-1}\sum_{t=1}^T u_t^\top d_t = o_p(1)$, implying $\text{plim}(T^{-1}\sum_{t=1}^T d_t^\top d_t) \leq 0$, so $T^{-1}\sum_{t=1}^T d_t^\top d_t \xrightarrow{\text{p}} 0$. In part (i2), we show that if $\hat{\rho} \overset{p}{\nrightarrow} \rho^0$, then, with positive probability, $T^{-1}\sum_{t=1}^T d_t^\top d_t > K$ for some $K > 0$, contradicting $T^{-1}\sum_{t=1}^T d_t^\top d_t \xrightarrow{\text{p}} 0$, and therefore yielding $\hat{\rho} \xrightarrow{\text{p}} \rho^0$.
Part (i1). For random variables s_t, let $\sum_\Delta s_t = \sum_{t=1}^T s_t \mathbf{1}[\rho \leq q_t \leq \rho^0]$. Then, for any $\rho \leq \rho^0$ instead of just $\hat{\rho}$, and using $\text{vec}(ABC) = (C^\top \otimes A)\,\text{vec}(B)$, we have:

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

By standard arguments, $\Pi_1^0 - \hat{\Pi}_{1\rho} = o_p(1)$, $\Pi_1^0 - \hat{\Pi}_{2\rho} = O_p(1)$, and $\Pi_2^0 - \hat{\Pi}_{2\rho} = O_p(1)$. By Lemma B.3(ii), $T^{-1}\sum_{1\rho} z_t \otimes u_t = o_p(1)$. Therefore, $T^{-1}\sum_{t=1}^T u_t^\top d_t = o_p(1)$ (uniformly in ρ). Similarly, $T^{-1}\sum_{t=1}^T d_t^\top d_t = O_p(1)$. 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 $T^{-1}\sum_{t=1}^T d_t^\top d_t \xrightarrow{\text{p}} 0$.
Part (i2). By the continuity assumption 4.1(e), there exist an $\epsilon > 0$ such that with positive probability, $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$. If $\hat{\rho} \overset{p}{\nrightarrow} \rho^0$, because $\hat{\rho} \leq \rho^0$, $\hat{\rho} < \rho^0 - \epsilon$. Because $\hat{\Pi}_{2\hat{\rho}}$ is the multivariate LS estimator in the sample $q_t > \rho^0 - \epsilon$, the residuals evaluated over the sub-sample $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$ will also be evaluated at $\hat{\Pi}_{2\hat{\rho}} = \hat{\Pi}_2$. 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 $\sum_{AB} = \sum \mathbf{1}[q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]]$. Then:

$$\begin{aligned}
T^{-1}\sum_{AB} d_t^\top d_t &= T^{-1}\sum\nolimits_{AB}(x_t - \hat{\Pi}_2^\top z_t)^\top(x_t - \hat{\Pi}_2^\top z_t) \\
&= (\Pi_1^0 - \hat{\Pi}_2)^\top(T^{-1}\sum\nolimits_A z_t z_t^\top)(\Pi_1^0 - \hat{\Pi}_2) + (\Pi_2^0 - \hat{\Pi}_2)^\top(T^{-1}\sum\nolimits_B z_t z_t^\top)(\Pi_2^0 - \hat{\Pi}_2) \\
&\geq ||\Pi_1^0 - \hat{\Pi}_2||^2||T^{-1}\sum\nolimits_A z_t z_t^\top|| + ||\Pi_2^0 - \hat{\Pi}_2||^2||T^{-1}\sum\nolimits_B z_t z_t^\top|| \\
&= ||\Pi_1^0 - \hat{\Pi}_2||^2||M_{1\rho^0} - M_{1,\rho^0-\epsilon}|| + ||\Pi_2^0 - \hat{\Pi}_2||^2||M_{2,\rho^0+\epsilon} - M_{2\rho^0}|| + o_p(1).
\end{aligned}$$

Let the minimum eigenvalues of $(M_{1\rho^0} - M_{1,\rho^0-\epsilon})$ and $(M_{2,\rho^0+\epsilon} - M_{2\rho^0})$ be η_1 and η_2 respectively. By Assumption 4.1(d), $\eta = \min(\eta_1, \eta_2) > 0$. Therefore,

$$T^{-1}\sum_{t=1}^{t} d_t^\top d_t \geq T^{-1}\sum_{AB} d_t^\top d_t > \eta(||\Pi_1^0 - \hat{\Pi}_2||^2 + ||\Pi_2^0 - \hat{\Pi}_2||^2)$$
$$+ o_p(1) \geq ||\Pi_1^0 - \Pi_2^0||^2\eta/2 + o_p(1) > K + o_p(1),$$

where for the second to last inequality, we used the fact that for any vectors a, b, c of the same length, $(a-b)^\top(a-b) + (c-b)^\top(c-b) \geq (a-c)^\top(a-c)/2$. This means that $\text{plim}\, T^{-1}\sum_{t=1}^{t} d_t^\top d_t > K$ with positive probability, reaching a contradiction.

Part (ii). For $\rho = \rho^0$, the proof is obvious. Therefore, we consider the case $\rho^0 - \hat{\rho} > 0$, and let $\rho^0 - \hat{\rho} \leq \epsilon$, now for a chosen small $\epsilon > 0$, as we know $\hat{\rho} \xrightarrow{p} \rho^0$. Let $V_\epsilon(C) = \{\rho : 0 < \rho^0 - \hat{\rho} \leq \epsilon, T(\rho^0 - \hat{\rho}) > C\}$ for some chosen large enough C. For the rest of the proof of part (ii) of this theorem, we write ρ instead of $\hat{\rho}$ for simplicity. Let S_1 be the multivariate sum of squared residuals in (2) evaluated at ρ and parameters $\hat{\Pi}_1, \hat{\Pi}_2$, S_2 be the multivariate sum of of squared residuals evaluated at ρ^0 with parameters $\hat{\Pi}_{1\rho^0}, \hat{\Pi}_{2\rho^0}$, and let S_3 be the multivariate sum of squared residuals when regressing x_t on $z_t\mathbf{1}[q_t \leq \rho]$, $z_t\mathbf{1}[\rho < q_t \leq \rho^0]$, and $z_t\mathbf{1}[q_t > \rho^0]$, evaluated at parameters $\hat{\Pi}_{1\rho}, \hat{\Pi}_\Delta, \hat{\Pi}_{2\rho^0}$, where $\hat{\Pi}_\Delta = (\sum_\Delta z_t z_t')^{-1}\sum_\Delta z_t x_t'$. Then, by definition, $\frac{S_1 - S_2}{T(\rho^0-\rho)} \leq 0$, We will now show that $\frac{S_1-S_2}{T(\rho^0-\rho)} = O_p(1)$, therefore plim $\frac{S_1-S_2}{T(\rho^0-\rho)} \leq 0$. We also show that if $\rho \in V_\epsilon(C)$, then, for sufficiently large C and small enough ϵ, plim $\frac{S_1-S_2}{T(\rho^0-\rho)} > K$, for some positive constant K, reaching a contradiction and establishing that $T(\rho^0 - \hat{\rho}) \leq C$.

Note that $S_1 - S_2 = (S_1 - S_3) - (S_2 - S_3)$. Letting $\hat{Q}_\Delta = \frac{1}{T(\rho^0-\rho)}\sum_\Delta z_t z_t^\top$ and $\hat{M}_\Delta = T^{-1}\sum_\Delta z_t z_t^\top$, by Bai and Perron (1998), pp. 70,

$$\frac{S_1 - S_2}{T(\rho^0 - \rho)} = \text{trace}[(\hat{\Pi}_{2\rho^0} - \hat{\Pi}_\Delta)^\top(\hat{Q}_\Delta - \hat{Q}_\Delta\hat{M}_{2\rho}^{-1}\hat{M}_\Delta)(\hat{\Pi}_{2\rho^0} - \hat{\Pi}_\Delta)]$$
$$\frac{S_2 - S_3}{T(\rho^0 - \rho)} = \text{trace}[(\hat{\Pi}_{1\rho} - \hat{\Pi}_\Delta)^\top(\hat{Q}_\Delta - \hat{Q}_\Delta\hat{M}_{1\rho}^{-1}\hat{M}_\Delta)(\hat{\Pi}_{1\rho} - \hat{\Pi}_\Delta)].$$

By standard arguments, $\hat{\Pi}_{2\rho^0} = \Pi_2^0 + o_p(1)$, $\hat{\Pi}_\Delta = \Pi_1^0 + o_p(1)$, and by Lemma B.3, $\Pi_{1\rho} = \Pi_1 + o_p(1)$. Therefore, $\hat{\Pi}_{2\rho^0} - \hat{\Pi}_\Delta = \Pi_2^0 - \Pi_1^0 + o_p(1) = O_p(1)$, and $\hat{\Pi}_{1\rho} - \hat{\Pi}_\Delta = o_p(1)$. For C large enough, $\hat{Q}_\Delta = O_p(1)$ and $\hat{M}_{i\rho} = O_p(1)$ for $i = 1, 2$. On the other hand,

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

where the last equality holds by continuity and boundedness of $f(x)$(Assumption 4.1(e)) and for some $b \in [\rho^0 - \epsilon, \rho^0]$. Therefore, $\hat{Q}_\Delta - \hat{Q}_\Delta\hat{M}_{2\rho}^{-1}\hat{M}_\Delta = \hat{Q}_\Delta + O_p(\epsilon) = O_p(1)$. This implies that $\frac{S_1-S_3}{T(\rho^0-\rho)} = O_p(1)$, and that $\frac{S_2-S_3}{T(\rho^0-\rho)} = o_p(1)$, therefore $\frac{S_1-S_2}{T(\rho^0-\rho)} = O_p(1)$, which completes the first part of the proof.

It also follows that the asymptotically dominant term in $\frac{S_1-S_2}{T(\rho^0-\rho)}$ is $N_1 = \text{trace}[(\Pi_2^0 - \Pi_1^0)^\top\hat{M}_\Delta(\Pi_2^0 - \Pi_1^0)]$. For $\rho \in V_\epsilon(C)$, and C large enough, $\hat{M}_\Delta = \frac{M_{1\rho^0} - M_{1\rho}}{\rho^0-\rho} + o_p(1)$, and because $M_{1\rho^0} - M_{1\rho}$ is p.d. by Assumption 4.1(d), it follows that $\frac{S_1-S_2}{T(\rho^0-\rho)} = N_1 + O_p(\epsilon) > K > 0$, for some positive constant K. This completes the second part of the proof of (i), because it contradicts plim $\frac{S_1-S_2}{T(\rho^0-\rho)} \leq 0$. Therefore, $\hat{\rho} \notin V_\epsilon(C)$, and so there is a C for which $|T(\rho^0 - \rho)| < C$.

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 $|T(\rho^0 - \hat{\rho})| < C$ terms, which are uniformly bounded by Assumption 4.1(c), it follows that $T^{1/2}\,\mathrm{vec}(\hat{\Pi}_i - \Pi_i^0) = T^{1/2}\,\mathrm{vec}(\hat{\Pi}_{i\rho^0} - \Pi_i^0) + o_p(1)$, for $i = 1, 2$. The rest of the proof follows standard arguments.

By Theorem 4.1 and its proof, wlog, we treat below $\hat{\rho}$ as if it was equal to ρ^0.

Notation specific for a TFS. 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 $w_t = A_1^0 z_t \mathbf{1}[q_t \leq \rho^0] + A_2^0 z_t \mathbf{1}[q_t > \rho^0] + (u_t^\top, \mathbf{0}_{1\times q_1})^\top$. Let $A_t^0 = A_1^0 \mathbf{1}[q_t \leq \rho^0] + A_2^0 \mathbf{1}[q_t > \rho^0]$, $\hat{A}_t = \hat{A}_1 \mathbf{1}[q_t \leq \rho^0] + \hat{A}_2 \mathbf{1}[q_t > \rho^0]$, $\Pi_t = \Pi_1^0 \mathbf{1}[q_t \leq \rho^0] + \Pi_2^0 \mathbf{1}[q_t > \rho^0]$, $\hat{\Pi}_t = \hat{\Pi}_1 \mathbf{1}[q_t \leq \rho^0] + \hat{\Pi}_2 \mathbf{1}[q_t > \rho^0]$.

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:

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

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

Definition B.2. *Let $V_\mathcal{B} = \sum_{i=1}^2 A_i^0\, D\, H_{i,\gamma}\, D^\top A_i^0$,*

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

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

and $V_{\mathcal{B},2\gamma} = V_\mathcal{B} + V_{\mathcal{B},1\gamma} - V_{\mathcal{B},12,\gamma} - V_{\mathcal{B},12,\gamma}^\top$. 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 $\hat{V}_\gamma$ 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 $\bar{D}_{i\gamma}$, $F_{1\gamma}$ 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 B.3-B.4 and Theorem 4.2, for x_t generated by the TFS (3) instead of the LFS (2).

Proof of Lemma B.3. Part (i). $T^{-1}\sum_{i\gamma} z_t z_t^\top \xrightarrow{\text{p}} M_{i\gamma}$ still holds, as the result is not specific to a LFS or TFS. So,

$$\begin{aligned}
T^{-1}\sum\nolimits_{1\gamma} \hat{w}_t\hat{w}_t^\top &= \hat{A}_1 \sum\nolimits_{1,\rho^0\wedge\gamma} z_t z_t^\top \hat{A}_1^\top + \hat{A}_2(T^{-1}\sum\nolimits_{1\gamma} z_t z_t^\top - T^{-1}\sum\nolimits_{1,\rho^0\wedge\gamma} z_t z_t^\top)\hat{A}_2^\top \\
&\xrightarrow{\text{p}} A_1^0 M_{1,\rho^0\wedge\gamma} A_1^{0\top} + A_2^0(M_{1\gamma} - M_{1,\rho^0\wedge\gamma})A_2^{0\top} = C_{1\gamma} \\
T^{-1}\sum\nolimits_{2\gamma} \hat{w}_t\hat{w}_t^\top &= \hat{A}_1(T^{-1}\sum\nolimits_{1,\rho^0\vee\gamma} z_t z_t^\top - T^{-1}\sum\nolimits_{1,\gamma} z_t z_t^\top)\hat{A}_1^\top + \hat{A}_2 T^{-1}\sum\nolimits_{2\rho^0\vee\gamma} z_t z_t^\top \hat{A}_2^\top \\
&\xrightarrow{\text{p}} A_1^0(M_{1,\rho^0\vee\gamma} - M_{1\gamma})A_1^{0\top} + A_2^0 M_{2\rho^0\vee\gamma} A_2^{0\top} = C_{2\gamma}.
\end{aligned}$$

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

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

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

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

For $\gamma > \rho^0$,

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

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

Proof of Theorem 4.2. 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 B.3.
Part (ii). Follows the same steps as for the LFS proof until equation (10). Then note that because $\hat{\Pi}_i - \Pi_i^0 = o_p(1)$,

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

following the same arguments as in the LFS proof.

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

Proof of Lemma B.4 and Theorem 4.2. As evident from the proof of Theorem B.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.