TESTING FOR A THRESHOLD IN MODELS WITH ENDOGENOUS REGRESSORS

By

Mario P. Rothfelder, Otilia Boldea

30 October 2019

ISSN 0924-7815
ISSN 2213-9532
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 known1, 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-

1See 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.

2For 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 \textit{apriori}.\textsuperscript{3}

Caner and Hansen (2004) (CH henceforth) proposed a Wald test for exactly this hypothesis based on GMM (generalized method of moments) estimation.\textsuperscript{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

\textsuperscript{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.

\textsuperscript{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 \( O \left( \frac{T - 1}{2} \right) \) 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 \( O \left( \frac{T - 1}{2} \right) \) 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 \( \gamma^0 \):

\[
y_t = (x_t^\top \theta_{1x}^0 + z_{1t}^\top \theta_{1z}^0) 1[q_t \leq \gamma^0] + (x_t^\top \theta_{2x}^0 + z_{1t}^\top \theta_{2z}^0) 1[q_t > \gamma^0] + \epsilon_t
\]

\[
= w_t^\top \theta_1^0 1[q_t \leq \gamma^0] + w_t^\top \theta_2^0 1[q_t > \gamma^0] + \epsilon_t,
\]

(1)

where \( q_t \) 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, \( 1[] \) is the indicator function, \( w_t = (x_t, z_{1t}^\top)^\top \) and \( \theta_i^0 = (\theta_{1x}^0, \theta_{1z}^0)^\top \). 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

\[
x_t = \Pi_0^0 z_t + u_t,
\]

(2)

or a threshold model:

\[
x_t = \Pi_1^0 z_t 1[q_t \leq \rho^0] + \Pi_2^0 z_t 1[q_t > \rho^0] + u_t,
\]

(3)
where \( \rho^0 \in \Gamma \), and \( z_t \) are \( q \times 1 \) strong and valid instruments, including \( z_{it} \), with \( q \geq p_1 \). We assume that \( E((\varepsilon_t, u_t)|F_t) = 0 \), where \( F_t = \{ q_t, q_{t-1}, \ldots, z_t, z_{t-1}, \ldots, x_{t-1}, \ldots \} \), 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. \( H_0 : \theta^0_1 = \theta^0_2 \equiv \theta^0 \). Such a test is already available from CH.\(^5\) They proposed a test based on GMM estimators of \( \theta^0_i \), (\( i = 1, 2 \)) for each \( \gamma \in \Gamma \). Because \( z_t \) and \( q_t \) are exogenous, the moment conditions (4) hold for all \( \gamma \in \Gamma \).

Based on these moment conditions, they construct the usual two-step GMM estimators:

\[
\hat{\theta}_{\gamma_t}(2) = (\hat{N}_{\gamma_t}^{-1}\hat{\Omega}_{\gamma_t}^{-1}\hat{N}_{\gamma_t}^{-T})\hat{N}_{\gamma_t}^{-1}(T^{-1}\sum_{t=1}^{T} z_{it} q_t),
\]

where \( \hat{N}_{\gamma_t} = T^{-1}\sum_{t=1}^{T} w_{it} z_{it}^T \), \( \hat{\Omega}_{\gamma_t} = T^{-1}\sum_{t=1}^{T} z_{it}^T z_{it} \), \( \sum_{t=1}^{T} (\cdot) = \sum_{t=1}^{T} 1[q_t \leq \gamma] \), \( \sum_{t=1}^{T} (\cdot) = \sum_{t=1}^{T} 1[q_t > \gamma] \) are the first step GMM residuals, and \( \hat{\theta}_{\gamma_t}(1) \) are consistent first-step versions of \( \hat{\theta}_{\gamma_t}(2) \), for example by replacing \( \hat{\Omega}_{\gamma_t} \) with \( M_{\gamma_t} = T^{-1}\sum_{t=1}^{T} z_{it} z_{it}^T \). These estimators can be used to construct a Wald test for each \( \gamma \), and taking the maximum of this sequence of Wald tests over \( \gamma \in \Gamma \) yields the test in CH:

\[
WG_T = \sup_{\gamma \in \Gamma} WG_T(\gamma) = \sup_{\gamma \in \Gamma} \left( \hat{\theta}_{\gamma_t}(2) - \hat{\theta}_{\gamma_t}(2) \right)^\top \hat{V}_{\gamma_t}(1) \left( \hat{\theta}_{\gamma_t}(2) - \hat{\theta}_{\gamma_t}(2) \right),
\]

where \( \hat{V}_{\gamma_t}(1) = \sum_{t=1}^{T} \hat{V}_{\gamma_t}(1) \) and \( \hat{V}_{\gamma_t}(1) = (\hat{N}_{\gamma_t}^{-1}\hat{\Omega}_{\gamma_t}^{-1}\hat{N}_{\gamma_t}^{-1})^{-1} \).

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 \( \gamma \), 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 \( \gamma^0 \) under the alternative hypothesis \( H_A : \theta^0_1 \neq \theta^0_2 \), 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 \( \gamma^0 \) 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 \( \gamma^0 \).
The LFS (2) is estimated over the full sample:

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

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

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

where \( \hat{\Pi}_{\rho} = \left( \sum_{t=1}^{T} z_t z_t^\top \right)^{-1} \sum_{t=1}^{T} z_t x_t^\top \) for \( i = 1, 2 \), and \( \hat{\Pi}_{\rho t} = \hat{\Pi}_{\rho} 1[q_t \leq \rho] + \hat{\Pi}_{\rho t} 1[q_t > \rho] \).

At this threshold estimator,

\[ \hat{\Pi}_i = \hat{\Pi}_{i\hat{\rho}}, \quad \hat{x}_t = \hat{\Pi}_{i\hat{\rho}}^T z_t. \]

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

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

We now propose two tests for \( H_0 : \theta_1^0 = \theta_2^0 \) and an unknown threshold. The first one is a likelihood-ratio type test:

\[ LR_T = \sup_{\gamma \in \Gamma} LR_T(\gamma) = \sup_{\gamma \in \Gamma} \frac{SSR_0 - SSR_1(\gamma)}{SSR_1(\gamma)/(T - 2p)} \]  

(6)

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

The second one is a Wald-type test:

\[ W_T = \sup_{\gamma \in \Gamma} W_T(\gamma) = \sup_{\gamma \in \Gamma} \left( \hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma} \right)^\top V_\gamma^{-1} \left( \hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma} \right), \]

(7)

where \( V_\gamma = \lim \text{Var} \left(T^{1/2}(\hat{\theta}_{1\gamma} - \hat{\theta}_{2\gamma})\right) \), and the explicit expression for \( V_\gamma \) 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 \( V_\gamma \) 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 \( \rho^0 \) and \( \gamma \).

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 \( \Gamma \), but only consider unique values of \( q_t \) taken in the sample, from the \( \eta \) to the \((1 - \eta)\) quantile, where typically \( \eta \in [0.10, 0.25] \).
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 \( \eta_t \) described in Assumption 4.2 to preserve the potential conditional heteroskedasticity in the data.\(^8\) A bootstrap sample is computed always under \( H_0 : \theta^0 = \theta^2 \), as follows:

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

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

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

Then the bootstrap equivalent of \( LRT \) is \( \sup_{\gamma \in \Gamma} LRT(\gamma) \) is \( \sup_{\gamma \in \Gamma} LRT^b(\gamma) \), where \( LRT^b(\gamma) \) is computed in the same way as \( LRT(\gamma) \), but with \( SSR_0 \) replaced by \( SSR^b_0 = \sum_{t=1}^{T} (\hat{y}_t^b - \hat{\theta}_T^b \hat{\eta}_t^b)^2 \) and \( SSR_1(\gamma) \) by \( SSR^b_1(\gamma) = \sum_{t=1}^{T} \sum_{i=1}^{T} (\hat{y}_t^b - \hat{\theta}_T^b \hat{\gamma}_i + \hat{\epsilon}_t^b)^2 \).

The bootstrap equivalent of \( W_T = \sup_{\gamma \in \Gamma} W_T(\gamma) \) is \( W^b_T = \sup_{\gamma \in \Gamma} W^b_T(\gamma) \), where \( W^b_T(\gamma) = T(\hat{\theta}_T^b - \hat{\theta}_T^b)^\top (\hat{\Pi}^b T z_{1_t}^b)^{-1} (\hat{\theta}_T^b - \hat{\theta}_T^b) \), with \( \hat{\theta}_T^b \) obtained as \( \hat{\gamma}_T \) 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 \( H_0 \). 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 \( \hat{\Omega} = T^{-1} \sum_{t=1}^{T} z_t z_t^\top, \hat{\epsilon}_{t(1)} = y_t - \hat{w}_t^\top \hat{\theta}^{(1)} \), 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 \( \hat{N} = T^{-1} \sum_{t=1}^{T} \hat{w}_t z_t^\top \);
- obtain the two-step residuals \( \hat{\epsilon}_{t(2)} = y_t - \hat{w}_t^\top \hat{\theta}^{(2)} \);
- let \( \hat{\epsilon}_t^b = \hat{\epsilon}_{t(2)} \eta_t \), where \( \eta_t \sim i.i.d.(0, 1) \);
- let \( \hat{y}_t^b = \hat{w}_t^\top \hat{\theta}^{(2)} + \hat{\epsilon}_t^b \).

\(^8\)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 \( \hat{\beta}_{(1)} \) instead of \( \hat{\epsilon}_{t,(1)} \) and \( \hat{\beta}_{(2)} \) instead of \( \hat{\epsilon}_{t,(2)} = y_t - \omega_T^\top (\hat{\beta}_{(1)} 1[q_t \leq \gamma] + \hat{\beta}_{(2)} 1[q_t > \gamma]) \); (2) that we add back the conditional mean for \( y_t \) instead of writing \( \hat{y}_t = \hat{\beta} \), for the purposes of better replicating the variation in \( y_t \). Then the bootstrap equivalent of the GMM test is \( WG_T^b = \sup_{\gamma \in T} W G_T^b \), with

\[
WG_T^b = T \left( \hat{\beta}_{(1),(2)}^b - \hat{\beta}_{(2),(2)}^b \right)^\top \left( \hat{V}_{\gamma,(1)}^b \right)^{-1} \left( \hat{\beta}_{(1),(2)}^b - \hat{\beta}_{(2),(2)}^b \right),
\]

where \( \hat{\beta}_{(1),(2)}^b = (\hat{\theta}_{(1),(1)}, \hat{\theta}_{(1),(2)}, \hat{\theta}_{(2),(1)}, \hat{\theta}_{(2),(2)}) \), \( \hat{\theta}_{(i),(j)}^b = (\hat{\theta}_{(i),(j)}(1)^b, \hat{\theta}_{(i),(j)}(2)^b) \), \( \hat{V}_{\gamma,(i),(j)}^b = \sum_{t=1}^T \hat{V}_{\gamma,(i),(j)}(1)^b + \hat{V}_{\gamma,(i),(j)}(2)^b \), \( \hat{V}_{\gamma,(i)}^b = (\hat{\Omega}_{(i)}(1)^b, \hat{\Omega}_{(i)}(2)^b) \), \( \hat{\Omega}_{(i)}(1)^b = \sum_{t=1}^T \hat{\theta}_{(i),(1)}(1)^b z_t z_t^\top \), and \( \hat{\Omega}_{(i)}(2)^b = \sum_{t=1}^T \hat{\theta}_{(i),(2)}(2)^b z_t z_t^\top \).

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 \( q_t = x_t - u_t \), \( h_t = y_t - \epsilon_t \), \( M_1(\gamma) = E[z_t z_t^\top 1[q_t \leq \gamma]] \), \( M = M(\gamma_{\max}) = E[z_t z_t^\top] \), where \( \Gamma_0 = [\gamma_{\min}, \gamma_{\max}] \) is the support of \( q_t \). \( M(\gamma) = M - M_1(\gamma) \), and \( v_t = (\epsilon_t, u_t)^\top \). Let \( \| \cdot \| \) be the Euclidean norm. The following assumptions are similar to CH.

Assumption 4.1.

(a) \( E[v_t | \mathcal{F}_t] = 0 \) with \( \mathcal{F}_t = \sigma\{z_{t-s}, v_{t-s-1}, q_{t-s}, s \geq 0\} \); (b) The series \( (v_t, q_t, h_t, g_t, z_t) \) is strictly stationary with \( \rho(m) = \mathcal{O}(m^{-A}) \) for some \( A > \frac{a}{2} \) and \( 1 < a \leq r \); (c) \( \sup_t E[\|z_t\|^{4r}] < \infty, \sup_t E[\|v_t\|^{4r}] < \infty \); (d) \( \inf_t \det M_1(\gamma) > 0 \), and if (3) holds, then \( M_1(\gamma) \) is strictly increasing in \( \gamma \) (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) \( E[v_t v_t^\top] \) and \( E[(v_t v_t^\top) \otimes (z_t z_t^\top)] \) are two p.d. matrices of constants; (g) The coefficient matrices \( \Pi_0^b \) (for the LFS (2)) or \( \Pi_1^b, \Pi_2^b \) (for the TFS (3)) are full rank, and \( \Pi_1^b - \Pi_2^b \neq 0 \).

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 \( 1[q_t \leq \gamma] \) 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 \( \gamma \), 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) $\hat{\rho} - \rho^0 \xrightarrow{p} 0$; (ii) $T(\hat{\rho} - \rho^0) = O_p(1)$; (iii) $T^{1/2} \text{vec}(\Pi_i - \Pi_i^0) = T^{1/2} \text{vec}(\Pi_i - \Pi_i^0) + o_p(1)$.

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 $\eta_t$ 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 $W^b_T$ test, in particular for $V^b_\gamma$ to weakly converge to $V_\gamma$ 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) $\sup_{c \in \mathbb{R}} P^b(\{|LR^b_T - LR_T| \leq c\}) \xrightarrow{p} 0$ under Assumptions 4.1-4.2(a); (ii) $\sup_{c \in \mathbb{R}} P^b(\{|W^b_T - W_T| \leq c\}) \xrightarrow{p} 0$, under Assumptions 4.1-4.2.

This theorem proves bootstrap validity of the proposed test statistics. The bootstrap validity of $WG^b_T$ 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 $t = 1, \ldots, T$:

$$y_t = 1 + x_t + \delta_x x_t 1[q_t > \gamma^0] + \epsilon_t, \quad x_t = 1 + z_t + \delta_{\Pi} z_t 1[q_t > \rho^0] + u_t,$$
of the bootstrap distribution of the test statistic, and if the test in the original sample bootstrap as described in section 3. Precisely estimate $\Omega$ realizations of $\eta$ 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 $\nu_t \sim N(0, 1)$. 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, $\epsilon_t^{i.i.d.} \sim N(0, \Sigma_t)$ with $\hat{\Sigma}_t = T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t^\top$ for 2SLS. For GMM, $\epsilon_t^{i.i.d.} \sim N(0, \sigma_t^2)$ with $\hat{\sigma}_t^2 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^{2,(1)}$, and $\hat{\epsilon}_t^{2,(1)} = y_t - w_t^\top \hat{\theta}$, where recall that $\hat{\theta}$ is the full-sample 2SLS estimator. Second, all heteroskedasticity-robust estimators are replaced by their homoskedastic analogs. For example, $E[z_t z_t^\top \epsilon_t^{2} 1\{q_t \leq \gamma\}]$ is no longer estimated by $T^{-1} \sum_{t=1}^T z_t z_t^\top \hat{\epsilon}_t^{2} 1[q_t \leq \gamma]$, but by $\hat{\sigma}_t^2(T^{-1} \sum_{t=1}^T z_t z_t^\top 1[q_t \leq \gamma])$.

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 - $\epsilon_t = \nu_t \cdot z_t/\sqrt{T}$ with $Var(\epsilon_t) = Var(\nu_t) = 1$ and $Cov(\nu_t, \nu_t) = 0.5$, and we use the same heteroskedasticity-robust variance estimators. In table 1 cases (b)-(c), the bootstrap is performed using $\eta_t \sim N(0, 1)$ 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. $\gamma$ 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 $\hat{\Omega}_\gamma$, for small and moderate sample sizes pertinent to applications. In particular, when $\gamma$ is close to the 15% or 85% quantiles of $q_t$, there is not enough data to precisely estimate $\hat{\Omega}_\gamma$. In the Online supplement, Section S1, we compute $\hat{\Omega}_\gamma$, $\hat{\Omega}_\gamma^*$, and $\hat{\Omega}_\gamma^{10}$ for a sample size of $T = 100$ and case (c). We show that the entries of $\hat{\Omega}_\gamma$ can be 100 times larger than their limits in $\hat{\Omega}_\gamma^*$, and also twice as large or five times smaller than $\hat{\Omega}_\gamma^{10}$.

9Note that because of just-identification, there is no difference between the first and the second-step GMM estimators, therefore $\hat{\theta}_t^{(2)} = \hat{\theta}_t^{(1)} = (\sum_{t=1}^T \nu_t z_t^\top)^{-1}(T^{-1} \sum_{t=1}^T \nu_t \nu_t^\top z_t y_t)$, and $\hat{\theta}_t^{(2), (1)} = (N_t^{-1})^{-1}(T^{-1} \sum_{t=1}^T z_t y_t)$. Also, $\hat{\nu}_t^{(1)} = (\sum_{t=1}^T \nu_t z_t^\top)^{-1}(T^{-1} \sum_{t=1}^T z_t y_t)$. $\hat{V}_{\gamma_t, t}^{(1)} = (\hat{\Omega}_\gamma^{-1} N_t^{-1})^{-1}$ and $\hat{\epsilon}_{t, (2)} = \hat{\epsilon}_{t, (1)} = y_n - w_i^\top \nu_i \gamma_i 1[q_n \leq \gamma] - w_i^\top \hat{\epsilon}_2 1[q_n > \gamma]$.
their bootstrap equivalents in $\hat{\Omega}_\gamma$, when $\gamma$ 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 $\gamma$ is equal to the median.\footnote{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.}

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 $\gamma$ 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 $\hat{\Omega}_\gamma$ 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.\footnote{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.} For simplicity, we first focus on the instantaneous government spending multiplier $\theta_{g,i}(i = 1, 2)$, estimated similarly to RZ from:

$$y_t = (\theta_{g,1} x_t + z_{1,t}^T \theta_{z,1}) 1[q_t \leq \gamma^0] + (\theta_{g,2} x_t + z_{1,t}^T \theta_{z,2}) 1[q_t > \gamma^0] + \epsilon_t$$

$$g_t = \Pi_1^T z_t 1[q_t \leq \rho^0] + \Pi_2^T z_t 1[q_t > \rho^0] + v_t$$

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 $z_{1,t}$ are also included in $z_t$ and contain an intercept and four lags of $g_t, y_t, m_t$. Thus, $z_t = [z_{1,t}^T, m_t^T]^T$.

The data is from the RZ replication package which can be found at \url{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.
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 $\theta_i^0$ 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).\footnote{All numbers referring to unemployment rates, such as 6.5, should be interpreted as percentages: 6.5%.} 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 $\gamma^0$ is known or even that there is a threshold $\gamma^0$; we instead test for the presence of $\gamma^0$ first.

Our methods require first estimating $\rho^0$ in equation (9).\footnote{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.} 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 $LR^T_{2b}$ and $W^b_T$ tests described in Section 3, which were proposed in Hansen (1996). The estimate of $\rho^0$ 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 $\rho^0$. 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).\footnote{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.} 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
multipliers $\theta^h_{g,i} (i = 1, 2)$ at horizon $h = 0, 1, \ldots, H$ from the IV regression:

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

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 $q_t \leq 4.0636$ (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 Cížek, 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 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.
### Table 1. Rejection frequencies under the null DGP and 5% nominal size

| Case | \( T \) | \( LFS \) | \( TFS, \delta = -0.5 \) | \( TFS, \delta = 0.5 \) | \( TFS, \delta = 1 \) |
|------|---------|---------|-----------------|-----------------|-----------------|
|      |         | \( LR \) | \( W \) | \( WG \) | \( LR \) | \( W \) | \( WG \) | \( LR \) | \( W \) | \( WG \) | \( LR \) | \( W \) | \( WG \) |
| (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% | 1.6% | 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, \( \eta_t \sim \mathcal{N}(0, 1) \). \( LR \) 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 | \( T \) | \( LFS \) | \( TFS, \delta = -0.5 \) | \( TFS, \delta = 0.5 \) | \( TFS, \delta = 1 \) |
|------|---------|---------|-----------------|-----------------|-----------------|
|      |         | \( LR \) | \( W \) | \( WG \) | \( LR \) | \( W \) | \( WG \) | \( LR \) | \( W \) | \( WG \) | \( LR \) | \( W \) | \( WG \) |
| (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% | 3.0% | 4.3% | 7.6% |
|      | 1000    | 4.3%   | 4.9% | 9.5% | 3.3% | 4.0% | 12.3% | 3.6% | 3.9% | 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% | 7.6% |
|      | 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 \( \eta_t \sim \mathcal{N}(0, 1) \) 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  | Yes    | 29.335 | 28.775 | Yes     | 178.296| 198.705| No     |
| 15%   | 3.5264           | 10.7000        | 78.158  | Yes    | 27.698 | 28.775 | No      | 66.523 | 67.292 | No     |
| 20%   | 3.7300           | 9.3443         | 75.332  | Yes    | 27.505 | 29.772 | No      | 66.523 | 67.292 | No     |
| 25%   | 4.0636           | 8.3363         | 65.719  | Yes    | 27.505 | 29.772 | No      | 66.523 | 67.292 | No     |

"5%CV," 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 | $p$-val. | AR $p$-val. | $h$ | 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 | 1.11 0.29 | -0.02 | 2.24  -1.92 1.54 | -7.95 4.10 | 0.04 0.24 |
| 1   | 0.89 0.19   | 0.13  | 1.64  | -0.17  0.25 | -1.16  0.81 | 0.00 0.24 | 0.64 0.12 | 0.17  | 1.12  0.46 0.14 | -0.09 1.01 | 0.26 0.39 |
| 2   | 0.71 0.14   | 0.15  | 1.28  | 0.22   0.16 | -0.42  0.87 | 0.01 0.25 | 0.62 0.09 | 0.26  | 0.99  0.59 0.11 | 0.17 1.01 | 0.81 0.82 |
| 3   | 0.64 0.12   | 0.17  | 1.12  | 0.46   0.14 | -0.09  1.01 | 0.26 0.39 | 0.59 0.09 | 0.24  | 0.95  0.60 0.10 | 0.23 0.97 | 0.95 0.95 |
| 4   | 0.62 0.09   | 0.26  | 0.99  | 0.59   0.11 | 0.17  1.01 | 0.81 0.82 | 0.65 0.09 | 0.23  | 0.95  0.65 0.09 | 0.23 0.95 | 0.82 0.82 |
| 5   | 0.67 0.09   | 0.19  | 1.15  | 0.68   0.05 | 0.47  0.89 | 0.99 0.99 | 0.59 0.09 | 0.24  | 1.13  0.68 0.05 | 0.47 0.89 | 0.99 0.99 |

"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 | $p$-val. | AR $p$-val. | $h$ | 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 | 1.14 0.25 | 0.17  | 2.11  -1.68 1.37 | -7.05 3.68 | 0.03 0.23 |
| 1   | 0.93 0.16   | 0.28  | 1.57  | -0.55  0.48 | -2.42  1.33 | 0.00 0.25 | 0.74 0.13 | 0.25  | 1.23  -0.02 0.20 | -0.80 0.76 | 0.00 0.26 |
| 2   | 0.67 0.11   | 0.24  | 1.09  | 0.32   0.17 | -0.36  1.00 | 0.07 0.31 | 0.65 0.10 | 0.28  | 1.03  0.52 0.18 | -0.19 1.22 | 0.45 0.52 |
| 3   | 0.63 0.10   | 0.26  | 1.01  | 0.62   0.19 | -0.14  1.38 | 0.94 0.94 | 0.60 0.09 | 0.24  | 0.96  0.66 0.20 | -0.13 1.45 | 0.77 0.78 |
| 4   | 0.60 0.09   | 0.25  | 0.94  | 0.66   0.18 | -0.04  1.37 | 0.71 0.73 | 0.62 0.09 | 0.29  | 0.96  0.66 0.15 | 0.07 1.25 | 0.84 0.85 |
| 5   | 0.65 0.09   | 0.31  | 0.10  | 0.66   0.13 | 0.14  1.18 | 0.96 0.96 | 0.65 0.09 | 0.32  | 1.02  0.65 0.12 | 0.17 1.13 | 0.87 0.87 |
| 6   | 0.67 0.09   | 0.32  | 1.03  | 0.64   0.12 | 0.17  1.11 | 0.85 0.85 | 0.66 0.09 | 0.30  | 1.03  0.67 0.11 | 0.22 1.11 | 0.98 0.98 |
| 7   | 0.65 0.10   | 0.27  | 1.03  | 0.70   0.11 | 0.28  1.11 | 0.75 0.76 | 0.65 0.10 | 0.27  | 1.03  0.70 0.11 | 0.28 1.11 | 0.75 0.76 |
| 8   | 0.64 0.11   | 0.22  | 1.05  | 0.71   0.10 | 0.31  1.18 | 0.58 0.60 | 0.66 0.11 | 0.21  | 1.03  0.71 0.10 | 0.31 1.18 | 0.58 0.60 |

See table 5 notes.
Figure 1. Size-adjusted power plots for $\delta_x = 0.25$.

(a) $\delta_{\Pi} = -0.5$

(b) $\delta_{\Pi} = 0.5$

(c) $\delta_{\Pi} = 1$
Figure 2. Plot of $LR_T(\gamma)$, $W_T(\gamma)$ and $WG_T(\gamma)$

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: $\| P \| = \sqrt{\text{tr}(P^TP)}$. 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 $\gamma$. “$\Rightarrow$” stands for weak convergence in Skorokhod metric, “$\Rightarrow_b$” for weak convergence in Skorokhod metric under the bootstrap measure, and “$p_b \rightarrow$” 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 $\rho$-mixing series, with $\rho$-mixing coefficient $\rho(m) = O(m^{-A})$ for some $A > a$ 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 $\sup_{\gamma \in \Gamma} |T^{-1} \sum_{t=1}^T a_t - E[a_t 1[q_t \leq \gamma]| L_0 0.$

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 $\rho$-mixing implies ergodicity. Second, set in the proof of their Lemma 1 $w_t = (a_t, q_t)$, $\phi(w_t) = a_t$, and $\{w_t \leq \gamma\} = 1[q_t \leq \gamma]$. Follow the steps in Hansen (1996), until their equation (15). Then note that $\sup_t \| a_t \|_{1+\delta} \leq K$ 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 $\epsilon$ 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 $q = \delta/(1 + \delta)$):

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

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

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

**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) $E[|a_t|1|q_t \geq \gamma] = F_\gamma$, a p.d. matrix of constants, (iii) $\inf_{x\in\mathbb{R}} \det F_\gamma > 0$, then: $T^{-1/2}\sum_{t=1}^T a_t \Rightarrow \mathcal{F}(\gamma)$, a vector of Gaussian processes with covariance function $E[a_ta_t']1(q_t \leq (\gamma_1 \wedge \gamma_2))$.

**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 $P_{\gamma} = E[P_11[q_t \leq \gamma]]$, and $\{P_1\}_{t=1}^T$ a matrix of random variables, we let $P = P_{\gamma_{\max}} = E[P_1]$, and $P_{\gamma} = P - P_{\gamma}$. For example, since $M_{1\gamma} = E[ztz_t'|1(q_t \leq \gamma)]$, we have $M = E[ztz_t|1]$ and $M_{2\gamma} = E[z_tz_t'|1(q_t > \gamma)]$. 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 $\hat{M}_{1\gamma} = T^{-1}\sum_{t=1}^T z_tz_t'$, for $M$ it is $\hat{M} = T^{-1}\sum_{t=1}^T z_tz_t'$, for $H_{1,1\gamma} = E[z_tz_t'|1(q_t \leq \gamma)]$, it is $\hat{H}_{1,1\gamma} = T^{-1}\sum_{t=1}^T z_tz_t'$, where $\hat{\epsilon_t}$ is an estimate of the residual $\epsilon_t$, and so on. When the notation $\hat{P}_{\gamma}$ 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 $H_{1\gamma} = E[(v_tv_t' \otimes z_tz_t')1(q_t \leq \gamma)]$, and covariance function $E[\mathcal{G}_1(\gamma)\mathcal{G}_1'(\gamma_2)] = E[(v_tv_t' \otimes z_tz_t')1(q_t \leq (\gamma_1 \wedge \gamma_2))]$. We denote $\mathcal{G} = \mathcal{G}(\gamma_{\max})$, where recall that $[\gamma_{\min}, \gamma_{\max}]$ is the support of $q_t$, and $\mathcal{G}(\gamma) = \mathcal{G} - \mathcal{G}_1(\gamma)$. Let $\sigma^2 = E(\epsilon_t + u_t^0 \theta_0^0)^2$, $\vartheta^2 = \text{vec}(1, \theta_0^0)^2$ and $\vartheta^3 = \text{vec}(0, \theta_0^0)^2$.

Throughout the text, quantities of the form $v_tv_t' \otimes z_tz_t'$ should be read as $(v_tv_t') \otimes (z_tz_t')$. Define $\hat{\epsilon_t} = \epsilon_t + (x_t - \tilde{x}_t)^T\theta_0^0$, where $\hat{x}_t$ is obtained either with a LFS or a TFS specification, depending on the context. Define $\tilde{C}_{1\gamma} = T^{-1}\sum_{t=1}^T \tilde{v}_t\tilde{v}_t'$, and $\tilde{C} = \tilde{C}_{1\gamma} + \tilde{C}_{2\gamma}$, both for a LFS and a TFS.

**LINEAR FIRST STAGE**

**Notation specific for a LFS:** Let $A_0 = [\Pi_0, S^T]^T$ be the augmented matrix of the FS slope parameters, where $S = [I_{p_2}, \Pi_0^T(q-p_2)]$ and let $\hat{A} = [\Pi, S^T]^T$. Hence, $z_t = Szt$ and $u_t = A_0zt + (u_t, 0_{1\times q_2})^T$. Define $C_{1\gamma} = A_0M_1A_0^T$, $R_{1\gamma} = M_1M^{-1}$, $C_{1\gamma} = [C_{1\gamma}, C_{2\gamma}]$, $D_{1\gamma} = [\vartheta^T \otimes I_q] - [\vartheta^T \otimes R_{1\gamma}]$, $D_{2\gamma} = [\vartheta^T \otimes I_q] - [\vartheta^T \otimes R_{2\gamma}]$, $\hat{R}_{1\gamma} = \hat{M}_1M^{-1}$, $D = D_{1\gamma} + D_{2\gamma} = [1, 0_{1\times p_2}] \otimes I_q$ and $Q_{\gamma} = C_{1\gamma}C_{1\gamma}^{-1}C_{2\gamma}$. Also, define the Gaussian
proceses $B_i(\gamma) = A^0 D_i \gamma, G_i(\gamma)$ for $i = 1, 2, B = B_1(\gamma) + B_2(\gamma) = A^0 D G$, as well as the processes $B(\gamma) = \text{vec}(B_1(\gamma), B_2(\gamma))$ and $G(\gamma) = C_1 B(\gamma)$, whose covariance matrices $V_{B,\gamma}$ and respectively $V_\gamma$ are defined below, along with $V_\gamma$ used in the construction of the sup$_{\gamma \in T} W_T(\gamma)$ test.

**Definition B.1.** Denote $V_{B,1,\gamma} = A^0 D_{1,\gamma} H_{1,\gamma} D_{1,\gamma}^T, A^0, V_B = A^0 D_H D^T, A^0, V_{B,12,\gamma} = A^0 D_{1,\gamma} H_{1,\gamma} D^T A^0, V_{B,2,\gamma} = V_{B_1,1,\gamma} - V_{B,12,\gamma} - V_{B,2,\gamma}^T$. Then $V_{B,\gamma} = \begin{bmatrix} V_{B,1,\gamma} & V_{B,12,\gamma} \\ V_{B,12,\gamma}^T & V_{B,2,\gamma} \end{bmatrix}$, and $V_\gamma = C_\gamma V_{B,\gamma} C_\gamma^T$. Define $D_{i,\gamma} = \begin{bmatrix} \hat{\theta}^T & 0 \\ 0 & \hat{\theta} \end{bmatrix}$, where $\hat{\theta} = [1, \theta_2]$ and $\theta = [0, \theta_2]$. Then $V_\gamma$ is defined as $V_\gamma$, but replacing $C_{i,\gamma}$ with $\hat{C}_{i,\gamma}$, $A^0$ with $\hat{A}$, $D_{i,\gamma}$ by $D_{i,\gamma}$, and $H_{1,\gamma}$ by $\hat{H}_{1,\gamma} = T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \otimes z_i z_i^T$, and $H_{ \hat{1}, \gamma} = T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \otimes z_i z_i^T$.

**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 $H_0$, and for $i = 1, 2$, (i) $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow B(\gamma)$, and $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow C(\gamma)$; (ii) $T^{-1/2} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow B(\gamma), T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow B(\gamma)$, and $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow C(\gamma)$.

**Proof of Lemma B.3.** Part (i). $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T = T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow B(\gamma)$, and $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow C(\gamma)$. By Assumption 4.1 and standard arguments, we have that $\hat{A} = A^0 + o_p(1)$, so $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T = (A^0 + o_p(1))T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T (A^0 + o_p(1))^T$. By Assumption 4.1(b),(c) and (e), the assumptions of Lemma B.1 are satisfied for elements of $x_t = z_t^T, T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow B(\gamma)$. By Lemma B.1(i), $M_{1,\gamma} \Rightarrow M_{1,\gamma}$, and $M_{2,\gamma} \Rightarrow M_{2,\gamma}$, and $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow B(\gamma)$. Because $A^0, A^0, T^{-1/2} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow A^0 D_{1,\gamma} G_1 \gamma) = B(\gamma)$. Because $T^{-1} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T = A T^{-1/2} \sum_{i,\gamma} \theta_i \hat{v}_i \hat{v}_i^T \Rightarrow A D (T^{-1/2} \sum_{i,\gamma} \hat{v}_i \hat{v}_i^T \Rightarrow A D G = B$.

**Theorem B.1.** (Asymptotic Distribution LFS) Let $y_t$ be generated by (1) and $x_t$ be generated by the LFS (2). Then, under $H_0$ and Assumption 4.1, (i) $T^{1/2}(\hat{\theta} - \theta_2) \Rightarrow \mathcal{E}(\gamma)$, (ii) sup$_{\gamma \in T} LR_T(\gamma) \Rightarrow \sup_{\gamma \in T} \left[ E^T(\gamma) Q_\gamma E(\gamma) / \sigma^2 \right]$, (iii) sup$_{\gamma \in T} W_T(\gamma) \Rightarrow \sup_{\gamma \in T} \left[ E^T(\gamma) V_{\gamma}^{-1} E(\gamma) \right]$, where $V_\gamma$ is in Definition B.1.

**Proof of Theorem B.1.** Part (i). By Lemma B.3, we have that $T^{-1/2}(\hat{\theta}_1 - \theta_2) = \ldots$
\[ \hat{C}_{\gamma}^{-1}(T^{-1/2} \sum_{i\gamma} \hat{\epsilon}_t \hat{\epsilon}_t^T) \Rightarrow C_{\gamma}^{-1} \mathcal{B}_1(\gamma), \] so

\[ T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2) \Rightarrow [C_{\gamma}^{-1} - C_{\gamma}^{-1}] \text{vec}(\mathcal{B}_1, \mathcal{B}_2) = C_{\gamma} \mathcal{B}(\gamma) = \mathcal{E}(\gamma). \]

**Part (ii).** Since \( \hat{\theta} = \hat{C}^{-1}(T^{-1} \sum_{t=1}^T \hat{w}_t y_t), \hat{C} \hat{\theta} = \sum_{i=1}^T \hat{C}_{i\gamma} \hat{\theta}_{i\gamma} \), so \( \hat{\theta} = \sum_{i=1}^T C_{i\gamma} \hat{\theta}_{i\gamma} + o_p(1) \), and therefore \( \hat{\theta}_1 - \hat{\theta}_2 = C_{\gamma}^{-1} C_{\gamma} \hat{\theta}_2 - \hat{\theta}_2 + o_p(1) \) and \( \hat{\theta}_2 - \hat{\theta}_2 = C_{\gamma}^{-1} C_{\gamma} (\hat{\theta}_2 - \hat{\theta}_2) + o_p(1) \). Hence,

\[ SSR_0 - SSR_1(\gamma) = \sum_{i=1}^2 \left( \sum_{i\gamma} (y_t - \hat{w}_t^T \hat{\theta})^2 - (y_t - \hat{w}_t^T \hat{\theta}_2)^2 \right) \]

\[ = \sum_{i=1}^2 \sum_{i\gamma} \left( \hat{\theta}_i - \hat{\theta}_2 \right)^T \left( 2 \sum_{i\gamma} \hat{w}_t \hat{w}_t^T (\hat{\theta}_i - \hat{\theta}_2) \right) \]

\[ = \sum_{i=1}^2 \sum_{i\gamma} \left( T^{1/2}(\hat{\theta}_i - \hat{\theta}_2) \right)^T \left( T^{-1} \sum_{i\gamma} \hat{w}_t \hat{w}_t^T \right) \left( T^{-1/2}(\hat{\theta}_i - \hat{\theta}_2) \right)^T \]

\[ = T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2)(C_{\gamma}^{-1} C_{\gamma} C_{\gamma}^{-1} C_{\gamma} + C_{\gamma}^{-1} C_{\gamma} C_{\gamma}^{-1} C_{\gamma}) T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2) + o_p(1) \]

\[ = T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2)^T Q_{\gamma} T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2) + o_p(1), \]

where the last line follows because \( C = \sum_{i=1}^2 C_{i\gamma}, \) therefore \( C_{\gamma}^{-1} C_{\gamma} = I_p - C_{\gamma}^{-1} C_{\gamma}, \)

\[ C_{\gamma}^{-1} C_{\gamma} = (C - C_{\gamma}^{-1} C_{\gamma}) = C - C_{\gamma}^{-1} C_{\gamma} - C_{\gamma}^{-1} C_{\gamma}. \]

And so \( C_{\gamma}^{-1} C_{\gamma} = C_{\gamma}^{-1} C_{\gamma} - C_{\gamma}^{-1} C_{\gamma} - C_{\gamma}^{-1} C_{\gamma} \). Since \( T^{1/2}(\hat{\theta}_1 - \hat{\theta}_2) \Rightarrow \mathcal{E}(\gamma), SSR_0 - SSR_1(\gamma) \Rightarrow \mathcal{E}(\gamma) Q_{\gamma} \mathcal{E}(\gamma). \]

Now \( SSR_1(\gamma)(T - 2p) = T^{-1} SSR_1(\gamma) + o_p(1), \) since show below \( T^{-1} SSR_1(\gamma) \overset{p}{\Rightarrow} \sigma^2: \)

\[ T^{-1} SSR_1(\gamma) = \sum_{i=1}^2 \sum_{i\gamma} (y_t - \hat{w}_t^T \hat{\theta}_i)^2 \]

\[ = \sum_{i=1}^2 \sum_{i\gamma} T^{-1} \sum_{i\gamma} (\hat{\theta}_i - \hat{\theta}_2)^T \hat{w}_t \hat{w}_t^T (\hat{\theta}_i - \hat{\theta}_2)^T \]

\[ = T^{-1} \sum_{t=1}^T \sum_{i\gamma} \hat{\epsilon}_t \hat{\epsilon}_t^T (\hat{\theta}_i - \hat{\theta}_2)^T \]

By Lemma B.3, \( T^{-1} \sum_{i\gamma} \hat{\epsilon}_t \hat{\epsilon}_t^T \overset{p}{\Rightarrow} C_{\gamma}, \) and \( T^{-1} \sum_{i\gamma} \hat{\epsilon}_t \hat{\epsilon}_t^T = o_p(1). \) Since \( \hat{\theta}_i - \hat{\theta}_2 = o_p(1), \)

\[ T^{-1} SSR_1(\gamma) = T^{-1} \sum_{t=1}^T (\hat{\epsilon}_t + u_t^T \theta_{0,2}^T z_t^T (\Pi - \Pi^0)^2 + o_p(1) \]

\[ = T^{-1} \sum_{t=1}^T (\hat{\epsilon}_t + u_t^T \theta_{0,2}^T)^2 + (\Pi - \Pi^0 T^{-1} \sum_{t=1}^T z_t z_t^T (\Pi - \Pi^0) \theta_{0,2}^T \]

\[ - 2T^{-1} \sum_{t=1}^T z_t (\hat{\epsilon}_t + u_t^T \theta_{0,2}^T) (\Pi - \Pi^0) \theta_{0,2}^T + o_p(1) \]

\[ = T^{-1} \sum_{t=1}^T (\epsilon_t + u_t^T \theta_{0,2}^T)^2 + o_p(1), \]

where the last equality used Lemma B.3(ii) and \( \Pi - \Pi^0 = o_p(1). \) We now apply Lemma B.1 to \( a_t = (\epsilon_t + u_t^T \theta_{0,2}^T)^2. \) First, \( E(\epsilon_t + u_t^T \theta_{0,2}^T)^2 = \sigma^2. \) Second, by Assumption 4.1(b), \( a_t \) is strictly stationary with -mixing coefficients satisfying condition (i) in Lemma B.1. Third, by Minkowski’s inequality, \( \|\epsilon_t + u_t^T \theta_{0,2}^T\|_2 \leq \|\epsilon_t\|_2 + \|u_t^T \theta_{0,2}^T\|_2 + 2\|\epsilon_t u_t^T \theta_{0,2}^T\|_2. \)

Note that \( \|\epsilon_t\|_2 = (E|\epsilon_t^4|)^{1/2} < K \) by Assumption 4.1(c). By the same assumption and Hölder and Minkowski inequalities, letting \( \theta_{x,j}^0 \) be the \( j \)-th element of \( \theta_{x}^0, \)

\[ \|\theta_{x,j}^0 u_t u_t^T \theta_{x,j}^0\|_2 \leq \sum_{j,k=1}^{p_1} \theta_{x,j}^0 \theta_{x,j}^0 u_t u_t^T u_t u_t, k \|_2 \leq \max_{j,k} \|\theta_{x,j}^0 \theta_{x,j}^0 u_t u_t^T \theta_{x,j}^0 \|_2 < K, \]
and similarly, \( \|e_t^0 \theta_t^0 \|^2 \leq K \). Therefore, by Lemma B.1, \( T^{-1} \sum_{t=1}^T (e_t + w_t^0 \theta_t^0)^2 \overset{p}{\to} \sigma^2 \), completing the proof of part (ii).

**Part (iii).** We are left to show \( \hat{V}_T \overset{p}{\to} V_\gamma \). Since \( \hat{C}_r \overset{p}{\to} C_r \), \( \hat{A}_r \overset{p}{\to} A_r \), \( \hat{\theta}_0 \overset{p}{\to} \theta_0 \), and therefore \( \hat{D}_\gamma \overset{p}{\to} D_\gamma \), to show that \( \hat{V}_T \overset{p}{\to} V_\gamma \), it suffices to show that \( \hat{H}_T = \text{Var}(T^{-1/2} \sum_{i=1}^T \hat{v}_i \otimes z_i) \overset{p}{\to} H_\gamma \) for \( i = 1, 2 \). We show \( \hat{H}_{a,1,1,1} = T^{-1} \sum_{i=1}^I z_i z_i^T e_i^2 \overset{p}{\to} H_{a,1,1,1} = E(z_i z_i^T e_i^2 I[q_i \leq \gamma]); \) the rest of the proof is similar.

\[
\hat{H}_{a,1,1} = T^{-1} \sum_{i=1}^I z_i z_i^T e_i^2 = T^{-1} \sum_{i=1}^I z_i z_i^T (e_i + w_i^0 (\theta_0 - \hat{\theta}))^2 = T^{-1} \sum_{i=1}^I z_i z_i^T e_i^2 + T^{-1} \sum_{i=1}^I z_i z_i^T [(\theta_0 - \hat{\theta})^T w_i w_i^0 (\theta_0 - \hat{\theta})] + 2T^{-1} \sum_{i=1}^I z_i z_i^T [(\theta_0 - \hat{\theta})^T w_i e_i].
\]

First, by Assumption 4.1(b),(c) and (e), and Lemma B.1, \( T^{-1} \sum_{i=1}^I z_i z_i^T e_i^2 \overset{p}{\to} H_{a,1,1} \). Second, \( T^{-1} \sum_{i=1}^I z_i z_i^T [(\theta_0 - \hat{\theta})^T w_i w_i^0 (\theta_0 - \hat{\theta})] = [I_q \otimes (\theta_0 - \hat{\theta})^T] T^{-1} \sum_{i=1}^I z_i z_i^T \otimes w_i w_i^0 [I_q \otimes (\theta_0 - \hat{\theta})] \). We know that \( \theta_0 - \hat{\theta} = \text{O}_p(T^{-1/2}) \). Also, it can be shown that for typical element \( \xi_i = z_i z_i^T w_i w_i^0 \), with \( i, j = 1, \ldots, q \) and \( k, l = 1, \ldots, p \), we have: \( \|\xi_i\|_r \leq K \max\{\|z_i\|_r \|z_i\|_r \} < \infty \), by Assumption 4.1(c), where \( r > 1 \). Therefore, by Assumption 4.1(b)-(c) and Lemma B.1, \( T^{-1} \sum_{i=1}^I z_i z_i^T \otimes w_i w_i^0 = \text{O}_p(1) \) uniformly in \( \gamma \). It follows that \( T^{-1} \sum_{i=1}^I z_i z_i^T [(\theta_0 - \hat{\theta})^T w_i w_i^0 (\theta_0 - \hat{\theta})] = \text{O}_p(1) \). Similarly, \( T^{-1} \sum_{i=1}^I z_i z_i^T [\theta_0 - \hat{\theta}]^T w_i e_i = \text{O}_p(1) \), therefore \( \hat{H}_{a,1,1} \overset{p}{\to} H_{a,1,1,1} \), so by similar arguments, \( \hat{H}_{a,2,1} \overset{p}{\to} H_{a,2,1} \), completing the proof of part (ii).

**BOOTSTRAP VALIDITY LFS**

**Lemmas 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 \( \mathbb{H}_b \), (i) \( T^{-1} \sum_{t=1}^T v_t \eta_t \otimes z_t = o_p(1) \); (ii) \( T^{-1} \sum_{t=1}^T z_t z_t^T \eta_t = o_p(1) \); (iii) \( T^{-1/2} \sum_{t=1}^T z_t \eta_t \otimes z_t = \text{G} \); (iv) \( T^{-1/2} \sum_{t=1}^T z_t \eta_t = \text{G} \); (v) \( T^{1/2}(\hat{\Pi} - \Pi) = T^{1/2}(\hat{\Pi} - \Pi) + o_p(1) \); (vi) \( T^{-1} \sum_{t=1}^T \hat{w}_t \hat{w}_t^T = C_\gamma \); (vii) \( T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{w}_t^T \overset{p}{\to} B(\gamma) \) and \( \text{vec}(T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{w}_t^T, T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{w}_t^T) \overset{p}{\to} B(\gamma) \), where \( \hat{w}_t = \hat{w}_t - \text{w}_t \theta_t \).

**Proof of Lemma B.4.** We 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), \( P(b(\|T^{-1} \sum_{t=1}^T v_t \eta_t \otimes z_t \| > c) \leq c^{-1} E[|T^{-1} \sum_{t=1}^T v_t \eta_t \otimes z_t|] = c^{-1} E[|T^{-1} \sum_{t=1}^T v_t \eta_t \otimes z_t|] \leq c^{-1} K \|T^{-1} \sum_{t=1}^T v_t \eta_t z_t \| = o_p(1) \), where the last equality follows from Lemma B.3(ii). Therefore, \( T^{-1} \sum_{t=1}^T v_t \eta_t \otimes z_t = o_p(1) \).

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

**Part (iii).** To show \( T^{-1/2} \sum_{t=1}^T v_t \eta_t \otimes z_t \overset{d}{\to} \text{G} \), we apply Lemma B.2 (FCLT) and verify that \( \text{Var}(T^{-1/2} \sum_{t=1}^T v_t \eta_t \otimes z_t) \overset{p}{\to} H_{1,1} \). 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 $E_b[|\gamma T_1 \delta_1|] \leq (E_b[|\gamma T_1|]^{1/2}) \leq O_p(1)$ by Assumptions 4.1(c) and 4.2. Finally, $Var_b(T^{-1/2} \sum_{\gamma} v_t \eta_t \otimes z_t) = T^{-1} \sum_{\gamma} v_t \eta_t \otimes z_t \mathfrak{d} \to H(\gamma)$ where the last statement follows by applying Lemma B.1(U LLN).

Part (iv). Since $\hat{u}_t = x_j^T - \hat{x}_j^T = u_t - (\bar{H} - \hat{H})^T z_t$,

$$T^{-1/2} \sum_{\gamma} u_t^T \otimes z_t = T^{-1/2} \sum_{\gamma} \hat{u}_t \eta_t \otimes z_t$$

$$= T^{-1/2} \sum_{\gamma} u_t \eta_t \otimes z_t - (T^{1/2} (\bar{H} - \hat{H})^T \otimes I_q) \left( T^{-1} \sum_{\gamma} z_t \eta_t \otimes z_t \right)$$

$$= T^{-1/2} \sum_{\gamma} u_t \eta_t \otimes z_t + O_p(1) \times \text{vec} \left( T^{-1} \sum_{\gamma} z_t \eta_t \otimes z_t \right)$$

$$= T^{-1/2} \sum_{\gamma} u_t \eta_t \otimes z_t + o_p(1).$$

Similarly, $T^{-1/2} \sum_{\gamma} u_t \eta_t \otimes z_t + o_p(1)$, therefore $T^{-1/2} \sum_{\gamma} u_t \eta_t \otimes z_t \Rightarrow G_1(\gamma)$.

Part (v). Since $u_t = (\epsilon_t, u_t^T)^T$, from part (iv) and Lemma B.3(ii), $T^{1/2}(\hat{H} - \bar{H}) = T^{1/2} \left( \sum_{t=1}^T \epsilon_t^2 u_t \right) = T^{1/2} \left( \sum_{t=1}^T \epsilon_t^2 u_t + o_p(1) \right) = T^{1/2}(\bar{H} - \hat{H}) + o_p(1)$.

Part (vi). $T^{-1} \sum_{\gamma} \hat{u}_t^b \gamma \hat{w}_t^b = \hat{A}_b \left( T^{-1} \sum_{\gamma} z_t \gamma \hat{e}_t^b \right) \hat{A}_b \gamma^T$, from part (v), $T^{1/2}(\bar{H} - \hat{H}) = o_p(1)$, so $\hat{A}_b \gamma \hat{w}_t^b \to \hat{A}_b$, therefore $\hat{A}_b \gamma \hat{w}_t^b \to \hat{A}_b$, and $T^{-1} \sum_{\gamma} \hat{u}_t^b \gamma \hat{w}_t^b \to C_1(\gamma)$.

Part (vii). By part (iv) and (v) and Lemma B.3(i)-(ii), and recalling that $D_1(\gamma) = \gamma^T G_1(\gamma) - \gamma^T \hat{\eta}^T R_1 - \gamma^T \hat{D}_1$, we have:

$$T^{-1/2} \sum_{\gamma} \hat{w}_t^b \gamma \hat{e}_t^b = \hat{A}_b \left( T^{-1/2} \sum_{\gamma} z_t \gamma \hat{e}_t^b \right) = T^{1/2}(\hat{H} - \hat{H}) \hat{\theta}_x$$

$$= \hat{A}_b \left( T^{-1/2} \sum_{\gamma} z_t \gamma \hat{e}_t^b \right) + o_p(1) = \hat{A}_b \left( T^{-1/2} \sum_{\gamma} z_t \gamma \hat{e}_t^b \right) + o_p(1) \Rightarrow B_1(\gamma).$$

Similarly, $T^{-1/2} \sum_{\gamma} \hat{u}_t^b \gamma \hat{e}_t^b \Rightarrow B_2(\gamma)$, so vec($T^{-1/2} \sum_{\gamma} \hat{u}_t^b \gamma \hat{e}_t^b$) $\Rightarrow B(\gamma)$.

Proof of Theorem 4.2. Part (i).

$$SSR_b^1 - SSR_b^1(\gamma) = \sum_{i=1}^2 \left[ \sum_{\gamma} \epsilon_i^T \left( y_i^b - \hat{u}_t^b \gamma \hat{\theta}_x \right) - \left( y_i^b \gamma \hat{\theta}_x \right) \right]^2$$

$$= \sum_{i=1}^2 \left( \hat{\theta}_x^\gamma - \hat{\theta}_x^b \right)^T \left[ 2 \sum_{i=1}^2 \hat{u}_t^b \gamma \hat{e}_t^b - \sum_{i=1}^2 \hat{u}_t^b \gamma \hat{e}_t^b \left( \hat{\theta}_x^b - \hat{\theta}_x^b \right) - \sum_{i=1}^2 \hat{u}_t^b \gamma \hat{e}_t^b \left( \hat{\theta}_x^b - \hat{\theta}_x^b \right) \right]$$

$$= \sum_{i=1}^2 T^{1/2} \left( \hat{\theta}_x^b \gamma - \hat{\theta}_x^b \right)^T \left( \hat{\theta}_x^b - \hat{\theta}_x^b \right) + o_p(1)$$

$$= \sum_{i=1}^2 T^{1/2} \left( \hat{\theta}_x^b \gamma - \hat{\theta}_x^b \right)^T C_{\gamma} \sum_{i=1}^2 T^{1/2} \left( \hat{\theta}_x^b \gamma - \hat{\theta}_x^b \right) + o_p(1)$$

$$= T^{1/2} \left( \hat{\theta}_x^b \gamma - \hat{\theta}_x^b \right)^T Q_x T^{1/2} \left( \hat{\theta}_x^b \gamma - \hat{\theta}_x^b \right) + o_p(1),$$

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 4.2(i). By Lemma
B.4(vi) and (vii),
\[
\begin{align*}
T^{3/2}(\hat{\theta}_{\gamma}^b - \hat{\theta}_{\gamma}^h) &= T^{3/2}(\hat{\theta}_{\gamma}^b - \hat{\theta}) - T^{3/2}(\hat{\theta}_{\gamma}^h - \hat{\theta}) \\
&= (T^{-1} \sum_{t=1}^T \tilde{w}_t^b \tilde{w}_t^b)^{1/2} T^{-1/2} \sum_{t=1}^T \theta_t^b + (T^{-1} \sum_{t=1}^T \tilde{w}_t^h \tilde{w}_t^h)^{1/2} T^{-1/2} \sum_{t=2}^T \theta_t^h \theta_t^h \\
&= C_{\gamma}^{-1} B_1(\gamma) - C_{\gamma}^{-1} B_2(\gamma) = \mathcal{E}(\gamma).
\end{align*}
\]

Using (12) into (11), we have: 
\[
SSR_0^b - SSR_1^b(\gamma) \rightarrow \mathcal{E}^T(\gamma) Q_\gamma \mathcal{E}(\gamma).
\]

It remains to show that 
\[
SSR_0^b (T - 2p) \rightarrow \sigma^2,
\]
or, equivalently, that 
\[
T^{-1} SSR_1^b(\gamma) \rightarrow \sigma^2.
\]

By Lemma B.4(vii), 
\[
T^{-1} \sum_{t=1}^T \gamma_t \tilde{w}_t^b \tilde{w}_t^b \rightarrow C_{\gamma},
\]

and 
\[
T^{-1} \sum_{t=1}^T \gamma_t \tilde{w}_t^b \tilde{w}_t^T = O_p(T^{-1/2}).
\]

From above,
\[
\hat{\theta}_{\gamma}^b - \hat{\theta} = O_p^b(T^{-1/2}) = o_p^b(1).
\]

Therefore, 
\[
T^{-1} SSR_1^b(\gamma) = T^{-1} \sum_{t=1}^T \gamma_t \tilde{w}_t^b \tilde{w}_t^T + o_p(1) = T^{-1} \sum_{t=1}^T (\gamma_t \tilde{w}_t^b \tilde{w}_t^b) + o_p(1) = T^{-1} \sum_{t=1}^T (\gamma_t \tilde{w}_t^b \tilde{w}_t^b),
\]

which equals Lemma B.4(vi), which implies 
\[
\hat{\Pi}^b - \Pi = o_p(1).
\]

We now show that 
\[
T^{-1} \sum_{t=1}^T (\gamma_t \tilde{w}_t^b \tilde{w}_t^b) \rightarrow \sigma^2
\]

which then would complete the proof of part (i).

Since 
\[
T^{-1} \sum_{t=1}^T (\gamma_t \tilde{w}_t^b \tilde{w}_t^b) = \hat{\theta}_{\gamma}^b T^{-1} \sum_{t=1}^T (\tilde{v}_t \tilde{v}_t^T) \tilde{v}_t^b\tilde{v}_t^T
\]

we analyze 
\[
T^{-1} \sum_{t=1}^T (\tilde{v}_t \tilde{v}_t^T) \tilde{v}_t^b \tilde{v}_t^T
\]

First consider 
\[
T^{-1} \sum_{t=1}^T \gamma_t^2 \eta_t^2,
\]

the first element of this matrix. We have:
\[
\begin{align*}
T^{-1} \sum_{t=1}^T \gamma_t^2 \eta_t^2 &= T^{-1} \sum_{t=1}^T (\gamma_t \tilde{w}_t^b (\hat{\theta} - \theta))^2 \eta_t^2 \\
&= T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2 \\
&= (\hat{\theta} - \theta)^T A^T(\tilde{v}_t \tilde{v}_t^T) A^T(\tilde{v}_t \tilde{v}_t^T) (\hat{\theta} - \theta) \\
&= T^{-1} \sum_{t=1}^T (c_t^2 \eta_t^2) + o_p(1) T^{-1} \sum_{t=1}^T (z_t \tilde{v}_t \tilde{v}_t^T \eta_t^2) A^T(\tilde{v}_t \tilde{v}_t^T) (\hat{\theta} - \theta) \\
&= T^{-1} \sum_{t=1}^T (c_t^2 \eta_t^2) + o_p(1) T^{-1} \sum_{t=1}^T (z_t \tilde{v}_t \tilde{v}_t^T \eta_t^2) \eta_t^2 \eta_t^2 = o_p(1) - T^{-1} \sum_{t=1}^T (c_t^2 \eta_t^2) \eta_t^2 = o_p(1).
\end{align*}
\]

First, we show that 
\[
T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2 \rightarrow E(c_t^2).\]

Note that 
\[
E^b(T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2) = T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2.
\]

Also, let 
\[
K^* = E^b(c_t^2),\]

\[
E^b(T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2) = T^{-2} \sum_{t,s=1,t \neq s} c_t^2 K^* + T^{-2} \sum_{t,s=1,t \neq s} c_t^2 e_t^2 c_s^2.
\]

By Assumption 4.1(b)-(c), 
\[
e_t^4 \rightarrow E(ne_t^4),\]

and therefore 
\[
T^{-2} \sum_{t,s=1,t \neq s} c_t^2 e_t^2 = (T^{-1} \sum_{t=1} c_t^2)^2 - T^{-2} \sum_{t,s=1,t \neq s} c_t^2 \rightarrow 0.
\]

Hence, 
\[
E^b(T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2) = 0,
\]

and 
\[
E^b(T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2) = o_p(1).
\]

Therefore, 
\[
T^{-1} \sum_{t=1}^T c_t^2 \eta_t^2 \rightarrow E(c_t^2).
\]
Second, we show that $T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2} = O_{b}(1)$. We have $E^{b}(z_{t} z_{t}^{\top} \eta_{t}^{2}) = z_{t} z_{t}^{\top}$. Also,

$$E^{b}((\text{vec}(T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2})) \text{vec}(T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2})) = T^{-2} \sum_{t=1}^{T} (z_{t} z_{t}^{\top} \otimes z_{t} z_{t}^{\top}) K + (T^{-1} \sum_{t=1}^{T} z_{t} \otimes z_{t})(T^{-1} \sum_{t=1}^{T} z_{t} \otimes z_{t})$$

$$- T^{-2} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \otimes z_{t} z_{t}^{\top},$$

By Assumption 4.1(b)-(c) and Lemma B.1, $T^{-2} \sum_{t=1}^{T} (z_{t} z_{t}^{\top} \otimes z_{t} z_{t}^{\top}) = O_{b}(1)$. Also, by Lemma B.3(a), $T^{-1} \sum_{t=1}^{T} z_{t} \otimes z_{t} = \text{vec} M \overset{p}{\rightarrow} \text{vec} M$. So, $E^{b} \text{vec}(T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2}) = z_{t} \otimes z_{t}$, and $\text{Var}^{b}(\text{vec}(T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2} - z_{t} z_{t}^{\top})) = 0$. Therefore, by Markov's inequality,

$$T^{-2} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2} = O_{b}(1),$$

hence $T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2} = T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} + o_{p}(1) = M + o_{p}(1) = O_{b}(1)$.

Third, we show that $\xi_{t} = T^{-1} \sum_{t=1}^{T} \epsilon_{t} z_{t} \eta_{t}^{2} = O_{b}(1)$. 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, $E^{b}(\xi_{t}) \overset{p}{\rightarrow} 0$, so, by Markov's inequality, $\xi_{t} \overset{p}{\rightarrow} 0$. Substituting these results into (14), it follows that $T^{-1} \sum_{t=1}^{T} \epsilon_{t} \eta_{t}^{2} \overset{p}{\rightarrow} E(\epsilon_{t}^{2})$.

Next,

$$T^{-1} \sum_{t=1}^{T} \hat{u}_{t} \hat{u}_{t}^{\top} \eta_{t}^{2} = T^{-1} \sum_{t=1}^{T} (u_{t} + z_{t}^{\top} (\hat{U} - \Pi^{0}))(u_{t} + z_{t}^{\top} (\hat{U} - \Pi^{0}))^{\top} \eta_{t}^{2}$$

$$= T^{-1} \sum_{t=1}^{T} u_{t} u_{t}^{\top} \eta_{t}^{2} + (\hat{U} - \Pi^{0})^{\top} T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \eta_{t}^{2} (\hat{U} - \Pi^{0}) + [T^{-1} \sum_{t=1}^{T} u_{t} z_{t} \eta_{t}^{2} (\hat{U} - \Pi^{0})]$$

$$+ [T^{-1} \sum_{t=1}^{T} u_{t} z_{t} \eta_{t}^{2} (\hat{U} - \Pi^{0})]^{\top} \overset{p}{\rightarrow} E(u_{t} u_{t}^{\top}),$$

by similar arguments as for $T^{-1} \sum_{t=1}^{T} \epsilon_{t} \eta_{t}^{2} \overset{p}{\rightarrow} E(\epsilon_{t}^{2})$. Similarly, $T^{-1} \sum_{t=1}^{T} \hat{v}_{t} \hat{v}_{t}^{\top} \eta_{t}^{2} \overset{p}{\rightarrow} E(\hat{v}_{t} \hat{v}_{t}^{\top})$. Therefore,

$$T^{-1} \sum_{t=1}^{T} \{u_{t}^{\top} + u_{t}^{\top} \theta_{t}^{0}\} = \theta_{t}^{0} T^{-1} \sum_{t=1}^{T} \hat{v}_{t} \hat{v}_{t}^{\top} \eta_{t}^{2} \theta_{t}^{0} = \theta_{t}^{0} T^{-1} \sum_{t=1}^{T} \hat{v}_{t} \hat{v}_{t}^{\top} \eta_{t}^{2} + o_{p}(1) = \sigma_{t}^{2} + o_{p}(1),$$

completing the proof of part (i).

Part (ii). Let $D_{1}^{b} = [\hat{\theta}_{1}^{b} \otimes I_{\gamma}] - [\hat{\theta}_{1}^{b} \otimes \hat{R}_{1\gamma}]$, where $\hat{\theta}_{1}^{b} = [1, \hat{\theta}_{2}^{b}]$ and $\hat{R}_{1\gamma} = [\hat{R}_{1\gamma}]$. From (12), $T^{1/2}(\hat{\theta}_{1\gamma}^{b} - \hat{\theta}_{2\gamma}^{b}) \overset{p}{\rightarrow} E(\gamma)$, so it remains to show that $\hat{V}_{1\gamma}^{b} \overset{p}{\rightarrow} V_{1\gamma}$. We will just show that $\hat{V}_{1\gamma}^{b} = \hat{A}^{0} \hat{D}_{1\gamma}^{b} \hat{H}_{1\gamma}^{b} \hat{D}_{1\gamma}^{b} \hat{A}^{0 \top} \hat{D}_{1\gamma}^{b} \hat{H}_{1\gamma}^{b} \hat{D}_{1\gamma}^{b} \hat{A}^{0 \top} = V_{1\gamma}$, where $\hat{V}_{1\gamma}^{b} = T^{-1} \sum_{t=1}^{T} \hat{v}_{t} \hat{v}_{t}^{\top} \otimes z_{t} z_{t}^{\top}$; the rest follows by similar arguments. Since $\hat{A}^{b} \overset{p}{\rightarrow} A$, $\hat{D}_{1\gamma}^{b} \overset{p}{\rightarrow} D_{1\gamma}$, and the proof for $\hat{H}_{1\gamma}^{b} \overset{p}{\rightarrow} H_{1\gamma}$, is similar to $\hat{H}_{1\gamma}^{b} \overset{p}{\rightarrow} H_{1\gamma}$, where $\hat{H}_{1\gamma}^{b} = T^{-1} \sum_{t=1}^{T} [(e_{t}^{2})] z_{t} z_{t}^{\top}$, we only show $\hat{H}_{1\gamma}^{b} \overset{p}{\rightarrow} H_{1\gamma}$.

$$\hat{H}_{1\gamma}^{b} = T^{-1} \sum_{t=1}^{T} \{z_{t} z_{t}^{\top} (\hat{b}_{t} - w_{t}^{T} \hat{b}_{t})^{2}\} = T^{-1} \sum_{t=1}^{T} \{z_{t} z_{t}^{\top} [e_{t}^{2} + w_{t}^{T} (\hat{\theta} - \hat{\theta}_{t}^{b})]^{2}\}$$

$$= T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} (e_{t}^{2})^{2} + T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} [w_{t}^{T} (\hat{\theta} - \hat{\theta}_{t}^{b})]^{2} + 2T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} [e_{t}^{T} w_{t}^{T} (\hat{\theta} - \hat{\theta}_{t}^{b})]$$

$$= T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} (e_{t}^{2})^{2} + [I_{\gamma} \otimes ((\hat{\theta} - \hat{\theta}_{t}^{b})^{T} \hat{A})] (T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} \otimes z_{t} z_{t}^{\top}) [I_{\gamma} \otimes ((\hat{A}^{0 \top} (\hat{\theta} - \hat{\theta}_{t}^{b}))^{T})]$$

$$+ 2[I_{\gamma} \otimes ((\hat{\theta} - \hat{\theta}_{t}^{b})^{T} \hat{A})] (T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\top} z_{t} \otimes z_{t}).$$
We already showed that $T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top = O_p(1)$, $\hat{\theta} - \theta_0 = O_p(T^{-1/2})$, and $A = A_0 + o_p(1)$. Therefore, 

$$
\hat{R}^b_{\alpha,\gamma} = T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \eta_t^2 + O_p(T^{-1/2})(T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \otimes z_t) + o_p(1)
$$

$$
\equiv L_1^b + L_2^b + o_p(1).
$$

$$
L_1^b = T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \eta_t^2 = T^{-1} \sum_{t=1}^T z_t z_t^\top (\epsilon_t - w_t ^\top (\hat{\theta} - \theta_0))^2 \eta_t^2 = T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^2 \eta_t^2 + [I_q \otimes ((\hat{\theta} - \theta)^{\top} A^0)] (T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t^2) [I_q \otimes (A_0^{\top} (\hat{\theta} - \theta_0))]
$$

$$
= T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^2 \eta_t^2 + O_p(T^{-1/2}) (T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t^2) O_p(T^{-1/2})
$$

$$
+ O_p(T^{-1/2}) (T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t^2) \equiv \sum_{i=1}^3 L_{1,i}^b.
$$

For any element $\epsilon_t$ of $z_t z_t^\top \otimes z_t z_t^\top$, $E^b[\epsilon_t \eta_t^2] = [\epsilon_t] E^b(\eta_t^2) = [\epsilon]$. So, $T^{-1-\alpha} \sum_{t=1}^T E^b[\epsilon_t \eta_t^2] \leq T^{-\alpha} \sum_{t=1}^T E^b[\epsilon_t^2] \leq o_p(T^{\alpha})$. Therefore, $L_1^b = O_p(T^{-1/2}) (T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t^2) O_p(T^{-1/2}) = o_p(1)$, choosing $\alpha < 1$. Similarly, it can be shown that $L_2^b = o_p(1)$. Also, $E^b(L_{1,3}^b) = T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \eta_t^2$. Recalling that $E^b(\eta^2_t) = K^*$, noting that 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 $a$,

$$
Var^b(vec L_{1,1}^b vec L_{1,1}^b)^{\top} = T^{-2} \sum_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^b \eta_t^2 \otimes K^*
$$

$$
+ T^{-2} \sum_{t=1}^T \sum_{s=1,s \neq t}^T z_t z_t^\top \otimes z_s z_s^\top \epsilon_t^b \epsilon_s^b \eta_t \eta_s 1[\gamma_t \leq \gamma_s] 1[q_s \leq \gamma]
$$

$$
- (T^{-1} \sum_{t=1}^T z_t \otimes z_t \epsilon_t^b) (T^{-1} \sum_{t=1}^T z_t \otimes z_t \epsilon_t^b) = T^{-2} \sum_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^b (K^* - 1).
$$

By Hölder’s inequality applied for $p = r$ and $q = r$, and Assumption 4.2(b),

$$
E\|z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^b \eta_t^2\| \leq (E\|z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^b \eta_t^2\|^{r/(r-1)}(r-1)/r) (E\|\epsilon_t\|^{4r})^{1/4r}
$$

$$
\leq sup(E\|\epsilon_t\|^{4r})^{1/4r} (r-1)/r K = o(T).
$$

Therefore, by Markov’s inequality, $T^{-2} \sum_{t=1}^T z_t z_t^\top \otimes z_t z_t^\top \epsilon_t^b (K^* - 1) = o_p(1)$ From Cheby-

shev’s inequality, it follows that $L_{1,1}^b = T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \eta_t^2 = o_p(1)$. By Assumption 4.1(b),(c)

and (e) and Lemma B.1, $T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \eta_t^2 \leq H_{a,1,\gamma}$, therefore $L_{1,1}^b \overset{p}{\longrightarrow} H_{a,1,\gamma}$, so $L_{1,1}^b \overset{p}{\longrightarrow} H_{a,1,\gamma}$.

We now analyze $L_2^b = O_p(T^{-1/2})(T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \otimes z_t)$.

$$
T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t^b \otimes z_t = T^{-1} \sum_{t=1}^T z_t z_t^\top \epsilon_t \eta_t \otimes z_t = T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t(\epsilon_t - w_t ^\top (\hat{\theta} - \theta_0)) \eta_t
$$

$$
= T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t - (T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t) [I_q \otimes (A_0^{\top} (\hat{\theta} - \theta_0))]
$$

$$
= T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t - (T^{-1} \sum_{t=1}^T z_t z_t^\top \otimes z_t \epsilon_t \eta_t) O_p(T^{-1/2}).
$$
For any element $\xi_t$ of $z_tz_t^\top \otimes z_t e_t$, $E[\rho|T^{-1}\sum_{t=1}^T(1)\sum_{t=1}^T(1)\sup_\theta E^\theta|\eta_t| = T^{-1}O_p(1)O(1) = o_p(1)$, give any $\alpha > 0$. Therefore, $T^{-1}\sum_{t=1}^Tz_tz_t^\top \otimes z_t e_t = o_p(T^{1/2})$. Similarly, $T^{-1}\sum_{t=1}^Tz_tz_t^\top \otimes z_t e_t = o_p(T^{1/2})$. Choose $\alpha < 1/2$; then $L_2^A = O_p(T^{-1/2})o_p(T^{1/2}) = o_p(1)$. Substituting this and $L_1^b = H_{a,1}\gamma$ into (15), it follows that $\hat{R}_{a,1}^b$, $H_{a,1}\gamma$, completing the proof.

**Proof of Theorem 4.1. Part (i).**

Let $Q(\hat{\Pi}_{\hat{\rho}}, \hat{\rho}) \equiv T^{-1}\sum_{t=1}^T u_t^\top d_t = T^{-1}\sum_{t=1}^T v_t^\top (\Pi^0_0 - \hat{\Pi}_{\hat{\rho}})^\top z_t + T^{-1}\sum_{t=1}^T u_t^\top (\Pi^0_0 - \hat{\Pi}_{\hat{\rho}})^\top z_t$.

By standard arguments, $\Pi^0_0 - \hat{\Pi}_{\hat{\rho}} = o_p(1), \Pi^0_0 - \hat{\Pi}_{\hat{\rho}} = o_p(1), (\Pi^0_0 - \hat{\Pi}_{\hat{\rho}}) = o_p(1)$. By Lemma B.3(ii), $T^{-1}\sum_{t=1}^T u_t^\top d_t = o_p(1)$ (uniformly in $\rho$).

Similarly (iii), $T^{-1}\sum_{t=1}^T d_t^\top d_t = O_p(1)$. Because these results hold uniformly over $\rho$, we have $o_p(1) + T^{-1}\sum_{t=1}^T d_t^\top d_t \leq 0$ uniformly over $\rho$, and therefore also at $\hat{\rho}$, so $T^{-1}\sum_{t=1}^T d_t^\top d_t \leq 0$.

**Part (ii).** 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 $\hat{\rho} \leq \rho^0$, because $\hat{\rho} \leq \rho^0$, $\hat{\rho} < \rho^0 - \epsilon$. Because $\hat{\Pi}_{\hat{\rho}}$ is the multivariate LS estimator in the sample $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$, the residuals evaluated over the sub-sample $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$ will also be evaluated at $\hat{\Pi}_{\hat{\rho}} = \hat{\Pi}_2$. However, the true parameter values are $\Pi^0_0$ for $q_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]$, and $\Pi^0_0$ for $q_t \in (\rho^0, \rho^0 + \epsilon)$. Let $A_1 = \sum 1[\theta_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]], B_1 = \sum 1[\theta_t \in (\rho^0, \rho^0 + \epsilon)]$, and $A_{1B} = \sum 1[\theta_t \in [\rho^0 - \epsilon, \rho^0 + \epsilon]]$.

Then:

$$T^{-1}\sum_{t=1}^T d_t^\top d_t = T^{-1}\sum_{t=1}^T (x_t - \hat{\Pi}_2 z_t)(x_t - \hat{\Pi}_2 z_t)$$

$$= (\Pi^0_0 - \hat{\Pi}_2)^\top (T^{-1}\sum_{t=1}^T z_t z_t^\top)(\Pi^0_0 - \hat{\Pi}_2)$$

$$\geq ||\Pi^0_0 - \hat{\Pi}_2||^2 ||T^{-1}\sum_{t=1}^T z_t z_t^\top|| + ||\Pi^0_0 - \hat{\Pi}_2||^2 ||T^{-1}\sum_{t=1}^T z_t z_t^\top||$$

$$= ||\Pi^0_0 - \hat{\Pi}_2||^2 ||M_{1,\rho} - M_{1,\rho^0 + \epsilon}|| + ||\Pi^0_0 - \hat{\Pi}_2||^2 ||M_{2,\rho^0 + \epsilon} - M_{2,\rho^0}|| + o_p(1).$$
Let the minimum eigenvalues of \((M_{1,p} - M_{1,p^0})\) and \((M_{2,p} - M_{2,p^0})\) be \(\eta_1\) and \(\eta_2\) respectively. By Assumption 4.1(d), \(\eta = \min(\eta_1, \eta_2) > 0\). Therefore,

\[
T^{-1} \sum_{t=1}^{T} d_t^T d_t \geq T^{-1} \sum_{AB} d_t^T d_t > \eta(||\hat{\Pi}_1^0 - \hat{\Pi}_2||^2 + ||\Pi_0^0 - \hat{\Pi}_2||^2)
\]

\[+ o_p(1) \geq ||\Pi_1^0 - \Pi_0^0||^2 \eta/2 + o_p(1) > K + o_p(1),\]

where for the second to last inequality, we used the fact that for any vectors \(a, b, c\) of the same length, \((a - b)^T(a - b) + (c - b)^T(c - b) \geq (a - c)^T(a - c)/2\). This means that \(\lim T^{-1} \sum_{t=1}^{T} d_t^T d_t > K\) 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 \(\rho^0 - \hat{\rho} \leq \epsilon, \text{now for a chosen small } \epsilon > 0, \text{as we know } \hat{\rho} \geq \rho^0\). 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 \(\rho\) instead of \(\hat{\rho}\) for simplicity. Let \(S_1\) be the multivariate sum of squared residuals in (2) evaluated at \(\rho\) and parameters \(\hat{\Pi}_1, \hat{\Pi}_2, S_2\) be the multivariate sum of squared residuals evaluated at \(\rho^0\) with parameters \(\Pi_{1,\rho^0}, \Pi_{2,\rho^0}\), and let \(S_3\) be the multivariate sum of squared residuals when regressing \(x_t\) on \(z_t [q_t \leq \rho], z_t [p < q_t \leq \rho^0], \text{and } z_t [q_t > \rho^0]\), evaluated at parameters \(\hat{\Pi}_{1,\rho}, \hat{\Pi}_2, \Pi_{2,\rho^0}\), where \(\hat{\Pi}_\Delta = (\sum_\Delta z_t z_t')^{-1} \sum_\Delta z_t x_t'\). Then, by definition, \(\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} \leq 0\). We will now show that \(\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} = O_p(1)\), then \(\lim \frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} \leq 0\). We also show that if \(\rho \in V_\epsilon(C)\), then, for sufficiently large \(C\) and small enough \(\epsilon\), \(\lim \frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} > K\), 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 \(\hat{Q}_\Delta = \frac{1}{T(\rho^0 - \rho)} \sum_\Delta z_t z_t'^T\) and \(\hat{M}_\Delta = T^{-1} \sum_\Delta z_t z_t'^T\), by Bai and Perron (1998), pp. 70,

\[
\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} = \text{trace}[(\hat{\Pi}_{2,\rho^0} - \hat{\Pi}_\Delta)^T (\hat{Q}_\Delta - \hat{Q}_\Delta \hat{M}_{2,\rho^0}^{-1} \hat{M}_\Delta)(\hat{\Pi}_{2,\rho^0} - \hat{\Pi}_\Delta)]
\]

\[
\frac{\hat{S}_2 - \hat{S}_3}{T(\rho^0 - \rho)} = \text{trace}[(\hat{\Pi}_{1,\rho} - \hat{\Pi}_\Delta)^T (\hat{Q}_\Delta - \hat{Q}_\Delta \hat{M}_{1,\rho}^{-1} \hat{M}_\Delta)(\hat{\Pi}_{1,\rho} - \hat{\Pi}_\Delta)].
\]

By standard arguments, \(\hat{\Pi}_{2,\rho^0} = \Pi_0^0 + o_p(1), \hat{\Pi}_\Delta = \Pi_0^0 + o_p(1)\), and by Lemma B.3, \(\Pi_{1,\rho} = \Pi_1 + o_p(1)\). Therefore, \(\hat{\Pi}_{2,\rho^0} - \hat{\Pi}_\Delta = \Pi_0^0 - \Pi_0^0 + o_p(1) = O_p(1)\), and \(\hat{\Pi}_{1,\rho} - \hat{\Pi}_\Delta = o_p(1)\). For \(C\) large enough, \(\hat{Q}_\Delta = O_p(1)\) and \(\hat{M}_\Delta = O_p(1)\) for \(i = 1, 2\). On the other hand,

\[
||\hat{M}_\Delta|| = ||E[z_t z_t'^T (I[q_t \leq \rho^0] - I[q_t \leq \rho])]| + o_p(1)|| \leq (E||z_t||^4)^1/2 \left(\int_{\rho^0}^{\rho^0 - \epsilon} f(x) dx\right)^{1/2}
\]

\[+ o_p(1) \leq M \left(\int_{\rho^0 - \epsilon}^{\rho^0} f(x) dx\right)^{1/2} + o_p(1) = M f(b) + o_p(1) \leq O_p(\epsilon),\]

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, \(\hat{Q}_\Delta - \hat{Q}_\Delta \hat{M}_{2,\rho^0}^{-1} \hat{M}_\Delta = \hat{Q}_\Delta + O_p(\epsilon) = O_p(1)\). This implies that \(\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} = O_p(1)\), and that \(\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} = o_p(1)\), therefore \(\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)} = O_p(1)\), which completes the first part of the proof.

It also follows that the asymptotically dominant term in \(\frac{\hat{S}_1 - \hat{S}_2}{T(\rho^0 - \rho)}\) is \(N_1 = \text{trace}[(\Pi_0^0 - \Pi_0^0)^T \hat{M}_\Delta (\Pi_0^0 - \Pi_0^0)]\). For \(\rho \in V_\epsilon(C)\), and \(C\) large enough, \(\hat{M}_\Delta = \frac{M_1, p - M_1, p^0}{\rho^0 - \rho} + o_p(1)\), and
because $M_{t,\rho} - M_{t,\rho}$ is p.d. by Assumption 4.1(d), it follows that $$\frac{S_i - S_j}{T(\rho - \rho)} = N_1 + O_p(\epsilon) > K > 0,$$ for some positive constant $K$. This completes the second part of the proof of (i), because it contradicts $\lim \frac{S_i - S_j}{T(\rho - \rho)} \leq 0$. Therefore, $\hat{\rho} \notin V_i(C)$, and so there is a $C$ for which $|T(\rho - \rho)| < C$.

**Part (iii).** Since any partial sum in the expression of $\hat{\Pi}_i$ differs by the partial sum in the expression of $\Pi_{t,\rho}$ by $|T(\rho - \rho)| < C$ terms, which are uniformly bounded by Assumption 4.1(c), it follows that $T^{1/2} \text{vec}(\Pi_{t,\rho}) = T^{1/2} \text{vec}(\Pi_{t,\rho} - \Pi_{t,\rho}) + o_p(1), \text{ for } i = 1, 2$. The rest of the proof follows standard arguments.

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

**Notation specific for a TFS.** Let $A_i^0 = [\Pi_{i,\rho}^0, S^T]^{\top}$ be the augmented matrices of the FS slope parameters, where $S = [I_{p_2}, 0_{p_2 \times q_1}]$, $q_1 = q - p_2$, and $\hat{A}_i = [\hat{\Pi}_{i,\rho}^0, S^T]^{\top}$. Hence, $z_{t,i} = S z_{t} + w_{t} = A_i^0 z_{t} \{q_t \leq \rho^0\} + A_i^0 z_{t} \{q_t > \rho^0\} + (u_t, 0_{q_1 \times 1})^{\top}$. Let $A_i^0 = A_i^0 \{q_t \leq \rho^0\} + A_i^0 \{q_t > \rho^0\}$. $A_i^0 = A_i^0 \{q_t \leq \rho^0\} + A_i^0 \{q_t > \rho^0\}$, $A_i^0 = A_i^0 \{q_t \leq \rho^0\} + A_i^0 \{q_t > \rho^0\}$, $A_i^0 = A_i^0 \{q_t \leq \rho^0\} + A_i^0 \{q_t > \rho^0\}$. Let $\land$ and $\lor$ define the minimum and maximum operators. Let $C_{1,\gamma} = A_i^0 M_{1,\rho} \land A_i^0$ and $C_{2,\gamma} = A_i^0 (M_{1,\rho} \lor A_i^0) + A_i^0 M_{2,\rho} \lor A_i^0$. Also, $C_\gamma = [C_{1,\gamma}, -C_{2,\gamma}]$, $C = C_{1,\gamma} + C_{2,\gamma}$, $R_\gamma = M_{1,\rho} M_{1,\rho}^{-1}$ for $i = 1, 2$, $D = [1, 0_{q_1 \times 1}] \otimes I_q$, and $Q_\gamma = C_{1,\gamma} C^{-1} C_{2,\gamma}$. Let $F_{\gamma} = [\hat{\theta}^{\top} \otimes R_{\gamma}]$ and $D_{\gamma} = [\hat{\theta}^{\top} \otimes I_q]$ - $F_{\gamma}$. Also define the Gaussian processes:

$$B_{1,\gamma}(\gamma) = \begin{cases} A_1^0 (D_{1,\gamma} G_1(\gamma) - F_{1,\gamma} (G_1(\rho^0) - G_1(\gamma))), & \gamma \leq \rho^0 \\ B - A_1^0 (D_{2,\gamma} G_2(\gamma) - F_{2,\gamma} (G_2(\gamma) - G_2(\rho^0))), & \gamma > \rho^0 \\ \end{cases}$$

where $B = \sum_{i=1}^{2} A_i^0 D G_i(\rho^0)$.

**Definition B.2.** Let $V_{B} = \sum_{i=1}^{2} A_i^0 D H_{1,\gamma} D^{\top} A_1^0$,

$$V_{B,1,1} = A_1^0 [D_{1,\gamma} G_{1,\gamma} + F_{1,\gamma} (H_{1,\rho} - H_{1,\gamma})] D^{\top} A_1^0, \quad \gamma \leq \rho^0,$$

$$V_{B,1,1} = A_1^0 [D_{1,\gamma} G_{1,\gamma} + F_{1,\gamma} (H_{1,\rho} - H_{1,\gamma})] D^{\top} A_1^0, \quad \gamma > \rho^0,$$

$$V_{B,2,2} = A_1^0 [D_{2,\gamma} G_{2,\gamma} + F_{2,\gamma} (H_{1,\rho} - H_{1,\gamma})] D^{\top} A_1^0, \quad \gamma < \rho^0,$$

$$V_{B,12} = A_1^0 [D_{1,\gamma} G_{1,\gamma} + F_{1,\gamma} (H_{1,\rho} - H_{1,\gamma})] D^{\top} A_1^0 - V_{B,1,1}, \quad \gamma \leq \rho^0,$$

$$V_{B,2,1} = A_1^0 [D_{2,\gamma} G_{2,\gamma} + F_{2,\gamma} (H_{1,\rho} - H_{1,\gamma})] D^{\top} A_1^0 - V_{B,1,1}, \quad \gamma > \rho^0,$$

and $V_{B,1,2} = V_{B} + V_{B,1,1} - V_{B,12,1} - V_{B,12,2}^{\top}$. Then $V_{B,\gamma} = \begin{bmatrix} V_{B,\gamma}^{11} & V_{B,\gamma}^{12} \\ V_{B,\gamma}^{12} & V_{B,\gamma}^{22} \end{bmatrix}$, and $V_{\gamma} = C_{1,\gamma} V_{B,\gamma} C_{1,\gamma}^{\top}$.

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

**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). $T^{-1} \sum_{t=1}^T z_t z_t^\top \overset{P}{\rightarrow} M_{\gamma}$ still holds, as the result is not specific to a LFS or TFS. So,

$$T^{-1} \sum_{t=1}^T \hat{w}_t \hat{w}_t^\top = \hat{A}_1 \sum_{\rho, \gamma} z_t z_t^\top \hat{A}_1^\top + \hat{A}_2 \left( T^{-1} \sum_{t=1}^T z_t z_t^\top \right) \hat{A}_2^\top$$

Because of part (i) of this theorem, the desired result still holds. But now,

$$\tilde{\epsilon}_t = \epsilon_t + (\hat{x}_t - x_t) \gamma^\top \theta_2^\rho = \epsilon_t + u_t^\top \theta_2^\rho - 1 [q_t \leq \rho^0] - 1 [q_t > \rho^0] [z_t^\top (\hat{\Pi}_1 - \Pi_1^0) \theta_2^\rho].$$

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

$$T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{\epsilon}_t = A_1^0 \left( T^{-1/2} \sum_{t=1}^T z_t (\epsilon_t + u_t^\top \theta_2^\rho) - M_{\gamma} \sum_{t=1}^T z_t u_t^\top \theta_2^\rho ) \right)$$

$$\Rightarrow A_1^0 \left( \gamma_0 \otimes I_1 \right) \mathcal{G}_1 (\gamma) - A_1^0 \left( \gamma_0 \otimes R_1 \right) \mathcal{G}_1 (\rho^0) = A_1^0 (D_1 \mathcal{G}_1 (\gamma) - F_1 \mathcal{G}_1 (\rho^0) - G_1 (\gamma)) = B_1 (\gamma).$$

For $\gamma > \rho^0$,

$$T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{\epsilon}_t = T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{\epsilon}_t - T^{-1/2} \sum_{t=2}^T \hat{w}_t \hat{\epsilon}_t$$

$$\Rightarrow A_1^0 D \mathcal{G}_1 (\rho^0) + A_2^0 D \mathcal{G}_2 (\rho^0) - A_2^0 (\rho^0 \otimes I_1) \mathcal{G}_2 (\gamma) - F_2 \mathcal{G}_2 (\rho^0)$$

$$= B - A_2^0 (\gamma_0 \otimes I_1) \mathcal{G}_2 (\gamma) - F_2 \mathcal{G}_2 (\rho^0) - G_1 (\gamma) - B_2 (\gamma).$$

Because $T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{\epsilon}_t \Rightarrow A_1^0 D \mathcal{G}_1 (\rho^0) + A_2^0 D \mathcal{G}_2 (\rho^0), T^{-1/2} \sum_{t=2}^T \hat{w}_t \hat{\epsilon}_t \Rightarrow A_1^0 D \mathcal{G}_1 (\rho^0) + A_2^0 D \mathcal{G}_2 (\rho^0) - B_1 (\gamma) = B - B_1 (\gamma) = B_2 (\gamma)$, and vec($T^{-1/2} \sum_{t=1}^T \hat{w}_t \hat{\epsilon}_t, T^{-1/2} \sum_{t=2}^T \hat{w}_t \hat{\epsilon}_t$) $\Rightarrow B(\gamma).$

Proof of Theorem 4.2. Part (i). Because $T^{-1/2} \hat{\theta}_1 \gamma - \hat{\theta}_2 \gamma = \hat{\theta}_1 \gamma - \hat{\theta}_2 \gamma$, 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 $\Pi_1 - \Pi_1^0 = o_p(1),$

$$T^{-1} \text{SSR}_{R_1} (\gamma) = T^{-1} \sum_{i=1}^2 \left( \sum_{i \in \rho^0} (\epsilon_t + u_t^\top \theta_2^\rho) \right)^2 - 2 \sum_{i \in \rho^0} (\epsilon_t + u_t^\top \theta_2^\rho) \gamma_1^\top (\Pi - \Pi_1^0) \theta_2^\rho$$

$$+ \sum_{i \in \rho^0} \theta_2^\rho \gamma_1^\top (\Pi - \Pi_1^0) \theta_2^\rho \gamma_1 = \sigma^2 + o_p(1),$$

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 $\gamma \leq \rho^0$ and $\gamma \geq \rho^0$, and taking to account the different parameter estimates in different regimes, that $V_{\gamma} \overset{P}{\rightarrow} V_{\gamma}$. 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 $\Pi$ with $\Pi_1$, and $\Pi^b$ with $\Pi_1^b$, 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.