Supplement Not for Online Publication to "Dynamic Local Average Treatment Effects in Time Series"

Alessandro Casini
University of Rome Tor Vergata

Adam McCloskey
University of Colorado at Boulder

Luca Rolla
University of Rome Tor Vergata

Raimondo Pala
University of Rome Tor Vergata

6th October 2025

Abstract

This supplemental material is structured as follows. Section N.A describes the collection of articles and specifications for the figure reported in the Introduction of the main article. Section N.B includes the theoretical results about the estimators of the true sub-population proposed in Section 5. Section N.C presents additional results on testing for full sample identification failure under homoskedasticity, primitive conditions for the assumptions of Section 6, results on consistent covariance matrix estimation for the tests introduced in Section 6 and results about identification-robust inference under strong IV and local or fixed alternatives. Section N.D presents additional Monte Carlo simulations.

N.A Publication Selection Criterion

We select recent publications from the following five economics journals: American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics and Review of Economic Studies. We first identify articles published in these journals between January 2019 and December 2022 that contain the keyword "instrument" in their text. We then exclude articles that do not estimate linear instrumental variables (IV) models or that are based solely on cross-sectional data. This results in 18 articles, listed in Table 1. From these 18 articles, we collect all IV specifications reported in their main text. Articles 1–3, 6–7, 9–12, and 14–16 use time series data, while articles 4–5, 8, 13, and 17–18 use panel data. Since the F^{*}statistic requires time series data, for panel data applications we treat each cross-sectional unit separately. In cases where an application includes thousands of cross-sectional units, we select only a subset to prevent a single panel application from dominating the distribution of the F statistics. The median cross-sectional size across applications is 34, so for panel specifications with more than 34 units we randomly select 34 units, while for those with fewer than 34 units we include all available units. When a specification involves multiple endogenous regressors, we run separate first-stage regressions for each endogenous regressor. This yields a total of 214 time series specifications and 1,346 panel data specifications.

Table 1: Selected Publications

Article ID	Year, Vol.(Issue)	Title	# of time series specifications	# of cross-sections in panel, # of specifications (# of cross-sections selected)
1	2019, AER	The Social Value of Financial Expertise	1	0
2	2020, AER	Turnover Liquidity and the Transmission of Monetary Policy	2	0
3	2021, AER	The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements	7	0
4	2021, AER	Stock Market Wealth and the Real Economy: A Local Labor Market Approach	0	2092, 2 (68)
5	2022, AER	Convex Supply Curves	0	21, 11 (231)
6	2022, ECMA	Monetary Policy, Redistribution, and Risk Premia	1	0
7	2022, JPE	Instrumental Variable Identification of Dynamic Variance Decompositions	2	0
8	2019, Restud	Innovation and Top Income Inequality	0	50, 8 (272)
9	2019, Restud	The Changing Returns to Crime: Do Criminals Respond to Prices?	8	0
10	2020, Restud	Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996–2016	2	0
11	2021, Restud	Bank Capital Redux: Solvency, Liquidity, and Crisis	0	15, 1 (15)
12	2022, Restud	Understanding the Size of the Government Spending Multiplier: It's in the Sign	120	0
13	2021, Restud	Housing Wealth Effects: The Long View	0	4400, 16 (544)
14	2022, Restud	Fiscal Multipliers and Foreign Holdings of Public Debt	48	0
15	2022, Restud	The Real Effects of Monetary Expansions: Evidence from a Large-scale Historical Experiment	5	6, 1 (6)
16	2020, Restud	Identifying Modern Macro Equations with Old Shocks	3	0
17	2022, QJE	Taxation and Innovation in the Twentieth Century	0	52, 1 (34)
18	2022, QJE	The Slope of the Phillips Curve: Evidence from U.S. States	0	34, 4 (191)

N.B Theoretical Results on Estimation of LATE and Sub-populations

In this section we establish theoretical results about the estimator $\widehat{\mathbf{S}}_{T,FGLS}$ from which we deduce the same results for $\widehat{\mathbf{S}}_{T,OLS}$ as a special case with $\vec{y} = D$ and $\widehat{\Omega}_{\varepsilon,\mathbf{S}} = I_T$. We first rewrite the model in matrix format as follows. We have 2 equations and T observations, excluding the initial conditions if lagged dependent variables are included among the regressors. The number of regimes that define the π sub-population is m, while the total number of regimes in the full sample is $\widetilde{m} \geq m$. For example, if $\pi = 1$ then $\widetilde{m} = m$, else $\widetilde{m} > m$. The break dates are denoted by the $\widetilde{m}$ vector $(T_1, \ldots, T_{\widetilde{m}})$ and we use the usual convention that $T_0 = 1$ and $T_{\widetilde{m}+1} = T$. A subscript i indexes a regime $(i = 1, \ldots, \widetilde{m} + 1)$, a subscript t indexes a temporal observation $(t = 1, \ldots, T)$ and a subscript j indexes the equation $(j = 1, 2)$ to which a scalar dependent variable y_{jt} belongs. According to our model in Section 5 $y_{1t} = Y_t$ and $y_{2t} = D_t$. $q + p$ is the number of regressors and z_t is the set that includes the regressors from all equations $z_t = (z_{1,t}, \ldots, z_{q+p,t})' = (Z_t', X_t')'$. The model considered in (5.1) can be written as

$$y_t = (I_2 \otimes z_t')\, \alpha_i + v_t, \tag{N.1}$$

where v_t has mean zero and covariance matrix Σ. The parameters in regime i are the $p + q$ vector $\alpha_i = (\beta\theta_i', \gamma_1' + \gamma_2'\beta, \theta_i', \gamma_2')'$, where $\theta_i = \theta$ for $T_{i-1} + 1 \leq t \leq T_i$ with $t \in \mathbf{S}_{0,T}$ and $\theta_i = 0$ for $T_{i-1} + 1 \leq t \leq T_i$ with $t \notin \mathbf{S}_{0,T}$. Let $\alpha = (\alpha_1', \ldots, \alpha_{\widetilde{m}+1}')'$.

To ease notation, define the $(q + p) \times 2$ matrix x_t by $x_t' = (I \otimes z_t')$ and rewrite (N.1) as

$$y_t = x_t'\alpha_i + v_t, \tag{N.2}$$

for $T_{i-1} + 1 \leq t \leq T_i$ $(i = 1, \ldots, \widetilde{m} + 1)$. We now express the model in matrix form. Let $\vec{Y} = (y_1', \ldots, y_T')'$ be the $2T$ vector of dependent variables, let $V = (v_1', \ldots, v_T')'$ be the error vector, and let the $2T \times 2\,(q + p)$ matrix of regressors be $\vec{X} = (x_1, \ldots, x_T)'$. For a given partition $\mathbf{S}$ with associated breaks $(T_1, \ldots, T_{\widetilde{m}})$, we define the block partition of the matrix $\vec{X}$ as the $2T \times 2\,(q + p)\,(\widetilde{m} + 1)$ matrix $\overline{X}\,(\mathbf{S}) = \mathrm{diag}(\overline{X}_1, \ldots, \overline{X}_{\widetilde{m}+1})$, where $\overline{X}_i$ $(i = 1, \ldots, \widetilde{m} + 1)$ is the $2\,(T_i - T_{i-1}) \times 2\,(q + p)$ subset of $\vec{X}$ that corresponds to observations in regime i. We also define the subvector V_i of V similarly. Then the regression (N.2) can be written as $\vec{Y} = \overline{X}\,(\mathbf{S})\,\alpha + V$. The true values of the parameters are denoted with a 0 superscript so that the data generating process is assumed to be $\vec{Y} = \overline{X}\,(\mathbf{S}_{0,T})\,\alpha_0 + V$, where $\overline{X}\,(\mathbf{S}_{0,T})$ is the diagonal partition of $\vec{X}$ using the partition $\mathbf{S}_{0,T}$, i.e. $(T_1^0, \ldots, T_{\widetilde{m}}^0)$. Let $\widehat{\Omega}_{\mathbf{S}}$ be the rearrangement of $\widehat{\Omega}_{\varepsilon,\mathbf{S}}$ in the main text corresponding to the rearrangement $\overrightarrow{Y}$ of $\overrightarrow{y}$. We make the following assumptions.

Assumption N.B.1. *$\sup_{\mathbf{S}} ||\overline{X}(\mathbf{S})' \widehat{\Omega}_{\mathbf{S}}^{-1/2}|| = O_{\mathbb{P}}(T^{1/2})$, $\sup_{\mathbf{S},\mathbf{S}'} ||\overline{X}(\mathbf{S})' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}(\mathbf{S}')/T|| = O_{\mathbb{P}}(1)$ and $\sup_{\mathbf{S}} ||V' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}(\mathbf{S})|| = O_{\mathbb{P}}(T^{1/2})$.*

Assumption N.B.2. *There exists an $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $(1/l)\sum_{t=T_i^0+1}^{T_i^0+l} \sum_{s=T_i^0+1}^{T_i^0+l} x_t [\widehat{\Omega}_{\mathbf{S}}^{-1}]_{(t,s)} x_s'$ and $(1/l)\sum_{t=T_i^0+1}^{T_i^0+l} \sum_{s=T_i^0+1}^{T_i^0+l} x_t [\widehat{\Omega}_{\mathbf{S}}^{-1}]_{(t,s)} x_s'$ are bounded away from zero uniformly over $i = 1, \dots, \widetilde{m}$ and $\mathbf{S}$ where $[\widehat{\Omega}_{\mathbf{S}}^{-1}]_{(t,s)}$ denotes the (t, s)-th element of $\widehat{\Omega}_{\mathbf{S}}^{-1}$.*

Assumption N.B.3. *The matrix $\sum_{t=k}^{l} \sum_{s=k}^{l} x_t [\widehat{\Omega}_{\mathbf{S}}^{-1}]_{(t,s)} x_s'$ is invertible for $l - k \geq k_0$ for some $0 < k_0 < \infty$.*

Assumption N.B.4. *We have $0 < \lambda_1^0 < \cdots < \lambda_{\widetilde{m}}^0 < 1$ with $T_i^0 = \lfloor T\lambda_i^0 \rfloor$.*

Assumption N.B.5. *The minimization search is taken over all partitions that satisfy $|\lambda_{i+1} - \lambda_i| \geq \epsilon$, $|\lambda_1| \geq \epsilon$, $|\lambda_{\widetilde{m}}| \leq 1 - \epsilon$.*

Assumption N.B.6. *For $\Omega = \mathbb{E}[VV'|\overline{X}(\mathbf{S}_{0,T})]$, $\sup_{\mathbf{S},\mathbf{S}'} T^{-1} \overline{X}(\mathbf{S}')' (\widehat{\Omega}_{\mathbf{S}}^{-1} - \Omega^{-1}) \overline{X}(\mathbf{S}) \xrightarrow{\mathbb{P}} 0$, $\sup_{\mathbf{S}} T^{-1} \overline{X}(\mathbf{S})' (\widehat{\Omega}_{\mathbf{S}}^{-1} - \Omega^{-1}) V \xrightarrow{\mathbb{P}} 0$ and $T^{-1} V' (\widehat{\Omega}_{\mathbf{S}}^{-1} - \Omega^{-1}) V \xrightarrow{\mathbb{P}} 0$.*

N.B.1 Consistency Under Fixed Shifts

Let $\widehat{\lambda}$ be the estimate of the break fractions $\lambda_0 = (\lambda_1^0, \lambda_2^0, \dots, \lambda_{\widetilde{m}}^0)$ that corresponds to $\widehat{\mathbf{S}}_{T,FGLS}$. The following proposition states the consistency of $\widehat{\lambda}$ for λ_0.

Proposition N.B.1. *Let Assumptions N.B.1-N.B.6 hold. Then, $\widehat{\lambda}_i \xrightarrow{\mathbb{P}} \lambda_i^0$, $i = 1, \dots, \widetilde{m}$.*

We now consider the rate of convergence of $\widehat{\lambda}$.

Proposition N.B.2. *Let Assumptions N.B.1-N.B.6 hold, for every $\eta > 0$, there exists a $C < \infty$, such that for all large T,*

$$\mathbb{P}\left(\left|T\left(\widehat{\lambda}_i - \lambda_i^0\right)\right| > C\right) < \eta, \qquad (i = 1, \dots, \widetilde{m}).$$

Let $\widehat{\alpha}(\cdot)$ be defined in analogy with $\widehat{\xi}_{FGLS}(\cdot)$ in the main text upon rearrangement of $\vec{y}$ to $\vec{Y}$. The T rate of convergence of $\widehat{\lambda}_i$ allows us to obtain the asymptotic equivalence between the estimated slope coefficients with the estimated subpopulation $\widehat{\alpha}(\widehat{\mathbf{S}}_{T,FGLS})$ and the estimated slope coefficients with known subpopulation $\widehat{\alpha}(\mathbf{S}_{0,T})$ so that standard results feasbilbe generalized least squares results implying $\sqrt{T}$ asymptotic normality for the latter also immediately apply to the former.

Proposition N.B.3. *Let Assumptions N.B.1-N.B.5 hold. We have $\sqrt{T}(\widehat{\alpha}(\widehat{\mathbf{S}}_{T,FGLS}) - \alpha_0) = O_{\mathbb{P}}(1)$.*

N.B.2 Proofs

N.B.2.1 Proof of Proposition N.B.1

We first outline the main steps of the proof using a few lemmas that are proved below. By the definition of $\widehat{\mathbf{S}}_{T,FGLS}$ and Assumption N.B.6,

$$\frac{1}{T}\widehat{V}\left(\widehat{\mathbf{S}}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\widehat{V}\left(\widehat{\mathbf{S}}\right) \leq \frac{1}{T}V'\Omega^{-1}V, \tag{N.3}$$

with probability approaching one, where $\widehat{V}(\widehat{\mathbf{S}}) = \vec{Y} - \overline{X}\left(\widehat{\mathbf{S}}\right)\widehat{\alpha}\left(\widehat{\mathbf{S}}\right)$ with $\widehat{\mathbf{S}} = \widehat{\mathbf{S}}_{T,FGLS}$. Note that

$$\begin{aligned}
&\widehat{V}\left(\widehat{\mathbf{S}}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\widehat{V}\left(\widehat{\mathbf{S}}\right) \qquad\qquad (\text{N.4})\\
&= \left(V - \overline{X}\left(\widehat{\mathbf{S}}\right)\left(\widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \alpha_0\right) - \left(\overline{X}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_0\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\\
&\quad \times \left(V - \overline{X}\left(\widehat{\mathbf{S}}\right)\left(\widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \alpha_0\right) - \left(\overline{X}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_0\right)\\
&= V'\Omega^{-1}V + \left(V'\widehat{\Omega}_{\mathbf{S}}^{-1}V - V'\Omega^{-1}V\right) + \left(\widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \alpha_0\right)'\overline{X}\left(\widehat{\mathbf{S}}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\widehat{\mathbf{S}}\right)\left(\widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \alpha_0\right)\\
&\quad + \alpha_0'\left(\overline{X}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\left(\overline{X}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_0\\
&\quad + 2\left(\widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \alpha_0\right)'\overline{X}\left(\widehat{\mathbf{S}}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\left(\overline{X}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_0\\
&\quad - 2V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\widehat{\mathbf{S}}\right)\left(\widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \alpha_0\right) - 2V'\widehat{\Omega}_{\mathbf{S}}^{-1}\left(\overline{X}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_0\\
&\equiv V'\Omega^{-1}V + \sum_{j=1}^{6} E_j.
\end{aligned}$$

The proof of Proposition N.B.1 uses (N.3)-(N.4) and the limit of $E_1, \ldots, E_6$. By Assumption N.B.6, $T^{-1}E_1 \xrightarrow{\mathbb{P}} 0$ and we show that $T^{-1}E_j \xrightarrow{\mathbb{P}} 0$ for $j = 5$ and 6, in Lemma N.B.1 below. These results combined with (N.3) imply that $T^{-1}\left(E_2 + E_3 + E_4\right) \xrightarrow{\mathbb{P}} 0$. The proof follows by showing that the latter imply $\widehat{\lambda} \xrightarrow{\mathbb{P}} \lambda_0$ via Lemma N.B.2. We proceed with a couple of lemmas.

Lemma N.B.1. *Let Assumptions N.B.1 and N.B.3 hold. We have $T^{-1}E_j \xrightarrow{\mathbb{P}} 0$ for $j = 5$ and 6.*

Proof of Lemma N.B.1. To prove the lemma, it suffices to show that

$$\sup_{\mathbf{S}} \frac{1}{T}\left|V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\left(\widehat{\alpha}\left(\mathbf{S}\right) - \alpha_0\right)\right| = O_{\mathbb{P}}\left(T^{-1/2}\right) = o_{\mathbb{P}}\left(1\right), \tag{N.5}$$

$$\sup_{\mathbf{S}} \frac{1}{T}\left|V'\widehat{\Omega}_{\mathbf{S}}^{-1}\left(\overline{X}\left(\mathbf{S}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_0\right| = O_{\mathbb{P}}\left(T^{-1/2}\right) = o_{\mathbb{P}}\left(1\right). \tag{N.6}$$

First consider (N.5). We can rewrite

$$\begin{aligned}
&V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\widehat{\alpha}\left(\mathbf{S}\right)-V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\alpha_{0}\\
&\quad=V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\left(\overline{X}\left(\mathbf{S}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\right)^{-1}\overline{X}\left(\mathbf{S}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}_{0,T}\right)\alpha_{0}\\
&\qquad+V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\left(\overline{X}\left(\mathbf{S}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\right)^{-1}\overline{X}\left(\mathbf{S}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}V\\
&\qquad-V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\alpha_{0}.
\end{aligned}$$

Using Assumptions N.B.1 and N.B.3, the first term on the right-hand side is $O_{\mathbb{P}}(T^{1/2})\,O_{\mathbb{P}}(T^{-1})O_{\mathbb{P}}\left(T\right)=O_{\mathbb{P}}(T^{1/2})$ uniformly over all partitions. The second term is $O_{\mathbb{P}}(T^{1/2})O_{\mathbb{P}}\left(T^{-1}\right)O_{\mathbb{P}}(T^{1/2})=O_{\mathbb{P}}\left(1\right)$ and the third term is $O_{\mathbb{P}}(T^{1/2})$, both uniformly over all partitions. Then, (N.5) follows. Next, consider (N.6). Using Assumption N.B.1, we have

$$\begin{aligned}
V'\widehat{\Omega}_{\mathbf{S}}^{-1}\left(\overline{X}\left(\mathbf{S}\right)-\overline{X}\left(\mathbf{S}_{0,T}\right)\right)\alpha_{0}&=V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}\right)\alpha_{0}-V'\widehat{\Omega}_{\mathbf{S}}^{-1}\overline{X}\left(\mathbf{S}_{0,T}\right)\alpha_{0}\\
&=O_{\mathbb{P}}\left(T^{1/2}\right)+O_{\mathbb{P}}\left(T^{1/2}\right).
\end{aligned}$$

This implies (N.6). $\square$

Lemma N.B.2. *Let Assumptions N.B.2-N.B.5 hold. If $\widehat{\lambda}_i \overset{\mathbb{P}}{\nrightarrow} \lambda_i^0$ for some i, then*

$$\liminf_{T\to\infty}\mathbb{P}\left(T^{-1}\left(E_{2}+E_{3}+E_{4}\right)>c\right)>\epsilon_{0}$$

for some $c>0$ and $\epsilon_0>0$.

Proof of Lemma N.B.2. We have for $T_{i-1}+1\leq t\leq T_i$,

$$\begin{aligned}
\widehat{v}_{t}\left(\widehat{\mathbf{S}}\right)&=\begin{bmatrix}Y_{t}\\ D_{t}\end{bmatrix}-\begin{bmatrix}Z_{t}'\widehat{\theta}_{\beta,\hat{i}}\\ Z_{t}'\widehat{\theta}_{\hat{i}}\end{bmatrix}-\begin{bmatrix}X_{t}'\widehat{\gamma}_{\beta}\\ X_{t}'\widehat{\gamma}_{2}\end{bmatrix}=\begin{bmatrix}Z_{t}'\theta_{\beta,i}\\ Z_{t}'\theta_{i}\end{bmatrix}+\begin{bmatrix}X_{t}'\gamma_{\beta}\\ X_{t}'\gamma_{2}\end{bmatrix}-\begin{bmatrix}Z_{t}'\widehat{\theta}_{\beta,\hat{i}}\\ Z_{t}'\widehat{\theta}_{\hat{i}}\end{bmatrix}-\begin{bmatrix}X_{t}'\widehat{\gamma}_{\beta}\\ X_{t}'\widehat{\gamma}_{2}\end{bmatrix}+v_{t}\\
&=\begin{bmatrix}Z_{t}'\left(\theta_{\beta,i}-\widehat{\theta}_{\beta,\hat{i}}\right)\\ Z_{t}'\left(\theta_{i}-\widehat{\theta}_{\hat{i}}\right)\end{bmatrix}+\begin{bmatrix}X_{t}'\left(\gamma_{\beta}-\widehat{\gamma}_{\beta}\right)\\ X_{t}'\left(\gamma_{2}-\widehat{\gamma}_{2}\right)\end{bmatrix}+v_{t},
\end{aligned}$$

where $\alpha_i=(\theta_{\beta,i},\,\gamma_\beta,\,\theta_i',\,\gamma_2')'$, $\widehat{\alpha}_{\hat{i}}=(\widehat{\theta}_{\beta,\hat{i}},\,\widehat{\gamma}_\beta,\,\widehat{\theta}_{\hat{i}}',\,\widehat{\gamma}_2')'$ and $\hat{i}$ corresponds to a regime in $\widehat{\mathbf{S}}$.

By Assumptions N.B.4 and N.B.5, if there exists a break, say λ_i^0, which cannot be consistently estimated, then with some probability $\epsilon_0>0$ there exists a $\eta>0$ such that no estimated break falls in the interval $[T(\lambda_j^0-\eta),\,T(\lambda_j^0+\eta)]$ for a subsequence of T. Suppose this interval is classified into the k-th regime, i.e., $\widehat{T}_{k-1}\leq T\left(\lambda_i^0-\eta\right)$ and $T\left(\lambda_i^0+\eta\right)\leq\widehat{T}_k$. Let d_t denote the difference between the fitted residuals and true errors. Then,

$$d_t = \begin{cases} \begin{bmatrix} Z_t' \left(\theta_{\beta,i} - \widehat{\theta}_{\beta,k}\right) \\ Z_t' \left(\theta_i - \widehat{\theta}_k\right) \end{bmatrix} + \begin{bmatrix} X_t' \left(\gamma_\beta - \widehat{\gamma}_\beta\right) \\ X_t' \left(\gamma_2 - \widehat{\gamma}_2\right) \end{bmatrix} & \text{for } t \in \left[T\left(\lambda_i^0 - \eta\right), T\lambda_i^0\right] \\ \begin{bmatrix} Z_t' \left(\theta_{\beta,i+1} - \widehat{\theta}_{\beta,k}\right) \\ Z_t' \left(\theta_{i+1} - \widehat{\theta}_k\right) \end{bmatrix} + \begin{bmatrix} X_t' \left(\gamma_\beta - \widehat{\gamma}_\beta\right) \\ X_t' \left(\gamma_2 - \widehat{\gamma}_2\right) \end{bmatrix} & \text{for } t \in \left[T\lambda_i^0, T\left(\lambda_i^0 + \eta\right)\right]. \end{cases}$$

For $t \in [T(\lambda_i^0 - \eta), T\lambda_i^0]$,

$$\begin{aligned} d_t' &= \begin{bmatrix} \theta_{\beta,i} - \widehat{\theta}_{\beta,k} & \theta_i - \widehat{\theta}_k \end{bmatrix} Z_t + \begin{bmatrix} \gamma_\beta - \widehat{\gamma}_\beta & \gamma_2 - \widehat{\gamma}_2 \end{bmatrix} X_t \\ &\equiv a_{\beta,i,k}' \boldsymbol{x}_t, \end{aligned}$$

where $\boldsymbol{x}_t = (Z_t', X_t')'$ while for $t \in [T\lambda_i^0, T\left(\lambda_i^0 + \eta\right)]$,

$$\begin{aligned} d_t' &= \begin{bmatrix} \theta_{\beta,i+1} - \widehat{\theta}_{\beta,k} & \theta_{i+1} - \widehat{\theta}_k \end{bmatrix} Z_t + \begin{bmatrix} \gamma_\beta - \widehat{\gamma}_\beta & \gamma_2 - \widehat{\gamma}_2 \end{bmatrix} X_t \\ &\equiv b_{\beta,i,k}' \boldsymbol{x}_t. \end{aligned}$$

We have

$$\begin{aligned} E_2 + E_3 + E_4 &= \left(\widehat{V}\left(\widehat{\mathbf{S}}\right) - V\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \left(\widehat{V}\left(\widehat{\mathbf{S}}\right) - V\right) \\ &= \left(\overline{X}\left(\widehat{\mathbf{S}}\right) \widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\alpha_0\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \left(\overline{X}\left(\widehat{\mathbf{S}}\right) \widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right)\alpha_0\right) \\ &= D\left(\widehat{\mathbf{S}}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} D\left(\widehat{\mathbf{S}}\right) \\ &\geq \sum_{t=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} \sum_{s=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} d_t' \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} d_s + \sum_{t=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} \sum_{s=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} d_t' \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} d_s \\ &= a_{\beta,i,k}' \sum_{t=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} \sum_{s=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} \boldsymbol{x}_t \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} \boldsymbol{x}_s' a_{\beta,i,k} + b_{\beta,i,k}' \sum_{t=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} \sum_{s=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} \boldsymbol{x}_t \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} \boldsymbol{x}_s' b_{\beta,i,k} \qquad \text{(N.7)} \\ &\geq \gamma_T \left[\left\|\theta_{\beta,i} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\gamma_\beta - \widehat{\gamma}_\beta\right\|^2 + \left\|\theta_i - \widehat{\theta}_k\right\|^2 + \left\|\gamma_2 - \widehat{\gamma}_2\right\|^2\right] \\ &\quad + \gamma_T^* \left[\left\|\theta_{\beta,i+1} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\gamma_\beta - \widehat{\gamma}_\beta\right\|^2 + \left\|\theta_{i+1} - \widehat{\theta}_k\right\|^2 + \left\|\gamma_2 - \widehat{\gamma}_2\right\|^2\right] \\ &\geq \min\left\{\gamma_T, \gamma_T^*\right\} \left(\left\|\theta_{\beta,i} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\theta_i - \widehat{\theta}_k\right\|^2 + \left\|\theta_{\beta,i+1} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\theta_{i+1} - \widehat{\theta}_k\right\|^2\right) \\ &\geq 2^{-1} \min\left\{\gamma_T, \gamma_T^*\right\} \left(\left\|\theta_{\beta,i} - \theta_{\beta,i+1}\right\|^2 + \left\|\theta_i - \theta_{i+1}\right\|^2\right), \end{aligned}$$

where $D(\widehat{\mathbf{S}})=[d_1'\,d_2'\cdots d_T']'$, γ_T and γ_T^* are the smallest eigenvalues of the first and second matrices on the left-hand side of the second inequality, and the last inequality follows from

$$(x-a)'\,A\,(x-a)+(x-b)'\,A\,(x-b)\geq\frac{1}{2}\,(a-b)'\,A\,(a-b)$$

for an arbitrary positive definite matrix A and for all x. Now, the first matrix in (N.7) can be written as

$$(T\eta)\,\frac{1}{T\eta}\sum_{T\left(\lambda_j^0-\eta\right)}^{T\lambda_j^0}\;\sum_{s=T\left(\lambda_i^0-\eta\right)}^{T\lambda_i^0}\boldsymbol{x}_t\left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)}\boldsymbol{x}_s'\equiv(T\eta)\,\boldsymbol{A}_T.$$

By Assumption N.B.2, the smallest eigenvalue of $\boldsymbol{A}_T$ is bounded away from zero. Thus, γ_T is of the order $(T\eta)$. A similar argument can be applied to γ_T^*. Therefore,

$$\sum_{j=2}^{4}E_j>T\eta c_1\min\{\|\theta_{\beta,i}-\theta_{\beta,i+1}\|^2\,,\,\|\theta_i-\theta_{i+1}\|^2\}=TC\min\{\|\theta_{\beta,i}-\theta_{\beta,i+1}\|^2\,,\,\|\theta_i-\theta_{i+1}\|^2\},$$

for some $C=\eta c_1>0$ with probability no less than $\epsilon_0>0$ as $T\to\infty$. □

Proof of Proposition N.B.1. Using (N.4), $T^{-1}E_1\xrightarrow{\mathbb{P}}0$ and Lemmas N.B.1-N.B.2, and under the supposition that some break date is not consistently estimated, we have the inequality

$$\frac{1}{T}\widehat{V}\left(\widehat{\mathbf{S}}\right)'\widehat{\Omega}_{\mathbf{S}}^{-1}\widehat{V}\left(\widehat{\mathbf{S}}\right)\geq\frac{1}{T}V'\Omega^{-1}V+C+o_{\mathbb{P}}(1)$$

for some $C>0$ holding with probability no less than some ϵ_0 as $T\to\infty$. This is in contradiction with (N.3). Hence, all break fractions are consistently estimated. □

N.B.2.2 Proof of Proposition N.B.2

Without loss of generality, we assume there are only three regimes ($\widetilde{m}=3$) and provide an explicit proof of T-consistency for $\widehat{\lambda}_2$ only. The analysis for $\widehat{\lambda}_1$ and $\widehat{\lambda}_3$ is virtually the same and is omitted.

By Proposition N.B.1, for each $\epsilon>0$ and T large, we have $|\widehat{T}_i-T_i^0|\leq\epsilon T$ with probability approaching one. For each $\epsilon>0$, let $\mathbf{T}_\epsilon=\{(T_1,\,T_2,\,T_3):\,|\widehat{T}_i-T_i^0|\leq\epsilon T$ for $i=1,\ldots,3\}$ so that $\mathbb{P}(\{\widehat{T}_1,\,\widehat{T}_2,\,\widehat{T}_3\}\in\mathbf{T}_\epsilon)\to1$. Therefore we only need to examine the behavior of the objective function, $Q_T(T_1,\,T_2,\,T_3)=\widehat{V}\,(\mathbf{S})'\,\widehat{\Omega}_{\mathbf{S}}^{-1}\widehat{V}\,(\mathbf{S})$, for those T_i corresponding to $\mathbf{S}$ that are close to the true breaks such that $|T_i-T_i^0|<\epsilon T$ for all i. Also using an argument of symmetry, we can without loss of generality, restrict attention to the case $T_2 < T_2^0$. For $C > 0$, define

$$\mathbf{T}_\epsilon(C) = \left\{(T_1, T_2, T_3) : \left|T_i - T_i^0\right| < \epsilon T,\, 1 \leq i \leq 3,\, T_2 - T_2^0 < -C\right\}.$$

Note that $\mathbf{T}_\epsilon(C) \subset \mathbf{T}_\epsilon$. Because $Q_T(\widehat{T}_1, \widehat{T}_2, \widehat{T}_3) \leq Q_T(\widehat{T}_1, T_2^0, \widehat{T}_3)$ with probability 1, to prove the proposition it is enough to show that for each $\eta > 0$, there exist $C > 0$ and $\epsilon > 0$ such that for large T,

$$\mathbb{P}\left(\min_{\mathbf{T}_\epsilon(C)} \left\{Q_T(T_1, T_2, T_3) - Q_T\left(T_1, T_2^0, T_3\right)\right\} \leq 0\right) < \eta, \tag{N.8}$$

or equivalently,

$$\mathbb{P}\left(\min_{\mathbf{T}_\epsilon(C)} \left\{\left[Q_T(T_1, T_2, T_3) - Q_T\left(T_1, T_2^0, T_3\right)\right] / \left(T_2^0 - T_2\right)\right\} \leq 0\right) < \eta. \tag{N.9}$$

That would imply that for a large C, global minimization cannot be achieved on $\mathbf{T}_\epsilon(C)$. Thus with probability approaching one, $|\widehat{T}_2 - T_2^0| \leq C$. Now denote

$$\begin{aligned}
Q_{1,T} &= Q_T(T_1, T_2, T_3) \\
Q_{2,T} &= Q_T\left(T_1, T_2^0, T_3\right) \\
Q_{3,T} &= \widehat{V}(\mathbf{S}_{3,T})' \widehat{\Omega}_{\mathbf{S}}^{-1} \widehat{V}(\mathbf{S}_{3,T})
\end{aligned}$$

where $\mathbf{S}_{3,T}$ is the partition based on (T_1, T_2, T_2^0, T_3). Subtracting and adding $Q_{3,T}$, we have

$$Q_{1,T} - Q_{2,T} = Q_{1,T} - Q_{3,T} - (Q_{2,T} - Q_{3,T}).$$

This latter relation is useful because it allows us to perform the analysis in terms of two problems involving a single break. Indeed, $Q_{1,T} - Q_{3,T}$ is the difference in the objective function allowing an additional fourth break at time T_2^0 between the breaks T_2 and T_3. Similarly, $Q_{2,T} - Q_{3,T}$ is the difference in the objective function allowing an additional fourth break at time T_2 between the breaks T_1 and T_2^0. Consider $Q_{1,T} - Q_{3,T}$ first. Let $(\widehat{\alpha}_1^*, \widehat{\alpha}_2^*, \widehat{\alpha}_\Delta^*, \widehat{\alpha}_3^*, \widehat{\alpha}_4^*)$ denote the estimator of $(\alpha_1^0, \alpha_2^0, \alpha_2^0, \alpha_3^0, \alpha_4^0)$. In particular, $\widehat{\alpha}_2^*$ is an estimate of α_2^0 associated with the regressors $(0, \ldots 0, x_{T_1+1}, \ldots, x_{T_2}, 0, \ldots 0)'$, $\widehat{\alpha}_\Delta^*$ is the vector of estimated coefficients associated with the regressors $X_\Delta = (0, \ldots 0, x_{T_2+1}, \ldots, x_{T_2^0}, 0, \ldots 0)'$.

From the argument on p. 31 in Amemiya (1985),

$$\begin{aligned} Q_{1,T} - Q_{3,T} &= \widehat{V}\left(\mathbf{S}_{1,T}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \widehat{V}\left(\mathbf{S}_{1,T}\right) - \widehat{V}\left(\mathbf{S}_{3,T}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \widehat{V}\left(\mathbf{S}_{3,T}\right) \\ &= \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right)' X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1/2} M_{\widehat{\Omega}_{\mathbf{S}}^{-1/2}\overline{X}\left(\mathbf{S}_{1,T}\right)} \widehat{\Omega}_{\mathbf{S}}^{-1/2} X_\Delta \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right), \end{aligned}$$

where $M_X = I - X\left(X'X\right)^{-1} X'$ for a matrix X and $\widehat{\alpha}_3^*$ is the vector of estimated coefficients associated with the regressors $(0, \ldots 0, x_{T_2^0+1}, \ldots, x_{T_3}, 0, \ldots 0)'$. Similarly, we have for $Q_{2,T} - Q_{3,T}$,

$$Q_{2,T} - Q_{3,T} = \left(\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\right)' X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1/2} M_{\widehat{\Omega}_{\mathbf{S}}^{-1/2}\overline{X}\left(\mathbf{S}_{2,T}\right)} \widehat{\Omega}_{\mathbf{S}}^{-1/2} X_\Delta \left(\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\right).$$

Thus,

$$\begin{aligned} Q_{1,T} - Q_{2,T} \geq & \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right)' X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1/2} M_{\widehat{\Omega}_{\mathbf{S}}^{-1/2}\overline{X}\left(\mathbf{S}_{1,T}\right)} \widehat{\Omega}_{\mathbf{S}}^{-1/2} X_\Delta \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right) \\ & - \left(\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\right)' X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta \left(\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\right), \end{aligned} \tag{N.10}$$

where we used

$$X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1/2} M_{\widehat{\Omega}_{\mathbf{S}}^{-1/2}\overline{X}\left(\mathbf{S}_{2,T}\right)} \widehat{\Omega}_{\mathbf{S}}^{-1/2} X_\Delta \leq X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta.$$

From the definition of $M_{\widehat{\Omega}_{\mathbf{S}}^{-1/2}\overline{X}\left(\mathbf{S}_{1,T}\right)}$, we have

$$\begin{aligned} \frac{Q_{1,T} - Q_{2,T}}{T_2^0 - T_2} \geq & \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right)' \frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right) \\ & - \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right)' \frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}\left(\mathbf{S}_{1,T}\right)}{T_2^0 - T_2} \left[\frac{\overline{X}\left(\mathbf{S}_{1,T}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}\left(\mathbf{S}_{1,T}\right)}{T}\right]^{-1} \\ & \times \frac{\overline{X}\left(\mathbf{S}_{1,T}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T} \left(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*\right) \\ & - \left(\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\right)' \frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} \left(\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\right) + Q_{2,T} - Q_{2,T} \\ \equiv & L_1 - L_2 - L_3. \end{aligned} \tag{N.11}$$

Consider the limiting behavior of L_1. Note first that for small ϵ, the estimates $\widehat{\alpha}_i^*$ will be close to α_i^0 with high probability for large T given that, on the set $\mathbf{T}_\epsilon(C)$, the distance between T_i and T_i^0 can be made small by choosing a small ϵ. Further, $\widehat{\alpha}_\Delta^*$ is estimated using observations from the second true regime only and it is close to α_2^0 in probability on $\mathbf{T}_\epsilon(C)$ for a large enough C. Hence, for large C, large T and small ϵ, L_1 is larger than

$$(\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*)' \frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} (\widehat{\alpha}_3^* - \widehat{\alpha}_\Delta^*) \geq \frac{1}{2} \left(\alpha_3^0 - \alpha_2^0\right)' \frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} \left(\alpha_3^0 - \alpha_2^0\right)$$

with high probability.

Next consider L_2. It is easy to see that on $\mathbf{T}_\epsilon(C)$, $\widehat{\alpha}_3^*$ and $\widehat{\alpha}_\Delta^*$ are $O_{\mathbb{P}}(1)$ uniformly. Also on $\mathbf{T}_\epsilon(C)$,

$$\frac{\overline{X}(\mathbf{S}_{1,T})' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}(\mathbf{S}_{1,T})}{T} = O_{\mathbb{P}}(1),$$

and

$$\frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}(\mathbf{S}_{1,T})}{T_2^0 - T_2} = O_{\mathbb{P}}(1)$$

by Assumption N.B.1 since $X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X}(\mathbf{S}_{1,T})$ involves less than $T_2^0 - T_2$ observations. Furthermore,

$$\left\| \frac{\overline{X}(\mathbf{S}_{1,T})' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T} \right\| = \left\| \frac{\overline{X}(\mathbf{S}_{1,T})' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} \frac{T_2^0 - T_2}{T} \right\| \leq \epsilon O_{\mathbb{P}}(1).$$

Thus L_2 is no larger than $\epsilon O_{\mathbb{P}}(1)$. Consider finally L_3. Because both $\widehat{\alpha}_2^*$ and $\widehat{\alpha}_\Delta^*$ are close to α_2^0, $\|\widehat{\alpha}_2^* - \widehat{\alpha}_\Delta^*\| < \rho$ with probability approaching one for any given small number $\rho > 0$. We also have

$$\left\| \frac{\overline{X}(\mathbf{S}_{1,T})' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} \right\| = O_{\mathbb{P}}(1),$$

uniformly on $\mathbf{T}_\epsilon(C)$. Thus L_3 is no larger than $\rho O_{\mathbb{P}}(1)$. In summary, the following inequality holds with probability approaching one on $\mathbf{T}_\epsilon(C)$:

$$\frac{Q_{1,T} - Q_{2,T}}{T_2^0 - T_2} \geq \frac{1}{2} \left(\alpha_3^0 - \alpha_2^0\right)' \frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2} \left(\alpha_3^0 - \alpha_2^0\right) - \epsilon O_{\mathbb{P}}(1) - \rho O_{\mathbb{P}}(1). \tag{N.12}$$

By Assumption N.B.2,

$$\frac{X_\Delta' \widehat{\Omega}_{\mathbf{S}}^{-1} X_\Delta}{T_2^0 - T_2}$$

has its minimum eigenvalue bounded away from zero on $\mathbf{T}_\epsilon(C)$. Thus, the first term on the right-hand side of (N.12) is positive and larger in absolute value than the other two terms. Thus, we have

$$\frac{Q_{1,T} - Q_{2,T}}{T_2^0 - T_2} > 0$$

with probability approaching one. This proves (N.9) and the proposition. $\square$

N.B.3 Proof of Proposition N.B.3

Begin by noting that

$$\widehat{\Omega}_{\mathbf{S}}^{-1/2} \left(\overline{X} \left(\widehat{\mathbf{S}} \right) - \overline{X} \left(\mathbf{S}_{0,T} \right) \right)$$

involves $\sum_{i=1}^{\widetilde{m}} |\widehat{T}_i - T_i^0| = O_{\mathbb{P}} \left(\widetilde{m} \right)$ nonzero observations by Proposition N.B.2 so that, after applying Assumption N.B.1,

$$T^{-1} \overline{X} \left(\widehat{\mathbf{S}} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X} \left(\widehat{\mathbf{S}} \right) = T^{-1} \overline{X} \left(\mathbf{S}_{0,T} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X} \left(\mathbf{S}_{0,T} \right) + O_{\mathbb{P}} \left(T^{-1/2} \right).$$

Similarly,

$$T^{-1/2} \overline{X} \left(\widehat{\mathbf{S}} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X} \left(\mathbf{S}_{0,T} \right) \alpha_0 - \alpha_0 = T^{-1/2} \overline{X} \left(\widehat{\mathbf{S}} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \left(\overline{X} \left(\mathbf{S}_{0,T} \right) - \overline{X} \left(\widehat{\mathbf{S}} \right) \right) \alpha_0 = O_{\mathbb{P}} \left(T^{-1/2} \right)$$

and

$$T^{-1/2} \overline{X} \left(\widehat{\mathbf{S}} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} V - T^{-1/2} \overline{X} \left(\mathbf{S}_{0,T} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} V = T^{-1/2} \left(\overline{X} \left(\widehat{\mathbf{S}} \right) - \overline{X} \left(\mathbf{S}_{0,T} \right) \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} V = O_{\mathbb{P}} \left(T^{-1/2} \right)$$

so that another application of Assumption N.B.1 yields

$$\begin{aligned} \sqrt{T} \left(\widehat{\alpha} \left(\widehat{\mathbf{S}}_{T,FGLS} \right) - \widehat{\alpha} \left(\mathbf{S}_{0,T} \right) \right) &= \left(\left(T^{-1} \overline{X} \left(\mathbf{S}_{0,T} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X} \left(\mathbf{S}_{0,T} \right) \right)^{-1} + o_{\mathbb{P}} (1) \right) \\ \times T^{-1/2} \left(\overline{X} \left(\widehat{\mathbf{S}} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \overline{X} \left(\mathbf{S}_{0,T} \right) \alpha_0 - \alpha_0 + \overline{X} \left(\widehat{\mathbf{S}} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} V - \overline{X} \left(\mathbf{S}_{0,T} \right)' \widehat{\Omega}_{\mathbf{S}}^{-1} V \right) &= O_{\mathbb{P}} \left(T^{-1/2} \right). \quad \square \end{aligned}$$

N.C Additional Results

N.C.1 Identification of Compliers for Continuous Instrument

N.C.1.1 Identification of Compliers

We say that observation $t \in \mathbf{S}_{0,T}$ is a complier if and only if $\mathbb{E}(D_t(z) \mid \widetilde{V}_t)$ is strictly increasing in z almost surely. In the present case of a continuous instrument, the policy and control samples $\mathbf{P}$

and $\mathbf{C}$ need to be redefined relative to the simpler case of a binary instrument in the main text. Let $\mathbf{P} \subset \{1, \ldots, T\}$ and $\mathbf{C} \subset \{1, \ldots, T\}$ satisfy $\mathbf{P} \cap \mathbf{C} = \emptyset$ and $\min_{t \in \mathbf{P}} Z_t > \max_{t \in \mathbf{C}} Z_t$. Let $\mathbf{Z_P}$ denote the values in $\mathbf{Z}$ such that $Z_t \in \mathbf{Z_P}$ is equivalent to $t \in \mathbf{P}$ and similarly for $\mathbf{Z_C}$. Construct $\mathbf{P}$ and $\mathbf{C}$ such that $\underline{z}_{\mathbf{P}} \equiv \inf(\mathbf{Z_P}) > \sup(\mathbf{Z_C}) \equiv \overline{z}_{\mathbf{C}}$. For example, a simple way to define the policy and control samples is $\mathbf{P} = \{t \in \{1, \ldots, T\} : Z_t \geq \tilde{z} + \epsilon\}$ and $\mathbf{C} = \{t \in \{1, \ldots, T\} : Z_t \leq \tilde{z} - \epsilon\}$ for some $\tilde{z}$ and small $\epsilon > 0$. With these definitions, we impose the continuous instrument-analogs of Assumptions 2.7 and 2.8 in the main text.

Assumption N.C.1. *(i) For any $t \in \mathbf{C}$, $\overline{D}_{C,t,n} \xrightarrow{\mathbb{P}} \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_C})$ as $n \to \infty$ with $n/|\mathbf{C}| \to 0$. (ii) For $t \in \mathbf{P}$, $\mathbb{E}[D_{t-1}(Z_{t-1})|Z_{t-1} \in \mathbf{Z_C}] = \mathbb{E}[D_t(Z_t)|Z_t \in \mathbf{Z_C}]$.*

Assumption N.C.2. *(i) For any $t \in \mathbf{P}$, $\overline{D}_{P,t,n} \xrightarrow{\mathbb{P}} \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_P})$ as $n \to \infty$ with $n/|\mathbf{P}| \to 0$. (ii) For $t \in \mathbf{C}$, $\mathbb{E}[D_t(Z_t)|Z_t \in \mathbf{Z_P}] = \mathbb{E}[D_{s^*(t)}(Z_{s^*(t)})|Z_{s^*(t)} \in \mathbf{Z_P}]$, where $s^*(t) = \arg\min_{s \in \mathbf{P}} |t - s|$.*

Proposition N.C.1. *Let Assumptions 2.1, N.C.1 and N.C.2 hold and $n_0, n_1 \to \infty$ with $n_0/|\mathbf{C}|, n_1/|\mathbf{P}| \to 0$. We have:*
(i) if $t \in \mathbf{P}$ is a complier, then $\overline{D}_{P,t,n_1} - \overline{D}_{C,t-1,n_0} \xrightarrow{\mathbb{P}} c$ where $c > 0$.
(ii) if $t \in \mathbf{C}$ is a complier, then $\overline{D}_{P,s^*(t),n_1} - \overline{D}_{C,t,n_0} \xrightarrow{\mathbb{P}} \widetilde{c}$ where $\widetilde{c} > 0$.

Proof of Proposition N.C.1. Consider first the policy sample. Suppose $t \in \mathbf{P}$ is a complier. Then, by Assumptions N.C.1(i) and N.C.2(i), as $n_0, n_1 \to \infty$,

$$
\begin{aligned}
\overline{D}_{P,t,n_1} - \overline{D}_{C,t-1,n_0} &\xrightarrow{\mathbb{P}} \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_P}) - \mathbb{E}(D_{t-1}(Z_{t-1})|Z_{t-1} \in \mathbf{Z_C}) \\
&= \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_P}) - \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_C}) \\
&= \int_{\tilde{v}} \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_P}, \widetilde{V}_t = \tilde{v}) dF_{\widetilde{V}_t}(\tilde{v}) - \int_{\tilde{v}} \mathbb{E}(D_t(Z_t)|Z_t \in \mathbf{Z_C}, \widetilde{V}_t = \tilde{v}) dF_{\widetilde{V}_t}(\tilde{v}) \\
&= \int_{\tilde{v}} \int_{z \in \mathbf{Z_P}} \mathbb{E}(D_t(z)|\widetilde{V}_t = \tilde{v}) dF_{Z_t|\widetilde{V}_t = \tilde{v}, Z_t \in \mathbf{Z_P}}(z) dF_{\widetilde{V}_t}(\tilde{v}) \\
&\quad - \int_{\tilde{v}} \int_{z \in \mathbf{Z_C}} \mathbb{E}(D_t(z)|\widetilde{V}_t = \tilde{v}) dF_{Z_t|\widetilde{V}_t = \tilde{v}, Z_t \in \mathbf{Z_C}}(z) dF_{\widetilde{V}_t}(\tilde{v}) \\
&\geq \int_{\tilde{v}} \int_{z \in \mathbf{Z_P}} \mathbb{E}(D_t(\underline{z}_{\mathbf{P}})|\widetilde{V}_t = \tilde{v}) dF_{Z_t|\widetilde{V}_t = \tilde{v}, Z_t \in \mathbf{Z_P}}(z) dF_{\widetilde{V}_t}(\tilde{v}) \\
&\quad - \int_{\tilde{v}} \int_{z \in \mathbf{Z_C}} \mathbb{E}(D_t(\overline{z}_{\mathbf{C}})|\widetilde{V}_t = \tilde{v}) dF_{Z_t|\widetilde{V}_t = \tilde{v}, Z_t \in \mathbf{Z_C}}(z) dF_{\widetilde{V}_t}(\tilde{v}) \\
&= \mathbb{E}(D_t(\underline{z}_{\mathbf{P}})) - \mathbb{E}(D_t(\overline{z}_{\mathbf{C}})) > 0,
\end{aligned}
$$

where $F_{\widetilde{V}_t}(\cdot)$ is the distribution function of $\widetilde{V}_t$, $F_{Z_t|\widetilde{V}_t = \tilde{v}, Z_t \in \mathbf{Z_P}}(\cdot)$ is the conditional distribution function of Z_t given $\widetilde{V}_t = \tilde{v}$ and $Z_t \in \mathbf{Z_P}$, $F_{Z_t|\widetilde{V}_t = \tilde{v}, Z_t \in \mathbf{Z_C}}(\cdot)$ is the conditional distribution function of Z_t given $\widetilde{V}_t = \tilde{v}$ and $Z_t \in \mathbf{Z_C}$, the first equality follows from Assumption N.C.1(ii), the third equality follows from Assumption 2.1 and the inequalities follow from the definition of a complier. The proof for the control sample is entirely analogous and therefore omitted. □

Assumption N.C.3. *(Continuous Case Monotonicity)* $\mathbb{E}(D_t(z) \,|\, \widetilde{V}_t)$ *is monotonic in* z *almost surely.*

Under Assumption N.C.3, assume without loss of generality that $\mathbb{E}(D_t(z) \,|\, \widetilde{V}_t)$ is increasing in z almost surely. We obtain the following characterization of compliers under monotonicity for the continuous instrument case.

Corollary N.C.1. *Let Assumptions 2.3 and N.C.3 hold. Then, the set of compliers coincides with* $\mathbf{S}_{0,T}$.

Proof of Proposition N.C.1. Assumption N.C.3 rules out defiers, i.e., $\mathbb{E}(D_t(z) \,|\, \widetilde{V}_t)$ being strictly decreasing in z almost surely, so that non-compliers are characterized by $\mathbb{E}(D_t(z) \,|\, \widetilde{V}_t)$ being constant in z almost surely. Therefore, a non-complier cannot belong to $\mathbf{S}_{0,T}$ by definition. And any complier belongs to $\mathbf{S}_{0,T}$ by definition. □

N.C.2 Primitive Conditions on IVs, Exogenous Regressors and Errors for the Assumptions of Section 6.2

Assumptions 6.1-6.3 of Section 6.2 are implied by any one of the following assumptions:

Assumption N.C.4. $\{(v_t,\, w_t) : t \geq 1\}$ *are i.i.d.,* $\mathbb{E}\,(v_t \otimes w_t) = 0$, $\mathbb{E}||v_t||^2 + \mathbb{E}||w_t||^2 + \mathbb{E}||v_t \otimes w_t||^2 < \infty$, $\Sigma_v = \mathbb{E}(v_t v_t')$ *is positive definite, and uniformly in* $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$, *for* $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$ *and* $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$, $T^{-1}\sum_{t=1}^{T} \mathbb{E}(w_t(\mathbf{S}_T)\, w_t(\mathbf{S}_T')') \to Q(\mathbf{S}, \mathbf{S}')$ *for some* $(q+p) \times (q+p)$ *matrix* $Q(\mathbf{S}, \mathbf{S}')$ *for which* $Q(\mathbf{S}, \mathbf{S})$ *is positive definite and* $T^{-1}\sum_{t=1}^{T} \mathbb{E}((v_t \otimes w_t(\mathbf{S}_T))\, (v_t \otimes w_t(\mathbf{S}_T'))') \to \Psi(\mathbf{S}, \mathbf{S}')$ *for some* $2\,(q+p) \times 2\,(q+p)$ *matrix* $\Psi(\mathbf{S}, \mathbf{S}')$.

Assumption N.C.5. $\{(v_t,\, w_t) : t \geq 1\}$ *are independent,* $\mathbb{E}\,(v_t \otimes w_t) = 0$ *for all* $t \geq 1$, $\sup_{t\geq 1}(\mathbb{E}||v_t||^{2+\varsigma} + \mathbb{E}||w_t||^{2+\varsigma} + \mathbb{E}||v_t \otimes w_t||^{2+\varsigma}) < \infty$ *for some* $\varsigma > 0$, $T^{-1}\sum_{t=1}^{T} \mathbb{E}\,(v_t v_t') \to \Sigma_v$ *for some positive definite* 2×2 *matrix* Σ_v, *and uniformly in* $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$, *for* $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$ *and* $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$, $T^{-1}\sum_{t=1}^{T} \mathbb{E}(w_t(\mathbf{S}_T)\, w_t(\mathbf{S}_T')') \to Q(\mathbf{S}, \mathbf{S}')$ *for some* $(q+p) \times (q+p)$ *matrix* $Q(\mathbf{S}, \mathbf{S}')$ *for which* $Q(\mathbf{S}, \mathbf{S})$ *is positive definite and* $T^{-1}\sum_{t=1}^{T} \mathbb{E}((v_t \otimes w_t(\mathbf{S}_T))\, (v_t \otimes w_t(\mathbf{S}_T'))') \to \Psi(\mathbf{S}, \mathbf{S}')$ *for some* $2\,(q+p) \times 2\,(q+p)$ *matrix* $\Psi(\mathbf{S}, \mathbf{S}')$.

Assumption N.C.6. $\{(v_t \otimes w_t,\, \mathscr{F}_t) : t \geq 1\}$ *is a martingale difference sequence, where* $\mathscr{F}_t = \sigma(v_t,\, w_t,\, v_{t-1},\, w_{t-1},\, \ldots)$, $\{(v_t \otimes w_t) : t \geq 1\}$ *is an ergodic sequence,* $\sup_{t\geq 1}(\mathbb{E}||v_t||^2 + \mathbb{E}||w_t||^2 + \mathbb{E}||v_t \otimes w_t||^2) < \infty$,

$\Sigma_v = \mathbb{E}(v_t v_t')$ is positive definite, and uniformly in $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$, for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$ and $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$, $T^{-1}\sum_{t=1}^{T}\mathbb{E}(w_t(\mathbf{S}_T)\, w_t(\mathbf{S}_T')') \to Q(\mathbf{S},\mathbf{S}')$ for some $(q+p)\times(q+p)$ matrix $Q(\mathbf{S},\mathbf{S}')$ for which $Q(\mathbf{S},\mathbf{S})$ is positive definite and $T^{-1}\sum_{t=1}^{T}\mathbb{E}((v_t\otimes w_t(\mathbf{S}_T))\,(v_t\otimes w_t(\mathbf{S}_T'))') \to \Psi(\mathbf{S},\mathbf{S}')$ for some $2(q+p)\times 2(q+p)$ matrix $\Psi(\mathbf{S},\mathbf{S}')$.

Assumption N.C.7. *$\{(v_t,\, w_t) : t = \ldots 0, 1, \ldots\}$ is a doubly infinite ergodic sequence with $\mathbb{E}(v_t \otimes w_t) = 0$, $\sup_{t\geq 1}(\mathbb{E}||v_t||^2 + \mathbb{E}||w_t||^2 + \mathbb{E}||v_t\otimes w_t||^2) < \infty$, $\sup_{t\geq 1}\sum_{j=1}^{\infty}(\mathbb{E}||\mathbb{E}(v_t\otimes w_t|\, \mathscr{F}_{t-j})||^2)^{1/2} < \infty$ where $\mathscr{F}_t = \sigma(v_t,\, w_t,\, v_{t-1},\, w_{t-1}, \ldots)$, $T^{-1}\sum_{t=1}^{T}\mathbb{E}(v_t v_t') \to \Sigma_v$ for some positive definite 2×2 matrix Σ_v, and uniformly in $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$, for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$ and $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$, $T^{-1}\sum_{t=1}^{T}\mathbb{E}(w_t(\mathbf{S}_T)\, w_t(\mathbf{S}_T')') \to Q(\mathbf{S},\mathbf{S}')$ for some $(q+p)\times(q+p)$ matrix $Q(\mathbf{S},\mathbf{S}')$ for which $Q(\mathbf{S},\mathbf{S})$ is positive definite and $T^{-1}\sum_{t=1}^{T}\sum_{j=-\infty}^{\infty}\mathbb{E}(v_t\otimes w_t(\mathbf{S}_T))(v_{t-j}\otimes w_{t-j}(\mathbf{S}_T'))' \to \Psi(\mathbf{S},\mathbf{S}') = \int_0^1 \Psi_u(\mathbf{S},\mathbf{S}')\,du$, where $\Psi_u(\mathbf{S},\mathbf{S}')$ is the local long-run covariance matrix of $v_t\otimes w_t(\mathbf{S}_T)$ and $v_t\otimes w_t(\mathbf{S}_T')$.*

The random vectors $\{(v_t,\, w_t) : t = \ldots 0, 1, \ldots\}$ are uncorrelated under Assumptions N.C.4-N.C.6, but are (possibly) correlated under Assumption N.C.7. Assumptions N.C.5-N.C.7 allow for nonstationarity (i.e., time-varying moments). In particular, they are satisfied by segmented local stationarity [see Casini (2024, 2023)].

If the errors are conditionally homoskedastic and $\{(v_t,\, w_t) : t \geq 1\}$ are uncorrelated, the following assumption holds.

Assumption N.C.8. *$\Psi(\mathbf{S},\mathbf{S}') = \Sigma_v \otimes Q(\mathbf{S},\mathbf{S}')$, where $Q(\cdot)$ is defined in Assumption 6.1 and $\Psi(\cdot)$ is defined in Assumption 6.3.*

This assumption is implied by any one of Assumptions N.C.4, N.C.5, and N.C.6 plus the following.

Assumption N.C.9. *$\mathbb{E}((v_t v_t')\otimes(w_t(\mathbf{S}_T)\, w_t(\mathbf{S}_T')')) = \Sigma_v \otimes \widetilde{Q}(\mathbf{S}_T,\mathbf{S}_T')$ for all $t\geq 1$ and $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$.*

By iterated expectations, a sufficient condition for Assumption N.C.9 is $\mathbb{E}(v_t v_t'|\, w_t(\mathbf{S}_T), w_t(\mathbf{S}_T')) = \mathbb{E}(v_t v_t') = \Sigma_v$ a.s. for all $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$ and all $t \geq 1$. Note that Assumptions N.C.6 and N.C.7 allow for intertemporal conditional heteroskedasticity even when Assumption N.C.9 holds. The following lemma summarizes the relations between the assumptions.

Lemma N.C.1. *(i) Any one of Assumptions N.C.4, N.C.5, N.C.6 and N.C.7 implies Assumptions 6.1-6.3;*
(ii) Any one of Assumptions N.C.4, N.C.5, N.C.6 plus Assumption N.C.9 imply Assumption N.C.8.

Proof of Lemma N.C.1. Although $w(\mathbf{S}_T)$ is a function of the partition $\mathbf{S}_T$, we do not need to rely on laws of large numbers for partial sum processes or functional central limit theorems. The reason is that Assumptions 6.1-6.3 and N.C.9 involve full-sample averages. Thus, the lemma follows from Lemma 4 in Andrews, Moreira, and Stock (2004). These authors required stationarity in their Assumptions INID, MDS and CORR but this is not required for the lemma to hold. $\square$

N.C.3 Consistent Covariance Matrix Estimation

Let $V_{b,t}(\mathbf{S}_T) = v_t' b_0 \overline{Z}_t(C_T)$ and $V_{a,t}(\mathbf{S}_T) = v_t' \Sigma_v^{-1} a_{0,\beta} \overline{Z}_t(C_T)$. We have

$$\Sigma_{N_1}(\mathbf{S}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{r=1}^{T} \mathbb{E}\left(V_{b,t}(\mathbf{S}_T) V_{b,r}(\mathbf{S}_T)'\right),$$
$$\Sigma_{N_1 N_2}(\mathbf{S}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{r=1}^{T} \mathbb{E}\left(V_{b,t}(\mathbf{S}_T) V_{a,r}(\mathbf{S}_T)'\right),$$
$$\Sigma_{N_2}^{*}(\mathbf{S}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{r=1}^{T} \mathbb{E}\left(V_{a,t}(\mathbf{S}_T) V_{a,r}(\mathbf{S}_T)'\right).$$

Let $\widehat{V}_{b,t}(\mathbf{S}_T) = \widehat{v}_t(\mathbf{S}_T)' b_0 \overline{Z}_t(C_T)$ and $\widehat{V}_{a,t}(s) = \widehat{v}_t(\mathbf{S}_T)' \widehat{\Sigma}_v^{-1} a_{0,\beta} \overline{Z}_t(C_T)$. We consider both HAC and double-kernel HAC (DK-HAC) estimators of $\Sigma_{v\overline{Z}}(\mathbf{S})$. Here we discuss the HAC estimators of Newey and West (1987) and Andrews (1991). The DK-HAC estimator was recently proposed by Casini (2023). It is consistent under both the null and the alternative so that tests based on it do not suffer from power losses induced by nonstationarity [cf. Casini, Deng, and Perron (2025)].

The HAC estimators are defined as

$$\widehat{\Sigma}_{N_1}(\mathbf{S}_T) = \frac{T}{T-q-p} \sum_{k=-T+1}^{T-1} K_1(b_{1,T} k) \widehat{\Gamma}_{bb}(k, \mathbf{S}_T),$$

$$\text{with} \qquad \widehat{\Gamma}_{bb}(k, \mathbf{S}_T) = \begin{cases} T^{-1} \sum_{t=k+1}^{T} \widehat{V}_{b,t}(\mathbf{S}_T) \widehat{V}_{b,t-k}'(\mathbf{S}_T), & k \geq 0 \\ T^{-1} \sum_{t=-k+1}^{T} \widehat{V}_{b,t+k}(\mathbf{S}_T) \widehat{V}_{b,t}'(\mathbf{S}_T), & k < 0 \end{cases},$$

and $\widehat{\Sigma}_{N_2}^{*}(\mathbf{S}_T)$ and $\widehat{\Sigma}_{N_1 N_2}(\mathbf{S}_T)$ are defined analogously to $\widehat{\Sigma}_{N_1}(\mathbf{S}_T)$ after replacing $\widehat{\Gamma}_{bb}(k, \mathbf{S}_T)$ with analogous quantities $\widehat{\Gamma}_{aa}(k, \mathbf{S}_T)$ and $\widehat{\Gamma}_{ab}(k, \mathbf{S}_T)$, respectively. We consider the following class of kernels

$$\boldsymbol{K}_1 = \Bigg\{ K_1(\cdot) : \mathbb{R} \to [-1,\, 1] : \, K_1(0) = 1, \, K_1(x) = K_1(-x), \, \forall x \in \mathbb{R}, \int_{-\infty}^{\infty} |K_1(x)| \, dx < \infty \tag{N.1}$$
$$\int_{-\infty}^{\infty} K_1^2(x) \, dx < \infty, \;\; K_1(\cdot) \text{ is continuous at 0 and at all but a finite number of points} \Bigg\}.$$

The class $\boldsymbol{K}_1$ was considered by Andrews (1991) and Casini (2023). Examples of kernels in $\boldsymbol{K}_1$ include the Truncated, Bartlett, Parzen, Quadratic Spectral (QS) and Tukey-Hanning kernels.

The DK-HAC estimators are defined as

$$\widehat{\Sigma}_{N_1}(\mathbf{S}_T) = \frac{T}{T-q-p} \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k) \, \widehat{\Gamma}_{\mathrm{DK}}(k, \, \mathbf{S}_T), \quad \text{with}$$
$$\widehat{\Gamma}_{\mathrm{DK}}(k, \, \mathbf{S}_T) = \frac{n_T}{T-n_T} \sum_{r=0}^{\lfloor (T-n_T)/n_T \rfloor} \widehat{c}_{bb}(rn_T/T, \, k, \, \mathbf{S}_T),$$

where $n_T \to \infty$ satisfies the conditions given below, and

$$\widehat{c}_{bb}(rn_T/T, \, k, \, \mathbf{S}_T) = \begin{cases} (Tb_{2,T})^{-1} \sum_{t=k+1}^{T} K_2\left(\frac{((r+1)n_T-(t-k/2))/T}{b_{2,T}}\right) \widehat{V}_{b,t}(\mathbf{S}_T) \, \widehat{V}'_{b,t-k}(\mathbf{S}_T), & k \geq 0 \\ (Tb_{2,T})^{-1} \sum_{t=-k+1}^{T} K_2\left(\frac{((r+1)n_T-(t+k/2))/T}{b_{2,T}}\right) \widehat{V}_{b,t+k}(\mathbf{S}_T) \, \widehat{V}'_{b,t}(\mathbf{S}_T), & k < 0 \end{cases}, \tag{N.2}$$

with K_2 being a kernel and $b_{2,T}$ is a bandwidth sequence. The DK-HAC estimators $\widehat{\Sigma}^*_{N_2}(\mathbf{S}_T)$ and $\widehat{\Sigma}_{N_1 N_2}(\mathbf{S}_T)$ are defined analogously to $\widehat{\Sigma}_{N_1}(\mathbf{S}_T)$ after replacing $\widehat{c}_{bb}(rn_T/T, \, k, \, \mathbf{S}_T)$ with analogous quantities $\widehat{c}_{aa}(rn_T/T, \, k, \, \mathbf{S}_T)$, and $\widehat{c}_{ab}(rn_T/T, \, k, \, \mathbf{S}_T)$, respectively. Casini (2023) considered the following class of kernels

$$\boldsymbol{K}_2 = \Bigg\{ K_2(\cdot) : \mathbb{R} \to [0, \, \infty] : \, K_2(x) = K_2(1-x), \int K_2(x) \, dx = 1, \tag{N.3}$$
$$K_2(x) = 0 \text{ for } x \notin [0, \, 1], \int_{-\infty}^{\infty} |K_2(x)| \, dx < \infty, \; K_2(\cdot) \text{ is continuous} \Bigg\}.$$

The QS kernel was shown to be optimal in the class $\boldsymbol{K}_1$ for HAC estimators under the mean-squared error (MSE) criterion by Andrews (1991) and for DK-HAC estimators under a sequential and global MSE criterion by Casini (2023) and Belotti, Casini, Catania, Grassi, and Perron (2023).

The QS kernel is defined as

$$K_1^{\mathrm{QS}}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

Casini (2023) showed that the optimal kernel in the class $\boldsymbol{K}_2$ is a quadratic-type kernel, $K_2^{\mathrm{opt}}(x) = 6x(1-x)$, $0 \leq x \leq 1$.

For both HAC and DK-HAC estimators, define

$$\begin{aligned}\widehat{\Sigma}_{v\overline{Z}}(\mathbf{S}_T) &= \begin{bmatrix}\widehat{\Sigma}_{N_1}(\mathbf{S}_T) & \widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T)' \\ \widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T) & \widehat{\Sigma}^*_{N_2}(\mathbf{S}_T)\end{bmatrix} \\ \widehat{\Sigma}_{N_2}(\mathbf{S}_T) &= \widehat{\Sigma}^*_{N_2}(\mathbf{S}_T) - \widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T)\widehat{\Sigma}^{-1}_{N_1}(\mathbf{S}_T)\widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T).\end{aligned}$$

We now provide sufficient conditions under which the HAC and DK-HAC estimators are uniformly consistent for $\Sigma_{v\overline{Z}}(\mathbf{S})$, and therefore $\Sigma_{N_2}(\mathbf{S})$, over $\mathbf{S} \in \mathcal{S}$. Let $V_t(\mathbf{S}_T) = v_t \otimes w_t(\mathbf{S}_T)$ where $w_t(\mathbf{S}_T)$ is the tth row of $w = [C_T Z : X]$ written as a column vector.

Assumption N.C.10. *((i) $\{V_t(\mathbf{S}_T)\}$ satisfies*

$$\sum_{j=-\infty}^{\infty} \sup_{t\geq 1}\sup_{\mathbf{S}_T\in\mathcal{S}} ||\mathbb{E}(V_t(\mathbf{S}_T)V'_{t-j}(\mathbf{S}_T))|| < \infty,$$

and for all conformable $a_1, a_2, a_3, a_4 \in \mathbb{Z}_+$, $\sum_{n=1}^{\infty}\sum_{j=1}^{\infty}\sum_{m=1}^{\infty}\sup_{t\geq 1}|\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(n, j, m, \mathbf{S}_T)| < \infty$ where $\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(n, j, m, \mathbf{S}_T)$ is the time-t fourth-order cumulant of

$$(V_t^{(a_1)}(\mathbf{S}_T), V_{t+n}^{(a_2)}(\mathbf{S}_T), V_{t+j}^{(a_3)}(\mathbf{S}_T), V_{t+m}^{(a_4)}(\mathbf{S}_T)).$$

(ii) $\sup_{t\geq 1}\sup_{\mathbf{S}_T\in\mathcal{S}}\mathbb{E}||V_t(\mathbf{S}_T)||^2 < \infty$.

Assumption N.C.11. *$b_{1,T} \to 0$ with $Tb_{1,T}^2 \to \infty$ and $K_1(\cdot) \in \boldsymbol{K}_1$.*

Assumption N.C.10 imposes conditions on the temporal dependence of the instruments and errors. It is a standard assumption in the literature, see Andrews (1991) and Casini (2023). Note that Assumption N.C.10 allows for nonstationary random variables (i.e., time-varying moments). The condition on the bandwidth in Assumption N.C.11 is from Andrews (1991). Assumptions N.C.10-N.C.11 are sufficient for the consistency of HAC estimators.

Assumption N.C.12. *$b_{1,T}, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$, $1/Tb_{1,T}b_{2,T} \to 0$, $\sqrt{T}b_{1,T} \to \infty$, $K_1(\cdot) \in \boldsymbol{K}_1$ and $K_2(\cdot) \in \boldsymbol{K}_2$.*

The conditions on the bandwidths $b_{1,T}$, $b_{2,T}$ and on n_T are from Casini (2023). Assumptions N.C.10 and N.C.12 are sufficient for the consistency of the DK-HAC estimators.

Lemma N.C.2. *Let Assumptions 6.1-6.3 hold. We have:* $\widehat{\Sigma}_{v\overline{Z}}(\mathbf{S}_T) \xrightarrow{\mathbb{P}} \Sigma_{v\overline{Z}}(\mathbf{S})$ *for* $\mathbf{S} = T^{-1}\mathbf{S}_T$ *uniformly in* $\mathbf{S}_T \in \mathcal{S}$ *under Assumptions N.C.10-N.C.11 for the HAC estimator and under Assumptions N.C.10 and N.C.12 for the DK-HAC estimator.*

The proof of Lemma N.C.2 is omitted. For the HAC estimator the proof follows from the discussion in Section 8 in Andrews (1991) who extended the consistency result in Theorem 1 in Andrews (1991) to nonstationary random variables. See also Casini (2022) who provided a solution to some issues in Section 8 in Andrews (1991). The proof for the DK-HAC estimator follows from Theorem 4.2 in Casini and Perron (2024). □

N.C.4 Strong IV and Local Alternative (SIV-LA) Asymptotics for Identification-Robust Tests

We analyze the strong IV asymptotic properties of the tests considered above for local alternatives. Under strong IV asymptotics, $\theta \neq 0$ is fixed. For local alternatives, β is local to β_0.

Assumption N.C.13. *(SIV-LA) (i)* $\beta = \beta_0 + r/T^{1/2}$ *for some constant* $r \in \mathbb{R}$*; (ii)* θ *is a fixed non-zero q-vector; (iii) There exists an estimator* $\widehat{\mathbf{S}}_T$ *such that* $T^{-1}\widehat{\mathbf{S}}_T \xrightarrow{\mathbb{P}} \mathbf{S}_0$.

Under strong IVs, part (iii) is satisfied by, for example, $\widehat{\mathbf{S}}_T$ in (6.5), $\widehat{\mathbf{S}}_{T,OLS}$ and $\widehat{\mathbf{S}}_{T,FGLS}$ where the optimization is over $\Xi_{\epsilon,\pi_0,m_0,T}$. Under SIV-LA asymptotics, $N_{1,T}(\mathbf{S}_T)$ and $N_{2,T}(\mathbf{S}_T)$ depend asymptotically on $\zeta_{N_1}(\mathbf{S}) \sim \mathscr{N}(\alpha_{N_1}(\mathbf{S}), I_q)$, $\alpha_{N_1}(\mathbf{S}) = \Sigma_{N_1}^{-1/2}(\mathbf{S})\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)\theta r$, and $\alpha_{N_2}(\mathbf{S}) = \Sigma_{N_2}^{-1/2}(\mathbf{S})\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)\theta\left(a_0'\Omega^{-1}a_0\right)^{-1/2}$.

We now determine the asymptotic distributions of the LR, LM and AR test statistics.

Theorem N.C.1. *Let Assumptions 6.1-6.4 and N.C.13 hold. We have: (i)* $AR_T(\widehat{\mathbf{S}}_T) \xrightarrow{d} \zeta_{N_1}(\mathbf{S}_0)'\zeta_{N_1}(\mathbf{S}_0) \sim \chi_q^2(\alpha_{N_1}(\mathbf{S}_0)'\alpha_{N_1}(\mathbf{S}_0))$*; (ii)* $LM_T(\widehat{\mathbf{S}}_T) \xrightarrow{d} (\alpha_{N_2}(\mathbf{S}_0)'\zeta_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2 \sim \chi_1^2((\alpha_{N_2}(\mathbf{S}_0)'\alpha_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2)$*; (iii)* $LR_T(\widehat{\mathbf{S}}_T) = LM_T(\mathbf{S}_0) + o_{\mathbb{P}}(1) \xrightarrow{d} (\alpha_{N_2}(\mathbf{S}_0)'\zeta_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2 \sim \chi_1^2((\alpha_{N_2}(\mathbf{S}_0)'\alpha_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2)$.

Since $T^{-1}\widehat{\mathbf{S}}_T \xrightarrow{\mathbb{P}} \mathbf{S}_0$ under strong IVs, the test statistics above are evaluated at $\mathbf{S}_0$ asymptotically. Akin to the case of known partition, the LM and LR test statistics are asymptotically equivalent under SIV-LA asymptotics for any value of q. When $q = 1$, $AR_T(\mathbf{S}_0)$, $LM_T(\mathbf{S}_0)$ and $LR_T(\mathbf{S}_0)$ are the same and so the three tests are asymptotically equivalent.

Under SIV-LA asymptotics and i.i.d. normal errors with unknown covariance matrix Σ_v and known $\mathbf{S}_0$, the model for y is a regular parametric model in the sense of standard likelihood theory. Hence, $LR_T(\mathbf{S}_0)$ and $LM_T(\mathbf{S}_0)$ are asymptotically efficient. This means they have standard large-sample optimality properties such as uniformly maximizing asymptotic power among asymptotically unbiased tests. Adapting the proof of Theorem 7 in Andrews, Moreira, and Stock (2006) while using $T^{-1}\widehat{\mathbf{S}}_T \xrightarrow{\mathbb{P}} \mathbf{S}_0$ it follows that $LM_T(\widehat{\mathbf{S}}_T)$ and $LR_T(\widehat{\mathbf{S}}_T)$ are asymptotically efficient under SIV-LA asymptotics and i.i.d. normal errors.

N.C.5 Strong IV and Fixed Alternative (SIV-FA) Asymptotics for Identification-Robust Tests

We now consider strong IV-fixed alternative (SIV-FA) asymptotics to determine the consistency, or lack thereof, of the tests.

Assumption N.C.14. *SIV-FA. (i) $\beta \neq \beta_0$ is fixed; (ii) θ is a fixed non-zero q-vector; (iii) There exists an estimator $\widehat{\mathbf{S}}_T$ such that $T^{-1}\widehat{\mathbf{S}}_T \xrightarrow{\mathbb{P}} \mathbf{S}_0$.*

Let $\Sigma_{\overline{Z}}(\mathbf{S}_0) = \Sigma_{\overline{Z}}(\mathbf{S}_0, \mathbf{S}_0)$. Define $\varphi_{N_1}(\mathbf{S}_0) = \Sigma_{N_1}^{-1/2}(\mathbf{S}_0)\,\Sigma_{\overline{Z}}(\mathbf{S}_0)\,\theta\,(\beta - \beta_0)$,

$$\varphi_{N_2}(\mathbf{S}_0) = \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\left(\Sigma_{\overline{Z}}(\mathbf{S}_0)\,\theta a_\beta' \Sigma_v^{-1} a_{0,\beta} - \Sigma_{N_1N_2}(\mathbf{S}_0,\, \mathbf{S}_0)\,\Sigma_{N_1}^{-1}\varphi_{N_1}(\mathbf{S}_0)\right), \quad \varsigma_q \sim \mathscr{N}(0,\, I_q)\,. \tag{N.4}$$

We now determine the asymptotic behavior of the test statistics under SIV-FA asymptotics.

Theorem N.C.2. *Let Assumptions 6.1-6.4 and N.C.14. We have: (i) $AR_T(\widehat{\mathbf{S}}_T)/T \xrightarrow{\mathbb{P}} \varphi_{N_1}(\mathbf{S}_0)'\,\varphi_{N_1}(\mathbf{S}_0) > 0$, (ii) $LM_T(\widehat{\mathbf{S}}_T)/T \xrightarrow{\mathbb{P}} (\varphi_{N_1}(\mathbf{S}_0)'\,\varphi_{N_2}(\mathbf{S}_0))^2/\varphi_{N_2}(\mathbf{S}_0)'\,\varphi_{N_2}(\mathbf{S}_0) > 0$ provided $\varphi_{N_2}(\mathbf{S}_0) \neq 0$; (iii)*

$$\begin{aligned} 2LR_T(\widehat{\mathbf{S}}_T)/T \xrightarrow{\mathbb{P}} & \ \varphi_{N_1}(\mathbf{S}_0)'\,\varphi_{N_1}(\mathbf{S}_0) - \varphi_{N_2}(\mathbf{S}_0)'\,\varphi_{N_2}(\mathbf{S}_0) \\ & - \sqrt{(\varphi_{N_1}(\mathbf{S}_0)'\,\varphi_{N_1}(\mathbf{S}_0) - \varphi_{N_1}(\mathbf{S}_0)'\,\varphi_{N_1}(\mathbf{S}_0))^2 - 4\left(\varphi_{N_1}(\mathbf{S}_0)'\,\varphi_{N_2}(\mathbf{S}_0)\right)^2}. \end{aligned}$$

The theorem shows that the test $AR_T(\widehat{\mathbf{S}}_T)$ is consistent against any alternative $\beta \neq \beta_0$, $LM_T(\widehat{\mathbf{S}}_T)$ is consistent against any alternative $\beta \neq \beta_0$ such that $\varphi_{N_2}(\mathbf{S}_0) \neq 0$, and $LR_T(\widehat{\mathbf{S}}_T)$ is consistent against any alternative for which the limit value given in the theorem is non-zero.

N.C.6 Proofs of Section N.C.4 and N.C.5

N.C.6.1 Proof of Theorem N.C.1

We begin with the following lemma.

Lemma N.C.3. *Let Assumptions 6.1-6.4 and N.C.13 hold. We have: (i)* $(N_{1,T}(\widehat{\mathbf{S}}_T),\ T^{-1/2}N_{2,T}(\widehat{\mathbf{S}}_T)) \xrightarrow{d} (\zeta_{N_1}(\mathbf{S}_0),\ \alpha_{N_2}(\mathbf{S}_0))$ *and (ii)* $(M_{1,T}(\widehat{\mathbf{S}}_T),\ T^{-1/2}M_{1,2,T}(\widehat{\mathbf{S}}_T),\ T^{-1}M_{2,T}(\widehat{\mathbf{S}}_T)) \xrightarrow{d} (\zeta_{N_1}(\mathbf{S}_0)'\zeta_{N_1}(\mathbf{S}_0),\ \alpha_{N_2}(\mathbf{S}_0)'\zeta_{N_1}(\mathbf{S}_0),\ \alpha_{N_2}(\mathbf{S}_0)'\alpha_{N_2}(\mathbf{S}_0))$.

Proof of Lemma N.C.3. Part (i) for $N_{1,T}(\widehat{\mathbf{S}}_T)$ follows from

$$
\begin{aligned}
N_{1,T}\left(\widehat{\mathbf{S}}_T\right) &= \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1/2}\overline{Z}\left(\widehat{C}_T\right)' yb_0 \\
&= \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1/2}\overline{Z}\left(\widehat{C}_T\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a_\beta' + v\right)b_0 \\
&= \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1}\overline{Z}\left(\widehat{C}_T\right)'\overline{Z}\left(C_{0,T}\right)\theta r + \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1/2}\overline{Z}\left(\widehat{C}_T\right)' vb_0 \\
&= \widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)T^{-1}\overline{Z}\left(C_{0,T}\right)'\overline{Z}\left(C_{0,T}\right)\theta r + \widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)T^{-1/2}\overline{Z}\left(C_{0,T}\right)' vb_0 + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta r + \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)T^{-1/2}\overline{Z}\left(C_{0,T}\right)' vb_0 + o_{\mathbb{P}}(1) \\
&\Rightarrow \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta r + \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\left[I_q : -Q_{12}\left(\mathbf{S}_0\right)Q_{22}^{-1}\right]\left(b_0'\otimes I_{q+p}\right)\mathscr{G}\left(\mathbf{S}_0\right) \\
&\sim \zeta_{N_1}\left(\mathbf{S}_0\right),
\end{aligned}
$$

where the third and fourth equalities hold by Assumption N.C.13, the final equality holds by Assumptions 6.1 and 6.4 and the convergence holds by Assumption 6.3. Under Assumptions 6.2 and N.C.13(iii), $\widehat{\Sigma}_v(\widehat{\mathbf{S}}_T) \xrightarrow{\mathbb{P}} \Sigma_v$ by the same arguments as when $\mathbf{S}_0$ is known. Then, part (i) for $N_{2,T}(\widehat{\mathbf{S}}_T)$ holds using

$$
T^{-1/2}N_{2,T}\left(\widehat{\mathbf{S}}_T\right) = \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\widehat{\Sigma}_v^{-1}\left(\widehat{\mathbf{S}}_T\right)a_{0,\beta} - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) \tag{N.5}
$$

$$
\begin{aligned}
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\Sigma_v^{-1}a_{0,\beta} - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) + o_{\mathbb{P}}(1) \\
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_{0,\beta} - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\mathbf{S}_0\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)N_{1,T}\left(\mathbf{S}_0\right)\right) + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_{0,\beta} - T^{-1/2}\Sigma_{N_1N_2}\left(\mathbf{S}_0\right)\Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)N_{1,T}\left(\mathbf{S}_0\right)\right) + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a_\beta' + v\right)\Sigma_v^{-1}a_{0,\beta}\right) + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta a_{0,\beta}'\Sigma_v^{-1}a_{0,\beta} + o_{\mathbb{P}}(1),
\end{aligned}
$$

where the third equality follows from Assumption N.C.13(iii), the fourth by Assumption 6.4, the fifth by part (i) for $N_{1,T}(\widehat{\mathbf{S}}_T)$ and the final equality holds by Assumption 6.1. Part (ii) holds by part (i) and the continuous mapping theorem. □

Proof of Theorem N.C.1. Parts (i) and (ii) of the theorem follow immediately from Lemma N.C.3(ii). Part (iii) of the theorem is established as follows. Following the argument based on a mean-value expansion of the LR_T statistic in the proof of Theorem 9 in Andrews, Moreira, and Stock (2004) (see eq. (14.50)-(14.53)) with references to Lemma 9-(b) there replaced by references to Lemma N.C.3, we have

$$\begin{aligned} LR_T(\widehat{\mathbf{S}}_T) &= \frac{1}{2}\left(2M_{1,T}\left(\widehat{\mathbf{S}}_T\right) - 2\left(M_{1,T}\left(\widehat{\mathbf{S}}_T\right) - M_{1,2,T}\left(\widehat{\mathbf{S}}_T\right)^2 / M_{2,T}\left(\widehat{\mathbf{S}}_T\right)\right)\right) + o_{\mathbb{P}}(1) \\ &= M_{1,2,T}^2\left(\mathbf{S}_0\right) / M_{2,T}\left(\mathbf{S}_0\right) + o_{\mathbb{P}}(1) \\ &= LM_T(\mathbf{S}_0) + o_{\mathbb{P}}(1), \end{aligned} \tag{N.6}$$

where we used Assumption N.C.13(iii). □

N.C.6.2 Proof of Theorem N.C.2

We begin with the following lemma.

Lemma N.C.4. *(i) Under Assumptions 6.1-6.4 and N.C.14, (i)* $(N_{1,T}(\widehat{\mathbf{S}}_T)/T^{1/2},\, N_{2,T}(\widehat{\mathbf{S}}_T)/T^{1/2}) \xrightarrow{\mathbb{P}} (\varphi_{N_1}(\mathbf{S}_0), \varphi_{N_2}(\mathbf{S}_0))$ *and (ii)* $(M_{1,T}(\widehat{\mathbf{S}}_T)/T,\, M_{1,2,T}(\widehat{\mathbf{S}}_T)/T,\, M_{2,T}(\widehat{\mathbf{S}}_T)/T) \xrightarrow{\mathbb{P}} (\varphi_{N_1}(\mathbf{S}_0)'\varphi_{N_1}(\mathbf{S}_0),\, \varphi_{N_1}(\mathbf{S}_0)' \varphi_{N_2}(\mathbf{S}_0),\, \varphi_{N_2}(\mathbf{S}_0)'\varphi_{N_2}(\mathbf{S}_0))$.

Proof of Lemma N.C.4. Part (i) of the lemma is established as follows:

$$\begin{aligned} T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y b_0 &= T^{-1}\overline{Z}\left(\widehat{C}_T\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a_\beta' + X\eta + v\right) b_0 \\ &= T^{-1}\overline{Z}\left(\widehat{C}_T\right)'\overline{Z}\left(C_{0,T}\right)\theta a_\beta' b_0 + T^{-1}\overline{Z}\left(\widehat{C}_T\right)' v b_0 \xrightarrow{\mathbb{P}} \Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta a_\beta' b_0, \end{aligned}$$

using Assumptions 6.1, 6.3, N.C.14(iii) and $\overline{Z}(\widehat{C}_T)'X = 0$. Hence, by Assumptions 6.1, 6.4 and

N.C.14(iii), we have

$$
\begin{aligned}
N_{1,T}\left(\widehat{\mathbf{S}}_T\right)/T^{1/2} &= \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1}\overline{Z}\left(\widehat{C}_T\right)' yb_0 \qquad (\text{N.7})\\
&= \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1}\overline{Z}\left(\widehat{C}_T\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a' + v\right)b_0\\
&= \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1}\overline{Z}\left(\widehat{C}_T\right)'\overline{Z}\left(C_{0,T}\right)\theta a'_\beta b_0 + \widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)T^{-1}\overline{Z}\left(\widehat{C}_T\right)' vb_0\\
&= \widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)T^{-1}\overline{Z}\left(C_{0,T}\right)'\overline{Z}\left(C_{0,T}\right)\theta a'_\beta b_0 + \widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)T^{-1}\overline{Z}\left(C_{0,T}\right)' vb_0 + o_{\mathbb{P}}\left(1\right)\\
&= \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta a'_\beta b_0 + \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)T^{-1}\overline{Z}\left(C_{0,T}\right)' vb_0 + o_{\mathbb{P}}\left(1\right)\\
&= \Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta\left(\beta-\beta_0\right) + o_{\mathbb{P}}\left(1\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
T^{-1/2}N_{2,T}\left(\widehat{\mathbf{S}}_T\right) &= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\widehat{\Sigma}_v^{-1}\left(\widehat{\mathbf{S}}_T\right)a_{0,\beta} - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) \qquad (\text{N.8})\\
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\Sigma_v^{-1}a_{0,\beta} - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) + o_{\mathbb{P}}\left(1\right)\\
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_{0,\beta} - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\mathbf{S}_0\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)N_{1,T}\left(\mathbf{S}_0\right)\right) + o_{\mathbb{P}}\left(1\right)\\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_{0,\beta} - T^{-1/2}\Sigma_{N_1N_2}\left(\mathbf{S}_0\right)\Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)N_{1,T}\left(\mathbf{S}_0\right)\right) + o_{\mathbb{P}}\left(1\right)\\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a'_\beta + v\right)\Sigma_v^{-1}a_{0,\beta}\right)\\
&\quad - \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{N_1N_2}\left(\mathbf{S}_0\right)\Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\varphi_{N_1}\left(\mathbf{S}_0\right) + o_{\mathbb{P}}\left(1\right)\\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta a'_\beta\Sigma_v^{-1}a_{0,\beta} - \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{N_1N_2}\left(\mathbf{S}_0\right)\Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)\varphi_{N_1}\left(\mathbf{S}_0\right) + o_{\mathbb{P}}\left(1\right)\\
&= \varphi_{N_2}\left(\mathbf{S}_0\right) + o_{\mathbb{P}}\left(1\right).
\end{aligned}
$$

Part (ii) of the lemma follows from part (i) and Slutsky's Theorem. □

Proof of Theorem N.C.2. Parts (i)-(iii) of the theorem hold by Lemma N.C.4 and simple calculations. In the case of $LM_T(\widehat{\mathbf{S}}_T)$, the convergence only holds if β is such that $\varphi_{N_2}(\mathbf{S}_0) \neq 0$ because $\varphi_{N_2}(\mathbf{S}_0)$ appears in the denominator. □

N.D Additional Monte Carlo Simulations

We consider the performance of the identification-robust tests under serial correlation in the errors. Specifically, we examine DGP (S.C.2)–(S.C.3) and model the error terms as $u_t = \rho_u u_{t-1} + v_{u,t}$ and $e_t = \rho_e e_{t-1} + v_{e,t}$, where $v_{u,t}$ and $v_{e,t}$ are jointly normally distributed with mean zero and covariance matrix Σ_{ue} as in (S.C.4) with $\rho \in \{0,\, 0.25,\, 0.5,\, 0.75\}$ and $\rho_e = \rho_u \in \{0.25,\, 0.5,\, 0.75\}$. We set the significance level to 5% and number of Monte Carlo replications to $10,000$. Table 2 and Figure 1 report the null rejection frequencies and size-adjusted power of the tests, respectively. Under strong serial dependence ($\rho_e = \rho_u = 0.75$) all tests exhibit rejection rates that exceed the nominal significance level. Specifically, $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ are a bit more oversized than LM_T and CLR_T but similar to qLL-S. Under weak serial dependence ($\rho_e = \rho_u = 0.25$), the proposed tests $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ are only slightly more oversized than their full sample counterparts, LM_T and CLR_T. Figure 1 shows that the size-adjusted power of the proposed tests is higher than that of the existing tests, similar to the i.i.d. case.

Finally, we consider a model with multiple instruments:

$$Y_t = \beta D_t + u_t, \tag{N.1}$$

where

$$D_t = \begin{cases} \theta_1 \left(Z_{1,t} + Z_{2,t}\right) + \widetilde{\theta}_1 Z_{3,t} + e_t, & t \leq \lfloor T/4 \rfloor \\ \theta_2 \left(Z_{1,t} + Z_{2,t} + Z_{3,t}\right) + e_t, & \lfloor T/4 \rfloor + 1 \leq t \leq \lfloor T/4 \rfloor + \lfloor (1-\pi_0)\, T \rfloor \\ \theta_3 \left(Z_{1,t} + Z_{2,t}\right) + \widetilde{\theta}_3 Z_{3,t} + e_t, & \lfloor T/4 \rfloor + \lfloor (1-\pi_0)\, T \rfloor + 1 \leq t \leq T, \end{cases} \tag{N.2}$$

$Z_{i,t} \sim$ i.i.d. $\mathscr{N}(1,\, 1)$ for $i = 1,\, 2,\, 3$, and u_t and e_t are i.i.d. jointly normal with mean zero and covariance

$$\Sigma_{ue} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{N.3}$$

with $\rho \in \{0.25,\, 0.75\}$. Under the null hypothesis we set $\theta_1 = \theta_2 = \theta_3 = \widetilde{\theta}_1 = \widetilde{\theta}_3 = 0$. Under the alternative hypothesis we set $\theta_1 = \theta_3 = dT^{-1/2}$ with dß$\{2,\, 4,\, 8\}$, $\theta_2 = 0$ and $\widetilde{\theta}_1 = \widetilde{\theta}_3 = 16/\sqrt{T}$. We set $\pi_0 \in \{0.6,\, 0.8\}$ and $T = 200$.

Table 3 reports the null rejection frequencies. The AR_T, $AR_T(\widehat{\mathbf{S}}_T)$, Split-S, qLL-S, ave-S and exp-S are severely undersized across all values of d and ρ. LM_T, $LM_T(\widehat{\mathbf{S}}_T)$, CLR_T, $CLR_T(\widehat{\mathbf{S}}_T)$ lead to quite accurate null rejection rates whereas Split-CLR displays null rejection rates substantially beyond the nominal level. Figure 2 plots the power functions. $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ are the most powerful, followed by $AR_T(\widehat{\mathbf{S}}_T)$ and then by the full sample counterparts of these tests. qLL-S displays the lowest power across all configurations. The power gains of $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ are substantial across all configurations.

Table 2: Finite-Sample Null Rejection Frequencies of Tests

$\rho = 0.50$	$\rho_e = \rho_u = 0.25$			$\rho_e = \rho_u = 0.50$			$\rho_e = \rho_u = 0.75$		
$T = 200$, $\pi_0 = 0.8$	$d = 10$	$d = 16$	$d = 24$	$d = 10$	$d = 16$	$d = 24$	$d = 10$	$d = 16$	$d = 24$
LM_T	0.073	0.073	0.073	0.089	0.089	0.089	0.158	0.158	0.158
CLR_T	0.079	0.077	0.076	0.096	0.094	0.093	0.179	0.169	0.166
$LM_T(\widehat{\mathbf{S}}_T)$	0.086	0.081	0.078	0.118	0.107	0.101	0.204	0.185	0.176
$CLR_T(\widehat{\mathbf{S}}_T)$	0.088	0.082	0.078	0.126	0.108	0.101	0.214	0.191	0.179
$\mathrm{split-S}$	0.054	0.055	0.055	0.071	0.074	0.076	0.139	0.148	0.155
$\mathrm{split-CLR}$	0.146	0.149	0.150	0.174	0.186	0.187	0.278	0.291	0.300
$\mathrm{qqL-S}$	0.028	0.028	0.028	0.047	0.052	0.052	0.191	0.198	0.199
$\mathrm{ave-S}$	0.047	0.046	0.047	0.068	0.071	0.070	0.153	0.161	0.171
$\mathrm{exp-S}$	0.019	0.020	0.020	0.034	0.034	0.034	0.109	0.109	0.107
$\rho = 0.50$	$\rho_e = \rho_u = 0.25$			$\rho_e = \rho_u = 0.50$			$\rho_e = \rho_u = 0.75$		
$T = 400$, $\pi_0 = 0.6$	$d = 10$	$d = 16$	$d = 24$	$d = 10$	$d = 16$	$d = 24$	$d = 10$	$d = 16$	$d = 24$
LM_T	0.061	0.061	0.061	0.074	0.074	0.074	0.124	0.124	0.124
CLR_T	0.070	0.067	0.067	0.088	0.083	0.083	0.153	0.142	0.136
$LM_T(\widehat{\mathbf{S}}_T)$	0.076	0.068	0.068	0.109	0.091	0.091	0.174	0.158	0.142
$CLR_T(\widehat{\mathbf{S}}_T)$	0.081	0.070	0.069	0.106	0.093	0.093	0.190	0.166	0.145
$\mathrm{split-S}$	0.046	0.044	0.043	0.060	0.059	0.059	0.113	0.113	0.113
$\mathrm{split-CLR}$	0.133	0.134	0.134	0.150	0.154	0.154	0.234	0.240	0.243
$\mathrm{qqL-S}$	0.042	0.043	0.039	0.042	0.063	0.063	0.169	0.177	0.180
$\mathrm{ave-S}$	0.048	0.050	0.048	0.048	0.073	0.068	0.130	0.133	0.137
$\mathrm{exp-S}$	0.0243	0.023	0.024	0.023	0.034	0.037	0.098	0.098	0.095

Table 3: Finite-Sample Null Rejection Frequencies of Tests for the model (N.1)-(N.2)

	$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
$T = 200$, $\pi_0 = 0.6$	$d = 2$	$d = 4$	$d = 8$	$d = 2$	$d = 4$	$d = 8$	$d = 2$	$d = 4$	$d = 8$
AR_T	0.000	0.000	0.001	0.008	0.000	0.001	0.001	0.000	0.001
LM_T	0.063	0.062	0.061	0.059	0.064	0.062	0.060	0.069	0.060
CLR_T	0.070	0.070	0.056	0.070	0.073	0.068	0.066	0.076	0.067
$AR_T(\widehat{\mathbf{S}}_T)$	0.002	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001
$LM_T(\widehat{\mathbf{S}}_T)$	0.074	0.067	0.072	0.067	0.064	0.073	0.064	0.072	0.070
$CLR_T(\widehat{\mathbf{S}}_T)$	0.075	0.068	0.071	0.067	0.073	0.073	0.065	0.072	0.070
$\mathrm{split}-\mathrm{S}$	0.015	0.015	0.015	0.015	0.015	0.016	0.016	0.015	0.016
$\mathrm{split}-\mathrm{CLR}$	0.096	0.090	0.086	0.094	0.092	0.082	0.091	0.094	0.087
$\mathrm{qqL}-\mathrm{S}$	0.008	0.009	0.008	0.008	0.009	0.010	0.010	0.009	0.009
$\mathrm{ave}-\mathrm{S}$	0.020	0.025	0.023	0.020	0.024	0.023	0.021	0.027	0.023
$\mathrm{exp}-\mathrm{S}$	0.004	0.034	0.005	0.005	0.004	0.004	0.005	0.004	0.004
	$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
$T = 200$, $\pi_0 = 0.8$	$d = 2$	$d = 4$	$d = 8$	$d = 2$	$d = 4$	$d = 8$	$d = 2$	$d = 4$	$d = 8$
AR_T	0.000	0.000	0.000	0.001	0.004	0.001	0.001	0.000	0.001
LM_T	0.061	0.063	0.058	0.056	0.066	0.059	0.060	0.066	0.059
CLR_T	0.048	0.050	0.062	0.059	0.070	0.061	0.063	0.068	0.062
$AR_T(\widehat{\mathbf{S}}_T)$	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.004	0.002
$LM_T(\widehat{\mathbf{S}}_T)$	0.064	0.061	0.065	0.067	0.072	0.067	0.070	0.079	0.055
$CLR_T(\widehat{\mathbf{S}}_T)$	0.064	0.062	0.065	0.067	0.073	0.068	0.069	0.079	0.054
$\mathrm{split}-\mathrm{S}$	0.020	0.017	0.021	0.022	0.019	0.021	0.018	0.018	0.020
$\mathrm{split}-\mathrm{CLR}$	0.102	0.098	0.086	0.103	0.100	0.098	0.103	0.108	0.098
$\mathrm{qqL}-\mathrm{S}$	0.009	0.074	0.009	0.010	0.008	0.011	0.010	0.008	0.009
$\mathrm{ave}-\mathrm{S}$	0.020	0.022	0.018	0.023	0.022	0.019	0.019	0.025	0.018
$\mathrm{exp}-\mathrm{S}$	0.004	0.026	0.004	0.005	0.003	0.005	0.005	0.005	0.004

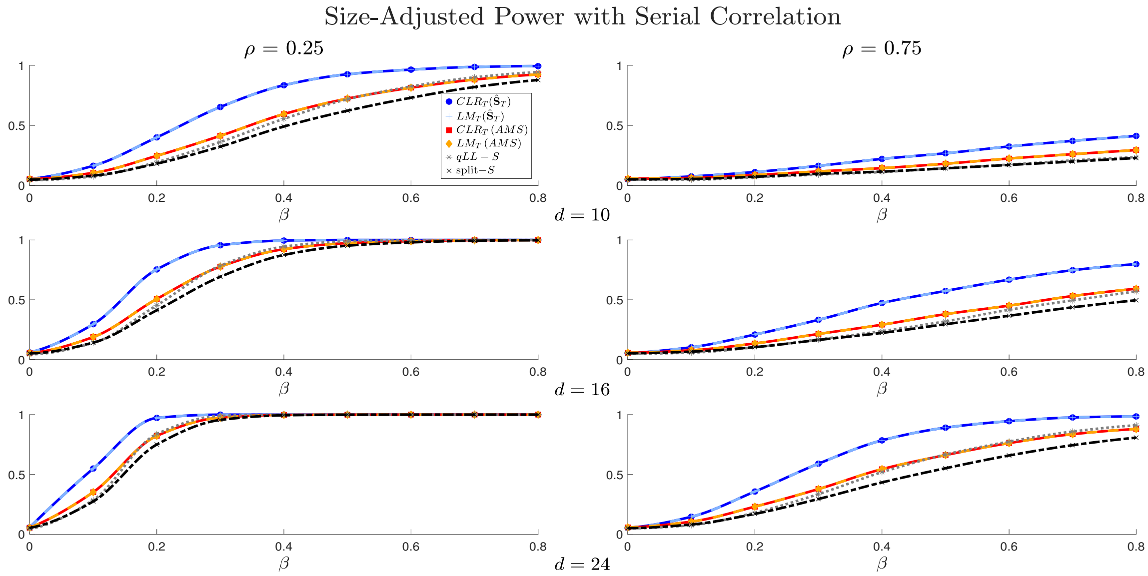

Figure 1: Size-adjusted power of identification robust tests for $T = 400$ and $\pi_0 = 0.6$ and $\rho = 0.25$.

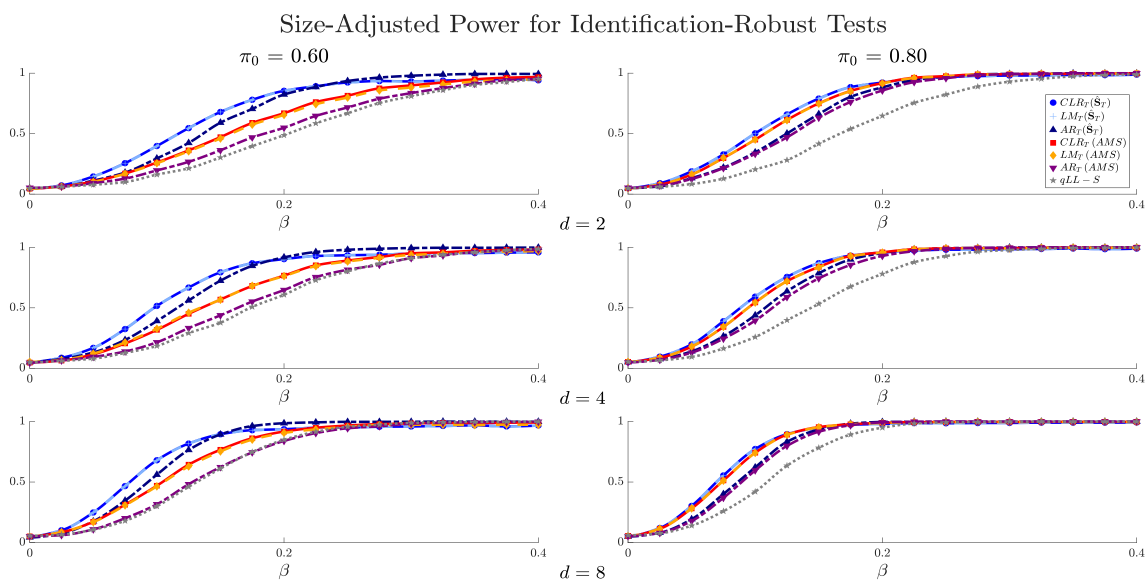

Figure 2: Size-adjusted power of identification robust tests for $T = 200$ for model in (N.1)-in(N.2).

References

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Andrews, D. W. K., M. Moreira, and J. H. Stock (2004): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," Discussion Paper 1476, Cowles Foundation, Yale University.

——— (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74(3), 715–752.

Belotti, F., A. Casini, L. Catania, S. Grassi, and P. Perron (2023): "Simultaneous Bandwidths Determination for Double-Kernel HAC Estimators and Long-Run Variance Estimation in Nonparametric Settings," *Econometric Reviews*, 42(3), 281–306.

Casini, A. (2022): "Comment on Andrews (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation"," *Econometrica*, 90(4), 1–2.

——— (2023): "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models," *Journal of Econometrics*, 235(2), 372–392.

——— (2024): "The Fixed-b Limiting Distribution and the ERP of HAR Tests Under Nonstationarity," *Journal of Econometrics*, 238(2), 105625.

Casini, A., T. Deng, and P. Perron (2025): "Theory of Low Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference," *Econometric Theory*, forthcoming.

Casini, A., and P. Perron (2024): "Prewhitened Long-Run Variance Estimation Robust to Nonstationarity," *Journal of Econometrics*, 242(1), 105794.

Newey, W. K., and K. D. West (1987): "A Simple Positive Semidefinite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.