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

16th September 2025

Abstract

This paper discusses identification, estimation, and inference on dynamic local average treatment effects (LATEs) in instrumental variables (IVs) settings. First, we show that compliers—observations whose treatment status is affected by the instrument—can be identified *individually* in time series data using smoothness assumptions and local comparisons of treatment assignments. Second, we show that this result enables not only better interpretability of IV estimates but also direct testing of the exclusion restriction by comparing outcomes among identified non-compliers across instrument values. Third, we document pervasive weak identification in applied work using IVs with time series data by surveying recent publications in leading economics journals. However, we find that strong identification often holds in large subsamples for which the instrument induces changes in the treatment. Motivated by this, we introduce a method based on dynamic programming to detect the most strongly-identified subsample and show how to use this subsample to improve estimation and inference. We also develop new identification-robust inference procedures that focus on the most strongly-identified subsample, offering efficiency gains relative to existing full sample identification-robust inference when identification fails over parts of the sample. Finally, we apply our results to heteroskedasticity-based identification of monetary policy effects. We find that about 75% of observations are compliers (i.e., cases where the variance of the policy shifts up on FOMC announcement days), and we fail to reject the exclusion restriction. Estimation using the most strongly-identified subsample helps reconcile conflicting IV and GMM estimates in the literature.

JEL Classification: B41, C12, C32.

Keywords: Compliers, Conditional inference, Exclusion, LATE.

*Casini acknowledges financial support from EIEF through 2022 EIEF Grant. McCloskey acknowledges support from the National Science Foundation under Grant SES-2341730.

1 Introduction

Economists work hard to extract plausibly exogenous variation in order to identify causal effects. Many identification strategies used in applied work either rely directly on instrumental variables (IVs) or can be reframed in terms of IV identification. This holds also in dynamic settings where, for example, external IVs may be constructed using a narrative approach or heteroskedasticity is exploited to yield additional identifying equations. Since Imbens and Angrist (1994), it has been well-known that IV-based approaches identify the local average treatment effect (LATE)—the average treatment effect for the sub-population of compliers, i.e., those whose treatment status is influenced by the policy intervention (the instrument).

In the LATE framework, the sub-population of compliers is unobserved. This means that although a LATE can be identified, the specific sample observations this effect represents is unknown. This limitation is often described informally as the inability to observe an observation's treatment status under both the intervention and non-intervention scenarios. From a practical interpretability perspective, this presents a challenge that has been widely discussed in the literature [see, e.g., Angrist, Imbens, and Rubin (1996), Heckman (1996), Imbens (2010) and Robins and Greenland (1996)]. Some progress has been made by Imbens and Rubin (1997) and Abadie (2003) who show that the proportion of compliers and some of their statistical characteristics can be identified, provided these characteristics can be expressed as functions of moments of the joint distribution of observed data. Using these results, Bhuller, Dahl, Løken, and Mogstad (2020) conduct a detailed analysis of compliers in the context of interpreting IV estimates of the effect of incarceration on recidivism and subsequent labor market outcomes. Their work, along with many other studies, highlights the importance of identifying the (characteristics of) compliers when drawing policy implications.

This paper considers IV identification in dynamic settings and shows how compliers can be identified individually in this context. We first show that the notion of compliers can be equivalently rewritten in terms of an inequality involving the difference in means of the potential treatment under different instrument values. Under assumptions of continuity over time in the mean of the potential treatment assignment process—conditional on a fixed hypothetical value of the instrument—it is possible to recover counterfactual values by averaging observations in a neighborhood around a given time point.

For example, consider heteroskedasticity-based identification of the causal effects of monetary policy [cf. Rigobon (2003) and Nakamura and Steinsson (2018)] where the instrument indicates whether there was an FOMC announcement on each date in the sample and the

treatment variable is equal to the variance of a short-term interest rate. Here compliers are defined as observations for which the volatility of the policy variable (change in short-term interest rate) increases if and only if there is an FOMC announcement. Suppose that there is an FOMC announcement on a given date of interest so that we do not observe the potential treatment assignment under the counterfactual instrument value indicating the absence of an announcement. Although the mean treatment assignment under non-announcement is unobserved at this date, it can be recovered if mean treatment assignments are a smooth function of time by computing an average of nearby days without an announcement. Under an additional assumption of deterministic complier status, the complier status of the date in question can be estimated and tested by comparing local means of the treatment variable, one corresponding to nearby dates for which an announcement occurred and the other corresponding to nearby dates for which it did not.[1] Applying our identification results and tests to the heteroskedasticity-based identification of monetary policy effects, we find that about 75% of observations are compliers while the non-compliers are primarily concentrated in the early zero lower bound period, when the central bank could no longer lower interest rates and forward guidance was not aggressive.

Identification of compliers is not only valuable in its own right. It also enables us to test the exclusion restriction, a key condition for valid IV estimation that is typically untestable in practice. By identifying compliers, and thus also non-compliers, we show that the exclusion restriction can be tested using a t-test that compares the average outcomes of non-compliers across different instrument values.

A key condition for identification of the LATE in the IV framework is instrument relevance, entailing nontrivial correlation between the endogenous variable and the instrument. We begin by analyzing the problem of weak instruments, entailing low correlation between the endogenous variable and instrument, in empirical work through a survey of articles using IVs published from 2019 to 2022 in five leading journals: American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies. Our sample includes 1,560 specifications from 18 papers, with 199 involving time series and 1361 involving panel data.[2] The left panels of Figure 1 show histograms of full sample first-stage F-statistics for the specifications in our survey, truncated

[1]Even though we focus on a time series setting, our identification results immediately apply to cross-sectional settings with spatial data provided that the temporal distance between observations is interpreted as geographical distance, and analogous continuity assumptions are imposed over space.

[2]See the supplement for the full list of papers and inclusion criteria.

above 100 for visibility. Many F-statistics concentrate around the χ^2_1 critical values and fall below the conventional thresholds of 10 and 23.1 suggested by Staiger and Stock (1997) and Montiel Olea and Pflueger (2013), raising serious concerns about weak instruments.[3] These findings align with those of Andrews, Stock, and Sun (2019), who analyze cross-sectional studies. For example, we find that 75% of time series and 72% of panel data specifications have first stage F-statistics below 24. The median F-statistic is 12.63 for time series and 9.29 for panel data.[4]

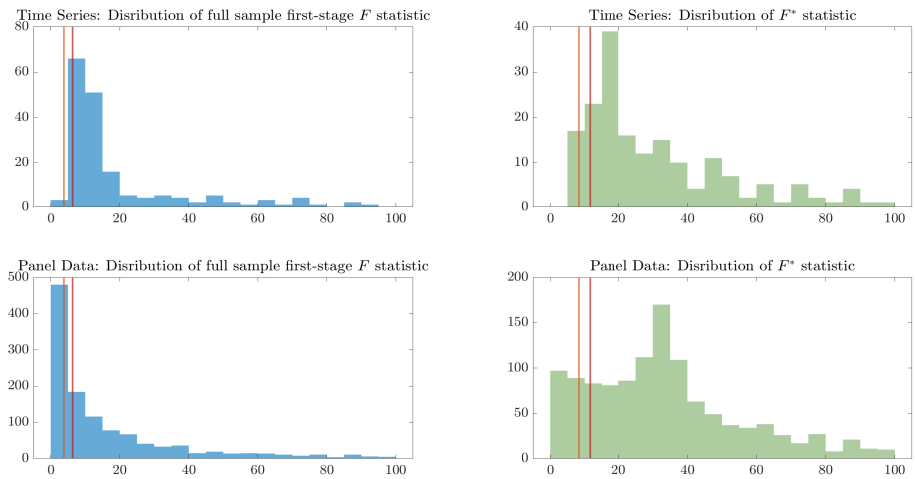

Figure 1: Distributions of the first-stage F (left panels) and F^* statistics (right panels). The top panels apply to time series specifications and the bottom panels apply to panel data specifications. The orange and red vertical lines correspond to the 5% and 1% level asymptotic critical values of the first-stage F (χ^2_1 for left panels) and F^* statistics (8.28 and 11.63) for right panels) under identification failure.

When identification fails or is weak, IV estimators can be severely biased for LATEs and conventional inference methods are rendered invalid. These problems have prompted extensive research on detecting weak instruments and constructing identification-robust confidence sets.[5] However, there has been little work on estimation and inference in a general LATE setting when identification may be stronger over subsamples. The second main contribution

[3]Indeed, Staiger and Stock (1997) derive the threshold of 10 under the homoskedasticity-only assumption—the relevant thresholds for time series data are larger [see Montiel Olea and Pflueger (2013)].

[4]For panel data specifications we consider each cross-sectional unit individually to enable comparison to our proposed time series test shown on the right panels.

[5]See, e.g., Andrews et al. (2006), Kleibergen (2002), Moreira (2003) and Staiger and Stock (1997).

of this paper is to develop a framework for identification, estimation, and inference on LATEs that accommodates time-varying instrument relevance. Within this framework, we propose a first-stage F-test to detect whether identification fails over all nontrivial subsamples. To solve the computationally intensive problem of searching for maximal identification strength among all possible sample partitions, we employ dynamic programming. This optimization is more complex than that in the structural break literature since evaluating identification strength requires more than comparing parameter changes across regimes.

In an attempt to understand the sources of weak IVs we plot the histograms of the F^* statistic proposed in this paper (cf. Section 4) in the right panels of Figure 1. The statistic F^* searches for the subsample with maximal identification strength among all possible subsamples of size at least $\pi_L T$.[6] The idea is that while the IVs may appear weak in the full sample, they may be strong in a possibly large subsample. Figure 1 shows that this is indeed often the case. The red vertical lines in Figure 1 mark the 95th percentile of the asymptotic distributions of the F and F^* statistics under the null of identification failure. Although its quantiles are larger, the F^* statistics have substantially more mass in the upper quantiles of their null distribution. This has at least three implications. First, it confirms substantial time variation in the instruments' strength. Second, strong identification appears to be frequently present in a sizeable subsample even when the instruments appear weak in the full sample. The median F^* is 27.22 for time series and 33.81 for panel data specifications. These are substantially higher than their full sample counterparts and this difference cannot be simply attributed to the different null asymptotic distributions of the two test statistics given that the difference in the asymptotic critical values is relatively small while the empirical distributions of the two test statistics is markedly different. About one half of the specifications that appear to suffer from weak IVs in the full sample seem better characterized by strong IVs in the subsample with maximal identification strength. Third, the subsamples where instruments appear strong tend to be large. From an empirical perspective, this is encouraging: although weak instruments in the full sample are common, researchers can often succeed in identifying large subsamples where instruments appear strong.

Motivated by this survey evidence, we construct consistent estimators of LATEs when subsamples are strongly-identified. It is commonly believed that if IVs are strong only in some portion of the sample, the full sample IV estimator remains consistent for a LATE. We show that this belief is unwarranted unless the LATE of interest is time-invariant (i.e.,

[6] We set $\pi_L = 0.6$ in Figure 1. We discuss the choice of π_L below.

homogeneous). If this condition fails, one can at best identify a LATE corresponding to the strongly-identified subsample. Even when the LATE is homogeneous, the full sample IV estimator may still be severely biased if instruments are irrelevant over parts of the sample.

Our approach differs from that of Magnusson and Mavroeidis (2014) and Antoine and Boldea (2018), who use time variation in IV strength to add moment conditions in a GMM context, enabling more efficient inference and estimation. In contrast, we exploit this time variation to identify the subsample where IVs are strongest and base our estimation on this subsample. This insight allows for *consistent estimation even when subsamples suffer from identification failure.*[7] If the parameter of interest is heterogeneous, our estimator remains valid but is interpretable only within the strongly-identified sub-population.

We apply our methodology to the heteroskedasticity-based identification strategy used to estimate the causal effects of monetary policy from high-frequency data [e.g., Nakamura and Steinsson (2018)]. The key identification condition for this strategy is that the volatility of the daily changes in short-term interest rates is higher on FOMC announcement days than on non-FOMC days. Lewis (2022) provides evidence of weak full sample identification and shows that IV and GMM estimates even differ in sign. We find that identification is substantially stronger over a subsample comprising 80–90% of the data, with the excluded subsample centered around the financial crisis, during which volatility was high even on non-FOMC days. Estimation using the most strongly-identified subsample yields IV and GMM estimates that have the same sign and similar magnitudes. We recommend reporting the most strongly-identified subsample estimates in addition to the full sample estimates when strong full sample identification may be in question.

Although our new methods are able to find the most strongly-identified subsample, this subsample may still fail to be strongly-identified. For our final theoretical contribution, we develop identification-robust inference procedures using the most strongly-identified subsample. We propose versions of the Anderson-Rubin, Lagrange Multiplier, and conditional likelihood ratio tests, which depend only on this subsample. These tests are more efficient than their full sample counterparts, which include noise from regimes suffering from identification failure. When instruments are strong throughout the sample, our tests coincide with the conventional ones. When instruments are irrelevant over parts of the sample, our tests achieve higher efficiency by focusing on stronger segments. In the worst case, when IVs

[7] Another major difference from Magnusson and Mavroeidis (2014) is that we address the computational challenge for the case of multiple breaks in the first-stage coefficient. Magnusson and Mavroeidis (2014) did not attempt to address this issue and refer to it as "computationally demanding".

are weak everywhere, our methods are no less efficient than existing ones. While there is a trade-off between using fewer observations and more strongly identified subsamples, simulations show that our tests have higher power, indicating that the efficiency loss from a smaller sample size is outweighed by the gain in identification strength.

The paper is organized as follows. Section 2 introduces the potential outcome framework and dynamic causal effects, and presents identification results. Section 3 discusses issues pertaining to heteroskedasticity-based identification of monetary policy. Section 4 presents an F-test for full sample identification failure. Estimation and inference robust to weak identification are discussed in Sections 5-6. An empirical application is considered in Section 7. Section 8 concludes. The supplements Casini, McCloskey, Rolla, and Pala (2025b, 2025a) include the Monte Carlo simulations, proofs and additional results.

2 Identification of Dynamic Causal Effects

A growing literature in macroeconomics uses IVs to identify dynamic causal effects when the policy variable of interest is endogenous.[8] Many existing identification approaches can be reframed in terms of IVs, either derived from the modeling approach [e.g., heteroskedasticity-based identification as in Rigobon (2003) and Nakamura and Steinsson (2018)] or through external IVs constructed using a narrative approach [cf., Montiel Olea, Stock, and Watson (2021)]. For example, Romer and Romer (1989) study the FOMC minutes to pinpoint dates when monetary policy actions were arguably exogenous. This allows the construction of exogenous variables that can be interpreted as IVs for some structural shock of interest.[9]

We adopt a potential outcomes framework, as introduced by Rubin (1974) and extended to time series settings by Angrist and Kuersteiner (2011) and Rambachan and Shephard (2021). Let the stochastic process $V_t = (Y_t, X_t, D_t, Z_t)$ be defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Y_t is a vector of outcome variables, D_t is a policy variable, X_t is a vector of other exogenous and/or lagged endogenous variables, and Z_t is a vector of instruments. Let $\vec{X}_t = \{\ldots, X_{t-1}, X_t, \}$ denote the covariate path up to time t, with analogous definitions

[8] See, e.g., Gertler and Karadi (2015), Jordà, Schularick, and Taylor (2015), Mertens and Montiel Olea (2018), Mertens and Ravn (2013), Plagborg-Møller and Wolf (2022), Ramey and Zubairy (2018) and Stock and Watson (2012, 2018).

[9] See Ramey and Shapiro (1998) for unanticipated defense spending shocks, Kuttner (2001), Nakamura and Steinsson (2018) and Romer and Romer (2004) for monetary policy shocks, Hamilton (2003), Känzig (2021a) and Kilian (2009) for oil market shocks, Känzig (2021b) for carbon pricing shocks, Romer and Romer for tax shocks, and Ramey (2011) for government spending shocks.

for $\vec{Y}_t$, $\vec{D}_t$ and $\vec{Z}_t$. Let the policy-relevant information set at time t denoted by $\mathscr{F}_t = \sigma(\widetilde{V}_t)$ where $\sigma(\widetilde{V}_t)$ is the σ-algebra generated by the history of V_t, $\widetilde{V}_t = (\vec{Y}_{t-1}, \vec{X}_t, \vec{D}_{t-1}, \vec{Z}_{t-1})$.

Policy decisions depend on past observable variables and the contemporaneous outcome through a systematic component and on idiosyncratic information available to the policy-maker (i.e., the random component). The systematic component, denoted $D(\widetilde{V}_t, Y_t, Z_t, t)$, is a time-varying non-stochastic function of the observed random variables $\widetilde{V}_t$, contemporaneous outcome Y_t, and the contemporaneous instrument Z_t. The idiosyncratic information is represented by a scalar stochastic shock e_t that is not observed by the researcher. The policy action is determined by $D_t = \varphi(D(\widetilde{V}_t, Y_t, Z_t, t), e_t, t)$, where φ is a general mapping. In a SVAR context, e_t is the structural shock to the policy variable D_t. For example, if the monetary authority follows a simple Taylor rule for the nominal interest rate, then φ is linear and $\widetilde{V}_t$ includes inflation, output and the natural rate of interest.

We define two types of potential outcomes. The first, $Y_t\left((\epsilon_{1:t}), (z_{1:t})\right)$, denotes the counterfactual values of Y_t under hypothetical sequences of the policy shocks $\epsilon_{1:t}$ and instruments $z_{1:t}$, where $a_{1:t} = \{a_s\}_{s=1}^{t}$.

Definition 2.1. A generalized potential outcome, $Y_t\left((\epsilon_{1:t}), (z_{1:t})\right)$, is defined as the value assumed by Y_t if $e_s = \epsilon_s$ and $Z_s = z_s$ for $s = 1, \ldots, t$.

This definition excludes dependence on future shocks or instruments. The potential outcome process should not be confused with the observed outcome $\{Y_t\}_{t\geq 1} = \{Y_t(e_{1:t}, Z_{1:t})\}_{t\geq 1}$. For $h \geq 0$ and any given ϵ and z, write the time-$t+h$ potential outcome along the path $\left((e_{1:t-1}, \epsilon, e_{t+1:t+h}), (Z_{1:t-1}, z, Z_{t+1:t+h})\right)$ as

$$Y_{t,h}(\epsilon, z) = Y_{t+h}\left((e_{1:t-1}, \epsilon, e_{t+1:t+h}), (Z_{1:t-1}, z, Z_{t+1:t+h})\right),$$

where $Y_{t,h}(e_t, Z_t) = Y_{t+h}$. Definition 2.1 captures the property that $Y_{t,h}(\epsilon, z)$ also depends on policy shocks that occur between time $t+1$ and $t+h$. The notation $Y_{t,h}(e, z)$ focuses on the effect of a single policy shock on current and future outcomes akin to the idea underlying an impulse response. When the potential outcomes do not depend on the instruments, $Y_{t,h}(\epsilon, z) = Y_{t,h}(\epsilon)$, and for $\epsilon \neq \epsilon'$, $Y_{t,h}(\epsilon) - Y_{t,h}(\epsilon')$ for $h = 0, 1, \ldots$ are the dynamic causal effects of a policy shock on the outcome. In a SVAR setting, one is often interested in these dynamic causal effects which are in fact the impulse responses.

The second potential outcome that we discuss, $Y_t^*\left((d_{1:t}), (z_{1:t})\right)$, is defined as the counterfactual values of Y_t under hypothetical sequences of treatments $d_{1:t}$ and instruments $z_{1:t}$.

The distinction with $Y_t(\epsilon_{1:t},\, z_{1:t})$ is that this formulation focuses on causal effects of the policy variable D, not the policy shock e. For $t \geq 1$, we assume that $d_t \in \mathbf{D}$, $z_t \in \mathbf{Z}$ for some sets $\mathbf{D}$ and $\mathbf{Z}$. In many applications outside SVARs, the causal effects of the policy are of interest. Think about the slope of demand functions, price elasticities, response coefficients or reaction functions of, for example, asset prices to monetary policy, and so on. Typically these causal effects are analyzed using event-studies, quasi-experiments, IV regressions, etc. The recent literature on causal effects in time series [e.g., Rambachan and Shephard (2021)] focuses on the identification of causal effects of the structural shocks. In this paper, we consider identification of causal effects of the policy variable. We illustrate the difference between these two causal effects and an application to SVAR using the following two examples.

Example 2.1. Consider the following system of simultaneous equations,

$$Y_t = \beta D_t + \eta_t \qquad \text{and} \qquad D_t = aY_t + e_t, \tag{2.1}$$

where the first equation is the demand curve, the second is the supply curve, Y_t and D_t are the observed price and quantity, and η_t and e_t are the structural shocks. The parameter β captures the slope of the demand function, which corresponds to the causal effect $\partial Y_t^*(d)/\partial d = \beta$. On the other hand, in a SVAR context one may be interested in the impulse response of Y_t given a shock to supply e_t. Solving for the reduced-form of (2.1),

$$Y_t = \frac{\beta}{1-\alpha\beta} e_t + \frac{1}{1-\alpha\beta}\eta_t,$$

shows that the lag-0 impulse response is $\mathrm{d}Y_t(e)/\mathrm{d}e = \beta/(1-a\beta)$, which differs from β.

Example 2.2. Consider the following reduced-form VAR,

$$V_t = A_1 V_{t-1} + A_2 V_{t-2} + \ldots + A_p V_{t-p} + u_t,$$

where $V_t = (D_t,\, Y_t')'$ is $n \times 1$, D_t is a scalar, and u_t is a vector of reduced-form VAR innovations. The latter are related to structural shocks, $\varepsilon_t = (e_t,\, \eta_t')'$, via $u_t = B_0 \varepsilon_t$ where B_0 is a non-singular matrix. Under suitable conditions, V_t admits a moving-average representation $V_t = \sum_{j=0}^{\infty} C_j(A) B_0 \varepsilon_{t-j}$, where $C_j(A) = \sum_{i=1}^{j} C_{j-i}(A) A_i$ for $j = 1, 2, \ldots$ with $C_0(A) = I_n$

and $A_i = 0$ for $i > p$. Then, the outcome variable admits a moving-average representation,

$$Y_t = \sum_{j=0}^{\infty} c_{ye,j} e_{t-j} + \sum_{j=0}^{\infty} c_{y\eta,j} \eta_{t-j},$$

where $c_{ye,j}$ and $c_{y\eta,j}$ are blocks of $C_j(A) B_0$ partitioned conformably to Y_t, e_t and η_t. If e_t is the policy shock, the potential outcomes here are defined as

$$Y_{t,h}(\epsilon) = Y_{t,h}(\epsilon, z) = \sum_{j=0, j\neq h}^{\infty} c_{ye,j} e_{t+h-j} + \sum_{j=0}^{\infty} c_{y\eta,j} \eta_{t+h-j} + c_{ye,h}\epsilon.$$

The potential outcome $Y_{t,h}(\epsilon)$ tells us what Y_{t+h} would be if $e_t = \epsilon$ and it does not depend upon z since the instrument Z_t is excluded from the VAR. Here the absence of causal effects means that $c_{ye,h} = 0$ for all h, coinciding with the canonical condition that the impulse responses are identically equal to zero.

The potential outcome framework is useful because it allows the study of nonparametric conditions such that common statistical estimands (e.g., impulse responses) have a causal interpretation. Montiel Olea, Stock, and Watson (2021) show how to use the instrument Z_t to identify the impulse response coefficient $\phi_{r,e,h} = \partial Y_{t+h}^{(r)} / \partial e_t$ (the effect of e_t on the rth variable in Y_{t+h}). From the moving-average representation we have $\phi_{r,e,h} = \iota_r' C_h(A) B_0 \iota_1$ where ι_s denotes the s-th standard basis vector. This shows that $\phi_{r,e,h}$ depends on the A's and the first column of B_0. The following assumptions are needed for the identification of $\phi_{r,e,h}$: (i) $\mathbb{E}(Z_t e_t) = \theta \neq 0$ (instrument relevance) and (ii) $\mathbb{E}(Z_t \eta_t) = 0$ (instrument exogeneity). By (i)-(ii), $B_0^{(:,1)} = B_0 \iota_1$ is identified up to scale by the covariance between Z_t and the reduced-form innovations u_t: $\Gamma = \mathbb{E}(Z_t u_t) = \mathbb{E}(Z_t B_0 \varepsilon_t) = \theta B_0^{(:,1)}$. Using the scale normalization $B_0^{(1,1)} = 1$ [see Stock and Watson (2018) for a discussion] we have $\Gamma^{(1,1)} = \mathbb{E}(Z_t e_t) = \theta$ and $B_0^{(:,1)} = \Gamma / \Gamma^{(1,1)} = \Gamma / \iota_1' \Gamma$. It follows that $\phi_{r,e,h}$ is identified since $\phi_{r,e,h} = \iota_r' C_h(A) \Gamma / \iota_1' \Gamma$, where A can be estimated consistently from the reduced-form VAR and Γ can be estimated consistently by using the VAR residuals $\widehat{u}_t$ in place of u_t. On the other hand, identifying the causal effects of the policy D_t here would require additional identification restrictions.

Montiel Olea, Stock, and Watson (2021) use shortfalls in OPEC oil production associated with wars and civil disruptions as an instrument for the oil supply shock in the SVAR of Kilian (2009) who investigates the effect of oil supply and demand shocks on oil production and prices. This variable is plausibly correlated with the oil supply shock and, because the shortfalls are associated with political events such as wars in the Middle East, it is

plausibly uncorrelated with the demand shocks. Using the analog of the nonparametric conditions we provide below, applied to the shock e_t rather than the policy D_t, permits a causal interpretation of the impulse response even when $\mathbb{E}(Z_t e_t) = 0$ for a sub-population.

In the following, we discuss identification of causal effects of the policy via IV estimands.

2.1 Identification Conditions

We explicitly allow for endogeneity and rely on IVs. We assume that the instrument only has a contemporaneous effect on D_t so that we may write $D_t = D_t(Z_t)$ where $D_t(z) = \varphi(D(\widetilde{V}_t,\, Y_t,\, z,\, t),\, e_t,\, t)$ is the potential treatment assignment at time t when Z_t is set equal to $z \in \mathbf{Z}$. The instrument Z_t is assumed to be (conditionally) independent of the potential outcomes $Y^*_{t,j}(d,\, z)$ and treatments $D_t(z)$ but correlated with the observed treatment D_t.

Assumption 2.1. (*Independence*) *For all $d \in \mathbf{D}$, $z \in \mathbf{Z}$ and $t \geq 1$, we have*

$$\left\{\left\{Y^*_{t,h}(d,\, z)\right\}_{h\geq 0},\, D_t(z)\right\} \perp Z_t \,|\, \widetilde{V}_t. \tag{2.2}$$

Assumption 2.1 states that, given $\widetilde{V}_t$, the instrument is as good as randomly assigned.

The second assumption is that potential outcomes $Y^*_{t,h}(d,\, z)$ are a function of d but not of z. In studies of causal effects of monetary policy such as Nakamura and Steinsson (2018), $Z_t = 1$ if there is an FOMC announcement on day t and $Z_t = 0$ otherwise. Then, potential realizations of expected output growth respond to changes in the monetary policy variable regardless of whether the change is associated with an FOMC announcement or not.

Assumption 2.2. (*Exclusion*) *For all $d \in \mathbf{D}$, $t \geq 1$ and $h \geq 0$, we have*

$$\left\{Y^*_{t,h}(d,\, z) = Y^*_{t,h}(d,\, z')\right\} \,|\, \widetilde{V}_t, \qquad \text{for all } z,\, z' \in \mathbf{Z}. \tag{2.3}$$

In a dynamic simultaneous equations model (e.g., a SVAR) the exclusion restriction requires the instrument not to appear in the causal equation of interest. In Example 2.2, Assumption 2.2 corresponds to condition (ii), i.e., $\mathbb{E}(Z_t \eta_t) = 0$ where η_t is composed of the structural shocks other than e_t. Under Assumption 2.2 we write $Y^*_{t,h}(d,\, z) = Y^*_{t,h}(d)$.

Identification based on IVs requires instrument relevance or "existence of a first-stage". The latter means that $\mathbb{E}(D_t(z) \,|\, \widetilde{V}_t)$ is a non-trivial function of z. In cross-sectional settings, the existence of a first-stage is typically assumed to hold for all units to guarantee strong

identification. Strong identification of this form often fails to hold in applications involving time series data due to temporary misspecification, bad luck, rare events or parameter instability. The analysis based on articles in five leading journals that we report earlier suggests that there are time periods for which the first-stage exists (strong identification) and others for which it does not (identification failure). Standard first-stage F-tests are then likely to indicate weak identification since they are based on averaging these two sub-populations.

We provide a theoretical framework to address this identification problem by assuming that there are two sub-populations. One comprises a fraction $\pi_0 \in [0, 1]$ of the overall population for which the first-stage exists. For the second sub-population, which comprises a fraction $1 - \pi_0$ of the population, the first-stage does not exist. This leads to a new notion of LATE, which we name π-LATE, the LATE for the (unknown) π_0 fraction of the population for which the first-stage exists. If $\pi_0 = 1$, then one recovers LATE.

Denote by $|\mathbf{S}_{0,T}|$ the cardinality of $\mathbf{S}_{0,T}$ (i.e., the number of indices in $\mathbf{S}_{0,T}$).

Assumption 2.3. (*Partial first-stage*) *Assume there exists* $\mathbf{S}_{0,T} \subseteq \{1, \ldots, T\}$ *such that* $|\mathbf{S}_{0,T}| = \lfloor \pi_0 T \rfloor$ *with* $\pi_0 \in (0, 1]$ *and for* $t \in \mathbf{S}_{0,T}$, $\mathbb{E}(D_t(z) \,|\, \widetilde{V}_t)$ *is a non-trivial function of* z, *i.e., for* $t \in \mathbf{S}_{0,T}$, $\mathbb{E}(D_t(z') \,|\, \widetilde{V}_t) - \mathbb{E}(D_t(z) \,|\, \widetilde{V}_t) \neq 0$ *for* z', $z \in \mathbf{Z}$ *such that* $z \neq z'$.[10]

Assumption 2.3 implies that there are two sub-populations: one for which the first-stage exists and one for which it does not. An average treatment effect can only be identified via IVs for the fraction π_0 of the population for which a first-stage exists.

The next assumption is monotonicity which, under heteroskedasticity-based identification of monetary policy (see Section 3), means that while for some days the FOMC announcement does not coincide with higher volatility in the policy variable, all of those days in which the announcement affects the volatility of the policy variable, volatility is shifted up.

Assumption 2.4. (*Monotonicity*) $\mathbf{D} \subseteq \mathbb{R}$. *For all* z, $z' \in \mathbf{Z}$ *and* $t \in \mathbf{S}_{0,T}$, *either* $D_t(z) \geq D_t(z')$ *or* $D_t(z') \geq D_t(z)$ *with probability 1.*

If $\pi_0 = 1$ (so $|\mathbf{S}_{0,T}| = T$), the condition reduces to that in Imbens and Angrist (1994).

Following Kolesár and Plagborg-Møller (2025), we impose the following assumption.

Assumption 2.5. *For all* $t \geq 1$ *and* $h \geq 0$, *(i)* $Y^*_{t,h}(\cdot)$ *is locally absolutely continuous on* $\mathbf{D}$ *and (ii)* $\mathbb{E}\left[\int_{\mathbf{D}} |\partial Y^*_{t,h}(d)/\partial d| \mathrm{d}d \,\middle|\, \widetilde{V}_t\right] < \infty$.

[10] We assume that all expectations exist.

Assumption 2.5 allows D_t to be either discrete, continuous or mixed. When D_t is discrete or mixed, it is implicitly assumed that to deal with the gaps in the support of D_t one extends $Y^*_{t,h}(\cdot)$ to $\mathbf{D}$ such that the extension is locally absolutely continuous. The support of D_t is allowed to be unbounded. These conditions are weaker than counterparts imposed in the recent literature [cf. Casini and McCloskey (2025) and Rambachan and Shephard (2021)], in particular local absolute continuity replaces differentiability of $Y^*_{t,h}(\cdot)$ plus bounded support of D_t. It allows the application of the fundamental theorem of calculus to $Y^*_{t,h}(\cdot)$ without requiring the support of D_t to be bounded.

2.2 Identification Results

2.2.1 Identification of Causal Effects

We first discuss the case of a discrete instrument. When the first-stage does not exist for all t, it is useful to define an IV estimand corresponding to the sub-population for which it does. Let the generalized Wald estimand be defined for all z', $z \in \mathbf{Z}$ by

$$\beta_{\pi,t,h}(\widetilde{v}) = \frac{\mathbb{E}\left(Y_{t+h} \middle| Z_t = z', \widetilde{V}_t = \widetilde{v}\right) - \mathbb{E}\left(Y_{t+h} \middle| Z_t = z, \widetilde{V}_t = \widetilde{v}\right)}{\mathbb{E}\left(D_t \middle| Z_t = z', \widetilde{V}_t = \widetilde{v}\right) - \mathbb{E}\left(D_t \middle| Z_t = z, \widetilde{V}_t = \widetilde{v}\right)}, \qquad \text{for } t \in \mathbf{S}_{0,T}, \tag{2.4}$$

where $\widetilde{v} \in \mathbf{V}$. This is the ratio of a reduced-form generalized impulse response to a first-stage generalized impulse response for $t \in \mathbf{S}_{0,T}$. We show that for $t \in \mathbf{S}_{0,T}$, the estimand $\beta_{\pi,t,h}(\widetilde{v})$ identifies a weighted average of causal effects for the compliers. Recall that $t \in \mathbf{S}_{0,T}$ and π_0 are related by $|\mathbf{S}_{0,T}| = \lfloor \pi_0 T \rfloor$. When $\pi_0 = 1$ and there is no conditioning on $\widetilde{V}_t = \widetilde{v}$, $\beta_{1,t,h}$ reduces to the Wald estimand considered by Rambachan and Shephard (2021). For $t \notin \mathbf{S}_{0,T}$, $\beta_{\pi,t,h}$ does not identify a causal effect because the denominator of (2.4) is equal to zero.

We show that for $t \in \mathbf{S}_{0,T}$, the generalized Wald estimand is equal to a weighted average of marginal effects where the latter are the derivatives $\partial Y^*_{t,h}(d)/\partial d$.

Proposition 2.1. *(π-LATE) Let Assumptions 2.1-2.5 hold. For $t \in \mathbf{S}_{0,T}$, $h \geq 0$, $\widetilde{v} \in \mathbf{V}$ and z', $z \in \mathbf{Z}$, we have*

$$\beta_{\pi,t,h}(\widetilde{v}) = \int_{\mathbf{D}} \mathbb{E}\left[\frac{\partial Y^*_{t,h}(d)}{\partial d}\middle| D_t(z) \leq d \leq D_t(z'), \widetilde{V}_t = \widetilde{v}\right] w_t(d|\,\widetilde{v})\,\mathrm{d}d, \quad \text{where} \tag{2.5}$$

$$w_t(d|\,\widetilde{v}) = \frac{\mathbb{P}\left(D_t(z) \leq d \leq D_t(z') \,|\, \widetilde{V}_t = \widetilde{v}\right)}{\int_{\mathbf{D}} \mathbb{P}\left(D_t(z) \leq d \leq D_t(z') \,|\, \widetilde{V}_t = \widetilde{v}\right)\mathrm{d}r} \geq 0 \quad \text{and} \quad \int_{\mathbf{D}} w_t(d|\,\widetilde{v})\,\mathrm{d}d = 1.$$

Proposition 2.1 shows that $\beta_{\pi,t,h}(\widetilde{v})$ identifies a weighted average of causal effects for compliers, characterized by $D_t(z') > D_t(z)$, for observations with a first-stage, with weights $w_t(d|\,\widetilde{v})$ determined by the (conditional) likelihood that $D_t(z) \leq d \leq D_t(z')$. We refer to the average treatment effect on the right-hand side of (2.5) as the time-t π-LATE since it is the LATE for the observations in this sub-population, which is a fraction π_0 of the whole population. In practice, the IV estimand $\beta_{\pi,t,h}(\widetilde{v})$ is characterized by two types of averaging. First, there is averaging over time. For any treatment d, the average involves only those observations whose treatment variable can be induced to change by a change in the instrument and is computed only over those observations that satisfy the first-stage (i.e., $t \in \mathbf{S}_{0,T}$). The second averaging is over different treatment values d at the same date t. This is reflected in the weight $w_t(\cdot)$ which is proportional to the number of observations in $\mathbf{S}_{0,T}$ for which $D_t(z) \leq d \leq D_t(z')$. Indeed, under regularity conditions permitting one to change the order of differentiation and integration, viz.,

$$\mathbb{E}\left[\frac{\partial Y^*_{t,h}(d)}{\partial d}\middle|\, D_t(z) \leq d \leq D_t(z'),\, \widetilde{V}_t = \widetilde{v}\right] = \frac{\partial}{\partial d}\mathbb{E}\left[Y^*_{t,h}(d)\middle|\, D_t(z) \leq d \leq D_t(z'),\, \widetilde{V}_t = \widetilde{v}\right],$$

$\beta_{\pi,t,h}(\widetilde{v})$ can be interpreted as a local average marginal effect.

Stationarity of the conditional joint distribution of the average potential outcome and treatment assignment functions for observations with a first-stage lends further interpretability to the generalized Wald estimand. Specifically, if $\{Y^*_{t,h}(d), D_t(z)\}|\widetilde{V}_t$ is identically distributed across t for all $t \in \mathbf{S}_{0,T}$, $d \in \mathbf{D}$ and $z \in \mathbf{Z}$, Proposition 2.1, immediately implies that $\beta_{\pi,t,h}$ is equal for all $t \in \mathbf{S}_{0,T}$. Given this, we can write $\beta_{\pi,t,h} = \beta_{\pi,h}$, making explicit that the generalized Wald estimand (2.4) equals a weighted average of causal effects for members of the sub-population with a first-stage, which represents a π_0-sized fraction of the total population. Under this assumption, we refer to the average causal effect inside of the integral as π-LATE since it is a LATE for a member of the $\mathbf{S}_{0,T}$ sub-population whose treatment variable can be induced to change by a change in the instrument.

The sample counterpart to the generalized Wald estimand (2.4) involves replacing the conditional expectations with sample estimates based upon observations $t \in \mathbf{S}_{0,T}$, yielding an estimator of a causal effect. When Assumption 2.3 holds with $\pi_0 \in (0, 1)$, the full sample estimand, i.e., the ratio of the time averages of the numerator and denominator of (2.4), is a poor representative of the full sample average treatment effects because it includes observations for which the instrument is not relevant in the averaging. We caution that the

usual practice of estimating the conditional expectations in (2.4) with full sample estimates will not estimate the full sample LATE, but π-LATE.

Angrist, Graddy, and Imbens (2000) and Rambachan and Shephard (2021) consider related results in cross-sectional and time series settings. The difference here is that we do not require D_t to be continuous or that the first-stage holds for all t. Kolesár and Plagborg-Møller (2025) established a similar result for the slope coefficient in the population version of the "reduced-form" regression of the outcome Y_{t+h} onto Z_t where they imposed no restriction on the first-stage and allowed for a continuous instrument.

A connection to program evaluation with binary policy actions arises when we map a dynamic problem with continuous variables into one with binary policy actions and instruments. For example, consider the analysis of causal effects of monetary policy using heteroskedasticity-based identification [cf. Nakamura and Steinsson (2018) and Rigobon and Sack (2003)]. Define a binary instrument Z_t with $Z_t = 1$ if there is a scheduled announcement on day t and $Z_t = 0$ otherwise. The policy Δi_t typically reflects changes in short-term interest rates. Identification relies on higher volatility in Δi_t during announcement days (policy sample) compared to non-announcement days (control sample). Think about mapping $|\Delta i_t|$ into a binary treatment such that $D_t = 1$ if $|\Delta i_t| \geq \delta$ for some threshold $\delta > 0$ and $D_t = 0$ if $|\Delta i_t| < \delta$ [cf. Rigobon and Sack (2003)]. Here π-LATE captures the average treatment effect for the sub-population whose interest rate changes exceed δ only when there is an announcement (i.e., when $Z_t = 1$). Observations where $|\Delta i_t| < \delta$ regardless of announcements are "never-takers," while those with $|\Delta i_t| \geq \delta$ regardless of announcements are "always-takers." Under monotonicity, these groups form the non-compliers, whose responses are driven by idiosyncratic factors other than announcement-specific effects. In Section 3 we document regimes where the volatility of Δi_t is high even in the absence of announcements.

Sojitra and Syrgkanis (2025) study dynamic treatment regimes with one-sided compliance where treatments in each period may depend on past instruments, treatments, outcomes, and confounding factors, while instruments in each period are generated based on prior instruments, treatments, and states. This setting encompasses applications such as digital recommendation systems and adaptive medical trials. Their focus is on the causal effect of treatment histories on long term outcomes, rather than of one-time shocks or single policy shifts on outcomes at horizon h. Under binary instruments and treatments, they establish nonparametric identification of the expected values of multi-period treatment effect contrasts for the corresponding complier subpopulations, which they refer to as dynamic LATE.

2.2.2 Identification of Compliers and Exclusion Restriction

A practical challenge for the π-LATE framework, and LATE frameworks in general, is that the sub-population of compliers is unknown. However, in time series settings with binary instruments, we show below that one can identify the compliers individually, i.e., to determine whether each observation t is a complier. In this section, we consider a binary instrument, e.g., $Z_t = 1$ if t is an FOMC meeting day and $Z_t = 0$ otherwise. Under Assumption 2.4, assume without loss of generality that $D_t(1) \geq D_t(0)$ for all t. Then, observation $t_0 \in \mathbf{S}_{0,T}$ is a complier if and only if $D_{t_0}(1) > D_{t_0}(0)$ with probability one—if the treatment changes in response to the instrument.

We begin with the following assumption which states that each observation is either a complier or a non-complier with certainty.

Assumption 2.6. *(Deterministic complier status) For each t either $\mathbb{P}(D_t(1) > D_t(0)) = 1$ or $\mathbb{P}(D_t(1) > D_t(0)) = 0$.*

Assumption 2.6 rules out cases where $\mathbb{P}(D_t(1) > D_t(0)) = p$ for some $p \in (0, 1)$. A non-complier cannot be characterized by $\mathbb{P}(D_t(1) > D_t(0)) > 0$. The latter probability must be zero. Under Assumption 2.6, Lemma S.D.2 in the supplement shows that $\mathbb{P}(D_{t_0}(1) > D_{t_0}(0)) = 1$ is equivalent to $\mathbb{E}(D_{t_0}(1)) > \mathbb{E}(D_{t_0}(0))$. This equivalence implies that compliers can be identified by comparing the expected treatment values under different instrument values.[11] Under mild smoothness assumptions that we discuss below, the latter two expected values can be estimated consistently from the sample so that we can determine whether t_0 is a complier in large samples by looking at the corresponding inequality based on sample quantities.

Let $\mathbf{P} \subset \{1, \ldots, T\}$ denote the "policy sample", the set of observations for which $Z_t = 1$ so that $D_t = D_t(1)$ for all $t \in \mathbf{P}$, and let $\mathbf{C} = \{1, \ldots, T\} \backslash \mathbf{P}$ denote the "control sample", where $D_t = D_t(0)$. It is reasonable to assume that, for a given value of the instrument, the potential treatment assignments vary smoothly over time. Suppose we wish to determine whether an observation $t_0 \in \mathbf{P}$ is a complier. Since $D_{t_0}(0)$ is not observed, under time-smoothness we approximate $\mathbb{E}(D_{t_0}(0))$ by averaging nearby observations in the control sample. Letting

[11] Note that this result is different from that in Lemma 2.1 in Abadie (2003) who shows that under several assumptions the proportion of compliers can be identified by $\mathbb{E}(D_i(1)) - \mathbb{E}(D_i(0))$ in a cross-sectional setting. He uses this lemma to show that any statistical characteristic that can be defined in terms of moments of the joint distribution of (Y_i, D_i, Z_i) is identified for compliers. He then remarks that it is not possible to identify compliers individually under these assumptions.

$N_0(t_0)$ denote the n_0 largest indices $s \in \mathbf{C}$ such that $s \leq t_0 - 1$, this implies

$$\overline{D}_{C,t_0-1,n_0} \equiv n_0^{-1} \sum_{s \in N_0(t_0)} D_s \xrightarrow{\mathbb{P}} \mathbb{E}\left(D_{t_0-1}(0)\right)$$

as $n_0 \to \infty$ with $n_0/|\mathbf{C}| \to 0$ under mild conditions. In addition, it follows that $\mathbb{E}\left(D_{t_0-1}(0)\right)$ is close to $\mathbb{E}\left(D_{t_0}(0)\right)$. A similar argument can be applied to $\mathbb{E}\left(D_{t_0}(1)\right)$ using adjacent days in the policy sample: we have $\overline{D}_{P,t_0,n_1} \xrightarrow{\mathbb{P}} \mathbb{E}\left(D_{t_0}(1)\right)$ as $n_1 \to \infty$ with $n_1/|\mathbf{P}| \to 0$, where $\overline{D}_{P,t_0,n_1} = n_1^{-1} \sum_{s \in N_1(t_0)} D_s$ and $N_1(t_0)$ denotes the n_1 largest indices $s \in \mathbf{P}$ such that $s \leq t_0$. Thus, observation $t_0 \in \mathbf{P}$ is a complier if and only if $\overline{D}_{P,t_0,n_1} - \overline{D}_{C,t_0-1,n_0} \xrightarrow{\mathbb{P}} c$ as $n_0, n_1 \to \infty$ with $n_0/|\mathbf{C}|, n_1/|\mathbf{P}| \to 0$ for any $c > 0$.

Intuitively, even though $D_{t_0}(0)$ is not observed when $t_0 \in \mathbf{P}$, observations close to t_0 characterized by no FOMC announcement provide information about what $\mathbb{E}\left(D_{t_0}(0)\right)$ would have been in the absence of an FOMC announcement.[12] There are about six weeks in between any two FOMC meetings, and so $n_0 \approx 30$. Alternatively, following Nakamura and Steinsson (2018) the control sample could include all Tuesdays and Wednesdays that are not FOMC meeting days. Nevertheless, one can skip the observation that pertains to the previous meeting, say $D_{t_{-1}}(0)$, whose realization is not observed, and continue averaging using the observations prior to that meeting as well to construct the average $\overline{D}_{C,t_0-1,n_0}$ possibly applying down-weighting for observations further in time from t_0, i.e., use $\ldots, D_{t_{-1}-1}, D_{t_{-1}+1}, D_{t_{-1}+2} \ldots, D_{t_0-2}, D_{t_0-1}$. Similarly, observations in $\mathbf{P}$ close to t_0 provide information about what $\mathbb{E}\left(D_{t_0}(1)\right)$ would have been, though here the successive observations are separated chronologically by the observations in the control sample $\mathbf{C}$.

We now present the formal result for identification of the compliers. The following two assumptions can be justified in large samples when the mean (potential) treatment assignments in both the control and policy samples vary smoothly over time. Under an infill asymptotic embedding where the original observations indexed by $t = 1, \ldots, T$ are mapped into the unit interval [0, 1] via $u = t/T$, if $\lim_{T\to\infty} \mathbb{E}(D_{Tu}(z))$ is continuous in u under a fixed instrument value $z \in \mathbf{Z}$, the following assumptions hold. This type of continuity accommodates general forms of smoothly time-varying means but not abrupt breaks in mean.[13]

Assumption 2.7. *(i) For any $t \in \mathbf{C}$, $\overline{D}_{C,t,n} \xrightarrow{\mathbb{P}} \mathbb{E}\left(D_t\right)$ as $n \to \infty$ with $n/|\mathbf{C}| \to 0$. (ii) For $t \in \mathbf{P}$ $\mathbb{E}(D_{t-1}(0)) = \mathbb{E}(D_t(0))$.*

[12] One could also use the observations to the right of t_0 to construct $\overline{D}_{C,t_0+1,n}$, i.e., $D_{t_0+1}, \ldots, D_{t_0+n}$.

[13] However, breaks in the mean of the assignment process can be estimated under some conditions as we explain below. Then, time-smoothness is required to hold only in regimes defined by successive break dates.

Assumption 2.8. *(i) For any* $t \in \mathbf{P}$ $\overline{D}_{P,t,n} \xrightarrow{\mathbb{P}} \mathbb{E}(D_t)$ *as* $n \to \infty$ *with* $n/|\mathbf{P}| \to 0$. *(ii) For* $t \in \mathbf{C}$ $\mathbb{E}(D_t(1)) = \mathbb{E}(D_{s^*(t)}(1))$ *where* $s^*(t) = \text{argmin}_{s \in \mathbf{P}} |t - s|$.

Assumption 2.7(i) requires a law of large numbers to apply to the rolling-window sample average of D_t at the points of continuity of $\mathbb{E}(D_t)$. It is a minimal technical assumption. Assumption 2.7(ii) strengthens part (i) a bit by requiring that for $t \in \mathbf{P}$ the potential treatment assignment under the trajectory $Z_t = 0$ has a locally constant mean. Assumption 2.8(i) adapts Assumption 2.7(i) to the observations in $\mathbf{P}$. This is a stronger assumption since two successive observations in the policy sample are separated by several observations in the control sample. Assumption 2.8(ii) requires that $\mathbb{E}(D_t(1))$ for $t \in \mathbf{C}$ is equal to the mean of the potential treatment assignment at the closest date in the policy sample $s^*(t)$. This is a first moment constancy assumption on the potential treatment assignment under the trajectory $Z_t = 1$. Assumption 2.7 is used to identify the compliers in the policy sample, while Assumption 2.8 is used to identify the compliers in the control sample.

Theorem 2.1. *Let Assumptions 2.6-2.8 hold and* $n_0, n_1 \to \infty$ *with* $n_0/|\mathbf{C}|, n_1/|\mathbf{P}| \to 0$. *Then:*
(i) $t \in \mathbf{P}$ *is a complier if and only if* $\overline{D}_{P,t,n_1} - \overline{D}_{C,t-1,n_0} \xrightarrow{\mathbb{P}} c$ *where* $c > 0$.
(ii) $t \in \mathbf{C}$ *is a complier if and only if* $\overline{D}_{P,s^*(t),n_1} - \overline{D}_{C,t,n_0} \xrightarrow{\mathbb{P}} \widetilde{c}$ *where* $\widetilde{c} > 0$.

Theorem 2.1 shows that the compliers can be identified individually. To the best of our knowledge, there is no equivalent result in the cross-sectional setting. The assumptions of the theorem are easily satisfied in time series applications. Using Theorem 2.1 is straightforward: one computes the difference between two sample averages and check whether it is greater than zero. Given the sampling uncertainty associated with the two averages, one can conduct inference using a t-statistic for the null hypothesis $\mathbb{E}(D_{t_0}(1)) - \mathbb{E}(D_{t_0}(0)) = 0$ (t_0 is not a complier) versus the alternative hypothesis that $\mathbb{E}(D_{t_0}(1)) - \mathbb{E}(D_{t_0}(0)) > 0$ (t_0 is a complier).

An additional challenge specific to the π-LATE framework is that the set of observations with a first stage $\mathbf{S}_{0,T}$ is also unknown. However, as the following result states, under Assumption 2.4, in the absence of covariates $\widetilde{V}_t$, $\mathbf{S}_{0,T}$ is equal to the (identified) set of compliers —i.e., observations for which the first-stage holds individually.

Proposition 2.2. *Suppose* Z_t *is binary and let Assumptions 2.3 without conditioning on* $\widetilde{V}_t$, *2.4 and 2.6 hold. Then, the set of compliers coincide with* $\mathbf{S}_{0,T}$.

Knowledge of the compliers sub-population (and hence of the non-compliers sub-population) can be used to test the exclusion restriction (cf. Assumption 2.2) by comparing the mean

outcomes of groups of non-compliers across different values of the instrument. For example, one can divide any large subset of non-compliers into two groups according to their assignment status. If one can reject the hypothesis that the average outcomes in these two groups is the same, then the exclusion restriction cannot hold.

Under Assumption 2.4 with $D_t(1) \geq D_t(0)$, the set of non-compliers is $\mathcal{NC} = \{t \in \{1, \ldots, T\} : D_t(1) = D_t(0) = D_t\}$. Let $\mathcal{NC}^s$ be any non-empty subset of $\mathcal{NC}$ such that $\mathcal{NC}^s_{\mathbf{P}} = \mathcal{NC}^s \cap \mathbf{P} \neq \emptyset$ and $\mathcal{NC}^s_{\mathbf{C}} = \mathcal{NC}^s \cap \mathbf{C} \neq \emptyset$. We can test the exclusion restriction in Assumption 2.2 under the following assumption on the subsets $\mathcal{NC}^s_{\mathbf{P}}$ and $\mathcal{NC}^s_{\mathbf{C}}$.

Assumption 2.9. *(i)* $\mathbb{E}[Y^*_t(d,z)|t \in \mathcal{NC}^s_{\mathbf{P}}] = \mathbb{E}[Y^*_r(d,z)|r \in \mathcal{NC}^s_{\mathbf{C}}]$ *for all* $t, r \geq 1$, $d \in \mathbf{D}$ *and* $z \in \mathbf{Z}$. *(ii)* $\{D_t, \widetilde{V}_t\}|t \in \mathcal{NC}^s_{\mathbf{P}} \sim \{D_r, \widetilde{V}_r\}|r \in \mathcal{NC}^s_{\mathbf{C}}$ *for all* $t, r \geq 1$. *(iii) For* $\mathbf{R} = \mathbf{C}$ *or* $\mathbf{P}$, $|\mathcal{NC}^s_{\mathbf{R}}|^{-1} \sum_{t \in \mathcal{NC}^s_{\mathbf{R}}} Y_t \xrightarrow{\mathbb{P}} \mathbb{E}\left[Y_t | t \in \mathcal{NC}^s_{\mathbf{R}}\right]$ *as* $|\mathcal{NC}^s_{\mathbf{R}}| \to \infty$.

Condition (i) states that the potential outcome for non-compliers is mean-stationary and the mean is the same across control and policy subsamples. Condition (ii) states that the policy variable and past observables for non-compliers are distributed identically across the control and policy subsamples. Condition (iii) states that a law of large numbers holds for non-compliers observations in both the control and policy subsamples. As long as the policy sample does not tend to contain systematic different values of the policy variable D_t among non-compliers than the control sample, these are relatively mild conditions.

Proposition 2.3. *Suppose* Z_t *is binary and let Assumptions 2.4 and 2.9 hold. If Assumption 2.2 holds, then as* $|\mathcal{NC}^s_{\mathbf{P}}|, |\mathcal{NC}^s_{\mathbf{C}}| \to \infty$,

$$|\mathcal{NC}^s_{\mathbf{P}}|^{-1} \sum_{t \in \mathcal{NC}^s_{\mathbf{P}}} Y_t - |\mathcal{NC}^s_{\mathbf{C}}|^{-1} \sum_{t \in \mathcal{NC}^s_{\mathbf{C}}} Y_t \xrightarrow{\mathbb{P}} 0.$$

Using Proposition 2.3 to test Assumption 2.2 is simple: since non-compliers can be identified individually using Theorem 2.1, one can immediately compute the sample averages specified in Proposition 2.3 and conduct inference using a t-statistic for the null hypothesis that the population mean of $t \in \mathcal{NC}^s_{\mathbf{P}}$ is equal to that of $t \in \mathcal{NC}^s_{\mathbf{C}}$. The researcher has the ability to choose the subset of non-compliers $\mathcal{NC}^s$ when implementing this test. The simplest choice is to set $\mathcal{NC}^s = \mathcal{NC}$, however, the researcher also has the ability to direct the power of the test toward particular types of non-compliers they may suspect of being more likely to violate the exclusion restriction. For example, one may wish to focus on $\mathcal{NC}^s = \{t \in \mathcal{NC} : D_t \geq d^*\}$ or $\mathcal{NC}^s = \{t \in \mathcal{NC} : D_t < d^*\}$ for some d^* value, such as

$d^* = |\mathcal{NC}|^{-1} \sum_{t \in \mathcal{NC}} D_t$, in order to test violations of the exclusion restriction for observations roughly corresponding to "always-takers" or "never-takers" in the case of a binary treatment.[14]

3 High-Frequency Identification of Monetary Policy Effects

To study the effects of monetary policy on real variables, a large literature has relied on high-frequency identification. This exploits the fact that at the time of an FOMC meeting a large amount of economic news is revealed. Here we discuss Rigobon's (2003) heteroskedasticity identification approach which uses a 1-day window [see, e.g., Nakamura and Steinsson (2018)], and can be reformulated as IV-based identification. In Section 3.1 we explain when the resulting reduced-form estimands have a causal meaning within the potential outcome framework of Section 2. In Section 3.2 we discuss the weak identification problem of current approaches and show how the π-LATE framework can be used to strengthen identification.

3.1 Heteroskedasticity-Based Identification

Consider the following system of equations:

$$\tilde{Y}_t = \beta_0 \tilde{D}_t + \eta_t, \qquad \text{and} \qquad \tilde{D}_t = a\tilde{Y}_t + e_t, \tag{3.1}$$

where $\tilde{Y}_t$ is the (demeaned) daily change in an outcome variable, (e.g., an asset price or a bond yield) and $\tilde{D}_t$ is the (demeaned) daily change in the unexpected component of a short-term interest rate or policy news (e.g., Δi_t as discussed after Proposition 2.1), η_t is a shock to $\tilde{Y}_t$, e_t is the monetary policy shock and a and β_0 are scalar parameters. The errors η_t and e_t have no serial correlation and are mutually uncorrelated. The parameter of interest is β_0 which represents the causal effect of monetary policy on the outcome variable. The model in (3.1) could arise from a bivariate VAR. In fact, one could add a vector X_t of exogenous variables to the model in (3.1). However, to focus on the main intuition, we follow Nakamura and

[14] Under an analogous assumption to Assumption 2.9 for a function of outcomes $f(Y_t)$, an analogous result to Proposition 2.3 holds. One may use this fact, for example, to test if the variances of groups of non-compliers are equal across different values of the instrument, which is implied by Assumption 2.2, or to test the equality of a set of moments across different values of the instrument. Taking this logic even further, one could invoke Gilvenko-Cantelli theorems to show that the difference between the empirical distribution functions of observations in $\mathcal{NC}^s_{\mathbf{P}}$ and $\mathcal{NC}^s_{\mathbf{C}}$ converge uniformly to zero under Assumption 2.2 and use a test for the equality of distributions such as the Kolmogorov-Smirnov test. However, we focus here on testing the equality of means across instrument values because the level of the outcomes, rather than functions of them, are likely to be of primary importance in practice.

Steinsson (2018) and we omit X_t and lagged terms of $\tilde{Y}_t$ and $\tilde{D}_t$. See Casini and McCloskey (2025) for a detailed discussion of why the lags can be omitted in this setting.

The model in (3.1) is a special case of the generalized framework studied in Section 2. It is useful because it directly motivates a particular IV estimand. However, we study the causal interpretation of this estimand in the general case for which the linear model with stable parameters is not the correct specification.

Heteroskedasticity-based identification requires that the variance of the monetary shock increases in the days of FOMC announcements, while the variance of other shocks is unchanged. Let T_P denote the number of days containing an FOMC announcement (policy sample), and let T_C denote the number of days that do not contain an FOMC announcement (control sample). Let $\sigma^2_{e,P} = T_P^{-1} \sum_{t \in \mathbf{P}} \mathbb{E}\left(e_t^2\right)$ and $\sigma^2_{e,C} = T_C^{-1} \sum_{t \in \mathbf{C}} \mathbb{E}\left(e_t^2\right)$ be the average variance of the monetary policy shock in the policy and control samples. Define $\sigma^2_{\eta,P}$ and $\sigma^2_{\eta,C}$ similarly. Then, the identification condition is

$$\sigma_{e,P} > \sigma_{e,C} \qquad \text{and} \qquad \sigma_{\eta,P} = \sigma_{\eta,C}. \tag{3.2}$$

Identification can be shown analytically by first solving for the reduced-form of (3.1):

$$\tilde{Y}_t = \frac{1}{1 - a\beta_0}\left(\eta_t + \beta_0 e_t\right), \qquad \tilde{D}_t = \frac{1}{1 - a\beta_0}\left(a\eta_t + e_t\right).$$

Let Σ_i denote the covariance matrix of $[\tilde{Y}_t,\, \tilde{D}_t]'$ in the subsample $i = P,\, C$. It follows that

$$\Sigma_i = \frac{1}{\left(1 - a\beta_0\right)^2} \begin{bmatrix} \sigma^2_{\eta,i} + \beta_0^2 \sigma^2_{e,i} & \beta_0 \sigma^2_{e,i} + a\sigma^2_{\eta,i} \\ \beta_0 \sigma^2_{e,i} + a\sigma^2_{\eta,i} & \sigma^2_{e,i} + a^2 \sigma^2_{\eta,i} \end{bmatrix}, \qquad i = P,\, C.$$

It is typical in the literature to assume within-regime covariance-stationarity, i.e., $\mathbb{E}\left(e_t^2\right)$ and $\mathbb{E}\left(\eta_t^2\right)$ are constant within each subsample $\mathbf{P}$ and $\mathbf{C}$ which is, however, restrictive for economic time series. It turns out that this is not necessary for identification. Volatilities can be time-varying as long as the average volatilities $\sigma_{e,i}$ and $\sigma_{\eta,i}$ $(i = P,\, C)$ satisfy (3.2).

When (3.1) is correctly specified, i.e., the true model is linear with stable parameters, the parameter β_0 can be identified using (3.2) by taking the difference between the covariance

matrices in the policy and control samples:

$$\beta_0 = \frac{\Delta\Sigma^{(1,2)}}{\Delta\Sigma^{(2,2)}} = \frac{T_P^{-1}\sum_{t\in\mathbf{P}}\operatorname{Cov}\left(\tilde{Y}_t,\,\tilde{D}_t\right) - T_C^{-1}\sum_{t\in\mathbf{C}}\operatorname{Cov}\left(\tilde{Y}_t,\,\tilde{D}_t\right)}{T_P^{-1}\sum_{t\in\mathbf{P}}\operatorname{Var}\left(\tilde{D}_t\right) - T_C^{-1}\sum_{t\in\mathbf{C}}\operatorname{Var}\left(\tilde{D}_t\right)}, \qquad \text{where} \tag{3.3}$$

$$\Delta\Sigma \triangleq \Sigma_P - \Sigma_C = \frac{\sigma_{e,P}^2 - \sigma_{e,C}^2}{(1-a_1\beta_0)^2}\begin{bmatrix}\beta_0^2 & \beta_0\\ \beta_0 & 1\end{bmatrix}.$$

To determine which average treatment effect this approach identifies in the general framework, we re-frame this problem in terms of instrumental variables as follows. Let $Z_t = 1$ for $t \in \mathbf{P}$ and $Z_t = 0$ for $t \in \mathbf{C}$. Multiply both sides of (3.1) by $\tilde{D}_t$ to yield $\tilde{D}_t\tilde{Y}_t = \beta_0\tilde{D}_t^2 + \tilde{D}_t\eta_t$. We can use Z_t as an instrument for $\tilde{D}_t^2$. The first-stage is $\tilde{D}_t^2 = \theta Z_t + \varepsilon_t$, where ε_t is some error term satisfying $\varepsilon_t \geq -\theta Z_t$. The resulting Wald estimand is

$$\beta^*_{\pi,t,0} = \frac{\mathbb{E}\left(\tilde{D}_t\tilde{Y}_t|\,Z_t=1\right) - \mathbb{E}\left(\tilde{D}_t\tilde{Y}_t|\,Z_t=0\right)}{\mathbb{E}\left(\tilde{D}_t^2|\,Z_t=1\right) - \mathbb{E}\left(\tilde{D}_t^2|\,Z_t=0\right)}, \tag{3.4}$$

which corresponds to the Wald estimand (2.4) for $h = 0$, $Y_t = \tilde{D}_t\tilde{Y}_t$, $D_t = \tilde{D}_t^2$ and no conditioning variable $\tilde{V}_t$. Under covariance stationarity within subsamples $\mathbf{P}$ and $\mathbf{C}$, the right-hand side of (3.4) is equal to the right-hand side of (3.3). The following corollary of Proposition 2.1 presents the causal meaning of $\beta^*_{\pi,t,0}$ under the general setting of Section 2.

Corollary 3.1. (*LATE in heteroskedasticity-based identification*) *Let Assumptions 2.1-2.5 hold for $Y_t = \tilde{D}_t\tilde{Y}_t$ and $D_t = \tilde{D}_t^2$ with $\tilde{D}_t(1)^2 \geq \tilde{D}_t(0)^2$. For $t \in \mathbf{S}_{0,T}$, we have*

$$\beta^*_{\pi,t,0} = \frac{\int_{\mathbf{D}}\mathbb{E}\left(\frac{\partial\left(\tilde{d}\tilde{Y}^*_{t,0}(\tilde{d})\right)}{\partial(\tilde{d}^2)}\middle|\,\tilde{D}_t(1)^2 \geq \tilde{d}^2 \geq \tilde{D}_t(0)^2\right)\mathbb{P}\left(\tilde{D}_t(1)^2 \geq \tilde{d}^2 \geq \tilde{D}_t(0)^2\right)\mathrm{d}(\tilde{d}^2)}{\int_{\mathbf{D}}\mathbb{P}\left(\tilde{D}_t(1)^2 \geq \tilde{d}^2 \geq \tilde{D}_t(0)^2\right)\mathrm{d}(\tilde{d}^2)}. \tag{3.5}$$

Corollary 3.1 shows that the Wald estimand in (3.4) has a causal meaning because it is the ratio of a reduced-form generalized impulse response of $\tilde{D}_t\tilde{Y}_t$ to a first-stage generalized impulse response of $\tilde{D}_t^2$. More specifically, $\beta^*_{\pi,t,0}$ identifies a weighted average of the derivative of the product between the potential outcome and policy variable for compliers. Hence, contrary to popular belief, the causal interpretation of the heteroskedasticity-based estimator (i.e., Rigobon's estimator) estimator is not the same as that of a standard IV estimator—though it remains local in nature as it averages over compliers. We continue to refer to it as LATE with the understanding that it is a LATE for $\tilde{D}_t\tilde{Y}_t$, not $\tilde{Y}_t$ itself.

Here the compliers are the observations for which the announcement induces a higher volatility of the policy $\tilde{D}_t$. In contrast, the non-compliers are characterized by idiosyncratic or general equilibrium factors that dominate the news specific to the announcement. That is, regimes where $\tilde{D}_t^2$ remains low regardless of the presence of an announcement correspond to "never-takers," while regimes where $\tilde{D}_t^2$ remains high even in the absence of an announcement correspond to "always-takers." Noting that $D_t = \tilde{D}_t^2$ in this context, we can apply Theorem 2.1 to identify the compliers individually. We do so in the empirical application in Section 7.

It is important to consider how the interpretation of the causal effect identified by $\beta^*_{\pi,t,0}$ in Corollary 3.1 varies with the functional relationship between $\tilde{Y}_t$ and $\tilde{D}_t$. Let us begin with the linear case with stable parameters as in (3.1). From (3.4), simple algebra shows that $\beta^*_{\pi,t,0}$ reduces to β_0 when the denominator of (3.5) is nonzero, which means that Rigobon's estimator identifies the causal effect of the policy (i.e., the slope coefficient in (3.1)). This result does not generally extend to the case where β_0 is time-varying or the first-stage is zero. At most one could identify a π-LATE provided that Rigobon's estimator is computed over the sub-population where the first-stage is nonzero. We will return to this in Section 3.2.

Let us turn to analyzing the consequences of nonlinearities. When D_t and the shock η_t are additively separable (i.e., $Y_t = \varphi_D(D_t) + \varphi_\eta(\eta_t)$ for some nonlinear functions $\varphi_D(\cdot)$ and $\varphi_\eta(\cdot)$), Kolesár and Plagborg-Møller (2025) show that the estimand resulting from a regression of Y_t on D_t using $Z_t = (W_t - \mathbb{E}(W_t))D_t$ as an instrument for which $\mathrm{Cov}(D_t^2, W_t) \neq 0$ identifies a weighted average of marginal effects of the policy shock e_t with weights that are not guaranteed to be positive. As a result, the researcher may infer an incorrect sign for the marginal effects. Thus, this estimand is not weakly causal [cf. Blandhol, Bonney, Mogstad, and Torgovitsky (2025)]. The authors also note that for the case $Y_t = e_t\varphi_\eta(\eta_t)$ with $\mathbb{E}[\varphi_\eta(\eta_t)] = 0$ and $e_t \perp \eta_t$ the estimand is nonzero while the true causal effect of the policy shock is zero since $\mathbb{E}[Y_t | e_t] = 0$.

Corollary 3.1 provides even more negative news about the effect of nonlinearities for heteroskedasticity-based identification than that shown by Kolesár and Plagborg-Møller (2025): in a general nonparametric model, Corollary 3.1 implies that Rigobon's Wald estimand $\beta^*_{\pi,t,0}$, which is in general different from the IV estimand examined by Kolesár and Plagborg-Møller (2025), does not necessarily equal a weighted average of marginal effects. The intuition is that the instrument affects $\mathrm{Var}(D_t)$ and not $\mathbb{E}(D_t)$, so variation in the instrument induces exogenous variation in D_t^2, which has a causal effect on D_tY_t not just Y_t. In short, it is generally difficult to interpret $\beta^*_{\pi,t,0}$ when the true model is nonlinear. Thus, we concur with the

recommendation of Kolesár and Plagborg-Møller (2025) that the linearity assumption should be checked carefully when using heteroskedasticity-based identification. This is likely even more important in the context of SVARs and local projections than in the the current event study setting since the former aggregates data over a month or a quarter while the latter uses relatively higher frequency data (e.g., a 30-minute or 1-day change in policy and outcome variables around an announcement), where linearity may be a more credible assumption since a nonlinear function can be locally well approximated by a linear one.[15]

3.2 Weak or Lack of Identification and the Usefulness of π-LATE

The key identification condition that the volatility of the policy variable is higher during FOMC announcement days appears reasonable in principle, since each announcement day is likely to be associated with substantial monetary news. However, the volatility of monetary policy variables can be high for other reasons. There are multi-year periods during which the volatility of several macroeconomic variables is elevated. In this case, general equilibrium factors dominate the news specific to the announcement. For example, during the 2007-09 financial crisis and the Covid-19 pandemic, volatility was high across many macroeconomic and financial variables. These facts pose serious challenges for identification, as the first-stage condition may not hold for all t. To see this, examine the denominator of β_0 in (3.3). If the first-stage does not hold for all t, we may have

$$T_P^{-1} \sum_{t \in \mathbf{P}} \mathrm{Var}\left(\tilde{D}_t\right) - T_C^{-1} \sum_{t \in \mathbf{C}} \mathrm{Var}\left(\tilde{D}_t\right) \approx 0, \tag{3.6}$$

which would render the estimate of the average treatment effect highly imprecise.

Using an F-test for weak identification, Lewis (2022) shows that the monetary policy effects based on a 1-day window in Nakamura and Steinsson (2018) appear to be weakly-identified. We show that this arises from significant time variation in the volatility of the policy variable within both policy and control samples. Figure 2 plots $\tilde{D}_t$ (2-Year Treasury yields) for the control and policy samples. The policy sample includes all regularly scheduled FOMC meeting days from 1/1/2000 to 3/19/2014. The control sample includes all Tuesdays

[15] The differences between our results on identification via heteroskedasticity and those in Kolesár and Plagborg-Møller (2025) are: (i) they consider the causal effect of the policy shock e_t while we consider the causal effect of the policy variable D_t; (ii) they consider an IV estimand while we explicitly consider Rigobon's estimand motivated by $\Delta\Sigma^{(1,2)}/\Delta\Sigma^{(2,2)}$ in (3.3) and as usually implemented in empirical work based on event studies; (iii) they consider specific nonlinear restrictions and allow the instrument to be continuous whereas we allow for a general nonlinear model and consider a binary instrument as motivated by (3.3).

and Wednesdays that are not FOMC meeting days between 1/1/2000 and 12/31/2012.

There appear to be multiple volatility regimes. Using the structural break test from Casini and Perron (2024), which allows for stable or smoothly varying volatility under the null and abrupt breaks under the alternative, we detect three breaks in the control sample. The first break (April 24, 2007) marks the start of the 2007-09 financial crisis. The second (July 28, 2009) captures the crisis period itself, characterized by the highest volatility. Afterward, volatility returns to pre-crisis levels until the third break (February 2, 2011), which aligns with the zero lower bound (ZLB) period and the start of unconventional monetary policy. The final regime shows the lowest volatility, reflecting initial policy effects and stabilization.[16]

These findings show significant time variation in $\text{Var}(\tilde{D}_t)$. In the second regime, control-sample volatility is close to the policy-sample average, contributing to the weak identification in (3.6). Nakamura and Steinsson (2018) find their estimates imprecise and not economically meaningful for some of the interest rates they use as outcome variables. Lewis (2022) reports a first-stage F-statistic of 8.11—well below the 23 critical value—suggesting weak identification.

We propose to focus on π-LATE. The fraction π_0 of the sample (i.e., all $t \in \mathbf{S}_{0,T}$) that has a first-stage corresponds to the regimes in the control sample where $\text{Var}(\tilde{D}_t)$ is low (relative to its average level). For example, it is likely that the regime $[\widehat{T}_1 + 1, \widehat{T}_2]$ does not belong to $\mathbf{S}_{0,T}$ since $\text{Var}(\tilde{D}_t)$ within this regime appears close to the average volatility in the policy sample. By construction, it is easier to identify π-LATE than full sample LATE. The usefulness of π-LATE depends on the magnitude of π_0: a small π_0 implies that identification is achievable only in a small portion of the population, whereas a large π_0 indicates that the identified π-LATE is representative of a substantial part of the population.[17]

The π-LATE parameter is the same as the LATE parameter (3.3) in Section 3.1 but instead of supposing that a first-stage exists, only uses observations with a nonzero first-stage. Let $T_{P,S}$ denote the number of days in $\mathbf{S}_{0,T}$ that contain an FOMC announcement, and let $T_{C,S}$ the number of days in $\mathbf{S}_{0,T}$ that do not contain an FOMC announcement. This means $T_{P,S} + T_{C,S} = \pi_0 T$.[18] Let $\Sigma_{i,S}$ denote the covariance matrix of $[\tilde{Y}_t, \tilde{D}_t]'$ in the subsample

[16] We do not test for breaks in the policy sample due to small size ($T_P = 74$), treating it as a single regime.

[17] It is possible that in practice the π_0 fraction of the sample contains a mixture of strong and weak identification. We discuss weak identification in the context of π-LATE formally in Section 6.

[18] For notational simplicity we assume that $\pi_0 T$ is an integer so that we avoid using the notation $\lfloor \pi_0 T \rfloor$, where $\lfloor \cdot \rfloor$ denotes the largest smaller integer function.

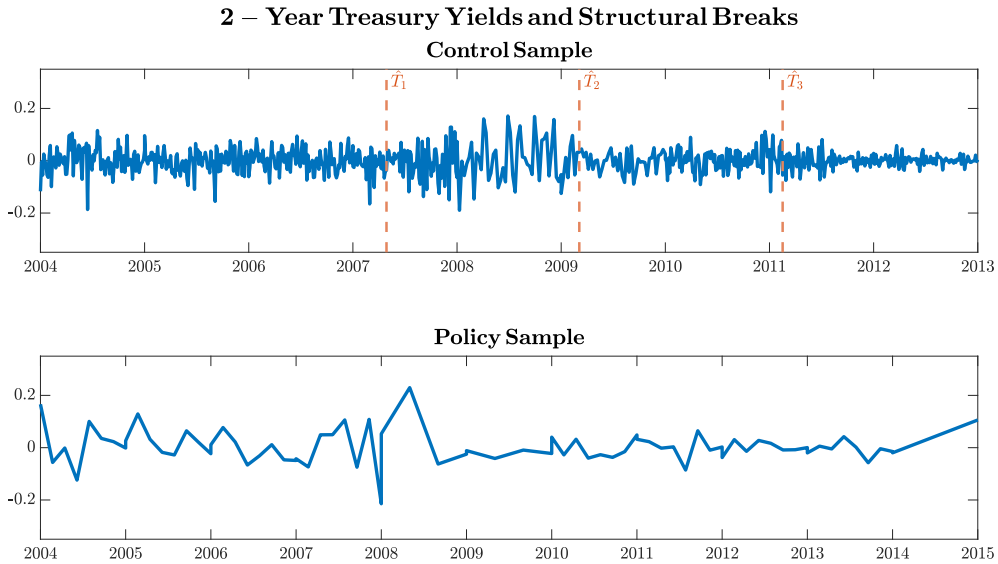

Figure 2: Plot of 2-years Treasury yields in control (top panel) and policy sample (bottom panel). Vertical broken lines are the estimated break dates using Casini and Perron's (2024) test.

$i = \mathbf{P}$, $\mathbf{C}$ using only observations $t \in \mathbf{S}_{0,T}$. We have

$$\widetilde{\beta}_{\pi,0} = \frac{\Delta\Sigma_S^{(1,2)}}{\Delta\Sigma_S^{(2,2)}} = \frac{T_{P,S}^{-1}\sum_{t\in\mathbf{P_S}}\mathrm{Cov}\left(\tilde{Y}_t,\,\tilde{D}_t\right) - T_{C,S}^{-1}\sum_{t\in\mathbf{C_S}}\mathrm{Cov}\left(\tilde{Y}_t,\,\tilde{D}_t\right)}{T_{P,S}^{-1}\sum_{t\in\mathbf{P_S}}\mathrm{Var}\left(\tilde{D}_t\right) - T_{C,S}^{-1}\sum_{t\in\mathbf{C_S}}\mathrm{Var}\left(\tilde{D}_t\right)}, \qquad \text{where} \tag{3.7}$$

$$\Delta\Sigma_S = \Sigma_{P,S} - \Sigma_{C,S} = \frac{\sigma_{e,P}^2 - \sigma_{e,C}^2}{(1-a_1\beta_0)^2}\begin{bmatrix}\beta_0^2 & \beta_0\\ \beta_0 & 1\end{bmatrix},$$

with $\mathbf{P_S} = \mathbf{P} \cap \mathbf{S}_{0,T}$ and $\mathbf{C_S} = \mathbf{C} \cap \mathbf{S}_{0,T}$. Proceeding as for LATE, the Wald estimand is

$$\widetilde{\beta}^*_{\pi,t,0} = \frac{\mathbb{E}\left(\tilde{D}_t\tilde{Y}_t|t\in\mathbf{P_S}\right) - \mathbb{E}\left(\tilde{D}_t\tilde{Y}_t|t\in\mathbf{C_S}\right)}{\mathbb{E}\left(\tilde{D}_t^2|t\in\mathbf{P_S}\right) - \mathbb{E}\left(\tilde{D}_t^2|t\in\mathbf{C_S}\right)}, \tag{3.8}$$

to which Corollary 3.1 immediately applies without the (now redundant) qualifier "for $t \in \mathbf{S}_{0,T}$." Under within subsample covariance stationarity, the right-hand side of (3.8) is equal to that of (3.7), implying $\tilde{\beta}_{\pi,0} = \tilde{\beta}^*_{\pi,t,0}$. Therefore, $\tilde{\beta}_{\pi,0}$ identifies the same π-LATE, as defined explicitly in Corollary 3.1. π-LATE is the average treatment effect for the sub-population for which a first-stage holds: observations for which $\tilde{D}_t^2$ is induced to be higher by the announcement (i.e., the sub-population of compliers in $\mathbf{S}_{0,T}$).

If the treatment effect is constant across the population [e.g., as in (3.1)], then the π-LATE for the sub-population $\mathbf{S}_{0,T}$ is equal to both the LATE and ATE in the full population. To determine which treatment effect is identified, we must determine which parts of the sample belong to $\mathbf{S}_{0,T}$.We discuss this in Sections 4-5.

4 Testing for Full Population Identification Failure

In this section, we introduce a test of the null hypothesis that no subpopulation exists for which a LATE can be identified, even weakly. In other words, the test assesses whether identifying a sub-population LATE is possible at all. However, we strongly caution against using this as a pretest before estimation or inference, as doing so may introduce pretest bias and invalidates standard inference unless the inference method is modified to account for the pretest [see, e.g., Andrews (2018)]. Instead, the test should be viewed as a diagnostic tool for evaluating whether there is evidence of identifiable sub-population LATEs in a given application. We apply it for this purpose to several existing studies that appear to face identification challenges. Notably, such a pretest is unnecessary for conducting identification-robust inference on sub-population LATEs, which we discuss in Section 6.

In accord with the analysis of Section 2, consider an IV regression model with a single endogenous variable and multiple instruments. In matrix format, the structural equation is

$$Y = D\beta + X\gamma_1 + u, \qquad t = 1, \ldots, T, \tag{4.1}$$

where Y is a $T \times 1$ vector of outcome variables, D is $T \times 1$ vector of endogenous variables, X is a $T \times p$ matrix of p exogenous regressors, u is a $T \times 1$ vector of error terms, and $\beta \in \mathbb{R}$ and $\gamma_1 \in \mathbb{R}^p$ are unknown parameters. The reduced-form equation is

$$D_t = Z_t'\theta \mathbf{1}\{t \in \mathbf{S}_{0,T}\} + X_t'\gamma_2 + e_t, \tag{4.2}$$

where Z_t is a $q \times 1$ vector of instruments, e_t is an error term, and $\theta \in \mathbb{R}^q$ and $\gamma_2 \in \mathbb{R}^p$ are unknown parameters. For $t \notin \mathbf{S}_{0,T}$, the instrument Z_t is irrelevant. For $t \in \mathbf{S}_{0,T}$, the instrument Z_t is relevant if $\theta \neq 0$. We assume that $|\mathbf{S}_{0,T}| = \pi_0 T$ for some $\pi_0 \in (0, 1]$, noting that this is without loss of generality since it does not rule out complete identification failure

which occurs when $\theta = 0$ for any $\pi_0 \in (0, 1]$. The hypothesis testing problem is

$$H_{\theta,0} : \theta = 0 \quad \text{versus} \quad H_{\theta,1} : \theta \neq 0.$$

We discuss both the cases for which the sub-population $\mathbf{S}_{0,T}$ is known and unknown. For the sake of the exposition, we focus on homogeneous θ in $\mathbf{S}_{0,T}$.[19]

Consider the $(\pi T \times T)$ selection matrix S_T that selects the πT rows of a matrix corresponding to the indices in $\mathbf{S}_T$. That is, for an arbitrary $T \times k$ matrix A, $S_T A$ is the $(\pi T \times k)$ matrix whose elements are the rows of A that correspond to the indices in $\mathbf{S}_T$. For example, if $\mathbf{S}_T = \{1, \ldots, 0.25T, 0.75T + 1, \ldots, T\}$,

$$S_T A = \left[A^{(1,:)\prime} : \cdots : A^{(0.25T,:)\prime} : A^{(0.75T+1,:)\prime} : \cdots : A^{(T,:)\prime}\right]',$$

where $A^{(r,:)}$ denotes the rth row of the matrix A. Using the standard projection matrix notation, $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$, let $\widetilde{A}(S_T) = M_{S_T X} S_T A$ for any arbitrary $T \times k$ matrix A. The following F test statistic is useful for testing whether $\theta = 0$ in the regression (4.2) when the sub-population $\mathbf{S}_{0,T}$ is known:

$$F_T(\mathbf{S}_T) = \frac{\widetilde{D}(S_T)' \widetilde{Z}(S_T) \widehat{J}(S_T)^{-1} \widetilde{Z}(S_T)' \widetilde{D}(S_T)}{q(\pi T - p - q)},$$

for $\mathbf{S}_T = \mathbf{S}_{0,T}$ and $Z = [Z_1 : \cdots : Z_T]'$ and $\widehat{J}(S_T)$ a consistent estimate of the long-run variance,

$$\lim_{T \to \infty} (T\pi)^{-1} \mathrm{Var}(\widetilde{Z}(S_T)' S_T e)$$

with $e = [e_1 : \cdots : e_T]'$. HAC or DK-HAC estimators can be used to estimate the long-run variance [cf. Andrews (1991), Casini (2023) and Newey and West (1987)].

For the case of an unknown sub-population, we follow the structural break literature and search for maximal identification strength over all sub-populations of minimal size $\pi_L T$ that can be partitioned into m distinct smaller sub-populations, where $\pi_L > 0$ and $1 \leq m \leq m_+$ for some upper bound on the number of regimes $m_+ > 0$:

$$F_T^* = \sup_{\pi \in [\pi_L, 1]} \max_{1 \leq m \leq m_+} \sup_{\mathbf{S}_T \in \Xi_{\epsilon,\pi,m,T}} F_T(\mathbf{S}_T),$$

[19] We could allow for $\theta_t \neq 0$ for $t \in \mathbf{S}_{0,T}$ at the expense of additional notation and longer proofs, though the key insights would not change. Actually, the computational procedures we develop to implement our methods allow $\theta_t \neq 0$ for $t \in \mathbf{S}_{0,T}$.

where $\Xi_{\epsilon,\pi,m,T}$ denotes the set of all possible partitions of a fraction π of $\{1,\ldots,T\}$ that involve m regimes $((\lambda_{L,1}T,\, \lambda_{R,1}T),\ldots,(\lambda_{L,m}T,\, \lambda_{R,m}T))$ for $\lambda_{L,i},\, \lambda_{R,i} \in [0,\, 1]$ such that (i) $\lambda_{L,i} < \lambda_{R,i}$ for all i, (ii) $\lambda_{R,i} < \lambda_{L,i+1}$ for $i = 1,\ldots,\, m-1$, (iii) $|\lambda_{R,i} - \lambda_{L,i}| \geq \epsilon$ for all i and some (small) $\epsilon > 0$ and (iv) $\sum_{i=1}^{m}(\lambda_{R,i} - \lambda_{L,i}) = \pi$. Conditions (i) and (ii) correspond to $T\lambda_{L,i}$ ($T\lambda_{R,i}$) denoting the start (end) date of regime i within the sub-population $\mathbf{S}_T$ while condition (iii) implies that each regime involves a non-negligible fraction of the sample. The statistic F_T^* thus implicitly searches for maximal identification strength over all possible sub-populations of size $\pi_L T$ and larger with less than m_+ distinct regimes that are at least a ϵ fraction of the overall sample size.

The tuning parameters π_L and ϵ determine the types of sub-populations for which the test can detect identification: smaller values of π_L allow detection in smaller sub-populations, while smaller values of ϵ enable detection in sub-populations with shorter regimes. The choice of these lower bounds should be guided by the empirical context, reflecting the smallest sub-population and regime sizes for which LATE inference remains meaningful in the application.[20] In our simulations and empirical applications we set $\pi_L = 0.6$ and $\epsilon = 0.05$.

For X_t' the tth row of X, let $w_t = (X_t',\, Z_t')'$ and $W_r(\cdot)$ denote a r-vector of independent Wiener processes on $[0,\, 1]$. We derive the asymptotic null distributions of $F_T(\mathbf{S}_T)$ and F_T^* under the following standard high-level assumptions that permit both heteroskedastic and serially correlated errors. Sufficient conditions for them can be found in the supplement.

Assumption 4.1. *$T^{-1}\sum_{t=1}^{\lfloor Ts \rfloor} w_t w_t' \xrightarrow{\mathbb{P}} sQ$, uniformly in $s \in [0,\, 1]$ for some p.d. matrix Q.*

Assumption 4.2. *$T^{-1/2}\sum_{t=1}^{\lfloor Ts \rfloor} w_t e_t \Rightarrow \Omega_{we}^{1/2} W_{p+q}(s)$ for some p.d. variance matrix Ω_{we}.*

Assumption 4.3. *$\widehat{J}(S_T)$ is p.d. for all T, $\mathbf{S}_T \in \Xi_{\epsilon,\pi,m,T}$ and $\widehat{J}(S_T) \xrightarrow{\mathbb{P}} \lim_{T\to\infty} T^{-1}\mathrm{Var}(e' S_T' \widetilde{Z}(S_T))$ uniformly in $\mathbf{S}_T \in \Xi_{\epsilon,\pi,m,T}$.*

Theorem 4.1. *Let Assumptions 4.1-4.3 hold. Under $H_{\theta,0}$,*

$$F_T(\mathbf{S}_T) \Rightarrow F(\mathbf{S}) \quad \text{if} \quad \mathbf{S}_T \in \Xi_{\epsilon,\pi,m,T}, \qquad \text{and} \qquad F_T^* \Rightarrow \sup_{\pi\in[\pi_L,\,1]} \max_{1\leq m\leq m_+} \sup_{\mathbf{S}\in\Xi_{\epsilon,\pi,m}} F(\mathbf{S}),$$

where $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, $\Xi_{\epsilon,\pi,m} = \lim_{T\to\infty} T^{-1}\Xi_{\epsilon,\pi,m,T}$ and

$$F(\mathbf{S}) = \frac{1}{q\pi}\sum_{i=1}^{m} \left\|\left(W_q(\lambda_{R,i}) - W_q(\lambda_{L,i})\right)\right\|^2.$$

[20] In the structural break literature, common recommendations for ϵ are 0.05, 0.10 and 0.15. See Casini and Perron (2019) for a review.

When $\pi = 1$ ($\pi_L = 1$ and $m_+ = 1$), $F_T(\mathbf{S}_T)$ (F_T^*) reduces to the usual first-stage F-statistic for $\theta = 0$ in (4.2). For $\pi \in (0, 1)$ ($\pi_L \in (0, 1)$), the consistency of tests against $H_{\theta,1}$ using $F_T(\mathbf{S}_{0,T})$ (F_T^*) follows from similar arguments as for the $\pi_0 = 1$ case. The asymptotic null distributions of both $F(\mathbf{S})$ and F_T^* are free of nuisance parameters. The critical values are obtained via simulations and reported in Table 4 for up to $m_+ = 6$ and up to $q = 6$.

5 Estimation of LATE and Identified Sub-Populations

We discuss estimation of the LATE parameter β in (4.1) in both the cases of a known and unknown sub-population $\mathbf{S}_{0,T}$, as well as estimation of $\mathbf{S}_{0,T}$ itself in the latter case. When $\mathbf{S}_{0,T}$ is known, estimation of β is an application of IV estimation for which $Z_t\mathbf{1}\{t \in \mathbf{S}_{0,T}\}$ is treated as the vector of instruments. Let this estimator be denoted as $\widehat{\beta}(\mathbf{S}_{0,T})$.

On the other hand, when the sub-population $\mathbf{S}_{0,T}$ is unknown, we must estimate it first. Although $\mathbf{S}_{0,T}$ can be estimated consistently in the special case of a binary instrument under the conditions of Proposition 2.2 and Theorem 2.1, it can also be estimated more generally. We discuss two methods. The first is more computationally straightforward but the second is more efficient because it uses the information in both structural and reduced-form equations (4.1)-(4.2). We follow the structural change literature and assume that π_0 and m_0 are known, i.e., the practitioner has previously used the tests from Section 4 to determine π_0 and m_0.

We begin with the first estimator. Consider the $T \times T$ matrix C_T that selects the πT rows of a matrix corresponding to the indices in $\mathbf{S}_T$ while setting the remaining $(1-\pi)T$ rows to zero. For example, for a $T \times k$ matrix A, if $\mathbf{S}_T = \{1, \ldots, 0.25T, 0.75T+1, \ldots, T\}$,

$$C_T A = \left[A^{(1,:)\prime} : \cdots : A^{(0.25T,:)\prime} : 0_{k\times 1} : \cdots : 0_{k\times 1} : A^{(0.75T+1,:)\prime} : \cdots : A^{(T,:)\prime}\right]'.$$

Let $\overline{A}(C_T) = M_X C_T A$ so that for a given $\mathbf{S}_T$, the OLS estimators of θ and γ_2 in (4.2) can be expressed as $\widehat{\theta}_{OLS}(\mathbf{S}_T) = (\overline{Z}(C_T)'\overline{Z}(C_T))^{-1}\overline{Z}(C_T)'D$ and $\widehat{\gamma}_{2,OLS}(\mathbf{S}_T) = (X'M_{C_T Z}X)^{-1}X'M_{C_T Z}D$. Our first estimator of $\mathbf{S}_{0,T}$ minimizes the sum of squared residuals of the reduced-form:

$$\widehat{\mathbf{S}}_{T,OLS} = \underset{\mathbf{S}_T \in \Xi_{\epsilon,\pi_0,m_0,T}}{\operatorname{argmin}} \left(D - C_T Z\widehat{\theta}_{OLS}(\mathbf{S}_T) - X\widehat{\gamma}_{2,OLS}(\mathbf{S}_T)\right)' \left(D - C_T Z\widehat{\theta}_{OLS}(\mathbf{S}_T) - X\widehat{\gamma}_{2,OLS}(\mathbf{S}_T)\right).$$

Correspondingly, we estimate β with $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$.

For the second estimator of the sub-population $\mathbf{S}_{0,T}$, we propose a GLS criterion that minimizes an efficiently weighted combination of the sum of squared residuals of both the

reduced-form representation of the structural equation (4.1) and the reduced-form equation (4.2). That is, the system of equations (4.1)-(4.2) can be written in reduced-form as

$$\vec{y} = W(\mathbf{S}_T)\xi + \varepsilon, \tag{5.1}$$

where $\vec{y} = (Y', D')'$, $W(\mathbf{S}_{0,T}) = I_2 \otimes [C_{0,T}Z : X]$, $\xi = (\beta\theta', \gamma_1' + \beta\gamma_2', \theta', \gamma_2')'$ and $\varepsilon = (u' + \beta e', e')'$ with $C_{0,T}$ defined as C_T but corresponding to the indices in $\mathbf{S}_{0,T}$. This is a system of two seemingly unrelated regressions. Let

$$\widehat{\xi}_{FGLS}(\mathbf{S}_T) = (W(\mathbf{S}_T)'\widehat{\Omega}_\varepsilon(\mathbf{S}_T)^{-1}W(\mathbf{S}_T))^{-1}W(\mathbf{S}_T)'\widehat{\Omega}_\varepsilon(\mathbf{S}_T)^{-1}\vec{y},$$

denote a feasible GLS estimator of ξ, where $\widehat{\Omega}_\varepsilon(\mathbf{S}_T)$ is a consistent estimator of $\mathbb{E}[\varepsilon\varepsilon'|W(\mathbf{S}_T)]$. Our second estimator of $\mathbf{S}_{0,T}$ minimizes the following GLS criterion based upon (5.1):

$$\widehat{\mathbf{S}}_{T,FGLS} = \underset{\mathbf{S}_T \in \Xi_{\epsilon,\pi_0,m_0,T}}{\operatorname{argmin}} \left(\vec{y} - W(\mathbf{S}_T)\widehat{\xi}_{FGLS}(\mathbf{S}_T)\right)' \widehat{\Omega}_{\varepsilon,\mathbf{S}}^{-1} \left(\vec{y} - W(\mathbf{S}_T)\widehat{\xi}_{FGLS}(\mathbf{S}_T)\right).$$

Correspondingly, we estimate β with $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$. In order for $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ to be provably more efficient than $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$, $\widehat{\Omega}_{\varepsilon,\mathbf{S}}$ must be a consistent estimator of $\mathbb{E}[\varepsilon\varepsilon'|W(\mathbf{S}_{0,T})]$. When ε_t does not exhibit conditional serial correlation or heteroskedasticity, i.e., $\mathbb{E}[\varepsilon\varepsilon'|W(\mathbf{S}_{0,T})] = \Sigma_\varepsilon \otimes I_T$, this is feasible since one could simply use $\widehat{\Omega}_{\varepsilon,\mathbf{S}} = \widehat{\Sigma}_\varepsilon \otimes I_T$, where $\widehat{\Sigma}_{\varepsilon,i,j} = (T-q-p)^{-1}\hat{\varepsilon}^{i\prime}\hat{\varepsilon}^j$ for $i,j = 1,2$ with $\hat{\varepsilon}^1$ ($\hat{\varepsilon}^2$) equal to the first (last) T elements of $\vec{y} - W(\widehat{\mathbf{S}}_{T,OLS})\widehat{\xi}_{OLS}(\widehat{\mathbf{S}}_{T,OLS})$, as is standard in seemingly unrelated regression. For serially dependent ε_t, consistent estimation of $\mathbb{E}[\varepsilon\varepsilon'|W(\mathbf{S}_{0,T})]$ requires a correctly-specified model for the dependence in ε_t, a strong assumption in some empirical applications. In the supplement Casini et al. (2025b) we present the consistency results about $\widehat{\mathbf{S}}_{T,OLS}$, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$, $\widehat{\mathbf{S}}_{T,FGLS}$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$.

In model (4.1) the LATE parameter β is constant, so π-LATE is the full population LATE and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ are consistent for the LATE parameter β. They can be precise estimates even when a first-stage F test detects full sample weak identification because they use the most-strongly identified subsample of the data. When the model (4.1) is misspecified, so that LATEs may be nonlinear and time-varying, the estimators $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ are still consistent for a weighted average the of the LATEs in the $\mathbf{S}_{0,T}$ subsample if the $\mathbf{S}_{0,T}$ subsample exhibits strong identification.

The estimators $\widehat{\mathbf{S}}_{T,OLS}$ and $\widehat{\mathbf{S}}_{T,FGLS}$ and the test statistic F_T^* solve an optimization problem over many partitions. This is computationally more complex than problems in

the structural breaks literature, as it involves optimizing both over sample partitions and identification strength. We address this challenge by proposing an efficient algorithm based on dynamic programming, extending the approach of Bai and Perron (2003) to our setting.[21]

6 Identification-Robust Inference

We consider tests on β in (4.1) that are robust to weak identification in both the cases for which the sub-population $\mathbf{S}_{0,T}$ is known and unknown. The hypothesis testing problem is $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$. Here we present results for the case of unknown sub-population $\mathbf{S}_{0,T}$ and weak instruments. We also briefly discuss the case of known $\mathbf{S}_{0,T}$ and strong instruments and defer their formal treatment to the supplement. We rewrite (5.1) as

$$y = \overline{Z}(C_{0,T})\theta a' + X\eta + v, \text{ where } y = [Y : D]\,,\ v = [v_1 : e]\,,\ a = (\beta,\, 1)'\,,\ \eta = [\gamma : \phi]\,, \quad (6.1)$$

with $v_1 = u + \beta e$, $\gamma = \gamma_1 + \phi\beta$ and $\phi = \gamma_2 + (X'X)^{-1}X'C_{0,T}Z\theta$. When $\mathbf{S}_{0,T}$ is known, it is straightforward to use existing tests in the identification-robust linear IVs literature to test H_0 [cf. Anderson and Rubin (1949), Andrews, Moreira, and Stock (2006), Kleibergen (2002) and Moreira (2003)]. However, Proposition S.B.1 in the supplement shows that $Z'M_Xy$ is not a sufficient statistic for $(\beta, \theta')'$ but $\overline{Z}(C_{0,T})'y$ is, implying that existing tests suffer a loss in efficiency because they treat Z rather than $C_{0,T}Z$ as the matrix of IVs. Efficient tests are therefore functions of $\overline{Z}(C_{0,T})'y$. Magnusson and Mavroeidis (2014) consider a model similar to (6.1). Our model specifies that θ is nonzero in the sub-population $\mathbf{S}_{0,T}$ and is zero in $\mathbf{S}^c_{0,T}$ where $\mathbf{S}^c_{0,T}$ is the complement of $\mathbf{S}_{0,T}$. Magnusson and Mavroeidis (2014) allow the first-stage coefficient θ_t to be generally time-varying for some of their tests. Their tests are based on the full sample of observations whereas our tests are based on a lower-dimensional statistic since we do not use the sub-population $\mathbf{S}^c_{0,T}$. This allows us to obtain gains in efficiency.

When $\mathbf{S}_{0,T}$ is known we can apply the results of Andrews, Moreira, and Stock (2006) to form identification-robust tests of H_0 vs H_1 that are functions of $\overline{Z}(C_{0,T})'y$ and are robust to both heteroskedasticity and autocorrelation (HAR) in the reduced-form errors $\{v_t\}$. Suppose $\widehat{\Sigma}_{N_1}(\mathbf{S}_{0,T})$, $\widehat{\Sigma}_{N_1,N_2}(\mathbf{S}_{0,T})$ and $\widehat{\Sigma}_{N_2}(\mathbf{S}_{0,T})$ are consistent estimators of $\Sigma_{N_1}(\mathbf{S}_0)$, $\Sigma_{N_1,N_2}(\mathbf{S}_0)$ and

[21] While Antoine and Boldea (2018) consider the case of a single break, and Magnusson and Mavroeidis (2014) study a related context, neither provide a computational solution—referring to the problem as "computationally demanding."

$\Sigma_{N_2}(\mathbf{S}_0)$ under H_0, where these latter quantities are defined by

$$\Sigma_{v\overline{Z}}(\mathbf{S}_0) = \begin{bmatrix} \Sigma_{N_1}(\mathbf{S}_0) & \Sigma_{N_1N_2}(\mathbf{S}_0)' \\ \Sigma_{N_1N_2}(\mathbf{S}_0) & \Sigma^*_{N_2}(\mathbf{S}_0) \end{bmatrix}, \tag{6.2}$$
$$\Sigma_{N_2}(\mathbf{S}_0) = \Sigma^*_{N_2}(\mathbf{S}_0) - \Sigma_{N_1N_2}(\mathbf{S}_0)\,\Sigma^{-1}_{N_1}(\mathbf{S}_0)\,\Sigma_{N_1N_2}(\mathbf{S}_0)'$$

for $\Sigma_{v\overline{Z}}(\mathbf{S}_0) = \Sigma_{v\overline{Z}}(\mathbf{S}_0, \mathbf{S}_0)$, with

$$\Sigma_{v\overline{Z}}(\mathbf{S}, \mathbf{S}') = \lim_{T\to\infty} \operatorname{Cov}\left(T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix} v_t' b_0 \overline{Z}_t(C_T) \\ v_t'\Sigma_v^{-1} a_0 \overline{Z}_t(C_T) \end{bmatrix}, T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix} v_t' b_0 \overline{Z}_t(C_T') \\ v_t'\Sigma_v^{-1} a_0 \overline{Z}_t(C_T') \end{bmatrix}\right)$$

for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$ $b_0 = (1, -\beta_0)'$ and $a_0 = (\beta_0, 1)'$, where and v_t and $\overline{Z}_t(C_T)$ are the tth rows v and $\overline{Z}(C_T)$.[22] Let $\widehat{\Sigma}_v(\mathbf{S}_{0,T}) = (T-q-p)^{-1}\,\widehat{v}(\mathbf{S}_{0,T})'\,\widehat{v}(\mathbf{S}_{0,T})$ with $\widehat{v}(\mathbf{S}_{0,T}) = y - P_{\overline{Z}(C_{0,T})}y - P_X y$. Define

$$N_{1,T}(\mathbf{S}_{0,T}) = \widehat{\Sigma}^{-1/2}_{N_1}(\mathbf{S}_{0,T})\,T^{-1/2}\overline{Z}(C_{0,T})'\,y b_0 \qquad \text{and} \tag{6.3}$$
$$N_{2,T}(\mathbf{S}_{0,T}) = \widehat{\Sigma}^{-1/2}_{N_2}(\mathbf{S}_{0,T})\left(T^{-1/2}\overline{Z}(C_{0,T})'\,y\widehat{\Sigma}^{-1}_v(\mathbf{S}_{0,T})\,a_0 - \widehat{\Sigma}_{N_1N_2}(\mathbf{S}_{0,T})\,\widehat{\Sigma}^{-1/2}_{N_1}(\mathbf{S}_{0,T})\,N_{1,T}(\mathbf{S}_{0,T})\right).$$

Consider the following HAR versions of the Anderson-Rubin (AR), Lagrange multiplier (LM) and likelihood ratio statistics based on the sufficient statistic $\overline{Z}(C_{0,T})'y$:

$$AR_T(\mathbf{S}_{0,T}) = M_{1,T}(\mathbf{S}_{0,T}), \qquad LM_T(\mathbf{S}_{0,T}) = \frac{M_{1,2,T}(\mathbf{S}_{0,T})^2}{M_{2,T}(\mathbf{S}_{0,T})}, \tag{6.4}$$
$$LR_T(\mathbf{S}_{0,T}) = \frac{1}{2}\left(M_{1,T}(\mathbf{S}_{0,T}) - M_{2,T}(\mathbf{S}_{0,T}) + \sqrt{(M_{1,T}(\mathbf{S}_{0,T}) - M_{2,T}(\mathbf{S}_{0,T}))^2 + 4M_{1,2,T}(\mathbf{S}_{0,T})^2}\right),$$

where $M_{1,T}(\mathbf{S}_{0,T}) = N_{1,T}(\mathbf{S}_{0,T})'\,N_{1,T}(\mathbf{S}_{0,T})$, $M_{1,2,T}(\mathbf{S}_{0,T}) = N_{1,T}(\mathbf{S}_{0,T})'\,N_{2,T}(\mathbf{S}_{0,T})$ and $M_{2,T}(\mathbf{S}_{0,T}) = N_{2,T}(\mathbf{S}_{0,T})'\,N_{2,T}(\mathbf{S}_{0,T})$. The conditional likelihood ratio (CLR) test of level α rejects H_0 when $LR_T(\mathbf{S}_{0,T}) > \kappa_\alpha(N_{2,T}(\mathbf{S}_{0,T}))$, where the critical value function $\kappa_\alpha(\cdot)$ is defined such that $\kappa_\alpha(n_2)$ is the $1-\alpha$ quantile of the large-sample conditional distribution of $LR_T(\mathbf{S}_{0,T})$ under H_0, given $N_{2,T}(\mathbf{S}_{0,T}) = n_2$:

$$\frac{1}{2}\left(\mathcal{Z}_q'\mathcal{Z}_q - n_2'n_2 + \sqrt{\left(\mathcal{Z}_q'\mathcal{Z}_q - n_2'n_2\right)^2 + 4(\mathcal{Z}_q'n_2)^2}\right),$$

[22] See the supplement for details on how to construct these estimators and for consistency results.

where $\mathcal{Z}_q \sim \mathcal{N}(0, I_q)$. The critical value function $\kappa_\alpha(\cdot)$ is approximated in Moreira (2003). The LM and AR tests reject H_0 when $LM_T > \chi_1^2(1-\alpha)$ and $AR_T > \chi_q^2(1-\alpha)$, where $\chi_q^2(1-\alpha)$ denotes the $1-\alpha$ quantile of a chi-squared distribution with q degrees of freedom.

When $\mathbf{S}_{0,T}$ is known the results of Andrews et al. (2006) imply that the CLR, LM and AR tests have limiting null rejection probabilities equal to α under weak IV asymptotics, $\theta = c/T^{1/2}$ for some nonstochastic $c \in \mathbb{R}^q$, under a weakening of Assumptions 6.1-6.4 below for which these assumptions need only hold pointwise in $\mathbf{S}_T$. These tests are asymptotically similar and therefore have asymptotically correct size in the presence of weak IVs.

For the case of an unknown sub-population, the identification-robust tests in the extant literature no longer apply because the set of instruments $C_{0,T}Z$ is unknown and must be estimated. In this section, we show how to form HAR CLR, LM and AR tests with correct asymptotic null rejection probabilities under both weak and strong IV asymptotics. To estimate the true sub-population $\mathbf{S}_{0,T}$ when constructing these tests let

$$\widehat{\mathbf{S}}_T = \arg\max_{\mathbf{S}_T \in \mathcal{S}} M_{2,T}(\mathbf{S}_T), \qquad \text{where} \qquad \mathcal{S} = \bigcup_{1 \le m \le m_+} \bigcup_{\pi \in (\epsilon, 1]} \Xi_{\epsilon,\pi,m,T}. \tag{6.5}$$

Proposition S.B.2 in the supplement shows that the process $\{\overline{Z}(C_T)'y\}_{\mathbf{S}_T \in \mathcal{S}}$ is sufficient for $(\beta, \theta')'$ in a canonical Gaussian setting analogous to that in Andrews et al. (2006) so that there is no loss in efficiency from using the unknown sub-population AR, LM and LR statistics, $LR_T(\widehat{\mathbf{S}}_T)$, $LM_T(\widehat{\mathbf{S}}_T)$ and $AR_T(\widehat{\mathbf{S}}_T)$, which are only functions of the process $\{\overline{Z}(C_T)'y\}_{\mathbf{S}_T \in \mathcal{S}}$.

We establish the asymptotic validity of the HAR CLR, LM and AR tests in the unknown sub-population setting under a weak set of high-level sufficient conditions on the IVs, exogenous variables and errors. Define $w(\mathbf{S}_T) = [C_T Z : X]$.

Assumption 6.1. *$T^{-1} w(\mathbf{S}_T)' w(\mathbf{S}_T') \xrightarrow{\mathbb{P}} Q(\mathbf{S}, \mathbf{S}')$ uniformly in $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$ for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$ and some p.d. $(q+p)\times(q+p)$ matrix $Q(\mathbf{S}, \mathbf{S}')$.*

Assumption 6.2. *$T^{-1} v'v \xrightarrow{\mathbb{P}} \Sigma_v$ for some 2×2 p.d. matrix Σ_v.*

Assumption 6.3. *For $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$ and $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, $\mathbf{S}' = \lim_{T\to\infty} T^{-1}\mathbf{S}_T'$, $T^{-1/2}\mathrm{vec}(w(\mathbf{S}_T)' v) \Rightarrow \mathscr{G}(\mathbf{S})$, where $\mathscr{G}(\cdot)$ is a mean-zero Gaussian process indexed by $\mathbf{S} \subseteq (0,1]$ with $2(q+p) \times 2(q+p)$ covariance function $\Psi(\mathbf{S}, \mathbf{S}') = \lim_{T\to\infty} T^{-1}\mathrm{Cov}(\mathrm{vec}(w(\mathbf{S}_T)' v), \mathrm{vec}(w(\mathbf{S}_T')' v))$.*

In Assumption 6.3, $\mathrm{vec}(\cdot)$ denotes the vec operator. The quantities $Q(\cdot)$, Σ_v, and $\Psi(\cdot)$ are assumed to be unknown. Assumptions 6.1-6.2 hold under suitable conditions by a (uniform) law of large numbers. Assumption 6.3 holds under suitable conditions by a functional

central limit theorem. Assumptions 6.1-6.3 are consistent with non-normal, heteroskedastic, autocorrelated errors and IVs and regressors that may be random or non-random.[23]

We assume that we can consistently estimate $\Sigma_{v\overline{Z}}(\mathbf{S}) \equiv \Sigma_{v\overline{Z}}(\mathbf{S},\mathbf{S})$ uniformly in $\mathbf{S}_T$.

Assumption 6.4. *We have an estimator $\widehat{\Sigma}_{v\overline{Z}}(\mathbf{S}_T)$ such that $\widehat{\Sigma}_{v\overline{Z}}(\mathbf{S}_T) \xrightarrow{\mathbb{P}} \Sigma_{v\overline{Z}}(\mathbf{S})$ uniformly in $\mathbf{S}_T \in \mathcal{S}$ for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$.*

Note that this assumption immediately implies the uniform consistency of $\widehat{\Sigma}_{N_2}(\mathbf{S}_T) = \widehat{\Sigma}^*_{N_2}(\mathbf{S}_T) - \widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T)\widehat{\Sigma}^{-1}_{N_1}(\mathbf{S}_T)\widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T)'$ as well. Consistent estimators of $\Sigma_{v\overline{Z}}$ are HAC and DK-HAC estimators.[24]

Finally, we impose a second-order stationarity condition for $v_t'b_0\overline{Z}_t(C_T)$ and $v_t'\Sigma_v^{-1}a_0\overline{Z}_t(C_T)$.

Assumption 6.5. *Let $\pi(\mathbf{S})$ equal the Lebesgue measure of $\mathbf{S} \subseteq (0,1]$. Assume that $\Sigma_{v\overline{Z}}(\mathbf{S},\mathbf{S}') = \pi(\mathbf{S}\cap\mathbf{S}')\Sigma_{v\overline{Z}}$ where $\mathbf{S},\mathbf{S}' \subseteq (0,1]$ and $\Sigma_{v\overline{Z}}$ is p.d.*

Assumption 6.5 is implied by a uniform law of large numbers and functional central limit theorem for partial sum processes under second-order stationarity. Under weak IV asymptotics, $T^{-1}\widehat{\mathbf{S}}_T$ is not consistent for $\mathbf{S}_0$. Assumption 6.5 is needed in order to show that $N_{1,T}(\cdot)$ and $N_{2,T}(\cdot)$ are asymptotically independent processes. Under strong IV asymptotics we can dispense with Assumption 6.5 because $T^{-1}\widehat{\mathbf{S}}_T \xrightarrow{\mathbb{P}} \mathbf{S}_0$ and the limit of the processes $N_{1,T}(\cdot)$ and $N_{2,T}(\cdot)$ have zero covariance when evaluated at a fixed $\mathbf{S}_0$.

Define the LR, LM and AR statistics in this context according to (6.4), replacing $\mathbf{S}_{0,T}$ with $\widehat{\mathbf{S}}_T$. We now establish the correct asymptotic null rejection probabilities of the sub-population-estimated plug-in HAR CLR, LM and AR tests under weak identification.

Theorem 6.1. *Let Assumptions 6.1-6.5 hold and suppose $\theta = c/T^{1/2}$ for some nonstochastic $c \in \mathbb{R}^q$. We have: (i) $AR_T(\widehat{\mathbf{S}}_T) \xrightarrow{d} \chi^2_q$ under H_0; (ii) $LM_T(\widehat{\mathbf{S}}_T) \xrightarrow{d} \chi^2_1$ under H_0; (iii) $\mathbb{P}_{\beta_0}(LR_T(\widehat{\mathbf{S}}_T) > \kappa_\alpha(N_{2,T}(\widehat{\mathbf{S}}_T)) \to \alpha$ where $\mathbb{P}_{\beta_0}(\cdot)$ is the probability computed under H_0.*

The key to establishing these asymptotic validity results is to show that each of the above statements hold conditional on the realization of $N_{2,T}(\cdot)$. This can be readily established from the facts that the stochastic processes $N_{1,T}(\cdot)$ and $N_{2,T}(\cdot)$ are asymptotically independent by construction, $\widehat{\mathbf{S}}_T$ is a function of $N_{2,T}(\cdot)$ and $N_{1,T}(\mathbf{S}_T) \Rightarrow \mathcal{N}(0, I_q)$ under H_0.[25]

[23] In the supplement we provide primitive sufficient conditions for Assumptions 6.1-6.3.

[24] In the supplement we provide weak sufficient conditions, even allowing for certain forms of nonstationarity, that ensure this assumption holds.

[25] In addition to identification-robust tests of H_0 vs H_1, since the causal interpretation of β depends upon

7 Empirical Evidence on LATE of Monetary Policy

We illustrate our methods by revisiting the identification of monetary policy effects in the framework of Nakamura and Steinsson (2018), introduced in Section 3. They use a bivariate model (3.1) to estimate the causal effect of $\widetilde{D}_t$ on $\widetilde{Y}_t$, employing both event-study and heteroskedasticity-based identification approaches. The dependent variable is the daily change in instantaneous U.S. Treasury forward rates. For the policy news $\widetilde{D}_t$ they use three variables: the daily change in nominal 2-Year Treasury yields, and the 30-minute or 1-day change in a "policy news" series—constructed as the first principal component of the unanticipated 30-minute changes in five selected interest rates. Heteroskedasticity-based identification assumes the variance of the monetary shock rises on FOMC announcement days, while the variance of other shocks remains constant [cf. eq. (3.2)]. FOMC dates define the policy sample $\mathbf{P}$, and analogous non-FOMC dates define the control sample $\mathbf{C}$. We consider specifications where $\widetilde{D}_t$ is either the 30-minute policy news series or 1-day change in Treasury yields, and $\widetilde{Y}_t$ is either the nominal or real 2-Year instantaneous Treasury forward rate. Nakamura and Steinsson's instrument for $\widetilde{D}_t^2$ is defined as $Z_t = \mathbf{1}\{t \in \mathbf{P}\}$, corresponding to the model in Section 3. We focus on the same period: January 1, 2004, to March 19, 2014.

Lewis (2022) recently analyzes this problem by developing a first-stage F-test for weak identification. He finds that weak identification is not rejected when $\widetilde{D}_t$ is the 1-day change in nominal 2-Year Treasury yields, but is strongly rejected when $\widetilde{D}_t$ is the 30-minute policy news series. This supports Nakamura and Steinsson's (2018) observation that the daily policy variable may suffer from weaker identification. Unlike Nakamura and Steinsson (2018), Lewis (2022) estimates the model using GMM and does not impose the assumption that the non-monetary policy shock η_t has equal variance across the treatment and control samples.

Section 7.1 reports results of our test for full sample identification failure. Section 7.2 presents causal effect estimates based on the most strongly-identified subsample. Section 7.3 provides identification-robust inference results, and Section 7.4 estimates compliers at the individual level and tests the exclusion restriction.

the sub-population $\mathbf{S}_{0,T}$, practitioners may wish to simultaneously report the result of these tests along with a corresponding estimate of the sub-population. More specifically, failure to reject H_0 should be interpreted as failure to reject that the estimand is equal to β_0, where the estimand is interpreted as a weighted average of the LATEs for the estimated sub-population $\widehat{\mathbf{S}}_T$. Given that the tests of H_0 remain asymptotically valid conditional on the realization of $N_{2,T}(\cdot)$ and the fact that $\widehat{\mathbf{S}}_T$ is a function of $N_{2,T}(\cdot)$, the tests remain asymptotically valid when interpreted conditional on the value of $\widehat{\mathbf{S}}_T$.

7.1 Testing for Identification Failure

We present the results of our test for identification failure over all sub-populations from Section 4 in Table 1 considering values of π_L from 0.6 to 1. For the 30-minute policy news variable, the F_T^* statistic is very large and identification failure is rejected at any common significance level. This supports the finding in Lewis (2022) and intuition in Nakamura and Steinsson (2018) that the 30-minute policy news variable leads to stronger identification in the full sample. In contrast, for the 1-day change in nominal Treasury yields, identification failure cannot be strongly rejected in the full sample: the F_T^* statistic at $\pi_L = 1$ (i.e., full sample) is only slightly larger than the 1% critical value. The F_T^* statistic increases substantially as π_L decreases and it is very far from the critical values. This is clear evidence that identification is much stronger over subsamples. At $\pi_L = 0.9$ it reaches 33.87, clearly rejecting identification failure in the π-subsample (with $\pi = 0.9$ or 0.95) over which the supremum of $F_T(\mathbf{S}_T)$ is computed. The F_T^* statistic increases monotonically with smaller π_L due to the increasing number of partitions considered. For example, at $\pi_L = 0.8$, F_T^* is 54.78—nearly seven times the full sample value. Overall, the results indicate that strong identification may hold when using a 1-day window, but only within subsamples comprising at most 90% of the data. The weak identification reported by Lewis (2022) using a 1-day window around FOMC announcements likely does not stem solely from volatility returning to normal after announcements. Rather, a small subsample (10–20% of the data) exhibits weak or failed identification, contributing to the weaker identification exhibited in the full sample.

Table 1: Tests for Identification Failure over all Sub-Populations

	F_T^* statistic and critical values				
			F_T^*		
$D_t \backslash \pi_L$	0.6	0.7	0.8	0.9	1
30-minute "policy news"	$10^4 \times 95.36$	$10^4 \times 56.50$	$10^4 \times 32.45$	$10^4 \times 18.75$	$10^4 \times 7.42$
1-day nominal Treasury yields	155.69	88.22	54.78	33.88	8.09
1% critical values	11.63	10.94	9.73	8.68	6.68
5% critical values	8.28	7.55	6.84	6.04	3.85

F_T^* statistics for first-stage identification failure. D_t is either the 30-minute policy news series or 1-day change in nominal Treasury yields. π_L is the minimum fraction of the sample over which the supremum of the $F(\mathbf{S}_T)$ is computed. Maximum number of breaks is set to $m_+ = 5$.

7.2 Estimation in Strongly-Identified Subsample

We turn to estimation of π_0 and $\mathbf{S}_{0,T}$ using the methods from Section 5, and then to estimating the LATE of monetary policy based on the strongly-identified subsample, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$, or simply, $\widehat{\pi}$-sample, where $\widehat{\pi} = |\widehat{\mathbf{S}}_{T,OLS}|/T$. We focus on $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$; results using $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ are similar. Figure 3 plots the 1-day changes in 2-Year yields for the control and policy samples and highlights the regimes included in the strongly-identified subsample $\widehat{\mathbf{S}}_{T,OLS}$. The estimate $\widehat{\pi} = 0.8$ implies that in 80% of the sample, the first-stage is strong and identification holds. In the control sample, the excluded periods include the first seven months of 2005 and the regime surrounding the financial crisis (2007-2009). As shown in the figure, volatility during the crisis period is much higher than in the rest of the control group and higher than the average volatility in the treatment group. This subsample appears to drive the apparent full sample weak identification. Since our method searches for maximum identification strength, it correctly excludes this period when computing π-LATE.[26] The interpretation is that in both excluded regimes—especially during the financial crisis—market uncertainty was elevated even on non-FOMC days, violating the identification assumption.

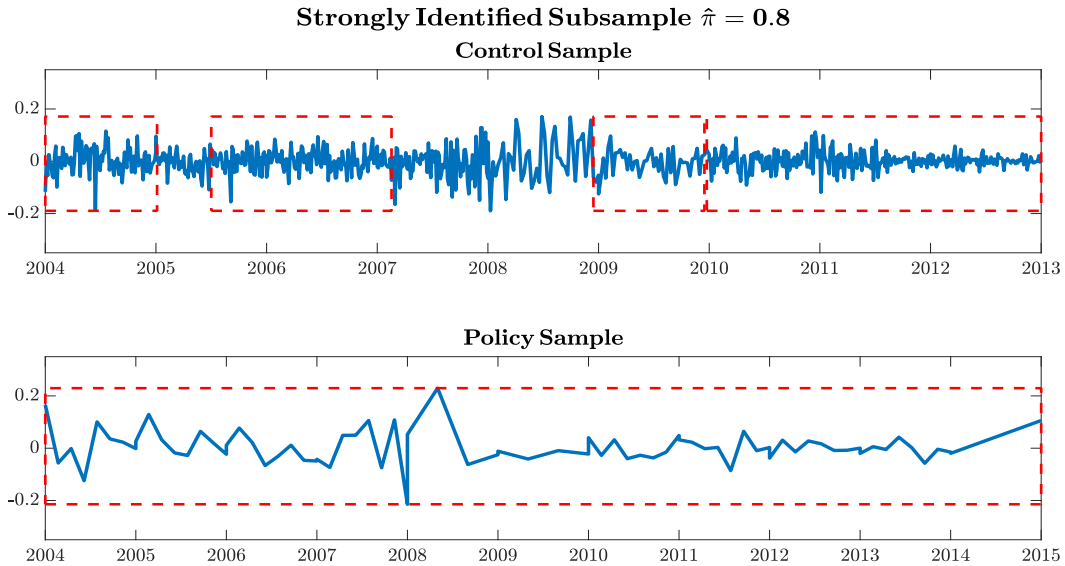

Figure 3: Plot of D_t (2-Years Treasury yields) in the control sample (top panel) and policy sample (bottom panel). The red rectangles indicate subsamples included in the strongly-identified subsample $\widehat{\mathbf{S}}_{T,OLS}$ where $\widehat{\pi} = 0.8$.

[26] The other excluded period (January to July 2005) does not display obviously high volatility but shows some persistence, with a short-duration cluster below the mean toward the end.

We now estimate the causal effect of monetary policy using the $\widehat{\pi}$-sample, where by construction the LATE is most strongly-identified. We compare these results with full sample estimates obtained using two-stage least squares (TSLS) and GMM, following Nakamura and Steinsson (2018) and Lewis (2022), respectively. Table 2 presents the results. Starting with the full sample estimates: when the policy variable is the 30-minute policy news series, TSLS and GMM yield very similar point estimates for both nominal and real forward rates, and both are statistically significant using standard and robust confidence intervals.[27]

As noted by Lewis (2022), the assumption that non-monetary shocks have equal variance across treatment and control groups does not bias the TSLS estimates, as they closely match the GMM ones. One explanation is that the GMM estimate of a (capturing reverse causality from forward rates to policy news) is both near zero and statistically significant (not reported). Since potential bias from this assumption is proportional to $a(\sigma^2_{\eta,P} - \sigma^2_{\eta,C})$, and a is close to zero, the resulting bias is negligible even if the variances $\sigma^2_{\eta,P}$ and $\sigma^2_{\eta,C}$ differ.

Turning to the case where the policy variable is the 1-day change in 2-Year Treasury yields, the TSLS and GMM estimates differ markedly from each other and from those based on the 30-minute policy news series. Notably, the GMM estimate of β is negative for nominal forwards and positive for real forwards, but in neither case is it statistically significant—whether using standard or robust confidence intervals.

As discussed by Lewis (2022), these estimates are difficult to interpret in economically meaningful terms. He also shows that the GMM estimates of a are nonzero and proposed a second dimension of policy news to account for the findings. However, the opposing signs of β across nominal and real forwards complicate this interpretation. Ultimately, he concludes that these results are inconsistent with Nakamura and Steinsson's (2018) "background noise" view of the non-monetary shock η_t which assumes that its volatility remains unchanged between FOMC and non-FOMC days.

We contribute to this discussion by presenting TSLS and GMM estimates based on the most strongly-identified $\widehat{\pi}$-sample. We focus first on standard confidence intervals and defer weak identification-robust inference to Table 3. The bottom panel of Table 2 shows that, for the 30-minute policy news variable, the TSLS and GMM estimates, including their statistical significance, are virtually unchanged. As expected—given the apparent strong identification

[27] The robust confidence intervals for the GMM estimates are based on the subset K-test in Lewis (2022).

in the full sample—results are broadly similar when using the $\widehat{\pi}$-sample.[28]

Table 2: Estimation of β

	30-minute Policy News		1-day 2-Year Yield	
dep. var.	Nominal	Real	Nominal	Real
		Full Sample		
		TSLS		
β	1.10**	0.96***	1.14***	0.97***
standard CI	[0.17, 2.02]	[0.41, 1.51]	[0.83, 1.45]	[0.40, 1.565]
		GMM		
β	1.07**	0.94***	-0.27	1.31
standard CI	[0.17, 1.98]	[0.36, 1.51]	[-4.90, 4.36]	[-3.74, 6.35]
robust CI	[0.27, 3.25]	[0.44, 2.38]	[-77.27, 0.94]	[-253.70, 1.92]
	π-sample based on $\widehat{\mathbf{S}}_{T,OLS}$ with $\widehat{\pi} = 0.8$			
		TSLS		
β	1.11**	0.97***	1.13***	0.92***
standard CI	[0.19, 2.02]	[0.42, 1.51]	[0.92, 1.30]	[0.56, 1.28]
		GMM		
β	1.07**	0.94***	0.65*	0.86**
standard CI	[0.17, 1.96]	[0.38, 1.50]	[-0.02, 131]	[0.29, 1.43]

TSLS estimates of β and GMM estimates of $\beta/(1-a\beta)$. The GMM estimates allow for changes also in the variance of η_t across regimes. The dependent variable is the 1-day change in either nominal or real 2-Year instantaneous Treasury forward rate. The policy variable is either the 30-minute changes in the "policy news" variable or 1-day changes in the 2-Year nominal Treasury yield. The standard 95% confidence interval is based on the standard normal critical values. For the GMM estimates, the robust 95% confidence interval is based on the subset K-test in Lewis (2022). Asterisks indicate statistical significance at the 10%, 5%, or 1% level based on standard intervals.

Finally, we turn to the $\widehat{\pi}$-sample estimates using the 1-day window for the policy. The GMM estimates differ sharply from those in the full sample: for both nominal and real forwards, they now have the same sign and are statistically significant. This suggests that the opposite signs reported by Lewis (2022) likely stemmed from weak identification, rendering those estimates unreliable.[29] Notably, the GMM estimates are now similar in magnitude to those based on the 30-minute policy variable, supporting a more meaningful interpretation.[30]

[28] The confidence intervals in the $\widehat{\pi}$-sample are even slightly tighter.

[29] While the TSLS estimates are nearly unchanged from the full sample, this should not be taken as evidence of their reliability. Under weak IVs, their similarity to the $\widehat{\pi}$-sample results may simply be coincidental.

[30] We also verified that the GMM estimate of a is 0.70 for nominal forwards and -0.91 for real forwards. It is intuitive that the estimate of a is close to zero when using a 30-minute window but significantly different from zero with a 1-day window. In the narrow 30-minute window around an FOMC announcement, reverse causality from $\widetilde{Y}_t$ to $\widetilde{D}_t$ is limited, as monetary news is more pronounced than other shocks—though some endogeneity may still arise from omitted factors affecting both. In contrast, over a full day, asset price movements can influence short-term interest rates, making reverse causality more likely.

Overall, this analysis highlights the advantage of using the most strongly-identified $\widehat{\pi}$-sample. Given weak identification in the full sample when using 1-day Treasury yields as the policy variable, the corresponding estimates should be discarded. In contrast, evidence from the $\widehat{\pi}$-sample shows that TSLS and GMM produce similar, positive estimates for β, consistent with monetary policy affecting real forward rates, as predicted by New Keynesian models, and supporting the existence of a forward guidance channel.[31]

7.3 Weak Identification-Robust Inference

We apply the weak identification-robust tests proposed in Section 6 and compare them to existing full sample tests LM_T and LR_T.[32] We test the the null hypothesis $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$, and extend the analysis to include 5-Year forward rates, in addition to the 2-Year forwards. Results are shown in Table 3. When the policy variable is the 30-minute policy news, identification is strong in the full sample. Accordingly, both the proposed and existing tests yield similar results: all tests reject at the 5% level for both nominal and real forwards. Nakamura and Steinsson (2018) showed that the effect of policy news peaks at the 2-Year maturity and declines with longer maturities. Consistent with this, we find weaker statistical significance for the 5-Year. In line with theoretical predictions, the long-run impact of monetary policy shocks on real interest rates (i.e., the 10 Year forwards) approaches zero (not reported). Our proposed tests confirm this, showing some rejection for the 5-Year real forwards but not for the 10-Year.

Let us instead consider the 1-day change in 2-Year yields as the policy variable. The existing LM_T and LR_T tests do not reject the null at any standard significance level for nominal forwards, and at the 1% level for real forwards. In contrast, our proposed tests based on $\widehat{\mathbf{S}}_T$ show much stronger rejections, aligning with the results using the 30-minute policy news series, which indicate a positive causal effect on 2-Year forwards.

[31] However, the results do not yet support a second meaningful dimension of news, as proposed by Lewis (2022), since the sign of the GMM estimate of a is unstable across nominal and real forwards. Regarding Nakamura and Steinsson's (2018) "background noise" interpretation of non-monetary shocks, we find no clear evidence against it: in the $\widehat{\pi}$-sample, identification appears strong, and TSLS and GMM estimates consistently share the same sign and similar magnitudes.

[32] We do not report the AR_T test since for $q = 1$ it is equivalent to the LM_T test.

Table 3: Identification-Robust Inference on β

	30-minute Policy News						1-day change in 2-Year Yields					
2-Year Forwards	Nominal			Real			Nominal			Real		
α	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
LM_T	✓	✓	×	✓	✓	✓	×	×	×	✓	✓	×
CLR_T	✓	✓	×	✓	✓	✓	×	×	×	✓	✓	×
$LM_T(\widehat{\mathbf{S}}_T)$	✓	✓	×	✓	✓	✓	✓	✓	×	✓	✓	✓
$CLR_T(\widehat{\mathbf{S}}_T)$	✓	✓	×	✓	✓	×	✓	✓	✓	✓	✓	✓
5-Year Forwards	Nominal			Real			Nominal			Real		
α	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
LM_T	×	×	×	×	×	×	×	×	×	×	×	×
CLR_T	×	×	×	×	×	×	×	×	×	×	×	×
$LM_T(\widehat{\mathbf{S}}_T)$	×	×	×	×	×	×	✓	×	×	✓	✓	×
$CLR_T(\widehat{\mathbf{S}}_T)$	×	×	×	×	×	×	✓	✓	×	✓	✓	×

Weak identification-robust tests on β. The dependent variable Y_t is either the 2-Year forward rates (top panel) or the 5-Year forward rates (bottom panel). D_t is either the 30-minute policy news series or the 1-day nominal Treasury yields. Significance levels are $\alpha = 0.10, 0.05, 0.01$. A ✓ means indicates rejection H_0; a × non-rejection.

7.4 Identification and Estimation of Compliers, and Exclusions Restriction

We now identify compliers individually by applying Theorem 2.1. Under heteroskedasticity-based identification, the sample rolling window averages in Theorem 2.1 correspond to rolling window variances, i.e., $\overline{D}_{P,t,n_1}$ and $\overline{D}_{C,t,n_0}$ are equal to $\overline{\sigma}^2_{P,t,n_1}$ and $\overline{\sigma}^2_{C,t,n_0}$ in this context, where $\overline{\sigma}^2_{P,t,n_1}$ and $\overline{\sigma}^2_{C,t,n_0}$ are defined analogously but using $\tilde{D}^2_t$ for D_t.[33] We use two-sided rolling windows with $n_0 = 101$ and $n_1 = 15$. For each t_0 we test the null hypothesis $H_0 : \mathbb{E}(D^2_{t_0}(1)) - \mathbb{E}(D^2_{t_0}(0)) = 0$ (t_0 is a non-complier) versus the one tailed alternative $H_1 : \mathbb{E}(D^2_{t_0}(1)) - \mathbb{E}(D^2_{t_0}(0)) > 0$ (t_0 is a complier). We use the t-statistic

$$t_{t_0} = \begin{cases} \frac{\sqrt{n_0}\left(\overline{\sigma}^2_{P,t_0,n_1} - \overline{\sigma}^2_{C,t_0-1,n_0}\right)}{\sqrt{J_{\mathrm{HAC},t_0-1}}} & t_0 \in \mathbf{P} \\ \frac{\sqrt{n_0}\left(\overline{\sigma}^2_{P,s^*(t_0),n_1} - \overline{\sigma}^2_{C,t_0,n_0}\right)}{\sqrt{J_{\mathrm{HAC},t_0}}} & t_0 \in \mathbf{C}, \end{cases}$$

where J_{HAC,t_0} is the Newey-West estimator with $\left\lfloor n_0^{1/3} \right\rfloor$ lags applied to $\tilde{D}^2_s - n_0^{-1}\sum_{k\in N_0(t_0+1)} \tilde{D}^2_k$.

The results are shown in Figure 4. Approximately 75% of observations are classified as compliers. The non-compliers are mostly concentrated in the period from 2010 to mid-

[33] More specifically, we have $\overline{\sigma}^2_{C,t_0-1,n_0} = \frac{1}{n_0}\sum_{s\in N_0(t_0)} \tilde{D}^2_s$, and $\overline{\sigma}^2_{P,t_0,n_1} = \frac{1}{n_1}\sum_{s\in N_1(t_0)} \tilde{D}^2_s$.

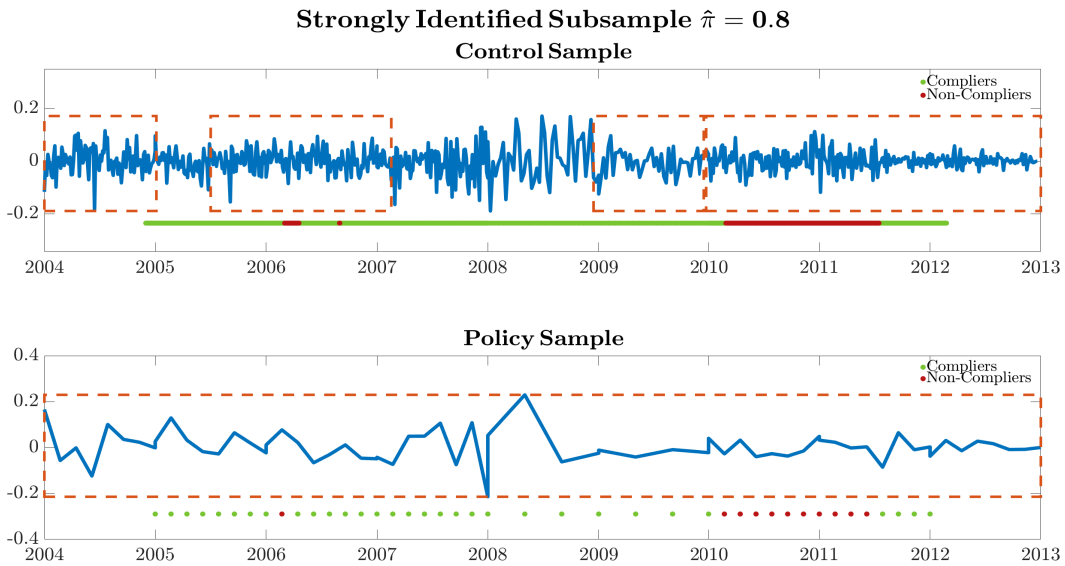

Figure 4: Plot of $\widetilde{D}_t$ (2-Years Treasury yields) in the control sample (top panel) and policy sample (bottom panel). The orange rectangles indicate subsamples included in the strongly-identified set $\widehat{\mathbf{S}}_{T,OLS}$ where $\widehat{\pi} = 0.8$. Green filled circles indicate compliers; red filled circles indicate non-compliers. Time points without colored markers correspond to cases where rolling sample variances could not be computed due to proximity to the start or end of the sample.

2011, which corresponds to the early phase of the zero lower bound (ZLB) period following the 2008–09 recession. During this time, the Fed relied primarily on qualitative forward guidance—e.g., stating that economic conditions were "likely to warrant exceptionally low levels of the federal funds rate for some time." In August 2011, the Fed shifted to more explicit, calendar-based guidance, stating that such conditions were "likely to warrant exceptionally low levels of the federal funds rate at least through mid-2013." Thus, the non-complier period aligns with the phase of the ZLB when forward guidance was less aggressive as the policy announcements by then only imply a near zero-rate horizon for the following three to four quarters, significantly shorter than what the ZLB constraint would actually have implied.[34]

Finally, we use Theorem 2.1 and Proposition 2.3 to test the exclusion restriction (cf. Assumption 2.2). We consider the whole set of compliers $\mathcal{NC}$. We test the null hypothesis

[34]The set of compliers does not coincide with the set of observations in the strongly-identified $\widehat{\pi}$-sample. However, this does not necessarily imply a violation of monotonicity [cf. Proposition 2.2]. First, the complier status is determined via a t-test whereas the strongly-identified $\widehat{\pi}$-sample is determined via estimation. Second, the complier status is determined by the rolling window variances at each t, whereas inclusion in the $\widehat{\pi}$-sample depends on how these variances contribute to the average volatility in the control sample relative to that in the policy sample.

that the exclusion restriction holds by using the following t-statistic:

$$t_{\text{exclusion}} = \frac{\sqrt{|\mathcal{NC}_{\mathbf{C}}|}\left(\overline{Y}_{\mathcal{NC}^s_{\mathbf{P}}} - \overline{Y}_{\mathcal{NC}^s_{\mathbf{C}}}\right)}{\sqrt{J_{\text{HAC},\mathcal{NC}^s}}}$$

where $\overline{Y}_{\mathcal{NC}_{\mathbf{P}}} = \frac{1}{|\mathcal{NC}_{\mathbf{P}}|}\sum_{t\in\mathcal{NC}_{\mathbf{P}}} Y_t$, $\overline{Y}_{\mathcal{NC}_{\mathbf{C}}} = \frac{1}{|\mathcal{NC}_{\mathbf{C}}|}\sum_{t\in\mathcal{NC}_{\mathbf{C}}} Y_t$, $Y_t = \tilde{D}_t\tilde{Y}_t$ and $J_{\text{HAC},\mathcal{NC}^s}$ is the Newey-West estimator applied to $\overline{Y}_{\mathcal{NC}_{\mathbf{P}}} - \overline{Y}_{\mathcal{NC}_{\mathbf{C}}}$. For the real (nominal) 2-Year forward rate, we find $t_{\text{exclusion}} = 0.91$ ($t_{\text{exclusion}} = 1.02$), and thus fail to reject the exclusion restriction.

8 Conclusions

This paper discusses identification, estimation and inference on dynamic LATE. We show that compliers can be identified individually and the exclusion restriction can be tested using a t-test. While weak identification is common in the full sample in practice, strong identification often appears to hold in a sizable subsample. We propose a method to isolate this strongly-identified subsample, enabling consistent estimation and inference.

Supplemental Materials: The online supplement [cf. Casini et al. (2025a)] includes Monte Carlo simulations, proofs of the results of Sections 2-4 and 6. The non-online supplement [cf. Casini et al. (2025b)] contains the theoretical results and corresponding proofs for the estimators in Section 5 and additional results.

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231–263.

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20(1), 46–63.

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

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, 100(2), 337–348.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instrumentsin IV Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434), 444–455.

Angrist, J. D., and G. M. Kuersteiner (2011): "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score," *Review of Economics and Statistics*, 93(3), 725–747.

Antoine, B., and O. Boldea (2018): "Efficient Estimation with Time-Varying Information and the New Keynesian Phillips Curve," *Journal of Econometrics*, 204(2), 268–300.

Bai, J., and P. Perron (2003): "Computation and Analysis of Multiple Structural Changes," *Journal of Applied Econometrics*, 18, 1–22.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020): "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 128(4), 1269–1324.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2025): "When is TSLS Actually LATE?," *NBER Working Paper 29709*.

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

Casini, A., and A. McCloskey (2025): "Identification, Estimation and Inference in High-Frequency Event Study Regressions," *arXiv preprint arXiv 2406.15667*.

Casini, A., A. McCloskey, L. Rolla, and R. Pala (2025a): "Online Supplement to "Dynamic Local Average Treatment Effects in Time Series"," *Unpublish manuscript*.

——— (2025b): "Supplement Not For Online Publication to "Dynamic Local Average Treatment Effects in Time Series"," *Unpublish manuscript*.

Casini, A., and P. Perron (2019): "Structural Breaks in Time Series," *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.

——— (2024): "Change-Point Analysis of Time Series with Evolutionary Spectra," *Journal of Econometrics*, 242(2), 105811.

Gertler, M., and P. Karadi (2015): "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7(1), 44–76.

Hamilton, J. D. (2003): "What Is an Oil Shock?," *Journal of Econometrics*, 113(2), 363–398.

Heckman, J. J. (1996): "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association*, 91(434), 459–462.

Imbens, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48(2), 399–423.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Imbens, G. W., and D. B. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64(4), 555–574.

Jordà, O., M. Schularick, and A. M. Taylor (2015): "Betting the House," *Journal of International Economics*, 96(S1), S2–S18.

Känzig, D. R. (2021a): "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements," *American Economic Review*, 111(4), 1092–1125).

——— (2021b): "The Unequal Economic Consequences of Carbon Pricing," *Unpublished Manuscript,*, London Business School.

Kilian, L. (2009): "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 99(3), 1053–1069.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

Kolesár, M., and M. Plagborg-Møller (2025): "Dynamic Causal Effects in a Nonlinear World:the Good, the Bad, and the Ugly," *arXiv preprint arXiv 2411.10415*.

Kuttner, K. N. (2001): "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market," *Journal of Monetary Economics*, 47(3), 523–544.

Lewis, D. J. (2022): "Robust Inference in Models Identified via Heteroskedasticity," *Review of Economics and Statistics*, 104(3), 510–524.

Magnusson, L. M., and S. Mavroeidis (2014): "Identification Using Stability Restrictions," *Econometrica*, 82(5), 1799–1851.

Mertens, K., and J. L. Montiel Olea (2018): "Marginal Tax Rates and Income: New Time Series Evidence," *Quarterly Journal of Economics*, 133(4), 1803–1884.

Mertens, K., and M. O. Ravn (2013): "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, 103(4), 1212–1247.

Montiel Olea, J. L., and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics*, 31(3), 358–369.

Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2021): "Inference in SVARs Identified with External Instruments," *Journal of Econometrics*, 225(1), 74–87.

Moreira, M. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

Nakamura, E., and J. Steinsson (2018): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133(3), 1283–1330.

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

Plagborg-Møller, M., and C. K. Wolf (2022): "Instrumental Variable Identification of Dynamic Variance Decompositions," *Journal of Political Economy*, 130(8), 2164–2202.

Rambachan, A., and N. Shephard (2021): "When Do Common Time Series Estimands Have Nonparametric Causal Meaning?," *Unpublished Manuscript, Harvard University.*

Ramey, V. A. (2011): "Identifying Government Spending Shocks: It's All in the Timing," *Quarterly Journal of Economics*, 126(1), 1–50.

Ramey, V. A., and M. D. Shapiro (1998): "Costly Capital Reallocation and the Effects of Government Spending," *Carnegie-Rochester Conference Series*, 48, 145–194.

Ramey, V. A., and S. Zubairy (2018): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126(2), 850–901.

Rigobon, R. (2003): "Identification Through Heteroskedasticity," *Review of Economics and Statistics*, 85(4), 777–792.

Rigobon, R., and B. Sack (2003): "Measuring the Reaction of Monetary Policy to the Stock Market," *The Quarterly Journal of Economics*, 118(2), 639–669.

Robins, J. M., and S. Greenland (1996): "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association*, 91(434), 456–458.

Romer, C. D., and D. H. Romer (1989): "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, 4, 121–184.

——— (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94(4), 1055–1084.

——— (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*.

Rubin, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66(5), 688–701.

Sojitra, R. B., and V. Syrgkanis (2025): "Dynamic Local Average Treatment Effects," *arXiv preprint arXiv:2405.01463*.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

Stock, J. H., and M. W. Watson (2012): "Disentangling the Channels of the 2007-2009 Recession," *Brookings Papers on Economic Activity*, (1), 81–135.

——— (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *Economic Journal*, 128(610), 917–948.

Online Supplement 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

16th September 2025

Abstract

This supplemental material is for online publication. It contains additional results on identification-robust inference, Monte Carlo simulations and proofs of the results.

S.A Critical Values of F_T^*

Table 4: Critical values of F_T^*

$\alpha = 0.10$						
$\pi_L \backslash q$	1	2	3	4	5	10
0.50	7.44	5.18	4.20	3.70	3.35	2.53
0.60	6.92	4.76	3.94	3.46	3.14	2.40
0.70	6.19	4.44	3.71	3.28	2.96	2.31
0.80	5.51	4.02	3.37	2.79	2.77	2.18
0.90	4.81	3.59	3.08	2.78	2.52	2.04
1.00	2.70	2.32	2.09	1.94	1.80	1.59
$\alpha = 0.05$						
$\pi_L \backslash q$	1	2	3	4	5	10
0.50	8.90	6.03	4.75	4.14	3.74	2.74
0.60	8.28	5.60	4.49	3.91	3.51	2.62
0.70	7.55	5.21	4.26	3.70	3.31	2.53
0.80	6.84	4.71	3.83	3.43	3.13	2.39
0.90	6.04	4.31	3.58	3.19	2.89	2.25
1.00	3.85	3.00	2.57	2.37	2.16	1.82
$\alpha = 0.01$						
$\pi_L \backslash q$	1	2	3	4	5	10
0.50	12.27	7.91	6.08	5.12	4.56	3.19
0.60	11.63	7.28	5.73	4.81	4.37	3.08
0.70	10.94	6.97	5.56	4.67	4.19	3.04
0.80	9.73	6.41	5.06	4.34	3.89	2.84
0.90	8.68	5.94	4.62	4.15	3.65	2.69
1.00	6.68	4.60	3.70	3.31	2.99	2.31

S.B Additional Results on Identification-Robust Inference

We present the sufficiency results referenced in Section 6 as well as other results.

S.B.1 Known Sub-Population

When $\mathbf{S}_{0,T}$ is known, it is straightforward to use existing tests in the identification-robust linear IVs literature to test H_0 with known optimality properties under certain conditions. However, one must be careful in defining the appropriate statistics when applying existing tests in this setting in order to maintain efficiency.

With fixed regressors X and Z and reduced-form errors v that are i.i.d. across rows with each row being bivariate normally distributed with a mean of zero and a known nonsingular covariance matrix Σ_v, careful application of the results of Andrews, Moreira, and Stock (2006) imply that $\overline{Z}(C_{0,T})'y$ is a sufficient statistic for $(\beta, \theta')'$. This implies that there can be no loss in efficiency from focusing on tests that are functions of only $\overline{Z}(C_{0,T})'y$. On the other hand, the following proposition implies a loss in efficiency from tests that are functions of only $Z'M_Xy$, which a casual user may be tempted to use when constructing identification-robust tests for β.

Proposition S.B.1. *For the model in (6.1) with fixed regressors X and Z and reduced-form errors v that are i.i.d. across rows with each row being bivariate normal with a zero mean and known p.d. covariance matrix Σ_v, $Z'M_Xy$ is not sufficient for $(\beta,\, \theta')'$ if $\pi_0 < 1$.*

This result implies that existing tests (i.e.,. CLR, LM and AR tests) and extensions thereof are not efficient when Z is treated as the matrix of IVs rather than $C_{0,T}Z$.

The results of Andrews, Moreira, and Stock (2006) imply that the CLR, LM and AR tests have limiting null rejection probabilities equal to α under weak instrument asymptotics. In addition, results in Andrews, Moreira, and Stock (2006) imply asymptotic near-optimality properties of the CLR test under a stronger set of assumptions that may not hold in the presence of serial correlation in $\{v_t\}$.

S.B.2 Unknown Sub-Population

In Section 6 we propose CLR, LM and AR statistics where we plug-in the estimate for the unknown sub-population: $CLR_T(\widehat{\mathbf{S}}_T)$, $LM_T(\widehat{\mathbf{S}}_T)$ and $AR_T(\widehat{\mathbf{S}}_T)$. To motivate their use in the unknown sub-population setting, we establish the analog of the sufficiency result of Andrews, Moreira, and Stock (2006) in this setting.

Proposition S.B.2. *For the model in (6.1) with fixed regressors X and Z and reduced form errors v that are i.i.d. across rows with each row being bivariate normal with a zero mean, known p.d. covariance matrix Σ_v and an unknown sub-population $\mathbf{S}_{0,T} \in \mathcal{S}$, the Gaussian process $\{\overline{Z}(C_T)'y\}_{\mathbf{S}_T \in \mathcal{S}}$ is a sufficient statistic for $(\beta, \theta')'$.*

Since the AR, LM and CLR statistics are only functions of $\{\overline{Z}(C_T)'y\}_{\mathbf{S}_T \in \mathcal{S}}$, this result implies that these tests entail no loss in efficiency relative to tests using the entire data.

We finally show that under weak IV asymptotics, $\widehat{\mathbf{S}}_T$ is the maximum likelihood estimator of $\mathbf{S}_T$ under H_0, so that the sub-population estimate we use when constructing and interpreting the identification-robust tests is efficient.

Proposition S.B.3. *For the model in (6.1) with fixed regressors X and Z and reduced form errors v that are i.i.d. across rows with each row being bivariate normal with a zero mean, p.d. covariance matrix Σ_v and unknown sub-population $\mathbf{S}_T \in \mathcal{S}$, if Assumptions 6.1 and 6.3-6.4 hold and $\theta = c/T^{1/2}$ for some nonstochastic $c \in \mathbb{R}^q$, $\widehat{\mathbf{S}}_T$ is asymptotically equivalent to the maximum likelihood estimator of $\mathbf{S}_T$ under H_0.*

The result in Proposition S.B.3 continues to hold under strong IVs, i.e., $\theta \neq 0$ is fixed. See Casini, McCloskey, Rolla, and Pala (2025b).

Magnusson and Mavroeidis (2014) propose tests for β that are robust to changes in θ whether through persistent time variation or breaks. For the latter case, they assumed the number of breaks is known, whereas our approach does not require this prior knowledge. For the weak IVs case, Magnusson and Mavroeidis (2014) consider a single break and build split-sample tests based on the sufficient statistic $\{Z(\tau)' y\}_{\tau\in[0,1]}$ where $Z(\tau) = [[\{Z_t'\}_{t=1}^{\lfloor \tau T\rfloor}\ \mathbf{0}']' \vdots [\mathbf{0}'\ \{Z_t'\}_{t=\lfloor \tau T\rfloor+1}^{T}]']$. This high-dimensional statistic effectively uses the full data sequence evaluated at all potential split points. In contrast, our framework focuses on subsamples where θ is nonzero, allowing us to construct a lower-dimensional sufficient statistic. In other words, their statistic is not minimal sufficient [cf. Lehmann and Romano (2005)], whereas ours is—making our tests more efficient in this setting.

S.C Monte Carlo Simulations

S.C.1 Finite-Sample Size and Power of F_T^*

We study the finite-sample rejection frequencies of the F_T^* test using a simulation experiment calibrated to real data from the analysis in Nakamura and Steinsson (2018) introduced in Section 3. We use the same sequences of policy dates and control dates, $\mathbf{P}$ and $\mathbf{C}$, to construct the instrument as

$$Z_t = \begin{cases} \frac{T}{T_P}, & t \in \mathbf{P} \\ -\frac{T}{T_C}, & t \in \mathbf{C}. \end{cases}$$

Under heteroskedasticity-based identification the first-stage equation can be equivalently written as $D_t = \theta Z_t D_t + e_t$ for some θ and e_t [cf. Rigobon and Sack (2003) and Lewis (2022)]. Hence, we generate D_t according to the following data-generating process (DGP):

$$D_t = \begin{cases} \frac{e_t}{1-\theta_1 Z_t}, & t \leq \lfloor T/4 \rfloor \\ \frac{e_t}{1-\theta_2 Z_t}, & \lfloor T/4 \rfloor + 1 \leq t \leq \lfloor T/4 \rfloor + \lfloor (1-\pi_0) T \rfloor \\ \frac{e_t}{1-\theta_3 Z_t}, & \lfloor T/4 \rfloor + \lfloor (1-\pi_0) T \rfloor + 1 \leq t \leq T, \end{cases} \tag{S.C.1}$$

where $\pi_0 = 0.4$, 0.6, 0.8 and $T = 400$. We set T_P and T_C equal to the number of policy and control dates that occur in the first T observations in Nakamura and Steinsson's (2018) sample. We specify $e_t = \rho_e e_{t-1} + v_{e,t}$, where $\rho_e \in \{0, 0.25, 0.5, 0.75\}$, $v_{e,t} \sim$ i.i.d. $\mathcal{N}(0, \sigma_v^2)$ and σ_v^2 is set equal to the sample variance of the policy variable (2-years nominal Treasury yields). We set $\theta_1 = \theta_2 = \theta_3 = 0$ under the null hypothesis. Under the alternative, $\theta_1 = \theta_3 > 0$ and $\theta_2 = 0$.

We also consider the following DGP:[35]

$$Y_t = \beta D_t + \gamma_1 X_t + u_t, \tag{S.C.2}$$

where $X_t = 1$ for all t and

$$D_t = \begin{cases} \theta_1 Z_t + \gamma_2 X_t + e_t, & t \leq \lfloor T/4 \rfloor \\ \theta_2 Z_t + \gamma_2 X_t + e_t, & \lfloor T/4 \rfloor + 1 \leq t \leq \lfloor T/4 \rfloor + \lfloor (1-\pi_0) T \rfloor \\ \theta_3 Z_t + \gamma_2 X_t + e_t, & \lfloor T/4 \rfloor + \lfloor (1-\pi_0) T \rfloor + 1 \leq t \leq T, \end{cases} \tag{S.C.3}$$

$Z_t \sim$ i.i.d. $\mathcal{N}(1, 1)$, and u_t and e_t are i.i.d. jointly normal with mean zero and covariance

$$\Sigma_{ue} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{S.C.4}$$

with $\rho \in \{0.25, 0.75\}$ and $\gamma_1 = \gamma_2 \in \{0, 1\}$. Under the null hypothesis we set $\theta_1 = \theta_2 = \theta_3 = 0$. Under the alternative hypothesis we set $\theta_1 = \theta_3 = dT^{-1/2}$ with $d \in \{4, 16, 32\}$ and $\theta_2 = 0$. We also consider two additional specifications for θ_2. In the first, $\theta_2 = -0.5dT^{-1/2}$, so θ_2 has the opposite value to θ_1 and θ_3. In the second, $\theta_2 = -0.5$. In both cases the instrument is

[35] This is also used to compare the performance of the estimators of β discussed in Section 5 and of the identification-robust tests discussed in Section 6.

relevant throughout the sample (no identification failure). However, in the second regime the first-stage effect tends to offset instrument relevance in the other regimes because of the sign reversal. We set $\pi_0 \in \{0.6, 0.8\}$ and $T = 200$. We also consider a variant of the DGP with serially correlated data. We assume $u_t = \rho_u u_{t-1} + v_{u,t}$ and $e_t = \rho_e e_{t-1} + v_{e,t}$ with $v_{u,t}$ and $v_{e,t}$ being jointly normal with mean zero and covariance Σ_{ue} as in (S.C.4) with $\rho \in \{0.25, 0.75\}$.

Throughout the simulation study, F_T^* is implemented with $\pi_L = 0.6$, $\epsilon = 0.05$, $m_+ = 5$. For both F_T^* and the full sample F_T we use the Newey-West estimator with bandwidth equal to the popular rule $\left\lfloor T^{-1/3} \right\rfloor$ for $\widehat{J}(\cdot)$.[36] The significance level is 5% and the number of simulations is 5,000. Figure 5 plots the rejection rates of F_T^* and the full sample F_T for the calibrated DGP in (S.C.1). F_T^* yields accurate rejection rates whereas F_T is undersized, providing evidence for the reliability of our empirical results in Section 7. F_T^* is more powerful than F_T by about 10% across all values of ρ_e.

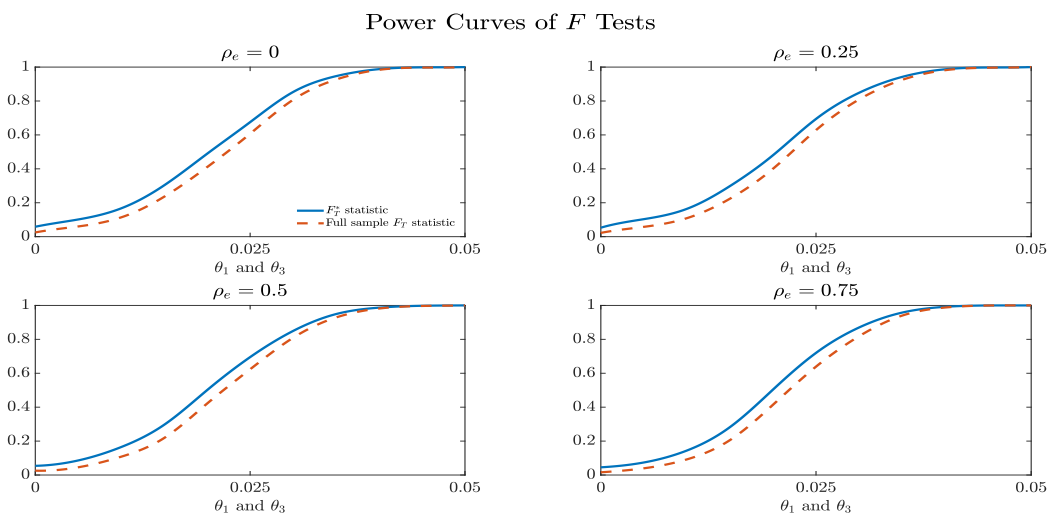

Figure 5: Power curves of F_T^* and F_T for the DGP calibrated to Nakamura and Steinsson's (2018) sample with $T = 400$.

Moving on to the DGP specified in (S.C.2)-(S.C.3), Table 5 presents the size and size-adjusted power of the F tests for the case $\theta_2 = -0.5$ under the alternative. The F_T^* statistic is more oversized than F_T for $\rho = 0.25$ and its size improves for $\rho = 0.75$, possibly due to small-sample bias in the long-run variance estimator. However, the over-sizing is relatively modest. The size-adjusted power of F_T^* is much higher than that of the the full sample F_T for

[36] We also consider data-dependent bandwidths. The results are similar and not reported.

all values of d. Power gains are larger when $\pi_0 = 0.6$ than when 0.8 since in the former case the strongly-identified subsample is smaller, making the full sample F_T rely more heavily on subsamples that suffer from identification failure.

Table 5: Size and size-adjusted power of F tests under alternative hypothesis with $\theta_2 = -0.5$

$\rho = 0.25$, $\pi_0 = 0.6$	$\theta_1 = \theta_2 = \theta_3 = 0$ (null)	$d = 4$	$d = 8$	$d = 12$	$d = 16$	$d = 20$	$d = 24$
full sample F_T	0.061	0.023	0.123	0.478	0.797	0.933	0.981
F_T^*	0.110	0.167	0.441	0.847	0.973	0.991	1.000
$\rho = 0.75$, $\pi_0 = 0.6$	$\theta_1 = \theta_2 = \theta_3 = 0$ (null)	$d = 4$	$d = 8$	$d = 12$	$d = 16$	$d = 20$	$d = 24$
full sample F_T	0.063	0.034	0.107	0.394	0.712	0.879	0.957
F_T^*	0.083	0.163	0.410	0.791	0.946	0.991	0.999
$\rho = 0.25$, $\pi_0 = 0.8$	$\theta_1 = \theta_2 = \theta_3 = 0$ (null)	$d = 2$	$d = 6$	$d = 10$	$d = 14$	$d = 18$	$d = 22$
full sample F_T	0.061	0.038	0.642	0.969	0.998	1.000	1.000
F_T^*	0.110	0.107	0.872	0.994	1.000	1.000	1.000
$\rho = 0.75$, $\pi_0 = 0.8$	$\theta_1 = \theta_2 = \theta_3 = 0$ (null)	$d = 2$	$d = 6$	$d = 10$	$d = 14$	$d = 18$	$d = 22$
full sample F_T	0.063	0.049	0.468	0.907	0.993	0.999	1.000
F_T^*	0.083	0.091	0.709	0.984	0.999	1.000	1.000

Figure 6 plots the size-adjusted power of the F tests for the specification $\theta_2 = -0.5dT^{-1/2}$.[37] In this specification, under the alternative the instrument is relevant throughout the sample. However, because θ_2 has opposite sign to θ_1 and θ_3 the contribution of the second regime tends to offset those of the first and third. Consistent with this, the plots show that F_T^* is substantially more powerful than the full sample F_T. The resulting power gains exceed those obtained when θ_2 shares the same sign as θ_1 and θ_3.

S.C.2 Finite-Sample Properties of $\widehat{S}_{T,OLS}$, $\widehat{S}_{T,FGLS}$, $\widehat{\beta}(\widehat{S}_{T,OLS})$ and $\widehat{\beta}(\widehat{S}_{T,FGLS})$

We study the finite-sample bias and mean-squared error (MSE) of the proposed estimators of $S_{0,T}$ and β. For the latter, we compare them with $\widehat{\beta}_{FS}$, the full sample IV estimator of β. We consider the same DGP as in (S.C.2)-(S.C.3) where $X_t \sim$ i.i.d. $\mathcal{N}(1, 1)$. We set $\beta = 1$.[38] The number of simulations is 5,000.

Table 6 reports the MSE of $\widehat{\mathbf{S}}_{T,OLS}$ and $\widehat{\mathbf{S}}_{T,FGLS}$ for the case $\theta_1 = \theta_2 = \theta_3 = dT^{-1/2}$, where $\widehat{\mathbf{S}}_{T,FGLS}$ is constructed using $\widehat{\Omega}_\varepsilon(\mathbf{S}_T) = \widehat{\Sigma}_\varepsilon \otimes I_T$, which is misspecified in the presence

[37] Note that the size of the tests is that reported in Table 5 since the DGPs for the two specifications are the same under the null.

[38] The results do not change with other values of β.

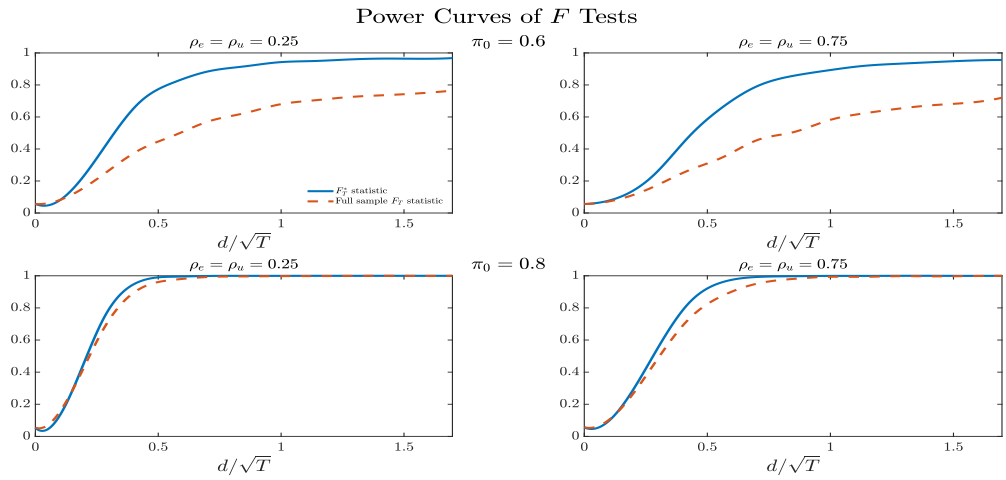

Figure 6: Power curves of F_T^* and full sample F_T for $\theta_2 = -0.5dT^{-1/2}$ under the alternative hypothesis with $T = 200$.

of serial correlation. Under i.i.d. errors, FGLS yields a lower MSE when the endogeneity is high ($\rho = 0.75$) and the instrument is not weak ($d = 16$ and 32). This is intuitive, as FGLS expolits information from the cross-equation correlation in the errors. However, when the errors are serially correlated ($\rho_e = \rho_u > 0$), FGLS exhibits a higher MSE than OLS except for a few instances (e.g., $\rho = 0.75$ and $d = 32$). This pattern arises because the relative efficiency gains of FGLS over OLS are not guaranteed in the presence of a misspecified covariance.

Table 6: MSE of $\widehat{S}_{T,OLS}$ and $\widehat{S}_{T,FGLS}$

	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.25$, $\pi_0 = 0.6$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
MSE($\widehat{S}_{T,OLS}$)	31.58	2.24	0.71	37.44	4.03	1.07
MSE($\widehat{S}_{T,FGLS}$)	55.29	2.83	0.76	55.97	9.21	1.29
	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.75$, $\pi_0 = 0.6$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
MSE($\widehat{S}_{T,OLS}$)	31.58	2.24	0.71	37.44	4.03	1.07
MSE($\widehat{S}_{T,FGLS}$)	52.71	1.84	0.47	54.69	4.56	0.64
	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.25$, $\pi_0 = 0.8$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
MSE($\widehat{S}_{T,OLS}$)	22.61	2.26	0.92	26.70	3.84	1.21
MSE($\widehat{S}_{T,FGLS}$)	30.43	2.81	1.02	32.56	8.81	1.53
	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.75$, $\pi_0 = 0.8$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
MSE($\widehat{S}_{T,OLS}$)	22.44	2.27	0.96	26.70	3.84	1.21
MSE($\widehat{S}_{T,FGLS}$)	24.24	1.55	0.77	31.63	3.42	0.92

To facilitate readability each value is multiplied by 10^2.

Next, we examine the bias and MSE of $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ and $\widehat{\beta}_{FS}$ for the case $\theta_1 = \theta_2 = \theta_3 = dT^{-1/2}$, where $\widehat{\mathbf{S}}_{T,FGLS}$ is again constructed using $\widehat{\Omega}_\varepsilon(\mathbf{S}_T) = \widehat{\Sigma}_\varepsilon \otimes I_T$. Table 7 reports the results. When the instrument is weak throughout the sample ($d = 4$) no estimator uniformly dominates in terms of bias or MSE, and the results vary considerably across settings. In contrast, when the instrument is strong in parts of the sample ($d = 16$, 32), $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ generally exhibits lower bias and MSE than both $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ and $\widehat{\beta}_{FS}$. The reduction in bias and MSE can be substantial—often at least 50% in many configurations. Moreover, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ tends to outperform $\widehat{\beta}_{FS}$ when $d = 16$ or 32, and occasionally even when $d = 4$. The results for the cases $\theta_2 = -0.5dT^{-1/2}$ and $\theta_2 = -0.5$ are similar and omitted.

We now evaluate the performance of the estimators of β in a setting where the instrument is strong in the subsample $S_{0,T}$ but weak in the remainder of the sample. This allows us to assess whether the full sample estimator—which combines the strongly identified subsample with the weakly identified subsample—outperforms estimators based solely on the strongly identified subsample. We consider the DGP specified in (S.C.2)-(S.C.3), setting $\theta_2 = d_2/\sqrt{T}$ with $d_2 \in \{4, 8\}$ and $\theta_1 = \theta_3 = d/\sqrt{T}$ with $d \in \{16, 24, 32\}$.

The results are reported in Table 8. When the instrument is weak in the remaining part of the sample ($d_2 = 4$), $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ consistently yield lower MSE for both $\rho = 0.25$ and 0.75. The bias of the full sample estimator is smaller than that of $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ only when $d = 16$. For larger values of d, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ exhibit lower bias.

When the instrument has intermediate identification strength in the remaining sample ($d_2 = 8$), $\widehat{\beta}_{FS}$ delivers lower bias and MSE than $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ when $d = 16$. However, for $d = 24$ and 32, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ exhibits lower bias and MSE than both $\widehat{\beta}_{FS}$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$.

Overall, the results indicate that when the instrument is weak in the remaining part of the sample, $\widehat{\beta}(\widehat{\mathbf{S}}_{T,OLS})$ and $\widehat{\beta}(\widehat{\mathbf{S}}_{T,FGLS})$ outperform the full sample estimator that also incorporates the weakly identified subsample, provided that the instrument is sufficiently strong in the strongly identified subsample.

Table 7: Bias and MSE of $\widehat{\beta}_{FS}$, $\widehat{\beta}(\widehat{S}_{T,OLS})$ and $\widehat{\beta}(\widehat{S}_{T,FGLS})$

$\theta_2 = dT^{-1/2}$	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.25$, $\pi_0 = 0.6$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
Bias($\widehat{\beta}_{FS}$)	-12.00	-0.18	-0.09	-46.77	-0.24	-0.08
Bias($\widehat{\beta}(\widehat{S}_{T,OLS})$)	18.70	0.21	-0.08	8.75	0.13	-0.01
Bias($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	-2.16	-0.33	-0.18	-8.13	-0.70	-0.18
MSE($\widehat{\beta}_{FS}$)	2801.97	1.19	0.29	71860.15	1.57	0.38
MSE($\widehat{\beta}(\widehat{S}_{T,OLS})$)	4868.16	0.70	0.16	26217.10	0.92	0.22
MSE($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	179845.84	7.90	0.18	25547.87	1.22	0.25
	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.75$, $\pi_0 = 0.6$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
Bias($\widehat{\beta}_{FS}$)	-18.00	-0.82	-0.22	29.23	-1.13	-0.29
Bias($\widehat{\beta}(\widehat{S}_{T,OLS})$)	30.88	0.19	-0.07	54.22	0.22	-0.11
Bias($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	126.40	-0.64	-0.18	9.22	-1.04	-0.29
MSE($\widehat{\beta}_{FS}$)	6879.05	1.27	0.29	114152.48	1.68	0.40
MSE($\widehat{\beta}(\widehat{S}_{T,OLS})$)	12408.54	0.71	0.17	98469.13	0.98	0.23
MSE($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	397451.09	0.79	0.19	38287.46	1.18	0.25
	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.25$, $\pi_0 = 0.8$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
Bias($\widehat{\beta}_{FS}$)	-16.50	-0.07	-0.05	0.58	-0.08	-0.03
Bias($\widehat{\beta}(\widehat{S}_{T,OLS})$)	13.51	0.11	-0.07	2.60	0.20	-0.00
Bias($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	-11.13	-0.30	-0.13	-34.74	-0.41	-0.12
MSE($\widehat{\beta}_{FS}$)	7589.57	0.64	0.16	19.70	0.83	0.21
MSE($\widehat{\beta}(\widehat{S}_{T,OLS})$)	8934.70	0.51	0.12	290.25	0.67	0.17
MSE($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	1174.15	0.56	0.13	52974.88	0.80	0.18
	$\rho_e = \rho_u = 0$			$\rho_e = \rho_u = 0.50$		
$\rho = 0.75$, $\pi_0 = 0.8$	$d = 4$	$d = 16$	$d = 32$	$d = 4$	$d = 16$	$d = 32$
Bias($\widehat{\beta}_{FS}$)	-11.70	-0.43	-0.06	-6.43	-0.58	-0.16
Bias($\widehat{\beta}(\widehat{S}_{T,OLS})$)	-11.75	0.11	-0.03	-1.45	0.25	-0.07
Bias($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	-24.69	-0.50	-0.13	-49.61	-0.77	-0.23
MSE($\widehat{\beta}_{FS}$)	607.85	0.62	0.16	3808.63	0.86	0.21
MSE($\widehat{\beta}(\widehat{S}_{T,OLS})$)	4083.34	0.50	0.13	1577.48	0.70	0.17
MSE($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	8027.84	0.54	0.14	33900.00	0.77	0.18

To facilitate readability each value is multiplied by 10^2.

Table 8: Bias and MSE of $\widehat{\beta}_{FS}$, $\widehat{\beta}\left(\widehat{S}_{T,OLS}\right)$ and $\widehat{\beta}\left(\widehat{S}_{T,FGLS}\right)$

	$\rho = 0.25$			$\rho = 0.75$		
$d_2 = 4, \pi_0 = 0.6$	$d = 16$	$d = 24$	$d = 32$	$d = 16$	$d = 24$	$d = 32$
Bias($\widehat{\beta}_{FS}$)	-0.13	-0.12	-0.14	-0.55	-0.26	-0.25
Bias($\widehat{\beta}(\widehat{S}_{T,OLS})$)	0.45	0.03	-0.05	1.02	0.12	-0.05
Bias($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	-0.34	-0.19	-0.16	-0.43	-0.31	-0.21
MSE($\widehat{\beta}_{FS}$)	0.87	0.40	0.24	0.89	0.40	0.24
MSE($\widehat{\beta}(\widehat{S}_{T,OLS})$)	0.68	0.30	0.17	0.70	0.30	0.17
MSE($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	0.81	0.35	0.19	0.78	0.34	0.18
	$\rho = 0.25$			$\rho = 0.75$		
$d_2 = 8, \pi_0 = 0.6$	$d = 16$	$d = 24$	$d = 32$	$d = 16$	$d = 24$	$d = 32$
Bias($\widehat{\beta}_{FS}$)	-0.09	-0.10	-0.13	-0.40	-0.21	-0.22
Bias($\widehat{\beta}(\widehat{S}_{T,OLS})$)	1.00	0.09	-0.03	-2.70	0.36	0.08
Bias($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	0.28	-0.22	-0.16	0.55	-0.26	-0.20
MSE($\widehat{\beta}_{FS}$)	0.65	0.33	0.20	0.67	0.33	0.20
MSE($\widehat{\beta}(\widehat{S}_{T,OLS})$)	0.71	0.30	0.17	0.75	0.30	0.17
MSE($\widehat{\beta}(\widehat{S}_{T,FGLS})$)	0.96	0.35	0.19	0.82	0.34	0.18

To facilitate readability each value is multiplied by 10^2.

S.C.3 Finite-sample Size and Power of Weak Identification-Robust Tests

We study the finite-sample rejection frequencies of the proposed tests and compare their performance to the tests analyzed by Andrews, Moreira, and Stock (2006) and Magnusson and Mavroeidis (2014). The analysis is based on the same DGP described in (S.C.2)-(S.C.3) with parameters $\rho \in \{0.25, 0.5, 0.75\}$, $\gamma_1 = \gamma_2 \in \{0, 1\}$, $\theta_1 = \theta_3 = dT^{-1/2}$ with $d \in \{4, 8, 10, 16, 24\}$ and $\theta_2 = 0$. We consider values of $\pi_0 \in \{0.4, 0.6, 0.8\}$ and $T = \{200, 400\}$. Under the null hypothesis we set $\beta = 0$.

We compare the performance of our proposed test statistics $AR_T(\widehat{\mathbf{S}}_T)$, $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ with their full sample counterparts and with the test statistics Split-S, Split-CLR, qLL-S, ave-S and exp-S analyzed by Andrews, Moreira, and Stock (2006) and with the tests statistics Split-S, Split-CLR, qLL-S, ave-S and exp-S proposed by Magnusson and Mavroeidis (2014). For all tests, we consider heteroskedasticity and autocorrelation-robust versions using the Newey-West estimator with bandwidth equal to the popular rule $\left\lfloor T^{-1/3} \right\rfloor$.[39] For qLL-S, the tuning parameters c and $\widetilde{c}$ are set to 10, following the recommendation in Magnusson and Mavroeidis (2014). The significance level is fixed at 5%, and the number of Monte Carlo replications is set to 10,000 throughout the analysis.

[39] We also experiment with data-dependent bandwidths, though results are similar and therefore omitted.

Tables 9-10 report the null rejections frequencies of the various test statistics. In Table 9 the sample size is set to $T = 400$, $\rho = 0.25$ and the errors are assumed to be i.i.d.[40] The results show that qLL-S and exp-S tests systematically yield rejection rates below the nominal significance level, while the Split-CLR is systematically oversized. LM_T and $LM_T(\widehat{\mathbf{S}}_T)$ show accurate null rejection frequencies for all values of d, whereas CLR_T and $CLR_T(\widehat{\mathbf{S}}_T)$ tend to produce slightly oversized rejection rates when the instrument is weak (i.e., $d = 4$ and $d = 8$). We note that $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ have often more accurate rejection rates than their full sample counterparts. These findings are consistent across values of π_0. In Table 10 we consider a smaller sample size of $T = 200$ and $\rho \in \{0.25, 0.50, 0.75\}$. The qualitative patterns remain similar. As ρ increases $LM_T(\widehat{\mathbf{S}}_T)$ and $CLR_T(\widehat{\mathbf{S}}_T)$ become slightly more oversized than their full sample counterparts for $d = 4$, although the opposite occurs for larger d. Thus, the proposed tests demonstrate good size control, even in small samples.

In the supplement, we examine the impact of serial correlation on the null rejection rates. 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.

Finally, we turn to the comparison of size-adjusted power, as reported in Figures 7-8. Neither test achieves unit power when the instrument is weak. The results indicate that the proposed tests consistently achieve the highest size-adjusted power across all specifications considered. The power gains are substantial, averaging around approximately 20–30% and reaching 40-50% in the most favorable cases. The latter coincide with the specifications $\theta_2 = -0.5dT^{-1/2}$ and $\theta_2 = -0.5$. Consistent with the theoretical predictions, the gains are more pronounced for smaller values of π_0, provided that π_0 is not too small. Overall, these finite-sample results support our theoretical relative efficiency results.

[40] When the number of instruments is one, the AR tests are not reported, as they are numerically equivalent to the LM tests. For results with serially correlated errors, see the supplement. The conclusions are similar to i.i.d. errors.

Table 9: Finite-Sample Null Rejection Frequencies of Tests

$T = 400$	$\gamma_1 = \gamma_2 = 0$ and $\rho_e = \rho_u = 0$				$\gamma_1 = \gamma_2 = 1$ and $\rho_e = \rho_u = 0$			
$\rho = 0.25$, $\pi_0 = 0.6$	$d = 4$	$d = 8$	$d = 12$	$d = 16$	$d = 4$	$d = 8$	$d = 12$	$d = 16$
LM_T	0.057	0.057	0.053	0.056	0.059	0.060	0.058	0.056
CLR_T	0.078	0.065	0.060	0.062	0.083	0.074	0.067	0.065
$LM_T(\widehat{\mathbf{S}}_T)$	0.061	0.052	0.049	0.050	0.065	0.055	0.054	0.048
$CLR_T(\widehat{\mathbf{S}}_T)$	0.072	0.054	0.051	0.050	0.079	0.074	0.055	0.049
split − S	0.039	0.038	0.036	0.036	0.042	0.042	0.038	0.034
split − CLR	0.115	0.122	0.117	0.115	0.110	0.130	0.122	0.123
qqL − S	0.027	0.031	0.028	0.056	0.031	0.029	0.031	0.026
ave − S	0.044	0.039	0.043	0.038	0.043	0.043	0.042	0.038
exp − S	0.019	0.020	0.017	0.019	0.017	0.022	0.021	0.019
$\rho = 0.25$, $\pi_0 = 0.4$	$d = 4$	$d = 8$	$d = 12$	$d = 16$	$d = 4$	$d = 8$	$d = 12$	$d = 16$
LM_T	0.058	0.058	0.058	0.058	0.059	0.058	0.056	0.060
CLR_T	0.087	0.079	0.073	0.072	0.083	0.081	0.072	0.073
$LM_T(\widehat{\mathbf{S}}_T)$	0.061	0.060	0.065	0.064	0.065	0.065	0.062	0.066
$CLR_T(\widehat{\mathbf{S}}_T)$	0.081	0.071	0.073	0.071	0.079	0.076	0.070	0.072
split − S	0.036	0.036	0.038	0.037	0.042	0.039	0.039	0.038
split − CLR	0.104	0.119	0.120	0.121	0.110	0.119	0.119	0.122
qqL − S	0.027	0.028	0.029	0.031	0.028	0.030	0.032	0.032
ave − S	0.040	0.043	0.047	0.043	0.044	0.043	0.037	0.044
exp − S	0.015	0.017	0.017	0.018	0.017	0.020	0.020	0.019
$\rho = 0.25$, $\pi_0 = 0.8$	$d = 4$	$d = 8$	$d = 12$	$d = 16$	$d = 4$	$d = 8$	$d = 12$	$d = 16$
LM_T	0.055	0.056	0.057	0.058	0.056	0.060	0.058	0.060
CLR_T	0.069	0.061	0.060	0.060	0.076	0.067	0.062	0.062
$LM_T(\widehat{\mathbf{S}}_T)$	0.060	0.061	0.053	0.049	0.061	0.059	0.056	0.052
$CLR_T(\widehat{\mathbf{S}}_T)$	0.068	0.063	0.054	0.049	0.071	0.062	0.058	0.053
split − S	0.038	0.041	0.047	0.045	0.043	0.047	0.044	0.041
split − CLR	0.117	0.128	0135	0.135	0.122	0.133	0.133	0.128
qqL − S	0.028	0.032	0.029	0.030	0.025	0.030	0.032	0.031
ave − S	0.045	0.044	0.047	0.040	0.042	0.046	0.044	0.041
exp − S	0.028	0.019	0.017	0.020	0.018	0.022	0.019	0.018

Model M1 and M2. The null hypothesis is $H_0 : \beta = 0$.

Table 10: Finite-Sample Null Rejection Frequencies of Tests

	$\gamma_1 = \gamma_2 = 0$ and $\rho_e = \rho_u = 0$								
	$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
$T = 200$, $\pi_0 = 0.6$	$d = 4$	$d = 10$	$d = 16$	$d = 4$	$d = 10$	$d = 16$	$d = 4$	$d = 10$	$d = 16$
LM_T	0.061	0.061	0.061	0.062	0.062	0.059	0.062	0.062	0.062
CLR_T	0.084	0.073	0.070	0.085	0.073	0.063	0.080	0.071	0.070
$LM_T(\widehat{\mathbf{S}}_T)$	0.068	0.054	0.053	0.075	0.057	0.059	0.082	0.057	0.054
$CLR_T(\widehat{\mathbf{S}}_T)$	0.081	0.057	0.054	0.086	0.059	0.059	0.090	0.058	0.054
split − S	0.034	0.035	0.034	0.035	0.034	0.035	0.036	0.035	0.034
split − CLR	0.105	0.113	0.111	0.106	0.115	0.115	0.110	0.115	0.115
qqL − S	0.015	0.017	0.017	0.019	0.019	0.014	0.017	0.018	0.018
ave − S	0.035	0.038	0.035	0.038	0.036	0.039	0.036	0.039	0.042
exp − S	0.012	0.012	0.012	0.012	0.012	0.014	0.012	0.012	0.012
	$\gamma_1 = \gamma_2 = 1$ and $\rho_e = \rho_u = 0$								
	$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
$T = 200$, $\pi_0 = 0.6$	$d = 4$	$d = 10$	$d = 16$	$d = 4$	$d = 10$	$d = 16$	$d = 4$	$d = 10$	$d = 16$
LM_T	0.066	0.066	0.066	0.064	0.064	0.064	0.069	0.069	0.069
CLR_T	0.089	0.077	0.075	0.089	0.075	0.073	0.089	0.080	0.078
$LM_T(\widehat{\mathbf{S}}_T)$	0.070	0.060	0.056	0.082	0.061	0.056	0.095	0.058	0.056
$CLR_T(\widehat{\mathbf{S}}_T)$	0.085	0.063	0.057	0.094	0.064	0.056	0.105	0.060	0.057
split − S	0.040	0.040	0.039	0.041	0.040	0.038	0.040	0.037	0.038
split − CLR	0.110	0.123	0.121	0.112	0.112	0.119	0.118	0.119	0.120
qqL − S	0.020	0.021	0.021	0.024	0.019	0.024	0.025	0.023	0.024
ave − S	0.041	0.040	0.039	0.042	0.036	0.037	0.040	0.041	0.038
exp − S	0.014	0.015	0.015	0.015	0.016	0.015	0.014	0.016	0.014

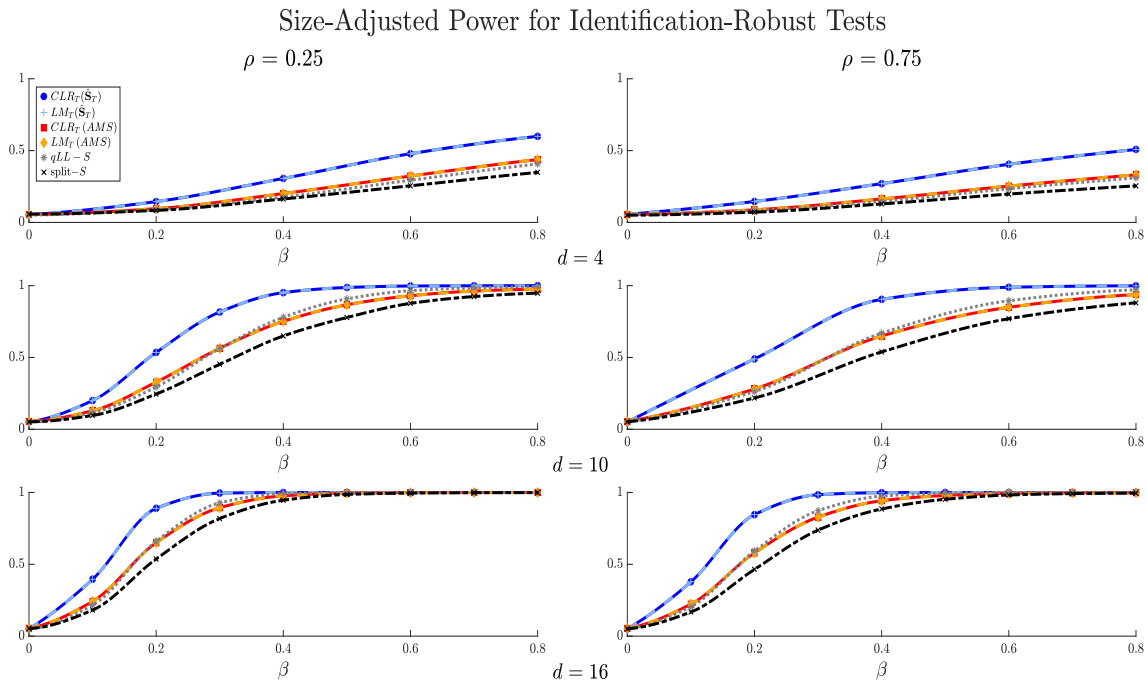

Figure 7: Size-adjusted power of identification robust tests for $\theta_2 = dT^{-1/2}$ with $T = 200$ and $\pi_0 = 0.6$.

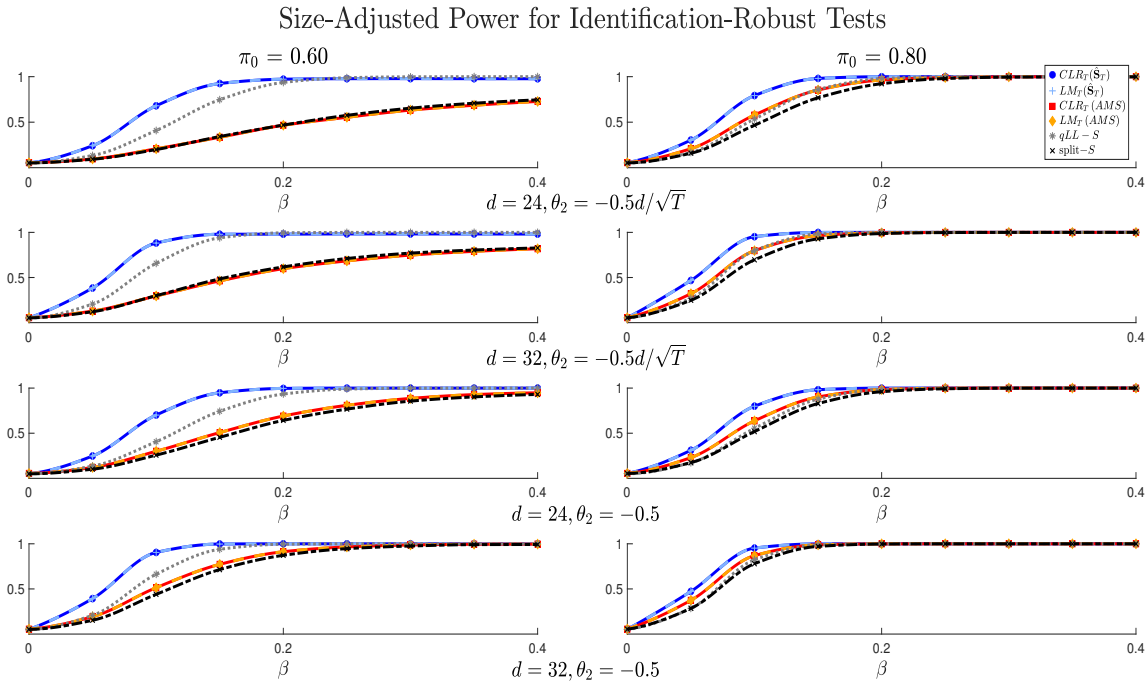

Figure 8: Size-adjusted power of identification robust tests for $\theta_2 = -0.5dT^{-1/2}$ and $\theta_2 = -0.5$ wirh $T = 200$ and $\rho = 0.25$.

S.D Mathematical Proofs

S.D.1 Proofs of the Results of Sections 2-3

S.D.1.1 Preliminary Lemmas

Lemma S.D.1. *Let Assumption 2.4 hold, $g_t(\cdot)$ be locally absolutely continuous on $\mathbf{D} \subseteq \mathbb{R}$ and $\mathbb{E}[\int_{\mathbf{D}} |\partial g_t(d)/\partial d| \mathrm{d}d | \widetilde{V}_t] < \infty$. For $t \in \mathbf{S}_{0,T}$, $\widetilde{v}$ in the support of $\widetilde{V}_t$, and z, $z' \in \mathbf{Z}$, we have*

$$\begin{aligned}
&\mathbb{E}\left(g_t\left(D_t\left(z'\right)\right) - g_t\left(D_t\left(z\right)\right) \mid \widetilde{V}_t = \widetilde{v}\right) \\
&\quad = \int_{\mathbf{D}} \mathbb{E}\left(\frac{\partial}{\partial d} g_t\left(d\right) \mid D_t\left(z\right) \leq d \leq D_t\left(z'\right), \widetilde{V}_t = \widetilde{v}\right) \mathbb{P}\left(D_t\left(z\right) \leq d \leq D_t\left(z'\right) \mid \widetilde{V}_t = \widetilde{v}\right) \mathrm{d}d.
\end{aligned}$$

Proof of Lemma S.D.1. Suppose that Assumption 2.4 holds with $D_t(z') \geq D_t(z)$. We have

$$\begin{aligned}
&\mathbb{E}\left(g_t\left(D_t\left(z'\right)\right) - g_t\left(D_t\left(z\right)\right) \mid \widetilde{V}_t = \widetilde{v}\right) \\
&= \mathbb{E}\left(\int_{D_t(z)}^{D_t(z')} \frac{\partial g_t}{\partial d}\left(d\right) \mathrm{d}d \mid \widetilde{V}_t = \widetilde{v}\right) \\
&= \mathbb{E}\left(\int_{\mathbf{D}} \frac{\partial g_t}{\partial d}\left(d\right) \mathbf{1}\left\{D_t\left(z\right) \leq d \leq D_t\left(z'\right)\right\} \mathrm{d}d \mid \widetilde{V}_t = \widetilde{v}\right) \\
&= \int_{\mathbf{D}} \mathbb{E}\left(\frac{\partial g_t}{\partial d}\left(d\right) \mathbf{1}\left\{D_t\left(z\right) \leq d \leq D_t\left(z'\right)\right\} \mid \widetilde{V}_t = \widetilde{v}\right) \mathrm{d}d \\
&= \int_{\mathbf{D}} \mathbb{E}\left(\frac{\partial g_t}{\partial d}\left(d\right) \mid D_t\left(z\right) \leq d \leq D_t\left(z'\right), \widetilde{V}_t = \widetilde{v}\right) \mathbb{P}\left(D_t\left(z\right) \leq d \leq D_t\left(z'\right) \mid \widetilde{V}_t = \widetilde{v}\right) \mathrm{d}d,
\end{aligned}$$

where the first equality follows from local absolute continuity and the fundamental theorem of calculus and the third equality follows from Fubini's theorem and integrability. □

S.D.1.2 Proof of Proposition 2.1

Consider first the denominator of $\beta_{\pi,t,h}(\widetilde{v})$. We appeal to Assumption 2.1 and apply Lemma S.D.1 with $g_t(d) = d$ to obtain for $t \in \mathbf{S}_{0,T}$,

$$\mathbb{E}\left(D_t \mid Z_t = z', \widetilde{V}_t = \widetilde{v}\right) - \mathbb{E}\left(D_t \mid Z_t = z, \widetilde{V}_t = \widetilde{v}\right) = \int_{\mathbf{D}} \mathbb{P}\left(D_t\left(z\right) \leq d \leq D_t\left(z'\right) \mid \widetilde{V}_t = \widetilde{v}\right) \mathrm{d}d.$$

Noting that $Y_{t,h}$ is a function of $D_t = d$, we appeal to Assumptions 2.1-2.2 and 2.5(i) and apply Lemma S.D.1 with $g_t(d) = Y^*_{t,h}(d)$ to the numerator of $\beta_{\pi,t,h}(\widetilde{v})$ to obtain,

$$\begin{aligned}
&\mathbb{E}\left(Y_{t+h} \mid Z_t = z', \widetilde{V}_t = \widetilde{v}\right) - \mathbb{E}\left(Y_{t+h} \mid Z_t = z, \widetilde{V}_t = \widetilde{v}\right) \\
&\quad = \int_{\mathbf{D}} \mathbb{E}\left(\frac{\partial Y^*_{t,h}}{\partial d}(d) \mid D_t(z) \leq d \leq D_t(z'), \widetilde{V}_t = \widetilde{v}\right) \times \mathbb{P}\left(D_t(z) \leq d \leq D_t(z') \mid \widetilde{V}_t = \widetilde{v}\right) \mathrm{d}d.
\end{aligned}$$

Using the derived expressions for the numerator and denominator of $\beta_{\pi,t,h}(\widetilde{v})$ yields the expression given in the proposition, which is well-defined by Assumption 2.3. Note that by definition $w_t(d \mid \widetilde{v}) \geq 0$ and $\int_{\mathbf{D}} w_t(d \mid \widetilde{v})\, \mathrm{d}d = 1$. □

S.D.1.3 Proof of Theorem 2.1

Lemma S.D.2. *Let Assumption 2.6 hold. For each t, $D_t(1) > D_t(0)$ with probability one if and only if $\mathbb{E}(D_t(1)) > \mathbb{E}(D_t(0))$.*

Proof of Lemma S.D.2. First, note that $\mathbb{P}(D_t(1) > D_t(0)) = 1$ immediately implies $\mathbb{E}(D_t(1)) > \mathbb{E}(D_t(0))$. To see the reverse implication, suppose $\mathbb{E}(D_t(1)) > \mathbb{E}(D_t(0))$. Then it must be the case that $\mathbb{P}(D_t(1) > D_t(0)) > 0$. But then Assumption 2.6 immediately implies $\mathbb{P}(D_t(0) < D_t(1)) = 1$. □

Consider first the policy sample. Given $t \in \mathbf{P}$ we have $Z_t = 1$ and $D_t = D_t(1)$. By Lemma S.D.2 $t \in \mathbf{P}$ is a complier if and only if $\mathbb{E}(D_t(1)) > \mathbb{E}(D_t(0))$. By Assumption 2.7(ii) we have $\mathbb{E}(D_t(0)) - \mathbb{E}(D_{t-1}(0)) = 0$. The latter implies $t \in \mathbf{P}$ is a complier if and only if

$$\begin{aligned}
\mathbb{E}(D_t(1)) - \mathbb{E}(D_{t-1}(0)) &= \mathbb{E}(D_t(1)) - \mathbb{E}(D_t(0)) \\
&> \mathbb{E}(D_t(0)) - \mathbb{E}(D_t(0)) = 0.
\end{aligned}$$

By Assumption 2.7(i), $\overline{D}_{C,t-1,n_0} \xrightarrow{\mathbb{P}} \mathbb{E}(D_{t-1}(0))$ as $n_0 \to \infty$. By Assumption 2.8(i), $\overline{D}_{P,t,n_1} \xrightarrow{\mathbb{P}} \mathbb{E}(D_t(1))$ as $n_1 \to \infty$. Thus, $\overline{D}_{P,t,n_1} - \overline{D}_{C,t-1,n_0} \xrightarrow{\mathbb{P}} c$ as $n_0, n_1 \to \infty$, where $c > 0$ if and only if $t \in \mathbf{P}$ is a complier.

Now consider the control sample. Since $t \in \mathbf{C}$ we have $Z_t = 0$ and $D_t = D_t(0)$. Using Assumption 2.8(ii) and Lemma S.D.2 we have $t \in \mathbf{C}$ is a complier if and only if

$$\mathbb{E}\left(D_{s^*(t)}(1)\right) - \mathbb{E}(D_t(0)) = \mathbb{E}(D_t(1)) - \mathbb{E}(D_t(0))$$

$$> \mathbb{E}\left(D_t\left(0\right)\right) - \mathbb{E}\left(D_t\left(0\right)\right) = 0.$$

By Assumption 2.7(i), $\overline{D}_{C,t,n_0} \xrightarrow{\mathbb{P}} \mathbb{E}\left(D_t\left(0\right)\right)$ as $n_0 \to \infty$. By Assumption 2.8(i), $\overline{D}_{P,s^*(t),n_1} \xrightarrow{\mathbb{P}} \mathbb{E}(D_{s^*(t)}\left(1\right))$ as $n_1 \to \infty$. Thus, $\overline{D}_{P,s^*(t),n_1} - \overline{D}_{C,t,n_0} \xrightarrow{\mathbb{P}} \tilde{c}$ as $n_0, n_1 \to \infty$, where $\tilde{c} > 0$ if and only if $t \in \mathbf{C}$ is a complier. □

S.D.1.4 Proof of Proposition 2.2

Under Assumption 2.4 with $D_t(1) \geq D_t(0)$, non-compliers are characterized by $\mathbb{P}(D_t(1) = D_t(0)) = 1$ using Assumption 2.6 so that any non-complier cannot belong to $\mathbf{S}_{0,T}$. On the other hand, if t is a complier, $\mathbb{E}[D_t(1)] \neq \mathbb{E}[D_t(0)]$ since $\mathbb{P}(D_t(1) > D_t(0)) = 1$ by the definition of a complier so that $t \in \mathbf{S}_{0,T}$. □

S.D.1.5 Proof Proposition 2.3

For $t \in \mathcal{NC}^s_{\mathbf{C}}$, $Z_t = 0$ so that $Y_t = Y_t^*(D_t(0), 0)$ and Assumption 2.9(iii) implies

$$|\mathcal{NC}^s_{\mathbf{C}}|^{-1} \sum_{t \in \mathcal{NC}^s_{\mathbf{C}}} Y_t \xrightarrow{\mathbb{P}} \mathbb{E}[Y_t | t \in \mathcal{NC}^s_{\mathbf{C}}] = \mathbb{E}[Y_t^*(D_t(0), 0) | t \in \mathcal{NC}^s_{\mathbf{C}}] = \mathbb{E}[Y_t^*(D_t, 0) | t \in \mathcal{NC}^s_{\mathbf{C}}],$$

as $|\mathcal{NC}^s_{\mathbf{C}}| \to \infty$ since t is a non-complier. Similarly, Assumption 2.9(iii) implies

$$|\mathcal{NC}^s_{\mathbf{P}}|^{-1} \sum_{t \in \mathcal{NC}^s_{\mathbf{P}}} Y_t \xrightarrow{\mathbb{P}} \mathbb{E}[Y_t^*(D_t, 1) | t \in \mathcal{NC}^s_{\mathbf{P}}]$$

as $|\mathcal{NC}^s_{\mathbf{P}}| \to \infty$. But Assumption 2.9(i)–(ii) implies $\mathbb{E}[Y_t^*(D_t, 0) | t \in \mathcal{NC}^s_{\mathbf{C}}] = \mathbb{E}[Y_t^*(D_t, 0) | t \in \mathcal{NC}^s_{\mathbf{P}}]$, which is in turn equal to $\mathbb{E}[Y_t^*(D_t, 1) | t \in \mathcal{NC}^s_{\mathbf{P}}]$ under Assumption 2.2. □

S.D.2 Proofs of the Results of Sections 4, 6 and S.B

S.D.2.1 Proof of Theorem 4.1

Suppose that $\mathbf{S}_T \in \Xi_{\epsilon,\pi,m,T}$. Note that $\widetilde{Z}\left(S_T\right) = M_{S_T X} S_T Z$ and

$$\widetilde{D}\left(S_T\right) = M_{S_T X} S_T D = M_{S_T X} S_T \left(X \gamma_2 + e\right) = M_{S_T X} S_T e$$

under $H_{\theta,0}$. Thus,

$$\widetilde{D}\left(S_T\right)'\widetilde{Z}\left(S_T\right) = (S_T e)' M_{S_T X} M_{S_T X} S_T Z = (S_T e)' M_{S_T X} S_T Z = (S_T e)' \widetilde{Z}(S_T).$$

By Assumptions 4.1-4.2 we have

$$T^{-1/2}\widetilde{Z}\left(S_T\right)'\widetilde{D}\left(S_T\right) = T^{-1/2}\sum_{t\in\mathbf{S}_T}\widetilde{Z}_t e_t \Rightarrow B\left(\lambda_{L,1},\,\lambda_{R,m}\right),$$

where $B\left(\lambda_{L,1},\,\lambda_{R,m}\right)$ is q-dimensional Brownian motion with covariance $J(S) = \lim_{T\to\infty}(\pi T)^{-1}$ $\mathrm{Var}((S_T Z)' M_{S_T X} S_T e)$ and $S = \lim_{T\to\infty} S_T$. Since $\widehat{J}(S_T) \xrightarrow{\mathbb{P}} J\left(S\right)$ uniformly by Assumption 4.3, we have

$$F_T\left(\mathbf{S}_T\right) \Rightarrow \frac{B\left(\lambda_{L,1},\,\lambda_{R,m}\right)' J(S)^{-1} B\left(\lambda_{L,1},\,\lambda_{R,m}\right)}{q\pi} = \frac{1}{q\pi}\sum_{i=1}^{m}\left\|\left(W_q\left(\lambda_{R,i}\right) - W_q\left(\lambda_{L,i}\right)\right)\right\|^2.$$

The result then follows from the continuous mapping theorem. □

S.D.2.2 Proof of Theorem 6.1

Lemma S.D.3. *Let Assumptions 6.1-6.3 hold and suppose $\theta = c/T^{1/2}$ for some nonstochastic $c \in \mathbb{R}^q$. We have $\widehat{\Sigma}_v\left(\mathbf{S}_T\right) \xrightarrow{\mathbb{P}} \Sigma_v$ uniformly in $\mathbf{S}_T \in \mathcal{S}$.*

Proof of Lemma S.D.3. Recall that C_T is the selection matrix that corresponds to $\mathbf{S}_T$. Using $y = \overline{Z}\left(C_{0,T}\right)\theta a' + X\eta + v$, $P_X\overline{Z}\left(C_T\right) = 0$ and $P_{\overline{Z}(C_T)}X = 0$, we have

$$\begin{aligned}\widehat{v}\left(\mathbf{S}_T\right) &= y - P_{\overline{Z}(C_T)}y - P_X y = y - P_{\overline{Z}(C_T)}y - X\eta - P_X v \\ &= \overline{Z}\left(C_{0,T}\right)\theta a' + v - P_{\overline{Z}(C_T)}y - P_X v \\ &= \overline{Z}\left(C_{0,T}\right)\theta a' + v - P_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' - P_{\overline{Z}(C_T)}v - P_X v \\ &= M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' + v - P_{\overline{Z}(C_T)}v - P_X v.\end{aligned} \tag{S.D.1}$$

Then, using $\overline{Z}\left(C_T\right)' X = 0$, we have

$$\begin{aligned}&\widehat{v}\left(\mathbf{S}_T\right)'\widehat{v}\left(\mathbf{S}_T\right) \\ &\quad= \left(M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' + v - P_{\overline{Z}(C_T)}v - P_X v\right)' \\ &\qquad\times\left(M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' + v - P_{\overline{Z}(C_T)}v - P_X v\right)\end{aligned} \tag{S.D.2}$$

$$= a\theta'\overline{Z}\left(C_{0,T}\right)' M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' + a\theta'\overline{Z}\left(C_{0,T}\right)' M_{\overline{Z}(C_T)}v$$
$$+ v' M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' + v'v - v' P_{\overline{Z}(C_T)}v - v' P_X v.$$

Using Assumptions 6.1, 6.3 and $\theta = cT^{-1/2}$ we have

$$\begin{aligned} &T^{-1}v' M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' \qquad \text{(S.D.3)} \\ &\quad = T^{-1}v'\overline{Z}\left(C_{0,T}\right)\frac{c}{\sqrt{T}}a' - T^{-1}v'\overline{Z}\left(C_T\right)\left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_{0,T}\right)\frac{c}{\sqrt{T}}a' \\ &\quad = T^{-1/2}O_{\mathbb{P}}\left(1\right)\frac{c}{\sqrt{T}}a' - T^{-1/2}\left(T^{-1/2}v'\overline{Z}\left(C_T\right)\right)\left(T^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}T^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_{0,T}\right)\frac{c}{\sqrt{T}}a' \\ &\quad = O_{\mathbb{P}}\left(T^{-1}\right), \qquad \text{and} \end{aligned}$$

$$\begin{aligned} &T^{-1}a\theta'\overline{Z}\left(C_{0,T}\right)' M_{\overline{Z}(C_T)}\overline{Z}\left(C_{0,T}\right)\theta a' \qquad \text{(S.D.4)} \\ &= T^{-1}a\theta'\overline{Z}\left(C_{0,T}\right)'\overline{Z}\left(C_{0,T}\right)\theta a' - T^{-1}a\theta'\overline{Z}\left(C_{0,T}\right)'\overline{Z}\left(C_T\right)\left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)\overline{Z}\left(C_{0,T}\right)\theta a' \\ &= a\frac{c'}{\sqrt{T}}T^{-1}\overline{Z}\left(C_{0,T}\right)'\overline{Z}\left(C_{0,T}\right)\frac{c}{\sqrt{T}}a' \\ &\quad - a\frac{c'}{\sqrt{T}}T^{-1}\overline{Z}\left(C_{0,T}\right)'\overline{Z}\left(C_T\right)\left(T^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}T^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_{0,T}\right)\frac{c}{\sqrt{T}}a' = O_{\mathbb{P}}\left(T^{-1}\right) \end{aligned}$$

so that

$$T^{-1}\widehat{v}'\left(\mathbf{S}_T\right)\widehat{v}\left(\mathbf{S}_T\right) - \Sigma_v = \left(T^{-1}v'v - \Sigma_v\right) - T^{-1}v' P_{\overline{Z}(C_T)}v - T^{-1}v' P_X v + O_{\mathbb{P}}\left(T^{-1}\right). \quad \text{(S.D.5)}$$

By Assumption 6.2, the first term on the right-hand side of (S.D.5) converges in probability to zero. The second term satisfies

$$\begin{aligned} T^{-1}v' P_{\overline{Z}(C_T)}v &\leq T^{-1}v' P_{C_T Z}v \\ &= T^{-1}\left(T^{-1/2}v' C_T Z\right)\left(T^{-1}\left(C_T Z\right)' C_T Z\right)^{-1}\left(T^{-1/2}\left(C_T Z\right)' v\right) = o_{\mathbb{P}}\left(1\right), \end{aligned}$$

where the inequality holds because the span of $\overline{Z}\left(C_T\right) = M_X C_T Z$ is contained in the span of $C_T Z$ and the $o_{\mathbb{P}}\left(1\right)$ result follows from Assumptions 6.1 and 6.3. All convergence results above hold uniformly in $\mathbf{S}_T$. The argument for the third term of (S.D.5) is analogous. □

Lemma S.D.3 shows that $\widehat{\Sigma}_v\left(\mathbf{S}_T\right)$ is consistent for Σ_v for all $\mathbf{S}_T \in \mathcal{S}$. The convergence in the lemma occurs uniformly over all true parameters β, c, γ, ϕ and over $\mathbf{S}_T \in \mathcal{S}$. Thus,

estimation based on any partition $\mathbf{S}_T \in \mathcal{S}$ leads to residuals $\{\widehat{v}(\mathbf{S}_T)\}$ that can be used to construct a consistent estimate $\widehat{\Sigma}_v(\mathbf{S}_T)$ for Σ_v under weak instruments.

For some partitions $\mathbf{S}, \mathbf{S}' \subseteq (0,1]$, partition $Q(\mathbf{S}, \mathbf{S}')$ conformably with

$$w(\mathbf{S}_T)'w(\mathbf{S}_T') = \begin{bmatrix} Z(C_T)'Z(C_T') & Z(C_T)'X \\ X'Z(C_T') & X'X \end{bmatrix},$$

where $Z(C_T) = C_T Z$, so that

$$Q(\mathbf{S}, \mathbf{S}') = \begin{bmatrix} Q_{11}(\mathbf{S}, \mathbf{S}') & Q_{12}(\mathbf{S}) \\ Q_{21}(\mathbf{S}') & Q_{22} \end{bmatrix},$$

and let $Q(\mathbf{S}) = Q(\mathbf{S}, \mathbf{S})$. In addition, note that

$$\begin{aligned} &\Sigma_{N_1}(\mathbf{S}, \mathbf{S}') = J(\mathbf{S}) B_0 \Psi(\mathbf{S}, \mathbf{S}') B_0' J(\mathbf{S}')', \quad \Sigma_{N_1 N_2}(\mathbf{S}, \mathbf{S}') = J(\mathbf{S}) A_0 \Psi(\mathbf{S}, \mathbf{S}') B_0' J(\mathbf{S}')', \\ &\Sigma_{N_2}^*(\mathbf{S}, \mathbf{S}') = J(\mathbf{S}) A_0 \Psi(\mathbf{S}, \mathbf{S}') A_0' J(\mathbf{S}')' \end{aligned} \tag{S.D.6}$$

for $J(\mathbf{S}) = \left[I_q : -Q_{12}(\mathbf{S}) Q_{22}^{-1}\right]$, $B_0 = (b_0' \otimes I_{q+p})$ and $A_0 = (\Sigma_v^{-1} a_0)' \otimes I_{q+p}$.[41]

Finally, let $N_{1,\infty}(\cdot)$ and $N_{2,\infty}(\cdot)$ be independent q-dimensional Gaussian processes indexed by $\mathbf{S} \subseteq (0,1]$ with

$$\begin{aligned} &N_{1,\infty}(\mathbf{S}) \sim \mathscr{N}\left(\Sigma_{N_1}^{-1/2}(\mathbf{S}) \Sigma_{\overline{Z}}(\mathbf{S}, \mathbf{S}_0) c a' b_0,\ I_q\right), \\ &N_{2,\infty}(\mathbf{S}) \sim \mathscr{N}\left(\Sigma_{N_2}^{-1/2}(\mathbf{S}) \left(\Sigma_{\overline{Z}}(\mathbf{S}, \mathbf{S}_0) c a' \Sigma_v^{-1} a_0 - \Sigma_{N_1 N_2}(\mathbf{S}) \Sigma_{N_1}^{-1}(\mathbf{S}) \Sigma_{\overline{Z}}(\mathbf{S}, \mathbf{S}_0) c a' b_0\right),\ I_q\right), \end{aligned} \tag{S.D.7}$$

where $\Sigma_{\overline{Z}}(\mathbf{S}, \mathbf{S}') = Q_{11}(\mathbf{S}, \mathbf{S}') - Q_{12}(\mathbf{S}) Q_{22}^{-1} Q_{21}(\mathbf{S}')$.

Lemma S.D.4. *Let Assumptions 6.1-6.5 hold and suppose $\theta = c/T^{1/2}$ for some nonstochastic $c \in \mathbb{R}^q$. Then, for $\mathbf{S}_T \in \mathcal{S}$ and $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, we have $(N_{1,T}(\mathbf{S}_T), N_{2,T}(\mathbf{S}_T)) \Rightarrow (N_{1,\infty}(\mathbf{S}), N_{2,\infty}(\mathbf{S}))$.*

Proof of Lemma S.D.4. By Assumption 6.1,

$$T^{-1}\overline{Z}(C_T)'\overline{Z}(C_T') = T^{-1}Z(C_T)'Z(C_T') - T^{-1}Z(C_T)'P_X Z(C_T') \xrightarrow{\mathbb{P}} \Sigma_{\overline{Z}}(\mathbf{S}, \mathbf{S}'), \tag{S.D.8}$$

[41] See (S.D.9) and (S.D.12) for details.

uniformly in $\mathbf{S}_T, \mathbf{S}_T' \in \mathcal{S}$. By Assumptions 6.1 and 6.3, we have uniformly in $\mathbf{S}_T \in \mathcal{S}$,

$$\begin{aligned} T^{-1/2}\overline{Z}\left(C_T\right)' vb_0 &= T^{-1/2}\left(Z\left(C_T\right) - P_X Z\left(C_T\right)\right)' vb_0 \\ &= T^{-1/2}\left(Z\left(C_T\right) - XQ_{22}^{-1}Q_{21}\left(\mathbf{S}_T\right)\right)' vb_0 + o_{\mathbb{P}}\left(1\right) \\ &= \left[I_q : -Q_{12}\left(\mathbf{S}_T\right)Q_{22}^{-1}\right] T^{-1/2} w\left(\mathbf{S}_T\right)' vb_0 + o_{\mathbb{P}}\left(1\right) \\ &= \left[I_q : -Q_{12}\left(\mathbf{S}_T\right)Q_{22}^{-1}\right]\left(b_0' \otimes I_{q+p}\right) T^{-1/2}\mathrm{vec}\left(w(\mathbf{S}_T)'v\right), \\ &\Rightarrow J(\mathbf{S})B_0\mathscr{G}(\mathbf{S}). \end{aligned} \tag{S.D.9}$$

Using (S.D.6), (S.D.8), (S.D.9) and Assumption 6.4,

$$\begin{aligned} N_{1,T}\left(\mathbf{S}_T\right) &= \widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_T\right) T^{-1/2}\overline{Z}\left(C_T\right)'\left(T^{-1/2}\overline{Z}\left(C_{0,T}\right)ca' + v\right) b_0 \\ &\Rightarrow \Sigma_{N_1}^{-1/2}\left(\mathbf{S}\right)\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)ca'b_0 + \Sigma_{N_1}^{-1/2}\left(\mathbf{S}\right)J(\mathbf{S})B_0\mathscr{G}(\mathbf{S}) \sim N_{1,\infty}\left(\mathbf{S}\right) \end{aligned} \tag{S.D.10}$$

since $J(\mathbf{S})B_0\mathscr{G}(\mathbf{S}) \sim \mathscr{N}(0, \Sigma_{N_1}(\mathbf{S}))$. Similarly, using Lemma S.D.3, Assumptions 6.1, 6.3 and 6.4, (S.D.6) and (S.D.10), we have

$$\begin{aligned} &N_{2,T}\left(\mathbf{S}_T\right) \\ &\quad= \widehat{\Sigma}_{N_2}^{-1/2}\left(\mathbf{S}_T\right) \\ &\qquad\times\left(T^{-1/2}\overline{Z}\left(C_T\right)'\left(T^{-1/2}\overline{Z}\left(C_{0,T}\right)ca' + v\right)\widehat{\Sigma}_v^{-1}\left(\mathbf{S}_T\right)a_0 - \widehat{\Sigma}_{N_1N_2}\left(\mathbf{S}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_T\right)N_{1,T}\left(\mathbf{S}_T\right)\right) \\ &\quad= \widehat{\Sigma}_{N_2}^{-1/2}\left(\mathbf{S}_T\right) \\ &\qquad\times\left(\left(T^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_{0,T}\right)ca' + T^{-1/2}\overline{Z}\left(C_T\right)'v\right)\Sigma_v^{-1}a_0 - \widehat{\Sigma}_{N_1N_2}\left(\mathbf{S}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_T\right)N_{1,T}\left(\mathbf{S}_T\right)\right) + o_{\mathbb{P}}\left(1\right) \\ &\quad\Rightarrow \Sigma_{N_2}^{-1/2}\left(\mathbf{S}\right)\left(\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)ca'\Sigma_v^{-1}a_0\right) + \Sigma_{N_2}^{-1/2}\left(\mathbf{S}\right)J\left(\mathbf{S}\right)A_0\mathscr{G}(\mathbf{S}) \\ &\qquad- \Sigma_{N_2}^{-1/2}\left(\mathbf{S}\right)\left(\Sigma_{N_1N_2}\left(\mathbf{S}\right)\Sigma_{N_1}^{-1}\left(\mathbf{S}\right)\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)ca'b_0\right) \\ &\qquad- \Sigma_{N_2}^{-1/2}\left(\mathbf{S}\right)\left(\Sigma_{N_1N_2}\left(\mathbf{S}\right)\Sigma_{N_1}^{-1}\left(\mathbf{S}\right)J\left(\mathbf{S}\right)B_0\mathscr{G}(\mathbf{S})\right) \sim N_{2,\infty}\left(\mathbf{S}\right), \end{aligned} \tag{S.D.11}$$

where $T^{-1}\overline{Z}\left(C_T\right)'\overline{Z}\left(C_{0,T}\right) \xrightarrow{\mathbb{P}} \Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)$ uniformly over $\mathbf{S}_T \in \mathcal{S}$ by (S.D.8) and

$$T^{-1/2}\overline{Z}\left(C_T\right)' v\Sigma_v^{-1}a_0 \Rightarrow \left[I_q : -Q_{12}\left(\mathbf{S}\right)Q_{22}^{-1}\right]\left(a_0'\Sigma_v^{-1} \otimes I_{q+p}\right)\mathscr{G}(\mathbf{S}) \tag{S.D.12}$$

in analogy with the arguments that show (S.D.9). The distributional equivalence of the limit

in (S.D.11) can be seen after noting

$$\begin{aligned}
&\operatorname{Var}\left(J(\mathbf{S})\, A_0 \mathscr{G}(\mathbf{S}) - \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, J(\mathbf{S})\, B_0 \mathscr{G}(\mathbf{S})\right) && \text{(S.D.13)}\\
&\quad = J(\mathbf{S})\, A_0 \Psi(\mathbf{S},\mathbf{S}))\, A_0' J(\mathbf{S})' - J(\mathbf{S})\, A_0 \Psi(\mathbf{S},\mathbf{S}))\, B_0' J(\mathbf{S})'\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1 N_2}(\mathbf{S})'\\
&\qquad - \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, J(\mathbf{S})\, B_0 \Psi(\mathbf{S},\mathbf{S}))\, A_0' J(\mathbf{S})'\\
&\qquad + \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, J(\mathbf{S})\, B_0 \Psi(\mathbf{S},\mathbf{S}))\, B_0' J(\mathbf{S})'\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1 N_2}(\mathbf{S})' && \text{(S.D.14)}\\
&\quad = \Sigma_{N_2}^{*}(\mathbf{S}) - \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1 N_2}(\mathbf{S})' - \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1 N_2}(\mathbf{S})'\\
&\qquad + \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1 N_2}(\mathbf{S})'\\
&\quad = \Sigma_{N_2}^{*}(\mathbf{S}) - \Sigma_{N_1 N_2}(\mathbf{S})\, \Sigma_{N_1}^{-1}(\mathbf{S})\, \Sigma_{N_1 N_2}(\mathbf{S})' = \Sigma_{N_2}(\mathbf{S}).
\end{aligned}$$

The weak convergence in (S.D.9) occurs jointly with that in (S.D.11) since $N_{1,T}(\cdot)$ and $N_{2,T}(\cdot)$ are functions of the same data. And finally, they are asymptotically independent since they are asymptotically Gaussian and

$$\begin{aligned}
&\operatorname{Cov}\left(\Sigma_{N_1}^{-1/2}(\mathbf{S})\, J(\mathbf{S}) B_0 \mathscr{G}(\mathbf{S}), \Sigma_{N_2}^{-1/2}(\mathbf{S}')\left(J(\mathbf{S}')\, A_0 - \Sigma_{N_1 N_2}(\mathbf{S}')\, \Sigma_{N_1}^{-1}(\mathbf{S}')\, J(\mathbf{S}')\, B_0\right)\mathscr{G}(\mathbf{S}')\right)\\
&\quad = \Sigma_{N_1}^{-1/2}(\mathbf{S})\, J(\mathbf{S}) B_0 \Psi(\mathbf{S},\mathbf{S}')\left(A_0' J(\mathbf{S}')' - B_0' J(\mathbf{S}')'\, \Sigma_{N_1}^{-1}(\mathbf{S}')\, \Sigma_{N_1 N_2}(\mathbf{S}')'\right)\Sigma_{N_2}^{-1/2}(\mathbf{S}')\\
&\quad = \Sigma_{N_1}^{-1/2}(\mathbf{S})\left(\Sigma_{N_1 N_2}(\mathbf{S}',\mathbf{S})' - \Sigma_{N_1}(\mathbf{S},\mathbf{S}')\, \Sigma_{N_1}^{-1}(\mathbf{S}')\, \Sigma_{N_1 N_2}(\mathbf{S}')'\right)\Sigma_{N_2}^{-1/2}(\mathbf{S}')\\
&\quad = \pi(\mathbf{S}\cap\mathbf{S}')\Sigma_{N_1}^{-1/2}(\mathbf{S})\left(\Sigma_{N_1 N_2}' - \Sigma_{N_1}\Sigma_{N_1}^{-1}\Sigma_{N_1 N_2}'\right)\Sigma_{N_2}^{-1/2}(\mathbf{S}') = 0
\end{aligned}$$

for any $\mathbf{S}$, $\mathbf{S}' \subseteq (0,1]$ by Assumption 6.5. □

Inspection of the proof shows that the results hold uniformly over compact sets of true values of β and c (including the zero vector) and over arbitrary sets of true γ and ϕ values.

Proof of Theorem 6.1. Let

$$\begin{aligned}
&M_\infty(\mathbf{S}) = [N_{1,\infty}(\mathbf{S}) : N_{2,\infty}(\mathbf{S})]'\,[N_{1,\infty}(\mathbf{S}) : N_{2,\infty}(\mathbf{S})], && \text{(S.D.15)}\\
&\overline{M}_{1,\infty}(\mathbf{S}) = \left(N_{1,\infty}(\mathbf{S})'\, N_{1,\infty}(\mathbf{S}),\ N_{1,\infty}(\mathbf{S})'\, N_{2,\infty}(\mathbf{S})\right)',\\
&M_{2,\infty}(\mathbf{S}) = N_{2,\infty}(\mathbf{S})'\, N_{2,\infty}(\mathbf{S}), \quad M_{1,2,\infty}(\mathbf{S}) = N_{1,\infty}(\mathbf{S})'\, N_{2,\infty}(\mathbf{S}), \quad M_{1,\infty}(\mathbf{S}) = N_{1,\infty}(\mathbf{S})'\, N_{1,\infty}(\mathbf{S}).
\end{aligned}$$

By Lemma S.D.4 $N_{1,T}(\cdot)$ and $N_{2,T}(\cdot)$ are asymptotically independent. $\widehat{\mathbf{S}}_T$ depends on $N_{2,T}(\cdot)$ only. Thus, $N_{1,\infty}(\cdot)$ and $\widehat{\mathbf{S}}_T$ are asymptotically independent. Using Lemma S.D.4 we yield that under H_0 $N_{1,\infty}(\widehat{\mathbf{S}}_T)$ is Gaussian with zero mean and variance I_q. Thus, $M_{1,\infty}(\mathbf{S}_T) \sim M_{1,\infty}$ for all $\mathbf{S}_T \in \mathcal{S}$. Part (i) follows by using the continuous mapping theorem.

For part (ii), note that conditional on $N_{2,\infty}(\cdot)$ Lemma S.D.4 implies that $M_{1,2,\infty}(\widehat{\mathbf{S}}_T)$ is Gaussian with zero mean and variance $N_{2,\infty}(\widehat{\mathbf{S}}_T)'N_{2,\infty}(\widehat{\mathbf{S}}_T)$. The result then follows from the continuous mapping theorem.

We now move to part (iii). By Lemma S.D.4 $N_{1,T}(\cdot)$ and $N_{2,T}(\cdot)$ are asymptotically independent and $\widehat{\mathbf{S}}_T$ depends on $N_{2,T}(\cdot)$ only. $\widehat{\mathbf{S}}_T$ is inconsistent and converges in distribution to a random variable $\mathbf{S}_\infty$. Thus, $LR_T(\widehat{\mathbf{S}}_T)$ has asymptotically the same distribution as

$$\begin{aligned} &CLR_\infty\left(M_{1,\infty}(\mathbf{S}_\infty),\, M_{2,\infty}(\mathbf{S}_\infty),\, \Sigma_v,\, \beta_0\right) \\ &\quad = \frac{1}{2}\left(M_{1,\infty}(\mathbf{S}_\infty) - M_{2,\infty}(\mathbf{S}_\infty) + \sqrt{\left(M_{1,\infty}(\mathbf{S}_\infty) - M_{2,\infty}(\mathbf{S}_\infty)\right)^2 + 4M_{1,2,\infty}^2(\mathbf{S}_\infty)}\right). \end{aligned}$$

Conditional on $N_{2,T}(\cdot)$ $\widehat{\mathbf{S}}_T$ is fixed. Define $\kappa_{CLR,\alpha}(M_{1,\infty}(\mathbf{S}_\infty),\, m_2,\, \Sigma_v,\, \beta_0)$ to be the $1-\alpha$ quantile of the null distribution of $CLR_\infty(M_{1,\infty}(\mathbf{S}_\infty),\, m_2,\, \Sigma_v,\, \beta_0)$. Since under H_0 $\kappa_{CLR,\alpha}(\widehat{\mathbf{S}}_T)$ has asymptotically the same distribution as $\kappa_{\mathrm{CLR},\alpha}(M_{1,\infty}(\mathbf{S}_\infty),\, M_{2,\infty}(\mathbf{S}_\infty),\, \Sigma_v,\, \beta_0)$, we have that the distribution of $LR_T(\widehat{\mathbf{S}}_T) - \kappa_{CLR,\alpha}(\widehat{\mathbf{S}}_T)$ is asymptotically the same distribution as

$$CLR_\infty\left(M_{1,\infty}(\mathbf{S}_\infty),\, M_{2,\infty}(\mathbf{S}_\infty),\, \Sigma_v,\, \beta_0\right) - \kappa_{CLR,\alpha}(M_{1,\infty}(\mathbf{S}_\infty),\, M_{2,\infty}(\mathbf{S}_\infty),\, \Sigma_v,\, \beta_0).$$

Conditional on $N_{2,\infty}(\cdot)$, $N_{1,\infty}(\widehat{\mathbf{S}}_T) \sim \mathscr{N}(0,\, I_q)$ under H_0. This implies that the conditional null distribution of CLR_∞ given $N_{2,\infty}(\cdot)$ does not depend on θ or c. Thus, the test that rejects H_0 when $LR_T(\widehat{\mathbf{S}}_T) - \kappa_{CLR,\alpha}(\widehat{\mathbf{S}}_T) > 0$ is similar at significance level α. □

S.D.2.3 Proof of Proposition S.B.1

The distribution of y is multivariate normal with

$$\mathbb{E}\left(y\right) = \overline{Z}(C_{0,T})\theta a' + X\eta, \tag{S.D.16}$$

independence across rows, and covariance matrix Σ_v for each row. Thus, the density of y is

$$\begin{aligned} &(2\pi)^{-T/2}\left|\Sigma_v\right|^{-T/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(y_t - a\theta'\overline{Z}_t(C_{0,T}) - \eta' X_t\right)'\Sigma_v^{-1}\left(y_t - a\theta'\overline{Z}_t(C_{0,T}) - \eta' X_t\right)\right) \\ &\qquad = (2\pi)^{-T/2}\left|\Sigma_v\right|^{-T/2}\exp\left(-\frac{1}{2}\left[\sum_{t=1}^{T} y_t'\Sigma_v^{-1}y_t - 2\theta'\left(\sum_{t=1}^{T}\overline{Z}_t(C_{0,T})y_t'\right)\Sigma_v^{-1}a\right.\right. \end{aligned} \tag{S.D.17}$$

$$- 2\mathrm{Tr}\left(\left(\sum_{t=1}^{T} X_t y_t'\right)\Sigma_v^{-1}\eta'\right) + \sum_{t=1}^{T}\left(a\theta'\overline{Z}_t(C_{0,T}) - \eta' X_t\right)'\Sigma_v^{-1}\left(a\theta'\overline{Z}_t(C_{0,T}) - \eta' X_t\right)\Bigg]\Bigg).$$

By the Fisher–Neyman factorization theorem $N(y)$ is a sufficient statistic for $\psi = (\beta,\, \theta',\, \gamma',\, \phi')'$ if and only if the density $\mathscr{L}(y; \beta,\, \theta,\, \gamma,\, \phi,\, \mathbf{S}_{0,T})$ can be factorized as $\mathscr{L}(y; \beta,\, \theta,\, \gamma,\, \phi,\, \mathbf{S}_{0,T}) = f_\psi(N(y))\, h(y)$ for nonnegative functions $f_\psi(\cdot)$ and $h(\cdot)$. Note that $\overline{Z}(C_{0,T}) = M_X C_{0,T} Z \neq M_X Z$ if $\pi_0 < 1$ and so $\mathscr{L}(y; \beta,\, \theta,\, \gamma,\, \phi,\, \mathbf{S}_{0,T})$ cannot be factorized as above for $N(y) = [y' M_X Z : y' X]$. Thus, $Z' M_X y$ and $X' y$ are not sufficient statistics for ψ if $\pi_0 < 1$. Therefore when $\pi_0 < 1$, $Z' M_X y$ cannot be sufficient for $(\beta,\, \theta')'$ if (i) $Z' M_X y$ and $X' y$ are independent, (ii) the distribution of $X' y$ does not depend on $(\beta,\, \theta')'$ and (iii) the distribution of $Z' M_X y$ does not depend on $(\gamma',\, \phi')'$.

To complete the proof, we verify that (i)–(iii) above hold. For (i), note that $Z' M_X y$ and $X' y$ are (jointly) multivariate normal random matrices and for any $b_1,\, b_2 \in \mathbb{R}^2$, we have

$$\mathrm{Cov}\left(Z' M_X y b_1,\, X' y b_2\right) = Z' M_X \mathrm{Cov}(y b_1, y b_2) X = Z' M_X b_1' \Sigma_v b_2 I_T X = b_1' \Sigma_v b_2 Z' M_X X = 0,$$

where the second equality uses the independence of the rows of y. Lemma 1(c) of Andrews, Moreira, and Stock (2006) implies (ii) and for (iii), note that the normality of $Z' M_X y$ has

$$\begin{aligned}
\mathbb{E}\left(Z' M_X y\right) &= Z' M_X \mathbb{E}\left(y\right) = Z' M_X \left(\overline{Z}(C_{0,T})\theta a' + X\eta\right) = Z'\overline{Z}(C_{0,T})\theta a', \\
\mathrm{Var}\left(Z' M_X y b\right) &= Z' M_X \mathrm{Var}\left(y b\right) M_X Z = Z' M_X b' \Sigma_v b I_T M_X Z,
\end{aligned}$$

for any $b \in \mathbb{R}^2$, which do not depend on $(\gamma',\, \phi')'$. □

S.D.2.4 Proof of Proposition S.B.2

The density of y is given by $\mathscr{L}(y; \beta,\, \theta,\, \gamma,\, \phi,\, \mathbf{S}_T)$, where $\mathscr{L}(\cdot)$ is defined in (S.D.17) and $\mathbf{S}_T$ is an unknown parameter. Using the same logic as in the proof of Proposition S.B.1, by the Fisher–Neyman factorization theorem $\{\overline{Z}\left(C_T\right)' y\}_{\mathbf{S}_T \in \mathcal{S}}$ is a sufficient statistic for $(\beta,\, \theta')'$ if (i) $\{\overline{Z}\left(C_T\right)' y\}_{\mathbf{S}_T \in \mathcal{S}}$ and $X' y$ are independent, (ii) the distribution of $X' y$ does not depend on $(\beta,\, \theta')'$ and (iii) the distribution of $\{\overline{Z}\left(C_T\right)' y\}_{\mathbf{S}_T \in \mathcal{S}}$ does not depend on $(\gamma',\, \phi')'$. For (i), note that $\overline{Z}\left(C_T\right)' y$ and $X' y$ are (jointly) multivariate normal random matrices and for any $b_1,\, b_2 \in \mathbb{R}^2$, we have

$$\mathrm{Cov}\left(\overline{Z}\left(C_T\right)' y b_1,\, X' y b_2\right) = Z' C_T' M_X \mathrm{Cov}(y b_1, y b_2) X = Z' C_T' M_X b_1' \Sigma_v b_2 I_T X = b_1' \Sigma_v b_2 Z' M_X X = 0,$$

where the second equality uses the independence of the rows of y. Lemma 1(c) of Andrews, Moreira, and Stock (2006) implies (ii) and for (iii), note that

$$\mathbb{E}\left(Z'C_T'M_Xy\right) = Z'C_T'M_X\mathbb{E}\left(y\right) = Z'C_T'M_X\left(\overline{Z}(C_{0,T})\theta a' + X\eta\right) = Z'C_T'\overline{Z}(C_{0,T})\theta a',$$
$$\mathrm{Var}\left(Z'C_T'M_Xyb\right) = Z'C_T'M_X\mathrm{Var}\left(yb\right)M_XC_TZ = Z'C_T'M_Xb'\Sigma_vbI_TM_XC_TZ,$$

for any $b \in \mathbb{R}^2$, which does not depend on $(\gamma', \phi')'$. □

S.D.2.5 Proof of Proposition S.B.3

Under the conditions of the proposition, the log-likelihood of y is given (up to a constant) by

$$\begin{aligned}
&\ell\left(\beta, \theta, \gamma, \phi, \mathbf{S}_T\right) \qquad \text{(S.D.18)}\\
&= -\frac{1}{2}\mathrm{Tr}\left(\Sigma_v^{-1}\left(\left(y - \overline{Z}\left(C_T\right)\theta a' - X\eta\right)'\left(y - \overline{Z}\left(C_T\right)\theta a' - X\eta\right)\right)\right)\\
&= \mathrm{Tr}\left(\Sigma_v^{-1}a\theta'\overline{Z}\left(C_T\right)'y\right) - \frac{1}{2}\mathrm{Tr}\left(\Sigma_v^{-1}a\theta'\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\theta a'\right) - \frac{1}{2}\mathrm{Tr}\left(\Sigma_v^{-1}(y - X\eta)'(y - X\eta)\right).
\end{aligned}$$

Maximizing this log-likelihood with respect to θ under H_0 yields

$$\widetilde{\theta}(C_T) = \left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)'y\Sigma_v^{-1}a_0(a_0'\Sigma_v^{-1}a_0)^{-1},$$

so that the concentrated likelihood function under H_0 is

$$\begin{aligned}
&\ell\left(y; \beta_0, \widetilde{\theta}(C_T), \gamma, \phi, \mathbf{S}_T\right) \qquad \text{(S.D.19)}\\
&= (a_0'\Sigma_v^{-1}a_0)^{-1}\mathrm{Tr}\left(\Sigma_v^{-1}a_0a_0'\Sigma_v^{-1}y'\overline{Z}\left(C_T\right)\left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)'y\right)\\
&\quad - \frac{1}{2}(a_0'\Sigma_v^{-1}a_0)^{-2}\mathrm{Tr}\left(\Sigma_v^{-1}a_0a_0'\Sigma_v^{-1}y'\overline{Z}\left(C_T\right)\left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)'y\Sigma_v^{-1}a_0a_0'\right)\\
&\quad - \frac{1}{2}\mathrm{Tr}\left(\Sigma_v^{-1}(y - X\eta)'(y - X\eta)\right)\\
&= \frac{1}{2}(a_0'\Sigma_v^{-1}a_0)^{-1}a_0'\Sigma_v^{-1}y'\overline{Z}\left(C_T\right)\left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)'y\Sigma_v^{-1}a_0\\
&\quad - \frac{1}{2}\mathrm{Tr}\left(\Sigma_v^{-1}(y - X\eta)'(y - X\eta)\right).
\end{aligned}$$

Maximizing (S.D.19) with respect to $\mathbf{S}_T$ is equivalent to maximizing

$$(a_0'\Sigma_v^{-1}a_0)^{-1}a_0'\Sigma_v^{-1}y'\overline{Z}\left(C_T\right)\left(\overline{Z}\left(C_T\right)'\overline{Z}\left(C_T\right)\right)^{-1}\overline{Z}\left(C_T\right)'y\Sigma_v^{-1}a_0, \qquad \text{(S.D.20)}$$

making the MLE of $\mathbf{S}_T$ equal to the maximizer of (S.D.20) over $\mathbf{S}_T \in \mathcal{S}$.

To complete the proof, we show that maximizing $M_{2,T}(\mathbf{S}_T) = N_{2,T}(\mathbf{S}_T)' N_{2,T}(\mathbf{S}_T)$ over $\mathbf{S}_T \in \mathcal{S}$ is asymptotically equivalent to maximizing (S.D.20) over $\mathbf{S}_T \in \mathcal{S}$ under the conditions of the proposition. To see this, first note that

$$\begin{aligned}\Sigma_{N_1 N_2}(\mathbf{S}) &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[v_t b_0 \overline{Z}_t(C_T)' \overline{Z}_t(C_T) a_0' \Sigma_v^{-1} v_t'\right] \\ &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \overline{Z}_t(C_T)' \overline{Z}_t(C_T) a_0' \Sigma_v^{-1} \mathbb{E}\left[v_t' v_t\right] b_0 \\ &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \overline{Z}_t(C_T)' \overline{Z}_t(C_T) a_0' b_0 = 0\end{aligned} \tag{S.D.21}$$

for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, where the first equality follows from the i.i.d. assumption, the second from the assumption of fixed regressors, the third from $\mathbb{E}[v_t' v_t] = \Sigma_v$ and the fourth from $a_0' b_0 = 0$. Thus, applying Lemma S.D.3, $N_{2,T}(\mathbf{S}_T)$ is asymptotically equivalent to $\Sigma_{N_2}^{-1/2}(\mathbf{S}_T) T^{-1/2} \overline{Z}(C_T)' y \Sigma_v^{-1} a_0$ under Assumption 6.4, implying that $M_{2,T}(\mathbf{S}_T)$ is asymptotically equivalent to $T^{-1} a_0' \Sigma_v^{-1} y' \overline{Z}(C_T) \Sigma_{N_2}^{-1}(\mathbf{S}_T) \overline{Z}(C_T)' y \Sigma_v^{-1} a_0$. The result then follows from

$$\begin{aligned}\Sigma_{N_2}(\mathbf{S}) &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[v_t \Sigma_v^{-1} a_0 \overline{Z}_t(C_T)' \overline{Z}_t(C_T) a_0' \Sigma_v^{-1} v_t'\right] \\ &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \overline{Z}_t(C_T)' \overline{Z}_t(C_T) a_0' \Sigma_v^{-1} \mathbb{E}\left[v_t' v_t\right] \Sigma_v^{-1} a_0 = \lim_{T\to\infty} T^{-1} \overline{Z}(C_T)' \overline{Z}(C_T) a_0' \Sigma_v^{-1} a_0\end{aligned}$$

for $\mathbf{S} = \lim_{T\to\infty} T^{-1}\mathbf{S}_T$, in analogy with (S.D.21). □

References

Andrews, D. W. K., M. Moreira, and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74(3), 715–752.

Casini, A., A. McCloskey, L. Rolla, and R. Pala (2025): "Supplement Not For Online Publication to "Dynamic Local Average Treatment Effects in Time Series"," *Unpublish manuscript*.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer-Verlag New York, 3 edn.

Lewis, D. J. (2022): "Robust Inference in Models Identified via Heteroskedasticity," *Review of Economics and Statistics*, 104(3), 510–524.

Magnusson, L. M., and S. Mavroeidis (2014): "Identification Using Stability Restrictions," *Econometrica*, 82(5), 1799–1851.

Nakamura, E., and J. Steinsson (2018): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133(3), 1283–1330.

Rigobon, R., and B. Sack (2003): "Measuring the Reaction of Monetary Policy to the Stock Market," *The Quarterly Journal of Economics*, 118(2), 639–669.

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

16th September 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 $i = 1, \ldots, \widetilde{m}$ 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 $\lambda_0 = (\lambda_1^0, \lambda_2^0, \ldots, \lambda_{\widetilde{m}}^0)$ 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, $\widehat{\lambda}_i \xrightarrow{\mathbb{P}} \lambda_i^0$, $i = 1, \ldots, \widetilde{m}$.*

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,*

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

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 (\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). □

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

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

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

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

We have

$$\begin{aligned} E_2 + E_3 + E_4 &= \left(\widehat{V}\left(\widehat{\mathbf{S}}\right) - V\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \left(\widehat{V}\left(\widehat{\mathbf{S}}\right) - V\right) \\ &= \left(\overline{X}\left(\widehat{\mathbf{S}}\right) \widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right) \alpha_0\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \left(\overline{X}\left(\widehat{\mathbf{S}}\right) \widehat{\alpha}\left(\widehat{\mathbf{S}}\right) - \overline{X}\left(\mathbf{S}_{0,T}\right) \alpha_0\right) \\ &= D\left(\widehat{\mathbf{S}}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} D\left(\widehat{\mathbf{S}}\right) \\ &\geq \sum_{t=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} \sum_{s=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} d_t' \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} d_s + \sum_{t=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} \sum_{s=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} d_t' \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} d_s \\ &= a_{\beta,i,k}' \sum_{t=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} \sum_{s=T\left(\lambda_i^0 - \eta\right)}^{T\lambda_i^0} \boldsymbol{x}_t \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} \boldsymbol{x}_s' a_{\beta,i,k} + b_{\beta,i,k}' \sum_{t=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} \sum_{s=T\lambda_i^0+1}^{T(\lambda_i^0+\eta)} \boldsymbol{x}_t \left[\widehat{\Omega}_{\mathbf{S}}^{-1}\right]_{(t,s)} \boldsymbol{x}_s' b_{\beta,i,k} \end{aligned} \tag{N.7}$$

$$\begin{aligned} &\geq \gamma_T \left[\left\|\theta_{\beta,i} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\gamma_\beta - \widehat{\gamma}_\beta\right\|^2 + \left\|\theta_i - \widehat{\theta}_k\right\|^2 + \left\|\gamma_2 - \widehat{\gamma}_2\right\|^2\right] \\ &\quad + \gamma_T^* \left[\left\|\theta_{\beta,i+1} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\gamma_\beta - \widehat{\gamma}_\beta\right\|^2 + \left\|\theta_{i+1} - \widehat{\theta}_k\right\|^2 + \left\|\gamma_2 - \widehat{\gamma}_2\right\|^2\right] \\ &\geq \min\left\{\gamma_T, \gamma_T^*\right\} \left(\left\|\theta_{\beta,i} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\theta_i - \widehat{\theta}_k\right\|^2 + \left\|\theta_{\beta,i+1} - \widehat{\theta}_{\beta,k}\right\|^2 + \left\|\theta_{i+1} - \widehat{\theta}_k\right\|^2\right) \\ &\geq 2^{-1} \min\left\{\gamma_T, \gamma_T^*\right\} \left(\left\|\theta_{\beta,i} - \theta_{\beta,i+1}\right\|^2 + \left\|\theta_i - \theta_{i+1}\right\|^2\right), \end{aligned}$$

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 $\mathbf{T}_\epsilon = \{(T_1, \, T_2, \, T_3) \, : \, |\widehat{T}_i - T_i^0| \leq \epsilon T \text{ for } i = 1, \ldots, 3\}$ so that $\mathbb{P}(\{\widehat{T}_1, \, \widehat{T}_2, \, \widehat{T}_3\} \in \mathbf{T}_\epsilon) \to 1$. Therefore we only need to examine the behavior of the objective function, $Q_T(T_1, \, T_2, \, T_3) = \widehat{V}\left(\mathbf{S}\right)' \widehat{\Omega}_{\mathbf{S}}^{-1} \widehat{V}\left(\mathbf{S}\right)$, 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

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

Note that $\mathbf{T}_\epsilon(C) \subset \mathbf{T}_\epsilon$. Because $Q_T(\widehat{T}_1,\, \widehat{T}_2,\, \widehat{T}_3) \le Q_T(\widehat{T}_1,\, T_2^0,\, \widehat{T}_3)$ 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,

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

or equivalently,

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

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

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

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 $\mathbf{T}_\epsilon(C)$,

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

and

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

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

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

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

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

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. □

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 $\sum_{i=1}^{\widetilde{m}} |\widehat{T}_i - T_i^0| = O_{\mathbb{P}}(\widetilde{m})$ 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

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

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 $\overline{D}_{P,s^*(t),n_1} - \overline{D}_{C,t,n_0} \xrightarrow{\mathbb{P}} \tilde{c}$ where $\tilde{c} > 0$.

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)* $\mathbb{E}(D_t(z) \mid \widetilde{V}_t)$ *is monotonic in* z *almost surely.*

Under Assumption N.C.3, assume without loss of generality that $\mathbb{E}(D_t(z) \mid \widetilde{V}_t)$ 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., $\mathbb{E}(D_t(z) \mid \widetilde{V}_t)$ being strictly decreasing in z almost surely, so that non-compliers are characterized by $\mathbb{E}(D_t(z) \mid \widetilde{V}_t)$ 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. □

N.C.3 Consistent Covariance Matrix Estimation

Let $V_{b,t}(\mathbf{S}_T) = v_t' b_0 \overline{Z}_t(C_T)$ and $V_{a,t}(\mathbf{S}_T) = v_t' \Sigma_v^{-1} a_0 \overline{Z}_t(C_T)$. We have

$$\Sigma_{N_1}(\mathbf{S}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{r=1}^{T} \mathbb{E}\left(V_{b,t}(\mathbf{S}_T) V_{b,r}(\mathbf{S}_T)'\right),$$
$$\Sigma_{N_1N_2}(\mathbf{S}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{r=1}^{T} \mathbb{E}\left(V_{b,t}(\mathbf{S}_T) V_{a,r}(\mathbf{S}_T)'\right),$$
$$\Sigma_{N_2}^{*}(\mathbf{S}) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T}\sum_{r=1}^{T} \mathbb{E}\left(V_{a,t}(\mathbf{S}_T) V_{a,r}(\mathbf{S}_T)'\right).$$

Let $\widehat{V}_{b,t}(\mathbf{S}_T) = \widehat{v}_t(\mathbf{S}_T)' b_0 \overline{Z}_t(C_T)$ and $\widehat{V}_{a,t}(s) = \widehat{v}_t(\mathbf{S}_T)' \widehat{\Sigma}_v^{-1} a_0 \overline{Z}_t(C_T)$. 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 $\widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T)$ 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 $\widehat{\Sigma}_{N_1N_2}(\mathbf{S}_T)$ 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, $K_2^{\mathrm{opt}}(x) = 6x(1-x),\ 0 \le x \le 1$.

For both HAC and DK-HAC estimators, define

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

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*

$$\sum_{j=-\infty}^{\infty} \sup_{t\ge1} \sup_{\mathbf{S}_T\in\mathcal{S}} ||\mathbb{E}(V_t(\mathbf{S}_T)\,V'_{t-j}(\mathbf{S}_T))|| < \infty,$$

and for all conformable $a_1,\ a_2,\ a_3,\ a_4 \in \mathbb{Z}_+$, $\sum_{n=1}^{\infty}\sum_{j=1}^{\infty}\sum_{m=1}^{\infty}\sup_{t\ge1}|\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(n,\ j,\ m,\mathbf{S}_T)| < \infty$ 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) $\sup_{t\ge1}\sup_{\mathbf{S}_T\in\mathcal{S}}\mathbb{E}||V_t(\mathbf{S}_T)||^2 < \infty$.

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 $\zeta_{N_1}(\mathbf{S}) \sim \mathcal{N}(\alpha_{N_1}(\mathbf{S}), I_q)$, $\alpha_{N_1}(\mathbf{S}) = \Sigma_{N_1}^{-1/2}(\mathbf{S})\,\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)\,\theta r$, and $\alpha_{N_2}(\mathbf{S}) = \Sigma_{N_2}^{-1/2}(\mathbf{S})\,\Sigma_{\overline{Z}}(\mathbf{S},\mathbf{S}_0)\,\theta\,(a_0'\Omega^{-1}a_0)^{-1/2}$.

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)* $AR_T(\widehat{\mathbf{S}}_T) \xrightarrow{d} \zeta_{N_1}(\mathbf{S}_0)'\,\zeta_{N_1}(\mathbf{S}_0) \sim \chi^2_q(\alpha_{N_1}(\mathbf{S}_0)'\,\alpha_{N_1}(\mathbf{S}_0))$*; (ii)* $LM_T(\widehat{\mathbf{S}}_T) \xrightarrow{d} (\alpha_{N_2}(\mathbf{S}_0)'\,\zeta_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2 \sim \chi^2_1((\alpha_{N_2}(\mathbf{S}_0)'\,\alpha_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2)$*; (iii)* $LR_T(\widehat{\mathbf{S}}_T) = LM_T(\mathbf{S}_0) + o_{\mathbb{P}}(1) \xrightarrow{d} (\alpha_{N_2}(\mathbf{S}_0)'\,\zeta_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2 \sim \chi^2_1((\alpha_{N_2}(\mathbf{S}_0)'\,\alpha_{N_1}(\mathbf{S}_0))^2/||\alpha_{N_2}(\mathbf{S}_0)||^2)$.

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

$$\varphi_{N_2}(\mathbf{S}_0) = \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\left(\Sigma_{\overline{Z}}(\mathbf{S}_0)\,\theta a' \Sigma_v^{-1} a_0 - \Sigma_{N_1 N_2}(\mathbf{S}_0,\, \mathbf{S}_0)\,\Sigma_{N_1}^{-1}\varphi_{N_1}(\mathbf{S}_0)\right), \quad \varsigma_q \sim \mathscr{N}(0,\, I_q). \tag{N.4}$$

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

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

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

$$\begin{aligned}
T^{-1/2}N_{2,T}\left(\widehat{\mathbf{S}}_T\right) &= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\widehat{\Sigma}_v^{-1}\left(\widehat{\mathbf{S}}_T\right)a_0 - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) \qquad \text{(N.5)} \\
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\Sigma_v^{-1}a_0 - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) + o_{\mathbb{P}}(1) \\
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_0 - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\mathbf{S}_0\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\mathbf{S}_0\right)N_{1,T}\left(\mathbf{S}_0\right)\right) + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_0 - T^{-1/2}\Sigma_{N_1N_2}\left(\mathbf{S}_0\right)\Sigma_{N_1}^{-1/2}\left(\mathbf{S}_0\right)N_{1,T}\left(\mathbf{S}_0\right)\right) + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a' + v\right)\Sigma_v^{-1}a_0\right) + o_{\mathbb{P}}(1) \\
&= \Sigma_{N_2}^{-1/2}\left(\mathbf{S}_0\right)\Sigma_{\overline{Z}}\left(\mathbf{S}_0\right)\theta a_0'\Sigma_v^{-1}a_0 + o_{\mathbb{P}}(1),
\end{aligned}$$

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:

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

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

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

Similarly,

$$
\begin{aligned}
T^{-1/2}N_{2,T}\left(\widehat{\mathbf{S}}_T\right) &= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\widehat{\Sigma}_v^{-1}\left(\widehat{\mathbf{S}}_T\right)a_0 - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) \qquad \text{(N.8)}\\
&= \widehat{\Sigma}_{N_2}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)\left(T^{-1}\overline{Z}\left(\widehat{C}_T\right)' y\Sigma_v^{-1}a_0 - T^{-1/2}\widehat{\Sigma}_{N_1N_2}\left(\widehat{\mathbf{S}}_T\right)\widehat{\Sigma}_{N_1}^{-1/2}\left(\widehat{\mathbf{S}}_T\right)N_{1,T}\left(\widehat{\mathbf{S}}_T\right)\right) + o_{\mathbb{P}}(1)\\
&= \widehat{\Sigma}_{N_2}^{-1/2}(\mathbf{S}_0)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_0 - T^{-1/2}\widehat{\Sigma}_{N_1N_2}(\mathbf{S}_0)\widehat{\Sigma}_{N_1}^{-1/2}(\mathbf{S}_0)N_{1,T}(\mathbf{S}_0)\right) + o_{\mathbb{P}}(1)\\
&= \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)' y\Sigma_v^{-1}a_0 - T^{-1/2}\Sigma_{N_1N_2}(\mathbf{S}_0)\Sigma_{N_1}^{-1/2}(\mathbf{S}_0)N_{1,T}(\mathbf{S}_0)\right) + o_{\mathbb{P}}(1)\\
&= \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\left(T^{-1}\overline{Z}\left(C_{0,T}\right)'\left(\overline{Z}\left(C_{0,T}\right)\theta a' + v\right)\Sigma_v^{-1}a_0\right)\\
&\quad - \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\Sigma_{N_1N_2}(\mathbf{S}_0)\Sigma_{N_1}^{-1/2}(\mathbf{S}_0)\varphi_{N_1}(\mathbf{S}_0) + o_{\mathbb{P}}(1)\\
&= \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\Sigma_{\overline{Z}}(\mathbf{S}_0)\theta a'\Sigma_v^{-1}a_0 - \Sigma_{N_2}^{-1/2}(\mathbf{S}_0)\Sigma_{N_1N_2}(\mathbf{S}_0)\Sigma_{N_1}^{-1/2}(\mathbf{S}_0)\varphi_{N_1}(\mathbf{S}_0) + o_{\mathbb{P}}(1)\\
&= \varphi_{N_2}(\mathbf{S}_0) + o_{\mathbb{P}}(1).
\end{aligned}
$$

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
split − S	0.054	0.055	0.055	0.071	0.074	0.076	0.139	0.148	0.155
split − CLR	0.146	0.149	0.150	0.174	0.186	0.187	0.278	0.291	0.300
qqL − S	0.028	0.028	0.028	0.047	0.052	0.052	0.191	0.198	0.199
ave − S	0.047	0.046	0.047	0.068	0.071	0.070	0.153	0.161	0.171
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
split − S	0.046	0.044	0.043	0.060	0.059	0.059	0.113	0.113	0.113
split − CLR	0.133	0.134	0.134	0.150	0.154	0.154	0.234	0.240	0.243
qqL − S	0.042	0.043	0.039	0.042	0.063	0.063	0.169	0.177	0.180
ave − S	0.048	0.050	0.048	0.048	0.073	0.068	0.130	0.133	0.137
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
split − S	0.015	0.015	0.015	0.015	0.015	0.016	0.016	0.015	0.016
split − CLR	0.096	0.090	0.086	0.094	0.092	0.082	0.091	0.094	0.087
qqL − S	0.008	0.009	0.008	0.008	0.009	0.010	0.010	0.009	0.009
ave − S	0.020	0.025	0.023	0.020	0.024	0.023	0.021	0.027	0.023
exp − 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
split − S	0.020	0.017	0.021	0.022	0.019	0.021	0.018	0.018	0.020
split − CLR	0.102	0.098	0.086	0.103	0.100	0.098	0.103	0.108	0.098
qqL − S	0.009	0.074	0.009	0.010	0.008	0.011	0.010	0.008	0.009
ave − S	0.020	0.022	0.018	0.023	0.022	0.019	0.019	0.025	0.018
exp − S	0.004	0.026	0.004	0.005	0.003	0.005	0.005	0.005	0.004

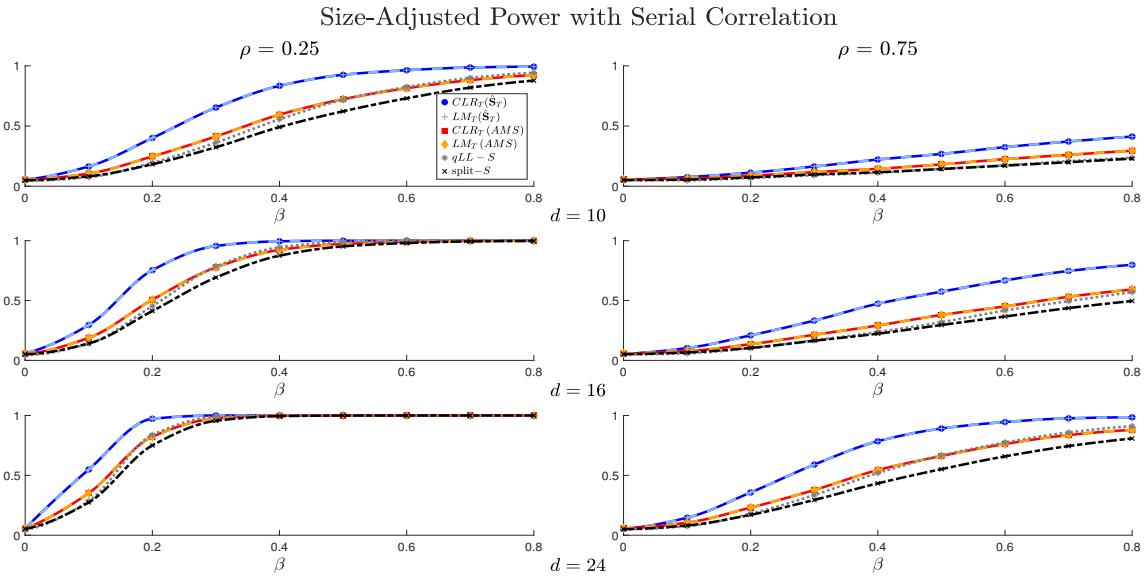

Figure 1: Size-adjusted power of identification robust tests for $T = 400$ and $\pi_0 = 0.6$ and $\rho = 0.25$.

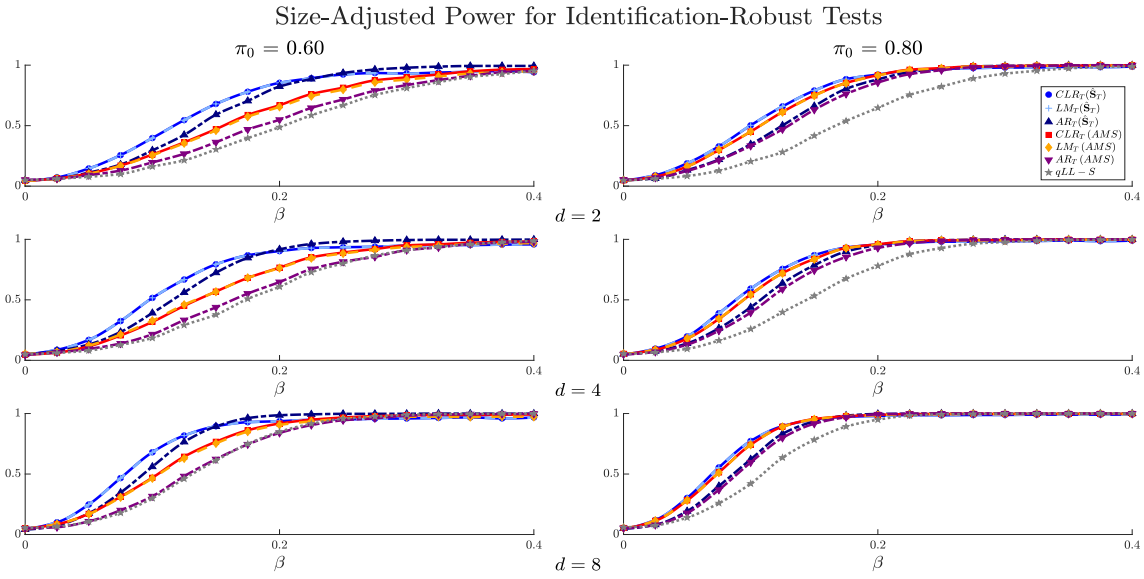

Figure 2: Size-adjusted power of identification robust tests for $T = 200$ for model in (N.1)-in(N.2).

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.