The Divide-and-Conquer Estimator for High-Dimensional Linear Regression under Weak Exogeneity

Yassin Alaya *

November 1, 2026

Abstract

Mikusheva and Sølvsten (2025) show that ordinary least squares is generally inconsistent for weakly exogenous time-series regressions when the number of regressors is proportional to the sample size and propose a new estimator for high-dimensional time-series regressions. Consistency of this estimator, however, relies on homoskedasticity, and inference results are limited. We propose the Divide-and-Conquer (DaC) estimator as an alternative. It is consistent under heteroskedasticity when feedback from the error term to future regressors has a known, finite horizon. The idea is simple: partition the sample into subsamples, spaced so that strict exogeneity holds within each, estimate OLS on every subsample, and average. The resulting estimator is unbiased and consistent for linear contrasts of the slope parameter when the number of regressors is proportional to the sample size. Asymptotic normality in the proportional regime is established and a conservative inference procedure is proposed. Simulations compare the DaC estimator, OLS, and the Mikusheva–Sølvsten estimator.

*PhD candidate at Toulouse School of Economics. I am grateful to Eric Gautier, Koen Jochmans, Pascal Lavergne and Nour Meddahi for their advice. Artificial Intelligence was involved in proving the results of this paper as well as in the programming of the Monte Carlo simulations and the creation of tables.

1 Introduction

In linear time-series regressions, identification often relies on the assumption of weak exogeneity. Weak exogeneity requires the error term to be mean independent of the past and current regressors only and thus allows for the conditional mean of the error term to depend on future values of the regressors. Such dependence between the error term and the regressors of future periods might arise as a consequence of *feedback* from the error term to the regressors, i.e., the error term might affect the regressors in the following periods. The ordinary least squares (OLS) estimator has been known to have a finite-sample bias under weak exogeneity at least since Stambaugh (1999). However, in the standard asymptotic framework, this bias vanishes—OLS is consistent. Thus OLS is commonly used to estimate time-series regressions even where weak exogeneity is suspected, as the bias is thought to be negligible when the sample size is large enough. Mikusheva and Sølvsten (2025) (henceforth M&S) show that, under weak exogeneity, OLS is inconsistent in an asymptotic framework where the number of regressors is proportional to the sample size (the *proportional regime*) if regressors are in addition autocorrelated. Thus, the bias of OLS may be non-negligible if the number of regressors relative to the number of observations (the *aspect ratio*) is non-negligible.

In this paper, we propose an estimator that is consistent in the proportional regime under heteroskedasticity, given an exogeneity assumption that is slightly stronger than classical weak exogeneity: we allow for feedback from the error term to future regressors with a *finite horizon* of h periods only. That is, we assume the error term at time t to be mean independent of regressors from the past, the present and the future—except those from periods $t+1$ through $t+h$. If the magnitude of feedback declines fast enough, this might be a reasonable approximation to the true data generating process (DGP). The idea of the estimator is simple: Under the assumption of a finite feedback horizon, if h is taken as known, the sample can be partitioned into $h+1$ subsamples within each of which strict exogeneity holds. This is done by skipping h periods between observations. If each subsample still contains more observations than regressors, we can obtain an OLS estimate for each subsample and average these across subsamples. We call the resulting average the Divide-and-Conquer (DaC) estimator. Due to strict exogeneity within subsamples, the DaC estimator is unbiased.

The theoretical framework under which asymptotic results are derived in this paper largely mirrors that of M&S. In particular, we assume that regressors can be decomposed into a strictly exogenous part ($\tilde{\mathbf{X}}$) on the one hand and feedback from the error terms of the past h periods on the other hand. As in M&S, the strictly exogenous component is treated as non-stochastic, which is necessary for analytical tractability when the number of regressors is proportional to the sample size. We show consistency and asymptotic normality for individual linear contrasts of the slope parameter in the proportional regime under appropriate assumptions. These assumptions are imposed on the deterministic sequence of the strictly exogenous parts of the regressors. Inference on the DaC estimator involves an important complication: While the variances of the OLS estimators of each subsample can be estimated consistently using adequate estimators robust to high dimensionality and heteroskedasticity, their covariances cannot be estimated without knowledge of the feedback parameters or good enough estimators for them. It is, however, possible to estimate an upper bound on the variance, which can be employed for conservative inference. In simulations, the widths of the resulting confidence intervals do not exceed the widths of the exact confidence intervals by much.[1] We also provide simulation results that compare OLS, the DaC estimator and the estimator of M&S (henceforth the M&S estimator). While the M&S estimator generally performs well even under heteroskedasticity, it does exhibit a non-negligible bias when heteroskedasticity is linked to the magnitude of the regressors and regressors are heavy-tailed.

This paper is most closely related to M&S. Under a particular assumption on the form of weak exogeneity, they show that bias in the OLS estimate for an individual parameter arises due to interactions between the regression errors that are weighted by off-diagonal elements of the projection matrix associated with the remaining regressors. When regressors are autocorrelated, the resulting bias term is of the order K/T, where K is the number of regressors and T is the sample size. The reader is referred to M&S for a more thorough explanation. M&S go on to propose an alternative estimator. Unlike that of the DaC estimator, consistency of this estimator requires homoskedasticity. Another limitation is the absence of inference results when $h > 1$. In contrast, the conservative inference results

[1] However, the upper bound we present in this paper may not be the tightest estimable upper bound.

presented here for the DaC estimator apply to any h for which each subsample contains more observations than regressors.

The second main theme of this paper is high-dimensional regression. Results on the asymptotics of OLS in the proportional regime exist only under highly restrictive assumptions on the distribution of the regressors. Lei, Bickel, and El Karoui (2018) provide conditions on the asymptotic behaviour of a (non-random) design matrix for asymptotic normality of the estimator of an individual coefficient or of a linear combination of a fixed number of coefficients in the proportional regime. They then provide DGPs for the regressors under which these conditions are met with high probability. These DGPs are very restrictive and, in particular, do not allow for time-series dependence. Relevant results from the statistical literature that could be applied to time-series regressions are scarce. In the current paper, we therefore follow M&S in treating the strictly exogenous component of the regressors as non-random. We impose the necessary assumptions on the eigenvalues of the Gram matrix as well as on the leverage of individual observations with respect to the relevant directions.

While the results of this paper do not touch on efficiency, it should be noted that, when K is proportional to T, the DaC estimator incurs some loss in efficiency relative to OLS, at least in a cross-sectional setting: Dobriban and Sheng (2021) show that, in the cross-sectional framework with stochastic but i.i.d. regressors, i.i.d. error terms (and strict exogeneity), the DaC estimator is less efficient than the standard OLS estimator when the dimension is proportional to the sample size. (In the standard asymptotic framework with a fixed number of regressors, there is no asymptotic loss of efficiency.) In the time-series setting, the loss in efficiency might be less pronounced if regressors are positively autocorrelated.

Lastly, as estimation of the variances of the individual subsample estimators is a part of the proposed inference procedure, the current paper touches on the problem of heteroskedasticity-robust variance estimation in high dimensions. Cattaneo, Jansson, and Newey (2018) show that the usual Eicker–Huber–White standard errors are not consistent under the proportional regime, nor are their modified versions HC1–HC4. They also propose a consistent alternative, which, however, requires the aspect ratio to be smaller than 1/2. An alternative estimator is the one proposed by Jochmans (2022), but it comes at the cost of additional assumptions. We provide assumptions under which each of these two

estimators can be used to yield consistent standard errors for subsample OLS estimators under the weak exogeneity model of the present paper.

General notation. We use C and c as generic positive constants throughout, which may take different values at each occurrence. All limits are taken as $T \to \infty$. For a sequence a_T and a positive sequence b_T, we write $a_T = O(b_T)$ if $|a_T|/b_T$ is bounded for all large T, $a_T = o(b_T)$ if $a_T/b_T \to 0$, $a_T = \Omega(b_T)$ if $\liminf_{T\to\infty} a_T/b_T > 0$, and $a_T \asymp b_T$ if both $a_T = O(b_T)$ and $a_T = \Omega(b_T)$. (The relation $a_T \asymp b_T$ is sometimes written $a_T = \Theta(b_T)$.) For a sequence of random variables X_T, we write $X_T = O_p(b_T)$ if X_T/b_T is bounded in probability, i.e., for every $\epsilon > 0$ there is a $C < \infty$ such that $\mathbb{P}[|X_T|/b_T > C] < \epsilon$ for all large T, and $X_T = o_p(b_T)$ if X_T/b_T converges to zero in probability. For positive $a \in \mathbb{R}$, we use the notation $\lfloor a \rfloor$ to denote the largest integer less than or equal to a. For integers $a, b \in \mathbb{N}$, $a \bmod b$ denotes the remainder of the integer division of a by b, i.e., $a \bmod b = a - \lfloor a/b \rfloor b$. Let $u(t) = (t-1) \bmod (h+1) + 1$ be the subsample to which period t belongs. Let $r(v,t) = v + (t-1)(h+1)$. For $v \in \{1, \ldots, h+1\}$, this is the period the t–th observation from subsample v corresponds to but it is also sometimes used with $v \in \{1-h, \ldots, 0\}$.

Matrix notation. In the following, for a square matrix $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote its largest and smallest eigenvalues, respectively, and $\operatorname{tr}(\mathbf{A})$ denotes its trace (the sum of its diagonal elements). $\|\cdot\|$ denotes the Euclidean norm when applied to vectors. Let $\|\cdot\|_2$ denote the spectral norm, i.e., for a $p \times q$ matrix $\mathbf{B}$, $\|\mathbf{B}\|_2 := \sup_{x \in \mathbb{R}^q,\, \|x\| \le 1} \|\mathbf{B}x\| = \sqrt{\lambda_{\max}(\mathbf{B}'\mathbf{B})}$. As a consequence, when $\mathbf{B}$ is symmetric and positive semidefinite, $\|\mathbf{B}\|_2 = \lambda_{\max}(\mathbf{B})$. (More generally, when $\mathbf{B}$ is symmetric, $\|\mathbf{B}\|_2 = \max\{\lambda_{\max}(\mathbf{B}), -\lambda_{\min}(\mathbf{B})\}$.) Let $\|\cdot\|_F$ denote the Frobenius norm, i.e., for a $p \times q$ matrix $\mathbf{B}$, $\|\mathbf{B}\|_F := \sqrt{\sum_{i,j} \mathbf{B}_{ij}^2} = \sqrt{\operatorname{tr}(\mathbf{B}'\mathbf{B})}$. For symmetric square matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimension, the notation $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. Many useful inequalities on eigenvalues and the matrix norms used in this paper can be found in Appendix B.

2 The Divide-and-Conquer estimator

We observe a sample of size T from the model

$$y_t = x_t'\beta + \varepsilon_t,$$

where the dimension of x_t is denoted by K. We are interested in conducting inference on the slope parameter $\beta \in \mathbb{R}^K$, or more precisely on some linear contrast of the slope parameter, $r'\beta$ (for some $r \in \mathbb{R}^K$). Strict exogeneity may not hold, but suppose that the feedback has a known, finite horizon of h periods, as formalised by the following assumption:

Assumption *Finite Feedback Horizon.* *There is an $h \in \mathbb{N}$ such that*

$$\mathbb{E}\left[\varepsilon_t \mid \dots, x_{t-1}, x_t, x_{t+h+1}, x_{t+h+2}, \dots\right] = 0$$

for all $t \in \mathbb{Z}$.

Then, strict exogeneity holds within all subsamples where all observations are at least $h+1$ periods apart from each other. We can partition the sample into $h+1$ subsamples of equally spaced observations. Thus we obtain the subsamples $\{1, h+2, 2h+3, \dots\}, \{2, h+3, 2h+4, \dots\}, \dots, \{h+1, 2h+2, \dots\}$. For ease of notation, we suppose that T is a multiple of $h+1$ so that each subsample has the same size. Let $\bar{T} = T/(h+1)$ denote this subsample size. We denote by $\mathbf{X}_{(v)}$ the submatrix of observations of the regressors from subsample $v \in \{1, \dots, h+1\}$:

$$\mathbf{X}_{(v)} = \begin{pmatrix} x_v' \\ x_{v+h+1}' \\ x_{v+2h+2}' \\ \vdots \end{pmatrix}.$$

Likewise, the subvector of outcome values for periods belonging to subsample $v \in \{1, \dots, h+1\}$ is $y_{(v)} = (y_v, y_{v+h+1}, y_{v+2h+2}, \dots)'$.

Suppose that the subsample size exceeds the dimension, i.e., $\bar{T} = T/(h+1) > K$, and that

each $\mathbf{X}_{(v)}$ has full column rank. Then we can obtain an OLS estimator for each subsample:

$$\hat{\beta}_{(v)} = \left(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}\right)^{-1}\mathbf{X}'_{(v)}y_{(v)}.$$

Each of these is an unbiased estimator of β, so their average

$$\hat{\beta} = \frac{1}{h+1}\sum_{v=1}^{h+1}\hat{\beta}_{(v)}$$

is unbiased as well. We call $\hat{\beta}$ the DaC estimator.

3 Asymptotic theory

To obtain asymptotic results about the DaC estimator that can be relied upon in high-dimensional settings, we adopt an asymptotic framework in which the number of regressors grows with the sample size. More specifically, we consider a sequence of data generating processes—one for each sample size T—with a generally different set of regressors for each sample size and potentially even a different outcome variable. Accordingly, the identity of the parameter of interest is not necessarily stable. In full generality, we consider one linear contrast $r'\beta$ for some sequence $r \in \mathbb{R}^K$. In the simplest case, $r = (1, 0, \ldots, 0)' \in \mathbb{R}^K$, and the parameter of interest is the first coefficient of β, but the assumptions and results below are general enough to allow for any (sequence of) linear combination(s) r with bounded Euclidean norm. Although K, β, r and even the individual observations x_t now depend on T, we suppress the subscript T to keep notation simple.

To deal with a framework that combines increasing dimension and regressors that are not fully exogenous, we follow the framework of M&S. In particular, the regressors can be decomposed into a strictly exogenous part and feedback from past periods. As M&S do, we treat the strictly exogenous part of the regressors as non-random.

Assumption *Feedback Decomposition.* *The observed regressors x_t admit the decomposition*

$$x_t = \tilde{x}_t + \sum_{l=1}^{h}\alpha_l\varepsilon_{t-l},$$

where the K-vectors $\tilde{x}_1, \ldots, \tilde{x}_T$ are non-random, the K-vectors $\alpha_1, \ldots, \alpha_h$ are non-random and $\mathbb{E}[\varepsilon_t] = 0$. The horizon h is fixed (it does not depend on T).

The vectors $\alpha_1, \ldots, \alpha_h$ provide the feedback strength for each regressor at each feedback horizon. They too depend on the sample size. We write $\tilde{\mathbf{X}}_{(v)} = (\tilde{x}_v, \tilde{x}_{v+h+1}, \ldots)'$ for the matrix assembling the regressors of each subsample v. The next assumption provides bounds on the sequence of data generating processes. The eigenvalue bounds on $\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}$ (the subsample Gram matrices) ensure that variances of the regressors do not grow without bound and do not vanish either, thus imposing basic bounds on the sequence of data generating processes. The assumption also introduces a bound on the linear combination of interest, r, which is of course needed for consistency and any subsequent asymptotic result since the estimation error of $r'\hat{\beta}$ scales with r.

Assumption *Bounds*. *For all (large) T,*

$$\infty > C > \lambda_{\max}\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}\right) \geq \lambda_{\min}\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}\right) > c > 0$$

for all subsamples. In addition, $\|r\|, \|\alpha_1\|, \ldots, \|\alpha_h\| < C$ for all T, and

$$r'(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}r > c > 0$$

for all (large) T and at least one subsample $v \in \{1, \ldots, h+1\}$.

To assess the plausibility of the eigenvalue bounds when $\tilde{\mathbf{X}}_{(v)}$ is random, imagine for a moment that the individual entries of each $\tilde{\mathbf{X}}_{(v)}$ are i.i.d. standardised random variables with bounded fourth moments and that $K/\bar{T} \to c \in (0, 1)$. In this case, the empirical spectral distribution of $\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}$ converges to a Marčenko–Pastur law (Marčenko and Pastur 1967), whose support is $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$. Furthermore, with probability one, no eigenvalue of $\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}$ lies outside this limiting support for all large $\bar{T}$ (Bai and Silverstein 1998). Only limited generalisations of results like these to data that are not i.i.d. across variables and observations exist. Therefore, we sidestep the random matrix aspect by treating the exogenous part of the regressors as fixed.

Regarding the error term, we allow for heteroskedasticity within bounds and require uniformly bounded fourth moments. As the $\tilde{x}_t$ are treated as non-random, the relevant notion of heteroskedasticity is unconditional heteroskedasticity. We assume error distributions to be independent and to differ only by their scale. Let $\sigma_t^2 := \mathbb{E}[\varepsilon_t^2]$. Given the decomposition model of weak exogeneity embodied by Assumption *Feedback Decomposition*, serial correlation of the errors would not be compatible with the original definition of weak exogeneity as even the x_t and possibly $x_{t-1}, ...$ would then generally be correlated with each ε_t.[2]

Assumption *Regular Errors*. *$0 < c < \sigma_t^2 < C < \infty$ for all $t \geq 1 - h$. The normalised errors $\{\varepsilon_t/\sigma_t\}_t$ are i.i.d. and $\mathbb{E}[\varepsilon_t^4] \leq C^2$ for all $t \geq 1 - h$.*

The following assumption allows the aspect ratio to be proportional to sample size subject to the constraint that there be more than K observations in each subsample.

Assumption *Proportional Regime*. *$\limsup K/T < 1/(h+1)$ and $K \geq h + 1$.*

Asymptotic results are obtained as follows. Using the decomposition of $\mathbf{X}$ stipulated by Assumption *Feedback Decomposition* into a non-random, strictly exogenous part $\tilde{\mathbf{X}}$ and feedback, the estimation error of each subsample OLS estimator can be written asymptotically as the sum of a linear and a quadratic form in the $(T+h)$-vector of normalised error terms $\tilde{\varepsilon} := (\varepsilon_{1-h}/\sigma_{1-h}, \ldots, \varepsilon_T/\sigma_T)'$:

$$r'(\hat{\beta}_{(v)} - \beta) = \omega_{(v)}'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}_{(v)}\tilde{\varepsilon} + O_p(T^{-1}), \tag{1}$$

for some non-random conformable vectors and matrices of weights $\omega_{(v)}$ and $\mathbf{B}_{(v)}$ defined in Appendix A. Define $U_v = \omega_{(v)}'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}_{(v)}\tilde{\varepsilon}$ and $\Sigma_v := \mathrm{Var}(U_v)$. Thus the estimation error of the DaC estimator can be written as

$$(h+1)r'(\hat{\beta} - \beta) = \omega'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}\tilde{\varepsilon} + O_p(T^{-1}), \tag{2}$$

[2]It should be noted, though, that weak exogeneity is not generally incompatible with serially correlated error terms. An alternative model would decompose ε_t into two independent parts, one of which only affects future x_s, and one of which only affects future ε_s.

where $\omega = \sum_{v=1}^{h+1} \omega_{(v)}$ and $\mathbf{B} = \sum_{v=1}^{h+1} \mathbf{B}_{(v)}$. Define

$$U = \omega'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}\tilde{\varepsilon} = \sum_{v=1}^{h+1} U_v.$$

A key property of the DaC estimator is that the diagonal elements of $\mathbf{B}$ (and of all of the $\mathbf{B}_{(v)}$) are each individually 0; this property makes the DaC estimator unbiased.[3] Another consequence is that the linear part and the quadratic part of (2) are uncorrelated. Thus, the variance $\Sigma := \mathrm{Var}\,(U)$ is given by

$$\begin{aligned}\Sigma =& \|\omega\|^2 + \sum_{t>s} (\mathbf{B}_{ts} + \mathbf{B}_{st})^2 \\ =& \|\omega\|^2 + \frac{1}{2}\left\|\mathbf{B} + \mathbf{B}'\right\|_F^2 \\ =& \|\omega\|^2 + \mathrm{tr}\,(\mathbf{B}\mathbf{B}') + \mathrm{tr}\left(\mathbf{B}^2\right).\end{aligned}$$

Lemma 7 in Appendix A then shows that $\Sigma = O(T^{-1})$, which immediately entails the following result.

Theorem 1 (Consistency)**.** *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*, and* Proportional Regime*,*

$$r'(\hat{\beta} - \beta) \xrightarrow{p} 0.$$

We turn to the question of asymptotic normality. Corollary A2.8 from Sølvsten (2020) provides a central limit theorem applicable to the self-normalised sum $U/\sqrt{\Sigma}$ under conditions on ω and $\mathbf{B}$. It turns out that the necessary conditions on $\mathbf{B}/\sqrt{\Sigma}$ are implied by the assumptions we have already made. Another assumption is, however, required to ensure that no individual weight among the elements of ω dominates the overall variance of the linear form, i.e., $\max_t |\omega_t| = o(\|\omega\|)$. Towards formulating the necessary assumptions, we define a vector of *direction-specific leverages*. Let $\mathbf{A} := (r, \alpha_1, \ldots, \alpha_h)' \in \mathbb{R}^{(h+1)\times K}$. Then the direction-specific leverages are the elements of the vector $f_t := \mathbf{A}(\tilde{\mathbf{X}}'_{(u(t))}\tilde{\mathbf{X}}_{(u(t))}/\bar{T})^{-1}\tilde{x}_t \in \mathbb{R}^{h+1}$ for

[3]This contrasts with the M&S estimator, whose quadratic form has diagonal elements that are asymptotically 0 only on average; this makes the estimator asymptotically unbiased under homoskedasticity.

each period $t \in \{1, \ldots, T\}$. Within our framework, the condition that no individual weight dominate the linear form amounts to the following assumption on f_t:[4]

Assumption *No Influential Observations*. A *has full row rank uniformly in* T*, i.e.,* $\lambda_{\min}(\mathbf{A}\mathbf{A}') \geq c > 0$*, and*

$$\max_{1 \leq t \leq T} \|f_t\| = o(T^{1/2}).$$

When $h = 0$, so that $\tilde{\mathbf{X}}_{(v)} = \mathbf{X}_{(v)}$, Assumption *No Influential Observations* reduces to $\max_t |r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}/T)^{-1}x_t| = o(T^{1/2})$. The relevance of this quantity can be seen by noting that the estimation error takes the form of a weighted average of regression errors with deterministic weights $r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)})^{-1}x_t$,

$$r'(\hat{\beta} - \beta) = \frac{1}{T}\sum_t r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}/T)^{-1}x_t\varepsilon_t,$$

so that $\omega_t = T^{-1}r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}/T)^{-1}x_t\,\sigma_t$ and $\mathbf{B} = 0$ in this case. The condition

$$\max_t |r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}/T)^{-1}x_t| = o(T^{1/2})$$

corresponds to the necessary and sufficient condition from Lemma 2.1 in Huber (1973) and ensures that the maximum leverage with respect to the relevant direction r among all observations vanishes. When $h \geq 1$, the same must hold jointly for the r-direction and all h feedback directions $\alpha_1, \ldots, \alpha_h$. While theoretical results relevant to showing convergence in probability of $T^{-1/2}\max_t \|f_t\|$ to zero in the proportional regime under general assumptions (in particular, under time-series assumptions) when treating $\tilde{x}_t$ as random are not available, simulation evidence in Appendix C confirms that this holds at least for the particular DGP simulated (the exogenous part of the regressors is generated as a stationary vector autoregression).

Under this additional condition, the DaC estimator is asymptotically normal:

[4] A weaker but more technical alternative to Assumption *No Influential Observations* is the following: Let $\bar{\mathbf{A}}$ be a full-row-rank matrix with the same row space as $\mathbf{A}$. As an alternative to Assumption *No Influential Observations*, assume

$$\max_t \tilde{x}_t'(\tilde{\mathbf{X}}'_{(u(t))}\tilde{\mathbf{X}}_{(u(t))}/\bar{T})^{-1}\bar{\mathbf{A}}'(\bar{\mathbf{A}}(\tilde{\mathbf{X}}'_{(u(t))}\tilde{\mathbf{X}}_{(u(t))}/\bar{T})^{-1}\bar{\mathbf{A}}')^{-1}\bar{\mathbf{A}}(\tilde{\mathbf{X}}'_{(u(t))}\tilde{\mathbf{X}}_{(u(t))}/\bar{T})^{-1}\tilde{x}_t = o(\bar{T}).$$

The advantage of this assumption is that it does not require $\lambda_{\min}(\mathbf{A}\mathbf{A}') > c$.

Theorem 2 (Asymptotic normality)**.** *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*, and* No Influential Observations*,*

$$\frac{r'(\hat{\beta}-\beta)}{\sqrt{\Sigma/(h+1)^2}} \xrightarrow{d} N(0,1).$$

By contrast, inference results for the M&S estimator in the proportional regime are only available under the additional assumption of normally distributed errors.

We now turn to the question of inference based on the DaC estimator in the proportional regime. Two distinct difficulties have to be addressed.

The first difficulty is to account for the covariances between individual subsample estimators. Note that in the current model, the estimators from different subsamples are in general correlated even though errors are assumed serially uncorrelated. The covariances arise from the quadratic part of the estimators (i.e., the interactions between regression errors and feedback contained in regressors) as products of the form $\varepsilon_t\varepsilon_j$ (for $t \neq j$) appear in each individual subsample estimator. While these covariances are negligible when $K/\bar{T}$ vanishes, they are non-negligible in the proportional regime.[5] Unlike estimation of the variances of the individual subsample estimators, estimation of the covariance terms requires knowledge of the individual feedback vectors $\alpha_1, \ldots, \alpha_h$ or the availability of sufficiently accurate estimators of them. Instead of an estimation procedure involving estimation of $\alpha_1, \ldots, \alpha_h$, we propose to estimate an upper bound for $\Sigma/(h+1)^2$, which can be used for conservative inference. This upper bound takes the form

$$\frac{1+\Lambda}{(h+1)^2}\sum_{v=1}^{h+1}\Sigma_v,$$

where Λ is an estimable upper bound on the ratio of covariances to variances of the quadratic

[5]The fact that the covariances between subsample estimators are no longer negligible in the proportional regime has the same root as the inconsistency of the Eicker–Huber–White variance estimator in the proportional regime: the largest elements of the matrices $\tilde{\mathbf{X}}_{(v)}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)})^{-1}\tilde{\mathbf{X}}'_{(v)}$ no longer vanish.

parts of (1):[6]

$$\frac{\sum_{v\neq q}\operatorname{Cov}\left(\tilde{\varepsilon}'\mathbf{B}_{(v)}\tilde{\varepsilon},\tilde{\varepsilon}'\mathbf{B}_{(q)}\tilde{\varepsilon}\right)}{\sum_{v}\operatorname{Var}\left(\tilde{\varepsilon}'\mathbf{B}_{(v)}\tilde{\varepsilon}\right)}=\frac{\operatorname{tr}\left(\mathbf{B}^{2}\right)}{\operatorname{tr}\left(\mathbf{B}\mathbf{B}'\right)}\leq\Lambda\leq 1.$$

For the precise definition of Λ, the reader is referred to Appendix A.

For the variances Σ_v of the individual subsample estimators, we consider a sandwich estimator of the form

$$\hat{\Sigma}_{v}=r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)})^{-1}\sum_{t=1}^{\bar{T}}\hat{\sigma}^{2}_{r(v,t)}x_{r(v,t)}x'_{r(v,t)}(\mathbf{X}'_{(v)}\mathbf{X}_{(v)})^{-1}r,$$

where $\hat{\sigma}_t^2$ is one of the two alternative estimators $(\grave{\sigma}_t^2, \acute{\sigma}_t^2)$ introduced below. To construct the estimator of Λ, let

$$\mathbf{H}_{(v)}=\mathbf{I}_{\bar{T}}-\mathbf{X}_{(v)}(\mathbf{X}'_{(v)}\mathbf{X}_{(v)})^{-1}\mathbf{X}'_{(v)},$$

for each subsample $v \in \{1,\ldots,h+1\}$, as well as

$$\hat{d}_{v,q}=\sum_{t,j=1}^{\bar{T}}\mathbf{H}^{2}_{(v),tj}\hat{\sigma}^{2}_{r(v,t)}\hat{\sigma}^{2}_{r(q-h-1,j)},$$

$$\hat{d}_{q,v}=\sum_{t,j=1}^{\bar{T}}\mathbf{H}^{2}_{(q),tj}\hat{\sigma}^{2}_{r(q,t)}\hat{\sigma}^{2}_{r(v,j)},$$

and

$$\hat{c}_{v,q}=\sum_{t=1}^{\bar{T}}\sum_{j=1}^{\bar{T}-1}\mathbf{H}_{(v),t,j+1}\mathbf{H}_{(q),tj}\hat{\sigma}^{2}_{r(v,t)}\hat{\sigma}^{2}_{r(q,j)}$$

for $v < q$. The asymmetry between $\hat{d}_{v,q}$ and $\hat{d}_{q,v}$ is intended: in $\hat{d}_{v,q}$ the second subsample enters through its lagged block (see Appendix A). For $t \leq 0$, we simply use $\hat{\sigma}_t^2 = \hat{\sigma}_{t+h+1}^2$, which is not a good estimator of σ_t^2 but suffices for estimating Λ, as the resulting error is negligible.

[6]The properties of the $\mathbf{B}_{(v)}$ that lead from the first to the second expression are $\operatorname{tr}\left(\mathbf{B}^2_{(v)}\right)=0$ for all v and $\operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(q)}\right)=0$ for all $v \neq q$.

We then estimate Λ by

$$\hat{\Lambda} = \min\left\{1, \max_{1\le v<q\le h+1} \frac{|\hat{c}_{v,q}|}{\sqrt{\hat{d}_{v,q}\hat{d}_{q,v}}}\right\}.$$

This leaves us with the problem of estimating the σ_t^2 in such a way that the $\hat{\Sigma}_v$ and $\hat{\Lambda}$ are consistent, which brings us to the second difficulty: The failure of standard heteroskedasticity-robust variance estimators in the proportional regime due to overfitting. When K is proportional to T, the Eicker–Huber–White (HC0) variance estimator is no longer consistent and the same holds for all other members of the HCk class (Cattaneo, Jansson, and Newey 2018), even when focusing on just one or a constant number of linear combinations of the parameter. To understand the problem, let us temporarily put aside the subsampling aspect and suppose we are dealing with a standard cross-sectional or strictly exogenous context. The HC0 estimator simply estimates each σ_t^2 by $\hat{\varepsilon}_t^2$, the square of the standard OLS residual. If σ_t^2 could instead be estimated by ε_t^2, the resulting estimator would be consistent even in the proportional regime (under a somewhat strengthened condition on the most influential observation, see Assumption *Bounded Leverage* below). However, each $\hat{\varepsilon}_t^2$ depends not only on ε_t but also on the other ε_j. When K/T does not vanish, the influence of the most influential observation's ε_j on each other observation's $\hat{\varepsilon}_t$ does not vanish because the maximum leverage $\max_t x_t'(\mathbf{X}'\mathbf{X})^{-1}x_t$ is always at least K/T. For the HC0 estimator, this entails an asymptotic downward bias, whereas the HC3 estimator (the jackknife estimator of MacKinnon and White (1985)) is asymptotically upward biased (i.e., conservative). The fundamental problem is that each HCk estimator uses a squared residual (with different kinds of adjustments), which introduces a bias that does not vanish in the proportional regime. Several alternative heteroskedasticity-robust variance estimators have been proposed for the proportional regime. Cattaneo, Jansson, and Newey (2018) (henceforth CJN) propose a variance estimator (resembling that of Bera, Suprayitno, and Premaratne (2002)), which replaces the $\hat{\varepsilon}_t^2$ by linear combinations of all the $\hat{\varepsilon}_j^2$ in such a way that the effects of the $\hat{\varepsilon}_j^2$ for $j \neq t$ cancel in expectation. Concretely, returning to the weak exogeneity framework and subsampling idea of this paper and adapting the CJN estimator

to it leads to an estimator of the following form: Let $\hat{\varepsilon}_{r(v,t)}$ be the subsample OLS residual from the t-th observation of subsample v: $\hat{\varepsilon}_{r(v,t)} = y_{r(v,t)} - x'_{r(v,t)}\hat{\beta}_{(v)}$. That is, for each period we use the OLS estimator from its subsample rather than the DaC estimator to build the residual. Then for each $t \in \{1, \ldots, \bar{T}\}$ and each $v \in \{1, \ldots, h+1\}$, $\sigma^2_{r(v,t)}$ is estimated by

$$\grave{\sigma}^2_{r(v,t)} = \sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \hat{\varepsilon}^2_{r(v,j)}$$

for some weights $\boldsymbol{\kappa}_{(v),tj}$ that are functions of $\mathbf{X}_{(v)}$, chosen such that

$$\mathbb{E}\left[\sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \hat{\varepsilon}^2_{r(v,j)} \,\middle|\, \mathbf{X}_{(v)}\right] = \sigma^2_{r(v,t)}.$$

Using that $\hat{\varepsilon}_{r(v,t)} = \sum_{j=1}^{\bar{T}} \mathbf{H}_{(v),tj} \varepsilon_{r(v,j)}$, we have

$$\begin{aligned}
\mathbb{E}\left[\grave{\sigma}^2_{r(v,t)} | \mathbf{X}_{(v)}\right] =& \mathbb{E}\left[\sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \hat{\varepsilon}^2_{r(v,j)} | \mathbf{X}_{(v)}\right] \\
=& \sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \mathbb{E}\left[\hat{\varepsilon}^2_{r(v,j)} | \mathbf{X}_{(v)}\right] \\
=& \sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \mathbb{E}\left[\left(\sum_{l=1}^{\bar{T}} \mathbf{H}_{(v),jl} \varepsilon_{r(v,l)}\right)^2 | \mathbf{X}_{(v)}\right] \\
=& \sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \sum_{l=1}^{\bar{T}} \mathbf{H}^2_{(v),jl} \sigma^2_{r(v,l)} \\
=& \sum_{l=1}^{\bar{T}} \left(\sum_{j=1}^{\bar{T}} \boldsymbol{\kappa}_{(v),tj} \mathbf{H}^2_{(v),jl}\right) \sigma^2_{r(v,l)}.
\end{aligned}$$

Hence, assuming $\mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$ is invertible, choosing $\boldsymbol{\kappa}_{(v)} = \left(\mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}\right)^{-1}$ yields $\mathbb{E}\left[\grave{\sigma}^2_{r(v,t)} \mid \mathbf{X}_{(v)}\right] = \sigma^2_{r(v,t)}$. (The symbol $\odot$ denotes the Hadamard product, i.e., the elementwise product.)[7]

While this estimator ensures unbiasedness of $\hat{\Sigma}_v$ when $\hat{\sigma}^2_t = \grave{\sigma}^2_t$, consistency requires a bound on the norm of $\boldsymbol{\kappa}_{(v)}$. A sufficient condition for bounding $\boldsymbol{\kappa}_{(v)}$ is the following.

[7]Note that this choice of $\boldsymbol{\kappa}_{(v)}$ differs slightly from the one proposed by CJN. The difference is negligible.

Assumption *Sub-One-Half Leverage*. *There is a $c_\kappa > 0$ such that, for all T,*

$$\max_{1\le t\le T} \tilde{x}_t'(\tilde{\mathbf{X}}_{(u(t))}'\tilde{\mathbf{X}}_{(u(t))})^{-1}\tilde{x}_t \;\le\; \frac{1}{2} - c_\kappa.$$

This condition also ensures that $\mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$ is invertible with probability approaching 1. Finally, we also need to strengthen the assumption on influential observations somewhat:

Assumption *Bounded Leverage*.

$$\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}} \|f_{r(v,t)}\|^4 \le C$$

for all v.

(Note that Assumption *Bounded Leverage* implies the condition $\max_{1\le t\le T}\|f_t\| = o(T^{1/2})$ of Assumption *No Influential Observations*, but not the full-row-rank condition $\lambda_{\min}(\mathbf{A}\mathbf{A}') \ge c$ of the same assumption.) Assumption *Bounded Leverage* plays the role of a bounded fourth moment assumption on regressors. Under the above assumptions, $\grave{\Lambda}$ using the $\grave{\sigma}_t^2$ is consistent. In summary, we have the following result:

Theorem 3 (Conservative inference with the CJN estimator)**.** *For each $v \in \{1,\dots,h+1\}$, let $\grave{\Sigma}_v$ and $\grave{\Lambda}$ be $\hat{\Sigma}_v$ and $\hat{\Lambda}$ with $\hat{\sigma}_t^2 = \grave{\sigma}_t^2$, respectively. Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*,* No Influential Observations*,* Sub-One-Half Leverage*, and* Bounded Leverage*, we have*

$$\liminf_{T\to\infty} \mathbb{P}\left[r'\beta \in \left[r'\hat{\beta} \pm \frac{z_{1-\alpha/2}}{h+1}\sqrt{\left(1+\grave{\Lambda}\right)\sum_{v=1}^{h+1}\grave{\Sigma}_v}\right]\right] \;\ge\; 1-\alpha$$

for every $\alpha \in (0,1)$, where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

As $\max_{1\le t\le T}\ \tilde{x}_t'(\tilde{\mathbf{X}}_{(u(t))}'\tilde{\mathbf{X}}_{(u(t))})^{-1}\tilde{x}_t \ge K/\bar{T}$, an important implication of Assumption *Sub-One-Half Leverage* is that $K/\bar{T} < 1/2$, which may be too restrictive in many cases.

An alternative estimator was proposed by Jochmans (2022). For each $t \in \{1,\dots,\bar{T}\}$

and each $v \in \{1, \ldots, h+1\}$, it uses $\acute{\sigma}^2_{r(v,t)} := y_{r(v,t)} \frac{\hat{\varepsilon}_{r(v,t)}}{\mathbf{H}_{(v),tt}}$ to estimate the $\sigma^2_{r(v,t)}$.[8] The ratio $\hat{\varepsilon}_{r(v,t)}/\mathbf{H}_{(v),tt}$ is a leave-one-out estimator of $\varepsilon_{r(v,t)}$:

$$\frac{\hat{\varepsilon}_{r(v,t)}}{\mathbf{H}_{(v),tt}} = y_{r(v,t)} - x'_{r(v,t)}\hat{\beta}_{(v),-r(v,t)},$$

where $\hat{\beta}_{(v),-r(v,t)}$ is the OLS estimator from a regression in subsample v dropping period $r(v,t)$ (Kline, Saggio, and Sølvsten 2020). Multiplying the leave-one-out residual by $y_{r(v,t)} = x'_{r(v,t)}\beta + \varepsilon_{r(v,t)}$ circumvents the problematic squaring of the residual by introducing the product of the residual with the error directly ($\varepsilon_{r(v,t)}\hat{\varepsilon}_{r(v,t)}/\mathbf{H}_{(v),tt}$). It comes at the cost of introducing noise ($x'_{r(v,t)}\beta\hat{\varepsilon}_{r(v,t)}/\mathbf{H}_{(v),tt}$). Consistency of the resulting estimators requires a bound on this noise in terms of a bound on the fitted values $x'_t\beta$. We implement this bound by bounding the $\tilde{x}'_t\beta$. In addition, consistency of $\acute{\Lambda}$ (i.e., $\hat{\Lambda}$ using the $\acute{\sigma}^2_t$) requires the feedback to be a vanishing portion of the fitted values. This requirement arises as $\acute{\Lambda}$ involves products of $\acute{\sigma}^2_t$ from different subsamples, i.e., products of the form

$$\acute{\sigma}^2_{r(v,t)}\acute{\sigma}^2_{r(q,s)} = y_{r(v,t)} \frac{\hat{\varepsilon}_{r(v,t)}}{\mathbf{H}_{(v),tt}} y_{r(q,s)} \frac{\hat{\varepsilon}_{r(q,s)}}{\mathbf{H}_{(q),ss}}$$

for $q \neq v$. This product in turn involves a product of the feedback component $\varpi_{r(v,t)} := \sum_{l=1}^{h} \varepsilon_{r(v,t)-l}\alpha'_l\beta$ (contained in the $y_{r(v,t)} = \tilde{x}'_{r(v,t)}\beta + \varpi_{r(v,t)} + \varepsilon_{r(v,t)}$) with the residual $\hat{\varepsilon}_{r(q,s)}$ from another subsample. Both $\varpi_{r(v,t)}$ and $\hat{\varepsilon}_{r(q,s)}$ are functions of error terms from subsample q. Their product introduces a bias in $\acute{c}_{v,q}$ that vanishes only if the $\alpha'_l\beta$ vanish. When β and the α_l have bounded norms, the condition $\alpha'_l\beta = o(1)$ is, however, natural in the increasing K regime unless there are reasons to suspect alignment between α_l and β. The bound on the norms of the α_l is already imposed by Assumption *Bounds*. A bound on the norm of β is plausible if the average (squared) effects in models with more explanatory variables are smaller. We introduce the following assumption:

Assumption *Bounded Fitted Values*. $\max_t(\beta'\tilde{x}_t)^2 = o(T)$, $\alpha'_l\beta = o(1)$ *for all* $l = 1, \ldots, h$, *and* $\|\beta\| \leq C$.

Furthermore, for the leave-one-out approach to be feasible (or equivalently, for the $\mathbf{H}^{-1}_{(v),tt}$

[8] Note that this estimator differs slightly from the one proposed in Jochmans (2022), but the difference is negligible.

terms to remain stochastically bounded), the leave-one-out Gram matrices must be uniformly well-conditioned:

Assumption *Leave-One-Out Conditioning*. *For each subsample* $v \in \{1, \dots, h+1\}$ *and each* $t \in \{1, \dots, \bar{T}\}$, *define*

$$\tilde{\mathbf{X}}_{(v),-t} = (\tilde{x}_{r(v,1)}, \dots, \tilde{x}_{r(v,t-1)}, \tilde{x}_{r(v,t+1)}, \dots)'$$

($\tilde{\mathbf{X}}_{(v)}$ *with the* t*-th observation omitted). Then*

$$\lambda_{\min}\left(\frac{1}{\bar{T}}\tilde{\mathbf{X}}'_{(v),-t}\tilde{\mathbf{X}}_{(v),-t}\right) \geq c \quad \textit{for all } t.$$

To understand this condition, let us set the subsampling aside for a moment and consider a standard (maybe cross-sectional) regression without feedback ($h = 0$), i.e., $\tilde{\mathbf{X}} = \mathbf{X}$. If $\mathbf{X}'_{-t}\mathbf{X}_{-t}$ were singular while $\mathbf{X}'\mathbf{X}$ has full rank, this would mean that observation t always has a perfect fit regardless of the y_t so that nothing could be learned about ε_t or σ_t^2. To see this, note that $\mathbf{X}'_{-t}\mathbf{X}_{-t}$ and $\mathbf{X}'\mathbf{X}$ have full rank if and only if $\mathbf{X}_{-t}$ and $\mathbf{X}$ have full rank, respectively. Thus, if

$$\operatorname{rank}(\mathbf{X}'_{-t}\mathbf{X}_{-t}) = \operatorname{rank}(\mathbf{X}'\mathbf{X}) - 1 = K - 1,$$

x_t is not a linear combination of the other observations x_j, $j \neq t$. In other words, there is a $w \in \mathbb{R}^K$ satisfying $x_j'w = 0$ for all $j \neq t$ but $x_t'w \neq 0$. As a consequence, OLS regression of any vector y onto $\mathbf{X}$ always delivers a perfect fit for observation t: Let $b \in \mathbb{R}^K$ be such that $y_t - x_t'b \neq 0$. Then b cannot be the OLS estimator as there is a scalar γ such that adding γw to b sets the residual for observation t to zero ($y_t - x_t'(b + \gamma w) = 0$) without affecting any of the other residuals, proving that $\hat{\varepsilon}_t = 0$, so that it is uninformative about σ_t^2. When employing $\acute{\sigma}_t^2$, the problem manifests itself in the form of leverages $x_t'(\mathbf{X}'\mathbf{X})^{-1}x_t$ being close to 1 for some t (exactly 1 if $\mathbf{X}'_{-t}\mathbf{X}_{-t}$ is singular) so that the denominator of $\acute{\sigma}_t^2$ is close to 0 (or exactly 0 if $\mathbf{X}'_{-t}\mathbf{X}_{-t}$ is singular).

Under these two additional assumptions, we obtain the following result on conservative inference using $\acute{\sigma}_t^2$:

Theorem 4 (Conservative inference with the Jochmans estimator)**.** *For each* $v \in \{1, \dots, h+1\}$*, let* $\acute{\Sigma}_v$ *and* $\acute{\Lambda}$ *be* $\hat{\Sigma}_v$ *and* $\hat{\Lambda}$ *with* $\hat{\sigma}_t^2 = \acute{\sigma}_t^2$*, respectively. Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*,* No Influential Observations*,* Bounded Leverage*,* Bounded Fitted Values*, and* Leave-One-Out Conditioning*, we have*

$$\liminf_{T\to\infty} \mathbb{P}\left[r'\beta \in \left[r'\hat{\beta} \pm \frac{z_{1-\alpha/2}}{h+1}\sqrt{\left(1+\acute{\Lambda}\right)\sum_{v=1}^{h+1}\acute{\Sigma}_v}\right]\right] \geq 1-\alpha$$

for every $\alpha \in (0,1)$*, where* $z_{1-\alpha/2}$ *denotes the* $1-\alpha/2$ *quantile of the standard normal distribution.*

4 Simulations

We consider two types of simulations. The first set of simulations demonstrates the biases of OLS and the M&S estimator and contrasts their variances with that of the DaC estimator. The second set of simulations computes actual coverage of the conservative confidence intervals proposed in Section 3. Both sets of simulations are based on a DGP of the following form:

Data Generating Process

- Only the first regressor is affected by feedback.
- The strictly exogenous part of the regressors follows independent AR(1) processes with a common autoregressive parameter: $\tilde{x}_{k,t} = \rho\tilde{x}_{k,t-1} + \eta_{k,t}$, $|\rho| < 1$.
- $\eta_{k,t} \sim t(\nu)\sqrt{(\nu-2)/\nu}\,\sigma_\eta$ (where $t(\nu)$ is Student's t-distribution with ν degrees of freedom[a]) for some $\nu \in (2,\infty]$, so that $\mathrm{Var}\left(\eta_{k,t}\right) = \sigma_\eta^2$. $\nu = \infty$ means that $\eta_{k,t} \sim N(0,\sigma_\eta^2)$.
- Error terms are then drawn as follows: Let $\{\tilde{\varepsilon}_t\}_t \overset{\text{i.i.d.}}{\sim} N(0,1)$. The error terms are

$$\varepsilon_t = \tilde{\varepsilon}_t\sigma\frac{\|\tilde{x}_t\|^p}{\sqrt{\sum_{s=1}^T \|\tilde{x}_s\|^{2p}/T}}$$

for some $\sigma > 0$, $p \geq 1$ (with $\varepsilon_t = \tilde{\varepsilon}_t \sigma$ for $t \leq 0$). The factor $\|\tilde{x}_t\|^p$ introduces heteroskedasticity, while the denominator $\sqrt{\sum_{s=1}^{T} \|\tilde{x}_s\|^{2p}/T}$ normalises the variances to ensure that the average variance equals σ^2.

- h-period feedback is then added to the first regressor, $x_{1,t} = \tilde{x}_{1,t} + \sum_{l=1}^{h} a_l \varepsilon_{t-l}$, while the remaining regressors are strictly exogenous, i.e., $x_{k,t} = \tilde{x}_{k,t}$ for $k \geq 2$.

- The parameter of interest is β_1, i.e., $r = (1, 0, \ldots, 0)'$.

[a] For $\nu \in (0, 2]$, $Z \sim t(\nu)$ does not have a finite variance. For $\nu \in (2, 4]$, $Z \sim t(\nu)$ has $\mathrm{Var}(Z) = \nu/(\nu - 2)$ and infinite kurtosis.

This DGP violates the assumptions made in Section 3 in several ways. First, the exogenous part of the regressors is not treated as fixed. In the simulations, $\tilde{\mathbf{X}}$ is in fact redrawn in each replication to ensure that the results are not specific to a particular realisation of $\tilde{\mathbf{X}}$. Second, the heavy tails of the $\tilde{x}_t$, which we allow for in some configurations, may be incompatible even with reasonable probabilistic variants of the various deterministic assumptions made on $\tilde{\mathbf{X}}$ in Section 3. Third, error variances are not deterministically bounded above or below. The simulation results both on the variance of the DaC estimator and on the actual coverage of the confidence interval proposed in Section 3 suggest that the theoretical results presented in Section 3 are robust to these departures from the assumptions.

4.1 Bias and standard deviations across estimators

Tables 1 and 2 display the bias and standard deviation of the OLS, M&S, and DaC estimators of β_1 as approximated by Monte Carlo simulations for different sample sizes and parameter configurations. Note that the number of subsamples used for the DaC estimator is always $h+1$ and the assumed feedback horizon for the M&S estimator (which must also be specified) is always h. The M&S estimator exhibits a small but visible bias in the case of heavy-tailed regressors where the error standard deviation scales linearly ($p = 1$) or quadratically ($p = 2$) with the norm of the regressor vector. On the other hand, the DaC estimator, which we know to be unbiased, is often competitive in terms of variance with M&S (and OLS) and efficiency losses when the subsample aspect ratio is 0.8 (as for the case of $h = 3$ and $K/T = 0.2$)

are moderate. The results also show a sizeable bias of OLS. For the larger sample sizes simulated, the bias of OLS exceeds its standard deviation, in some configurations by a factor of more than three.

Table 1: Bias and standard deviation of the OLS, M&S, and DaC estimators of β_1, $h = 1$

	Bias			**Std. dev.**		
Sample size	OLS	M&S	DaC	OLS	M&S	DaC
$\nu = \mathbf{2.1}$, $p = \mathbf{1}$						
33	-0.728	-0.100	-0.069	0.882	0.830	1.032
100	-0.776	-0.090	0.022	0.570	0.530	0.590
300	-0.748	-0.082	-0.001	0.363	0.344	0.359
900	-0.731	-0.085	0.005	0.281	0.241	0.227
2700	-0.711	-0.076	0.002	0.204	0.191	0.148
$\nu = \mathbf{2.1}$, $p = \mathbf{2}$						
33	-0.777	-0.192	0.015	0.965	0.972	1.162
100	-0.846	-0.226	-0.017	0.675	0.675	0.675
300	-0.842	-0.190	0.003	0.551	0.545	0.495
900	-0.832	-0.226	0.020	0.541	0.527	0.464
$\nu = \mathbf{2.5}$, $p = \mathbf{1}$						
33	-0.530	-0.096	-0.021	0.713	0.711	0.875
100	-0.537	-0.063	-0.011	0.444	0.445	0.460
300	-0.505	-0.035	-0.002	0.265	0.248	0.265
900	-0.478	-0.028	0.003	0.168	0.144	0.154
$\nu = \mathbf{2.5}$, $p = \mathbf{2}$						
33	-0.508	-0.135	-0.028	0.785	0.760	0.897
100	-0.544	-0.094	0.021	0.509	0.463	0.493
300	-0.540	-0.069	-0.011	0.353	0.319	0.304
900	-0.516	-0.092	-0.007	0.276	0.252	0.221

Notes: $K/T = 0.25$, $a_1 = 0.2$, $\beta = 0.5$, $\sigma = 1$, $\rho = 0.9$, $\sigma_\eta = 0.2$.

Table 2: Bias and standard deviation of the OLS, M&S, and DaC estimators of β_1, $h = 3$

	Bias			**Std. dev.**		
Sample size	OLS	M&S	DaC	OLS	M&S	DaC
$\nu = \mathbf{2.1}$, $p = \mathbf{1}$						
33	-1.057	-0.316	-0.092	1.426	1.400	2.467
100	-1.016	-0.195	-0.026	0.851	0.815	1.250
300	-0.897	-0.144	-0.004	0.519	0.479	0.627
900	-0.874	-0.131	0.002	0.388	0.298	0.351
$\nu = \mathbf{2.1}$, $p = \mathbf{2}$						
33	-1.171	-0.469	-0.252	1.474	1.539	3.239
100	-1.179	-0.369	-0.003	0.992	0.976	1.199
300	-1.083	-0.345	0.021	0.761	0.723	0.634
900	-0.983	-0.388	0.005	0.614	0.585	0.375
$\nu = \mathbf{2.5}$, $p = \mathbf{1}$						
33	-0.522	-0.115	-0.017	0.923	0.916	1.562
100	-0.481	-0.081	0.037	0.546	0.534	0.923
300	-0.445	-0.026	0.008	0.313	0.300	0.447
900	-0.404	-0.026	0.001	0.187	0.171	0.237
$\nu = \mathbf{2.5}$, $p = \mathbf{2}$						
33	-0.530	-0.218	0.032	0.950	1.010	1.648
100	-0.540	-0.127	-0.004	0.608	0.592	0.824
300	-0.467	-0.099	0.010	0.371	0.333	0.438
900	-0.457	-0.078	-0.002	0.268	0.220	0.232

Notes: $K/T = 0.2$, $a_1 = 0.1$, $a_2 = 0.05$, $a_3 = 0.025$, $\beta = 0.5$, $\sigma = 1$, $\rho = 0.9$, $\sigma_\eta = 0.2$. Number of replications: 1000.

4.2 Coverage

Tables 3 and 4 show the coverage rates of different confidence intervals for different parameter configurations as simulated by Monte Carlo. The tables show the coverage of confidence intervals based only on the sum of the subsample variance estimates, $\sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $\sum_{v=1}^{h+1} \acute{\Sigma}_v$, respectively, as well as confidence intervals based on the estimated upper bounds for the full variance ($(1 + \grave{\Lambda}) \sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $(1 + \acute{\Lambda}) \sum_{v=1}^{h+1} \acute{\Sigma}_v$). We first notice that when the subsample aspect ratio $K/\bar{T}$ is large (0.8) and the sample size is small ($T = 100$), confidence intervals based on the CJN variance estimator undercover severely whether or not they are adjusted by an estimate of Λ. This is expected as Assumption *Sub-One-Half Leverage* requires the subsample aspect ratio to be below 0.5. Interestingly, however, CJN-based inference does well for larger sample sizes ($T = 900$) even when the subsample aspect ratio exceeds 0.5. The remaining cells mostly show that inference based on $\sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $\sum_{v=1}^{h+1} \acute{\Sigma}_v$ alone can occasionally undercover somewhat, especially when the number of subsamples is large ($h = 4$ in Table 4). The $1 + \grave{\Lambda}$ and $1 + \acute{\Lambda}$ factors mostly correct this, often leading to some degree of overcoverage as expected.

Table 3: Coverage of the DaC confidence intervals, $h = 1$

T	h	$K/\bar{T}$	ν	Cattaneo–Jansson–Newey		Jochmans		MC
				$\sum_{v=1}^{h+1} \grave{\Sigma}_v$	$(1+\grave{\Lambda}) \sum_{v=1}^{h+1} \grave{\Sigma}_v$	$\sum_{v=1}^{h+1} \acute{\Sigma}_v$	$(1+\acute{\Lambda}) \sum_{v=1}^{h+1} \acute{\Sigma}_v$	oracle
100	1	0.20	∞	0.944	0.954	0.946	0.959	0.947
300	1	0.20	∞	0.961	0.972	0.960	0.975	0.954
900	1	0.20	∞	0.950	0.963	0.950	0.964	0.949
100	1	0.80	∞	0.591	0.603	0.908	0.928	0.946
300	1	0.80	∞	0.818	0.857	0.939	0.951	0.948
900	1	0.80	∞	0.928	0.949	0.946	0.956	0.952
100	1	0.20	2.1	0.947	0.964	0.955	0.975	0.952
300	1	0.20	2.1	0.959	0.970	0.962	0.973	0.945
900	1	0.20	2.1	0.961	0.975	0.961	0.976	0.951
100	1	0.80	2.1	0.585	0.598	0.922	0.942	0.947
300	1	0.80	2.1	0.837	0.871	0.957	0.971	0.948
900	1	0.80	2.1	0.946	0.962	0.972	0.980	0.955

Empirical coverage of nominal 95% intervals for $r'\beta$, $r = e_1$, with $\rho = 0.90$, $\sigma_\eta = 0.20$, $\sigma = 1.00$, heteroskedasticity exponent $p = 1$, feedback loadings $a = (0.2)$, and β a constant vector with $\|\beta\| = 0.50$. The number of subsamples is $h+1$. A burn-in is simulated where necessary. 2000 replications per row. The DGP is that of Section 4. The columns headed $\sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $\sum_{v=1}^{h+1} \acute{\Sigma}_v$ use the uncorrected variance estimate and those headed $(1+\grave{\Lambda}) \sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $(1+\acute{\Lambda}) \sum_{v=1}^{h+1} \acute{\Sigma}_v$ the estimated upper bound; "oracle" is the coverage of the infeasible interval built on the Monte Carlo variance of the statistic. Coverage rates for the Cattaneo–Jansson–Newey estimator are conditional on invertibility of $\mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$, which held in all 2000 replications; the other columns use all replications. Monte Carlo standard error at the nominal level ≈ 0.005.

Table 4: Coverage of the DaC confidence intervals, $h = 4$

T	h	$K/\bar{T}$	ν	Cattaneo–Jansson–Newey		Jochmans		MC
				$\sum_{v=1}^{h+1} \grave{\Sigma}_v$	$(1+\grave{\Lambda})\sum_{v=1}^{h+1} \grave{\Sigma}_v$	$\sum_{v=1}^{h+1} \acute{\Sigma}_v$	$(1+\acute{\Lambda})\sum_{v=1}^{h+1} \acute{\Sigma}_v$	oracle
100	4	0.20	∞	0.942	0.976	0.944	0.975	0.950
300	4	0.20	∞	0.945	0.968	0.946	0.967	0.949
900	4	0.20	∞	0.953	0.962	0.954	0.964	0.953
100	4	0.80	∞	—	—	0.863	0.927	0.950
300	4	0.80	∞	0.668	0.700	0.934	0.963	0.949
900	4	0.80	∞	0.913	0.974	0.952	0.969	0.946
100	4	0.20	2.1	0.921	0.969	0.942	0.975	0.946
300	4	0.20	2.1	0.934	0.973	0.939	0.978	0.946
900	4	0.20	2.1	0.943	0.976	0.946	0.980	0.944
100	4	0.80	2.1	—	—	0.871	0.923	0.952
300	4	0.80	2.1	0.684	0.717	0.937	0.979	0.951
900	4	0.80	2.1	0.902	0.960	0.945	0.975	0.953

Empirical coverage of nominal 95% intervals for $r'\beta$, $r = e_1$, with $\rho = 0.90$, $\sigma_\eta = 0.20$, $\sigma = 1.00$, heteroskedasticity exponent $p = 1$, feedback loadings $a = (0.2)$, and β a constant vector with $\|\beta\| = 0.50$. The number of subsamples is $h + 1$. Here $a_2 = a_3 = a_4 = 0$, so that the true feedback horizon is 1 and $h = 4$ over-specifies it. A burn-in is simulated where necessary. 2000 replications per row. The DGP is that of Section 4. The columns headed $\sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $\sum_{v=1}^{h+1} \acute{\Sigma}_v$ use the uncorrected variance estimate and those headed $(1+\grave{\Lambda})\sum_{v=1}^{h+1} \grave{\Sigma}_v$ and $(1+\acute{\Lambda})\sum_{v=1}^{h+1} \acute{\Sigma}_v$ the estimated upper bound; "oracle" is the coverage of the infeasible interval built on the Monte Carlo variance of the statistic. Coverage rates for the Cattaneo–Jansson–Newey estimator are conditional on invertibility of $\mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$; a dash (—) indicates that $\mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$ was not invertible in any replication; the other columns use all replications. Monte Carlo standard error at the nominal level ≈ 0.005.

5 Conclusion

We have proposed a simple estimator for linear time-series regression models that is unbiased and consistent under weak exogeneity with a finite feedback horizon when the number of regressors is non-negligible as a fraction of the sample size. It is robust to heteroskedasticity in the errors. It is applicable when the number of regressors is large enough for the OLS estimator to be substantially biased but small enough to allow for splitting the sample into subsamples whose observations are at least $h + 1$ periods apart. Conservative asymptotic inference is available in the proportional regime using non-standard variance estimators.

References

Bai, Zhi-Dong and Jack W. Silverstein (1998). "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices". In: *The Annals of Probability* 26.1, pp. 316–345.

Bera, Anil K., Totok Suprayitno, and Gamini Premaratne (2002). "On some heteroskedasticity-robust estimators of variance–covariance matrix of the least-squares estimators". In: *Journal of Statistical Planning and Inference* 108.1–2, pp. 121–136.

Cattaneo, Matias D., Michael Jansson, and Whitney K. Newey (2018). "Inference in linear regression models with many covariates and heteroscedasticity". In: *Journal of the American Statistical Association* 113.523, pp. 1350–1361.

Dobriban, Edgar and Yue Sheng (2021). "Distributed linear regression by averaging". In: *The Annals of Statistics* 49.2, pp. 918–943.

Huber, Peter J. (1973). "Robust regression: asymptotics, conjectures and Monte Carlo". In: *The Annals of Statistics* 1.5, pp. 799–821.

Jochmans, Koen (2022). "Heteroscedasticity-robust inference in linear regression models with many covariates". In: *Journal of the American Statistical Association* 117.538, pp. 887–896.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2020). "Leave-out estimation of variance components". In: *Econometrica* 88.5, pp. 1859–1898.

Lei, Lihua, Peter J. Bickel, and Noureddine El Karoui (2018). "Asymptotics for high dimensional regression M-estimates: fixed design results". In: *Probability Theory and Related Fields* 172.3, pp. 983–1079.

MacKinnon, James G. and Halbert White (1985). "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties". In: *Journal of Econometrics* 29.3, pp. 305–325.

Marčenko, Vladimir A. and Leonid Andreevich Pastur (1967). "Distribution of eigenvalues for some sets of random matrices". In: *Mathematics of the USSR-Sbornik* 1.4, pp. 457–483.

Mikusheva, Anna and Mikkel Sølvsten (2025). "Linear regression with weak exogeneity". In: *Quantitative Economics* 16.2, pp. 367–403.

Sølvsten, Mikkel (2020). "Robust estimation with many instruments". In: *Journal of Econometrics* 214.2, pp. 495–512.

Stambaugh, Robert F. (1999). "Predictive regressions". In: *Journal of Financial Economics* 54.3, pp. 375–421.

Varah, James M. (1975). "A lower bound for the smallest singular value of a matrix". In: *Linear Algebra and its Applications* 11.1, pp. 3–5.

A Proofs

This preamble introduces the notation used throughout the appendix. Objects that are used only in a single statement or proof are defined there. Recall that T is assumed to be a multiple of $h+1$ and that the subsample size is denoted by $\bar{T} = T/(h+1)$.

Let $\varepsilon = (\varepsilon_{1-h}, \ldots, \varepsilon_T)'$ be the vector of all error terms affecting observed regressors or outcomes, and let e_l denote the lth canonical basis vector, i.e., the vector whose only non-zero element is its lth element, equal to 1. Its dimension is the one required by the context.

For $v \in \{1-h, \ldots, h+1\}$, define the $\bar{T} \times (T+h)$ selection matrix

$$\mathbf{S}_{(v)} = \begin{pmatrix} e'_{h+v} \\ e'_{v+2h+1} \\ e'_{v+3h+2} \\ \vdots \\ e'_{T+v-1} \end{pmatrix},$$

where $e_l \in \mathbb{R}^{T+h}$. For $v \geq 1$, $\mathbf{S}_{(v)}$ selects the periods belonging to the vth subsample from a vector containing periods $1-h$ through T. For $v \leq 0$, it selects their lagged counterparts. For example, $\mathbf{S}_{(1)}\varepsilon = (\varepsilon_1, \varepsilon_{2+h}, \varepsilon_{3+2h}, \ldots, \varepsilon_{T-h})'$ and $\mathbf{S}_{(-1)}\varepsilon = (\varepsilon_{-1}, \varepsilon_h, \varepsilon_{1+2h}, \ldots, \varepsilon_{T-h-2})'$. Moreover, $\mathbf{S}'_{(v)}\mathbf{S}_{(v)}$ is a diagonal matrix with diagonal entries in $\{0,1\}$, whence $\left\|\mathbf{S}_{(v)}\right\|_2 = 1$.

Let $\mathbf{J}$ denote the $\bar{T} \times \bar{T}$ shift matrix with ones on the first superdiagonal and zeros elsewhere, i.e., $\mathbf{J}_{tj} = 1$ if $j = t+1$ and $\mathbf{J}_{tj} = 0$ otherwise. Post-multiplying a matrix by $\mathbf{J}$ shifts its columns one place to the right, and pre-multiplying by $\mathbf{J}'$ shifts its rows one place down, in both cases filling the vacated column or row with zeros. For any $(T+h)$-vector b and $v, q \in \{1-h, \ldots, h+1\}$,

$$\mathbf{S}_{(v)}\mathrm{diag}(b)\mathbf{S}'_{(q)} = \begin{cases} \mathrm{diag}(\mathbf{S}_{(v)}b) & \text{if } v = q, \\ \mathrm{diag}(\mathbf{S}_{(v)}b)\,\mathbf{J} & \text{if } v \geq 1,\ q = v-(h+1), \\ \mathbf{J}'\mathrm{diag}(\mathbf{S}_{(q)}b) & \text{if } v \leq 0,\ q = v+h+1, \\ \mathbf{0}_{\bar{T}\times\bar{T}} & \text{otherwise,} \end{cases} \tag{3}$$

where, in block form,

$$\mathrm{diag}(\mathbf{S}_{(v)}b)\,\mathbf{J} = \begin{pmatrix} \mathbf{0}_{\bar{T}-1} & \mathrm{diag}(b_{r(v,1)}, \ldots, b_{r(v,\bar{T}-1)}) \\ 0 & \mathbf{0}'_{\bar{T}-1} \end{pmatrix},$$

$$\mathbf{J}'\mathrm{diag}(\mathbf{S}_{(q)}b) = \begin{pmatrix} \mathbf{0}'_{\bar{T}-1} & 0 \\ \mathrm{diag}(b_{r(q,1)}, \ldots, b_{r(q,\bar{T}-1)}) & \mathbf{0}_{\bar{T}-1} \end{pmatrix},$$

and $b_{r(v,t)}$ denotes the element of b belonging to period $r(v,t) = v + (t-1)(h+1)$. This

holds because the tth row of $\mathbf{S}_{(v)}$ selects period $r(v,t)$, so that $(\mathbf{S}_{(v)}\mathrm{diag}(b)\mathbf{S}'_{(q)})_{tj}$ is non-zero only if $r(v,t) = r(q,j)$, i.e., $v - q = (j-t)(h+1)$. For $|v-q| \leq h$, this requires $v = q$ and $j = t$. For $q = v-(h+1)$, it requires $j = t+1$, and for $q = v+h+1$, it requires $j = t-1$. For all other pairs, it is impossible.

We call the set of periods selected by $\mathbf{S}_{(v)}$ the block v. Blocks with $v \in \{1,\ldots,h+1\}$ are the $h+1$ subsamples, and blocks with $v \leq 0$ are the lagged blocks. We write u^* for the subsample to which block u belongs, that is, $u^* := u$ for $u \geq 1$ and $u^* := u+h+1$ for $u \leq 0$. For a lagged block $u \leq 0$,

$$r(u,t) = r(u^*, t-1) \quad \text{for } t \geq 2, \qquad r(u,1) = u \leq 0, \tag{4}$$

so that a lagged block consists of $\bar{T}-1$ periods of the subsample u^* together with one pre-sample period.

Let $\mathbf{\Phi}^2 = \mathrm{diag}(\sigma^2_{1-h},\ldots,\sigma^2_T)$ be the diagonal matrix of error variances (including the pre-sample error variances $\sigma^2_{1-h},\ldots,\sigma^2_0$), and, for every block u, let $\sigma^2_{(u)} := (\sigma^2_{r(u,1)},\ldots,\sigma^2_{r(u,\bar{T})})'$ and $\mathbf{\Phi}^2_{(u)} := \mathbf{S}_{(u)}\mathbf{\Phi}^2\mathbf{S}'_{(u)} = \mathrm{diag}(\sigma^2_{(u)})$. By Assumption *Regular Errors*, the variances of pre-sample periods also lie in $[c,C]$, so $\sigma^2_{(u)}$ is well defined for every block. A pre-sample period differs from the other periods only in that no residual is observed for it. Likewise, $\hat{\sigma}^2_{(u)} := (\hat{\sigma}^2_{r(u,1)},\ldots,\hat{\sigma}^2_{r(u,\bar{T})})'$ for a subsample $u \geq 1$ and a vector of estimates $(\hat{\sigma}^2_t)_t$. For $u \leq 0$, we let $\hat{\sigma}^2_{(u)} := (\hat{\sigma}^2_{u+h+1}, \hat{\sigma}^2_{u+h+1}, \hat{\sigma}^2_{u+2(h+1)},\ldots,\hat{\sigma}^2_{r(u,\bar{T})})'$, so that the pre-sample variance σ^2_u is proxied by $\hat{\sigma}^2_{r(u^*,1)}$. By (3), for $v,q \in \{1-h,\ldots,h+1\}$,

$$\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(q)} = \begin{cases} \mathbf{\Phi}^2_{(v)} & \text{if } v = q, \\ \mathbf{\Phi}^2_{(v)}\mathbf{J} & \text{if } v \geq 1,\ q = v-(h+1), \\ \mathbf{J}'\mathbf{\Phi}^2_{(q)} & \text{if } v \leq 0,\ q = v+h+1, \\ \mathbf{0}_{\bar{T}\times\bar{T}} & \text{otherwise.} \end{cases} \tag{5}$$

Under Assumption *Feedback Decomposition*, the matrices $\mathbf{X}_{(v)}$ and $\tilde{\mathbf{X}}_{(v)}$ of observed and

strictly exogenous regressors from subsample v (Sections 2 and 3) are related by

$$\mathbf{X}_{(v)} = \tilde{\mathbf{X}}_{(v)} + \sum_{l=1}^{h} \mathbf{S}_{(v-l)}\varepsilon\alpha_l'.$$

For each subsample v, let $\mathbf{\Theta}_{(v)}$ be the transformation of Lemma 1, let $\mathbf{\Theta}_{(v),1}$ denote its first $h+1$ columns and $\mathbf{\Theta}_{(v),2}$ the remaining columns. Let $\mathbf{Z}_{(v)} := \mathbf{X}_{(v)}\mathbf{\Theta}_{(v)}$, $\mathbf{Z}_{1,(v)} := \mathbf{X}_{(v)}\mathbf{\Theta}_{(v),1}$, $\mathbf{Z}_{2,(v)} := \mathbf{X}_{(v)}\mathbf{\Theta}_{(v),2}$, $\tilde{\mathbf{Z}}_{1,(v)} := \tilde{\mathbf{X}}_{(v)}\mathbf{\Theta}_{(v),1}$ and $\tilde{\mathbf{Z}}_{2,(v)} := \tilde{\mathbf{X}}_{(v)}\mathbf{\Theta}_{(v),2}$, and, for each period $t \in \{1,\ldots,T\}$, let $z_{1,t} := \mathbf{\Theta}'_{(u(t)),1}x_t$, $z_{2,t} := \mathbf{\Theta}'_{(u(t)),2}x_t$ and $\tilde{z}_{1,t} := \mathbf{\Theta}'_{(u(t)),1}\tilde{x}_t$. By Lemma 1, $\alpha_l'\mathbf{\Theta}_{(v),2} = 0$ for all l, so that $\mathbf{Z}_{2,(v)} = \tilde{\mathbf{Z}}_{2,(v)}$ is non-random, and $\tilde{\mathbf{Z}}'_{1,(v)}\tilde{\mathbf{Z}}_{2,(v)} = \mathbf{0}$. For $l \in \{1,\ldots,h\}$, let $r_{(v)}, \alpha_{(v,l)} \in \mathbb{R}^{h+1}$ be the first $h+1$ elements of $r'\mathbf{\Theta}_{(v)} = ({r_{(v)}}', 0'_{K-h-1})$ and $\alpha_l'\mathbf{\Theta}_{(v)} = (\alpha'_{(v,l)}, 0'_{K-h-1})$, respectively. Let $\mathbf{M}_{(v)} := \mathbf{I}_{\bar{T}} - \mathbf{Z}_{2,(v)}(\mathbf{Z}'_{2,(v)}\mathbf{Z}_{2,(v)})^{-1}\mathbf{Z}'_{2,(v)}$, which is well defined and non-random because $\mathbf{Z}'_{2,(v)}\mathbf{Z}_{2,(v)} = \bar{T}\mathbf{I}_{K-h-1}$ by Lemma 1. As $\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{Z}_{2,(v)} = \mathbf{0}$,

$$\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)}. \tag{6}$$

For each position $t \in \{1,\ldots,\bar{T}\}$, let $\hat{z}'_{1,r(v,t)}$ denote row t of $\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}$ and let $\psi_{(v,l),t} := \mathbf{M}_{(v),t\cdot}\mathbf{S}_{(v-l)}\varepsilon$ for $l \in \{1,\ldots,h\}$. By (6) and the decomposition of $\mathbf{X}_{(v)}$ above,

$$\hat{z}_{1,r(v,t)} = \tilde{z}_{1,r(v,t)} + \sum_{l=1}^{h} \psi_{(v,l),t}\,\alpha_{(v,l)}. \tag{7}$$

Let $\mathbf{Q}_{(v)} := \mathbb{E}\left[\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right]$, which is invertible by Lemma 4. For each subsample v and $l \in \{1,\ldots,h\}$, let

- $\gamma_{v,l} := r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\alpha_{(v,l)}$,
- $\omega_{(v)} := \left(r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}\right)'$,
- $\mathbf{B}_{(v)} := \mathbf{\Phi}\sum_{l=1}^{h}\gamma_{v,l}\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}$,
- $U_v := \omega'_{(v)}\mathbf{\Phi}^{-1}\varepsilon + \varepsilon'\mathbf{\Phi}^{-1}\mathbf{B}_{(v)}\mathbf{\Phi}^{-1}\varepsilon$, and
- $\Sigma_v := \mathrm{Var}\left(U_v\right) = \mathrm{Var}\left(r'_{(v)}\mathbf{Q}^{-1}_{(v)}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)$.

Moreover, let $\omega := \sum_{v=1}^{h+1} \omega_{(v)}$, $\mathbf{B} := \sum_{v=1}^{h+1} \mathbf{B}_{(v)}$, $U := \sum_{v=1}^{h+1} U_v = \omega' \mathbf{\Phi}^{-1} \varepsilon + \varepsilon' \mathbf{\Phi}^{-1} \mathbf{B} \mathbf{\Phi}^{-1} \varepsilon$ and $\Sigma := \mathrm{Var}\,(U)$. We note that $\|\omega\|^2 = \sum_{v=1}^{h+1} \|\omega_{(v)}\|^2$ because the $\omega_{(v)}$ have disjoint supports (see the proof of Lemma 7).

For each subsample v, let $\mathbf{\Psi}_{(v)} := \mathbf{Z}'_{1,(v)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)}$. The subsample variance estimates $\grave{\Sigma}_v$ and $\acute{\Sigma}_v$ are those of Theorems 3 and 4, respectively.

The estimates $\hat{c}_{v,q}$, $\hat{d}_{v,q}$ and $\hat{d}_{q,v}$ of Section 3 estimate the following traces: for $1 \leq v < q \leq h+1$, let

$$
\begin{aligned}
c_{v,q} &:= \mathrm{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(q)} \mathbf{\Phi}^2_{(q)} \mathbf{J}\right) = \sum_{t=1}^{\bar{T}} \sum_{j=1}^{\bar{T}-1} \mathbf{M}_{(v),t,j+1} \mathbf{M}_{(q),tj}\, \sigma^2_{r(v,t)}\, \sigma^2_{r(q,j)}, \\
d_{v,q} &:= \mathrm{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(q-h-1)}\right) = \sum_{t=1}^{\bar{T}} \sum_{j=1}^{\bar{T}} \mathbf{M}^2_{(v),tj}\, \sigma^2_{r(v,t)}\, \sigma^2_{r(q-h-1,j)}, \\
d_{q,v} &:= \mathrm{tr}\left(\mathbf{M}_{(q)} \mathbf{\Phi}^2_{(q)} \mathbf{M}_{(q)} \mathbf{\Phi}^2_{(v)}\right) = \sum_{t=1}^{\bar{T}} \sum_{j=1}^{\bar{T}} \mathbf{M}^2_{(q),tj}\, \sigma^2_{r(q,t)}\, \sigma^2_{r(v,j)}.
\end{aligned}
\tag{8}
$$

Both families of traces are traces of products of two of the blocks $\mathbf{B}_{(v)}$, one from each of the two parts of $\Sigma = \sum_{v=1}^{h+1} \Sigma_v + \mathrm{tr}\,(\mathbf{B}^2)$ (Lemma 7): with $l(v,q) := h + 1 - (q - v)$ and $m(v,q) := q - v$,

$$
\begin{aligned}
\mathrm{tr}\,(\mathbf{B}\mathbf{B}') &= \sum_{v=1}^{h+1} \sum_{l=1}^{h} \gamma^2_{v,l}\, \mathrm{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v-l)}\right) \\
&= \sum_{1 \leq v < q \leq h+1} \left(\gamma^2_{v,l(v,q)}\, d_{v,q} + \gamma^2_{q,m(v,q)}\, d_{q,v}\right), \\
\mathrm{tr}\left(\mathbf{B}^2\right) &= 2 \sum_{1 \leq v < q \leq h+1} \gamma_{v,h+1-(q-v)}\, \gamma_{q,q-v}\ \mathrm{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(q)} \mathbf{\Phi}^2_{(q)} \mathbf{J}\right) \\
&= 2 \sum_{1 \leq v < q \leq h+1} \gamma_{v,l(v,q)}\, \gamma_{q,m(v,q)}\, c_{v,q}.
\end{aligned}
\tag{9}
$$

The first equalities for $\mathrm{tr}\,(\mathbf{B}\mathbf{B}')$ and $\mathrm{tr}\,(\mathbf{B}^2)$ are Lemma 7(i) and (ii), respectively. The second equality for $\mathrm{tr}\,(\mathbf{B}^2)$ follows from the definitions of $l(v,q)$, $m(v,q)$ and $c_{v,q}$. The second equality for $\mathrm{tr}\,(\mathbf{B}\mathbf{B}')$ regroups the double sum by pairs of subsamples. Fix v and a lag l. If $l < v$, the block $v - l$ is the subsample $u := v - l$, and the summand equals $\gamma^2_{v,m(u,v)} d_{v,u}$ because $l = m(u,v)$. If $l \geq v$, the block $v - l$ is the lagged block $q - h - 1$ of the subsample

$q := v - l + h + 1 > v$, and the summand equals $\gamma^2_{v,l(v,q)} d_{v,q}$ because $l = l(v,q)$. Each ordered pair of distinct subsamples arises in exactly one of these two ways.

The design factor Λ of Section 3 is

$$\Lambda := \min\{1, \lambda\}, \qquad \lambda := \max_{1 \le v < q \le h+1} \frac{|c_{v,q}|}{\sqrt{d_{v,q}\, d_{q,v}}},$$

which is well defined because $d_{v,q} \ge c^2 \operatorname{tr}\left(\mathbf{M}_{(v)}\right) = c^2(\bar{T} - K + h + 1) > 0$ by Assumptions *Regular Errors* and *Proportional Regime* and the idempotency of $\mathbf{M}_{(v)}$, and likewise for $d_{q,v}$.

All three traces in (8) have the shape

$$\vartheta = \sum_{t=1}^{\bar{T}} \sum_{j=1}^{\bar{T}} w_{tj}\, \sigma^2_{(u_1),t}\, \sigma^2_{(u_2),j}, \tag{10}$$

with u_1 and u_2 two blocks coming from distinct subsamples and $\mathbf{W} = (w_{tj})$ a deterministic array. By the double sums in (8), $d_{v,q}$ has the pair $(u_1, u_2) = (v, q-h-1)$ and $w_{tj} = \mathbf{M}^2_{(v),tj}$, and $d_{q,v}$ has the pair $(u_1, u_2) = (q, v)$ and $w_{tj} = \mathbf{M}^2_{(q),tj}$. Moreover, $c_{v,q}$ has the pair $(u_1, u_2) = (v, q)$ and $w_{tj} = \mathbf{M}_{(v),t,j+1}\mathbf{M}_{(q),tj}$ for $j \le \bar{T} - 1$, with an empty last column. We call an array $\mathbf{W} = (w_{tj})$ admissible if

$$\begin{aligned}
&\text{(W1)} \quad \max_t \sum_j |w_{tj}| \le 1 \ \text{ and } \ \max_j \sum_t |w_{tj}| \le 1;\\
&\text{(W2)} \quad \|\mathbf{W}\|_F^2 = O(\bar{T});\\
&\text{(W3)} \quad \|\mathbf{W}\|_2 = O(1).
\end{aligned}$$

The estimates $\grave{\sigma}^2_t$ and $\acute{\sigma}^2_t$ depend on $\mathbf{H}_{(v)}$ and hence, through the feedback, on the errors of the h blocks preceding v. Estimates from different subsamples are therefore not independent. We will therefore compare each of them with an auxiliary construction that is built on the deterministic, but unobservable, $\mathbf{M}_{(v)}$ instead of $\mathbf{H}_{(v)}$. Write $\grave{\sigma}^2_{(u)}$ and $\acute{\sigma}^2_{(u)}$ for $\hat{\sigma}^2_{(u)}$ with $\hat{\sigma}^2_t = \grave{\sigma}^2_t$ and $\hat{\sigma}^2_t = \acute{\sigma}^2_t$, respectively.

The auxiliary construction for $\grave{\sigma}^2_{(v)}$, used in the proof of Theorem 3, is defined as follows.

For each subsample v, let

$$\dot{\varepsilon}_{(v)} := \mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon \qquad \text{and} \qquad \check{\boldsymbol{\kappa}}_{(v)} := \left(\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right)^{-1},$$

where $\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}$ is invertible under Assumption *Sub-One-Half Leverage* by Lemma 22(i), and let

$$\check{\sigma}^2_{(v)} := \check{\boldsymbol{\kappa}}_{(v)}\, \dot{\varepsilon}^{\odot 2}_{(v)},$$

where the power $\odot 2$ denotes the entrywise power. Thus $\check{\sigma}^2_{(v)}$ is $\grave{\sigma}^2_{(v)}$ with $\mathbf{M}_{(v)}$ in place of $\mathbf{H}_{(v)}$.

The auxiliary construction for $\acute{\sigma}^2_{(v)}$, used in the proof of Theorem 4, is defined as follows. For each subsample v, let $\tilde{\mu}_{(v)} := \tilde{\mathbf{X}}_{(v)}\beta$ be the non-random part of the fitted values, let $\mathbf{m}_{(v)} := (\mathbf{M}_{(v),11}, \ldots, \mathbf{M}_{(v),\bar{T}\bar{T}})'$, and let

$$\breve{\sigma}^2_{(v)} := \left(\tilde{\mu}_{(v)} + \mathbf{S}_{(v)}\varepsilon\right) \odot \left(\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right) \oslash \mathbf{m}_{(v)},$$

where $\oslash$ denotes entrywise division. The entries of $\mathbf{m}_{(v)}$ are positive under Assumption *Sub-One-Half Leverage* by Lemma 16 and under Assumption *Leave-One-Out Conditioning* by Lemma 28(i). Thus $\breve{\sigma}^2_{(v)}$ is $\acute{\sigma}^2_{(v)}$ with $\mathbf{M}_{(v)}$ in place of $\mathbf{H}_{(v)}$ and without the feedback component $\varpi_{r(v,t)} = \sum_{l=1}^{h} \varepsilon_{r(v,t)-l}\alpha_l'\beta$ of the fitted values. As a consequence, $\breve{\sigma}^2_{(v)}$ is a function of the errors of a single subsample. Lemma 29 shows that omitting $\varpi_{r(v,t)}$ is negligible under Assumption *Bounded Fitted Values*.

For a lagged block $u \leq 0$, $\check{\sigma}^2_{(u)}$ and $\breve{\sigma}^2_{(u)}$ are defined from those of the subsample $u^* = u + h + 1$ by $\check{\sigma}^2_{(u)} := (\sigma^2_u, \check{\sigma}^2_{(u^*),1}, \ldots, \check{\sigma}^2_{(u^*),\bar{T}-1})'$, and likewise for $\breve{\sigma}^2_{(u)}$. Their first entry is the true pre-sample variance σ^2_u, whereas the first entry of $\hat{\sigma}^2_{(u)}$ is the proxy $\hat{\sigma}^2_{r(u^*,1)}$. Lemmas 25 and 29 bound the resulting difference. For every block u, the covariance matrices of $\check{\sigma}^2_{(u)}$ and $\breve{\sigma}^2_{(u)}$ are denoted by $\check{\boldsymbol{\Gamma}}_{(u)} := \mathrm{Var}\left(\check{\sigma}^2_{(u)}\right)$ and $\breve{\boldsymbol{\Gamma}}_{(u)} := \mathrm{Var}\left(\breve{\sigma}^2_{(u)}\right)$.

Lemma 28 bounds $\breve{\boldsymbol{\Gamma}}_{(v)}$ through a split of $\breve{\sigma}^2_{(v)} - \sigma^2_{(v)}$ into a pure-error part and a fitted-value part. For each subsample v, let $\mathbf{D}_{(v)} := \mathrm{diag}(\tilde{\mu}_{(v)})$, and let $\theta^0_{(v)}$ and $\theta^{\mu}_{(v)}$ be the vectors

with entries

$$\theta^0_{(v),t} := \varepsilon^2_{r(v,t)} - \sigma^2_{r(v,t)} + \sum_{j \neq t} \frac{\mathbf{M}_{(v),tj}}{\mathbf{M}_{(v),tt}} \varepsilon_{r(v,t)} \varepsilon_{r(v,j)}, \qquad \theta^\mu_{(v),t} := \frac{\tilde{\mu}_{(v),t} \left(\mathbf{M}_{(v)} \mathbf{S}_{(v)} \varepsilon \right)_t}{\mathbf{M}_{(v),tt}},$$

so that $\breve{\sigma}^2_{(v)} - \sigma^2_{(v)} = \theta^0_{(v)} + \theta^\mu_{(v)}$. Let $\mathbf{\Gamma}^0_{(v)} := \mathrm{Var}\left(\theta^0_{(v)}\right)$ and

$$\mathbf{F}_{(v)} := \mathrm{diag}(\mathbf{m}_{(v)})^{-1} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \, \mathrm{diag}(\mathbf{m}_{(v)})^{-1},$$

so that $\mathbf{\Gamma}^\mu_{(v)} := \mathrm{Var}\left(\theta^\mu_{(v)}\right) = \mathbf{D}_{(v)} \mathbf{F}_{(v)} \mathbf{D}_{(v)}$. For a lagged block $u \leq 0$, let $\mathbf{D}_{(u)} := \mathbf{J}' \mathbf{D}_{(u^*)} \mathbf{J}$, $\mathbf{F}_{(u)} := \mathbf{J}' \mathbf{F}_{(u^*)} \mathbf{J}$, $\mathbf{\Gamma}^0_{(u)} := \mathbf{J}' \mathbf{\Gamma}^0_{(u^*)} \mathbf{J}$ and $\mathbf{\Gamma}^\mu_{(u)} := \mathbf{D}_{(u)} \mathbf{F}_{(u)} \mathbf{D}_{(u)}$.

The estimates $\grave{\sigma}^2_{(v)}$ and $\acute{\sigma}^2_{(v)}$ are compared with $\breve{\sigma}^2_{(v)}$ and $\breve{\sigma}^2_{(v)}$ through the difference $\mathbf{R}_{(v)}$ between $\mathbf{M}_{(v)}$ and $\mathbf{H}_{(v)}$. For each subsample v, let

$$\mathbf{T}_{(v)} := \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)} \left(\mathbf{Z}'_{1,(v)} \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)} \right)^{-1/2}, \qquad \mathbf{R}_{(v)} := \mathbf{M}_{(v)} - \mathbf{H}_{(v)}, \qquad \hat{\varepsilon}_{(v)} := \mathbf{H}_{(v)} \mathbf{S}_{(v)} \varepsilon,$$

where $\mathbf{T}_{(v)}$ is well defined with probability approaching one by Lemma 3, and let

$$\eta := \max_{1 \leq v \leq h+1} \max_{1 \leq t \leq \bar{T}} \| \mathbf{T}_{(v),t\cdot} \|.$$

By the Frisch–Waugh–Lovell decomposition of $\mathrm{col}(\mathbf{X}_{(v)})$,

$$\mathbf{R}_{(v)} = \mathbf{T}_{(v)} \mathbf{T}'_{(v)}, \tag{11}$$

an orthogonal projection onto a space of dimension at most $h+1$, so that $\left\| \mathbf{R}_{(v)} \right\|^2_F = \mathrm{tr}\left(\mathbf{R}_{(v)} \right) \leq h+1$ for every realisation. The entries of $\hat{\varepsilon}_{(v)}$ are the residuals of Section 3, $\hat{\varepsilon}_{(v),t} = \hat{\varepsilon}_{r(v,t)}$, and $\dot{\varepsilon}_{(v)} = \hat{\varepsilon}_{(v)} + \mathbf{R}_{(v)} \mathbf{S}_{(v)} \varepsilon$.

The weight matrices $\boldsymbol{\kappa}_{(v)}$ and $\check{\boldsymbol{\kappa}}_{(v)}$ are the inverses of

$$\mathbf{G}_{(v)} := \mathbf{H}_{(v)} \odot \mathbf{H}_{(v)} \qquad \text{and} \qquad \dot{\mathbf{G}}_{(v)} := \mathbf{M}_{(v)} \odot \mathbf{M}_{(v)},$$

respectively, that is, $\boldsymbol{\kappa}_{(v)} = \mathbf{G}^{-1}_{(v)}$ and $\check{\boldsymbol{\kappa}}_{(v)} = \dot{\mathbf{G}}^{-1}_{(v)}$. Entry by entry, $\mathbf{G}_{(v),tj} - \dot{\mathbf{G}}_{(v),tj} =$

$\left(\mathbf{M}_{(v),tj} - \mathbf{R}_{(v),tj}\right)^2 - \mathbf{M}_{(v),tj}^2$, that is,

$$\mathbf{G}_{(v)} - \dot{\mathbf{G}}_{(v)} = -2\left(\mathbf{M}_{(v)} \odot \mathbf{R}_{(v)}\right) + \left(\mathbf{R}_{(v)} \odot \mathbf{R}_{(v)}\right). \tag{12}$$

A.1 Preliminary lemmas

The following lemma is a version of Lemma 5 from Mikusheva and Sølvsten (2025). It will be used to transform the regressors in such a way that only a fixed number of them are affected by feedback, the remaining regressors are strictly exogenous, and $r'\beta$ depends only on a fixed number of parameters.

Lemma 1 (Normalising transformation)**.** *For $K \geq h+1$, let the $(h+1) \times K$ matrix $\mathbf{A}$ be given by $\mathbf{A} = (r, \alpha_1, \ldots, \alpha_h)'$. For each subsample $v \in \{1, \ldots, h+1\}$, there is an invertible $K \times K$ matrix $\mathbf{\Theta}_{(v)}$ satisfying:*

- $\mathbf{\Theta}'_{(v)}\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\mathbf{\Theta}_{(v)} = \bar{T}\mathbf{I}_K$,
- $\mathbf{A}\mathbf{\Theta}_{(v)} = \begin{bmatrix}\mathbf{A}^{\star\prime} & \mathbf{0}_{(h+1)\times(K-h-1)}\end{bmatrix}$ *where* $\mathbf{A}^\star \in \mathbb{R}^{(h+1)\times(h+1)}$ *satisfies* $\mathbf{A}^{\star\prime}\mathbf{A}^\star = \bar{T}\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)})^{-1}\mathbf{A}'$ *and* $\mathbf{0}_{(h+1)\times(K-h-1)}$ *is an $(h+1) \times (K-h-1)$ matrix of zeros,*
- *and* $\mathbf{\Theta}_{(v)}\mathbf{\Theta}'_{(v)} = \left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}\right)^{-1} = \bar{T}\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\right)^{-1}$. *Thus, the eigenvalues of $\mathbf{\Theta}'_{(v)}\mathbf{\Theta}_{(v)}$ are those of $(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}$, too.*

Proof. We construct $\mathbf{\Theta}_{(v)}$ as follows. Let $\mathbf{\Upsilon}$ be the symmetric square root of $\frac{1}{\bar{T}}\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}$. Let $\mathbf{L}\mathbf{A}^\star$ be the QR decomposition of $\mathbf{\Upsilon}^{-1}\mathbf{A}'$, meaning $\mathbf{L}$ is a $K \times (h+1)$ matrix whose columns are orthonormal, $\mathbf{A}^\star$ is an upper triangular $(h+1) \times (h+1)$ matrix and $\mathbf{L}\mathbf{A}^\star = \mathbf{\Upsilon}^{-1}\mathbf{A}'$. Let $g_1, \ldots, g_{K-h-1} \in \mathbb{R}^K$ be an orthonormal set of vectors, each of which is orthogonal to all columns of $\mathbf{L}$. Let $\mathbf{O} = [\mathbf{L}, g_1, \ldots, g_{K-h-1}]$ (note $\mathbf{O}' = \mathbf{O}^{-1}$) and $\bar{\mathbf{A}} = \begin{pmatrix}\mathbf{A}^{\star\prime} & \mathbf{0}_{(h+1)\times(K-h-1)}\end{pmatrix}'$. Note that $\mathbf{O}\bar{\mathbf{A}} = \mathbf{L}\mathbf{A}^\star = \mathbf{\Upsilon}^{-1}\mathbf{A}'$. Set $\mathbf{\Theta}_{(v)} := \mathbf{\Upsilon}^{-1}\mathbf{O}$. We verify the properties stated in the lemma:

$$\mathbf{\Theta}'_{(v)}\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\mathbf{\Theta}_{(v)} = \mathbf{O}'\mathbf{\Upsilon}^{-1}\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\mathbf{\Upsilon}^{-1}\mathbf{O} = \bar{T}\mathbf{O}'\mathbf{I}_K\mathbf{O} = \bar{T}\mathbf{I}_K,$$

$$\mathbf{A}\mathbf{\Theta}_{(v)} = \mathbf{A}\mathbf{\Upsilon}^{-1}\mathbf{O} = \bar{\mathbf{A}}'\mathbf{O}'\mathbf{O} = \bar{\mathbf{A}}' = \begin{pmatrix}\mathbf{A}^{\star\prime} & \mathbf{0}_{(h+1)\times(K-h-1)}\end{pmatrix},$$

$$\mathbf{A}^{\star\prime}\mathbf{A}^\star = \bar{\mathbf{A}}'\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Upsilon}^{-1}\mathbf{O}\mathbf{O}'\mathbf{\Upsilon}^{-1}\mathbf{A}' = \bar{T}\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)})^{-1}\mathbf{A}'.$$

Finally, since $\mathbf{O}$ is orthogonal ($\mathbf{OO}' = \mathbf{I}_K$) and $\mathbf{\Upsilon}^{-1}$ is symmetric,

$$\mathbf{\Theta}_{(v)}\mathbf{\Theta}'_{(v)} = \mathbf{\Upsilon}^{-1}\mathbf{OO}'\mathbf{\Upsilon}^{-1} = \mathbf{\Upsilon}^{-2} = \left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}\right)^{-1} = \bar{T}\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\right)^{-1}.$$

As $\mathbf{\Theta}_{(v)}$ is square, the eigenvalues of $\mathbf{\Theta}'_{(v)}\mathbf{\Theta}_{(v)}$ are exactly those of $\bar{T}\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\right)^{-1}$. □

The following lemma is a version of Lemma 7 from Mikusheva and Sølvsten (2025) with a slight change:

Lemma 2 (Stochastic order bounds)**.** *Suppose $\tilde{\mathbf{X}}$ is a $T\times(h+1)$ matrix with $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}/T = \mathbf{I}_{h+1}$, $\mathbf{B}$ is a $T\times T$ matrix, $\alpha_1,\dots,\alpha_h$ are non-random $(h+1)$-vectors, and $u_1,\dots,u_h$ are independent vectors of dimension T of independent mean-zero random variables with variances $0 < c \le \mathrm{Var}(u_{lt}) \le C < \infty$ and $\mathbb{E}[u_{lt}^4] < C^2 < \infty$ for all l,t. Then:*

(i) $\left\|T^{-1/2}\tilde{\mathbf{X}}'\mathbf{B}\sum_{l=1}^h u_l\alpha_l'\right\|_F = O_p(\max_l\|\alpha_l\|\|\mathbf{B}\|_2)$,

(ii) $u_1'\mathbf{B}u_1 - \mathbb{E}[u_1'\mathbf{B}u_1] = O_p(\|\mathbf{B}\|_F)$.

Proof. (i) Note that:

$$\begin{aligned}
\mathbb{E}\left[\left\|\frac{1}{\sqrt{T}}\tilde{\mathbf{X}}'\mathbf{B}\sum_{l=1}^h u_l\alpha_l'\right\|_F^2\right] &= \mathrm{tr}\left[\frac{\tilde{\mathbf{X}}\tilde{\mathbf{X}}'}{T}\mathbf{B}\sum_{l,j=1}^h\mathbb{E}\left[u_lu_j'\right]\alpha_l'\alpha_j\mathbf{B}'\right]\\
&\le \mathrm{tr}\left[\frac{\tilde{\mathbf{X}}\tilde{\mathbf{X}}'}{T}\right]\left\|\mathbf{B}\sum_{l=1}^h\mathbb{E}\left[u_lu_l'\right]\|\alpha_l\|^2\mathbf{B}'\right\|_2\\
&\le (h+1)\|\mathbf{B}\|_2^2\left(\sum_{l=1}^h\|\alpha_l\|^2\cdot\|\mathbb{E}\left[u_lu_l'\right]\|_2\right)\\
&\le C(h+1)h\max\|\alpha_l\|^2\|\mathbf{B}\|_2^2.
\end{aligned}$$

The first inequality follows from the trace-norm inequality (Lemma 34), and the last uses that the $\mathbb{E}[u_lu_{l'}]$ are diagonal matrices whose largest elements are at most C. Then Markov's inequality leads to the result.

(ii) Let $\sigma_t^2 := \mathrm{Var}\,(u_{1t})$. We have

$$\begin{aligned} u_1{}'\mathbf{B}u_1 - \mathbb{E}\,[u_1{}'\mathbf{B}u_1] &= \sum_t \sum_{s\neq t} \mathbf{B}_{ts} u_{1t} u_{1s} + \sum_t \mathbf{B}_{tt}(u_{1t}^2 - \sigma_t^2) \\ &= \sum_{t>s} (\mathbf{B}_{ts} + \mathbf{B}_{st}) u_{1s} u_{1t} + \sum_t \mathbf{B}_{tt}(u_{1t}^2 - \sigma_t^2). \end{aligned}$$

In the summations above, the summands are uncorrelated. Therefore,

$$\begin{aligned} \mathbb{E}\left[(u_1{}'\mathbf{B}u_1 - \mathbb{E}\,[u_1{}'\mathbf{B}u_1])^2\right] &= \sum_{t>s} (\mathbf{B}_{ts} + \mathbf{B}_{st})^2 \sigma_s^2 \sigma_t^2 + \sum_t \mathbf{B}_{tt}^2 \mathbb{E}\left[(u_{1t}^2 - \sigma_t^2)^2\right] \\ &\leq C^2 \left(\sum_{t>s} (\mathbf{B}_{ts} + \mathbf{B}_{st})^2 + \sum_t \mathbf{B}_{tt}^2 \right) \\ &\leq C^2 \frac{1}{2} \left\| \mathbf{B} + \mathbf{B}' \right\|_F^2 \\ &= O(\|\mathbf{B}\|_F^2). \end{aligned}$$

The result is then implied by Markov's inequality.

□

Lemma 3 (Eigenvalue lower bound for the partialled block)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime, *for every subsample* $v \in \{1, \ldots, h+1\}$,

$$\lambda_{\min}\Big(\frac{1}{\bar{T}} \mathbf{Z}_{1,(v)}{}' \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)}\Big) \;\geq\; 1 - o_p(1).$$

In particular, $\|(\mathbf{Z}_{1,(v)}{}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T})^{-1}\|_2 = O_p(1)$.

Proof. By (6) and Lemma 1,

$$\frac{1}{\bar{T}} \tilde{\mathbf{Z}}_{1,(v)}{}' \mathbf{M}_{(v)} \tilde{\mathbf{Z}}_{1,(v)} = \frac{1}{\bar{T}} \tilde{\mathbf{Z}}_{1,(v)}{}' \tilde{\mathbf{Z}}_{1,(v)} = \mathbf{I}_{h+1}. \tag{13}$$

Hence

$$\frac{1}{\bar{T}} \mathbf{Z}_{1,(v)}{}' \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)} = \mathbf{I}_{h+1} + \boldsymbol{\Delta}_{(v)}^{(1)} + \boldsymbol{\Delta}_{(v)}^{(2)} + \boldsymbol{\Delta}_{(v)}^{(3)}, \tag{14}$$

where

$$\boldsymbol{\Delta}^{(1)}_{(v)} := \frac{1}{\bar{T}}\tilde{\mathbf{Z}}_{1,(v)}{}'\mathbf{M}_{(v)}\sum_{l=1}^{h}\mathbf{S}_{(v-l)}\varepsilon\,\alpha'_{(v,l)} + \left(\frac{1}{\bar{T}}\tilde{\mathbf{Z}}_{1,(v)}{}'\mathbf{M}_{(v)}\sum_{l=1}^{h}\mathbf{S}_{(v-l)}\varepsilon\,\alpha'_{(v,l)}\right)',$$

$$\boldsymbol{\Delta}^{(2)}_{(v)} := \sum_{l=1}^{h}\alpha_{(v,l)}\alpha'_{(v,l)}\frac{1}{\bar{T}}\sum_{s=1}^{\bar{T}}\sigma^2_{v-l+(s-1)(h+1)}\mathbf{M}_{(v),ss},$$

$$\boldsymbol{\Delta}^{(3)}_{(v)} := \frac{1}{\bar{T}}\sum_{l,k=1}^{h}\alpha_{(v,l)}\alpha'_{(v,k)}\Big(\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\varepsilon - \mathbb{E}\left[\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\varepsilon\right]\Big).$$

The decomposition (14) separates the deterministic part $\mathbf{I}_{h+1} + \boldsymbol{\Delta}^{(2)}_{(v)}$ from the random perturbations $\boldsymbol{\Delta}^{(1)}_{(v)}$ and $\boldsymbol{\Delta}^{(3)}_{(v)}$. We apply Weyl's inequality (Lemma 36). Since all matrices in (14) are symmetric,

$$\lambda_{\min}\Big(\frac{1}{\bar{T}}\mathbf{Z}_{1,(v)}{}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\Big) \geq \lambda_{\min}(\mathbf{I}_{h+1} + \boldsymbol{\Delta}^{(2)}_{(v)}) + \lambda_{\min}(\boldsymbol{\Delta}^{(1)}_{(v)}) + \lambda_{\min}(\boldsymbol{\Delta}^{(3)}_{(v)}). \tag{15}$$

We bound the three terms on the right-hand side of (15) in turn, beginning with $\mathbf{I}_{h+1} + \boldsymbol{\Delta}^{(2)}_{(v)}$. The diagonal elements of $\mathbf{M}_{(v)}$ satisfy $\mathbf{M}_{(v),ss} \geq 0$ for all s (since $\mathbf{M}_{(v)}$ is an orthogonal projection), so each rank-one term $\alpha_{(v,l)}\alpha'_{(v,l)} \cdot \frac{1}{\bar{T}}\sum_s \sigma^2_{v-l+(s-1)(h+1)}\mathbf{M}_{(v),ss}$ is positive semidefinite. Therefore $\boldsymbol{\Delta}^{(2)}_{(v)} \succeq 0$ and

$$\lambda_{\min}(\mathbf{I}_{h+1} + \boldsymbol{\Delta}^{(2)}_{(v)}) \geq 1. \tag{16}$$

We next derive the rate of $\boldsymbol{\Delta}^{(1)}_{(v)}$. For any symmetric matrix $\mathbf{B}$, $\lambda_{\min}(\mathbf{B}) \geq -\|\mathbf{B}\|_2$. Applying Lemma 2 with $\widetilde{\mathbf{X}} \equiv \tilde{\mathbf{Z}}_{1,(v)}$ (satisfying $\tilde{\mathbf{Z}}_{1,(v)}{}'\tilde{\mathbf{Z}}_{1,(v)}/\bar{T} = \mathbf{I}_{h+1}$ by (13)), $\mathbf{B} \equiv \mathbf{M}_{(v)}$, and noting $\|\alpha_{(v,l)}\| \leq C$ under Assumption *Bounds*,

$$\|\boldsymbol{\Delta}^{(1)}_{(v)}\|_2 \leq 2\Big\|\frac{1}{\bar{T}}\tilde{\mathbf{Z}}_{1,(v)}{}'\mathbf{M}_{(v)}\sum_{l}\mathbf{S}_{(v-l)}\varepsilon\,\alpha'_{(v,l)}\Big\|_F = O_p\Big(\frac{\|\mathbf{M}_{(v)}\|_2}{\sqrt{\bar{T}}}\Big) = O_p\big(\bar{T}^{-1/2}\big), \tag{17}$$

so $\lambda_{\min}(\boldsymbol{\Delta}^{(1)}_{(v)}) \geq -\|\boldsymbol{\Delta}^{(1)}_{(v)}\|_2 = O_p(\bar{T}^{-1/2}) = o_p(1)$.

It remains to derive the rate of $\boldsymbol{\Delta}^{(3)}_{(v)}$. Again $\lambda_{\min}(\boldsymbol{\Delta}^{(3)}_{(v)}) \geq -\|\boldsymbol{\Delta}^{(3)}_{(v)}\|_2 \geq -\|\boldsymbol{\Delta}^{(3)}_{(v)}\|_F$. We bound the Frobenius norm term by term. Fix $l, k \in \{1, \ldots, h\}$ and set $a = \mathbf{S}_{(v-l)}\varepsilon$, $b = \mathbf{S}_{(v-k)}\varepsilon$. We bound the centred scalar $a'\mathbf{M}_{(v)}b - \mathbb{E}\left[a'\mathbf{M}_{(v)}b\right]$, distinguishing two cases.

If $l = k$, then $a = b$ and $a'\mathbf{M}_{(v)}a - \mathbb{E}\left[a'\mathbf{M}_{(v)}a\right]$ is a centred quadratic form in independent variables. By Lemma 2,

$$a'\mathbf{M}_{(v)}a - \mathbb{E}\left[a'\mathbf{M}_{(v)}a\right] = O_p(\|\mathbf{M}_{(v)}\|_F) = O_p(\bar{T}^{1/2}).$$

If $l \neq k$, then a and b are independent, so $\mathbb{E}\left[a'\mathbf{M}_{(v)}b\right] = 0$, and the summands $\mathbf{M}_{(v),ts}a_t b_s$ are pairwise uncorrelated. Hence

$$\operatorname{Var}\left(a'\mathbf{M}_{(v)}b\right) = \sum_{t,s}\mathbf{M}_{(v),ts}^2 \operatorname{Var}(a_t)\operatorname{Var}(b_s) \leq C^2\sum_{t,s}\mathbf{M}_{(v),ts}^2 = C^2\|\mathbf{M}_{(v)}\|_F^2,$$

so by Chebyshev's inequality $a'\mathbf{M}_{(v)}b = O_p(\|\mathbf{M}_{(v)}\|_F) = O_p(\bar{T}^{1/2})$.

In both cases the centred form is $O_p(\bar{T}^{1/2})$. With $\|\alpha_{(v,l)}\alpha'_{(v,k)}\|_F = \|\alpha_{(v,l)}\|\|\alpha_{(v,k)}\| \leq C^2$ and a fixed double sum over h^2 pairs,

$$\|\boldsymbol{\Delta}_{(v)}^{(3)}\|_F \leq \frac{1}{\bar{T}}\sum_{l,k=1}^{h}\|\alpha_{(v,l)}\alpha'_{(v,k)}\|_F \cdot O_p(\|\mathbf{M}_{(v)}\|_F) = \frac{h^2C^2\cdot O_p(\bar{T}^{1/2})}{\bar{T}} = O_p\big(\bar{T}^{-1/2}\big), \tag{18}$$

so $\lambda_{\min}(\boldsymbol{\Delta}_{(v)}^{(3)}) \geq -\|\boldsymbol{\Delta}_{(v)}^{(3)}\|_F = O_p(\bar{T}^{-1/2}) = o_p(1)$.

Finally, substituting (16), (17), and (18) into (15),

$$\lambda_{\min}\Big(\frac{1}{\bar{T}}\mathbf{Z}_{1,(v)}{}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\Big) \;\geq\; 1 - o_p(1).$$

For any fixed $\delta \in (0,1)$ it follows that $\mathbb{P}\left[\lambda_{\min}(\mathbf{Z}_{1,(v)}{}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}) \geq 1-\delta\right] \to 1$, so $(\mathbf{Z}_{1,(v)}{}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T})^{-1}$ exists with probability approaching 1 and

$$\|(\mathbf{Z}_{1,(v)}{}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T})^{-1}\|_2 \leq (1-\delta)^{-1} = O_p(1).$$

□

Lemma 4 (Eigenvalues of $\mathbf{Q}_{(v)}$)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime, *for each subsample* $v \in \{1,\ldots,h+1\}$, $\lambda_{\min}\left(\mathbf{Q}_{(v)}\right) = \Omega(T)$ *and* $\lambda_{\max}\left(\mathbf{Q}_{(v)}\right) = O(T)$.

Proof. Using

$$\mathbf{Q}_{(v)}/\bar{T} = \mathbf{I}_{h+1} + \sum_{l=1}^{h} \alpha_{(v,l)}\alpha'_{(v,l)} \frac{1}{\bar{T}} \sum_{t=1}^{\bar{T}} \sigma^2_{r(v-l,t)} \mathbf{M}_{(v),tt},$$

we use Weyl's inequality (Lemma 36) to establish that the smallest eigenvalue of $\mathbf{Q}_{(v)}/\bar{T}$ is bounded below:

$$\lambda_{\min}\left(\mathbf{Q}_{(v)}/\bar{T}\right) \geq \lambda_{\min}\left(\mathbf{I}_{h+1}\right) + \lambda_{\min}\left(\sum_{l=1}^{h} \alpha_{(v,l)}\alpha'_{(v,l)} \frac{1}{\bar{T}} \sum_{s=1}^{\bar{T}} \sigma^2_{v-l+(s-1)(h+1)} \mathbf{M}_{(v),ss}\right) \geq 1.$$

The second inequality follows from the fact that the diagonal elements of an orthogonal projection matrix are non-negative, which makes the second matrix positive semidefinite. To bound $\lambda_{\max}\left(\mathbf{Q}_{(v)}/\bar{T}\right)$ above, we use the triangle inequality:

$$\begin{aligned}
\lambda_{\max}\left(\mathbf{Q}_{(v)}/\bar{T}\right) \leq & \lambda_{\max}\left(\mathbf{I}_{h+1}\right) + \frac{1}{\bar{T}} \sum_{l=1}^{h} \|\alpha_{(v,l)}\|^2 \sum_{s=1}^{\bar{T}} \sigma^2_{v-l+(s-1)(h+1)} \mathbf{M}_{(v),ss} \\
\leq & 1 + \frac{C}{\bar{T}} \sum_{l=1}^{h} \|\alpha_{(v,l)}\|^2 (\bar{T} - K + h + 1) \\
\leq & 1 + C^3 h.
\end{aligned}$$

□

Lemma 5 (Convergence of the inverse partialled Gram matrix)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime,

$$\left\| \left(\mathbf{Z}'_{1,(v)} \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)}\right)^{-1} - \mathbf{Q}^{-1}_{(v)} \right\|_2 = O_p(\bar{T}^{-3/2}).$$

Proof. We will use the identity

$$\mathbf{N}_1^{-1} - \mathbf{N}_2^{-1} = \mathbf{N}_1^{-1}(\mathbf{N}_2 - \mathbf{N}_1)\mathbf{N}_2^{-1},$$

(where $\mathbf{N}_1$ and $\mathbf{N}_2$ are any two invertible matrices of the same dimension), which yields

$$\left\|\mathbf{N}_1^{-1} - \mathbf{N}_2^{-1}\right\|_2 \leq \left\|\mathbf{N}_1^{-1}\right\|_2 \left\|\mathbf{N}_2^{-1}\right\|_2 \left\|\mathbf{N}_2 - \mathbf{N}_1\right\|_2.$$

Applying this with $\mathbf{N}_1 = \mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}$ and $\mathbf{N}_2 = \mathbf{Q}_{(v)}/\bar{T}$:

$$\begin{aligned}
&\frac{1}{\bar{T}}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} - \mathbf{Q}_{(v)}\right) \\
=&\frac{1}{\bar{T}}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\sum_{l=1}^{h}\mathbf{S}_{(v-l)}\varepsilon\alpha'_{(v,l)} + \frac{1}{\bar{T}}\sum_{l=1}^{h}\alpha_{(v,l)}\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} \\
&+ \frac{1}{\bar{T}}\sum_{l,k=1}^{h}\alpha_{(v,l)}\alpha'_{(v,k)}\left(\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\varepsilon - \mathbb{E}\left[\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\varepsilon\right]\right).
\end{aligned}$$

Given Lemma 1, $\|\alpha_{(v,l)}\|^2 = \bar{T}\alpha'_l\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\right)^{-1}\alpha_l$, which is bounded for each v, l by Assumption *Bounds*. Therefore, applying Lemma 2 to these terms yields

$$\left\|\frac{1}{\bar{T}}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\sum_{l=1}^{h}\mathbf{S}_{(v-l)}\varepsilon\alpha'_{(v,l)}\right\|_F = O_p\left(\frac{\left\|\mathbf{M}_{(v)}\right\|_2}{\sqrt{\bar{T}}}\right) = O_p\left(\frac{1}{\sqrt{\bar{T}}}\right),$$

and

$$\begin{aligned}
\frac{1}{\bar{T}}\left(\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\varepsilon - \mathbb{E}\left[\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\varepsilon\right]\right) =& O_p\left(\frac{1}{\bar{T}}\left\|\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\right\|_F\right) \\
=& O_p\left(\frac{1}{\bar{T}}\left\|\mathbf{M}_{(v)}\right\|_F\right) = O_p\left(\frac{1}{\sqrt{\bar{T}}}\right).
\end{aligned}$$

Here we used that

$$\left\|\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}\right\|_F = \left\|\mathbf{M}_{(v)}\right\|_F$$

(each non-zero element of $\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v-k)}$ is a distinct entry of $\mathbf{M}_{(v)}$) and $\left\|\mathbf{M}_{(v)}\right\|_F = \sqrt{\bar{T}-K+h+1}$. Since $\left\|\alpha_{(v,l)}\alpha'_{(v,k)}\right\|_F = \sqrt{\mathrm{tr}\left(\alpha_{(v,l)}\alpha'_{(v,k)}\alpha_{(v,k)}\alpha'_{(v,l)}\right)} = \|\alpha_{(v,l)}\|\|\alpha_{(v,k)}\|$ is bounded, it follows that

$$\left\|\frac{1}{\bar{T}}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} - \mathbf{Q}_{(v)}\right)\right\|_2 \leq \left\|\frac{1}{\bar{T}}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} - \mathbf{Q}_{(v)}\right)\right\|_F = O_p(\bar{T}^{-\frac{1}{2}}).$$

We know from Lemma 4 that $\lambda_{\min}\left(\mathbf{Q}_{(v)}/\bar{T}\right)$ is bounded below by a positive constant, which implies that $\left\|(\mathbf{Q}_{(v)}/\bar{T})^{-1}\right\|_2 = \lambda_{\max}\left((\mathbf{Q}_{(v)}/\bar{T})^{-1}\right)$ is bounded above. Regarding $\left\|\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}\right)^{-1}\right\|_2$, Lemma 3 implies that it is bounded by any constant larger

than one with probability approaching 1, so it is $O_p(1)$. Combining these bounds with the inequality above gives $\left\|(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T})^{-1} - (\mathbf{Q}_{(v)}/\bar{T})^{-1}\right\|_2 = O_p(\bar{T}^{-1/2})$, which is the claim after rescaling by $\bar{T}^{-1}$. □

Lemma 6 (Linearisation)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime,

$$(h+1)\, r'(\hat{\beta} - \beta) = U + O_p(T^{-1}).$$

Proof. Applying the transformation $\mathbf{\Theta}_{(v)}$ to $\mathbf{X}_{(v)}$ yields:

$$\mathbf{Z}_{1,(v)} = \tilde{\mathbf{X}}_{(v)}\mathbf{\Theta}_{(v),1} + \sum_{l=1}^{h} \mathbf{S}_{(v-l)}\varepsilon\alpha'_{(v,l)},$$

and $\mathbf{Z}_{2,(v)} = \tilde{\mathbf{X}}_{(v)}\mathbf{\Theta}_{(v),2}$ is strictly exogenous as $\alpha'_l\mathbf{\Theta}_{(v),2} = 0$ for all l by Lemma 1.

$$\begin{aligned}
(h+1)r'(\hat{\beta} - \beta) &= \sum_{v=1}^{h+1} r'(\mathbf{X}'_{(v)}\mathbf{X}_{(v)})^{-1}\mathbf{X}'_{(v)}\mathbf{S}_{(v)}\varepsilon \\
&= \sum_{v=1}^{h+1} r'\mathbf{\Theta}_{(v)}(\mathbf{Z}'_{(v)}\mathbf{Z}_{(v)})^{-1}\mathbf{Z}'_{(v)}\mathbf{S}_{(v)}\varepsilon \\
&= \sum_{v=1}^{h+1} r'_{(v)} \left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1} \mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon.
\end{aligned}$$

The last equality uses the Frisch–Waugh–Lovell theorem. The matrix $\mathbf{Q}_{(v)}$ satisfies

$$\begin{aligned}
\mathbf{Q}_{(v)} &= \mathbb{E}\left[\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right] \\
&= \left(\tilde{\mathbf{Z}}'_{1,(v)}\tilde{\mathbf{Z}}_{1,(v)} + \sum_{l=1}^{h} \alpha_{(v,l)}\alpha'_{(v,l)} \sum_{s=1}^{\bar{T}} \sigma^2_{v-l+(s-1)(h+1)}\mathbf{M}_{(v)_{s,s}}\right) \\
&= \left(\bar{T}\mathbf{I}_{h+1} + \sum_{l=1}^{h} \alpha_{(v,l)}\alpha'_{(v,l)} \sum_{s=1}^{\bar{T}} \sigma^2_{v-l+(s-1)(h+1)}\mathbf{M}_{(v)_{s,s}}\right).
\end{aligned}$$

We can decompose the quantity of interest as follows:

$$
\begin{aligned}
(h+1)r'(\hat{\beta}-\beta) = &\sum_{v=1}^{h+1} r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon + \\
&\sum_{v=1}^{h+1} r'_{(v)}\Delta\frac{1}{\bar{T}}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon,
\end{aligned} \tag{19}
$$

where $\Delta := \left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}\right)^{-1} - \left(\mathbf{Q}_{(v)}/\bar{T}\right)^{-1}$. The first term is equal to U :

$$
\begin{aligned}
&\sum_{v=1}^{h+1} r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon \\
=&\sum_{v=1}^{h+1} r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon + \sum_{v=1}^{h+1} r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\sum_{l=1}^{h}\alpha_{(v,l)}\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon \\
=&\sum_{v=1}^{h+1} r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon + \varepsilon'\sum_{v=1}^{h+1}\sum_{l=1}^{h}\gamma_{v,l}\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon \\
=&\left(\sum_{v=1}^{h+1} r'_{(v)}{\mathbf{Q}_{(v)}}^{-1}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}\right)\mathbf{\Phi}^{-1}\varepsilon + \varepsilon'\mathbf{\Phi}^{-1}\left(\mathbf{\Phi}\sum_{v=1}^{h+1}\sum_{l=1}^{h}\gamma_{v,l}\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}\right)\mathbf{\Phi}^{-1}\varepsilon \\
=&\omega'\mathbf{\Phi}^{-1}\varepsilon + \varepsilon'\mathbf{\Phi}^{-1}\mathbf{B}\mathbf{\Phi}^{-1}\varepsilon \\
=&U.
\end{aligned}
$$

It remains to show that the second term of (19) is $O_p(\bar{T}^{-1})$. We know from Lemma 5 that $\|\Delta\|_2 = O_p(\bar{T}^{-1/2})$.

It remains to bound

$$
\frac{1}{\bar{T}}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon = \frac{1}{\bar{T}}\left(\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon + \sum_{l=1}^{h}\alpha_{(v,l)}\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right).
$$

Lemma 2 can be applied to these terms. Part (i), with $\alpha_1 = (1,0,0,\dots)'$ and $\alpha_2 = \cdots = \alpha_h = 0$, can be applied to $\frac{1}{\bar{T}}\left(\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)$ to show that

$$
\left\|\frac{1}{\bar{T}}\left(\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)\right\| = O_p\left(\frac{1}{\sqrt{\bar{T}}}\left\|\mathbf{M}_{(v)}\right\|_2\right) = O_p\left(\frac{1}{\sqrt{\bar{T}}}\right).
$$

Part (ii) can be applied to each term $\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon$ given that its expectation is equal to

zero. Together with the fact that the $\|\alpha_{(v,l)}\|$ are bounded, we obtain:

$$\left\|\frac{1}{\bar{T}}\alpha_{(v,l)}\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right\| = O_p\left(\frac{1}{\bar{T}}\left\|\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\right\|_F\right) = O_p\left(\frac{1}{\sqrt{\bar{T}}}\right).$$

In conclusion, since $\|r_{(v)}\|$ is bounded, the second term of (19) is $O_p(1/\bar{T})$. Multiplying by $h+1$ and using $T=(h+1)\bar{T}$ with h fixed proves the lemma. □

Lemma 7 (The overall and the subsample variances)**.** *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*, and* Proportional Regime*,*

(i) (subsample variances) $\Sigma_v = \|\omega_{(v)}\|^2 + \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(v)}\right)$ *for each* v*, so that* $\sum_{v=1}^{h+1}\Sigma_v = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}')$*, where*

$$\operatorname{tr}(\mathbf{B}\mathbf{B}') = \sum_{v=1}^{h+1}\sum_{l=1}^{h}\gamma_{v,l}^2\ \operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v-l)}\right);$$

(ii) (overall variance)

$$\Sigma = \|\omega\|^2 + \tfrac{1}{2}\|\mathbf{B}+\mathbf{B}'\|_F^2 = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}') + \operatorname{tr}\left(\mathbf{B}^2\right) = \sum_{v=1}^{h+1}\Sigma_v + \operatorname{tr}\left(\mathbf{B}^2\right),$$

where $\operatorname{tr}(\mathbf{B}^2) = \sum_{v\neq q}\operatorname{Cov}(U_v, U_q)$ *is the sum of the covariances between distinct subsamples and*

$$\operatorname{tr}\left(\mathbf{B}^2\right) = 2\sum_{1\leq v<q\leq h+1}\gamma_{v,h+1-(q-v)}\,\gamma_{q,q-v}\ \operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(q)}\mathbf{\Phi}^2_{(q)}\mathbf{J}\right),$$

with $\mathbf{J}$ *the shift matrix of* (3)*;*

(iii) (order) $\sum_{v=1}^{h+1}\Sigma_v \asymp T^{-1}$ *and* $\Sigma \asymp T^{-1}$*.*

Proof. Throughout, write $\tilde{\varepsilon} = \mathbf{\Phi}^{-1}\varepsilon$, so that $U_v = \omega'_{(v)}\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}_{(v)}\tilde{\varepsilon}$ and $U = \omega'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}\tilde{\varepsilon}$. By Assumption *Regular Errors*, the entries of $\tilde{\varepsilon}$ are i.i.d. with mean zero and unit variance.

We first derive a general covariance formula. Let a, b be deterministic vectors of the same dimension and $\mathbf{F}, \mathbf{G}$ deterministic square matrices with zero diagonal whose dimension equals the length of a and b. The linear and quadratic parts of $a'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{F}\tilde{\varepsilon}$ and $b'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{G}\tilde{\varepsilon}$

are uncorrelated, because $\mathbb{E}\left[\tilde{\varepsilon}_k\tilde{\varepsilon}_i\tilde{\varepsilon}_j\right] = 0$ whenever $i \neq j$. Moreover, for $i \neq j$ and $k \neq l$, the moment $\mathbb{E}\left[\tilde{\varepsilon}_i\tilde{\varepsilon}_j\tilde{\varepsilon}_k\tilde{\varepsilon}_l\right]$ is non-zero (and then equal to one) only if $(k,l) = (i,j)$ or $(k,l) = (j,i)$. Hence

$$\begin{aligned}\operatorname{Cov}\left(a'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{F}\tilde{\varepsilon},\, b'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{G}\tilde{\varepsilon}\right) &= a'b + \sum_{i\neq j}\mathbf{F}_{ij}\mathbf{G}_{ij} + \sum_{i\neq j}\mathbf{F}_{ij}\mathbf{G}_{ji} \\ &= a'b + \operatorname{tr}\left(\mathbf{F}\mathbf{G}'\right) + \operatorname{tr}\left(\mathbf{F}\mathbf{G}\right).\end{aligned} \tag{20}$$

All products required below are computed from the identity (5) for $\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(q)}$. Recall that, for $v \in \{1,\ldots,h+1\}$,

$$\begin{aligned}\mathbf{B}_{(v)} &= \mathbf{\Phi}\sum_{l=1}^{h}\gamma_{v,l}\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}, \qquad \mathbf{B}'_{(v)} = \mathbf{\Phi}\sum_{l=1}^{h}\gamma_{v,l}\mathbf{S}'_{(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v-l)}\mathbf{\Phi}, \\ \omega'_{(v)} &= r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi},\end{aligned}$$

and that, for $v, q \in \{1,\ldots,h+1\}$ and $l, m \in \{1,\ldots,h\}$, the indices $v-l$ and $q-m$ lie in $\{1-h,\ldots,h\}$.

The matrices $\mathbf{B}_{(v)}$ have zero diagonal. Indeed, for any $(T+h)$-vector b, cyclicity of the trace gives

$$\begin{aligned}\operatorname{tr}\left(\operatorname{diag}(b)\mathbf{B}_{(v)}\right) &= \sum_{l=1}^{h}\gamma_{v,l}\operatorname{tr}\left(\mathbf{M}_{(v)}\left(\mathbf{S}_{(v)}\mathbf{\Phi}\operatorname{diag}(b)\,\mathbf{\Phi}\mathbf{S}'_{(v-l)}\right)\right) \\ &= \sum_{l=1}^{h}\gamma_{v,l}\operatorname{tr}\left(\mathbf{M}_{(v)}\left(\mathbf{S}_{(v)}\operatorname{diag}(\mathbf{\Phi}^2 b)\mathbf{S}'_{(v-l)}\right)\right) = 0,\end{aligned}$$

because $\mathbf{S}_{(v)}\operatorname{diag}(\mathbf{\Phi}^2 b)\mathbf{S}'_{(v-l)} = \mathbf{0}_{\bar{T}\times\bar{T}}$ by (3), as $v-l \neq v$ and $v-l \neq v-(h+1)$ for $1 \leq l \leq h$. Choosing $b = e_i$ yields $(\mathbf{B}_{(v)})_{ii} = 0$ for all i, so that $\mathbf{B}_{(v)}$, and therefore $\mathbf{B}$, has zero diagonal.

Next, consider the products of the linear coefficients,

$$\omega'_{(v)}\omega_{(q)} = r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\left(\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(q)}\right)\mathbf{M}_{(q)}\tilde{\mathbf{Z}}_{1,(q)}\mathbf{Q}^{-1}_{(q)}r_{(q)}.$$

For $v \neq q$, $|v-q| \leq h$, so that $\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(q)} = \mathbf{0}_{\bar{T}\times\bar{T}}$ by (5) and $\omega'_{(v)}\omega_{(q)} = 0$. Hence $\|\omega\|^2 = \sum_v \|\omega_{(v)}\|^2$.

For the products $\mathbf{B}_{(v)}\mathbf{B}'_{(q)}$,

$$\mathbf{B}_{(v)}\mathbf{B}'_{(q)} = \sum_{l=1}^{h}\sum_{m=1}^{h} \gamma_{v,l}\gamma_{q,m}\, \mathbf{\Phi}\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)} \left(\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(q)}\right) \mathbf{M}_{(q)}\mathbf{S}_{(q-m)}\mathbf{\Phi}.$$

For $v \neq q$ the bracket vanishes by (5), so that $\mathbf{B}_{(v)}\mathbf{B}'_{(q)} = \mathbf{0}$. For $v = q$ the bracket equals $\mathbf{\Phi}^2_{(v)}$ and, by cyclicity of the trace,

$$\begin{aligned}\operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(v)}\right) &= \sum_{l=1}^{h}\sum_{m=1}^{h} \gamma_{v,l}\gamma_{v,m} \operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)} \left(\mathbf{S}_{(v-m)}\mathbf{\Phi}^2\mathbf{S}'_{(v-l)}\right)\right) \\ &= \sum_{l=1}^{h} \gamma_{v,l}^2 \operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v-l)}\right),\end{aligned}$$

since for $l \neq m$ we have $0 < |l-m| \leq h-1$, so that $\mathbf{S}_{(v-m)}\mathbf{\Phi}^2\mathbf{S}'_{(v-l)} = \mathbf{0}_{\bar{T}\times\bar{T}}$, while for $l = m$ it equals $\mathbf{\Phi}^2_{(v-l)}$. Consequently $\operatorname{tr}(\mathbf{B}\mathbf{B}') = \sum_{v,q}\operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(q)}\right) = \sum_{v}\operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(v)}\right)$, which is the stated expression.

Finally, for the products $\mathbf{B}_{(v)}\mathbf{B}_{(q)}$,

$$\operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}_{(q)}\right) = \sum_{l=1}^{h}\sum_{m=1}^{h} \gamma_{v,l}\gamma_{q,m} \operatorname{tr}\left(\mathbf{M}_{(v)} \left(\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(q-m)}\right) \mathbf{M}_{(q)} \left(\mathbf{S}_{(q)}\mathbf{\Phi}^2\mathbf{S}'_{(v-l)}\right)\right).$$

For $q = v$ the first bracket is $\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(v-m)} = \mathbf{0}_{\bar{T}\times\bar{T}}$ for every m (as $1 \leq m \leq h$), so that $\operatorname{tr}\left(\mathbf{B}^2_{(v)}\right) = 0$. For $v < q$, the first bracket is non-zero only if $q - m = v$ (the alternative $q - m = v - (h+1)$ would require $m = q - v + h + 1 > h$), that is, $m = q - v$, and it then equals $\mathbf{\Phi}^2_{(v)}$. The second bracket is non-zero only if $v - l = q - (h+1)$ (the alternative $v - l = q$ is impossible as $v < q$), that is, $l = h+1-(q-v)$, and it then equals $\mathbf{\Phi}^2_{(q)}\mathbf{J}$. Hence, for $v < q$,

$$\operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}_{(q)}\right) = \gamma_{v,h+1-(q-v)}\, \gamma_{q,q-v}\, \operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(q)}\mathbf{\Phi}^2_{(q)}\mathbf{J}\right),$$

and $\operatorname{tr}\left(\mathbf{B}_{(q)}\mathbf{B}_{(v)}\right) = \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}_{(q)}\right)$ by cyclicity of the trace.

(i) Applying (20) to U_v and using $\operatorname{tr}\left(\mathbf{B}^2_{(v)}\right) = 0$ gives

$$\Sigma_v = \|\omega_{(v)}\|^2 + \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(v)}\right) + \operatorname{tr}\left(\mathbf{B}^2_{(v)}\right) = \|\omega_{(v)}\|^2 + \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(v)}\right).$$

Summing over v and using the results for $\omega'_{(v)}\omega_{(q)}$ and $\mathbf{B}_{(v)}\mathbf{B}'_{(q)}$ yields $\sum_{v=1}^{h+1}\Sigma_v = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}')$.

(ii) Applying (20) with $a = b = \omega$ and $\mathbf{F} = \mathbf{G} = \mathbf{B}$ gives

$$\Sigma = \operatorname{Var}(U) = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}') + \operatorname{tr}\left(\mathbf{B}^2\right) = \|\omega\|^2 + \frac{1}{2}\|\mathbf{B} + \mathbf{B}'\|_F^2 .$$

For $v \neq q$, (20) gives $\operatorname{Cov}(U_v, U_q) = \omega'_{(v)}\omega_{(q)} + \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}'_{(q)}\right) + \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}_{(q)}\right) = \operatorname{tr}\left(\mathbf{B}_{(v)}\mathbf{B}_{(q)}\right)$. Summing over all pairs, $\sum_{v\neq q}\operatorname{Cov}(U_v, U_q) = \operatorname{tr}(\mathbf{B}^2) - \sum_v \operatorname{tr}\left(\mathbf{B}^2_{(v)}\right) = \operatorname{tr}(\mathbf{B}^2)$, which has the stated closed form by the result for $\mathbf{B}_{(v)}\mathbf{B}_{(q)}$. Together with (i), $\Sigma = \sum_{v=1}^{h+1}\Sigma_v + \operatorname{tr}(\mathbf{B}^2)$.

(iii) Consider first

$$\begin{aligned}
\|\omega\|^2 &= \sum_{v=1}^{h+1}\|\omega_{(v)}\|^2 \\
&= \sum_{v=1}^{h+1} r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(v)}\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\mathbf{Q}^{-1}_{(v)}r_{(v)} \\
&= \sum_{v=1}^{h+1} r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\mathbf{Q}^{-1}_{(v)}r_{(v)}.
\end{aligned}$$

The matrix $\mathbf{M}_{(v)}$ can be omitted by (6). This expression is bounded as follows:

$$\begin{aligned}
\sum_v r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\mathbf{Q}^{-1}_{(v)}r_{(v)} &= \frac{1}{\bar{T}}\sum_v r'_{(v)}\bar{T}\mathbf{Q}^{-1}_{(v)}\frac{1}{\bar{T}}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\bar{T}\mathbf{Q}^{-1}_{(v)}r_{(v)} \\
&\leq \frac{1}{\bar{T}}\sum_v \|r_{(v)}\|^2\lambda_{\max}\left(\bar{T}\mathbf{Q}^{-1}_{(v)}\right)^2\lambda_{\max}\left(\frac{1}{\bar{T}}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\right).
\end{aligned}$$

We know from Lemma 4 that $\lambda_{\max}\left(\bar{T}\mathbf{Q}^{-1}_{(v)}\right)$ is bounded. As $\tilde{\mathbf{Z}}_{1,(v)}/\sqrt{\bar{T}}$ has orthonormal columns, Lemma 37 gives

$$\lambda_{\max}\left(\frac{1}{\bar{T}}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\right) \leq \lambda_{\max}\left(\mathbf{\Phi}^2_{(v)}\right) < C^2.$$

Finally, from Lemma 1, $\|r_{(v)}\| \leq C^2$. In conclusion, $\|\omega\|^2 = O(T^{-1})$.

For the remaining terms, $\frac{1}{2}\|\mathbf{B} + \mathbf{B}'\|_F^2 = \operatorname{tr}(\mathbf{B}\mathbf{B}') + \operatorname{tr}(\mathbf{B}^2)$. Using the Cauchy–Schwarz inequality for the trace inner product ($|\operatorname{tr}(\mathbf{F}'\mathbf{G})| \leq \|\mathbf{F}\|_F\|\mathbf{G}\|_F$, applied with $\mathbf{F} = \mathbf{B}'$ and

$\mathbf{G} = \mathbf{B}$):

$$\begin{aligned} |\operatorname{tr}\left(\mathbf{B}^2\right)| &= |\operatorname{tr}\left((\mathbf{B}')'\mathbf{B}\right)| \leq \|\mathbf{B}'\|_F \|\mathbf{B}\|_F \\ &= \|\mathbf{B}\|_F^2 = \operatorname{tr}\left(\mathbf{B}\mathbf{B}'\right), \end{aligned} \tag{21}$$

where we used that the Frobenius norm is invariant under transposition and that $\operatorname{tr}(\mathbf{B}\mathbf{B}') = \|\mathbf{B}\|_F^2$. Hence

$$\tfrac{1}{2}\|\mathbf{B}+\mathbf{B}'\|_F^2 = \operatorname{tr}(\mathbf{B}\mathbf{B}') + \operatorname{tr}\left(\mathbf{B}^2\right) \leq \operatorname{tr}(\mathbf{B}\mathbf{B}') + |\operatorname{tr}\left(\mathbf{B}^2\right)| \overset{(21)}{\leq} 2\operatorname{tr}(\mathbf{B}\mathbf{B}').$$

It remains to bound $\operatorname{tr}(\mathbf{B}\mathbf{B}')$ itself. By (i),

$$\operatorname{tr}(\mathbf{B}\mathbf{B}') = \sum_{v=1}^{h+1}\sum_{l=1}^{h} \gamma_{v,l}^2 \operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}_{(v)}^2\mathbf{M}_{(v)}\mathbf{\Phi}_{(v-l)}^2\right)$$

where $\gamma_{v,l} = r_{(v)}'\mathbf{Q}_{(v)}^{-1}\alpha_{(v,l)} = O(T^{-1})$. On the other hand, iterated application of the trace-norm inequality (Lemma 34) shows that

$$|\operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{\Phi}_{(v)}^2\mathbf{M}_{(v)}\mathbf{\Phi}_{(v-l)}^2\right)| \leq C^2 \operatorname{tr}\left(\mathbf{M}_{(v)}\right) = C^2(\bar{T} - K + h + 1).$$

Hence $\operatorname{tr}(\mathbf{B}\mathbf{B}') = O(T^{-2})\cdot O(T) = O(T^{-1})$, and therefore $\sum_{v=1}^{h+1}\Sigma_v = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}') = O(T^{-1})$ and $\Sigma \leq \|\omega\|^2 + 2\operatorname{tr}(\mathbf{B}\mathbf{B}') = O(T^{-1})$.

For the matching lower bounds, note that the quadratic contributions in (i) and (ii) are both non-negative, so that

$$\sum_{v=1}^{h+1}\Sigma_v = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}') \geq \|\omega\|^2 \qquad \text{and} \qquad \Sigma = \|\omega\|^2 + \tfrac{1}{2}\|\mathbf{B}+\mathbf{B}'\|_F^2 \geq \|\omega\|^2,$$

and it suffices to bound $\|\omega\|^2$ from below. In the expression for $\|\omega\|^2$ obtained at the beginning of the proof of (iii), every summand is non-negative, so that, for each v,

$$\begin{aligned} \|\omega\|^2 &= \sum_{q=1}^{h+1} r_{(q)}'\mathbf{Q}_{(q)}^{-1}\tilde{\mathbf{Z}}_{1,(q)}'\mathbf{\Phi}_{(q)}^2\tilde{\mathbf{Z}}_{1,(q)}\mathbf{Q}_{(q)}^{-1}r_{(q)} \\ &\geq \|r_{(v)}\|^2 \lambda_{\max}\left(\mathbf{Q}_{(v)}\right)^{-2} \lambda_{\min}\left(\tilde{\mathbf{Z}}_{1,(v)}'\tilde{\mathbf{Z}}_{1,(v)}\right) \lambda_{\min}\left(\mathbf{\Phi}_{(v)}^2\right) = \Omega(T^{-1})\|r_{(v)}\|^2, \end{aligned} \tag{22}$$

using Lemma 37, $\lambda_{\min}\left(\mathbf{\Phi}^2_{(v)}\right) > c$ (Assumption *Regular Errors*), $\tilde{\mathbf{Z}}'_{1,(v)}\tilde{\mathbf{Z}}_{1,(v)} = \bar{T}\mathbf{I}_{h+1}$ (Lemma 1) and $\lambda_{\max}\left(\mathbf{Q}_{(v)}\right) = O(T)$ (Lemma 4).

Finally, $\|r_{(v)}\|^2 = r'(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}r$ according to Lemma 1. Assumption *Bounds* then gives $\|\omega\|^2 = \Omega(T^{-1})$. Together with the upper bounds, $\sum_{v=1}^{h+1}\Sigma_v \asymp T^{-1}$ and $\Sigma \asymp T^{-1}$. □

A.2 Consistency and asymptotic normality

Theorem 1 (Consistency)**.** *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*, and* Proportional Regime*,*

$$r'(\hat{\beta} - \beta) \xrightarrow{p} 0.$$

Proof. The result follows by combining Lemmas 6 and 7 and Chebyshev's inequality. □

Lemma 8 (Corollary A2.8 of Sølvsten, 2020)**.** *Let $v_1, \ldots, v_n$ be i.i.d. vectors, $v_i = (v'_{i1}, v_{i2}, v_{i3})'$, and let $\mathbf{C} = \mathrm{Var}\,(v_{11})$. Let w_{ni} be a sequence of deterministic vectors of the same dimension as v_{i1}. Let $\mathbf{K}$ be a sequence of $n \times n$ symmetric matrices and $\bar{\mathbf{K}} = \mathbf{K}^2$ (the dependence on n is suppressed). Let*

$$W_n = \sum_i w'_{ni}v_{i1} \; + \; \sum_i \sum_{j\neq i} \mathbf{K}_{ij}v_{i2}v_{j3}.$$

If the following conditions hold:

(i)

$$\sigma_n^2 = \sum_i w'_{ni}\mathbf{C}w_{ni} + \sum_i \sum_{j\neq i} \mathbf{K}_{ij}^2 \left(\mathbb{E}\left[v_{12}^2\right]\mathbb{E}\left[v_{13}^2\right] + \mathbb{E}\left[v_{12}v_{13}\right]^2\right) > c > 0;$$

(ii)

$$\mathbb{E}\left[v_{11}v'_{11}\right] = \mathbf{C}, \quad \sum_i w'_{ni}\mathbf{C}w_{ni} \leq 1, \quad \max_i \|w_{ni}\|^2 \to 0;$$

(iii)

$$\mathbb{E}\left[v_{12}^4 + v_{13}^4\right] < \infty, \quad \sum_i \sum_{j\neq i} \mathbf{K}_{ij}^2 \leq 1, \quad \max_i \sum_{j\neq i} \mathbf{K}_{ij}^2 \to 0, \quad \sum_i \sum_{j\neq i} \bar{\mathbf{K}}_{ij}^2 \to 0;$$

(iv)

$$\mathbb{E}\left[\|v_{11}\|^2(v_{12}^2+v_{13}^2)\right]<\infty;$$

then

$$\frac{W_n}{\sigma_n}\ \xrightarrow{d}\ N(0,1).$$

Lemma 9 (Condition (iii) holds in the proportional regime)**.** *Let*

$$\mathbf{B}^{\mathrm{s}}:=\tfrac{1}{2}(\mathbf{B}+\mathbf{B}'),\qquad \mathcal{B}:=\mathbf{B}^{\mathrm{s}}/\sqrt{\Sigma}.$$

Under Assumptions Feedback Decomposition*,* Bounds*,* Regular Errors *and* Proportional Regime*,*

$$\|\mathcal{B}\|_2^2=O(\bar{T}^{-1}),\max_i\sum_{j\neq i}\mathcal{B}_{ij}^2=O(\bar{T}^{-1}),\sum_i\sum_{j\neq i}(\mathcal{B}^2)_{ij}^2=O(\bar{T}^{-1}).$$

Moreover, let $\bar{\mathcal{B}}:=\sqrt{2}\mathbf{B}^{\mathrm{s}}/\left\|\mathbf{B}^{\mathrm{s}}\right\|_F$*. If, in addition,* $\left\|\mathbf{B}+\mathbf{B}'\right\|_F^2=\Omega(\|\omega\|^2)$*, then*

$$\left\|\bar{\mathcal{B}}\right\|_2^2=O(\bar{T}^{-1}),\max_i\sum_{j\neq i}\bar{\mathcal{B}}_{ij}^2=O(\bar{T}^{-1}),\sum_i\sum_{j\neq i}(\bar{\mathcal{B}}^2)_{ij}^2=O(\bar{T}^{-1}).$$

Proof. Write $\mathbf{B}^{\mathrm{s}}=\frac{1}{2}(\mathbf{B}+\mathbf{B}')$, which is symmetric. We show that $\|\mathcal{B}\|_2^2=\left\|\mathbf{B}^{\mathrm{s}}\right\|_2^2/\Sigma=O(\bar{T}^{-1})$. The three conditions then follow from the argument given at the end of this proof. Throughout, we use $\Sigma\geq\|\omega\|^2$ and set $\Pi_v:=\|\mathbf{Q}_{(v)}^{-1}r_{(v)}\|^2$.

We first bound the numerator $\left\|\mathbf{B}^{\mathrm{s}}\right\|_2^2$ by a constant multiple of $\sum_v\Pi_v$. By the triangle inequality and submultiplicativity of the spectral norm, and using that $\left\|\mathbf{M}_{(v)}\right\|_2=1$, $\left\|\mathbf{S}_{(v)}\right\|_2=\left\|\mathbf{S}'_{(v-l)}\right\|_2=1$ and $\|\mathbf{\Phi}\|_2^2\leq C$,

$$\left\|\mathbf{B}^{\mathrm{s}}\right\|_2\leq\|\mathbf{B}\|_2\leq C\sum_{v=1}^{h+1}\sum_{l=1}^{h}|\gamma_{v,l}|.$$

Each $|\gamma_{v,l}|=|r'_{(v)}\mathbf{Q}_{(v)}^{-1}\alpha_{(v,l)}|\leq\|\mathbf{Q}_{(v)}^{-1}r_{(v)}\|\,\|\alpha_{(v,l)}\|\leq C\,\|\mathbf{Q}_{(v)}^{-1}r_{(v)}\|=C\,\Pi_v^{1/2}$, using $\|\alpha_{(v,l)}\|^2=\bar{T}\alpha_l'(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)})^{-1}\alpha_l\leq\|\alpha_l\|^2/c\leq C^2/c$ (Lemma 1, Assumption *Bounds*). Hence, by Cauchy–

Schwarz,

$$\|\mathbf{B}^{\mathrm{s}}\|_2^2 \leq C^2\Big(\sum_{v,l}|\gamma_{v,l}|\Big)^2 \leq C^2 h(h+1)\sum_{v,l}\gamma_{v,l}^2 \leq C^2 h^2(h+1)\,C^4/c\sum_v \Pi_v = O\left(\sum_v \Pi_v\right).$$

For the denominator we show that $\|\omega\|^2 \geq c\,\bar{T}\sum_v \Pi_v$:

$$\begin{aligned}
\|\omega\|^2 =& \sum_v r'_{(v)}\mathbf{Q}_{(v)}^{-1}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\mathbf{Q}_{(v)}^{-1}r_{(v)} \\
\geq& c\sum_v r'_{(v)}\mathbf{Q}_{(v)}^{-1}\tilde{\mathbf{Z}}'_{1,(v)}\tilde{\mathbf{Z}}_{1,(v)}\mathbf{Q}_{(v)}^{-1}r_{(v)} \\
=& c\,\bar{T}\sum_v r'_{(v)}\mathbf{Q}_{(v)}^{-2}r_{(v)} \\
=& c\,\bar{T}\sum_v \Pi_v.
\end{aligned}$$

Combining the two bounds with $\Sigma \geq \|\omega\|^2$, and since $\sum_v \Pi_v > 0$,

$$\|\mathcal{B}\|_2^2 = \frac{\|\mathbf{B}^{\mathrm{s}}\|_2^2}{\Sigma} \leq \frac{\|\mathbf{B}^{\mathrm{s}}\|_2^2}{\|\omega\|^2} \leq \frac{O\left(\sum_v \Pi_v\right)}{\Omega(\bar{T}\sum_v \Pi_v)} = O(\bar{T}^{-1}). \tag{23}$$

It remains to deduce the other two conditions. For the maximal row sum, symmetry of $\mathcal{B}$ gives $\max_i \sum_{j\neq i}\mathcal{B}_{ij}^2 \leq \max_i \|\mathcal{B}_{i\cdot}\|^2 \leq \|\mathcal{B}\|_2^2 = O(\bar{T}^{-1}) \to 0$ by (23). For the condition on $\mathcal{B}^2$, write $\lambda_k(\mathcal{B})$ for the eigenvalues of the symmetric matrix $\mathcal{B}$. Then $\sum_i\sum_{j\neq i}(\mathcal{B}^2)_{ij}^2 \leq \|\mathcal{B}^2\|_F^2 = \sum_k \lambda_k(\mathcal{B})^4 \leq \|\mathcal{B}\|_2^2\,\|\mathcal{B}\|_F^2 \leq O(\bar{T}^{-1})\cdot 1 \to 0$, where the last step uses $\|\mathcal{B}\|_F^2 \leq 1$, which holds because of the normalisation by $\sqrt{\Sigma}$. The same statements follow for $\bar{\mathcal{B}}$ by the same reasoning under the stated assumption that $\|\mathbf{B}+\mathbf{B}'\|_F^2 = \Omega(\|\omega\|^2)$. □

Theorem 2 (Asymptotic normality)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, *and* No Influential Observations,

$$\frac{r'(\hat{\beta}-\beta)}{\sqrt{\Sigma/(h+1)^2}} \xrightarrow{d} N(0,1).$$

Proof. Given Lemma 6, we need only show that $U/\sqrt{\Sigma}$ is asymptotically normal and that

the remainder term is negligible. Defining $\tilde{\varepsilon} = \mathbf{\Phi}^{-1}\varepsilon$, U can be written as follows:

$$U = \omega'\tilde{\varepsilon} + \tilde{\varepsilon}'\mathbf{B}\tilde{\varepsilon}, \tag{24}$$

where the entries of $\tilde{\varepsilon}$ are i.i.d. with bounded fourth moments by Assumption *Regular Errors*. We show asymptotic normality of the self-normalised sum by applying Lemma 8. This result applies to expressions like the one in equation (24) if the matrix of weights for the quadratic part is symmetric and has only zeros on the diagonal. The latter condition is satisfied by design in our case. The former condition is met by rewriting

$$U = \omega'\tilde{\varepsilon} + \frac{1}{2}\tilde{\varepsilon}'(\mathbf{B} + \mathbf{B}')\tilde{\varepsilon}.$$

By Lemma 7, $\Sigma = \mathrm{Var}\,(U) = \|\omega\|^2 + \frac{1}{2}\|\mathbf{B} + \mathbf{B}'\|_F^2$.

Now we apply Lemma 8 to show that

$$\frac{U}{\sqrt{\Sigma}} = \frac{1}{\sqrt{\Sigma}}\left(\omega'\tilde{\varepsilon} + \frac{1}{2}\tilde{\varepsilon}'\left(\mathbf{B} + \mathbf{B}'\right)\tilde{\varepsilon}\right) \xrightarrow{d} N(0,1).$$

The weights w_{nt} and $\mathbf{K}_{st}$ of Lemma 8 correspond to $\omega_t/\sqrt{\Sigma}$ and $\frac{1}{2}(\mathbf{B}_{st}+\mathbf{B}_{ts})/\sqrt{\Sigma}$, respectively.

1. Condition (i) translates to $\frac{1}{\Sigma}\left(\|\omega\|^2 + \frac{1}{2}\|\mathbf{B} + \mathbf{B}'\|_F^2\right) > c > 0$, which is automatically satisfied because $\frac{1}{\Sigma}\left(\|\omega\|^2 + \frac{1}{2}\|\mathbf{B} + \mathbf{B}'\|_F^2\right) = 1$.

2. The second part of condition (ii) translates to $\frac{1}{\Sigma}\|\omega\|^2 \leq 1$, which is also automatically satisfied. The third part translates to $\max_t \omega_t^2/\Sigma \to 0$. Here ω_t denotes the element of ω belonging to period $t \in \{1-h, \ldots, T\}$. First note that the elements of ω belonging to the pre-sample periods $1-h, \ldots, 0$ are already equal to zero.

   Since $\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)}$ by (6), the s-th column of $\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v)}$ equals $\tilde{\mathbf{Z}}'_{1,(v)}$'s s-th column, namely $\tilde{z}_{1,r(v,s)}$, so that, for a period $t = r(v,s)$ of subsample $v = u(t)$, the element of ω belonging to period t takes the simple form

$$\omega_t = r'_{(v)}\mathbf{Q}_{(v)}^{-1}\tilde{z}_{1,t}\,\sigma_t.$$

We bound $\max_t \omega_t^2/\Sigma$ using three facts.

(a) By the Cauchy–Schwarz inequality, $\lambda_{\min}\left(\mathbf{Q}_{(v)}\right) \geq \bar{T}$ (Lemma 4), and $\sigma_t^2 \leq C$ (Assumption *Regular Errors*),

$$\omega_t^2 \leq \sigma_t^2 \left\|\mathbf{Q}_{(v)}^{-1} r_{(v)}\right\|^2 \left\|\tilde{z}_{1,t}\right\|^2 \leq C\bar{T}^{-2}\|r_{(v)}\|^2\|\tilde{z}_{1,t}\|^2.$$

(b) By equation (27) in the proof of Lemma 10,

$$\|\tilde{z}_{1,t}\|^2 = f_t' \left(\mathbf{A}(\tilde{\mathbf{X}}_{(u(t))}'\tilde{\mathbf{X}}_{(u(t))}/\bar{T})^{-1}\mathbf{A}'\right)^{-1} f_t \leq \frac{C}{c}\|f_t\|^2,$$

using $\lambda_{\min}(\mathbf{A}\mathbf{A}') \geq c$ (Assumption *No Influential Observations*) and Assumption *Bounds*.

(c) Since $\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)}$ and $\mathbf{\Phi}_{(v)}^2 \succeq c\,\mathbf{I}_{\bar{T}}$ (Assumption *Regular Errors*),

$$\tilde{\mathbf{Z}}_{1,(v)}'\mathbf{M}_{(v)}\mathbf{\Phi}_{(v)}^2\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)}'\mathbf{\Phi}_{(v)}^2\tilde{\mathbf{Z}}_{1,(v)} \succeq c\,\tilde{\mathbf{Z}}_{1,(v)}'\tilde{\mathbf{Z}}_{1,(v)} = c\bar{T}\mathbf{I}_{h+1}$$

(Lemma 1). Together with $\lambda_{\max}\left(\mathbf{Q}_{(v)}\right) \leq C\bar{T}$ (Lemma 4), which gives $\mathbf{Q}_{(v)}^{-2} \succeq (C\bar{T})^{-2}\mathbf{I}_{h+1}$, this yields, for every v,

$$\begin{aligned}\Sigma \geq \|\omega\|^2 &= \sum_q r_{(q)}'\mathbf{Q}_{(q)}^{-1}\tilde{\mathbf{Z}}_{1,(q)}'\mathbf{\Phi}_{(q)}^2\tilde{\mathbf{Z}}_{1,(q)}\mathbf{Q}_{(q)}^{-1}r_{(q)} \geq c\bar{T}\sum_q r_{(q)}'\mathbf{Q}_{(q)}^{-2}r_{(q)} \\ &\geq c\bar{T}\, r_{(v)}'\mathbf{Q}_{(v)}^{-2}r_{(v)} \geq \frac{c}{C^2\bar{T}}\|r_{(v)}\|^2,\end{aligned}$$

using that every summand in this sum is non-negative.

Combining (a)–(c), whenever $r_{(v)} \neq 0$, for t in subsample v,

$$\frac{\omega_t^2}{\Sigma} \leq \frac{C\bar{T}^{-2}\|r_{(v)}\|^2 \cdot \frac{C}{c}\|f_t\|^2}{\frac{c}{C^2\bar{T}}\|r_{(v)}\|^2} = C'\,\frac{\|f_t\|^2}{\bar{T}}$$

for a constant C' not depending on t, v or T (and, if $r_{(v)} = 0$, $\omega_t = 0$ by (a) so the

same bound holds trivially). Hence

$$\max_t \frac{\omega_t^2}{\Sigma} \leq C' \frac{\max_t \|f_t\|^2}{\bar{T}},$$

which vanishes by Assumption *No Influential Observations*.

3. The third condition translates to $\mathbb{E}[\tilde{\varepsilon}_t^4] < \infty$ (satisfied by Assumption *Regular Errors*), $\frac{1}{\Sigma}\left\|\frac{1}{2}(\mathbf{B}+\mathbf{B}')\right\|_F^2 \leq 1$ (satisfied by construction), $\max_i \frac{1}{\Sigma}\sum_{j\neq i}\left(\frac{\mathbf{B}_{ij}+\mathbf{B}_{ji}}{2}\right)^2 \to 0$ and, letting $\bar{\mathbf{B}} = \frac{1}{4}(\mathbf{B}+\mathbf{B}')^2$, $\frac{1}{\Sigma^2}\sum_i\sum_{j\neq i}\bar{\mathbf{B}}_{ij}^2 \to 0$. The latter two conditions are proven to hold by Lemma 9.

4. The fourth condition amounts to bounded fourth moments of the $\tilde{\varepsilon}_t$, which Assumption *Regular Errors* guarantees.

Thus, the conditions of Lemma 8 are satisfied, proving that $\frac{1}{\sqrt{\Sigma}}U \xrightarrow{d} N(0,1)$.

Let

$$V := \sum_{v=1}^{h+1} r'_{(v)}\left(\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1} - {\mathbf{Q}_{(v)}}^{-1}\right)\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon.$$

We show that $V/\sqrt{\Sigma} = o_p(1)$ by bounding V directly, which is possible because Lemma 7(iii) provides the lower bound $\Sigma = \Omega(T^{-1})$. By the triangle and the Cauchy–Schwarz inequalities,

$$|V| \leq \sum_{v=1}^{h+1} \|r_{(v)}\| \cdot \left\|\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1} - {\mathbf{Q}_{(v)}}^{-1}\right\|_2 \cdot \left\|\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right\|_2. \tag{25}$$

The three factors are $O(1)$, $O_p(T^{-3/2})$ and $O_p(T^{1/2})$, respectively.

The first factor is $O(1)$ because $\|r_{(v)}\|^2 = r'(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}r \leq \|r\|^2/c$ by Lemma 1 and Assumption *Bounds*. The second factor is $O_p(\bar{T}^{-3/2})$ by Lemma 5.

For the third factor, insert $\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)} + \sum_{l=1}^{h}\left(\mathbf{M}_{(v)}\mathbf{S}_{(v-l)}\varepsilon\right)\alpha'_{(v,l)}$ and use $\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)}$ (see (6)) to obtain

$$\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon = \tilde{\mathbf{Z}}'_{1,(v)}\mathbf{S}_{(v)}\varepsilon + \sum_{l=1}^{h}\alpha_{(v,l)}\,\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon.$$

The first term satisfies

$$\mathbb{E}\left[\left\|\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{S}_{(v)}\varepsilon\right\|_2^2\right] = \operatorname{tr}\left(\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\tilde{\mathbf{Z}}'_{1,(v)}\right) \leq C\operatorname{tr}\left(\tilde{\mathbf{Z}}'_{1,(v)}\tilde{\mathbf{Z}}_{1,(v)}\right) = C(h+1)\bar{T},$$

using the trace-norm inequality (Lemma 34), so that the first term is $O_p(\bar{T}^{1/2})$ by Markov's inequality. Each summand of the second term is a quadratic form in ε with mean $\operatorname{tr}\left(\mathbf{M}_{(v)}\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(v-l)}\right) = 0$, since $\mathbf{S}_{(v)}\mathbf{\Phi}^2\mathbf{S}'_{(v-l)} = \mathbf{0}_{\bar{T}\times\bar{T}}$ for $1 \leq l \leq h$ by (5), and

$$\left\|\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\right\|_F \leq \left\|\mathbf{S}_{(v-l)}\right\|_2 \left\|\mathbf{M}_{(v)}\right\|_F \left\|\mathbf{S}_{(v)}\right\|_2 = \sqrt{\bar{T}-K+h+1} = O(\bar{T}^{1/2}),$$

so that it is $O_p(\bar{T}^{1/2})$ by Lemma 2(ii). Moreover, $\|\alpha_{(v,l)}\|^2 = \alpha'_l(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}\alpha_l \leq \|\alpha_l\|^2/c = O(1)$ by Lemma 1 and Assumption *Bounds*. As h is fixed, the third factor of (25) is $O_p(T^{1/2})$.

Multiplying the three factors gives $V = O_p(T^{-1})$, and therefore

$$\frac{V}{\sqrt{\Sigma}} = \frac{O_p(T^{-1})}{\Omega(T^{-1/2})} = O_p(T^{-1/2}) = o_p(1).$$

The same argument applies to the $O_p(T^{-1})$ remainder of the linearisation in Lemma 6: it is negligible on the studentised scale precisely because $\sqrt{\Sigma} \asymp T^{-1/2}$.

□

A.3 Estimating variances of subsample estimators

Lemma 10 (Bounded average fourth moment of $z_{1,r(v,t)}$)**.** *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*,* No Influential Observations *and* Bounded Leverage*, for every* $v \in \{1,\ldots,h+1\}$*,*

$$\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}}\mathbb{E}\left[\|z_{1,r(v,t)}\|^4\right] = O(1) \qquad \textit{and} \qquad \frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}}\mathbb{E}\left[\|\hat{z}_{1,r(v,t)}\|^4\right] = O(1),$$

Proof. Fix a subsample $v \in \{1,\ldots,h+1\}$, and recall the decomposition (7) of $\hat{z}_{1,r(v,t)}$.

By the c_r-inequality $\|a+b\|^4 \leq 8(\|a\|^4 + \|b\|^4)$ applied to (7), and then again to the sum

over l,

$$\|\hat{z}_{1,r(v,t)}\|^4 \leq 8\,\|\tilde{z}_{1,r(v,t)}\|^4 + 8\,h^3 \sum_{l=1}^{h} |\psi_{(v,l),t}|^4\,\|\alpha_{(v,l)}\|^4. \tag{26}$$

We bound the average over t of each term separately.

Consider first the non-random part. Write $\boldsymbol{\Upsilon}_{(v)} := (\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{1/2}$. Let $\mathbf{L}_{(v)}$ be the matrix from the proof of Lemma 1, i.e., the columns of $\mathbf{L}_{(v)}$ form an orthonormal basis for $\mathrm{col}(\boldsymbol{\Upsilon}_{(v)}^{-1}\mathbf{A}')$, and the orthogonal projection onto this subspace is

$$\begin{aligned}\mathbf{L}_{(v)}\mathbf{L}'_{(v)} &= \boldsymbol{\Upsilon}_{(v)}^{-1}\mathbf{A}'(\mathbf{A}\boldsymbol{\Upsilon}_{(v)}^{-2}\mathbf{A}')^{-1}\mathbf{A}\boldsymbol{\Upsilon}_{(v)}^{-1} \\ &= (\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1/2}\mathbf{A}'\big(\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}\mathbf{A}'\big)^{-1}\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1/2}.\end{aligned}$$

Since $\boldsymbol{\Theta}_{(v),1} = \boldsymbol{\Upsilon}_{(v)}^{-1}\mathbf{L}_{(v)}$,

$$\begin{aligned}\|\tilde{z}_{1,r(v,t)}\|^2 &= \tilde{x}'_{r(v,t)}\boldsymbol{\Theta}_{(v),1}\boldsymbol{\Theta}'_{(v),1}\tilde{x}_{r(v,t)} = \tilde{x}'_{r(v,t)}\boldsymbol{\Upsilon}_{(v)}^{-1}\mathbf{L}_{(v)}\mathbf{L}'_{(v)}\boldsymbol{\Upsilon}_{(v)}^{-1}\tilde{x}_{r(v,t)} \\ &= f'_{r(v,t)}\big(\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}\mathbf{A}'\big)^{-1}f_{r(v,t)} \;\leq\; \frac{\|f_{r(v,t)}\|^2}{\lambda_{\min}\big(\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}\mathbf{A}'\big)}.\end{aligned} \tag{27}$$

For the denominator,

$$\lambda_{\min}\left(\mathbf{A}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T})^{-1}\mathbf{A}'\right) \geq \frac{c}{C},$$

using $\lambda_{\min}\left(\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}\right)^{-1}\right) = \lambda_{\max}\left(\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}/\bar{T}\right)\right)^{-1} \geq 1/C$ and $\lambda_{\min}(\mathbf{A}\mathbf{A}') \geq c > 0$ (Lemma 37). Thus,

$$\|\tilde{z}_{1,r(v,t)}\|^4 \;\leq\; \frac{C^2}{c^2}\,\|f_{r(v,t)}\|^4.$$

Averaging over t and applying Assumption *Bounded Leverage*,

$$\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}}\|\tilde{z}_{1,r(v,t)}\|^4 \;\leq\; \frac{C^2}{c^2}\cdot\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}}\|f_{r(v,t)}\|^4 \;=\; O(1). \tag{28}$$

We next turn to the feedback part. Fix $l \in \{1,\ldots,h\}$. Recall that

$$\psi_{(v,l),t} = \mathbf{M}_{(v),t\cdot}\mathbf{S}_{(v-l)}\varepsilon = \sum_{s=1}^{\bar{T}}\mathbf{M}_{(v),ts}\,\varepsilon_{v-l+(s-1)(h+1)},$$

a linear combination of independent, mean-zero random variables which have variances in $[c, C]$ and uniformly bounded fourth moments $\mathbb{E}[\varepsilon_t^4] \leq C^2$ (Assumption *Regular Errors*). We use Rosenthal's inequality: for independent, mean-zero random variables $\zeta_1, \ldots, \zeta_n$ and $p \geq 2$ there is a constant C_p, depending only on p, such that

$$\mathbb{E}\left[\left|\sum_j \zeta_j\right|^p\right] \;\leq\; C_p\left(\sum_j \mathbb{E}\left[|\zeta_j|^p\right] + \left(\sum_j \mathbb{E}\left[\zeta_j^2\right]\right)^{p/2}\right). \tag{29}$$

With $p = 4$, it gives

$$\begin{aligned}\mathbb{E}\left[\psi_{(v,l),t}^4\right] &\leq C_4\Big(\sum_s \mathbf{M}_{(v),ts}^4\,\mathbb{E}\left[(\mathbf{S}_{(v-l)}\varepsilon)_s^4\right] + \Big(\sum_s \mathbf{M}_{(v),ts}^2\,\mathbb{E}\left[(\mathbf{S}_{(v-l)}\varepsilon)_s^2\right]\Big)^2\Big) \\ &\leq C_4\big(C^2 + C^2\big)\Big(\sum_s \mathbf{M}_{(v),ts}^2\Big)^2,\end{aligned} \tag{30}$$

where the last step uses $\sum_s \mathbf{M}_{(v),ts}^4 \leq \big(\sum_s \mathbf{M}_{(v),ts}^2\big)^2$ to merge the two sums. Because $\mathbf{M}_{(v)}$ is an orthogonal projection,

$$\sum_s \mathbf{M}_{(v),ts}^2 = \|\mathbf{M}_{(v),\cdot t}\|^2 = \mathbf{M}_{(v),tt} \leq \big\|\mathbf{M}_{(v)}\big\|_2^2 = 1,$$

so (30) yields the uniform bound

$$\mathbb{E}\left[\psi_{(v,l),t}^4\right] \leq 2C_4C^2 =: C' = O(1), \qquad \text{uniformly in } v, l, t. \tag{31}$$

Recalling from Lemma 1 that $\|\alpha_{(v,l)}\|^2 = \bar{T}\,\alpha_l'(\tilde{\mathbf{X}}_{(v)}'\tilde{\mathbf{X}}_{(v)})^{-1}\alpha_l \leq \|\alpha_l\|^2/c \leq C^2/c$, the feedback term of (26) obeys

$$\frac{8h^3}{\bar{T}}\sum_{t=1}^{\bar{T}}\sum_{l=1}^{h}\mathbb{E}\left[|\psi_{(v,l),t}|^4\right]\,\|\alpha_{(v,l)}\|^4 \leq 8h^3 \cdot h \cdot C' \cdot \frac{C^4}{c^2} = \frac{8h^4C'C^4}{c^2} = O(1). \tag{32}$$

Taking expectations in (26) and combining (28) with (32) gives the second claim, uniformly in v.

Likewise,

$$z_{1,r(v,t)} = \tilde{z}_{1,r(v,t)} + \sum_{l=1}^{h} \varepsilon_{r(v,t)-l}\, \alpha_{(v,l)},$$

so the c_r-inequality, exactly as in (26), gives

$$\|z_{1,r(v,t)}\|^4 \leq 8\, \|\tilde{z}_{1,r(v,t)}\|^4 + 8\, h^3 \sum_{l=1}^{h} |\varepsilon_{r(v,t)-l}|^4\, \|\alpha_{(v,l)}\|^4. \tag{33}$$

The non-random part is identical to (28), and is hence $O(1)$ on average. For the feedback part, $\mathbb{E}\left[|\varepsilon_{r(v,t)-l}|^4\right] \leq C^2$ replaces the Rosenthal bound (31), so the same computation as in (32) yields

$$\frac{8h^3}{\bar{T}} \sum_{t=1}^{\bar{T}} \sum_{l=1}^{h} \mathbb{E}\left[|\varepsilon_{r(v,t)-l}|^4\right]\, \|\alpha_{(v,l)}\|^4 \leq 8h^3 \cdot h \cdot C^2 \cdot \frac{C^4}{c^2} = \frac{8h^4 C^6}{c^2} = O(1).$$

Taking expectations in (33) and averaging therefore gives $\frac{1}{\bar{T}} \sum_{t=1}^{\bar{T}} \mathbb{E}\left[\|z_{1,r(v,t)}\|^4\right] = O(1)$ uniformly in v, which is the first claim.

□

Lemma 11 (Any sandwich with consistent meat is consistent for $\sum_{v=1}^{h+1} \Sigma_v$)**.** *Let* $\hat{\mathbf{\Psi}}_{(1)}, \ldots, \hat{\mathbf{\Psi}}_{(h+1)}$ *be any (possibly random)* $(h+1) \times (h+1)$ *matrices satisfying*

$$\frac{1}{\bar{T}} \left\| \hat{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)} \right\|_2 \xrightarrow{p} 0, \qquad v = 1, \ldots, h+1. \tag{34}$$

Then, under Assumptions Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime,

$$\frac{\sum_{v=1}^{h+1} r'_{(v)} \left(\mathbf{Z}'_{1,(v)} \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)}\right)^{-1} \hat{\mathbf{\Psi}}_{(v)} \left(\mathbf{Z}'_{1,(v)} \mathbf{M}_{(v)} \mathbf{Z}_{1,(v)}\right)^{-1} r_{(v)}}{\sum_{v=1}^{h+1} \Sigma_v} \xrightarrow{p} 1.$$

Proof. The proof proceeds in three steps. In the first step, $\hat{\mathbf{\Psi}}_{(v)}$ is replaced by $\mathbf{\Psi}_{(v)}$. By

Lemmas 37 and 3 and Assumption *Regular Errors*,

$$\lambda_{\min}\left(\frac{1}{\bar{T}}\mathbf{\Psi}_{(v)}\right) = \lambda_{\min}\left(\frac{1}{\bar{T}}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right) \geq \lambda_{\min}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}\right)\lambda_{\min}\left(\mathbf{\Phi}^2_{(v)}\right)$$
$$\geq (1-o_p(1))\, c.$$

Writing $p_{(v)} := \left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1} r_{(v)}$, it follows from (34) that, for every v,

$$\frac{\left|p'_{(v)}\left(\hat{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)}\right)p_{(v)}\right|}{p'_{(v)}\mathbf{\Psi}_{(v)}p_{(v)}} \leq \frac{\left\|\hat{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)}\right\|_2 \|p_{(v)}\|^2}{\lambda_{\min}\left(\mathbf{\Psi}_{(v)}\right)\|p_{(v)}\|^2}$$
$$= \frac{o_p(\bar{T})}{c\,\bar{T}(1-o_p(1))} = o_p(1).$$

As the $h{+}1$ summands $p'_{(v)}\mathbf{\Psi}_{(v)}p_{(v)}$ are non-negative, the bound is preserved under summation over v, so that

$$\sum_v p'_{(v)}\hat{\mathbf{\Psi}}_{(v)}p_{(v)} = (1+o_p(1))\sum_v p'_{(v)}\mathbf{\Psi}_{(v)}p_{(v)}. \tag{35}$$

In the second step, $\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}$ is replaced by its expectation. Put $\mathbf{\Delta}_{(v)} := (\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1} - \mathbf{Q}^{-1}_{(v)}$, so that

$$(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}r_{(v)} = \mathbf{Q}^{-1}_{(v)}r_{(v)} + \mathbf{\Delta}_{(v)}r_{(v)}.$$

By Lemma 5, $\left\|\mathbf{\Delta}_{(v)}\right\|_2 = O_p(\bar{T}^{-3/2})$, whereas by Lemma 4,

$$\|\mathbf{Q}^{-1}_{(v)}r_{(v)}\| \geq \lambda_{\max}\left(\mathbf{Q}_{(v)}\right)^{-1}\|r_{(v)}\| \geq c\,\bar{T}^{-1}\|r_{(v)}\|.$$

Hence, writing $a_{(v)} := \mathbf{Q}^{-1}_{(v)}r_{(v)}$ and $b_{(v)} := \mathbf{\Delta}_{(v)}r_{(v)}$, so that $p_{(v)} = a_{(v)} + b_{(v)}$,

$$\frac{\|b_{(v)}\|}{\|a_{(v)}\|} \leq \frac{O_p(\bar{T}^{-3/2})\|r_{(v)}\|}{c\,\bar{T}^{-1}\|r_{(v)}\|} = O_p(\bar{T}^{-1/2}), \tag{36}$$

which is well defined because $\|r_{(v)}\| \asymp 1$ for every v by Lemma 1 and Assumption *Bounds*. Assumption *Bounds* bounds $\|r_{(v_0)}\|^2 = r'(\tilde{\mathbf{X}}'_{(v_0)}\tilde{\mathbf{X}}_{(v_0)}/\bar{T})^{-1}r$ from below for at least one subsample v_0, so that $\|r\|^2 \geq c\,\lambda_{\min}\left(\tilde{\mathbf{X}}'_{(v_0)}\tilde{\mathbf{X}}_{(v_0)}/\bar{T}\right) > c^2$, and then $\|r\|^2/C \leq \|r_{(v)}\|^2 \leq \|r\|^2/c$

for every v.

Since $\mathbf{\Psi}_{(v)} \succeq 0$, the Cauchy–Schwarz inequality for the inner product induced by $\mathbf{\Psi}_{(v)}$ gives

$$\begin{aligned}\frac{\left|p'_{(v)}\mathbf{\Psi}_{(v)}p_{(v)} - a'_{(v)}\mathbf{\Psi}_{(v)}a_{(v)}\right|}{a'_{(v)}\mathbf{\Psi}_{(v)}a_{(v)}} &\leq 2\sqrt{\frac{b'_{(v)}\mathbf{\Psi}_{(v)}b_{(v)}}{a'_{(v)}\mathbf{\Psi}_{(v)}a_{(v)}}} + \frac{b'_{(v)}\mathbf{\Psi}_{(v)}b_{(v)}}{a'_{(v)}\mathbf{\Psi}_{(v)}a_{(v)}} \\ &\leq 2\sqrt{\operatorname{cond}(\mathbf{\Psi}_{(v)})}\frac{\|b_{(v)}\|}{\|a_{(v)}\|} + \operatorname{cond}(\mathbf{\Psi}_{(v)})\frac{\|b_{(v)}\|^2}{\|a_{(v)}\|^2},\end{aligned}$$

where $\operatorname{cond}(\mathbf{\Psi}_{(v)}) := \lambda_{\max}\left(\mathbf{\Psi}_{(v)}\right)/\lambda_{\min}\left(\mathbf{\Psi}_{(v)}\right)$ is the condition number of $\mathbf{\Psi}_{(v)}$. It is $O_p(1)$: the lower bound of the first step gives $\lambda_{\min}\left(\mathbf{\Psi}_{(v)}/\bar{T}\right) \geq c(1-o_p(1))$, while $\mathbf{\Phi}^2_{(v)} \preceq C\,\mathbf{I}_{\bar{T}}$ and $\mathbf{M}^2_{(v)} = \mathbf{M}_{(v)}$ give $\mathbf{\Psi}_{(v)} \preceq C\,\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}$, whose largest eigenvalue is $O_p(\bar{T})$ because $\left\|\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} - \mathbf{Q}_{(v)}\right)/\bar{T}\right\|_2 = O_p(\bar{T}^{-1/2})$ (proof of Lemma 5) and $\lambda_{\max}\left(\mathbf{Q}_{(v)}/\bar{T}\right) = O(1)$ (Lemma 4). With (36), each quadratic form therefore changes by a factor $1+o_p(1)$ when $p_{(v)}$ is replaced by $a_{(v)} = \mathbf{Q}^{-1}_{(v)}r_{(v)}$, and, the $a'_{(v)}\mathbf{\Psi}_{(v)}a_{(v)}$ being non-negative, so does their sum. It therefore suffices to determine the limit of $\sum_v r'_{(v)}\mathbf{Q}^{-1}_{(v)}\mathbf{\Psi}_{(v)}\mathbf{Q}^{-1}_{(v)}r_{(v)}$.

In the third step, this limit is determined. Using $\mathbf{M}_{(v)}\tilde{\mathbf{Z}}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)}$, which holds because the transformation $\mathbf{\Theta}_{(v)}$ orthogonalises the regressors, and $\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} = \tilde{\mathbf{Z}}_{1,(v)} + \sum_{l=1}^{h}\left(\mathbf{M}_{(v)}\mathbf{S}_{(v-l)}\varepsilon\right)\alpha'_{(v,l)}$, we decompose

$$\sum_v r'_{(v)}\mathbf{Q}^{-1}_{(v)}\mathbf{\Psi}_{(v)}\mathbf{Q}^{-1}_{(v)}r_{(v)} = \text{(I)} + \text{(II)} + \text{(III)}, \tag{37}$$

where

$$\begin{aligned}\text{(I)} &:= \sum_v r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\tilde{\mathbf{Z}}_{1,(v)}\mathbf{Q}^{-1}_{(v)}r_{(v)}, \\ \text{(II)} &:= 2\sum_v\sum_{l=1}^{h}\gamma_{v,l}\,r'_{(v)}\mathbf{Q}^{-1}_{(v)}\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v-l)}\varepsilon, \\ \text{(III)} &:= \sum_v\sum_{l,m=1}^{h}\gamma_{v,l}\gamma_{v,m}\,\varepsilon'\mathbf{S}'_{(v-l)}\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{S}_{(v-m)}\varepsilon.\end{aligned}$$

The first term is the contribution of the linear part, (I) $= \|\omega\|^2$, and satisfies the lower bound

$$\begin{aligned}\|\omega\|^2 \geq & \sum_v \|r_{(v)}\|^2 \lambda_{\max}\left(\mathbf{Q}_{(v)}\right)^{-2} \lambda_{\min}\left(\tilde{\mathbf{Z}}'_{1,(v)}\tilde{\mathbf{Z}}_{1,(v)}\right) \lambda_{\min}\left(\mathbf{\Phi}^2_{(v)}\right) \\ \geq & c\,\bar{T}^{-1} \sum_v \|r_{(v)}\|^2,\end{aligned} \tag{38}$$

by Lemmas 37, 4 and 1.

Consider next the summands of (III) with $m = l$. Since

$$\begin{aligned}\mathbb{E}\left[\varepsilon' \mathbf{S}'_{(v-l)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{S}_{(v-l)} \varepsilon\right] &= \operatorname{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)}\, \mathbf{S}_{(v-l)} \mathbf{\Phi}^2 \mathbf{S}'_{(v-l)}\right) \\ &= \operatorname{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v-l)}\right),\end{aligned}$$

the expectation of these summands is

$$\sum_{v,l} \gamma^2_{v,l} \operatorname{tr}\left(\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v-l)}\right) = \operatorname{tr}\left(\mathbf{B}\mathbf{B}'\right),$$

by Lemma 7. Lemma 2(ii) applies to the deviation of each summand from its mean, and, since $\left\|\mathbf{S}_{(v-l)}\right\|_2 = 1$ and $\|\mathbf{F}\mathbf{G}\|_F \leq \|\mathbf{F}\|_2 \|\mathbf{G}\|_F$ for conformable matrices,

$$\left\|\mathbf{S}'_{(v-l)} \mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \mathbf{S}_{(v-l)}\right\|_F \leq \left\|\mathbf{M}_{(v)}\right\|_2 \left\|\mathbf{\Phi}^2_{(v)}\right\|_2 \left\|\mathbf{M}_{(v)}\right\|_F = O(\bar{T}^{1/2}),$$

so that the deviation is $O_p\big(\bar{T}^{1/2} \sum_{v,l} \gamma^2_{v,l}\big)$. The summands of (III) with $m \neq l$ have expectation zero, because the index sets of $\mathbf{S}_{(v-l)}$ and $\mathbf{S}_{(v-m)}$ are disjoint, and by the same Frobenius bound each of them is $O_p\big(\bar{T}^{1/2} |\gamma_{v,l}\gamma_{v,m}|\big)$. Both deviations are negligible relative to $\operatorname{tr}(\mathbf{B}\mathbf{B}')$. Indeed, since $\mathbf{\Phi}^2_{(v)} \succeq c\,\mathbf{I}_{\bar{T}}$, $\mathbf{\Phi}^2_{(v-l)} \succeq c\,\mathbf{I}_{\bar{T}}$ and $\mathbf{M}_{(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \succeq 0$, two applications of the trace lower bound (Lemma 35), together with $\operatorname{tr}\left(\mathbf{M}^2_{(v)}\right) = \operatorname{tr}\left(\mathbf{M}_{(v)}\right)$, give

$$\operatorname{tr}\left(\mathbf{B}\mathbf{B}'\right) \;\geq\; c^2 \left(\bar{T} - K + h + 1\right) \sum_{v,l} \gamma^2_{v,l}.$$

Under Assumption *Proportional Regime*, $\bar{T} - K + h + 1 \asymp \bar{T}$, so that both deviations are $O_p(\bar{T}^{-1/2} \operatorname{tr}(\mathbf{B}\mathbf{B}')) = o_p(\operatorname{tr}(\mathbf{B}\mathbf{B}'))$. Hence (III) $= \operatorname{tr}(\mathbf{B}\mathbf{B}') + o_p(\operatorname{tr}(\mathbf{B}\mathbf{B}'))$.

It remains to bound (II). By Lemma 2(i), applied with $\tilde{\mathbf{X}} \equiv \tilde{\mathbf{Z}}_{1,(v)}$ and $\mathbf{B} \equiv \mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}$,

$$\left\| \tilde{\mathbf{Z}}'_{1,(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \sum_l \mathbf{S}_{(v-l)} \varepsilon \alpha'_{(v,l)} \right\|_2 = O_p(\bar{T}^{1/2}),$$

and $\|\mathbf{Q}^{-1}_{(v)} r_{(v)}\| \leq \lambda_{\min}\left(\mathbf{Q}_{(v)}\right)^{-1} \|r_{(v)}\| = O(\bar{T}^{-1})\|r_{(v)}\|$. Since $\gamma_{v,l} = \alpha'_{(v,l)}\mathbf{Q}^{-1}_{(v)} r_{(v)}$, the term (II) is the quadratic form

$$\text{(II)} = 2\sum_v \left(\mathbf{Q}^{-1}_{(v)} r_{(v)}\right)' \tilde{\mathbf{Z}}'_{1,(v)} \mathbf{\Phi}^2_{(v)} \mathbf{M}_{(v)} \sum_{l=1}^{h} \mathbf{S}_{(v-l)} \varepsilon \alpha'_{(v,l)} \left(\mathbf{Q}^{-1}_{(v)} r_{(v)}\right),$$

so that

$$\begin{aligned} |\text{(II)}| &\leq O_p(\bar{T}^{1/2}) \sum_v \|\mathbf{Q}^{-1}_{(v)} r_{(v)}\|^2 = O_p(\bar{T}^{-1/2}) \cdot \bar{T}^{-1} \sum_v \|r_{(v)}\|^2 \\ &= o_p(\|\omega\|^2), \end{aligned}$$

by (38).

Combining the three terms, the right-hand side of (35) equals

$$\|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}') + o_p\left(\|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}')\right) = \sum_{v=1}^{h+1} \Sigma_v \left(1 + o_p(1)\right),$$

by Lemma 7, which together with (35) proves the claim. □

Lemma 12 (Maximum of variables with bounded $2+\delta$ moments)**.** *For each n, let $\xi_{n,1}, \ldots, \xi_{n,n}$ be random variables satisfying*

$$\sup_{n,i} \mathbb{E}\left[|\xi_{n,i}|^{2+\delta}\right] \leq C_\xi < \infty \quad \textit{for some } \delta > 0.$$

Then $\max_i |\xi_{n,i}| = o_p(n^{1/2})$.

Proof. Fix $\epsilon > 0$. By the union bound,

$$\mathbb{P}\left[\max_{1 \leq i \leq n} |\xi_{n,i}| > \epsilon n^{1/2}\right] \leq \sum_{i=1}^{n} \mathbb{P}\left[|\xi_{n,i}| > \epsilon n^{1/2}\right].$$

Apply Markov's inequality to $|\xi_{n,i}|^{2+\delta}$:

$$\mathbb{P}\left[|\xi_{n,i}| > \epsilon n^{1/2}\right] = \mathbb{P}\left[|\xi_{n,i}|^{2+\delta} > \epsilon^{2+\delta} n^{(2+\delta)/2}\right] \leq \frac{\mathbb{E}\left[|\xi_{n,i}|^{2+\delta}\right]}{\epsilon^{2+\delta}\, n^{(2+\delta)/2}} \leq \frac{C_\xi}{\epsilon^{2+\delta}\, n^{1+\delta/2}}.$$

Summing over $i = 1, \ldots, n$,

$$\mathbb{P}\left[\max_{1\leq i\leq n} |\xi_{n,i}| > \epsilon n^{1/2}\right] \leq n \cdot \frac{C_\xi}{\epsilon^{2+\delta}\, n^{1+\delta/2}} = \frac{C_\xi}{\epsilon^{2+\delta}}\, n^{-\delta/2}.$$

Since $\delta > 0$, the right-hand side tends to 0 as $n \to \infty$. As $\epsilon > 0$ was arbitrary, $\max_i |\xi_{n,i}| = o_p(n^{1/2})$. □

Lemma 13 (Maximum row norm of $\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}$)**.** *Under Assumptions* Feedback Decomposition, Proportional Regime, Bounds, Regular Errors, No Influential Observations *and* Bounded Leverage,

$$\max_{v\in\{1,\ldots,h+1\}} \max_{t\in\{1,\ldots,\bar{T}\}} \|\tilde{z}_{1,r(v,t)}\| = o_p(\bar{T}^{1/2}),$$

and

$$\max_{v\in\{1,\ldots,h+1\}} \max_{t\in\{1,\ldots,\bar{T}\}} \|\hat{z}_{1,r(v,t)}\| = o_p(\bar{T}^{1/2}).$$

Proof. Fix v and recall from (7) and (6) that

$$\hat{z}_{1,r(v,t)} = \tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v),\cdot t} + \sum_{l=1}^{h} \psi_{(v,l),t}\, \alpha_{(v,l)}.$$

By the triangle inequality and $\|\alpha_{(v,l)}\| \leq C$,

$$\max_t \|\hat{z}_{1,r(v,t)}\| \leq \max_t \|\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v),\cdot t}\| + C\sum_{l=1}^{h} \max_t |\psi_{(v,l),t}|.$$

Consider first the non-random part. From (27) in Lemma 10, $\|\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v),\cdot t}\|^2 \leq (C^2/c^2)\|f_{r(v,t)}\|^2$ because $\lambda_{\min}(\mathbf{A}\mathbf{A}') > c$. Because $\frac{1}{\bar{T}}\sum_t \|f_{r(v,t)}\|^4 < C$ by Assumption *Bounded Leverage*, we have $\max_t \|f_{r(v,t)}\|^4 \leq \sum_t \|f_{r(v,t)}\|^4 < C\bar{T}$, so that

$\max_t \|f_{r(v,t)}\|^2 = O(\bar{T}^{1/2}) = o(\bar{T})$. This establishes that

$$\max_t \|\tilde{z}_{1,r(v,t)}\|^2 = \max_t \|\tilde{\mathbf{Z}}'_{1,(v)}\mathbf{M}_{(v),\cdot t}\|^2 = o(\bar{T}).$$

We next bound the feedback part. Fix l. Since $\psi_{(v,l),t} = \sum_s \mathbf{M}_{(v),ts}(\mathbf{S}_{(v-l)}\varepsilon)_s$ is a linear combination of independent mean-zero random variables with variances in $[c, C]$ and bounded $2+\delta$ moments (for some $\delta \in (0, 2]$), Rosenthal's inequality is applicable and shows that $\psi_{(v,l),t}$ also has moments of order greater than two:

$$\begin{aligned}
\mathbb{E}\left[|\psi_{(v,l),t}|^{2+\delta}\right] \leq & C\left(\sum_s |\mathbf{M}_{(v),ts}|^{2+\delta}\mathbb{E}\left[|\varepsilon_{v-l+(s-1)(h+1)}|^{2+\delta}\right]\right. \\
& \left. + \Big(\sum_s \mathbf{M}^2_{(v),ts}\,\mathbb{E}\left[\varepsilon^2_{v-l+(s-1)(h+1)}\right]\Big)^{1+\delta/2}\right) \\
\leq & C(C^2 + C^{1+\delta/2})\max\Big\{\sum_s \mathbf{M}^2_{(v),ts}, \Big(\sum_s \mathbf{M}^2_{(v),ts}\Big)^{1+\delta/2}\Big\} \\
\leq & C(C^2 + C^{1+\delta/2}).
\end{aligned}$$

The second inequality uses $|\mathbf{M}_{(v),ts}| \leq 1$, and the last uses $\sum_s \mathbf{M}^2_{(v),ts} = \|\mathbf{M}_{(v),\cdot t}\|^2 \leq \left\|\mathbf{M}_{(v)}\right\|_2^2 = 1$. Lemma 12 then gives $\max_{v,l,t} |\psi_{(v,l),t}| = o_p(\bar{T}^{1/2})$, and the second claim follows from the triangle inequality, as h is fixed. □

Lemma 14 (Leverage bound)**.** *Let* $\varphi_t := 1 - x_t'(\mathbf{X}'_{(u(t))}\mathbf{X}_{(u(t))})^{-1}x_t$. *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*,* No Influential Observations*,* Bounded Leverage *and* Leave-One-Out Conditioning*,*

$$\max_{1\leq t\leq T} \frac{1}{\varphi_t} = O_p(1).$$

Proof. We show first that the statement holds for $\tilde{\varphi}_{r(v,s)} := \tilde{\mathbf{H}}_{(v),ss}$, $v \in \{1, \ldots, h+1\}$, $s \in \{1, \ldots, \bar{T}\}$, where

$$\tilde{\mathbf{H}}_{(v)} = \mathbf{I}_{\bar{T}} - \tilde{\mathbf{X}}_{(v)}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)})^{-1}\tilde{\mathbf{X}}'_{(v)}.$$

For each v and s, let $\mathbf{N}_{(v),-s} = \frac{1}{\bar{T}}\tilde{\mathbf{X}}'_{(v),-s}\tilde{\mathbf{X}}_{(v),-s}$, where $\tilde{\mathbf{X}}_{(v),-s}$ is as defined in Assumption *Leave-One-Out Conditioning*. Applying the Sherman–Morrison formula to the rank-one

decomposition $\frac{1}{\bar{T}}\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)} = \mathbf{N}_{(v),-s} + \frac{1}{\bar{T}}\tilde{x}_{r(v,s)}\tilde{x}'_{r(v,s)}$ gives

$$\tilde{\varphi}_{r(v,s)} = \frac{1}{1+\tau_{r(v,s)}}, \qquad \tau_{r(v,s)} := \frac{1}{\bar{T}}\,\tilde{x}'_{r(v,s)}\mathbf{N}^{-1}_{(v),-s}\tilde{x}_{r(v,s)} \;\geq\; 0.$$

Using $\lambda_{\min}(\mathbf{N}_{(v),-s}) > c$ and $\|\tilde{x}_{r(v,s)}\|^2 \leq \left\|\tilde{\mathbf{X}}_{(v)}\right\|_2^2 = \lambda_{\max}(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}) < \bar{T}C$, we obtain $\max_t \tau_t \leq C/c$, so that $\min_t \tilde{\varphi}_t \geq c/(c+C)$ is bounded away from zero.

Next, fix v and a position s, and recall that $\mathbf{m}_{(v),s} = \mathbf{M}_{(v),ss}$. As $\mathrm{col}(\mathbf{Z}_{2,(v)}) \subseteq \mathrm{col}(\mathbf{X}_{(v)})$, we can decompose $\mathbf{H}_{(v)}$ as

$$\mathbf{H}_{(v)} = \mathbf{M}_{(v)} - \mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)},$$

so that $\varphi_{r(v,s)} = \mathbf{m}_{(v),s} - \hat{z}'_{1,r(v,s)}\,(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}\hat{z}_{1,r(v,s)}$. The second term is bounded as follows:

$$\hat{z}'_{1,r(v,s)}\,(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}\hat{z}_{1,r(v,s)} \leq \frac{\|\hat{z}_{1,r(v,s)}\|^2}{\lambda_{\min}(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})} = \frac{1}{\bar{T}}\,\frac{\|\hat{z}_{1,r(v,s)}\|^2}{\lambda_{\min}(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T})}.$$

By Lemma 3, $\lambda_{\min}(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}) \geq 1 - o_p(1)$, and by Lemma 13, $\max_s \|\hat{z}_{1,r(v,s)}\|^2/\bar{T} = o_p(1)$, so that

$$\max_s \hat{z}'_{1,r(v,s)}\,(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}\hat{z}_{1,r(v,s)} \leq \bar{T}^{-1}\max_s \|\hat{z}_{1,r(v,s)}\|^2 \cdot O_p(1) = o_p(1).$$

Since $\mathrm{col}(\mathbf{Z}_{2,(v)}) \subseteq \mathrm{col}(\tilde{\mathbf{X}}_{(v)})$, $\mathbf{M}_{(v)}$ dominates $\tilde{\mathbf{H}}_{(v)}$ in the Loewner order, so that $\mathbf{m}_{(v),s} \geq \tilde{\varphi}_{r(v,s)} \geq c/(c+C)$. Hence $\min_t \varphi_t > c/(c+C) - o_p(1)$, so that the probability that $\max_t \frac{1}{\varphi_t} < 2(c+C)/c$ goes to 1. □

Lemma 15 (Consistency of $\sum_{v=1}^{h+1}\acute{\Sigma}_v$)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, No Influential Observations, Bounded Leverage, Bounded Fitted Values *and* Leave-One-Out Conditioning, $\frac{\sum_{v=1}^{h+1}\acute{\Sigma}_v}{\sum_{v=1}^{h+1}\Sigma_v} \xrightarrow{p} 1$.

Proof. By definition,

$$\begin{aligned}\sum_{v=1}^{h+1}\acute{\Sigma}_v =& r'\sum_v \left(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}\right)^{-1}\mathbf{X}'_{(v)}\acute{\mathbf{\Phi}}^2_{(v)}\mathbf{X}_{(v)}\left(\mathbf{X}'_{(v)}\mathbf{X}_{(v)}\right)^{-1}r \\ =& \sum_v r'\mathbf{\Theta}_{(v)}\left(\mathbf{Z}'_{(v)}\mathbf{Z}_{(v)}\right)^{-1}\mathbf{Z}'_{(v)}\acute{\mathbf{\Phi}}^2_{(v)}\mathbf{Z}_{(v)}\left(\mathbf{Z}'_{(v)}\mathbf{Z}_{(v)}\right)^{-1}\mathbf{\Theta}'_{(v)}r \\ =& \sum_v r'_{(v)}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1}\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\acute{\mathbf{\Phi}}^2_{(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1}r_{(v)},\end{aligned}$$

where $\acute{\mathbf{\Phi}}^2_{(v)} = \operatorname{diag}\left(\left(y_{r(v,t)}\frac{\hat{\varepsilon}_{r(v,t)}}{\mathbf{H}_{(v),tt}}\right)_{t=1,\ldots,\bar{T}}\right)$.

Let

$$\acute{\mathbf{\Psi}}_{(v)} := \sum_{t=1}^{\bar{T}}\hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}\frac{y_{r(v,t)}\hat{\varepsilon}_{r(v,t)}}{\mathbf{H}_{(v),tt}}$$

be the meat of $\acute{\Sigma}_v$, and recall that the meat of Σ_v is $\mathbf{\Psi}_{(v)} = \mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} = \sum_{t=1}^{\bar{T}}\hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}\sigma^2_{r(v,t)}$. We show that

$$\frac{1}{\bar{T}}\left\|\acute{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)}\right\|_2 \xrightarrow{p} 0 \qquad \text{for each } v \in \{1,\ldots,h+1\}, \tag{39}$$

which is condition (34) of Lemma 11. The rest of the proof follows from that lemma.

By definition,

$$\frac{1}{\bar{T}}\left(\acute{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)}\right) = \frac{1}{\bar{T}}\sum_t \hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}\left(\frac{y_{r(v,t)}\hat{\varepsilon}_{r(v,t)}}{\mathbf{H}_{(v),tt}} - \sigma^2_{r(v,t)}\right).$$

Throughout, set $\acute{\varepsilon}_{(v),t} := \hat{\varepsilon}_{r(v,t)}/\mathbf{H}_{(v),tt}$, so that the summand becomes

$$\hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}(y_{r(v,t)}\acute{\varepsilon}_{(v),t} - \sigma^2_{r(v,t)}).$$

Adding and subtracting $\varepsilon^2_{r(v,t)}$,

$$\begin{aligned}\frac{1}{\bar{T}}\sum_t \hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}(y_{r(v,t)}\acute{\varepsilon}_{(v),t} - \sigma^2_{r(v,t)}) =& \frac{1}{\bar{T}}\sum_t \hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}(\varepsilon^2_{r(v,t)} - \sigma^2_{r(v,t)}) \\ &+ \frac{1}{\bar{T}}\sum_t \hat{z}_{1,r(v,t)}\hat{z}'_{1,r(v,t)}(y_{r(v,t)}\acute{\varepsilon}_{(v),t} - \varepsilon^2_{r(v,t)}).\end{aligned} \tag{40}$$

Consider the first term of (40). Conditionally on $\mathbf{X}_{(v)}$, the $\varepsilon^2_{r(v,t)} - \sigma^2_{r(v,t)}$ are mean zero and uncorrelated across t within the subsample (Assumption *Regular Errors*, the normalised errors being i.i.d.). Consider the (k, k') entry. Its conditional second moment is

$$\frac{1}{\bar{T}^2} \sum_{t,s} \hat{z}_{1,r(v,t),k} \hat{z}_{1,r(v,t),k'} \hat{z}_{1,r(v,s),k} \hat{z}_{1,r(v,s),k'} \, \mathbb{E}\left[(\varepsilon^2_{r(v,t)} - \sigma^2_{r(v,t)})(\varepsilon^2_{r(v,s)} - \sigma^2_{r(v,s)}) \;\middle|\; \mathbf{X}_{(v)} \right]$$
$$= \frac{1}{\bar{T}^2} \sum_{t} \hat{z}^2_{1,r(v,t),k} \hat{z}^2_{1,r(v,t),k'} \, \mathrm{Var}\left(\varepsilon^2_{r(v,t)} \mid \mathbf{X}_{(v)} \right),$$

which by $\mathrm{Var}\left(\varepsilon^2_{r(v,t)} \mid \mathbf{X}_{(v)}\right) \leq \mathbb{E}\left[\varepsilon^4_{r(v,t)}\right] \leq C^2$ (Assumption *Regular Errors*) and $\hat{z}^2_{1,r(v,t),k} \hat{z}^2_{1,r(v,t),k'} \leq \|\hat{z}_{1,r(v,t)}\|^4$ is at most

$$\frac{C^2}{\bar{T}} \cdot \frac{1}{\bar{T}} \sum_{t} \|\hat{z}_{1,r(v,t)}\|^4 = \frac{C^2}{\bar{T}} \, O_p(1) = o_p(1),$$

by Lemma 10. Markov's inequality applied entrywise (the matrices are $(h+1) \times (h+1)$, a fixed dimension) gives $\left\| \frac{1}{\bar{T}} \sum_t \hat{z}_{1,r(v,t)} \hat{z}'_{1,r(v,t)} (\varepsilon^2_{r(v,t)} - \sigma^2_{r(v,t)}) \right\|_2 = o_p(1)$.

We now decompose the second term of (40). Define $\mathcal{A}_{(v),tj} := \mathbf{H}_{(v),tj} / \mathbf{H}_{(v),tt}$, the annihilator scaled by its diagonal. Since $y_{r(v,t)} = \beta' x_{r(v,t)} + \varepsilon_{r(v,t)}$, and $\hat{\varepsilon}_{r(v,t)} = \sum_j \mathbf{H}_{(v),tj} \varepsilon_{r(v,j)}$, we have $\acute{\varepsilon}_{(v),t} = \sum_j \mathcal{A}_{(v),tj} \varepsilon_{r(v,j)}$ and therefore

$$y_{r(v,t)} \acute{\varepsilon}_{(v),t} - \varepsilon^2_{r(v,t)} = \varepsilon_{r(v,t)} \sum_{j \neq t} \mathcal{A}_{(v),tj} \varepsilon_{r(v,j)} + \left(\mathcal{A}_{(v),tt} - 1 \right) \varepsilon^2_{r(v,t)} + \beta' x_{r(v,t)} \sum_{j} \mathcal{A}_{(v),tj} \varepsilon_{r(v,j)}. \quad (41)$$

We first note that the middle term is zero because $\mathcal{A}_{(v),tt} = \frac{\mathbf{H}_{(v),tt}}{\mathbf{H}_{(v),tt}} = 1$. Substituting what remains of (41) into the second term of (40) produces two matrix averages, which we denote $\mathcal{T}_1$ and $\mathcal{T}_2$ and bound in turn. We repeatedly use the relation

$$\sum_j \mathcal{A}^2_{(v),tj} = \sum_j \mathbf{H}^2_{(v),tj} / \mathbf{H}^2_{(v),tt} = \mathbf{H}_{(v),tt} / \mathbf{H}^2_{(v),tt} \leq \mathbf{H}^{-2}_{(v),tt} = O_p(1),$$

which follows from idempotency of $\mathbf{H}_{(v)}$ ($\sum_j \mathbf{H}^2_{(v),tj} = \mathbf{H}_{(v),tt}$) and $\mathbf{H}_{(v),tt} \leq 1$, together with $1/\min_t \mathbf{H}_{(v),tt} = O_p(1)$ (Lemma 14).

Consider first the quadratic cross term

$$\mathcal{T}_1 = \frac{1}{\bar{T}} \sum_t \sum_{j \neq t} \hat{z}_{1,r(v,t)} \hat{z}'_{1,r(v,t)} \, \varepsilon_{r(v,t)} \mathcal{A}_{(v),tj} \varepsilon_{r(v,j)},$$

whose conditional variance we now bound. By the uncorrelatedness, the only surviving cross-products in the conditional second moment of each entry of $\mathcal{T}_1$ are the "diagonal" pairings $(t_1, j_1) = (t_2, j_2)$ and the "transposed" pairings $(t_1, j_1) = (j_2, t_2)$, so the conditional second moment of any fixed entry is bounded by

$$\frac{C^2}{\bar{T}^2} \sum_t \sum_{j \neq t} \|\hat{z}_{1,r(v,t)}\|^4 \mathcal{A}^2_{(v),tj} + \frac{C^2}{\bar{T}^2} \sum_t \sum_{j \neq t} \|\hat{z}_{1,r(v,t)}\|^2 \|\hat{z}_{1,r(v,j)}\|^2 |\mathcal{A}_{(v),tj} \mathcal{A}_{(v),jt}|,$$

using $\mathbb{E}\left[\varepsilon^2_{r(v,t)} \varepsilon^2_{r(v,j)} \mid \mathbf{X}_{(v)}\right] \leq C^2$. For the first sum, $\sum_{j \neq t} \mathcal{A}^2_{(v),tj} \leq \mathbf{H}^{-2}_{(v),tt}$, so it is at most

$$\frac{C^2}{\bar{T}} \Big(\max_t \frac{1}{\mathbf{H}^2_{(v),tt}} \Big) \cdot \frac{1}{\bar{T}} \sum_t \|\hat{z}_{1,r(v,t)}\|^4 = \frac{1}{\bar{T}} \, O_p(1) \, O_p(1) = o_p(1),$$

by $1/\min_t \mathbf{H}_{(v),tt} = O_p(1)$ and Lemma 10. For the second, note that $\mathcal{A}_{(v),tj} \mathcal{A}_{(v),jt} = \mathbf{H}^2_{(v),tj} / (\mathbf{H}_{(v),tt} \mathbf{H}_{(v),jj})$. Bounding $\|\hat{z}_{1,r(v,j)}\|^2$ by its maximum over j and $1/(\mathbf{H}_{(v),tt} \mathbf{H}_{(v),jj})$ by $1/\min_t \mathbf{H}^2_{(v),tt}$, and using that, by the symmetry and idempotency of $\mathbf{H}_{(v)}$, $\sum_j \mathbf{H}^2_{(v),tj} = \mathbf{H}_{(v),tt} \leq 1$, we obtain

$$\begin{aligned}
&\frac{C^2}{\bar{T}^2} \sum_t \sum_{j \neq t} \|\hat{z}_{1,r(v,t)}\|^2 \|\hat{z}_{1,r(v,j)}\|^2 |\mathcal{A}_{(v),tj} \mathcal{A}_{(v),jt}| \\
&\quad \leq \frac{C^2}{\bar{T}^2} \, \frac{\max_t \|\hat{z}_{1,r(v,t)}\|^2}{\min_t \mathbf{H}^2_{(v),tt}} \sum_t \|\hat{z}_{1,r(v,t)}\|^2 \sum_j \mathbf{H}^2_{(v),tj} \\
&\quad \leq C^2 \, \frac{\max_t \|\hat{z}_{1,r(v,t)}\|^2}{\bar{T}} \cdot \frac{1}{\min_t \mathbf{H}^2_{(v),tt}} \cdot \frac{1}{\bar{T}} \sum_t \|\hat{z}_{1,r(v,t)}\|^2 = o_p(1),
\end{aligned}$$

by Lemmas 10 and 13.

It remains to bound the term involving $\beta' x_{r(v,t)}$,

$$\mathcal{T}_2 = \frac{1}{\bar{T}} \sum_t \hat{z}_{1,r(v,t)} \hat{z}'_{1,r(v,t)} \, \beta' x_{r(v,t)} \sum_j \mathcal{A}_{(v),tj} \varepsilon_{r(v,j)},$$

which has conditional mean zero (again $\mathbb{E}\left[\varepsilon_{r(v,j)} \mid \mathbf{X}_{(v)}\right] = 0$). Fix the (k, k') entry. Its conditional variance is

$$\operatorname{Var}\left(\frac{1}{\bar{T}}\sum_t \hat{z}_{1,r(v,t),k}\hat{z}_{1,r(v,t),k'}\beta' x_{r(v,t)} \textstyle\sum_j \mathcal{A}_{(v),tj}\varepsilon_{r(v,j)} \;\middle|\; \mathbf{X}_{(v)}\right) = \frac{1}{\bar{T}^2}\sum_j \sigma^2_{r(v,j)}\Big(\sum_t \hat{z}_{1,r(v,t),k}\hat{z}_{1,r(v,t),k'}\beta' x_{r(v,t)}\mathcal{A}_{(v),tj}\Big)^2.$$

Bound $\sigma^2_{r(v,j)} \le C$ and use the quadratic-form structure. Writing

$$a_t := \hat{z}_{1,r(v,t),k}\hat{z}_{1,r(v,t),k'}\beta' x_{r(v,t)},$$

the inner double sum is

$$\sum_j\Big(\sum_t a_t\mathcal{A}_{(v),tj}\Big)^2 = \sum_{j,t,s} a_t a_s \mathcal{A}_{(v),tj}\mathcal{A}_{(v),sj} = \sum_{t,s} a_t a_s \sum_j \mathcal{A}_{(v),tj}\mathcal{A}_{(v),sj} = \sum_{t,s} a_t a_s \frac{\mathbf{H}_{(v),ts}}{\mathbf{H}_{(v),tt}\mathbf{H}_{(v),ss}}.$$

Writing $b_t := a_t/\mathbf{H}_{(v),tt}$ and using that $\mathbf{H}_{(v)}$ is a projection,

$$\sum_{t,s} a_t a_s \frac{\mathbf{H}_{(v),ts}}{\mathbf{H}_{(v),tt}\mathbf{H}_{(v),ss}} = \sum_{t,s} b_t \mathbf{H}_{(v),ts} b_s \le \sum_t b_t^2 \le \Big(\max_t \frac{1}{\mathbf{H}^2_{(v),tt}}\Big)\sum_t a_t^2.$$

Thus the conditional variance is at most

$$\frac{C}{\bar{T}^2}\Big(\max_t \frac{1}{\mathbf{H}^2_{(v),tt}}\Big)\sum_t a_t^2 = \frac{C}{\bar{T}}\,O_p(1)\cdot\frac{1}{\bar{T}}\sum_t \hat{z}^2_{1,r(v,t),k}\hat{z}^2_{1,r(v,t),k'}(\beta' x_{r(v,t)})^2.$$

Finally $\frac{1}{\bar{T}^2}\sum_t \hat{z}^2_{1,r(v,t),k}\hat{z}^2_{1,r(v,t),k'}(\beta' x_{r(v,t)})^2 \le \frac{\max_t(\beta' x_{r(v,t)})^2}{\bar{T}}\cdot\frac{1}{\bar{T}}\sum_t \|\hat{z}_{1,r(v,t)}\|^4 = o_p(1)\cdot O_p(1)$ by Lemma 10 under Assumption *Bounded Leverage* and

$$\max_t(\beta' x_{r(v,t)})^2 = o_p(\bar{T}),$$

which follows from Assumption *Bounded Fitted Values*. Indeed, by Assumption *Feedback Decomposition*, $\beta' x_{r(v,t)} = \tilde{x}'_{r(v,t)}\beta + \varpi_{r(v,t)}$ with the feedback component $\varpi_s = \sum_{l=1}^h \varepsilon_{s-l}\alpha_l'\beta$

of Section 3, so

$$\max_t(\beta' x_{r(v,t)})^2 \le 2\max_{1\le s\le T}(\beta'\tilde{x}_s)^2 + 2\Big(\sum_{l=1}^{h}|\alpha_l'\beta|\Big)^2 \max_{1-h\le s\le T}\varepsilon_s^2.$$

The first term is $o(T) = o(\bar{T})$, h being fixed. In the second, $\sum_l |\alpha_l'\beta| = o(1)$, and $\max_s \varepsilon_s^2 = o_p(T)$: for any $\delta > 0$, the union bound, Markov's inequality and $\mathbb{E}[\varepsilon_s^4] \le C^2$ (Assumption *Regular Errors*) give $\mathbb{P}[\max_s \varepsilon_s^2 > \delta T] \le (T+h)C^2/(\delta^2 T^2) \to 0$.

The conditional variance of each part ($\mathcal{T}_1$ and $\mathcal{T}_2$) vanishes, implying that the second term in (40) is $o_p(1)$ in spectral norm as well. As both terms on the right-hand side of (40) are $o_p(1)$ in spectral norm, the triangle inequality gives $\|\acute{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)}\|_2/\bar{T} = o_p(1)$.

This establishes (39), so that Lemma 11 applies and yields $\frac{\sum_{v=1}^{h+1}\acute{\Sigma}_v}{\sum_{v=1}^{h+1}\Sigma_v} \xrightarrow{p} 1$. □

Lemma 16 (Lower bound on the diagonal of $\mathbf{M}_{(v)}$)**.** *Under Assumptions* Feedback Decomposition *and* Sub-One-Half Leverage*, for every subsample v and every position t,*

$$\mathbf{M}_{(v),tt} \;\ge\; 1 - \tilde{x}_{r(v,t)}'(\tilde{\mathbf{X}}_{(v)}'\tilde{\mathbf{X}}_{(v)})^{-1}\tilde{x}_{r(v,t)} \;\ge\; \frac{1}{2} + c_\kappa,$$

deterministically.

Proof. The first inequality follows from $\mathrm{col}(\mathbf{Z}_{2,(v)}) \subseteq \mathrm{col}(\tilde{\mathbf{X}}_{(v)})$, as $\mathbf{M}_{(v)}$ then dominates the annihilator of $\tilde{\mathbf{X}}_{(v)}$ in the Loewner order. The second inequality is Assumption *Sub-One-Half Leverage*. □

Lemma 17 (Lower bound on the diagonal of $\mathbf{H}_{(v)}$)**.** *Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*,* No Influential Observations*,* Bounded Leverage *and* Sub-One-Half Leverage*,*

$$\mathbb{P}\left[\min_{1\le t\le \bar{T}} \mathbf{H}_{(v),tt} \ge \frac{1}{2} + \frac{c_\kappa}{2}\right] \to 1.$$

Proof. By the Frisch–Waugh–Lovell decomposition of the full leverage into its nuisance- and contrast-block contributions (as in the proof of Lemma 14),

$$\mathbf{H}_{(v),tt} = \mathbf{M}_{(v),tt} - \hat{z}_{1,r(v,t)}'(\mathbf{Z}_{1,(v)}'\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}\hat{z}_{1,r(v,t)},$$

where we recall that $\hat{z}'_{1,r(v,t)}$ is row t of $\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}$. Hence

$$\max_t \hat{z}'_{1,r(v,t)}(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)})^{-1}\hat{z}_{1,r(v,t)} \leq \frac{\max_t \|\hat{z}_{1,r(v,t)}\|^2}{\bar{T}\,\lambda_{\min}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}/\bar{T}\right)} = \frac{o_p(\bar{T})}{\bar{T}\,(1-o_p(1))} = o_p(1),$$

using $\max_t \|\hat{z}_{1,r(v,t)}\|^2 = o_p(\bar{T})$ from Lemma 13, which requires only Assumptions *Bounded Leverage* and *No Influential Observations* and not vanishing leverage, and Lemma 3. Combining with $\mathbf{M}_{(v),tt} \geq \frac{1}{2} + c_\kappa$ (Lemma 16) gives the claim. □

Lemma 18 (Boundedness of $\boldsymbol{\kappa}_{(v)}$)**.** *On the event* $\{\min_t \mathbf{H}_{(v),tt} \geq \frac{1}{2} + \frac{c_\kappa}{2}\}$ *of Lemma 17,* $\mathbf{G}_{(v)}$ *is invertible and*

$$\left\|\boldsymbol{\kappa}_{(v)}\right\|_2 \;\leq\; \frac{2}{c_\kappa}.$$

In particular, $\boldsymbol{\kappa}_{(v)}$ *is well defined with probability approaching one and* $\left\|\boldsymbol{\kappa}_{(v)}\right\|_2 = O_p(1)$*. (On the complement of this event, whose probability vanishes, define* $\grave{\Sigma}_v$ *arbitrarily.)*

Proof. Recall that $\mathbf{G}_{(v)} = \mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$ and $\boldsymbol{\kappa}_{(v)} = \mathbf{G}_{(v)}^{-1}$. Since $\mathbf{H}_{(v)}$ is symmetric and idempotent, $\sum_s \mathbf{H}^2_{(v),ts} = (\mathbf{H}^2_{(v)})_{tt} = \mathbf{H}_{(v),tt}$, so the off-diagonal row sums are $\sum_{s\neq t} \mathbf{G}_{(v),ts} = \mathbf{H}_{(v),tt} - \mathbf{H}^2_{(v),tt}$. The dominance margin in row t is therefore

$$\mathbf{G}_{(v),tt} - \sum_{s\neq t} \mathbf{G}_{(v),ts} = \mathbf{H}^2_{(v),tt} - \big(\mathbf{H}_{(v),tt} - \mathbf{H}^2_{(v),tt}\big) = \mathbf{H}_{(v),tt}\big(2\mathbf{H}_{(v),tt} - 1\big) \;\geq\; \Big(\tfrac{1}{2} + \tfrac{c_\kappa}{2}\Big)c_\kappa \;\geq\; \tfrac{c_\kappa}{2},$$

which is positive precisely when $\mathbf{H}_{(v),tt} > 1/2$. As $\mathbf{G}_{(v)}$ is symmetric, Lemma 41 gives $\left\|\mathbf{G}_{(v)}^{-1}\right\|_2 \leq 2/c_\kappa$. □

Lemma 19 (Consistency of $\sum_{v=1}^{h+1} \grave{\Sigma}_v$)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, No Influential Observations, Bounded Leverage *and* Sub-One-Half Leverage, *for each* $v \in \{1, \ldots, h+1\}$, $\mathbf{G}_{(v)}$ *is invertible with probability approaching one, and*

$$\frac{\sum_{v=1}^{h+1} \grave{\Sigma}_v}{\sum_{v=1}^{h+1} \Sigma_v} \xrightarrow{p} 1.$$

Proof. Let $\grave{\mathbf{\Psi}}_{(v)} := \mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\grave{\mathbf{\Phi}}^2_{(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}$, where $\grave{\mathbf{\Phi}}^2_{(v)} := \operatorname{diag}\big(\boldsymbol{\kappa}_{(v)}\hat{\varepsilon}^{\odot 2}_{(v)}\big)$. We show that

$$\left\| \frac{1}{\bar{T}}(\grave{\mathbf{\Psi}}_{(v)} - \mathbf{\Psi}_{(v)}) \right\|_2 \xrightarrow{p} 0. \tag{42}$$

We assume that $\mathbf{G}_{(v)}$ is invertible, an event whose probability tends to one under Assumption *Sub-One-Half Leverage* by Lemmas 17 and 18. As explained in the main text, $\mathbb{E}\left[\grave{\mathbf{\Phi}}^2_{(v)} \mid \mathbf{X}_{(v)}\right] = \mathbf{\Phi}^2_{(v)}$, so that

$$\mathbb{E}\left[\grave{\mathbf{\Psi}}_{(v)} \mid \mathbf{X}_{(v)}\right] = \mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)} = \mathbf{\Psi}_{(v)}.$$

As this matrix is of fixed dimension, it suffices to show that each entry concentrates around its conditional mean at rate $o_p(\bar{T})$. Fix $k, k' \in \{1, \ldots, h+1\}$ and let a be the $\bar{T}$-vector with entries $a_t := \hat{z}_{1,r(v,t),k}\,\hat{z}_{1,r(v,t),k'}$, so that, using the symmetry of $\boldsymbol{\kappa}_{(v)}$,

$$\big(\grave{\mathbf{\Psi}}_{(v)}\big)_{kk'} = \sum_t a_t \big(\boldsymbol{\kappa}_{(v)}\hat{\varepsilon}^{\odot 2}_{(v)}\big)_t = b'\hat{\varepsilon}^{\odot 2}_{(v)} = \sum_j b_j\,\hat{\varepsilon}^2_{r(v,j)} = (\mathbf{S}_{(v)}\varepsilon)'\,\mathbf{E}\,\mathbf{S}_{(v)}\varepsilon,$$

with $b := \boldsymbol{\kappa}_{(v)}a$ and $\mathbf{E} := \mathbf{H}_{(v)}\operatorname{diag}(b)\,\mathbf{H}_{(v)}$, where the last step substituted $\hat{\varepsilon}_{(v)} = \mathbf{H}_{(v)}\mathbf{S}_{(v)}\varepsilon$. Both b and $\mathbf{H}_{(v)}$ are $\mathbf{X}_{(v)}$-measurable, and $\mathbf{S}_{(v)}\varepsilon$ is independent of $\mathbf{X}_{(v)}$ with independent entries, variances in $[c, C]$ and fourth moments bounded by C^2 (Assumption *Regular Errors*). We require the following two bounds.

(a)
$$\begin{aligned}\|a\|^2 &= \sum_t \left(\hat{z}_{1,r(v,t),k}\hat{z}_{1,r(v,t),k'}\right)^2 \\ &\le \sqrt{\sum_t \hat{z}^4_{1,r(v,t),k}\sum_t \hat{z}^4_{1,r(v,t),k'}} \\ &\le \sum_t \|\hat{z}_{1,r(v,t)}\|^4 = O_p(\bar{T}),\end{aligned}$$

by Lemma 10 under Assumption *Bounded Leverage*.

(b) $\|b\| \le \left\|\boldsymbol{\kappa}_{(v)}\right\|_2 \|a\| \le \frac{2}{c_\kappa}\|a\| = O_p(\bar{T}^{1/2})$, by Lemma 18 and (a).

By Lemma 2(ii) applied conditionally on $\mathbf{X}_{(v)}$ to the quadratic form $(\mathbf{S}_{(v)}\varepsilon)'\mathbf{E}\,\mathbf{S}_{(v)}\varepsilon$,

$$\big(\dot{\mathbf{\Psi}}_{(v)}\big)_{kk'} - \mathbb{E}\left[\big(\dot{\mathbf{\Psi}}_{(v)}\big)_{kk'} \mid \mathbf{X}_{(v)}\right] = O_p\big(\,\|\mathbf{E}\|_F\,\big),$$

where

$$\|\mathbf{E}\|_F \le \big\|\mathbf{H}_{(v)}\big\|_2^2\,\|\mathrm{diag}(b)\|_F = \|b\| = O_p(\bar{T}^{1/2}),$$

using $\big\|\mathbf{H}_{(v)}\big\|_2 \le 1$ and (b). This proves (42), which is condition (34) of Lemma 11 for the meat $\hat{\mathbf{\Psi}}_{(v)} = \dot{\mathbf{\Psi}}_{(v)}$. Lemma 11 therefore gives $\big(\sum_{v=1}^{h+1}\dot{\Sigma}_v\big)/\sum_{v=1}^{h+1}\Sigma_v \xrightarrow{p} 1$. □

A.4 Conservative inference

The results of this subsection establish the upper bound $1+\Lambda$ on the ratio $\Sigma/\sum_{v=1}^{h+1}\Sigma_v$, and prove the two theorems on conservative inference.

Lemma 20 (The design factor Λ)**.** *Under Assumptions* Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime, *the cross term of Lemma 7 has the closed form*

$$\mathrm{tr}\,\big(\mathbf{B}^2\big) = 2\sum_{1\le v<q\le h+1}\gamma_{v,l(v,q)}\,\gamma_{q,m(v,q)}\,c_{v,q}. \tag{43}$$

Moreover, $\mathrm{tr}\,(\mathbf{B}^2) \le \lambda\,\mathrm{tr}\,(\mathbf{B}\mathbf{B}')$*, and* λ *is the smallest constant with this property uniformly over the vector* $(\gamma_{v,l})$*. Consequently*

$$\Sigma \;\le\; \sum_{v=1}^{h+1}\Sigma_v + \lambda\,\mathrm{tr}\,(\mathbf{B}\mathbf{B}') \;\le\; (1+\Lambda)\sum_{v=1}^{h+1}\Sigma_v. \tag{44}$$

Proof. The closed form (43) is the expression for $\mathrm{tr}\,(\mathbf{B}^2)$ in Lemma 7(ii), with $c_{v,q} = \mathrm{tr}\,\big(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(q)}\mathbf{\Phi}^2_{(q)}\mathbf{J}\big)$, $l(v,q) = h+1-(q-v)$ and $m(v,q) = q-v$.

To establish the bound and its sharpness, set, for $1 \le v < q \le h+1$, $g_{v,q} := \gamma_{v,l(v,q)}\sqrt{d_{v,q}}$ and $g_{q,v} := \gamma_{q,m(v,q)}\sqrt{d_{q,v}}$, and let g denote the vector of these $h(h+1)$ entries, one for each ordered pair of distinct subsamples. Each $\gamma_{v,l}$ enters exactly one entry of g (see the regrouping following (9)), and $\mathrm{tr}\,(\mathbf{B}\mathbf{B}') = \|g\|^2$ by (9). With $\pi_{v,q} := c_{v,q}/\sqrt{d_{v,q}d_{q,v}}$, the closed

form (43) becomes

$$\operatorname{tr}\left(\mathbf{B}^2\right) = 2 \sum_{1 \le v < q \le h+1} \pi_{v,q}\, g_{v,q}\, g_{q,v} = g' \mathbf{N} g,$$

where $\mathbf{N}$ is the symmetric matrix with entries $\mathbf{N}_{(v,q),(q,v)} = \mathbf{N}_{(q,v),(v,q)} = \pi_{v,q}$ for $v < q$ and zeros elsewhere. Consequently, $\sup_{\gamma \neq 0} \operatorname{tr}(\mathbf{B}^2) / \operatorname{tr}(\mathbf{B}\mathbf{B}') = \lambda_{\max}(\mathbf{N})$. In $\mathbf{N}$, each ordered pair (v, q) is linked only to the reversed pair (q, v). The $h(h+1)$ indices therefore decompose into $h(h+1)/2$ disjoint pairs, and after a permutation $\mathbf{N}$ is block diagonal with 2×2 blocks $\left(\begin{smallmatrix} 0 & \pi_{v,q} \\ \pi_{v,q} & 0 \end{smallmatrix}\right)$, whose eigenvalues are $\pm|\pi_{v,q}|$. Hence $\lambda_{\max}(\mathbf{N}) = \max_{v<q} |\pi_{v,q}| = \lambda$. The supremum is attained by any γ whose only non-zero entries are $\gamma_{v,l(v,q)}$ and $\gamma_{q,m(v,q)}$ for a maximising pair (v, q), chosen such that $g_{v,q} = \operatorname{sgn}(\pi_{v,q})\, g_{q,v}$.

It remains to prove (44). By Lemma 7,

$$\Sigma = \sum_{v=1}^{h+1} \Sigma_v + \operatorname{tr}\left(\mathbf{B}^2\right) \le \sum_{v=1}^{h+1} \Sigma_v + \lambda \operatorname{tr}(\mathbf{B}\mathbf{B}').$$

Moreover, $\operatorname{tr}(\mathbf{B}^2) \le \operatorname{tr}(\mathbf{B}\mathbf{B}')$ by the Cauchy–Schwarz inequality, so that λ may be replaced by $\Lambda = \min\{1, \lambda\}$. Finally, $\operatorname{tr}(\mathbf{B}\mathbf{B}') \le \sum_{v=1}^{h+1} \Sigma_v$ because $\sum_{v=1}^{h+1} \Sigma_v = \|\omega\|^2 + \operatorname{tr}(\mathbf{B}\mathbf{B}')$. □

Lemma 21 (The weight arrays of $c_{v,q}$ and $d_{v,q}$ are admissible)**.** *Let* $\mathbf{U}$ *and* $\mathbf{Y}$ *be symmetric idempotent* $\bar{T} \times \bar{T}$ *matrices. Then both*

$$\mathbf{W}^d := \mathbf{U} \odot \mathbf{U} \qquad \textit{and} \qquad \mathbf{W}^c := (\mathbf{U}\mathbf{J}') \odot \mathbf{Y}$$

are admissible. In particular this holds for $(\mathbf{U}, \mathbf{Y}) = (\mathbf{M}_{(v)}, \mathbf{M}_{(q)})$ *and for* $(\mathbf{U}, \mathbf{Y}) = (\mathbf{H}_{(v)}, \mathbf{H}_{(q)})$.

Proof. Symmetry and idempotency give $\sum_j \mathbf{U}_{tj}^2 = \mathbf{U}_{tt} \le 1$ for every t, hence $|\mathbf{U}_{tj}| \le 1$, as well as $\|\mathbf{U}\|_2 \le 1$. The same holds for $\mathbf{Y}$. Consider first $\mathbf{W}^d$. Condition (W1) holds because $\sum_j \mathbf{U}_{tj}^2 = \mathbf{U}_{tt} \le 1$, and the column sums are bounded in the same way by symmetry. Condition (W2) follows from $\left\|\mathbf{W}^d\right\|_F^2 = \sum_{t,j} \mathbf{U}_{tj}^4 \le \sum_{t,j} \mathbf{U}_{tj}^2 = \operatorname{tr}(\mathbf{U}) \le \bar{T}$. Since the row and column norms of $\mathbf{U}$ equal $\sqrt{\mathbf{U}_{tt}} \le 1$, Lemma 42(iii) yields $\left\|\mathbf{W}^d\right\|_2 \le 1$, which is (W3).

Turning to $\mathbf{W}^c$, observe that $(\mathbf{U}\mathbf{J}')_{tj} = \mathbf{U}_{t,j+1}$ for $j \le \bar{T} - 1$ and that the last column of

$\mathbf{UJ}'$ is zero, because the last column of $\mathbf{J}'$ is zero. By the Cauchy–Schwarz inequality,

$$\sum_t \left|\mathbf{W}^c_{tj}\right| \le \left(\sum_t \mathbf{U}^2_{t,j+1}\right)^{1/2} \left(\sum_t \mathbf{Y}^2_{tj}\right)^{1/2} = \mathbf{U}^{1/2}_{j+1,j+1}\mathbf{Y}^{1/2}_{jj} \le 1,$$

and the row sums are bounded analogously, which establishes (W1). Condition (W2) follows from $\|\mathbf{W}^c\|_F^2 \le \max_{t,j} \mathbf{U}^2_{t,j+1} \sum_{t,j} \mathbf{Y}^2_{tj} \le \operatorname{tr}(\mathbf{Y}) \le \bar{T}$. Finally, each row of $\mathbf{UJ}'$ is a row of $\mathbf{U}$ with one entry removed and therefore has norm at most one, and the same bound applies to the columns of $\mathbf{Y}$. Lemma 42(iii) thus gives $\|\mathbf{W}^c\|_2 \le 1$, which is (W3). □

Lemma 22 (Properties of $\check{\sigma}^2_{(u)}$)**.** *Let $u \in \{1-h, \dots, h+1\}$ be a block and u^* the subsample it belongs to. Under Assumptions* Feedback Decomposition*,* Regular Errors *and* Sub-One-Half Leverage*:*

(i) for a subsample $u \ge 1$, $\mathbf{M}_{(u)} \odot \mathbf{M}_{(u)}$ is invertible with $\left\|\check{\boldsymbol{\kappa}}_{(u)}\right\|_2 \le 1/c_\kappa$, deterministically;

(ii) $\check{\sigma}^2_{(u)}$ is a function of $\mathbf{S}_{(u^*)}\varepsilon$ alone. In particular $\check{\sigma}^2_{(u)}$ and $\check{\sigma}^2_{(u')}$ are independent whenever $u^* \ne u'^*$;

(iii) $\mathbb{E}\left[\check{\sigma}^2_{(u)}\right] = \sigma^2_{(u)}$ exactly;

(iv) $\left\|\check{\mathbf{\Gamma}}_{(u)}\right\|_2 \le 6C^2/c_\kappa^2$, and $\left\|\check{\mathbf{\Gamma}}_{(u)}\right\|_2 \le \left\|\check{\mathbf{\Gamma}}_{(u^*)}\right\|_2$ for $u \le 0$.

Proof. (i) Let $v := u \ge 1$. By Lemma 16, $\mathbf{M}_{(v),tt} \ge \frac{1}{2} + c_\kappa$ for every t. Since $\dot{\mathbf{G}}_{(v)} = \mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}$ and $\mathbf{M}_{(v)}$ is idempotent, $\sum_{s \ne t} \dot{\mathbf{G}}_{(v),ts} = \mathbf{M}_{(v),tt} - \mathbf{M}^2_{(v),tt}$, so the diagonal dominance margin of row t is

$$\dot{\mathbf{G}}_{(v),tt} - \sum_{s \ne t} \dot{\mathbf{G}}_{(v),ts} = \mathbf{M}_{(v),tt}\left(2\mathbf{M}_{(v),tt} - 1\right) \ge \left(\tfrac{1}{2} + c_\kappa\right) 2c_\kappa \ge c_\kappa.$$

As $\dot{\mathbf{G}}_{(v)}$ is symmetric, Lemma 41 gives $\left\|\dot{\mathbf{G}}^{-1}_{(v)}\right\|_2 \le 1/c_\kappa$. All quantities involved are deterministic, since $\mathbf{M}_{(v)}$ is deterministic by Lemma 1.

(ii) and (iii) Consider first a subsample $v := u \ge 1$. Since $\mathbf{M}_{(v)}$ and $\check{\boldsymbol{\kappa}}_{(v)}$ are deterministic, $\check{\sigma}^2_{(v)}$ is a fixed function of $\mathbf{S}_{(v)}\varepsilon$, and the blocks of ε selected by different subsamples are independent. For the mean, $\mathbb{E}\left[\dot{\varepsilon}^2_{(v),t}\right] = \sum_j \mathbf{M}^2_{(v),tj}\sigma^2_{r(v,j)} = \left(\left(\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right)\sigma^2_{(v)}\right)_t$, so $\mathbb{E}\left[\check{\sigma}^2_{(v)}\right] = \check{\boldsymbol{\kappa}}_{(v)}\left(\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right)\sigma^2_{(v)} = \sigma^2_{(v)}$.

The lagged blocks are treated at the end of the proof.

(iv) Consider again a subsample $v := u \geq 1$. Since $\check{\sigma}^2_{(v)} = \check{\boldsymbol{\kappa}}_{(v)} \dot{\varepsilon}^{\odot 2}_{(v)}$ with both $\mathbf{M}_{(v)}$ and $\check{\boldsymbol{\kappa}}_{(v)}$ deterministic, the matrix $\check{\boldsymbol{\Gamma}}_{(v)} = \check{\boldsymbol{\kappa}}_{(v)} \operatorname{Var}\left(\dot{\varepsilon}^{\odot 2}_{(v)}\right) \check{\boldsymbol{\kappa}}_{(v)}$ is deterministic. For every period t put $\chi_t := \mathbb{E}[\varepsilon_t^4] - 3\sigma_t^4$, and let $\chi_{(v)}$ collect the $\chi_{r(v,j)}$. Fix two positions t and i and expand $\dot{\varepsilon}^2_{(v),t} \dot{\varepsilon}^2_{(v),i}$ as a fourfold sum over the independent, mean-zero $\varepsilon_{r(v,j)}$. A term has non-zero expectation only if no index appears exactly once, which leaves the patterns $j = j' \neq k = k'$, $j = k \neq j' = k'$, $j = k' \neq j' = k$ and $j = j' = k = k'$, of which the middle two contribute equally. Subtracting $\mathbb{E}\left[\dot{\varepsilon}^2_{(v),t}\right] \mathbb{E}\left[\dot{\varepsilon}^2_{(v),i}\right]$ and collecting the diagonal terms gives

$$\operatorname{Var}\left(\dot{\varepsilon}^{\odot 2}_{(v)}\right) = 2\,\boldsymbol{\Omega}_{(v)} \odot \boldsymbol{\Omega}_{(v)} + \left(\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right) \operatorname{diag}(\chi_{(v)}) \left(\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right), \tag{45}$$

where $\boldsymbol{\Omega}_{(v)} := \operatorname{Var}\left(\dot{\varepsilon}_{(v)}\right) = \mathbf{M}_{(v)} \boldsymbol{\Phi}^2_{(v)} \mathbf{M}_{(v)}$. For the first term, $\boldsymbol{\Omega}_{(v)} \succeq 0$, and by idempotency $\boldsymbol{\Omega}_{(v),tt} = \sum_j \mathbf{M}^2_{(v),tj} \sigma^2_{r(v,j)} \leq C \mathbf{M}_{(v),tt} \leq C$, while $\lambda_{\max}\left(\boldsymbol{\Omega}_{(v)}\right) \leq \left\|\mathbf{M}_{(v)}\right\|_2^2 \left\|\boldsymbol{\Phi}^2_{(v)}\right\|_2 \leq C$. Lemma 42(ii) then gives $\left\|2\boldsymbol{\Omega}_{(v)} \odot \boldsymbol{\Omega}_{(v)}\right\|_2 \leq 2C^2$. For the second, $|\chi_t| \leq \mathbb{E}[\varepsilon_t^4] + 3\sigma_t^4 \leq 4C^2$ by Assumption *Regular Errors*, and $\mathbf{M}_{(v)} \succeq 0$ with $\mathbf{M}_{(v),tt} \leq 1$, so Lemma 42(ii) gives $\left\|\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right\|_2 \leq 1$ and the second term has spectral norm at most $4C^2$. Hence $\left\|\operatorname{Var}\left(\dot{\varepsilon}^{\odot 2}_{(v)}\right)\right\|_2 \leq 6C^2$ and, with (i), $\left\|\check{\boldsymbol{\Gamma}}_{(v)}\right\|_2 \leq 6C^2/c_\kappa^2$.

It remains to establish (ii)–(iv) for a lagged block $u \leq 0$. Since $\mathbf{J}'$ is the $\bar{T} \times \bar{T}$ matrix with ones on the first subdiagonal and zeros elsewhere, by definition $\check{\sigma}^2_{(u)} = \mathbf{J}' \check{\sigma}^2_{(u^*)} + \sigma_u^2 e_1$. For (ii), the entries at positions $t \geq 2$ are entries of $\check{\sigma}^2_{(u^*)}$ and the entry at position one is a constant, so the vector is a function of $\mathbf{S}_{(u^*)}\varepsilon$ alone. For (iii), those entries have mean $\sigma^2_{r(u^*,t-1)} = \sigma^2_{(u),t}$ and position one is the constant σ_u^2. For (iv), the added term is constant, so $\check{\boldsymbol{\Gamma}}_{(u)} = \mathbf{J}' \check{\boldsymbol{\Gamma}}_{(u^*)} \mathbf{J}$, which is deterministic with $\left\|\check{\boldsymbol{\Gamma}}_{(u)}\right\|_2 \leq \|\mathbf{J}'\|_2^2 \left\|\check{\boldsymbol{\Gamma}}_{(u^*)}\right\|_2 \leq \left\|\check{\boldsymbol{\Gamma}}_{(u^*)}\right\|_2$. □

Lemma 23 (Fourth moments of the residuals)**.** *Let Assumptions* Feedback Decomposition, Bounds, Regular Errors *and* Proportional Regime *hold. Then there is a constant C', depending only on C, such that, for every subsample v and every position $t \leq \bar{T}$,*

$$\mathbb{E}\left[\dot{\varepsilon}^4_{(v),t}\right] \leq C' \qquad \textit{and} \qquad \mathbb{E}\left[\hat{\varepsilon}^4_{(v),t}\right] \leq C'.$$

Consequently $\|\dot{\varepsilon}^{\odot 2}_{(v)}\| = O_p(\bar{T}^{1/2})$ and $\|\hat{\varepsilon}^{\odot 2}_{(v)}\| = O_p(\bar{T}^{1/2})$.

Proof. The proof relies on Rosenthal's inequality (29), stated in the proof of Lemma 10, with constant C_p.

Consider first $\dot{\varepsilon}_{(v)}$. Since $\mathbf{M}_{(v)}$ is deterministic, $\dot{\varepsilon}_{(v),t} = \sum_j \mathbf{M}_{(v),tj}\varepsilon_{r(v,j)}$ is a linear form in independent mean-zero random variables with deterministic coefficients. Both sums in (29) with $p = 4$ are controlled by the idempotency identity $\sum_j \mathbf{M}^2_{(v),tj} = \mathbf{M}_{(v),tt} \leq 1$, which implies $|\mathbf{M}_{(v),tj}| \leq 1$ and hence $\mathbf{M}^4_{(v),tj} \leq \mathbf{M}^2_{(v),tj}$. Using $\mathbb{E}[\varepsilon_t^4] \leq C^2$ and $\sigma_t^2 \leq C$ from Assumption *Regular Errors*, we obtain

$$\sum_j \mathbb{E}\left[\left|\mathbf{M}_{(v),tj}\varepsilon_{r(v,j)}\right|^4\right] \leq C^2 \sum_j \mathbf{M}^2_{(v),tj} \leq C^2, \qquad \sum_j \mathbb{E}\left[\mathbf{M}^2_{(v),tj}\varepsilon^2_{r(v,j)}\right] \leq C\,\mathbf{M}_{(v),tt} \leq C.$$

Substituting into (29) gives $\mathbb{E}\left[\dot{\varepsilon}^4_{(v),t}\right] \leq C_4(C^2 + C^2) = 2C_4C^2 =: C'$, uniformly in t and v.

Turning to the feasible residuals, $\hat{\varepsilon}_{(v),t} = \sum_j \mathbf{H}_{(v),tj}\varepsilon_{r(v,j)}$ has random coefficients, so that (29) cannot be applied directly. It applies, however, conditionally on $\mathbf{X}_{(v)}$: the matrix $\mathbf{H}_{(v)}$ is a function of $\mathbf{X}_{(v)}$ alone, and $\mathbf{S}_{(v)}\varepsilon$ is independent of $\mathbf{X}_{(v)}$, because by Assumption *Feedback Decomposition* the feedback in $\mathbf{X}_{(v)}$ depends only on the errors of the h blocks preceding v. Conditionally on $\mathbf{X}_{(v)}$, the coefficients are fixed, the errors remain independent with mean zero, and, by idempotency of $\mathbf{H}_{(v)}$, $\sum_j \mathbf{H}^2_{(v),tj} = \mathbf{H}_{(v),tt} \leq 1$. The two bounds above therefore hold with $\mathbf{H}_{(v)}$ in place of $\mathbf{M}_{(v)}$, which yields $\mathbb{E}\left[\hat{\varepsilon}^4_{(v),t} \mid \mathbf{X}_{(v)}\right] \leq C'$ with a constant that does not depend on t or on the realisation of $\mathbf{X}_{(v)}$. The law of iterated expectations then gives $\mathbb{E}\left[\hat{\varepsilon}^4_{(v),t}\right] \leq C'$.

Finally, summing the pointwise bounds over t gives

$$\mathbb{E}\left[\sum_t \dot{\varepsilon}^4_{(v),t}\right] \leq C'\bar{T} \qquad \text{and} \qquad \mathbb{E}\left[\sum_t \hat{\varepsilon}^4_{(v),t}\right] \leq C'\bar{T},$$

so that both sums are $O_p(\bar{T})$ by Markov's inequality, and their square roots are $O_p(\bar{T}^{1/2})$. □

Lemma 24 (The two annihilators are close)**.** *Let Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, No Influential Observations *and* Bounded Leverage *hold. Then* $\eta = o_p(1)$. *Moreover, for every* $v \in \{1, \ldots, h+1\}$, $\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\| = O_p(1)$ *and*

$$\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\|_\infty \leq \eta\,\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\| = o_p(1).$$

Proof. Row t of $\mathbf{T}_{(v)}$ is $\hat{z}'_{1,r(v,t)}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{-1/2}$, so

$$\|\mathbf{T}_{(v),t\cdot}\| \;\le\; \frac{\|\hat{\hat{z}}_{1,r(v,t)}\|}{\lambda_{\min}\left(\mathbf{Z}'_{1,(v)}\mathbf{M}_{(v)}\mathbf{Z}_{1,(v)}\right)^{1/2}}.$$

The numerator is $o_p(\bar{T}^{1/2})$ uniformly in t and v by Lemma 13, and the denominator is $\left(\bar{T}(1-o_p(1))\right)^{1/2}$ by Lemma 3. Hence $\eta = o_p(1)$. The average squared row norm vanishes trivially, since $\sum_t \|\mathbf{T}_{(v),t\cdot}\|^2 = \operatorname{tr}\left(\mathbf{R}_{(v)}\right) \le h+1$, so the content of the claim lies in the maximum.

For the second claim, $\mathbf{Z}_{1,(v)}$ is a function of $\mathbf{S}_{(v-1)}\varepsilon, \ldots, \mathbf{S}_{(v-h)}\varepsilon$ and $\mathbf{M}_{(v)}$ is deterministic, so $\mathbf{R}_{(v)}$ is independent of $\mathbf{S}_{(v)}\varepsilon$. Therefore

$$\mathbb{E}\left[\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\|^2 \mid \mathbf{R}_{(v)}\right] = \operatorname{tr}\left(\mathbf{R}_{(v)}\mathbf{\Phi}^2_{(v)}\right) \le C\operatorname{tr}\left(\mathbf{R}_{(v)}\right) \le C(h+1),$$

and Markov's inequality gives $\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\| = O_p(1)$. Finally, $\left(\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)_t = \mathbf{T}_{(v),t\cdot}\mathbf{T}'_{(v)}\mathbf{S}_{(v)}\varepsilon$, so that

$$\left|\left(\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)_t\right| \le \eta\|\mathbf{T}'_{(v)}\mathbf{S}_{(v)}\varepsilon\| = \eta\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\|.$$

□

Lemma 25 ($\grave{\sigma}^2_{(u)}$ is close to $\check{\sigma}^2_{(u)}$)**.** *Under the assumptions of Lemmas 22 and 24, for every block u,*

$$\left\|\grave{\sigma}^2_{(u)} - \check{\sigma}^2_{(u)}\right\| = o_p(\bar{T}^{1/2}), \qquad \left\|\grave{\sigma}^2_{(u)}\right\| = O_p(\bar{T}^{1/2}).$$

Proof. Consider first a subsample $v \ge 1$, and write

$$\grave{\sigma}^2_{(v)} - \check{\sigma}^2_{(v)} = \boldsymbol{\kappa}_{(v)}\left(\hat{\varepsilon}^{\odot 2}_{(v)} - \dot{\varepsilon}^{\odot 2}_{(v)}\right) + \left(\boldsymbol{\kappa}_{(v)} - \check{\boldsymbol{\kappa}}_{(v)}\right)\dot{\varepsilon}^{\odot 2}_{(v)}.$$

The two terms are bounded separately. For the first term, $\hat{\varepsilon}_{(v)} - \dot{\varepsilon}_{(v)} = -\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon$, so

$$\begin{aligned}\left\|\hat{\varepsilon}^{\odot 2}_{(v)} - \dot{\varepsilon}^{\odot 2}_{(v)}\right\| &= \left\|\left(\hat{\varepsilon}_{(v)} - \dot{\varepsilon}_{(v)}\right) \odot \left(\hat{\varepsilon}_{(v)} + \dot{\varepsilon}_{(v)}\right)\right\| \le \left\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\right\|_\infty \left\|\hat{\varepsilon}_{(v)} + \dot{\varepsilon}_{(v)}\right\| \\ &\le o_p(1)\cdot 2\left\|\mathbf{S}_{(v)}\varepsilon\right\| = o_p(\bar{T}^{1/2}),\end{aligned}$$

using Lemma 24, $\left\|\mathbf{H}_{(v)}\right\|_2 \leq 1$, $\left\|\mathbf{M}_{(v)}\right\|_2 \leq 1$ and $\|\mathbf{S}_{(v)}\varepsilon\| = O_p(\bar{T}^{1/2})$. Since $\left\|\boldsymbol{\kappa}_{(v)}\right\|_2 = O_p(1)$ by Lemma 18, the first term is $o_p(\bar{T}^{1/2})$.

For the second term, recall (12). Now $|\mathbf{R}_{(v),tj}| = |\mathbf{T}_{(v),t\cdot}\mathbf{T}'_{(v),j\cdot}| \leq \|\mathbf{T}_{(v),t\cdot}\|\|\mathbf{T}_{(v),j\cdot}\|$, so

$$\begin{aligned}\left\|\mathbf{M}_{(v)} \odot \mathbf{R}_{(v)}\right\|_F^2 &= \sum_{t,j} \mathbf{M}_{(v),tj}^2 \mathbf{R}_{(v),tj}^2 \leq \left(\max_t \|\mathbf{T}_{(v),t\cdot}\|^2\right) \max_j \|\mathbf{M}_{(v),\cdot j}\|^2 \sum_j \|\mathbf{T}_{(v),j\cdot}\|^2 \\ &\leq \eta^2(h+1) = o_p(1),\end{aligned}$$

using $\sum_j \|\mathbf{T}_{(v),j\cdot}\|^2 = \left\|\mathbf{T}_{(v)}\right\|_F^2 \leq h+1$. Likewise,

$$\left\|\mathbf{R}_{(v)} \odot \mathbf{R}_{(v)}\right\|_F^2 \leq \left(\max_{t,j} \mathbf{R}_{(v),tj}^2\right) \left\|\mathbf{R}_{(v)}\right\|_F^2 \leq \eta^4(h+1) = o_p(1).$$

Hence

$$\left\|\mathbf{G}_{(v)} - \dot{\mathbf{G}}_{(v)}\right\|_2 \leq \left\|\mathbf{G}_{(v)} - \dot{\mathbf{G}}_{(v)}\right\|_F = o_p(1),$$

and

$$\left\|\boldsymbol{\kappa}_{(v)} - \check{\boldsymbol{\kappa}}_{(v)}\right\|_2 \leq \left\|\boldsymbol{\kappa}_{(v)}\right\|_2 \left\|\check{\boldsymbol{\kappa}}_{(v)}\right\|_2 \left\|\mathbf{G}_{(v)} - \dot{\mathbf{G}}_{(v)}\right\|_2 = o_p(1),$$

while $\|\dot{\varepsilon}_{(v)}^{\odot 2}\| = O_p(\bar{T}^{1/2})$ by Lemma 23, so that the second term is $o_p(\bar{T}^{1/2})$ as well. The second claim for a subsample follows from $\|\grave{\sigma}_{(v)}^2\| \leq \left\|\boldsymbol{\kappa}_{(v)}\right\|_2 \|\hat{\varepsilon}_{(v)}^{\odot 2}\| = O_p(1)O_p(\bar{T}^{1/2})$, by Lemmas 18 and 23.

It remains to consider a lagged block $u \leq 0$. For $t \geq 2$ the difference $\grave{\sigma}_{(u),t}^2 - \check{\sigma}_{(u),t}^2$ equals $\grave{\sigma}_{(u^*),t-1}^2 - \check{\sigma}_{(u^*),t-1}^2$, so those entries contribute at most the norm already bounded for the subsample u^*. At the first position, $\grave{\sigma}_{(u)}^2$ contains the proxy $\grave{\sigma}_{(u^*),1}^2$, whereas $\check{\sigma}_{(u)}^2$ contains the true variance σ_u^2, and

$$\left|\grave{\sigma}_{(u^*),1}^2 - \sigma_u^2\right| \leq \underbrace{\left|\check{\sigma}_{(u^*),1}^2 - \sigma_{r(u^*,1)}^2\right|}_{O_p(1)} + \underbrace{\left|\sigma_{r(u^*,1)}^2 - \sigma_u^2\right|}_{\leq C-c} + \underbrace{\left|\grave{\sigma}_{(u^*),1}^2 - \check{\sigma}_{(u^*),1}^2\right|}_{\leq \|\grave{\sigma}_{(u^*)}^2 - \check{\sigma}_{(u^*)}^2\|} = o_p(\bar{T}^{1/2}),$$

the first term because $\check{\sigma}_{(u^*),1}^2$ has mean $\sigma_{r(u^*,1)}^2$ and variance $\left(\check{\boldsymbol{\Gamma}}_{(u^*)}\right)_{11} \leq \left\|\check{\boldsymbol{\Gamma}}_{(u^*)}\right\|_2 = O(1)$ by Lemma 22(iii)–(iv) and Lemma 38, so that Chebyshev's inequality applies, and the second because both variances lie in $[c, C]$ by Assumption *Regular Errors*. Adding the

two contributions gives the first claim, and $\|\grave{\sigma}^2_{(u)}\| \leq \|\breve{\sigma}^2_{(u)}\| + o_p(\bar{T}^{1/2}) = O_p(\bar{T}^{1/2})$, using $\mathbb{E}\left[\|\breve{\sigma}^2_{(u)}\|^2\right] \leq \|\sigma^2_{(u)}\|^2 + \operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(u)}\right) = O(\bar{T})$. □

Lemma 26 (The feasible weight arrays)**.** *Let* $\mathbf{W}$ *be the weight array of* $c_{v,q}$ *or of* $d_{v,q}$*, built from* $\mathbf{M}_{(v)}$ *and* $\mathbf{M}_{(q)}$*, and let* $\hat{\mathbf{W}}$ *be its feasible counterpart, built from* $\mathbf{H}_{(v)}$ *and* $\mathbf{H}_{(q)}$ *in place of* $\mathbf{M}_{(v)}$ *and* $\mathbf{M}_{(q)}$*. Under the assumptions of Lemma 24,* $\left\|\hat{\mathbf{W}} - \mathbf{W}\right\|_F = o_p(1)$.

Proof. Consider first the weight array of $d_{v,q}$. Here $\mathbf{W} = \mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}$ and $\hat{\mathbf{W}} = \mathbf{H}_{(v)} \odot \mathbf{H}_{(v)}$, so $\hat{\mathbf{W}} - \mathbf{W}$ is the matrix $\mathbf{G}_{(v)} - \dot{\mathbf{G}}_{(v)}$ of (12), shown to be $o_p(1)$ in Frobenius norm in the proof of Lemma 25.

For the weight array of $c_{v,q}$, the two Hadamard factors are different matrices, so that (12) is not available and we use a telescoping argument instead. With $\mathbf{W} = \left(\mathbf{M}_{(v)}\mathbf{J}'\right) \odot \mathbf{M}_{(q)}$ and $\hat{\mathbf{W}} = \left(\mathbf{H}_{(v)}\mathbf{J}'\right) \odot \mathbf{H}_{(q)}$, adding and subtracting the mixed array $\left(\mathbf{M}_{(v)}\mathbf{J}'\right) \odot \mathbf{H}_{(q)}$ and using $\mathbf{H}_{(v)} - \mathbf{M}_{(v)} = -\mathbf{R}_{(v)}$,

$$\begin{aligned}\hat{\mathbf{W}} - \mathbf{W} &= \left[\left(\mathbf{H}_{(v)} - \mathbf{M}_{(v)}\right)\mathbf{J}'\right] \odot \mathbf{H}_{(q)} + \left(\mathbf{M}_{(v)}\mathbf{J}'\right) \odot \left[\mathbf{H}_{(q)} - \mathbf{M}_{(q)}\right] \\ &= -\left(\mathbf{R}_{(v)}\mathbf{J}'\right) \odot \mathbf{H}_{(q)} - \left(\mathbf{M}_{(v)}\mathbf{J}'\right) \odot \mathbf{R}_{(q)}.\end{aligned}$$

For the first term, $\left|\left(\mathbf{R}_{(v)}\mathbf{J}'\right)_{tj}\right| = \left|\mathbf{R}_{(v),t,j+1}\right| \leq \|\mathbf{T}_{(v),t\cdot}\|\|\mathbf{T}_{(v),j+1,\cdot}\|$ for $j \leq \bar{T} - 1$, while the last column of $\mathbf{R}_{(v)}\mathbf{J}'$ is zero, so

$$\begin{aligned}\left\|\left(\mathbf{R}_{(v)}\mathbf{J}'\right) \odot \mathbf{H}_{(q)}\right\|_F^2 &\leq \left(\max_t \|\mathbf{T}_{(v),t\cdot}\|^2\right) \sum_{j \leq \bar{T}-1} \left(\|\mathbf{T}_{(v),j+1,\cdot}\|^2 \sum_t \mathbf{H}^2_{(q),tj}\right) \\ &\leq \eta^2(h+1) = o_p(1),\end{aligned}$$

using $\sum_t \mathbf{H}^2_{(q),tj} = \mathbf{H}_{(q),jj} \leq 1$. Similarly, for the second term,

$$\begin{aligned}\left\|\left(\mathbf{M}_{(v)}\mathbf{J}'\right) \odot \mathbf{R}_{(q)}\right\|_F^2 &\leq \left(\max_t \|\mathbf{T}_{(q),t\cdot}\|^2\right) \sum_{j \leq \bar{T}-1} \left(\|\mathbf{T}_{(q),j\cdot}\|^2 \sum_t \left(\mathbf{M}_{(v)}\mathbf{J}'\right)^2_{tj}\right) \\ &\leq \eta^2(h+1) = o_p(1),\end{aligned}$$

since $\sum_t \left(\mathbf{M}_{(v)}\mathbf{J}'\right)^2_{tj} = \mathbf{M}_{(v),j+1,j+1} \leq 1$ for $j \leq \bar{T} - 1$, the last column of $\mathbf{M}_{(v)}\mathbf{J}'$ being zero. □

Lemma 27 (Consistency of the plug-in traces based on $\grave{\sigma}^2$)**.** *Let u_1 and u_2 be two blocks coming from distinct subsamples, $u_1^* \neq u_2^*$, let $\mathbf{W}$ be admissible, and let $\hat{\mathbf{W}}$ satisfy $\left\|\hat{\mathbf{W}} - \mathbf{W}\right\|_F = o_p(1)$. Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, No Influential Observations, Sub-One-Half Leverage *and* Bounded Leverage, *the following holds. Let ϑ be the form* (10) *for these blocks u_1 and u_2, and put*

$$\grave{\vartheta} := \sum_{t,j} \hat{w}_{tj}\, \grave{\sigma}^2_{(u_1),t}\, \grave{\sigma}^2_{(u_2),j}.$$

Then $\grave{\vartheta} - \vartheta = o_p(\bar{T})$.

Proof. Write $\grave{\sigma}^2_{(u_1)} = \sigma^2_{(u_1)} + \breve{\theta}_{(u_1)} + \delta_{(u_1)}$ with $\breve{\theta}_{(u_1)} := \breve{\sigma}^2_{(u_1)} - \sigma^2_{(u_1)}$ and $\delta_{(u_1)} := \grave{\sigma}^2_{(u_1)} - \breve{\sigma}^2_{(u_1)}$, and similarly for v, with entries $\breve{\theta}_{(u_1),t}$ and $\delta_{(u_1),t}$. By Lemma 22(iii), $\mathbb{E}\left[\breve{\theta}_{(u_1)}\right] = 0$. By Lemma 22(ii), $\breve{\theta}_{(u_1)}$ and $\breve{\theta}_{(u_2)}$ are independent, since $u_1^* \neq u_2^*$. By Lemma 22(iv) and Markov's inequality, $\|\breve{\theta}_{(u_1)}\| = O_p(\bar{T}^{1/2})$. Finally, by Lemma 25, $\|\delta_{(u_1)}\| = o_p(\bar{T}^{1/2})$ and $\|\grave{\sigma}^2_{(u_1)}\| = O_p(\bar{T}^{1/2})$. Throughout, we use that $|x'\mathbf{U}y| \leq \|\mathbf{U}\|_2 \|x\|\|y\| \leq \|\mathbf{U}\|_F \|x\|\|y\|$.

We first replace the estimated weights by the true weights. The resulting error is the bilinear form $\grave{\sigma}^{2\prime}_{(u_1)} \left(\hat{\mathbf{W}} - \mathbf{W}\right) \grave{\sigma}^2_{(u_2)}$, so that

$$\begin{aligned}\left|\sum_{t,j} (\hat{w}_{tj} - w_{tj})\, \grave{\sigma}^2_{(u_1),t} \grave{\sigma}^2_{(u_2),j}\right| &\leq \left\|\hat{\mathbf{W}} - \mathbf{W}\right\|_F \left\|\grave{\sigma}^2_{(u_1)}\right\| \left\|\grave{\sigma}^2_{(u_2)}\right\| \\ &= o_p(1) O_p(\bar{T}^{1/2}) O_p(\bar{T}^{1/2}) = o_p(\bar{T}),\end{aligned}$$

and it remains to show that $\sum_{t,j} w_{tj} \left(\grave{\sigma}^2_{(u_1),t} \grave{\sigma}^2_{(u_2),j} - \sigma^2_{(u_1),t} \sigma^2_{(u_2),j}\right) = o_p(\bar{T})$. Expanding the product of the two decompositions and subtracting $\sigma^2_{(u_1),t}\sigma^2_{(u_2),j}$ leaves the eight products that involve at least one error term:

$$\begin{aligned}\grave{\sigma}^2_{(u_1),t} \grave{\sigma}^2_{(u_2),j} - \sigma^2_{(u_1),t}\sigma^2_{(u_2),j} = {} & \underbrace{\sigma^2_{(u_1),t}\breve{\theta}_{(u_2),j} + \breve{\theta}_{(u_1),t}\sigma^2_{(u_2),j}}_{\text{(a)}} + \underbrace{\sigma^2_{(u_1),t}\delta_{(u_2),j} + \delta_{(u_1),t}\sigma^2_{(u_2),j}}_{\text{(b)}} \\ & + \underbrace{\breve{\theta}_{(u_1),t}\breve{\theta}_{(u_2),j}}_{\text{(c)}} + \underbrace{\breve{\theta}_{(u_1),t}\delta_{(u_2),j} + \delta_{(u_1),t}\breve{\theta}_{(u_2),j} + \delta_{(u_1),t}\delta_{(u_2),j}}_{\text{(d)}}.\end{aligned}$$

The grouping reflects the different properties of the two error terms: each entry of $\breve{\theta}$ is only

$O_p(1)$, but $\check{\theta}$ has mean zero and is independent across the two blocks, whereas δ has neither property but is small in norm. We show that the sum over t, j of each group, weighted by w_{tj}, is $o_p(\bar{T})$.

For the terms in (a), summation over t gives $\sum_{t,j} w_{tj}\sigma^2_{(u_1),t}\check{\theta}_{(u_2),j} = a'\check{\theta}_{(u_2)}$ with $a_j := \sum_t w_{tj}\sigma^2_{(u_1),t}$, which is deterministic because the σ^2_t and $\mathbf{W}$ are deterministic. By $\sigma^2_t \leq C$ and the column-sum condition in (W1), $|a_j| \leq C$ and hence $\|a\|^2 \leq C^2\bar{T}$. Since $\check{\theta}_{(u_2)}$ has mean zero, $\mathbb{E}\left[a'\check{\theta}_{(u_2)}\right] = 0$ and

$$\operatorname{Var}\left(a'\check{\theta}_{(u_2)}\right) = a'\check{\mathbf{\Gamma}}_{(u_2)}a \leq \left\|\check{\mathbf{\Gamma}}_{(u_2)}\right\|_2 \|a\|^2 = O(\bar{T}),$$

by Lemma 22(iv), so that the term is $O_p(\bar{T}^{1/2}) = o_p(\bar{T})$ by Chebyshev's inequality. The other term in (a) is treated in the same way with the roles of u_1 and u_2 exchanged, using the row-sum condition in (W1).

For the terms in (b), $\sigma^2_t \leq C$, the column-sum condition in (W1) and the Cauchy–Schwarz inequality in the form $\sum_j |\delta_{(u_2),j}| \leq \bar{T}^{1/2}\|\delta_{(u_2)}\|$ give

$$\begin{aligned}\left|\sum_{t,j} w_{tj}\sigma^2_{(u_1),t}\delta_{(u_2),j}\right| &\leq C\sum_j |\delta_{(u_2),j}| \sum_t |w_{tj}| \leq C\bar{T}^{1/2}\|\delta_{(u_2)}\| \\ &= C\bar{T}^{1/2}o_p(\bar{T}^{1/2}) = o_p(\bar{T}),\end{aligned}$$

and the other term is bounded symmetrically. This argument uses only the norm of δ, and neither a zero mean nor independence.

For the term in (c), put $\mathcal{S} := \sum_{t,j} w_{tj}\check{\theta}_{(u_1),t}\check{\theta}_{(u_2),j} = \check{\theta}'_{(u_1)}\mathbf{W}\check{\theta}_{(u_2)}$. The two vectors are independent and have mean zero, so that $\mathbb{E}[\mathcal{S}] = 0$. Squaring the double sum and factorising each expectation gives

$$\operatorname{Var}(\mathcal{S}) = \sum_{t,t'}\sum_{j,j'} w_{tj}w_{t'j'}\left(\check{\mathbf{\Gamma}}_{(u_1)}\right)_{tt'}\left(\check{\mathbf{\Gamma}}_{(u_2)}\right)_{jj'} = \operatorname{tr}\left(\mathbf{W}'\check{\mathbf{\Gamma}}_{(u_1)}\mathbf{W}\check{\mathbf{\Gamma}}_{(u_2)}\right),$$

where the last equality uses the symmetry of the covariance matrices. If $\mathbf{Y} \succeq 0$ and $\mathbf{U}$ is symmetric with $\mathbf{U} \preceq \tau\mathbf{I}$, then $\operatorname{tr}(\mathbf{YU}) = \operatorname{tr}\left(\mathbf{Y}^{1/2}\mathbf{U}\mathbf{Y}^{1/2}\right) \leq \tau\operatorname{tr}(\mathbf{Y})$, because $\mathbf{Y}^{1/2}(\tau\mathbf{I} - \mathbf{U})\mathbf{Y}^{1/2} \succeq 0$ has non-negative trace. Applying this inequality first with

$\mathbf{Y} = \mathbf{W}\check{\mathbf{\Gamma}}_{(u_2)}\mathbf{W}'$ and $\mathbf{U} = \check{\mathbf{\Gamma}}_{(u_1)}$, and then with $\mathbf{Y} = \mathbf{W}'\mathbf{W}$ and $\mathbf{U} = \check{\mathbf{\Gamma}}_{(u_2)}$, yields

$$\operatorname{Var}(\mathcal{S}) \leq \left\|\check{\mathbf{\Gamma}}_{(u_1)}\right\|_2 \operatorname{tr}\left(\left(\mathbf{W}'\mathbf{W}\right)\check{\mathbf{\Gamma}}_{(u_2)}\right) \leq \left\|\check{\mathbf{\Gamma}}_{(u_1)}\right\|_2 \left\|\check{\mathbf{\Gamma}}_{(u_2)}\right\|_2 \|\mathbf{W}\|_F^2 = O(\bar{T}),$$

by Lemma 22(iv) and (W2). Chebyshev's inequality gives $\mathcal{S} = O_p(\bar{T}^{1/2}) = o_p(\bar{T})$. The independence of $\check{\theta}_{(u_1)}$ and $\check{\theta}_{(u_2)}$ is essential at this point: the bound $|\mathcal{S}| \leq \|\mathbf{W}\|_2 \|\check{\theta}_{(u_1)}\|\|\check{\theta}_{(u_2)}\|$ yields only $O_p(\bar{T})$. Note also that the argument requires a bound on the spectral norms of the covariance matrices, as provided by Lemma 22(iv), together with a bound on the Frobenius norm of the weights.

Finally, each of the three sums in (d) is a bilinear form in $\mathbf{W}$ with at least one factor δ, so that, by (W3),

$$\left|\check{\theta}'_{(u_1)}\mathbf{W}\delta_{(u_2)}\right| \leq \|\mathbf{W}\|_2 \|\check{\theta}_{(u_1)}\|\|\delta_{(u_2)}\| = O(1)O_p(\bar{T}^{1/2})o_p(\bar{T}^{1/2}) = o_p(\bar{T}),$$

and likewise for $\delta'_{(u_1)}\mathbf{W}\check{\theta}_{(u_2)}$ and $\delta'_{(u_1)}\mathbf{W}\delta_{(u_2)}$. Combining the bounds for (a)–(d) with the bound on the weight error gives $\grave{\vartheta} - \vartheta = o_p(\bar{T})$. □

Theorem 3 (Conservative inference with the CJN estimator)**.** *For each* $v \in \{1, \ldots, h+1\}$, *let* $\grave{\Sigma}_v$ *and* $\grave{\Lambda}$ *be* $\hat{\Sigma}_v$ *and* $\hat{\Lambda}$ *with* $\hat{\sigma}_t^2 = \grave{\sigma}_t^2$, *respectively. Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, No Influential Observations, Sub-One-Half Leverage, *and* Bounded Leverage, *we have*

$$\liminf_{T\to\infty} \mathbb{P}\left[r'\beta \in \left[r'\hat{\beta} \pm \frac{z_{1-\alpha/2}}{h+1}\sqrt{\left(1+\grave{\Lambda}\right)\sum_{v=1}^{h+1}\grave{\Sigma}_v}\right]\right] \geq 1-\alpha$$

for every $\alpha \in (0,1)$, *where* $z_{1-\alpha/2}$ *denotes the* $1-\alpha/2$ *quantile of the standard normal distribution.*

Proof. Throughout, $\grave{c}_{v,q}$, $\grave{d}_{v,q}$ and $\grave{d}_{q,v}$ denote the estimates $\hat{c}_{v,q}$, $\hat{d}_{v,q}$ and $\hat{d}_{q,v}$ evaluated at $\hat{\sigma}^2 = \grave{\sigma}^2$, and $\grave{\lambda}$ denotes the ratio in the definition of $\hat{\Lambda}$ evaluated at $\grave{\sigma}^2$, so that $\grave{\Lambda} = \min\{1, \grave{\lambda}\}$.

We first show that the scaled traces are consistent. Fix a pair $v < q$. By Lemma 21 the weight arrays of $c_{v,q}$, $d_{v,q}$ and $d_{q,v}$ are admissible, and by Lemma 26 their feasible counterparts satisfy $\left\|\hat{\mathbf{W}} - \mathbf{W}\right\|_F = o_p(1)$. The two blocks entering each of the three traces come from

distinct subsamples. For $c_{v,q}$ they are v and q, for $d_{v,q}$ they are v and $q-h-1$, whose subsample is q, and for $d_{q,v}$ they are q and v. Lemma 27 therefore applies to each and gives

$$\frac{\grave{c}_{v,q}-c_{v,q}}{\bar{T}}=o_p(1),\qquad \frac{\grave{d}_{v,q}-d_{v,q}}{\bar{T}}=o_p(1),\qquad \frac{\grave{d}_{q,v}-d_{q,v}}{\bar{T}}=o_p(1). \tag{46}$$

We next show that $\grave{\Lambda}-\Lambda \xrightarrow{p} 0$. The targets $c_{v,q}/\bar{T}$ and $d_{v,q}/\bar{T}$ are deterministic sequences that need not converge, so the continuous mapping theorem in its usual form does not apply. Instead, we show that the map producing λ is uniformly continuous on a compact set that contains the targets for every large T and the estimates with probability approaching one.

For the denominators, writing out the trace, $d_{v,q}=\sum_{t,j}\mathbf{M}^2_{(v),tj}\sigma^2_{r(v,t)}\sigma^2_{r(q-h-1,j)}$, and idempotency gives $\sum_{t,j}\mathbf{M}^2_{(v),tj}=\operatorname{tr}\left(\mathbf{M}_{(v)}\right)=\bar{T}-K+h+1$. With $c\le\sigma_t^2\le C$ from Assumption *Regular Errors* this yields bounds in both directions,

$$\delta\ \le\ c^2\,\frac{\bar{T}-K+h+1}{\bar{T}}\ \le\ \frac{d_{v,q}}{\bar{T}}\ \le\ C^2\,\frac{\bar{T}-K+h+1}{\bar{T}}\ \le\ C^2,$$

the first for all large T and some $\delta>0$, since $\limsup K/\bar{T}<1$ under Assumption *Proportional Regime*. The same holds for $d_{q,v}$. For the numerators, $\sigma_t^2\le C$ and (W1) give $\sum_{t,j}|w_{tj}|\le\bar{T}$ and hence $|c_{v,q}|/\bar{T}\le C^2$. Put $\bar{C}:=C^2$. For every large T the vector of scaled targets $(c_{v,q},d_{v,q},d_{q,v})/\bar{T}$ therefore lies in the compact set

$$\mathcal{K}_{\delta,\bar{C}}:=\left\{(x_1,x_2,x_3)\in\mathbb{R}^3:\ |x_1|\le\bar{C},\ \delta\le x_2\le\bar{C},\ \delta\le x_3\le\bar{C}\right\}.$$

Let $\mathcal{E}_{0,T}$ be the event that, for all pairs, the vector of scaled estimates lies in $\mathcal{K}:=\mathcal{K}_{\delta/2,2\bar{C}}$. If it does not, then, since the target lies in $\mathcal{K}_{\delta,\bar{C}}$, some coordinate of the estimate differs from the corresponding coordinate of the target by more than $\min\{\bar{C},\delta/2\}$. Hence (46) and a union bound over the finitely many pairs give $\mathbb{P}\left[\mathcal{E}_{0,T}\right]\to 1$. On $\mathcal{E}_{0,T}$ every estimated denominator is positive, so $\grave{\lambda}$ is well defined.

Let $g(x_1,x_2,x_3):=|x_1|/\sqrt{x_2x_3}$. Since g is homogeneous of degree zero,

$$\lambda=\max_{v<q} g\left((c_{v,q},d_{v,q},d_{q,v})/\bar{T}\right),$$

and likewise for $\grave{\lambda}$ with the estimates. On $\mathcal{K}$ the product $x_2 x_3 \geq (\delta/2)^2 > 0$, so g is continuous there, and $\mathcal{K}$ is compact and does not depend on T. By the Heine–Cantor theorem g is uniformly continuous on $\mathcal{K}$, with a single modulus of continuity serving for every T. Fix $\epsilon > 0$ and let $\varrho > 0$ be such that $|g(p) - g(p')| < \epsilon$ whenever $p, p' \in \mathcal{K}$ and $\|p - p'\|_\infty < \varrho$. Since $|\max_i a_i - \max_i b_i| \leq \max_i |a_i - b_i|$, the event $\{|\grave{\lambda} - \lambda| \geq \epsilon\}$ can occur on $\mathcal{E}_{0,T}$ only if, for some pair, the scaled estimate and the scaled target are at $\|\cdot\|_\infty$-distance at least ϱ. Writing

$$\Delta_{v,q} := \frac{1}{\overline{T}} \max\left\{ |\grave{c}_{v,q} - c_{v,q}|,\ |\grave{d}_{v,q} - d_{v,q}|,\ |\grave{d}_{q,v} - d_{q,v}| \right\}$$

for this distance, we obtain

$$\mathbb{P}\left[|\grave{\lambda} - \lambda| \geq \epsilon \right] \leq \mathbb{P}\left[\mathcal{E}_{0,T}^c \right] + \mathbb{P}\left[\max_{v<q} \Delta_{v,q} \geq \varrho \right],$$

and both terms tend to zero, the first because $\mathbb{P}[\mathcal{E}_{0,T}] \to 1$ and the second by (46) and a union bound. As ϵ was arbitrary, $\grave{\lambda} - \lambda \xrightarrow{p} 0$, and since $x \mapsto \min\{1, x\}$ is 1-Lipschitz, $\grave{\Lambda} - \Lambda \xrightarrow{p} 0$.

It remains to establish the asymptotic coverage. Write $z := z_{1-\alpha/2}$ and

$$\zeta_T := \frac{(h+1)\, r'(\hat{\beta} - \beta)}{\sqrt{\Sigma}}, \qquad R_T := \frac{\left(1 + \grave{\Lambda}\right) \sum_{v=1}^{h+1} \grave{\Sigma}_v}{\Sigma},$$

so that $\zeta_T \xrightarrow{d} N(0,1)$ by Theorem 2. The interval

$$\mathcal{I}_T := \left[r'\hat{\beta} \pm \frac{z}{h+1} \sqrt{\left(1 + \grave{\Lambda}\right) \sum_{v=1}^{h+1} \grave{\Sigma}_v} \right]$$

contains $r'\beta$ if and only if $|\zeta_T| \leq z\sqrt{R_T}$, with $\mathcal{I}_T$ read as empty when $\sum_{v=1}^{h+1} \grave{\Sigma}_v < 0$, which the estimates $\grave{\sigma}_t^2$ do not preclude in finite samples, or when $\sum_{v=1}^{h+1} \grave{\Sigma}_v$ is undefined because $\mathbf{G}_{(v)}$ is singular, which by Lemma 19 has probability tending to zero. Note that $\Sigma > 0$, being the variance by which Theorem 2 normalises, and hence $\sum_{v=1}^{h+1} \Sigma_v \geq \Sigma/(1+\Lambda) > 0$ by Lemma 20.

Factor

$$R_T = \underbrace{\frac{1+\grave{\Lambda}}{1+\Lambda} \cdot \frac{\sum_{v=1}^{h+1} \grave{\Sigma}_v}{\sum_{v=1}^{h+1} \Sigma_v}}_{=:P_T} \cdot \frac{(1+\Lambda)\sum_{v=1}^{h+1} \Sigma_v}{\Sigma}, \qquad \frac{(1+\Lambda)\sum_{v=1}^{h+1} \Sigma_v}{\Sigma} \geq 1,$$

by (44) of Lemma 20. The last factor is deterministic. We show $P_T \xrightarrow{p} 1$. First, $1 + \Lambda \geq 1$ and $\left|\frac{1+\grave{\Lambda}}{1+\Lambda} - 1\right| = \frac{|\grave{\Lambda}-\Lambda|}{1+\Lambda} \leq |\grave{\Lambda} - \Lambda| = o_p(1)$ by the preceding argument. Second, $\left(\sum_{v=1}^{h+1} \grave{\Sigma}_v\right) / \sum_{v=1}^{h+1} \Sigma_v \xrightarrow{p} 1$ by Lemma 19 together with Lemma 7. Both factors converge in probability to the fixed limit one and $(x, y) \mapsto xy$ is continuous at $(1, 1)$, so $P_T \xrightarrow{p} 1$ by the continuous mapping theorem.

Fix $\epsilon \in (0, 1)$ and let $\mathcal{E}_{1,T} := \{|P_T - 1| < \epsilon\}$, so $\mathbb{P}[\mathcal{E}_{1,T}] \to 1$. On $\mathcal{E}_{1,T}$ we have $P_T > 1 - \epsilon > 0$, and multiplying by a factor of at least one can only increase a positive number, so $R_T > 1 - \epsilon$ there. In particular, $\mathcal{I}_T$ is non-empty on $\mathcal{E}_{1,T}$. Let $\mathcal{E}_{2,T} := \{|\zeta_T| \leq z\sqrt{1-\epsilon}\}$. On $\mathcal{E}_{2,T} \cap \mathcal{E}_{1,T}$ the interval $\mathcal{I}_T$ covers $r'\beta$, so

$$\mathbb{P}[r'\beta \in \mathcal{I}_T] \;\geq\; \mathbb{P}[\mathcal{E}_{2,T}] - \mathbb{P}\left[\mathcal{E}_{1,T}^c\right].$$

The second term tends to zero. For the first, the set $[-z\sqrt{1-\epsilon}, z\sqrt{1-\epsilon}]$ has boundary of probability zero under the standard normal law, so by $\zeta_T \xrightarrow{d} N(0, 1)$ and the portmanteau theorem $\mathbb{P}[\mathcal{E}_{2,T}] \to 2\Phi\left(z\sqrt{1-\epsilon}\right) - 1$, with Φ the standard normal distribution function. Hence

$$\liminf_{T\to\infty} \mathbb{P}[r'\beta \in \mathcal{I}_T] \;\geq\; 2\Phi\left(z\sqrt{1-\epsilon}\right) - 1.$$

The left-hand side does not depend on ϵ, so the inequality holds for every $\epsilon \in (0, 1)$. Letting $\epsilon \downarrow 0$ and using the continuity of Φ, the right-hand side tends to $2\Phi(z) - 1 = 1 - \alpha$. □

Lemma 28 (Properties of $\breve{\sigma}^2_{(u)}$)**.** *Let u be a block and u^* the subsample it belongs to. Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, Bounded Fitted Values *and* Leave-One-Out Conditioning*:*

(i) for a subsample $u \geq 1$, $\mathbf{M}_{(u),tt} \geq c_M := c/(c + C)$ for every t, deterministically;

(ii) $\breve{\sigma}^2_{(u)}$ is a function of $\mathbf{S}_{(u^*)}\varepsilon$ alone. In particular $\breve{\sigma}^2_{(u)}$ and $\breve{\sigma}^2_{(u')}$ are independent whenever

$u^* \neq u'^*$;

(iii) $\mathbb{E}\left[\breve{\sigma}^2_{(u)}\right] = \sigma^2_{(u)}$ *exactly;*

(iv) for a subsample v*, the covariance matrix* $\breve{\mathbf{\Gamma}}_{(v)}$ *satisfies*

$$\breve{\mathbf{\Gamma}}_{(v)} \preceq 2\mathbf{\Gamma}^0_{(v)} + 2\mathbf{\Gamma}^\mu_{(v)},$$
$$\left\|\mathbf{\Gamma}^0_{(v)}\right\|_2 \leq C_0 := C^2\left(1 + \frac{1}{c_M} + \frac{2}{c_M^2}\right), \qquad \left\|\mathbf{F}_{(v)}\right\|_2 \leq \frac{C}{c_M^2},$$

and consequently $\operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(v)}\right) = O(\bar{T})$*. For a lagged block* $u \leq 0$ *the same holds with* u *in place of* v*, and* $\operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(u)}\right) \leq \operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(u^*)}\right)$.

Proof. We first record two consequences of Assumption *Bounded Fitted Values*. First, by Assumption *Bounds*, $\|\tilde{\mu}_{(v)}\|^2 = \beta'\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\beta \leq C\bar{T}\|\beta\|^2$, and $\|\beta\| \leq C$ by Assumption *Bounded Fitted Values*. Hence

$$\|\tilde{\mu}_{(v)}\|^2 = O(\bar{T}). \tag{47}$$

Second, $\tilde{\mu}_{(v),t} = \beta'\tilde{x}_{r(v,t)}$, and the periods of subsample v are a subset of $\{1, \ldots, T\}$ with $\bar{T} = T/(h+1)$ and h fixed, so that $\max_t(\beta'\tilde{x}_t)^2 = o(T)$ implies

$$\max_t \tilde{\mu}^2_{(v),t} = o(\bar{T}). \tag{48}$$

(i) Let $v := u \geq 1$. Since $\operatorname{col}(\mathbf{Z}_{2,(v)}) \subseteq \operatorname{col}(\tilde{\mathbf{X}}_{(v)})$, $\mathbf{M}_{(v)}$ dominates the annihilator of $\tilde{\mathbf{X}}_{(v)}$ in the Loewner order, so $\mathbf{M}_{(v),tt} \geq 1 - \tilde{x}'_{r(v,t)}\left(\tilde{\mathbf{X}}'_{(v)}\tilde{\mathbf{X}}_{(v)}\right)^{-1}\tilde{x}_{r(v,t)}$. The first part of the proof of Lemma 14 shows, by the Sherman–Morrison formula applied to the leave-one-out Gram matrix and Assumption *Leave-One-Out Conditioning*, that this equals $1/(1+\tau_{r(v,t)})$ with $\tau_{r(v,t)} \leq C/c$. Hence $\mathbf{M}_{(v),tt} \geq 1/(1+C/c) = c_M$, deterministically.

(ii) and (iii) Consider first a subsample $v := u \geq 1$. Since $\mathbf{M}_{(v)}$ and $\tilde{\mu}_{(v)}$ are deterministic, $\breve{\sigma}^2_{(v)}$ is a fixed function of $\mathbf{S}_{(v)}\varepsilon$, and the error blocks of distinct subsamples are independent. For the mean, $\mathbb{E}\left[\left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)\left(\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)_t\right] = \mathbf{M}_{(v),tt}\sigma^2_{r(v,t)}$, because $\mathbb{E}\left[\varepsilon_{r(v,t)}\varepsilon_{r(v,j)}\right]$ vanishes for $j \neq t$ and $\tilde{\mu}_{(v),t}$ multiplies a mean-zero term. Dividing by $\mathbf{M}_{(v),tt}$, which is positive by (i), gives (iii).

The lagged blocks are treated at the end of the proof.

(iv) Consider again a subsample $v \geq 1$, and recall the split $\breve{\sigma}^2_{(v)} - \sigma^2_{(v)} = \theta^0_{(v)} + \theta^\mu_{(v)}$ into a pure-error part and a fitted-value part, both of mean zero. The quadratic piece of $\theta^0_{(v),t}$ is $\sum_{j\neq t}(\mathbf{M}_{(v),tj}/\mathbf{M}_{(v),tt})\varepsilon_{r(v,t)}\varepsilon_{r(v,j)}$. By the index-pattern argument used for (45), the two pieces of $\theta^0_{(v)}$ are uncorrelated. Moreover, $\mathrm{Cov}\left(\varepsilon^2_{r(v,t)}, \varepsilon^2_{r(v,j)}\right) = \mathbb{1}\{t = j\}\,\mathrm{Var}\left(\varepsilon^2_{r(v,t)}\right)$, and the quadratic piece has covariances $\mathbb{1}\{t = j\}\sum_{k\neq t}(\mathbf{M}_{(v),tk}/\mathbf{M}_{(v),tt})^2\sigma^2_{r(v,t)}\sigma^2_{r(v,k)} + \mathbb{1}\{t \neq j\}\mathbf{M}^2_{(v),tj}\sigma^2_{r(v,t)}\sigma^2_{r(v,j)}/(\mathbf{M}_{(v),tt}\mathbf{M}_{(v),jj})$. Hence $\mathbf{\Gamma}^0_{(v)} = \mathrm{Var}\left(\theta^0_{(v)}\right)$ is the sum of two diagonal matrices and one matrix with zero diagonal. The first has entries $\mathrm{Var}\left(\varepsilon^2_{r(v,t)}\right) \leq C^2$. For the second, idempotency gives $\sum_{j\neq t}\mathbf{M}^2_{(v),tj} = \mathbf{M}_{(v),tt} - \mathbf{M}^2_{(v),tt}$, so its entries are at most $C^2\left(\mathbf{M}_{(v),tt} - \mathbf{M}^2_{(v),tt}\right)/\mathbf{M}^2_{(v),tt} \leq C^2/c_M$ by (i). The third equals $\mathbf{V}_{(v)}\left(\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)} - \mathrm{diag}(\mathbf{M}^2_{(v),tt})\right)\mathbf{V}_{(v)}$ with $\mathbf{V}_{(v)} := \mathrm{diag}(\sigma^2_{(v)})\mathrm{diag}(\mathbf{M}_{(v),tt})^{-1}$ of norm at most C/c_M. Since $\left\|\mathbf{M}_{(v)} \odot \mathbf{M}_{(v)}\right\|_2 \leq 1$ by Lemma 42(ii) and $\max_t \mathbf{M}^2_{(v),tt} \leq 1$, its norm is at most $2C^2/c^2_M$. Adding the three bounds gives $\left\|\mathbf{\Gamma}^0_{(v)}\right\|_2 \leq C^2(1 + 1/c_M + 2/c^2_M) = C_0$.

The fitted-value part is linear in $\mathbf{S}_{(v)}\varepsilon$, $\theta^\mu_{(v)} = \mathbf{D}_{(v)}\,\mathrm{diag}(\mathbf{m}_{(v)})^{-1}\mathbf{M}_{(v)}\mathbf{S}_{(v)}\varepsilon$, so that its covariance matrix is indeed $\mathbf{\Gamma}^\mu_{(v)} = \mathbf{D}_{(v)}\mathbf{F}_{(v)}\mathbf{D}_{(v)}$, where $\mathbf{F}_{(v)} \succeq 0$ and, by (i), $\left\|\mathbf{F}_{(v)}\right\|_2 \leq C/c^2_M$. As $\mathrm{Var}\left(\theta^0_{(v)} + \theta^\mu_{(v)}\right) + \mathrm{Var}\left(\theta^0_{(v)} - \theta^\mu_{(v)}\right) = 2\mathbf{\Gamma}^0_{(v)} + 2\mathbf{\Gamma}^\mu_{(v)}$ and $\mathrm{Var}\left(\theta^0_{(v)} - \theta^\mu_{(v)}\right) \succeq 0$, the Loewner bound follows.

Since the trace is monotone in the Loewner order, $\mathrm{tr}\left(\breve{\mathbf{\Gamma}}_{(v)}\right) \leq 2\,\mathrm{tr}\left(\mathbf{\Gamma}^0_{(v)}\right) + 2\,\mathrm{tr}\left(\mathbf{\Gamma}^\mu_{(v)}\right)$. The first trace is a sum of $\bar{T}$ eigenvalues each at most C_0. For the second, a diagonal entry of a positive semidefinite matrix is at most its largest eigenvalue (Lemma 38), so that

$$\mathrm{tr}\left(\mathbf{\Gamma}^\mu_{(v)}\right) = \sum_t \tilde{\mu}^2_{(v),t}\mathbf{F}_{(v),tt} \leq \left\|\mathbf{F}_{(v)}\right\|_2 \left\|\tilde{\mu}_{(v)}\right\|^2 = O(\bar{T}),$$

by (47). Hence $\mathrm{tr}\left(\breve{\mathbf{\Gamma}}_{(v)}\right) = O(\bar{T})$. Note that the spectral norm of $\breve{\mathbf{\Gamma}}_{(v)}$ can be as large as $\max_t \tilde{\mu}^2_{(v),t}$, whereas its trace depends on the fitted values only through $\|\tilde{\mu}_{(v)}\|^2$.

It remains to treat a lagged block $u \leq 0$. As in the proof of Lemma 22, the definition of $\breve{\sigma}^2_{(u)}$ for lagged blocks reads $\breve{\sigma}^2_{(u)} = \mathbf{J}'\breve{\sigma}^2_{(u^*)} + \sigma^2_u e_1$, and (ii) and (iii) follow as in that proof. Conjugating the Loewner bound by $\mathbf{J}'$ preserves the order, and $\mathbf{J}'\mathbf{D}_{(u^*)} = \mathbf{D}_{(u)}\mathbf{J}'$, both matrices having $\tilde{\mu}_{(u^*),t-1}$ in position $(t, t-1)$ and zeros elsewhere, so $\mathbf{J}'\mathbf{\Gamma}^\mu_{(u^*)}\mathbf{J} = \mathbf{D}_{(u)}\mathbf{F}_{(u)}\mathbf{D}_{(u)} = \mathbf{\Gamma}^\mu_{(u)}$ with $\left\|\mathbf{F}_{(u)}\right\|_2 \leq \left\|\mathbf{F}_{(u^*)}\right\|_2$, and likewise $\left\|\mathbf{\Gamma}^0_{(u)}\right\|_2 \leq \left\|\mathbf{\Gamma}^0_{(u^*)}\right\|_2$. The trace claim follows because $\mathrm{tr}\,(\mathbf{J}'\mathbf{N}\mathbf{J}) \leq \mathrm{tr}\,(\mathbf{N})$ for any positive semidefinite $\bar{T} \times \bar{T}$ matrix $\mathbf{N}$, and the diagonal of $\mathbf{D}_{(u)}$

obeys (47) and (48) because its entries are entries of $\tilde{\mu}_{(u^*)}$. □

Lemma 29 ($\acute{\sigma}^2_{(u)}$ is close to $\breve{\sigma}^2_{(u)}$)**.** *Under the assumptions of Lemmas 24 and 28, for every block u,*

$$\left\|\acute{\sigma}^2_{(u)} - \breve{\sigma}^2_{(u)}\right\| = o_p(\bar{T}^{1/2}), \qquad \left\|\acute{\sigma}^2_{(u)}\right\| = O_p(\bar{T}^{1/2}).$$

Proof. Consider first a subsample $v \geq 1$. By Assumption *Feedback Decomposition*, $y_{r(v,t)} = \tilde{\mu}_{(v),t} + \varpi_{r(v,t)} + \varepsilon_{r(v,t)}$ with $\varpi_t = \sum_{l=1}^{h} \varepsilon_{t-l}\alpha_l'\beta$, so the difference splits into a term that compares the two annihilators and a term that contains the random fitted values:

$$\acute{\sigma}^2_{(v),t} - \breve{\sigma}^2_{(v),t} = \left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)\left(\frac{\hat{\varepsilon}_{(v),t}}{\mathbf{H}_{(v),tt}} - \frac{\dot{\varepsilon}_{(v),t}}{\mathbf{M}_{(v),tt}}\right) + \varpi_{r(v,t)}\frac{\hat{\varepsilon}_{(v),t}}{\mathbf{H}_{(v),tt}}.$$

For the first term, recall that $\hat{\varepsilon}_{(v)} = \dot{\varepsilon}_{(v)} - \mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon$ and $\mathbf{H}_{(v),tt} = \mathbf{M}_{(v),tt} - \mathbf{R}_{(v),tt}$, so

$$\frac{\hat{\varepsilon}_{(v),t}}{\mathbf{H}_{(v),tt}} - \frac{\dot{\varepsilon}_{(v),t}}{\mathbf{M}_{(v),tt}} = \dot{\varepsilon}_{(v),t}\frac{\mathbf{R}_{(v),tt}}{\mathbf{H}_{(v),tt}\mathbf{M}_{(v),tt}} - \frac{\left(\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\right)_t}{\mathbf{H}_{(v),tt}}.$$

Since $\mathbf{R}_{(v),tt} = \|\mathbf{T}_{(v),t\cdot}\|^2 \leq \eta^2$, Lemma 28(i) gives $\mathbf{H}_{(v),tt} \geq c_M - \eta^2$ with $c_M = c/(c+C)$, and as $\eta = o_p(1)$ by Lemma 24, the event $\mathcal{E}_{v,T} := \{\min_t \mathbf{H}_{(v),tt} \geq c_M/2\}$ has probability tending to one. On $\mathcal{E}_{v,T}$, using $(a_1+a_2+a_3)^2 \leq 3(a_1^2+a_2^2+a_3^2)$ for real a_1, a_2, a_3,

$$\begin{aligned}\left\|\acute{\sigma}^2_{(v)} - \breve{\sigma}^2_{(v)}\right\|^2 \leq{}& \frac{12\eta^4}{c_M^4}\sum_t \left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)^2 \dot{\varepsilon}^2_{(v),t} \\ &+ \frac{12\left\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\right\|_\infty^2}{c_M^2}\sum_t \left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)^2 \\ &+ \frac{12}{c_M^2}\sum_t \varpi^2_{r(v,t)}\hat{\varepsilon}^2_{(v),t}.\end{aligned} \tag{49}$$

We next bound the last sum in (49), which contains the random fitted values. The vector ϖ restricted to subsample v is a function of $\mathbf{X}_{(v)}$, since β and the α_l are non-random, and $\mathbf{S}_{(v)}\varepsilon$ is independent of $\mathbf{X}_{(v)}$. Hence, conditionally on $\mathbf{X}_{(v)}$,

$$\mathbb{E}\left[\sum_t \varpi^2_{r(v,t)}\hat{\varepsilon}^2_{(v),t} \;\middle|\; \mathbf{X}_{(v)}\right] = \sum_t \varpi^2_{r(v,t)}\left(\mathbf{H}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{H}_{(v)}\right)_{tt} \leq C\sum_t \varpi^2_{r(v,t)},$$

and $\mathbb{E}\left[\sum_t \varpi^2_{r(v,t)}\right] = \sum_t \sum_{l=1}^h (\alpha_l'\beta)^2 \sigma^2_{r(v,t)-l} \leq C\bar{T}\sum_l (\alpha_l'\beta)^2 = o(\bar{T})$ by Assumption *Bounded Fitted Values*, the cross terms vanishing because the blocks $\mathbf{S}_{(v-l)}\varepsilon$, $l = 1, \ldots, h$, are disjoint. Consequently, the last sum in (49) is $o_p(\bar{T})$ by Markov's inequality.

It remains to bound the first two sums in (49). First, $\sum_t \left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)^2 \leq 2\|\tilde{\mu}_{(v)}\|^2 + 2\|\mathbf{S}_{(v)}\varepsilon\|^2 = O_p(\bar{T})$ by (47). Second, using $\mathbb{E}\left[\dot{\varepsilon}^2_{(v),t}\right] = \left(\mathbf{M}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{M}_{(v)}\right)_{tt} \leq C$, the Cauchy–Schwarz inequality and Lemma 23,

$$\mathbb{E}\left[\sum_t \left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)^2 \dot{\varepsilon}^2_{(v),t}\right] \leq 2C\|\tilde{\mu}_{(v)}\|^2 + 2\sum_t \left(\mathbb{E}\left[\varepsilon^4_{r(v,t)}\right]\mathbb{E}\left[\dot{\varepsilon}^4_{(v),t}\right]\right)^{1/2} = O(\bar{T}),$$

so this sum is $O_p(\bar{T})$. Since $\eta^4 = o_p(1)$ and $\|\mathbf{R}_{(v)}\mathbf{S}_{(v)}\varepsilon\|^2_\infty = o_p(1)$ by Lemma 24, the first two terms of (49) are $o_p(\bar{T})$ as well, which proves the first claim for a subsample.

For the second claim, on $\mathcal{E}_{v,T}$,

$$\left\|\acute{\sigma}^2_{(v)}\right\|^2 \leq \frac{8}{C_M^2}\left(\sum_t \left(\tilde{\mu}_{(v),t} + \varepsilon_{r(v,t)}\right)^2 \hat{\varepsilon}^2_{(v),t} + \sum_t \varpi^2_{r(v,t)}\hat{\varepsilon}^2_{(v),t}\right).$$

The second sum is $o_p(\bar{T})$, as shown above. For the first sum, note that, conditionally on $\mathbf{X}_{(v)}$, the entries of $\mathbf{H}_{(v)}$ are fixed and $\mathbf{S}_{(v)}\varepsilon$ is independent of $\mathbf{X}_{(v)}$. Hence the bound on the expectation derived above remains valid with $\hat{\varepsilon}$ in place of $\dot{\varepsilon}$, using the conditional bound of Lemma 23 and $\left(\mathbf{H}_{(v)}\mathbf{\Phi}^2_{(v)}\mathbf{H}_{(v)}\right)_{tt} \leq C$. This bounds the conditional expectation by $O(\bar{T})$ uniformly in $\mathbf{X}_{(v)}$, and the claim follows by the tower property and Markov's inequality. Finally $\mathbb{P}[\mathcal{E}_{v,T}] \to 1$, so that the restriction to $\mathcal{E}_{v,T}$ does not affect stochastic orders.

It remains to consider a lagged block $u \leq 0$. For $t \geq 2$ the difference $\acute{\sigma}^2_{(u),t} - \breve{\sigma}^2_{(u),t}$ equals $\acute{\sigma}^2_{(u^*),t-1} - \breve{\sigma}^2_{(u^*),t-1}$, so those entries contribute at most the norm already bounded for the subsample u^*. For the first entry,

$$\left|\acute{\sigma}^2_{(u^*),1} - \sigma^2_u\right| \leq \left|\breve{\sigma}^2_{(u^*),1} - \sigma^2_{r(u^*,1)}\right| + \left|\sigma^2_{r(u^*,1)} - \sigma^2_u\right| + \left|\acute{\sigma}^2_{(u^*),1} - \breve{\sigma}^2_{(u^*),1}\right|.$$

The middle term is at most $C - c$ and the last at most $\|\acute{\sigma}^2_{(u^*)} - \breve{\sigma}^2_{(u^*)}\| = o_p(\bar{T}^{1/2})$. The first

has mean zero and variance

$$\left(\breve{\mathbf{\Gamma}}_{(u^*)}\right)_{11} \leq 2\left(\mathbf{\Gamma}^0_{(u^*)}\right)_{11} + 2\tilde{\mu}^2_{(u^*),1}\left\|\mathbf{F}_{(u^*)}\right\|_2 \leq 2\left\|\mathbf{\Gamma}^0_{(u^*)}\right\|_2 + 2\left\|\mathbf{F}_{(u^*)}\right\|_2 \max_t \tilde{\mu}^2_{(u^*),t} = o(\bar{T}),$$

by Lemma 28(iv), Lemma 38 and (48), so it is $o_p(\bar{T}^{1/2})$ by Chebyshev's inequality. In contrast to Lemma 25, this entry need not be $O_p(1)$, because the term of $\breve{\sigma}^2_{(u^*)}$ involving the fitted values is unbounded. The order $o_p(\bar{T}^{1/2})$ suffices for the present purpose. Combining these bounds yields the first claim. Moreover, $\|\acute{\sigma}^2_{(u)}\| \leq \|\breve{\sigma}^2_{(u)}\| + o_p(\bar{T}^{1/2}) = O_p(\bar{T}^{1/2})$, since

$$\mathbb{E}\left[\|\breve{\sigma}^2_{(u)}\|^2\right] \leq \|\sigma^2_{(u)}\|^2 + \operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(u)}\right) = O(\bar{T}).$$

□

Lemma 30 (Consistency of the plug-in traces based on $\acute{\sigma}^2$)**.** *Let u_1 and u_2 be two blocks coming from distinct subsamples, $u_1^* \neq u_2^*$, let $\mathbf{W}$ be admissible, and let $\hat{\mathbf{W}}$ satisfy $\left\|\hat{\mathbf{W}} - \mathbf{W}\right\|_F = o_p(1)$. Under Assumptions* Feedback Decomposition, Bounds, Regular Errors, Proportional Regime, No Influential Observations, Bounded Leverage, Bounded Fitted Values *and* Leave-One-Out Conditioning*,*

$$\acute{\vartheta} := \sum_{t,j} \hat{w}_{tj}\, \acute{\sigma}^2_{(u_1),t}\, \acute{\sigma}^2_{(u_2),j} \qquad \textit{satisfies} \qquad \acute{\vartheta} - \vartheta = o_p(\bar{T}),$$

with ϑ as in Lemma 27. Assumption Sub-One-Half Leverage *is not required.*

Proof. The proof follows that of Lemma 27, with $\acute{\sigma}^2$, $\breve{\sigma}^2$ and $\breve{\mathbf{\Gamma}}$ in place of $\grave{\sigma}^2$, $\breve{\sigma}^2$ and $\check{\mathbf{\Gamma}}$, and with $\breve{\theta}_{(u_1)} := \breve{\sigma}^2_{(u_1)} - \sigma^2_{(u_1)}$ and $\delta_{(u_1)} := \acute{\sigma}^2_{(u_1)} - \breve{\sigma}^2_{(u_1)}$. The properties of the error terms used there hold here as well: $\mathbb{E}\left[\breve{\theta}_{(u_1)}\right] = 0$ by Lemma 28(iii), and $\breve{\theta}_{(u_1)}$ and $\breve{\theta}_{(u_2)}$ are independent by Lemma 28(ii). Further, $\|\breve{\theta}_{(u_1)}\| = O_p(\bar{T}^{1/2})$ follows from $\mathbb{E}\left[\|\breve{\theta}_{(u_1)}\|^2\right] = \operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(u_1)}\right) = O(\bar{T})$ (Lemma 28(iv)) and Markov's inequality. Finally, $\|\delta_{(u_1)}\| = o_p(\bar{T}^{1/2})$ and $\|\acute{\sigma}^2_{(u_1)}\| = O_p(\bar{T}^{1/2})$ by Lemma 29. The bounds on the weight error and on the groups (b) and (d) in the proof of Lemma 27 rely on these properties only and therefore remain valid. The bounds on (a) and (c) relied on the spectral bound on $\check{\mathbf{\Gamma}}$ of Lemma 22(iv), which Lemma 28(iv) replaces by a Loewner bound containing an unbounded fitted-value component. These two bounds are

therefore established anew. Recall that $\breve{\mathbf{\Gamma}}_{(u_1)} \preceq 2\mathbf{\Gamma}^0_{(u_1)} + 2\mathbf{\Gamma}^\mu_{(u_1)}$ for every block u, by Lemma 28(iv).

Three elementary facts are used. (F1) Condition (W1) implies $|w_{tj}| \leq 1$, so that $\sum_j w_{tj}^2 \leq \sum_j |w_{tj}| \leq 1$, and likewise for the sums over t. (F2) For any positive semidefinite matrix $\mathbf{B}$ and any diagonal matrix $\mathbf{D}$ of the same dimension, $\mathbf{DBD} \preceq \|\mathbf{B}\|_2 \mathbf{D}^2$, since $\mathbf{D}(\|\mathbf{B}\|_2 \mathbf{I} - \mathbf{B})\mathbf{D} \succeq 0$. (F3) If $\mathbf{Y} \succeq 0$ and $\mathbf{U} \preceq \tau\mathbf{I}$ is symmetric, then $\operatorname{tr}(\mathbf{YU}) \leq \tau \operatorname{tr}(\mathbf{Y})$. Consequently, $\mathbf{U}_1 \preceq \mathbf{U}_2$ and $\mathbf{Y} \succeq 0$ imply $\operatorname{tr}(\mathbf{U}_1\mathbf{Y}) \leq \operatorname{tr}(\mathbf{U}_2\mathbf{Y})$.

Consider first the terms in (a). As in the proof of Lemma 27, each of them is of the form $a'\breve{\theta}_{(u_2)}$ with deterministic a satisfying $|a_j| \leq C$, and has mean zero. Moreover,

$$\begin{aligned} a'\breve{\mathbf{\Gamma}}_{(u_2)}a &\leq 2a'\mathbf{\Gamma}^0_{(u_2)}a + 2a'\mathbf{\Gamma}^\mu_{(u_2)}a \leq 2\left\|\mathbf{\Gamma}^0_{(u_2)}\right\|_2 \|a\|^2 + 2\left\|\mathbf{F}_{(u_2)}\right\|_2 \sum_j a_j^2 \tilde{\mu}^2_{(u_2),j} \\ &\leq 2C^2\bar{T}\left\|\mathbf{\Gamma}^0_{(u_2)}\right\|_2 + 2C^2\left\|\mathbf{F}_{(u_2)}\right\|_2 \|\tilde{\mu}_{(u_2)}\|^2 = O(\bar{T}), \end{aligned}$$

because $\left\|\mathbf{\Gamma}^0_{(u_2)}\right\|_2 = O(1)$ and $\left\|\mathbf{F}_{(u_2)}\right\|_2 = O(1)$ by Lemma 28(iv) and by (47), so that the term is $O_p(\bar{T}^{1/2}) = o_p(\bar{T})$ by Chebyshev's inequality.

Turning to the term in (c), put $\mathcal{S} := \sum_{t,j} w_{tj}\breve{\theta}_{(u_1),t}\breve{\theta}_{(u_2),j}$. The arguments of the proof of Lemma 27 again give $\mathbb{E}[\mathcal{S}] = 0$ and $\operatorname{Var}(\mathcal{S}) = \operatorname{tr}\left(\mathbf{W}'\breve{\mathbf{\Gamma}}_{(u_1)}\mathbf{W}\breve{\mathbf{\Gamma}}_{(u_2)}\right)$. By cyclicity, this trace equals $\operatorname{tr}\left(\breve{\mathbf{\Gamma}}_{(u_1)}\left(\mathbf{W}\breve{\mathbf{\Gamma}}_{(u_2)}\mathbf{W}'\right)\right)$ with $\mathbf{W}\breve{\mathbf{\Gamma}}_{(u_2)}\mathbf{W}' \succeq 0$, so that, by the second part of (F3), each covariance matrix may be replaced by its Loewner bound:

$$\operatorname{Var}(\mathcal{S}) \leq 4 \sum_{\iota,\iota' \in \{0,\mu\}} \operatorname{tr}\left(\mathbf{W}'\mathbf{\Gamma}^\iota_{(u_1)}\mathbf{W}\mathbf{\Gamma}^{\iota'}_{(u_2)}\right).$$

By two applications of (F3) and by (W2), the term with $(\iota,\iota') = (0,0)$ is at most

$$\left\|\mathbf{\Gamma}^0_{(u_1)}\right\|_2 \left\|\mathbf{\Gamma}^0_{(u_2)}\right\|_2 \|\mathbf{W}\|_F^2 = O(1)\|\mathbf{W}\|_F^2 = O(\bar{T}),$$

as in the proof of Lemma 27. For $(\iota,\iota') = (\mu,0)$, (F2) gives $\mathbf{\Gamma}^\mu_{(u_1)} \preceq \left\|\mathbf{F}_{(u_1)}\right\|_2 \mathbf{D}^2_{(u_1)}$, so that,

by (F3) and (F1),

$$\begin{aligned}\operatorname{tr}\left(\mathbf{\Gamma}^{\mu}_{(u_1)}\left(\mathbf{W}\mathbf{\Gamma}^{0}_{(u_2)}\mathbf{W}'\right)\right) &\leq \left\|\mathbf{F}_{(u_1)}\right\|_2 \left\|\mathbf{\Gamma}^{0}_{(u_2)}\right\|_2 \operatorname{tr}\left(\mathbf{W}'\mathbf{D}^2_{(u_1)}\mathbf{W}\right) \\ &= \left\|\mathbf{F}_{(u_1)}\right\|_2 \left\|\mathbf{\Gamma}^{0}_{(u_2)}\right\|_2 \sum_{t,j} \tilde{\mu}^2_{(u_1),t} w^2_{tj} \\ &\leq \left\|\mathbf{F}_{(u_1)}\right\|_2 \left\|\mathbf{\Gamma}^{0}_{(u_2)}\right\|_2 \|\tilde{\mu}_{(u_1)}\|^2 = O(\bar{T}),\end{aligned}$$

and the term with $(\iota, \iota') = (0, \mu)$ is bounded in the same way with the roles of u_1 and u_2 exchanged. For $(\iota, \iota') = (\mu, \mu)$, applying (F2) to each factor and then (F1) gives

$$\begin{aligned}\operatorname{tr}\left(\mathbf{W}'\mathbf{\Gamma}^{\mu}_{(u_1)}\mathbf{W}\mathbf{\Gamma}^{\mu}_{(u_2)}\right) &\leq \left\|\mathbf{F}_{(u_1)}\right\|_2 \left\|\mathbf{F}_{(u_2)}\right\|_2 \sum_{t,j} \tilde{\mu}^2_{(u_1),t} w^2_{tj} \tilde{\mu}^2_{(u_2),j} \\ &\leq \left\|\mathbf{F}_{(u_1)}\right\|_2 \left\|\mathbf{F}_{(u_2)}\right\|_2 \left(\max_j \tilde{\mu}^2_{(u_2),j}\right) \|\tilde{\mu}_{(u_1)}\|^2,\end{aligned}$$

which is $o(\bar{T})O(\bar{T}) = o(\bar{T}^2)$ by (48) and (47). Hence $\operatorname{Var}(\mathcal{S}) = o(\bar{T}^2)$, and $\mathcal{S} = o_p(\bar{T})$ by Chebyshev's inequality. This rate is slower than the rate $O_p(\bar{T}^{1/2})$ obtained in the proof of Lemma 27, but it suffices. The difference is due to the fitted-value component, and this is the only step of the proof that uses the bound (48) on the largest fitted value. □

Theorem 4 (Conservative inference with the Jochmans estimator)**.** *For each $v \in \{1, \ldots, h+1\}$, let $\acute{\Sigma}_v$ and $\acute{\Lambda}$ be $\hat{\Sigma}_v$ and $\hat{\Lambda}$ with $\hat{\sigma}^2_t = \acute{\sigma}^2_t$, respectively. Under Assumptions* Feedback Decomposition*,* Bounds*,* Regular Errors*,* Proportional Regime*,* No Influential Observations*,* Bounded Leverage*,* Bounded Fitted Values*, and* Leave-One-Out Conditioning*, we have*

$$\liminf_{T\to\infty} \mathbb{P}\left[r'\beta \in \left[r'\hat{\beta} \pm \frac{z_{1-\alpha/2}}{h+1}\sqrt{\left(1+\acute{\Lambda}\right)\sum_{v=1}^{h+1}\acute{\Sigma}_v}\right]\right] \geq 1-\alpha$$

for every $\alpha \in (0,1)$, where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

Proof. The proof of Theorem 3 uses the estimates $\grave{\sigma}^2_t$ only through three properties. These are the consistency of the scaled traces, which came from Lemma 27, the ratio $\left(\sum_{v=1}^{h+1}\grave{\Sigma}_v\right)/\sum_{v=1}^{h+1}\Sigma_v \xrightarrow{p} 1$, which came from Lemma 19, and the fact that the variance

estimate is well defined with probability tending to one. Let $\acute{c}_{v,q}$, $\acute{d}_{v,q}$ and $\acute{d}_{q,v}$ denote the estimates $\hat{c}_{v,q}$, $\hat{d}_{v,q}$ and $\hat{d}_{q,v}$ evaluated at $\hat{\sigma}^2 = \acute{\sigma}^2$, and $\acute{\lambda}$ the ratio in the definition of $\hat{\Lambda}$ evaluated at $\acute{\sigma}^2$, so that $\acute{\Lambda} = \min\{1, \acute{\lambda}\}$.

Lemma 30, together with Lemmas 21 and 26, supplies the first for $\acute{c}_{v,q}$, $\acute{d}_{v,q}$ and $\acute{d}_{q,v}$, so that the argument of that proof leading from (46) to $\grave{\Lambda} - \Lambda \xrightarrow{p} 0$ gives $\acute{\lambda} - \lambda \xrightarrow{p} 0$ and hence $\acute{\Lambda} - \Lambda \xrightarrow{p} 0$. Note that Lemma 26 uses only $\eta = o_p(1)$ and so does not require Assumption *Sub-One-Half Leverage*. Lemma 15 supplies $\left(\sum_{v=1}^{h+1} \acute{\Sigma}_v\right) / \sum_{v=1}^{h+1} \Sigma_v \xrightarrow{p} 1$. Finally $\acute{\sigma}_t^2$ is well defined whenever $\min_s \mathbf{H}_{(v),ss} > 0$ for every v, which holds on the events $\mathcal{E}_{v,T}$ of the proof of Lemma 29, whose probability tends to one. Alternatively, it follows from Lemma 14. Theorem 2 is used in the same way as before.

With these substitutions, the final part of the proof of Theorem 3, which establishes the coverage, applies without modification, with $\acute{\Lambda}$ and $\sum_{v=1}^{h+1} \acute{\Sigma}_v$ in place of $\grave{\Lambda}$ and $\sum_{v=1}^{h+1} \grave{\Sigma}_v$, and yields

$$\liminf_{T\to\infty} \mathbb{P}\left[r'\beta \in \left[r'\hat{\beta} \pm \frac{z_{1-\alpha/2}}{h+1} \sqrt{\left(1 + \acute{\Lambda}\right) \sum_{v=1}^{h+1} \acute{\Sigma}_v} \right] \right] \;\geq\; 1 - \alpha$$

for every $\alpha \in (0, 1)$. In particular this holds without Assumption *Sub-One-Half Leverage*, for any aspect ratio permitted by Assumption *Proportional Regime*. □

B Matrix basics

B.1 Matrix norms

A matrix norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$ satisfies $\|\mathbf{A}\| \geq 0$, with $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}$ (non-negativity), $\|\alpha\mathbf{A}\| = |\alpha| \, \|\mathbf{A}\|$ for all $\alpha \in \mathbb{R}$ (homogeneity), and $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ (triangle inequality). A matrix norm is *submultiplicative* if $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|$ whenever the product is defined.

B.1.1 The spectral norm

The spectral (operator) norm is

$$\|\mathbf{A}\|_2 = \sup_{\|x\|_2=1} \|\mathbf{A}x\|_2.$$

It satisfies

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}'\mathbf{A})}.$$

Thus, if $\mathbf{A}$ is symmetric, $\|\mathbf{A}\|_2 = \max_i |\lambda_i(\mathbf{A})|$, and $\|\mathbf{A}\|_2 = \lambda_{\max}(\mathbf{A})$ if $\mathbf{A}$ is in addition positive semidefinite. Moreover, $\|\mathbf{A}\|_2 = \|\mathbf{A}'\|_2$, and the spectral norm is submultiplicative.

B.1.2 Induced norms and the maximum row- and column-sum norms

Every vector norm $\|\cdot\|$ induces a matrix norm via

$$\|\mathbf{A}\| = \sup_{x \neq 0} \frac{\|\mathbf{A}x\|}{\|x\|},$$

so that $\|\mathbf{A}x\| \leq \|\mathbf{A}\| \|x\|$ for all x by construction; this immediately makes every induced norm submultiplicative, since $\|\mathbf{AB}x\| \leq \|\mathbf{A}\| \|\mathbf{B}x\| \leq \|\mathbf{A}\| \|\mathbf{B}\| \|x\|$. The spectral norm is the norm induced by the Euclidean vector norm. Two further induced norms are used in this paper.

Lemma 31 (Maximum absolute row sum)**.** *The matrix norm induced by the max-norm* $\|x\|_\infty = \max_i |x_i|$ *is the maximum absolute row sum,*

$$\|\mathbf{A}\|_\infty = \max_i \sum_j |\mathbf{A}_{ij}|.$$

Proof. For any x with $\|x\|_\infty \leq 1$,

$$\|\mathbf{A}x\|_\infty = \max_i \left| \sum_j \mathbf{A}_{ij} x_j \right| \leq \max_i \sum_j |\mathbf{A}_{ij}| \, |x_j| \leq \max_i \sum_j |\mathbf{A}_{ij}|,$$

so the induced norm is at most the maximum absolute row sum. Conversely, assume $\mathbf{A} \neq \mathbf{0}$, let $i^\star$ attain the maximum and set $x_j = \operatorname{sgn}(a_{i^\star j})$, so that $\|x\|_\infty \leq 1$ and the $i^\star$-th entry of

$\mathbf{A}x$ equals $\sum_j |a_{i^\star j}|$. Hence $\|\mathbf{A}x\|_\infty \geq \sum_j |a_{i^\star j}|$, and the two bounds coincide. The absolute values are essential: without the sign adjustment, entries of opposite sign within a row would cancel. □

Correspondingly, the maximum absolute column sum is

$$\|\mathbf{A}\|_1 = \max_j \sum_i |\mathbf{A}_{ij}| = \|\mathbf{A}'\|_\infty,$$

the second equality holding because the rows of $\mathbf{A}'$ are the columns of $\mathbf{A}$; an argument analogous to that of Lemma 31 shows that this is the norm induced by the vector norm $\|x\|_1 = \sum_i |x_i|$.

Lemma 32 (Eigenvalues are bounded by any induced norm)**.** *Let $\|\cdot\|$ be any induced matrix norm and let λ be any eigenvalue of $\mathbf{A} \in \mathbb{R}^{n\times n}$. Then $|\lambda| \leq \|\mathbf{A}\|$. In particular the spectral radius satisfies $\rho(\mathbf{A}) := \max_i |\lambda_i(\mathbf{A})| \leq \|\mathbf{A}\|$.*

Proof. Let $x \neq 0$ satisfy $\mathbf{A}x = \lambda x$. Then

$$|\lambda|\,\|x\| = \|\lambda x\| = \|\mathbf{A}x\| \leq \|\mathbf{A}\|\,\|x\|,$$

and dividing by $\|x\| > 0$ gives the claim. Only the defining property $\|\mathbf{A}x\| \leq \|\mathbf{A}\|\,\|x\|$ of an induced norm is used, so the bound holds for $\|\cdot\|_2$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ alike. (Eigenvalues and eigenvectors of a general square matrix may be complex, in which case the norm is read on $\mathbb{C}^n$; the formula of Lemma 31 is unchanged there, with the sign choice replaced by the unit-modulus phases $x_j = \overline{a_{i^\star j}}/|a_{i^\star j}|$.) □

Lemma 33 (Interpolation between the row- and column-sum norms)**.** *For any $\mathbf{A} \in \mathbb{R}^{n\times n}$,*

$$\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_1\,\|\mathbf{A}\|_\infty}.$$

If $\mathbf{A}$ is symmetric, then $\|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$ and hence $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_\infty$.

Proof. The matrix $\mathbf{A}'\mathbf{A}$ is symmetric positive semidefinite, so its eigenvalues are real and non-negative and $\|\mathbf{A}\|_2^2 = \lambda_{\max}(\mathbf{A}'\mathbf{A}) = \rho(\mathbf{A}'\mathbf{A})$. Applying Lemma 32 with the maximum-

row-sum norm and then submultiplicativity,

$$\|\mathbf{A}\|_2^2 = \rho(\mathbf{A}'\mathbf{A}) \le \|\mathbf{A}'\mathbf{A}\|_\infty \le \|\mathbf{A}'\|_\infty \|\mathbf{A}\|_\infty = \|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty.$$

If $\mathbf{A} = \mathbf{A}'$, its rows and columns coincide, so $\|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$ and the bound reads $\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_\infty$. □

This lemma transfers a bound in the max-norm to the spectral norm. The two norms are equivalent only up to a factor depending on the dimension, but for symmetric matrices the interpolation inequality reduces to $\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_\infty$, so that a row-sum bound applies to the spectral norm without such a factor.

B.1.3 Frobenius norm

The Frobenius norm is

$$\|\mathbf{A}\|_F = \left(\sum_{i,j} \mathbf{A}_{ij}^2 \right)^{1/2} = \sqrt{\operatorname{tr}(\mathbf{A}'\mathbf{A})}.$$

B.1.4 Inequalities between norms

The spectral and Frobenius norms are related by

$$\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_F \le \sqrt{\operatorname{rank}(\mathbf{A})}\, \|\mathbf{A}\|_2.$$

Both norms are submultiplicative,

$$\|\mathbf{AB}\|_2 \le \|\mathbf{A}\|_2 \|\mathbf{B}\|_2, \qquad \|\mathbf{AB}\|_F \le \|\mathbf{A}\|_F \|\mathbf{B}\|_F,$$

and for the Frobenius norm the tighter bound $\|\mathbf{AB}\|_F \le \|\mathbf{A}\|_2 \|\mathbf{B}\|_F$ holds. The trace inequality

$$|\operatorname{tr}(\mathbf{A}'\mathbf{B})| \le \|\mathbf{A}\|_F \|\mathbf{B}\|_F$$

is the Cauchy–Schwarz inequality for the trace inner product. The following two lemmas bound traces of products in terms of the trace of a positive semidefinite factor.

Lemma 34 (Trace-norm inequality)**.** *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be symmetric positive semidefinite and let* $\mathbf{B} \in \mathbb{R}^{n\times n}$ *be arbitrary. Then*

$$|\mathrm{tr}(\mathbf{AB})| \leq \|\mathbf{B}\|_2 \ \mathrm{tr}(\mathbf{A}).$$

Proof. Since $\mathbf{A}$ is symmetric positive semidefinite, it admits an eigenvalue decomposition

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i u_i u_i',$$

where $\lambda_i \geq 0$ for all i, and $\{u_i\}_{i=1}^n$ is an orthonormal basis of $\mathbb{R}^n$. Using linearity and cyclicity of the trace, we obtain

$$\mathrm{tr}(\mathbf{AB}) = \sum_{i=1}^{n} \lambda_i \,\mathrm{tr}(u_i u_i' \mathbf{B}) = \sum_{i=1}^{n} \lambda_i \, u_i' \mathbf{B} u_i,$$

and hence

$$|\,\mathrm{tr}(\mathbf{AB})| \leq \sum_{i=1}^{n} \lambda_i \, |u_i' \mathbf{B} u_i|.$$

By the Cauchy–Schwarz inequality, $|u_i' \mathbf{B} u_i| \leq \|u_i\| \, \|\mathbf{B}u_i\| \leq \|\mathbf{B}\|_2$ for each i. Therefore,

$$|\,\mathrm{tr}(\mathbf{AB})| \leq \|\mathbf{B}\|_2 \sum_{i=1}^{n} \lambda_i = \|\mathbf{B}\|_2 \,\mathrm{tr}(\mathbf{A}). \qquad \square$$

Lemma 35 (Trace lower bound)**.** *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be symmetric positive semidefinite and let* $\mathbf{B} \in \mathbb{R}^{n\times n}$ *be symmetric. Then*

$$\mathrm{tr}(\mathbf{AB}) \geq \lambda_{\min}(\mathbf{B}) \,\mathrm{tr}(\mathbf{A}).$$

Proof. Since $\mathbf{A}$ is symmetric positive semidefinite, it has the spectral decomposition $\mathbf{A} = \sum_{i=1}^n \lambda_i(\mathbf{A}) \, u_i u_i'$, where $\lambda_i(\mathbf{A}) \geq 0$ and $\{u_i\}_{i=1}^n$ is an orthonormal basis of $\mathbb{R}^n$. Therefore

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}\left(\sum_{i=1}^{n} \lambda_i(\mathbf{A}) \, u_i u_i' \mathbf{B} \right) = \sum_{i=1}^{n} \lambda_i(\mathbf{A}) \, u_i' \mathbf{B} \, u_i.$$

Since $\mathbf{B}$ is symmetric, the Rayleigh quotient satisfies $u_i'\mathbf{B}\,u_i \geq \lambda_{\min}(\mathbf{B})\|u_i\|^2 = \lambda_{\min}(\mathbf{B})$ for each i. Hence

$$\operatorname{tr}(\mathbf{AB}) \geq \lambda_{\min}(\mathbf{B}) \sum_{i=1}^{n} \lambda_i(\mathbf{A}) = \lambda_{\min}(\mathbf{B}) \operatorname{tr}(\mathbf{A}). \qquad \square$$

B.1.5 Bounds on columns and rows

Let $\mathbf{A} = [a_1, \ldots, a_n]$ with columns a_j. The largest column norm is bounded by the spectral norm, $\max_j \|a_j\|_2 \leq \|\mathbf{A}\|_2$, and similarly for rows, $\max_i \|\mathbf{A}_{i\cdot}\|_2 \leq \|\mathbf{A}\|_2$.

B.1.6 Spectral norm and Frobenius norm of projection matrices

Let $\mathbf{P}$ be an orthogonal projection matrix, that is, $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}' = \mathbf{P}$. Its eigenvalues are 0 or 1; hence $\|\mathbf{P}\|_2 = 1$ unless $\mathbf{P} = \mathbf{0}$. Moreover,

$$\|\mathbf{P}\|_F = \sqrt{\operatorname{tr}(\mathbf{P})} = \sqrt{\operatorname{rank}(\mathbf{P})}.$$

B.2 Eigenvalues

Lemma 36 (Weyl's inequality)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be symmetric* $n \times n$ *matrices, and let* $\lambda_1(\mathbf{Q}) \leq \cdots \leq \lambda_n(\mathbf{Q})$ *denote the eigenvalues of a symmetric* $n \times n$ *matrix* $\mathbf{Q}$*. Then, for* i, j *with* $1 \leq i + j - 1 \leq n$*,*

$$\lambda_i(\mathbf{A}) + \lambda_j(\mathbf{B}) \leq \lambda_{i+j-1}(\mathbf{A} + \mathbf{B}).$$

In particular, $\lambda_{\min}(\mathbf{A} + \mathbf{B}) \geq \lambda_{\min}(\mathbf{A}) + \lambda_{\min}(\mathbf{B})$*.*

For any real matrix $\mathbf{A}$, the non-zero eigenvalues of $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ coincide.

Lemma 37 (Smallest eigenvalue of a product)**.** *Let* $\mathbf{A} \in \mathbb{R}^{m\times k}$ *and* $\mathbf{B} \in \mathbb{R}^{n\times m}$*. Then*

$$\lambda_{\min}(\mathbf{A}'\mathbf{A})\lambda_{\min}(\mathbf{B}'\mathbf{B}) \leq \lambda_{\min}(\mathbf{A}'\mathbf{B}'\mathbf{BA}),$$

and

$$\lambda_{\max}(\mathbf{A}'\mathbf{B}'\mathbf{BA}) \leq \lambda_{\max}(\mathbf{A}'\mathbf{A})\lambda_{\max}(\mathbf{B}'\mathbf{B}).$$

Proof. For any vector x,

$$x'\mathbf{A}'\mathbf{B}'\mathbf{BA}x = (\mathbf{A}x)'(\mathbf{B}'\mathbf{B})(\mathbf{A}x).$$

Using the Rayleigh quotient bound,

$$(\mathbf{A}x)'(\mathbf{B}'\mathbf{B})(\mathbf{A}x) \geq \lambda_{\min}(\mathbf{B}'\mathbf{B})\|\mathbf{A}x\|^2.$$

Since

$$\|\mathbf{A}x\|^2 = x'\mathbf{A}'\mathbf{A}x \geq \lambda_{\min}(\mathbf{A}'\mathbf{A})\|x\|^2,$$

it follows that

$$x'\mathbf{A}'\mathbf{B}'\mathbf{B}\mathbf{A}x \geq \lambda_{\min}(\mathbf{B}'\mathbf{B})\lambda_{\min}(\mathbf{A}'\mathbf{A})\|x\|^2.$$

Dividing by $\|x\|^2$ and taking the minimum over $x \neq 0$ yields

$$\lambda_{\min}(\mathbf{A}'\mathbf{B}'\mathbf{B}\mathbf{A}) \geq \lambda_{\min}(\mathbf{A}'\mathbf{A})\lambda_{\min}(\mathbf{B}'\mathbf{B}).$$

Similarly,

$$x'\mathbf{A}'\mathbf{B}'\mathbf{B}\mathbf{A}x = (\mathbf{A}x)'(\mathbf{B}'\mathbf{B})(\mathbf{A}x) \leq \lambda_{\max}(\mathbf{B}'\mathbf{B})\|\mathbf{A}x\|^2,$$

and

$$\|\mathbf{A}x\|^2 \leq \lambda_{\max}(\mathbf{A}'\mathbf{A})\|x\|^2.$$

Therefore

$$x'\mathbf{A}'\mathbf{B}'\mathbf{B}\mathbf{A}x \leq \lambda_{\max}(\mathbf{B}'\mathbf{B})\lambda_{\max}(\mathbf{A}'\mathbf{A})\|x\|^2.$$

Dividing by $\|x\|^2$ and taking the maximum over $x \neq 0$ gives

$$\lambda_{\max}(\mathbf{A}'\mathbf{B}'\mathbf{B}\mathbf{A}) \leq \lambda_{\max}(\mathbf{A}'\mathbf{A})\lambda_{\max}(\mathbf{B}'\mathbf{B}).$$

□

Lemma 38 (Eigenvalue bounds for principal submatrices)**.** *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be symmetric, and let* $\mathbf{B}$ *be a principal submatrix of* $\mathbf{A}$*. Then*

$$\lambda_{\min}(\mathbf{A}) \leq \lambda_{\min}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{A}).$$

Proof. After a simultaneous permutation of rows and columns, which leaves the eigenvalues

of $\mathbf{A}$ unchanged, write

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D} \end{pmatrix},$$

where $\mathbf{B}$ is a $k \times k$ principal submatrix of $\mathbf{A}$.

For any $x \in \mathbb{R}^k$,

$$x'\mathbf{B}x = \begin{pmatrix} x \\ 0 \end{pmatrix}' \mathbf{A} \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Hence,

$$x'\mathbf{B}x \geq \lambda_{\min}(\mathbf{A})\|x\|^2.$$

Dividing by $\|x\|^2$ and taking the minimum over $x \neq 0$ yields

$$\lambda_{\min}(\mathbf{B}) \geq \lambda_{\min}(\mathbf{A}).$$

Similarly,

$$x'\mathbf{B}x \leq \lambda_{\max}(\mathbf{A})\|x\|^2,$$

and dividing by $\|x\|^2$ and taking the maximum over $x \neq 0$ gives

$$\lambda_{\max}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{A}).$$

Combining the two inequalities gives

$$\lambda_{\min}(\mathbf{A}) \leq \lambda_{\min}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{B}) \leq \lambda_{\max}(\mathbf{A}).$$

□

B.3 Diagonal dominance

A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is *strictly diagonally dominant* if each diagonal entry exceeds, in absolute value, the sum of the absolute values of the remaining entries in its row:

$$|\mathbf{A}_{ii}| > \sum_{j\neq i} |\mathbf{A}_{ij}|, \qquad i = 1, \dots, n.$$

We say that the dominance holds *with margin* $\alpha > 0$ if

$$|\mathbf{A}_{ii}| - \sum_{j\neq i} |\mathbf{A}_{ij}| \;\geq\; \alpha, \qquad i = 1, \dots, n.$$

Lemma 39 (Gershgorin's circle theorem)**.** *Every eigenvalue λ of $\mathbf{A}$ lies in one of the discs*

$$D_i = \Big\{ z \;:\; |z - \mathbf{A}_{ii}| \leq \sum_{j\neq i} |\mathbf{A}_{ij}| \Big\}, \qquad i = 1, \dots, n.$$

Proof. Let $\mathbf{A}x = \lambda x$ with $x \neq 0$ and choose i with $|x_i| = \max_j |x_j| > 0$. The i-th row of the eigenvalue equation reads $\sum_j \mathbf{A}_{ij} x_j = \lambda x_i$, that is,

$$(\lambda - \mathbf{A}_{ii})\, x_i = \sum_{j\neq i} \mathbf{A}_{ij} x_j.$$

Taking absolute values and using $|x_j| \leq |x_i|$ for all j,

$$|\lambda - \mathbf{A}_{ii}|\, |x_i| \;\leq\; \sum_{j\neq i} |\mathbf{A}_{ij}|\, |x_j| \;\leq\; |x_i| \sum_{j\neq i} |\mathbf{A}_{ij}|,$$

and dividing by $|x_i| > 0$ places λ in D_i. □

Gershgorin's theorem also gives a second proof of Lemma 32 for the max-norm, since every disc D_i is contained in the disc of radius $|\mathbf{A}_{ii}| + \sum_{j\neq i} |\mathbf{A}_{ij}| = \sum_j |\mathbf{A}_{ij}| \leq \|\mathbf{A}\|_\infty$ centred at the origin. Strict diagonal dominance means that no Gershgorin disc contains the origin, which yields the following lemma.

Lemma 40 (Strict diagonal dominance implies invertibility)**.** *If $\mathbf{A}$ is strictly diagonally dominant, then $\mathbf{A}$ is invertible.*

Proof. Suppose $\mathbf{A}x = 0$ for some $x \neq 0$. Normalising, we may assume $\|x\|_\infty = 1$; pick i with $|x_i| = 1$. The i-th row gives $\mathbf{A}_{ii}x_i = -\sum_{j\neq i} \mathbf{A}_{ij}x_j$, so

$$|\mathbf{A}_{ii}| \;\le\; \sum_{j\neq i} |\mathbf{A}_{ij}|\,|x_j| \;\le\; \sum_{j\neq i} |\mathbf{A}_{ij}|,$$

contradicting strict dominance. Hence the null space of $\mathbf{A}$ is trivial. Equivalently, by Lemma 39 every eigenvalue lies in a disc centred at $\mathbf{A}_{ii}$ of radius $\sum_{j\neq i}|\mathbf{A}_{ij}| < |\mathbf{A}_{ii}|$, and no such disc contains the origin, so 0 is not an eigenvalue of $\mathbf{A}$. □

The following result is due to Varah (1975).

Lemma 41 (Varah's bound)**.** *If* $\mathbf{A}$ *is strictly diagonally dominant with margin* $\alpha > 0$*, then*

$$\|\mathbf{A}^{-1}\|_\infty \;\le\; \frac{1}{\alpha}.$$

If in addition $\mathbf{A}$ *is symmetric, then* $\|\mathbf{A}^{-1}\|_2 \le 1/\alpha$*.*

Proof. Let x satisfy $\|x\|_\infty = 1$ and pick i with $|x_i| = 1$. Bounding $\|\mathbf{A}x\|_\infty$ below by the absolute value of its i-th entry,

$$\|\mathbf{A}x\|_\infty \;\ge\; \Big|\mathbf{A}_{ii}x_i + \sum_{j\neq i}\mathbf{A}_{ij}x_j\Big| \;\ge\; |\mathbf{A}_{ii}| - \sum_{j\neq i}|\mathbf{A}_{ij}| \;\ge\; \alpha,$$

using $|x_i| = 1$ and $|x_j| \le 1$. By homogeneity, $\|\mathbf{A}x\|_\infty \ge \alpha\|x\|_\infty$ for every x. The matrix $\mathbf{A}$ is invertible by Lemma 40, so substituting $x = \mathbf{A}^{-1}y$ gives $\|y\|_\infty \ge \alpha\,\|\mathbf{A}^{-1}y\|_\infty$ for every y, i.e., $\|\mathbf{A}^{-1}\|_\infty \le 1/\alpha$. If $\mathbf{A}$ is symmetric then so is $\mathbf{A}^{-1}$, and Lemma 33 upgrades the bound to the spectral norm, $\|\mathbf{A}^{-1}\|_2 \le \|\mathbf{A}^{-1}\|_\infty \le 1/\alpha$. □

B.4 Hadamard products

Lemma 42 (Hadamard products)**.** *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$*.*

(i) (Schur product theorem.) If $\mathbf{A} \succeq 0$ *and* $\mathbf{B} \succeq 0$*, then* $\mathbf{A} \odot \mathbf{B} \succeq 0$*.*

(ii) If $\mathbf{A} \succeq 0$ *and* $\mathbf{B} \succeq 0$, *then*

$$\lambda_{\max}(\mathbf{A} \odot \mathbf{B}) \leq \left(\max_i \mathbf{A}_{ii}\right) \lambda_{\max}(\mathbf{B}),$$

and in particular $\|\mathbf{A} \odot \mathbf{B}\|_2 \leq (\max_i \mathbf{A}_{ii}) \|\mathbf{B}\|_2$.

(iii) For arbitrary $\mathbf{A}$ *and* $\mathbf{B}$,

$$\|\mathbf{A} \odot \mathbf{B}\|_2 \leq \left(\max_i \|\mathbf{A}_{i\cdot}\|\right) \left(\max_j \|\mathbf{B}_{\cdot j}\|\right),$$

the largest row norm of $\mathbf{A}$ *times the largest column norm of* $\mathbf{B}$.

Proof. (i) Write $\mathbf{B} = \sum_k \lambda_k u_k u_k'$ with $\lambda_k \geq 0$. For any vector u,

$$\mathbf{A} \odot (uu') = \operatorname{diag}(u)\, \mathbf{A} \operatorname{diag}(u) \succeq 0.$$

Hence $\mathbf{A} \odot \mathbf{B} = \sum_k \lambda_k \mathbf{A} \odot (u_k u_k')$ is a non-negative combination of positive semidefinite matrices and therefore positive semidefinite.

(ii) Since $\lambda_{\max}(\mathbf{B})\, \mathbf{I}_n - \mathbf{B} \succeq 0$, part (i) gives $\mathbf{A} \odot (\lambda_{\max}(\mathbf{B})\, \mathbf{I}_n - \mathbf{B}) \succeq 0$, that is

$$\mathbf{A} \odot \mathbf{B} \preceq \lambda_{\max}(\mathbf{B})\, (\mathbf{A} \odot \mathbf{I}_n) = \lambda_{\max}(\mathbf{B})\, \operatorname{diag}(\mathbf{A}_{11}, \ldots, \mathbf{A}_{nn}) \preceq \lambda_{\max}(\mathbf{B}) \left(\max_i \mathbf{A}_{ii}\right) \mathbf{I}_n.$$

By (i), $\mathbf{A} \odot \mathbf{B} \succeq 0$, so its spectral norm equals its largest eigenvalue, which gives the second statement.

(iii) For unit vectors x and y, by the Cauchy–Schwarz inequality applied to the pairs (i, j),

$$|y'(\mathbf{A} \odot \mathbf{B})\, x| = \left|\sum_{i,j} (y_i \mathbf{A}_{ij})\, (\mathbf{B}_{ij} x_j)\right| \leq \left(\sum_i y_i^2 \|\mathbf{A}_{i\cdot}\|^2\right)^{1/2} \left(\sum_j x_j^2 \|\mathbf{B}_{\cdot j}\|^2\right)^{1/2},$$

which is bounded by $(\max_i \|\mathbf{A}_{i\cdot}\|)\, (\max_j \|\mathbf{B}_{\cdot j}\|)$. The spectral norm is the supremum of the left-hand side over unit vectors x and y. □

C Leverage simulation

This appendix provides simulation evidence on the maximum-leverage condition of Assumption *No Influential Observations*, under the DGP described in Section 4.

As noted in Section 3, the ordinary leverage $\mathrm{lev}_t := \tilde{x}_t'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{x}_t$ satisfies $\max_t \mathrm{lev}_t \geq K/T$, so it cannot vanish whenever the aspect ratio K/T does not vanish. However, for the asymptotic normality of the DaC estimator, Assumption *No Influential Observations* requires only that

$$\max_t \sqrt{T}\left|r'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{x}_t\right|$$

vanish (this is the $h = 0$ special case $\max_t |r'(\mathbf{X}'\mathbf{X})^{-1}x_t| = o(T^{-1/2})$ discussed below that assumption). We simulate this quantity with $r = e_1$ under the DGP of Section 4, with $K/T = 0.5$, $\rho = 0.9$, $\sigma_\eta = 0.2$ and $\nu \in \{2.1, 2.5\}$. We also include the case where the innovations of the AR(1) processes generating $\tilde{x}_t$ are Gaussian. Recall that we normalise innovations to have the same standard deviation ($\sigma_\eta = 0.2$) for all choices of ν. We ignore subsampling. Table 5 reports the mean and standard deviation of $\max_t \sqrt{T}\,|e_1'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{x}_t|$ across Monte Carlo replications, for a range of sample sizes. In all three cases the statistic declines steadily as T grows, consistent with Assumption *No Influential Observations*; the relative decline is slower the more heavy-tailed the regressor.

Table 5: $\sqrt{T}\max_t |e_1'(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{x}_t|$ under independent AR(1) regressors

Sample size	Mean	Std. dev.
$\boldsymbol{\nu = 2.1}$		
33	7.553	3.089
100	5.215	1.545
300	3.756	0.813
900	3.006	0.701
$\boldsymbol{\nu = 2.5}$		
33	4.261	1.524
100	2.904	0.723
300	2.089	0.440
900	1.619	0.367
$\boldsymbol{\nu = \infty}$		
33	3.119	0.858
100	2.103	0.359
300	1.389	0.193
900	0.889	0.099

Notes: the entries of $\tilde{x}_t$ follow independent AR(1) processes with $K/T = 0.5$, persistence $\rho = 0.9$, innovations rescaled to coordinate-wise standard deviation 0.2. $\nu = \infty$ denotes Gaussian innovations. Mean and standard deviation across replications of the statistic in the caption. Number of replications: 500.