Dissecting Anomalies in Conditional Asset Pricing
ONLINE APPENDIX

VALENTINA RAPONI PAOLO ZAFFARONI

IESE Business School Imperial College London

This Draft

July 12, 2023

Contents

1 Preliminary notation and main definitions

To facilitate the reading, this section presents the notation and the main definitions that we use throughout the paper.

We use the bold font to denote vectors (in lower-case) and matrices (in upper-cases). The notation $\mathbf{I}_a$ is used to define the identity matrix of dimension $a \times a$, while $\mathbf{1}_a$ denotes an $a \times 1$ vector of ones. The s-th coulumn (or row) of $\mathbf{I}_a$ is denoted by the $a \times 1$ vector $\boldsymbol{\imath}_{s,a}$. Similarly, let $\mathbf{0}_a$ and $\mathbf{0}_{a \times b}$ be the $a \times 1$ vector of zeros and the zero matrix of dimension $a \times b$, respectively. For any full-column-rank matrix $\mathbf{A}$ of dimension $a \times b$, we define the $a \times a$ matrix $\mathbb{M}_A \equiv \mathbf{I}_a - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \equiv \mathbf{I}_a - \mathbb{P}_A$, with $\mathbb{P}_A \equiv \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. We also write $\mathbf{A} > 0$ and $\mathbf{A} \geq 0$ whenever the matrix $\mathbf{A}$ is positive definite and semi-positive definite, respectively. We use $\mathbb{1}_{\{\cdot\}}$ to denote the indicator function, while $\mathrm{E}[\cdot]$ and $\mathrm{E}[\cdot|A]$ identify the the unconditional and conditional (on the event A) expectations, respectively. Finally, $\otimes$, $\mathrm{vec}(\cdot), \mathrm{vec}^{-1}(\cdot)$ denote the Kronecker product, the vec operator, and the inverse vec operator, respectively.[1] The quantity $\mathbf{A}^{(2)} \equiv (\mathbf{A} \odot \mathbf{A})$, where the symbol $\odot$ indicates the Hadamard product. Convergence in distribution and in probability are denoted by $\to_d$ and $\to_p$, respectively. Finally, $\mathcal{N}(\cdot\,,\,\cdot)$ denotes a Normally-distributed random vector.

All the moments throughout the Online Appendix are assumed to hold conditionally on the factors $\mathbf{F}$, even if not written explicitly, and all the limits below hold as $N \to \infty$.

DA FARE TIMEVARYING BETAS SECTION

OA.1 Notation for Anomalies

Dealing with anomalies also involves some specific notation, because such variables are both time- and asset-specific. In fact, assuming that we have K_{z} anomaly variables, it requires dealing with an array of dimension $N \times T - 1 \times K_{\mathrm{z}}$. We emphasize the suffix $T - 1$ to remind that our data sample is of size $T - 1$, due to the presence of the lagged anomaly variables. In particular we set $t = 2, ..., T$ whenever we refer to asset returns, while we use $t = 1..., T - 1$ to index the anomalies.

[1] The inverse vec operator reconstructs a matrix from a column vector. However, it necessarily requires defining the desired number of rows and columns of the resultant matrix. In this paper we will only use this operator to reconstruct matrices of dimension $(T-1)K_{\mathrm{z}} \times (T-1)$. Thus, the inverse vec operator can be unambiguously defined as $vec^{-1}(\mathbf{a}) = (vec(\mathbf{I}_{(T-1)K_{\mathrm{z}}})' \otimes \mathbf{I}_{T-1})(\mathbf{I}_{(T-1)K_{\mathrm{z}}} \otimes \mathbf{a})$ for every vector $\mathbf{a}$ of length $(T-1)^2 K_{\mathrm{z}} \times 1$.

We define the overall $N \times K_{\mathrm{z}}(T-1)$ matrix of anomalies as

$$\mathbf{Z} \equiv (\mathbf{z}_1, ..., \mathbf{z}_N)'$$

where we use the (lower case) notation $\mathbf{z}_i$ to define the $K_{\mathrm{z}}(T-1) \times 1$ vector

$$\mathbf{z}_i \equiv \left(z_{i,1}^{(1)}, \cdots, z_{i,T-1}^{(1)}, \cdots, z_{i,1}^{(K_{\mathrm{z}})}, \cdots, z_{i,T-1}^{(K_{\mathrm{z}})} \right)'.$$

The $N \times K_{\mathrm{z}}$ matrix of anomalies at time $t-1$ is defined as $\mathbf{Z}_{t-1} = (\mathbf{z}_{1,t-1}, \cdots, \mathbf{z}_{N,t-1})'$, while the $T \times K_{\mathrm{z}}$ matrix of anomalies specific for the i-th asset is $\mathbf{Z}_i = (\mathbf{z}_{i,1}, \cdots, \mathbf{z}_{i,T-1})'$, where $\mathbf{z}_{i,t-1}$ denotes the $K_{\mathrm{z}} \times 1$ vector $\mathbf{z}_{i,t-1} = \left(z_{i,t-1}^{(1)}, \cdots, z_{i,t-1}^{(K_{\mathrm{z}})} \right)'$. Taking the time-series average of the anomalies leads to the $N \times K_{\mathrm{z}}$ matrix of sample averages $\bar{\mathbf{Z}} = \frac{1}{T-1} \sum_{t=1}^{T-1} \mathbf{Z}_t$.

Finally, we define the following $K_{\mathrm{z}}(T-1) \times K_{\mathrm{z}}$ matrix of constants

$$\mathbb{J} = \frac{1}{T-1} \begin{bmatrix} \mathbf{1}_{T-1} & \mathbf{0}_{T-1} & \dots & \mathbf{0}_{T-1} \\ \mathbf{0}_{T-1} & \mathbf{1}_{T-1} & \dots & \mathbf{0}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{T-1} & \mathbf{0}_{T-1} & \dots & \mathbf{1}_{T-1} \end{bmatrix} = \left(\mathbf{I}_{K_{\mathrm{z}}} \otimes \frac{\mathbf{1}_{T-1}}{(T-1)} \right) = \frac{1}{T-1} \sum_{s=1}^{T-1} \mathbb{J}_s \qquad \text{(OA.1)}$$

with

$$\mathbb{J}_s = \begin{bmatrix} \boldsymbol{\iota}_{s,T-1} & \mathbf{0}_{T-1} & \dots & \mathbf{0}_{T-1} \\ \mathbf{0}_{T-1} & \boldsymbol{\iota}_{s,T-1} & \dots & \mathbf{0}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{T-1} & \mathbf{0}_{T-1} & \dots & \boldsymbol{\iota}_{s,T-1} \end{bmatrix} = (\mathbf{I}_{K_{\mathrm{z}}} \otimes \boldsymbol{\iota}_{s,T-1}) \text{ for } 1 \leq s \leq T-1 \qquad \text{(OA.2)}$$

The matrix $\mathbb{J}$ applied to $\mathbf{Z}$ generates the time averages of the K_{z} anomaly variables, i.e. $\mathbf{Z}\mathbb{J} = \bar{\mathbf{Z}}$, while $\mathbb{J}_s$ allows to select the s-th time-series element of each of the N assets and for all the K_{z} anomalies. Notice also that $\boldsymbol{\iota}_{i,N}' \mathbf{Z} \mathbb{J}_s = \mathbf{z}_{i,s}'$, for every $1 \leq i \leq N$ and $1 \leq s \leq T-1$.

OA.2 Assumptions

In this section, we recall all the assumptions required for the validity of our large-N asymptotic theory, together with some comments. All the moments below are assumed to hold conditionally on the factors $\mathbf{F}$, even if not written explicitly, and all the limits below hold as $N \to \infty$.

It is useful to recall the $N \times K_{\mathrm{z}}(T-1)$ matrix of anomalies $\mathbf{Z} \equiv (\mathbf{z}_1, ..., \mathbf{z}_N)'$, where $\mathbf{z}_i$ defines the $K_{\mathrm{z}}(T-1) \times 1$ vector $\mathbf{z}_i \equiv \left(z_{i,1}^{(1)}, \cdots, z_{i,T-1}^{(1)}, \cdots, z_{i,1}^{(K_{\mathrm{z}})}, \cdots, z_{i,T-1}^{(K_{\mathrm{z}})} \right)'$. The $N \times K_{\mathrm{z}}$ matrix of anomalies at time $t-1$ is defined as $\mathbf{Z}_{t-1} = (\mathbf{z}_{1,t-1}, \cdots, \mathbf{z}_{N,t-1})'$, while the $(T-1) \times K_{\mathrm{z}}$ matrix of anomalies specific for the i-th asset is $\mathbf{Z}_i = (\mathbf{z}_{i,1}, \cdots, \mathbf{z}_{i,T-1})'$, setting $\mathbf{z}_{i,t-1} = \left(z_{i,t-1}^{(1)}, \cdots, z_{i,t-1}^{(K_{\mathrm{z}})} \right)'$.

Assumption OA.1 (*smoothness of the premia parameters*)**.** *The following hold:*

$$\mathbf{P}'\gamma_0 = \mathbf{0}_{K_\mathrm{f}}, \qquad \mathbf{P}'\breve{\boldsymbol{\delta}}_\mathrm{f} = \mathbf{0}_{K_\mathrm{f}\times K_\mathrm{f}}, \qquad \text{and} \quad \mathbf{P}'\boldsymbol{\Delta}_z = \mathbf{0}_{K_\mathrm{f}\times N},$$

setting the $(T-1)\times K_\mathrm{f}$ *matrix* $\breve{\boldsymbol{\delta}}_\mathrm{f} = (\breve{\boldsymbol{\delta}}_{f,1}, \cdots, \breve{\boldsymbol{\delta}}_{\mathrm{f},t-1})'$, with $\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1} \equiv \boldsymbol{\delta}_{\mathrm{f},t-1} - \mathbf{f}_t = \boldsymbol{\gamma}_{\mathrm{f},t-1} - E(\mathbf{f}_t|I_{t-1}, \boldsymbol{\Pi})$*, and the* $(T-1)\times N$ *matrix*

$$\boldsymbol{\Delta}_z \equiv \begin{bmatrix} \boldsymbol{\gamma}'_{z,1} - \boldsymbol{\gamma}'_\mathrm{z} & \mathbf{0}'_{K_\mathrm{z}} & \cdots & \mathbf{0}'_{K_\mathrm{z}} \\ \mathbf{0}'_{K_\mathrm{z}} & \boldsymbol{\gamma}'_{z,2} - \boldsymbol{\gamma}'_\mathrm{z} & \cdots & \mathbf{0}'_{K_\mathrm{z}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}'_{K_\mathrm{z}} & \mathbf{0}'_{K_\mathrm{z}} & \cdots & \boldsymbol{\gamma}'_{\mathrm{z},t-1} - \boldsymbol{\gamma}'_\mathrm{z} \end{bmatrix} \begin{bmatrix} \mathbf{Z}'_1 \\ \mathbf{Z}'_1 \\ \vdots \\ \mathbf{Z}'_{T-1} \end{bmatrix},$$

for some constant $K_\mathrm{z}\times 1$ *vector* $\boldsymbol{\gamma}_\mathrm{z}$ *satisfying* $N^{-1}\sum_{i=1}^N (\mathbf{Z}'_i\mathbf{Z}_i)^{-1}\mathbf{Z}'_i\mathbf{R}_i \to_p \boldsymbol{\gamma}_\mathrm{z}$.

Remark OA.1. *When the risk factors are traded,* $\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1} = -\boldsymbol{\gamma}_0\mathbf{1}'_{K_\mathrm{f}}$ *for every* t*, and Assumption OA.1 only concerns the zero-beta rate. In the special case of constant premia parameters, when both the test assets and the risk factors are expressed as excess returns, and assuming that a risk-free asset is also traded, then Assumption OA.1 is always satisfied.*

One can avoid imposing the smoothness conditions of Assumption OA.1, and thus allowing for time-series dependence between the time-varying premia and the risk factors, but at the cost of more complicate expressions. In particular, (7) *can be expressed as a panel data model with interactive-fixed effects:*

$$\mathbf{R}_t = \boldsymbol{\alpha} + \mathbf{Z}_{t-1}\bar{\boldsymbol{\gamma}}_\mathrm{z} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t,$$

where the error term satisfies $\mathbf{u}_t = \boldsymbol{\xi}_t + \boldsymbol{\Delta}\mathbf{g}_t$ *for an asset-specific error* $\boldsymbol{\xi}_t$ *and a vector of zero-mean latent factors* $\mathbf{g}_t$ *possibly correlated with the observed risk factors* $\mathbf{f}_t$*, with loadings* $\boldsymbol{\Delta}$*, and where* $\bar{\boldsymbol{\gamma}}_\mathrm{z} = T^{-1}\sum_{t=1}^T \boldsymbol{\gamma}_{t-1,z}$*. Assumption OA.1 implies orthogonality between* $\mathbf{f}_t$ *and* $\mathbf{u}_t$*, resurrecting the OLS estimator* $\hat{\mathbf{B}}$*. However, an alternative estimator for* $\mathbf{B}$ *exists that avoids Assumption OA.1 but leads to a more involved analysis of the CSR in the second pass. Details are available upon request.*

Assumption OA.2 (*risk factors and anomalies*)**.** Set $\tilde{\mathbf{Z}}_i \equiv \mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{Z}_i$, and $\mathbf{D} \equiv (\mathbf{1}_{T-1}, \mathbf{F})$. Then, for every T, the $(T-1)\times(K+1)$ matrix $\tilde{\mathbf{D}}_i = (\mathbf{D}, \tilde{\mathbf{Z}}_i)$ satisfies

$$\tilde{\mathbf{D}}'_i\tilde{\mathbf{D}}_i > 0 \quad \text{for every } i = 1, ..., N.$$

Assumption OA.3 (*loadings*)**.**

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i \to \boldsymbol{\mu}_\beta \qquad \text{and} \qquad \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i\boldsymbol{\beta}_i' \to \boldsymbol{\Sigma}_\beta,$$

such that the matrix

$$\boldsymbol{\Sigma}_{\mathrm{X}} \equiv \begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta \end{bmatrix} > 0.$$

Remark OA.2. *Assumption OA.3 states that the limiting cross-sectional averages of the betas, and of the squared betas, exist. The second part of Assumption OA.3 rules out the possibility of spurious and weak factors and situations in which at least one of the elements of $\boldsymbol{\beta}_i$ is cross-sectionally constant. It implies that $\mathbf{X} = (\mathbf{1}_N, \mathbf{B})$ has full (column) rank for N sufficiently large. To simplify the exposition, we assume that the $\boldsymbol{\beta}_i$ are non-random.*[2]

Assumption OA.4 (*asset-specific components*)**.** The $N \times 1$ vector of error terms $\boldsymbol{\epsilon}_t$ is independently and identically distributed (i.i.d.) over time with

$$\mathrm{E}[\boldsymbol{\epsilon}_t] = \mathbf{0}_N \tag{OA.3}$$

and with the $N \times N$ variance-covariance matrix satisfying

$$\mathrm{Var}\,[\boldsymbol{\epsilon}_t] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{pmatrix} \equiv \boldsymbol{\Sigma} > 0, \tag{OA.4}$$

where σ_{ij} denotes the (i,j)-th element of $\boldsymbol{\Sigma}$, for every $i,j = 1,\ldots,N$, and with $\sigma_i^2 \equiv \sigma_{ii}$.

Remark OA.3. *The i.i.d. assumption over time is common to many studies, including* **?** *and Raponi, Robotti, and Zaffaroni (2020). Nonetheless, in principle, our large-N asymptotic theory allows the ϵ_{it} to be arbitrarily correlated over time, at the cost of more complicated expressions and derivations for the limiting distributions of the estimators. Condition (OA.4) is not imposing any specific structure on the elements of $\boldsymbol{\Sigma}$ beyond non-singularity. In particular, we are not assuming that the returns' innovations are uncorrelated across assets or exhibit the same variance. However, our large-N asymptotic theory needs to discipline the degree (as N increases) of cross-correlation among the $\boldsymbol{\epsilon}_t$, as indicated in Assumption OA.5 below.*

[2] See **?** and Raponi, Robotti, and Zaffaroni (2020) for the analysis of asset pricing models with random betas.

Assumption OA.5 (*cross-sectional moments of asset-specific components*). (i)

$$\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2-\sigma^2\right)=o\left(\frac{1}{\sqrt{N}}\right), \tag{OA.5}$$

for some $0<\sigma^2<\infty$.

(ii)

$$\sum_{i,j=1}^{N}\mid\sigma_{ij}\mid\mathbb{1}_{\{i\neq j\}}=o\left(N\right). \tag{OA.6}$$

(iii)

$$\frac{1}{N}\sum_{i=1}^{N}\mu_{4i}\rightarrow\mu_4, \tag{OA.7}$$

for some $0<\mu_4<\infty$, where $\mu_{4i}\equiv E[\epsilon_{i,t}^4]$.

(iv)

$$\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4\rightarrow\sigma_4, \tag{OA.8}$$

for some $0<\sigma_4<\infty$.

(v)

$$\sup_i\mu_{4i}\leq C<\infty, \tag{OA.9}$$

for a generic constant C.

(vi)

$$E[\epsilon_{i,t}^3]=0. \tag{OA.10}$$

(vii)

$$\frac{1}{N}\sum_{i=1}^{N}\kappa_{4,iiii}\rightarrow\kappa_4, \tag{OA.11}$$

for some $0\leq|\kappa_4|<\infty$, where $\kappa_{4,iiii}\equiv\kappa_4[\epsilon_{it},\epsilon_{it},\epsilon_{it},\epsilon_{it}]$ denotes the fourth-order cumulant of the asset-specific component $\{\epsilon_{i,t},\epsilon_{i,t},\epsilon_{i,t},\epsilon_{i,t}\}$.

(viii) For every $3\leq h\leq 8$, all the following mixed cumulants of order h satisfy

$$\sup_{i_1}\sum_{i_2,\ldots,i_h=1}^{N}|\kappa_{h,i_1i_2\ldots i_h}|=o\left(N\right), \tag{OA.12}$$

for at least one i_j $(2\leq j\leq h)$ different from i_1, where $\kappa_{h,i_1,i_2\ldots i_h}$ is the mixed cumulant in the $\{\epsilon_{i_1,s},\epsilon_{i_2,s},\cdots,\epsilon_{i_h,s}\}$ of order h.

Remark OA.4. *Assumption OA.5 describes the cross-sectional behavior of the asset-specific components. Specifically, Assumption OA.5(i) limits the cross-sectional heterogeneity of the returns' conditional variance, while Assumption OA.5(ii) sets the maximum degree of (conditional) cross-correlation among asset returns allowed by our theory. These assumptions are not very restrictive and allow for several forms of strong cross-sectional dependence among the* ϵ_{it}*'s, such as, for example, a factor structure of the following form:*

$$\epsilon_{it} = \lambda_i u_t + \eta_{i,t}, \tag{OA.13}$$

where $u_t \sim i.i.d.(0,1)$ *and* $\eta_{i,t} \sim i.i.d.(0, \sigma_\eta^2)$ *over time and across units, and where* u_t *and* $\eta_{i,s}$ *are mutually independent for every* i, s *and* t*. The coefficient* λ_i *is such that* $\sum_{i=1}^N |\lambda_i| = O(N^\delta)$, $0 \leq \delta < 1/4$, *and* $\lambda_1 + \cdots + \lambda_q \sim CN^\delta$, *for some fixed* $q < N$ *and some constant* C. *Although Assumptions OA.5(i) and OA.5(ii) are easily satisfied in the special case of* $\sigma^2 = \sigma_\eta^2$, *notice that the maximum eigenvalue of* $\mathbf{\Sigma}$ *is now unbounded as* $N \to \infty$.[3] *This is in contrast with the standard Asset Pricing Theory (APT), where instead boundedness of the maximum eigenvalue is the most common assumption (see, e.g., the generalization of the APT by* **?***). Therefore, our assumptions are milder than the ones postulated by the APT and thus more likely to be verified by the data.*[4] *Other special cases nested in Assumption OA.5 (for which the cross-covariances* σ_{ij} *are non-zero) are network and spatial measures of cross-dependence and a suitably modified version of the block-dependence structure of* **?**.[5]

SECONDO ME MEGLIO RIPETERLA, NON HA SENSO RIMANDARE A RRZ. ANZI METTIAMO UNA VERSIONE RIDOTTA DI TABELLA RRZ? Qui c'e' tutta la parte di cross-sectional dependence spiegata con la tabella della Monte Carlo simulaton, che pero' e' IDENTICA a quella di Raponi, Robotti, and Zaffaroni (2020). dobbiamo ripeterla o basta citarla? See my footnote in red above

Assumption OA.5(iii) simply posits the existence of the limit of the conditional fourth moment,

[3]The maximum eigenvalue of $\mathbf{\Sigma}$ is given by $\sup_{\mathbf{c}s.t.\|\mathbf{c}\|=1} \mathbf{c}'\mathbf{\Sigma}\mathbf{c}$.

[4]Specifically, Assumption OA.5 allows for the maximum eigenvalue of $\mathbf{\Sigma}$ to diverge at rate $o\left(\sqrt{N}\right)$ (see Raponi, Robotti, and Zaffaroni (2020), Proposition 3). This implies that the row-column norm of $\mathbf{\Sigma}$, namely $\sup_{1\leq i\leq N} \sum_{j=1}^N |\sigma_{ij}|$, can now diverge as $N \to \infty$. **?** allow for an even faster rate of divergence, equal to $o(N)$, but in their setting both T and N tend to infinity jointly.

[5]Assumption BD.2 of **?** on block sizes and block numbers requires that the largest block size shrinks with N and that there are not too many large blocks; that is, the partition in independent blocks is sufficiently fine-grained asymptotically. They show formally that such block-dependence structure is compatible with the unboundedness of the maximum eigenvalue of $\mathbf{\Sigma}$.

averaged across assets. In Assumption OA.5(iv), the magnitude of σ_4 reflects the degree of cross-sectional heterogeneity of the conditional variance of the asset returns. Assumption OA.5(v) is a bounded fourth-moment condition, uniform across assets, which implies that $\sup_i \sigma_i^2 \le C < \infty$. Assumption OA.5(vi) is a convenient symmetry assumption, but it is not strictly necessary for our results. Indeed, this assumption could be relaxed, even though the derivation of the asymptotic distribution would be more cumbersome, due to the presence of several extra terms involving the third moment of the disturbance. Assumptions OA.5(vii)-(viii) allow for non-Gaussianity of the asset returns whenever $\kappa_4 > 0$. For example, Assumptions OA.5(vii)-(viii) are satisfied when the marginal distribution of asset returns is a Student-t with more than 4 degrees of freedom. However, when estimating the asymptotic covariance matrix of our bias-adjusted estimators, one needs to set $\kappa_4 = 0$ merely for identification purposes. That said, higher-order cumulants are not constrained to be zero, implying that, even when $\kappa_4 = 0$, the distribution is not necessarily equivalent to the Gaussian one.

Assumption OA.6 (*CLT of asset-specific component*). (i)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i \to_d \mathcal{N}\left(\mathbf{0}_{T-1}, \sigma^2\mathbf{I}_{T-1}\right). \quad \text{(OA.14)}$$

(ii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\text{vec}\left(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}\right) \to_d \mathcal{N}\left(\mathbf{0}_{(T-1)^2}, \mathbf{U}_\epsilon\right), \quad \text{(OA.15)}$$

where $\mathbf{U}_\epsilon \equiv \lim \frac{1}{N}\sum_{i,j=1}^{N} E\Big[\text{vec}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1})\text{vec}(\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j' - \sigma_j^2\mathbf{I}_{T-1})'\Big]$.

(iii) For any $T \times 1$ vector $\mathbf{c}$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{c}' \otimes \begin{pmatrix}1\\ \boldsymbol{\beta}_i\end{pmatrix}\right)\boldsymbol{\epsilon}_i \to_d \mathcal{N}\left(\mathbf{0}_{K_f+1}, (\mathbf{c}'\mathbf{c})\sigma^2\boldsymbol{\Sigma}_X\right). \quad \text{(OA.16)}$$

Remark OA.5. *From* (OA.16), *it follows that* $N^{-\frac{1}{2}}\sum_{i=1}^{N}(\mathbf{c}'\otimes\boldsymbol{\beta}_i)\,\boldsymbol{\epsilon}_i \to_d \mathcal{N}(\mathbf{0}_{K_f}, (\mathbf{c}'\mathbf{c})\sigma^2\boldsymbol{\Sigma}_\beta)$. *Primitive conditions for Assumption OA.6 can be derived, but at the cost of raising the level of complexity of our proofs.*[6]

[6] For instance, when (OA.13) holds and all the above assumptions are satisfied, then (OA.14) follows by Theorem 2 of **?** when η_{it} satisfy their martingale difference assumptions (see their Assumptions 1 and 2.) This result extends easily to (OA.15)–(OA.16) under suitable additional assumptions. Details are available upon request.

Remark OA.6. *The expression for* $\mathbf{U}_\epsilon$ *in* (OA.15) *can be derived in closed form. In particular, Raponi, Robotti, and Zaffaroni (2020) established that the* $T^2 \times T^2$ *matrix* $\mathbf{U}_\epsilon$ *has the following form*

$$\mathbf{U}_\epsilon = \left[\begin{array}{c:c:c:c:c} U_{11} & \cdots & U_{1t} & \cdots & U_{1T} \\ \hdashline \vdots & \ddots & \vdots & \vdots & \vdots \\ \hdashline U_{t1} & \cdots & U_{tt} & \cdots & U_{tT} \\ \hdashline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hdashline U_{T1} & \cdots & U_{Tt} & \cdots & U_{TT} \end{array}\right].$$

Each block of U_ϵ *is a* $T \times T$ *matrix. The blocks along the main diagonal, denoted by* U_{tt}*,* $t = 1, 2, \ldots, T$*, are themselves diagonal matrixes with* $(\kappa_4 + 2\sigma_4)$ *in the* (t,t)*-th position and* σ_4 *in the* (s,s) *position for every* $s \neq t$*. The blocks outside the main diagonal, denoted by* U_{ts}*,* $s, t = 1, 2, \ldots, T$ *with* $s \neq t$*, are all made of zeros except for the* (s,t)*-th position that contains* σ_4*; that is,*

$$U_{tt} = \underset{t\text{-th row}}{\rightarrow} \begin{bmatrix} \sigma_4 & \cdots & 0 & \overset{\substack{\downarrow \\ t\text{-th column}}}{\cdots} & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sigma_4 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & \sigma_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \sigma_4 \end{bmatrix}, \quad U_{ts} = \underset{s\text{-th row}}{\rightarrow} \begin{bmatrix} 0 & \cdots & 0 & \overset{\substack{\downarrow \\ t\text{-th column}}}{\cdots} & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_4 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix}.$$

Assumption OA.7 (*moments and CLT of anomalies*)**.** *Define the* $K_{\mathrm{z}}(T-1)^2 \times 1$ *vector* $\mathbf{u}_i \equiv \boldsymbol{\epsilon}_i \otimes \mathbf{z}_i$*.*

(i)

$$\frac{\mathbf{Z}'\mathbf{1}_N}{N} \rightarrow_p (\boldsymbol{\mu}_{\mathrm{z}} \otimes \mathbf{1}_{T-1}) \equiv \boldsymbol{\mu}_{\mathrm{z,T-1}}$$

for a finite $K_{\mathrm{z}} \times 1$ vector $\boldsymbol{\mu}_{\mathrm{z}} = \left(\mu_{\mathrm{z}}^{(1)}, \ldots, \mu_{\mathrm{z}}^{(K_{\mathrm{z}})}\right)' \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\mu}_{\mathrm{z}_i}$, setting $\boldsymbol{\mu}_{\mathrm{z}_i} \equiv \mathrm{E}[\mathbf{z}_{i,s}]$.

(ii)

$$\frac{\mathbf{Z}'\mathbf{Z}}{N} \to_p \mathbf{\Sigma}_{\mathrm{Z}},$$

for a finite $K_{\mathrm{z}}(T-1)\times K_{\mathrm{z}}(T-1)$ matrix $\mathbf{\Sigma}_{\mathrm{Z}}$, such that $\mathbb{J}'\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J} > 0$ and $\mathbb{J}'_{t-1}\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1} > 0$, for every $2 \leq t \leq T$.

(iii)

$$\frac{\mathbf{Z}'\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{\mathrm{ZB}},$$

for a finite $K_{\mathrm{z}}(T-1) \times K_{\mathrm{f}}$ matrix $\mathbf{\Sigma}_{\mathrm{ZB}}$.

(iv) *Setting* $\boldsymbol{\mu}_{\mathrm{u}_i} \equiv \mathrm{E}[\mathbf{u}_i]$,

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\mu}_{\mathrm{u}_i} = o\left(\frac{1}{\sqrt{N}}\right).$$

(v) *Setting* $\mathbf{\Sigma}_{\mathrm{u},ij} \equiv \mathrm{Cov}[\mathbf{u}_i, \mathbf{u}_j]$, *for* $i, j = 1, ..., N$,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{\mathrm{u},ii} \to \mathbf{\Sigma}_{\mathrm{U}} \equiv (\sigma^2\mathbf{I}_{T-1}\otimes\mathbf{\Sigma}_{\mathrm{Z}}) \text{ and } \sum_{i,j=1}^{N}\mathbf{\Sigma}_{\mathrm{u},ij}\mathbb{1}_{i\neq j} = o(N).$$

(vi) For any $i, j = 1, ..., N$,

$$\mathrm{Cov}\left[\mathbf{z}_{i,t}, \boldsymbol{\epsilon}'_j \otimes \boldsymbol{\epsilon}'_j\right] = \mathbf{0}_{K_{\mathrm{z}}\times(T-1)^2}, \quad \mathrm{Cov}\left[\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}'_j \otimes (\boldsymbol{u}_j - \mathrm{E}[\boldsymbol{u}_j])'\right] = \mathbf{0}_{T-1\times K_{\mathrm{z}}(T-1)^3}.$$

(vii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{u}_i - \boldsymbol{\mu}_{u_i}) \to_d \mathcal{N}\left(\mathbf{0}_{K_{\mathrm{z}}(T-1)^2}, \mathbf{\Sigma}_{\mathrm{U}}\right).$$

(viii) *Setting* $\mathbf{\Sigma}_{\mathrm{u}\epsilon,ij} \equiv \mathrm{Cov}\left[\boldsymbol{\epsilon}_i \otimes \boldsymbol{\epsilon}_i, \mathbf{u}'_j\right]$,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{\mathrm{u}\epsilon,ii} \quad\to\quad \mathbf{\Sigma}_{\mathrm{u}\epsilon} = \mathbf{0}_{(T-1)^2\times K_{\mathrm{z}}(T-1)^2} \text{ and } \frac{1}{N}\sum_{i,j=1}^{N}\mathbf{\Sigma}_{\mathrm{u}\epsilon,ij} \to \mathbf{0}_{(T-1)^2\times K_{\mathrm{z}}(T-1)^2}.$$

(ix)

$$\frac{1}{N^2}\sum_{i,j=1}^{N}\mathrm{Cov}\left[\boldsymbol{u}_i \otimes \boldsymbol{u}_i, \boldsymbol{u}'_j \otimes \boldsymbol{u}'_j\right] \to \mathbf{0}_{K_{\mathrm{z}}^2(T-1)^4\times K_{\mathrm{z}}^2(T-1)^4}.$$

(x) *Let* $\mathbb{P}_{\tilde{Z}_i} = \tilde{\mathbf{Z}}_i(\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i)^{-1}\tilde{\mathbf{Z}}_i'$, *with its generic* (t,s) *element denoted by* $\mathbb{p}_{i,ts}$, *for* $t,s = 1,...,T-1$, *where* $\tilde{\mathbf{Z}}_i = \mathbb{M}_{1_{T-1}}\mathbf{Z}_i$. *Then, for every* $1 \leq t+1, s+1, v_a, u_a \leq (T-1)$, *with* $a = 1,...,4$, *the following hold:*

$$(x.1) \qquad \frac{1}{N}\sum_{i=1}^{N}\mathbb{P}_{\tilde{Z}_i} \to_p \mathbb{P}_{\tilde{Z}}, \text{ for a finite matrix } \mathbb{P}_{\tilde{Z}},$$

$$(x.2) \qquad \frac{1}{N}\sum_{i=1}^{N}(\mathbb{P}_{\tilde{Z}_i} \odot \mathbb{P}_{\tilde{Z}_i}) \to_p \mathbb{P}_{\tilde{Z}}^{(2)}, \text{ for a finite matrix } \mathbb{P}_{\tilde{Z}}^{(2)},$$

$$(x.3) \qquad \frac{1}{N}\sum_{i=1}^{N}\mathbb{P}_{\tilde{Z}_i}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}) = \mathbb{P}_{\tilde{Z}}\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}) + o_p\left(\frac{1}{\sqrt{N}}\right),$$

$$(x.4) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa_4\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\mathbb{p}_{j,s-1v_a}, \prod_{a=1}^{4}\epsilon_{i,u_a+1}, \prod_{a=1}^{4}\epsilon_{j,v_a+1}\right] = o(1),$$

$$(x.5) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa_3\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\mathbb{p}_{j,s-1v_a}, \prod_{a=1}^{4}\epsilon_{i,u_a+1}\right] = o(1),$$

$$(x.6) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa_3\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\epsilon_{i,u_a+1}, \prod_{a=1}^{4}\epsilon_{j,v_a+1}\right] = o(1),$$

$$(x.7) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\text{Cov}\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\epsilon_{j,v_a+1}\right] = o(1),$$

$$(x.8) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\text{Cov}\left[\mathbb{p}_{j,su_1}\mathbb{p}_{i,tv_1}, \epsilon_{i,t+1}\epsilon_{j,s+1}\epsilon_{iu_1+1}\epsilon_{jv_1+1}\right] = o(1),$$

$$(x.9) \qquad \frac{1}{N}\sum_{i=1}^{N}\text{Cov}\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\epsilon_{i,v_a+1}\right] = o(1).$$

where $\kappa_3[\cdot,\cdot,\cdot]$ and $\kappa_4[\cdot,\cdot,\cdot,\cdot]$ denote the mixed cumulants of order 3 and 4, respectively.

(xi) For every $3 \leq h \leq 8$, all the following mixed cumulants of order h satisfy

$$\sup_{i_1}\sum_{i_2,\ldots,i_h=1}^{N}|\kappa^{\mathbb{p}}_{h,i_1i_2\ldots i_h}| = o(N), \qquad \text{(OA.17)}$$

for at least one i_j $(2 \leq j \leq h)$ different from i_1, where $\kappa^{\mathbb{p}}_{h,i_1,i_2\ldots i_h}$ is the mixed cumulant in the $\{\mathbb{p}_{i_1,t_1-1u_1}, \mathbb{p}_{i_2,t-_21u_2}, \cdots, \mathbb{p}_{i_h,t_h-1u_h},\}$ of order h, for every $2 \leq t_1, \cdots, t_h, u_1, \cdots, u_h \leq T$.

Remark OA.7. *By Assumption OA.7 (iv),*

$$\frac{\mathbf{Z}'\boldsymbol{\epsilon}'}{N} \to_p \boldsymbol{\Sigma}_{Z\epsilon} = \mathbf{0}_{K_z(T-1)\times(T-1)}.$$

Moreover, Assumption OA.7 (vii) implies

$$\sqrt{N}\text{vec}\left(\frac{\mathbf{Z}'\boldsymbol{\epsilon}'}{N}\right) = \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_i \otimes \mathbf{z}_i)\right) \to_d N\left(\mathbf{0}_{K_{\text{z}}(T-1)^2}, \boldsymbol{\Sigma}_{\text{U}}\right).$$

Remark OA.8. *Notice that, using our assumptions,* $\frac{1}{N}\bar{\mathbf{Z}}'\mathbf{1}_N = \frac{1}{N}\mathbb{J}'\mathbf{Z}'\mathbf{1}_N \to_p \boldsymbol{\mu}_{\text{z}}$, *and also* $\frac{1}{N}\mathbb{J}_{t-1}'\mathbf{Z}'\mathbf{1}_N \to_p \boldsymbol{\mu}_{\text{z}}$. *Clearly the two estimators have the same limiting behavior, with the former being be a more efficient estimator for* $\boldsymbol{\mu}_{\text{z}}$.

Remark OA.9. *Assumption OA.7 imposes some basic regularity conditions on the behavior of the matrix* $\mathbf{Z}$, *and controls for the degree of cross-sectional dependence between anomalies and excess returns, which slowly dissipates as* N *increases.*

To get a sense of the degree of cross-sectional dependence allowed by our theory, we use a Monte Carlo experiment, where we consider the same data generating process described in the simulation study in Section OA.7, with the error term following the process in (OA.87). *We consider the simplest case of one anomaly* ($K_{\text{z}} = 1$) *and report the sample cross-sectional correlation (both in levels and in absolute terms) between returns' innovation* (ϵ_{it}) *and anomaly' innovation* (η_{it}), *averaged across the time-series dimension (either* $T = 36$ *and* $T = 72$), *for different values of the parameter* δ *in* (OA.87) - *which controls the strength of the cross-sectional correlation between shocks and anomalies. Specifically, we consider the following two measures of (average) sample correlation between anomalies and asset returns:*

$$\rho(\delta, N) \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sum_{i=1}^{N}\epsilon_{it}\eta_{it}}{\left(\sum_{i=1}^{N}\epsilon_{it}^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{N}\eta_{it}^2\right)^{\frac{1}{2}}}\right) \quad \textit{and} \quad \tau(\delta, N) \equiv \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\sum_{i=1}^{N}\epsilon_{it}\eta_{it}}{\left(\sum_{i=1}^{N}\epsilon_{it}^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{N}\eta_{it}^2\right)^{\frac{1}{2}}}\right|, \tag{OA.18}$$

and report the results in Table I, for $\delta = \{0, 0.10, 0.25, 0.50, 0.75, 1.00\}$, *and for* $T = 36$ *(panel A) and* $T = 72$ *(panel B). Remember that, by (OA.87), the degree of cross-correlation is* inversely *related to* δ, *where our theory requires* $\delta > 0$ *for consistency of the premia estimators, and* $\delta \geq 0.5$ *for their asymptotic normality. Table I confirms that our setup admits a sizable cross-sectional dependence between anomalies and returns. For example, when* $\delta = 0.50$, *then both* ρ *and* τ *are 0.67 when N=100 and 0.37 when N=500, which represents a significant degree of correlation even when* N *is sufficiently large. As expected, the strength of the correlation vanishes when both* N *and* δ *increases.*

Table I: Conditional average cross-correlations between anomalies and asset returns.

Panel A: $T = 36$												
	$\rho(\delta, N)$						$\tau(\delta, N)$					
δ	0.00	0.10	0.25	0.50	0.75	1 .0	0.00	0.10	0.25	0.50	0.75	1.00
N												
10	0.993	0.990	0.980	0.941	0.835	0.631	0.993	0.990	0.980	0.941	0.835	0.642
100	0.993	0.984	0.942	0.669	0.279	0.097	0.993	0.984	0.942	0.669	0.279	0.121
500	0.993	0.978	0.883	0.369	0.083	0.017	0.993	0.978	0.883	0.369	0.085	0.041
1000	0.993	0.976	0.848	0.282	0.060	0.019	0.993	0.976	0.848	0.282	0.060	0.028

Panel B: $T = 72$												
	$\rho(\delta, N)$						$\tau(\delta, N)$					
δ	0.00	0.10	0.25	0.50	0.75	1.00	0.00	0.10	0.25	0.50	0.75	1.00
N												
10	0.993	0.989	0.979	0.939	0.838	0.656	0.993	0.989	0.979	0.939	0.838	0.656
100	0.993	0.984	0.942	0.663	0.266	0.082	0.993	0.984	0.942	0.663	0.266	0.102
500	0.993	0.979	0.885	0.372	0.083	0.016	0.993	0.979	0.885	0.372	0.084	0.036
1000	0.994	0.976	0.849	0.273	0.046	0.007	0.994	0.976	0.849	0.273	0.052	0.029

Remark OA.10. *Primitive conditions for Assumption OA.7 can be readily obtained. For instance, one can assume that the* $K_z \times 1$ *vector of anomalies* ($\mathbf{z}_{it}$) *follow a linear process, such as:*

$$\mathbf{z}_{it} = \boldsymbol{\mu}_{iz} + \sum_{k=0}^{\infty} \boldsymbol{\Delta}_{ik} \boldsymbol{\eta}_{i,t-k}, \tag{OA.19}$$

with innovations $\boldsymbol{\eta}_{i,t} \sim (\mathbf{0}_{K_z}, \boldsymbol{\Sigma}_\eta)$ *i.i.d. across time, and such that* $N^{-1}\sum_{i=1}^{N}(\sum_{k=0}^{\infty} \parallel \boldsymbol{\Delta}_{ik} \parallel) \leq C < \infty, \boldsymbol{\Delta}_{i0} = \mathbf{I}_{K_z}$, $N^{-1}\sum_{i=1}^{N} \parallel \boldsymbol{\mu}_{iz} \parallel \leq C < \infty$ *and* $E(\boldsymbol{\eta}_{i,t}\boldsymbol{\epsilon}_t') \neq \mathbf{0}_{K_z \times N}$ *for every* $1 \leq i \leq N$.

The specification in (OA.19) *represents a common assumption in time-series econometrics, in which case the matrices* $\boldsymbol{\Sigma}_{\text{ZZ}}$ *and* $\boldsymbol{\Sigma}_{\text{ZB}}$ *will be constant, with their* (a, b) *entries being a general function of* $(a - b)$.[7] *Moreover, one can also easily allow for cyclical and trending behaviors (either deterministic or stochastic) for the* $\mathbf{z}_{it}$ *in (OA.19). This could be relevant, for instance, when considering firms' characteristics such as value and size, which typically display significant time-variation. In this case, the limit of the cross-sectional averages involving the anomaly variables might not be constant across time, a situation that our methodology can handle but at the cost of a much heavier notation and formalism.*

OA.2.1 Additional assumptions required for the WLS estimation

In this Section, we introduce additional assumptions that are required for the validity of the WLS estimation described in Section 6. Before stating the main assumptions, it is useful to introduce some preliminary notation. In the following, we denote by $\mathbf{w}_{i.} \equiv (\mathrm{w}_{i,1}, \cdots, \mathrm{w}_{i,T-1})'$ the $(T-1) \times 1$ vector of weights specific for the i-th asset, and by $\mathbf{w}_{.t-1} \equiv (\mathrm{w}_{1,t-1}, \cdots, \mathrm{w}_{N,t-1})'$ the $N \times 1$ vector of weights at time $(t-1)$, for every $2 \leq t \leq T$, with the $N \times T$ matrix $\mathbf{W} = (\mathbf{w}_{.1}, \cdots, \mathbf{w}_{.T-1}) = (\mathbf{w}_{1.}, \cdots, \mathbf{w}'_{N.})$.

Assumption OA.8. *(CSR WLS weights)*

[7] One can further generalize (OA.19), allowing for a (dynamic) factor structure such as $\mathbf{z}_{it} = \boldsymbol{\mu}_{iz} + \sum_{k=0}^{\infty} \boldsymbol{\Delta}_{ik}\boldsymbol{\eta}_{i,t-k} + \sum_{k=0}^{\infty} \Delta_{ik}^{\dagger} \boldsymbol{\eta}_{t-k}$, for an i.i.d sequence $\boldsymbol{\eta}_t = (\eta_t^1, \cdots, \eta_t^{K_z})' \sim (\mathbf{0}_{K_z}, \boldsymbol{\Sigma}_{\eta\dagger})$ of *common* shocks. However, in this case, the matrices $\boldsymbol{\Sigma}_{ZZ}$ and $\boldsymbol{\Sigma}_{ZB}$ could have a random component because they depend on terms like $\eta_t^k \eta_s^{k'}$, for every $k, k' = 1, ..., K_z$ and $t, s = 1, ..., T-1$, with their generic (t, s) element not being a function of $(t - s)$ only. However, despite this lack of stationarity, our results continue to hold (conditionally on the $\boldsymbol{\eta}_t$) thanks to the fixed T assumption. The asymptotic distribution of the estimators will be mixed-normal, implying that the test statistics will possess the conventional chi-square distribution and all the testing procedures will still be feasible. Details are available upon request.

(i)

$$\frac{\mathbf{1}_N'\mathbf{W}_{t-1}\mathbf{1}_N}{N} \to_p 1.$$

(ii) For any real number $h > 1$ then,

$$\frac{\mathbf{1}_N'\mathbf{W}_{t-1}^h\mathbf{1}_N}{N} \to_p \mu_{\mathrm{w},t-1}^h$$

(iii)

$$\frac{1}{N}\sum_{i=1}^N \mathbf{w}_{i.}\mathbf{w}_{i.}' \to_p \mathbf{\Sigma}_{\mathrm{W}}.$$

Remark OA.11. *By Assumption OA.8, granularity follows since* $N^{-1}\mathbf{1}_N'\mathbf{W}_{t-1}\mathbf{1}_N \to_p 1$ *by Assumption OA.8(i) and* $N^{-2}\mathbf{1}_N'\mathbf{W}_{t-1}^2\mathbf{1}_N = O(N^{-1}) = o(1)$ *by Assumption OA.8(ii).*

Remark OA.12. *An important example of commonly-used weights satisfying Assumption OA.8 is when*

$$\mathrm{w}_{i,t} = \frac{N\mathrm{w}_{i,t}^{\$}}{\sum_{j=1}^N \mathrm{w}_{j,t}^{\$}}, \tag{OA.20}$$

where $\mathrm{w}_{i,t}^{\$}$ *represents the dollar-value of the market capitalization of stock* i *at time* t*. Notice that the multiplication by* N *in the numerator of* (OA.20) *is just a normalization factor implied by Assumption OA.8(i), which requires that the sample average of the weights goes to 1 as* N *goes to infinity.*

Assumption OA.9. *(Weighted loadings)* Let $\mathbf{W}_{t-1}$ satisfy Assumption OA.8 and let the loadings $\boldsymbol{\beta}_i$ be a non-random sequence. As $N \to \infty$, then

$$\frac{1}{N}\mathbf{B}'\mathbf{W}_{t-1}\mathbf{1}_N \to_p \boldsymbol{\mu}_\beta \qquad \text{and} \qquad \frac{1}{N}\mathbf{B}'\mathbf{W}_{t-1}\mathbf{B} \to_p \mathbf{\Sigma}_\beta, \tag{OA.21}$$

such that

$$\mathbf{\Sigma}_{\mathrm{X}} \equiv \begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \mathbf{\Sigma}_\beta \end{bmatrix} > 0. \tag{OA.22}$$

Remark OA.13. *Assumption OA.9 generalizes Assumption OA.3 to weighted averages. Notice that, due to the granularity assumption, the weighted averages in* (OA.21) *achieve the same limits of un-weighted counterparts in Assumption OA.3.*

Assumption OA.10. *(Weighted cross-sectional moments of returns' innovations)* As $N \to \infty$,

(i) Let $0 < \sigma^2 < \infty$.Then, for every $2 \leq t \leq T$:

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{w}_{i,t-1} \left(\sigma_i^2 - \sigma^2 \right) = o_p \left(\frac{1}{\sqrt{N}} \right), \tag{OA.23}$$

(ii)

$$\sum_{i,j=1}^{N} \mathrm{w}_{i,t-1} \mid \sigma_{ij} \mid \mathbb{1}_{\{i \neq j\}} = o_p \left(N \right). \tag{OA.24}$$

(iii) Let $0 < \mu_4 < \infty$, and let $\mu_{4i} = \mathrm{E}[\epsilon_{it}^4]$. Then, for every $2 \leq t \leq T$:

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{w}_{i,t-1} \mu_{4i} \to_p \mu_4, \tag{OA.25}$$

(iv) Let $0 < \sigma_4 < \infty$. Then, for every $2 \leq t \leq T$:

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{w}_{i,t-1} \sigma_i^4 \to_p \sigma_4, \tag{OA.26}$$

(v) *Let $\kappa_3(a, b, c)$ denote the third-order cumulant of the random variables a, b, and c. Then,*

$$\kappa_3[\epsilon_{i,t}, \epsilon_{j,s}, \mathrm{w}_{j,h}] = 0, \quad \textit{and} \quad \kappa_3[\epsilon_{i,t}, \epsilon_{j,s}, \mathbf{z}_{j,h}] = \mathbf{0}_{K_z}. \tag{OA.27}$$

(vi) Let $\kappa_{4,iiii} = \kappa_4[\epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}]$ denote the fourth-order cumulant of the asset-specific error $\{\epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}\}$. Then, for some $0 \leq |\kappa_4| < \infty$ and for every $2 \leq t \leq T$:

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{w}_{i,t-1} \kappa_{4,iiii} \to_p \kappa_4. \tag{OA.28}$$

(vii) For every $3 \leq h \leq 8$, all the following mixed cumulants of order h satisfy

$$\sup_{i_1} \sum_{i_2,\ldots,i_h=1}^{N} |\kappa_{h,\mathrm{w}_{i_1,t-1} i_2 \ldots i_h}| = o\left(N \right), \tag{OA.29}$$

and

$$\sup_{i_1} \sum_{i_2,\ldots,i_h=1}^{N} |\kappa_{h \mathrm{w}_{i_1,t-1}, \mathbf{z}_{i_2,r}, i_3 \ldots i_h}| = o\left(N \right), \tag{OA.30}$$

for at least one i_j $(2 \leq j \leq h)$ different from i_1, where $\kappa_{h,\mathrm{w}_{i_1,t-1} i_2 \ldots i_h}$ is the mixed cumulant in the $\{\mathrm{w}_{i_1,t-1}, \epsilon_{i_2,s}, \cdots, \epsilon_{i_h,s}\}$ of order h, and $\kappa_{h,\mathrm{w}_{i_1,t-1}, \mathbf{z}_{i_2,r}, i_3 \ldots i_h}$ is the mixed cumulant in the $\{\mathrm{w}_{i_1,t-1}, \mathbf{z}_{i_2,r}, \epsilon_{i_3,s}, \cdots, \epsilon_{i_h,s}\}$ of order h.

Remark OA.14. *Assumption OA.10 extends Assumption OA.5 to the case of weighted averages. Notice that, due to the granularity assumption in Assumption OA.8, all the weighted averages in* (OA.23)–(OA.30) *the same limits of the corresponding un-weighted averages in Assumption OA.5.*

Assumption OA.11. *(Weighted moments and CLT of anomalies) We define the* $(T-1)^2 \times 1$ *vector* $\mathbf{v}_i \equiv (\boldsymbol{\epsilon}_i \otimes \mathbf{w}_i)$ *and the corresponding* $N \times (T-1)^2$ *matrix* $\mathbf{V} \equiv (\mathbf{v}_1, \cdots, \mathbf{v}_N)'$*, such that* $\mathrm{E}[\mathbf{v}_i] \equiv \boldsymbol{\mu}_{\mathrm{v}_i} < \infty$*, and* $\boldsymbol{\Sigma}_{\mathrm{v},ij} \equiv \mathrm{Cov}\left[\mathbf{v}_i, \mathbf{v}_j\right]$.

(i)

$$\frac{\boldsymbol{\epsilon}\left(\mathbf{W}_{t-1} - \mathrm{E}[\mathbf{W}_{t-1}]\right)\boldsymbol{\epsilon}'}{N} \to_p \mathbf{0}_{(T-1)\times(T-1)}.$$

(ii)

$$\frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{1}_N}{N} \quad \to_p \quad \boldsymbol{\mu}_{\mathrm{z},t-1} \text{ and } \frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}}{N} \to_p \boldsymbol{\Sigma}_{\mathrm{Z},t-1}.$$

(iii) *Let* $\boldsymbol{\Sigma}_{\mathrm{ZW}}$ *be a finite* $K_{\mathrm{z}}(T-1) \times (T-1)$ *matrix. Then,*

$$\frac{\mathbf{Z}'\mathbf{W}}{N} \to_p \boldsymbol{\Sigma}_{\mathrm{ZW}}.$$

(iv)

$$\frac{1}{N}\left(\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}]\right)'\left(\mathbf{W}_{t-1} - \mathrm{E}[\mathbf{W}_{t-1}]\right)\boldsymbol{\epsilon}' \to_p \mathbf{0}_{K_{\mathrm{z}}\times(T-1)}.$$

(v)

$$\frac{1}{N}\left(\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}]\right)'\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - \frac{1}{N}\left(\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}]\right)'\boldsymbol{\epsilon}' = o_p\left(N^{-\frac{1}{2}}\right).$$

(vi)

$$\frac{\mathbf{X}'\mathbb{M}_{1_N}\mathbf{V}}{N} = o_p\left(N^{-\frac{1}{2}}\right), \quad \text{and} \quad \frac{\mathbf{Z}'\mathbb{M}_{1_N}\mathbf{V}}{N} = o_p\left(N^{-\frac{1}{2}}\right).$$

(vii)

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathrm{v},ii} \to \boldsymbol{\Sigma}_{\mathrm{V}} \equiv \sigma^2\mathbf{I}_{T-1} \otimes \boldsymbol{\Sigma}_{\mathrm{W}}, \quad \text{and} \quad \sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathrm{v},ij}\mathbb{1}_{i\neq j} = o(N)$$

(viii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{v}_i-\boldsymbol{\mu}_{\mathrm{v}_i})\to_d N\left(\mathbf{0}_{(T-1)^2},\boldsymbol{\Sigma}_{\mathrm{V}}\right)\quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\mu}_{\mathrm{v_i}}=o\left(N^{-\frac{1}{2}}\right).$$

(ix)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i-\boldsymbol{\mu}_{\mathrm{z}_i})\to_d N(\mathbf{0}_{K_z(T-1)},\boldsymbol{\Sigma}_{ZZ})\ \textit{and}\ \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\mu}_{\mathrm{z}_i}-\boldsymbol{\mu}_z)=o(N^{-\frac{1}{2}}).$$

Remark OA.15. *Assumption OA.11 states the limit of several sample moments that involve the anomalies, the weighs and the asset-specific errors, together with their mutual correlation across time. Specifically, we assume that each asset-specific error is uncorrelated with past values of both anomaly variables and weights, but each* $\boldsymbol{\epsilon}_s$ *could be potentially correlated with contemporary or future values of* $\mathbf{Z}_t$ *and* $\mathbf{W}_v$*, whenever* $s \le t$ *or* $s \le v$*. This implies that, despite the granularity assumption, the limit of certain quantities will depend on the weights, as they are allowed to be correlated with both anomalies and error-specific components.*

Remark OA.16. *To simplify the analysis, we also impose zero mixed third-moment conditions. This sometimes involves de-meaning the anomaly variables which, in general, could not have zero mean. The main moment conditions are summarized in the following assumption.*

Remark OA.17. *Assumption (ix) is a strengthening of our previous assumptions required by the added difficulties associated with the weighted estimator.*

OA.2.2 Additional assumptions required for estimation under model misspecification

Assumption OA.12. *(mixed-moments of pricing errors)*

(i)

$$\frac{1}{N}\boldsymbol{\epsilon}\mathbf{m}_{t-1}\to_p\begin{bmatrix}\boldsymbol{\theta}_{t-1,m}\\ \mathbf{0}_{T-t+1}\end{bmatrix},$$

with $\boldsymbol{\theta}_{t-1,m}\equiv(\theta_{t-3,m},\theta_{t-4,m},\dots,\theta_{0,m})'$, defined as, for every $2\le s,t\le T$,

$$\frac{1}{N}\sum_{i=1}^{N}\epsilon_{i,s}m_{i,t-1}\to_p\theta_{t-1-s,m},\ \text{such that}\ \theta_{u,m}=0\ \text{for}\ u<0.$$

(ii)

$$\frac{1}{N}\mathbf{m}_{t-1}'\mathbf{m}_{t-1}\to_p\sigma_{t-1mm}.$$

(iii)

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_{\tilde{D}_i}\boldsymbol{\epsilon}_i m_{i,t-1} \to_p \mathbf{0}_{T-1}.$$

Remark OA.18. *Assumption (OA.12)-(i) is very mild, as the zeros arise only as a consequence of the temporal iid-ness of the $\epsilon_{i,t}$, and (OA.12)-(ii) is ruling out explosive limiting behaviour of the average of the squared pricing errors. Finally, (OA.12)-(iii) simplifies the formulae and is strengthening the notion that the loadings to the omitted risk factors and the omitted anomalies are cross-sectionally unrelated to the anomalies and to the loadings of the risk factors of the candidate model.*

OA.2.3 Additional assumptions required for the cross-sectional R-squared test

In this Section we introduce additional assumptions that are required to derive the R-squared test described in Section 8.

Assumption OA.13. (i)

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta = o\left(N^{-\frac{1}{2}}\right) \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i\boldsymbol{\beta}_i' - \boldsymbol{\Sigma}_\beta = o\left(N^{-\frac{1}{2}}\right).$$

(ii)

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left((\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}(\boldsymbol{\Sigma}_\text{Z})\right) \to_d N(\mathbf{0}_{K_\text{z}^2},\mathbf{U}_\text{Z}), \quad \text{with}\\
&\frac{1}{N}\sum_{i=1}^{N}\text{E}\left[(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}(\boldsymbol{\Sigma}_\text{Z})\right] = o\left(N^{-\frac{1}{2}}\right), \ \frac{1}{N}\sum_{i=1}^{N}\text{E}\left[(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}(\boldsymbol{\Sigma}_\text{Z})\right]\left[(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}(\boldsymbol{\Sigma}_\text{Z})\right]' \to \mathbf{U}_\text{Z},\\
&\sum_{\substack{i,j=1\\ i\neq j}}^{N}\text{E}\left[(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}(\boldsymbol{\Sigma}_\text{Z})\right]\left[(\mathbf{z}_j\otimes\mathbf{z}_j)-\text{vec}(\boldsymbol{\Sigma}_\text{Z})\right]' = o(N), \ \text{and} \ \frac{1}{N}\sum_{i,j=1}^{N}\text{Cov}\left[(\mathbf{z}_i\otimes\mathbf{z}_i),\mathbf{z}_j'\right] \to \boldsymbol{\Sigma}_{\text{z}\otimes\text{z}}.
\end{aligned}$$

(iii)

$$\sqrt{N}\left(\frac{\mathbf{Z}'\mathbf{1}_N}{N} - \boldsymbol{\mu}_{\text{z},T-1}\right) \to_d N\left(\mathbf{0}_{K_\text{z}(T-1)}, \boldsymbol{\Sigma}_\text{Z} - \boldsymbol{\mu}_{\text{z},T-1}\boldsymbol{\mu}_{\text{z},T-1}'\right).$$

(iv)

$$\frac{1}{N}\sum_{i,j=1}^{N}\text{Cov}\left((\mathbf{z}_i\otimes\mathbf{z}_i),(\boldsymbol{\epsilon}_j'\otimes\boldsymbol{\epsilon}_j')\right) \to \boldsymbol{\Sigma}_{\text{Z}\otimes\epsilon} = \mathbf{0}_{((T-1)K_z)^2\times(T-1)^2}.$$

(v)

$$\frac{1}{N}\sum_{i,j=1}^{N}\text{Cov}\left((\mathbf{z}_i\otimes\mathbf{z}_i),\mathbf{u}_j'\right)\to\mathbf{\Sigma}_{\text{ZU}}=\mathbf{0}_{((T-1)K_z)^2\times(T-1)^2K_z}.$$

OA.3 Preliminary Lemmas

In this section, we establish several results which are required to derive the asymptotic results of the new CSR OLS estimators. All the results below hold as $N\to\infty$.

Lemma 1 (Raponi, Robotti, and Zaffaroni (2020)). *Under Assumptions OA.1–OA.7,*

(i)

$$\hat{\sigma}^2-\sigma^2=O_p\left(\frac{1}{\sqrt{N}}\right).$$

(ii)

$$\hat{\mathbf{X}}'\hat{\mathbf{X}}=O_p(N).$$

(iii)

$$\hat{\mathbf{\Sigma}}_{\text{X}}\to_p\mathbf{\Sigma}_{\text{X}}+\mathbf{\Lambda}_1.$$

(iv)

$$\frac{(\hat{\mathbf{X}}-\mathbf{X})'(\hat{\mathbf{X}}-\mathbf{X})}{N}\to_p\mathbf{\Lambda}_1.$$

(v)

$$\mathbf{X}'\boldsymbol{\epsilon}_t=O_p\left(\sqrt{N}\right),\ \ \mathbf{X}'\bar{\boldsymbol{\epsilon}}=O_p\left(\sqrt{N}\right).$$

(vi)

$$(\hat{\mathbf{X}}-\mathbf{X})'\mathbf{X}\mathbf{\Gamma}_{\text{f},t}=O_p\left(\sqrt{N}\right),\quad(\hat{\mathbf{X}}-\mathbf{X})'\mathbf{X}\bar{\mathbf{\Gamma}}_{\text{f}}=O_p\left(\sqrt{N}\right).$$

(vii)

$$(\hat{\mathbf{X}}-\mathbf{X})'\bar{\boldsymbol{\epsilon}}=O_p\left(\sqrt{N}\right).$$

(viii)

$$(\hat{\mathbf{X}}-\mathbf{X})'\boldsymbol{\epsilon}_t=\begin{bmatrix}0\\-\sigma^2\mathbf{P}'\boldsymbol{\imath}_{t-1,T-1}\end{bmatrix}+O_p\left(\sqrt{N}\right).$$

(ix) When the identification assumption $\kappa_4 = 0$ holds, then

$$\hat{\sigma}_4 \to_p \sigma_4.$$

where $\hat{\sigma}_4$ is defined in Theorem 2.

Proof. All these results are already established - or are immediate extensions - of Lemmas 1–7 in Raponi, Robotti, and Zaffaroni (2020). ■

Lemma 2. *Under Assumptions OA.1–OA.7,*

(i)

$$\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{N} = \mathbb{J}'\frac{\mathbf{Z}'\mathbf{Z}}{N}\mathbb{J} \to_p \mathbb{J}'\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J}.$$

(ii)

$$\frac{\bar{\mathbf{Z}}'\hat{\mathbf{X}}}{N} \to_p \left[\boldsymbol{\mu}_{\mathrm{z}}, \mathbb{J}'\mathbf{\Sigma}_{\mathrm{ZB}}\right] \equiv \mathbf{\Sigma}_{\mathrm{ZX}}.$$

(iii)

$$\frac{\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}}{N} = \mathbb{J}_{t-1}'\frac{\mathbf{Z}'\mathbf{Z}}{N}\mathbb{J}_{t-1} \to_p \mathbb{J}_{t-1}'\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1}.$$

(iv)

$$\frac{\mathbf{Z}_{t-1}'\hat{\mathbf{X}}}{N} \to_p \left[\boldsymbol{\mu}_{\mathrm{z}}, \mathbb{J}_{t-1}'\mathbf{\Sigma}_{\mathrm{ZB}}\right] \equiv \mathbf{\Sigma}_{\mathrm{ZX,t-1}}.$$

Proof. Part (i) follows immediately from Assumption OA.7(ii), once one recognizes that $\mathbf{Z}\mathbb{J} = \bar{\mathbf{Z}}$. To prove part (ii), first notice that the first column of the matrix $\frac{\bar{\mathbf{Z}}'\hat{\mathbf{X}}}{N}$ coincides with $\frac{\bar{\mathbf{Z}}'\mathbf{1}_N}{N}$, which converges to $\boldsymbol{\mu}_{\mathrm{z},T-1}$, by Assumption OA.7(i). The remaining K_{f} columns of the matrix satisfy:

$$\frac{\bar{\mathbf{Z}}'\hat{\mathbf{B}}}{N} = \frac{\bar{\mathbf{Z}}'\mathbf{B}}{N} + \frac{\bar{\mathbf{Z}}'\boldsymbol{\epsilon}'}{N}\mathbf{P} \to_p \mathbb{J}'\mathbf{\Sigma}_{\mathrm{ZB}},$$

by Assumptions OA.7(iii) and OA.6(i). Parts (iii) and (iv) follow precisely from parts (i) and (ii), by replacing $\mathbb{J}$ with $\mathbb{J}_{t-1}$. ■

Lemma 3. *Under Assumptions OA.1–OA.7, (i)*

$$\frac{1}{N}\hat{\mathbf{X}}'\left(\bar{\boldsymbol{\epsilon}} + (\mathbf{X} - \hat{\mathbf{X}})\bar{\mathbf{\Gamma}}_{\mathrm{f}}\right) \to_p -\mathbf{\Lambda}_1\bar{\mathbf{\Gamma}}_{\mathrm{f}}. \tag{OA.31}$$

(ii)

$$\frac{1}{N}\bar{\mathbf{Z}}'\left(\bar{\boldsymbol{\epsilon}} + (\mathbf{X} - \hat{\mathbf{X}})\bar{\boldsymbol{\Gamma}}_{\mathrm{f}}\right) \to_p \mathbf{0}_{K_{\mathrm{z}}}. \quad \text{(OA.32)}$$

(iii)

$$\frac{1}{N}\hat{\mathbf{X}}'\left(\boldsymbol{\epsilon}_t + (\mathbf{X} - \hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1}\right) \to_p -\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_{\mathrm{f},t-1} + \boldsymbol{\Lambda}_{2,t-1}. \quad \text{(OA.33)}$$

(iv)

$$\frac{1}{N}\mathbf{Z}'_{t-1}\left(\boldsymbol{\epsilon}_t + (\mathbf{X} - \hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1}\right) \to_p \mathbf{0}_{K_{\mathrm{z}}}. \quad \text{(OA.34)}$$

Proof. Parts (i) and (iii) follow immediately by Lemma 1, where in part (iii) we set $\boldsymbol{\Lambda}_{2,t-1} \equiv \begin{bmatrix} 0 \\ -\sigma^2\mathbf{P}'\boldsymbol{\imath}_{t-1,T-1} \end{bmatrix}$. ■

OA.3.1 Additional Lemmas required for WLS Estimation

This section establishes several preliminary results which are needed to derive the asymptotic results of the WLS OLS estimator described in Section 6. All the results below hold as $N \to \infty$.

Lemma 4. *Under Assumptions OA.1–OA.7 and OA.8-OA.11,*

(i)

$$\frac{\boldsymbol{\epsilon}\mathbf{W}_{t-1}\mathbf{1}_N}{N} \to_p \mathbf{0}_{T-1}.$$

(ii)

$$\frac{\boldsymbol{\epsilon}\mathbf{W}_{t-1}\mathbf{B}}{N} \to_p \mathbf{0}_{(T-1)\times K_{\mathrm{f}}}$$

(iii)

$$\frac{\hat{\mathbf{B}}'\mathbf{W}_{t-1}\hat{\mathbf{B}}}{N} \to_p \boldsymbol{\Sigma}_\beta + \sigma^2\mathbf{P}'\mathbf{P}.$$

(iv)

$$\frac{\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{N} \to_p \sigma^2\mathbf{I}_{T-1}.$$

Proof. The results in parts (i) and (ii) follow from Assumptions OA.8-OA.11, since

$$\frac{\boldsymbol{\epsilon}\mathbf{W}_{t-1}\mathbf{1}_N}{N} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i \mathrm{w}_{i,t-1} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i\left(\mathrm{w}_{i,t-1} - \mathrm{E}[\mathrm{w}_{i,t-1}]\right) + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i \mathrm{E}[\mathrm{w}_{i,t-1}] + o_p(1) \to_p \mathbf{0}_{T-1},$$

and, by independence between $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\beta}_i$,

$$\begin{aligned}\frac{\boldsymbol{\epsilon}\mathbf{W}_{t-1}\mathbf{B}}{N} &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i \mathrm{w}_{i,t-1}\boldsymbol{\beta}_i' \\ &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i \mathrm{w}_{i,t-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i' + o_p(1) \to_p \mathbf{0}_{(T-1)\times K_{\mathrm{f}}}.\end{aligned}$$

To prove part (iii), notice that

$$\hat{\mathbf{B}}'\mathbf{W}_{t-1}\hat{\mathbf{B}} = \mathbf{B}'\mathbf{W}_{t-1}\mathbf{B} + (\hat{\mathbf{B}}-\mathbf{B})'\mathbf{W}_{t-1}(\hat{\mathbf{B}}-\mathbf{B}) + (\hat{\mathbf{B}}-\mathbf{B})'\mathbf{W}_{t-1}\mathbf{B} + \mathbf{B}'\mathbf{W}_{t-1}(\hat{\mathbf{B}}-\mathbf{B}).$$

By Assumption OA.9, $N^{-1}\mathbf{B}'\mathbf{W}_{t-1}\mathbf{B} \to_p \boldsymbol{\Sigma}_\beta$ and, using part (i) and (ii),

$$\frac{(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{W}_{t-1}\mathbf{B}}{N} = \mathbf{P}'\frac{1}{N}\sum_{i=1}^{N}\mathrm{w}_{i,t-1}\boldsymbol{\epsilon}_i\boldsymbol{\beta}_i' \to_p \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{f}}}.$$

Finally, using the same arguments

$$\frac{(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{W}_{t-1}(\hat{\mathbf{B}}-\mathbf{B})}{N} = \frac{1}{N}\mathbf{P}'\sum_{i=1}^{N}\mathrm{w}_{i,t-1}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'\mathbf{P} \to_p \sigma^2\mathbf{P}'\mathbf{P}.$$

■

Lemma 5. *Under Assumptions OA.1–OA.7 and OA.8-OA.11,*

(i)

$$\frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{B}}{N} \to_p \boldsymbol{\mu}_{\mathrm{z},t-1}\boldsymbol{\mu}_\beta'.$$

(ii)

$$\frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N} \to_p \mathbf{0}_{K_{\mathrm{z}}}.$$

(ii)

$$\frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{N} \to_p \mathbf{0}_{K_{\mathrm{z}}\times(T-1)}.$$

Proof. Part (i) follows immediately by taking into account the non-randomness of the $\boldsymbol{\beta}_i$

$$\begin{aligned}\frac{\mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{B}}{N} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_{i,t-1}\mathrm{w}_{i,t-1}\boldsymbol{\beta}'_i = \frac{1}{N}\sum_{i=1}^{N}\mathrm{w}_{i,t-1}\mathbf{z}_{i,t-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}'_i + o_p(1)\\ &\to_p \boldsymbol{\mu}_{\mathrm{z},t-1}\boldsymbol{\mu}'_{\beta},\end{aligned}$$

by Assumptions OA.9 and OA.11. Using the same arguments and by Lemma 4, part (ii) follows since

$$\frac{\mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_{i,t-1}\mathrm{w}_{i,t-1}\boldsymbol{\epsilon}'_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_{i,t-1}\mathrm{w}_{i,t-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\epsilon}'_i + o_p(1) \to_p \mathbf{0}_{K_{\mathrm{z}}}.$$

For part (iii), notice that, by Assumption OA.11,

$$\begin{aligned}\frac{\mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{N} &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{w}_{i,t-1}\mathbf{z}_{i,t-1}\boldsymbol{\epsilon}'_i\\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{w}_{i,t-1} - \mathrm{E}[\mathrm{w}_{i,t-1}] + \mathrm{E}[\mathrm{w}_{i,t-1}]\right)\left(\mathbf{z}_{i,t-1} - \mathrm{E}[\mathbf{z}_{i,t-1}] + \mathrm{E}[\mathbf{z}_{i,t-1}]\right)\boldsymbol{\epsilon}'_i\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[\mathrm{w}_{i,t-1}]\mathrm{E}[\mathbf{z}_{i,t-1}]\boldsymbol{\epsilon}'_i + \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{w}_{i,t-1} - \mathrm{E}[\mathrm{w}_{i,t-1}]\right)\mathrm{E}[\mathbf{z}_{i,t-1}]\boldsymbol{\epsilon}'_i\\ &+ \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[\mathrm{w}_{i,t-1}]\left(\mathbf{z}_{i,t-1} - \mathrm{E}[\mathbf{z}_{i,t-1}]\right)\boldsymbol{\epsilon}'_i + \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{w}_{i,t-1} - \mathrm{E}[\mathrm{w}_{i,t-1}]\right)\left(\mathbf{z}_{i,t-1} - \mathrm{E}[\mathbf{z}_{i,t-1}]\right)\boldsymbol{\epsilon}'_i\\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{w}_{i,t-1} - \mathrm{E}[\mathrm{w}_{i,t-1}]\right)\mathrm{E}[\mathbf{z}_{i,t-1}]\boldsymbol{\epsilon}'_i + \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[\mathrm{w}_{i,t-1}]\left(\mathbf{z}_{i,t-1} - \mathrm{E}[\mathbf{z}_{i,t-1}]\right)\boldsymbol{\epsilon}'_i + o_p(1)\\ &\to_p \mathbf{0}_{K_{\mathrm{z}}\times(T-1)}\end{aligned}$$

■

Lemma 6. *Under Assumptions OA.1–OA.7 and OA.8-OA.11,*

$$\frac{\hat{\mathbf{X}}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N} \to_p \sigma^2\begin{bmatrix}0\\ \mathbf{P}'\imath_{t,T-1}\end{bmatrix}.$$

Proof. Let us rewrite

$$\frac{\hat{\mathbf{X}}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N} = \frac{\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N} + \frac{(\hat{\mathbf{X}} - \mathbf{X})'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N}.$$

Now, by Lemma 4, $N^{-1}\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t \to_p \mathbf{0}_{K_f+1}$. Moreover,

$$\begin{aligned}\frac{(\hat{\mathbf{X}}-\mathbf{X})'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t}{N} &= \frac{(\hat{\mathbf{X}}-\mathbf{X})'\left(\mathbf{W}_{t-1}-\mathrm{E}[\mathbf{W}_{t-1}]+\mathrm{E}[\mathbf{W}_{t-1}]\right)\boldsymbol{\epsilon}_t}{N} \\ &= \frac{(\hat{\mathbf{X}}-\mathbf{X})'\mathrm{E}[\mathbf{W}_{t-1}]\boldsymbol{\epsilon}_t}{N} + o_p(1) \to_p \sigma^2 \begin{bmatrix} 0 \\ \mathbf{P}'\imath_{t,T-1} \end{bmatrix}.\end{aligned}$$

■

Lemma 7. *Under Assumptions OA.1–OA.7 and OA.8-OA.11, when $\kappa_4 = 0$, for every $2 \leq t \leq T$,*

$$\hat{\sigma}^{2(\mathrm{w})}_{t-1} \to_p \sigma^2 \quad \textit{and} \quad \hat{\sigma}^{(\mathrm{w})}_{4,t-1} \to_p \sigma_4,$$

setting $\hat{\sigma}^{(\mathrm{w})}_{4,t-1} \equiv N^{-1}\sum_{s=2}^{T}\sum_{i=1}^{N} \mathrm{w}_{i,t-1}\hat{\epsilon}^4_{i,s}/(3\mathrm{tr}\left(\overline{\mathbb{M}}^{(2)}_{\tilde{\mathrm{D}}}\right)),$

Proof. The result can be easily obtained by following the corresponding results in Raponi, Robotti, and Zaffaroni (2020) (see their Lemmas 1 and 6), by replacing simple averages (e.g. $\boldsymbol{\epsilon}\boldsymbol{\epsilon}'$) with the corresponding weighted averages (e.g., $\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'$) and using the Assumptions OA.8–OA.11 ■

OA.3.2 Additional Lemmas required for OLS Estimation under model misspecification

Lemma 8. *Under Assumptions OA.1–OA.7 and OA.12,*

$$\hat{\boldsymbol{\theta}}_{t-1,m} \equiv \mathbf{S}_{t-1}\left(\frac{1}{N}\hat{\boldsymbol{\epsilon}}'\hat{\mathbf{m}}_{t-1} - \hat{\sigma}^2\mathbf{M}_{\tilde{D}}\boldsymbol{\iota}_{t-1,T-1}\right) \to_p \boldsymbol{\theta}_{t-1,m},$$

setting $\hat{\mathbf{m}}_{t-1} \equiv \mathbf{R}_t - (\hat{\mathbf{X}}, \mathbf{Z}_{t-1})(\hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1}, \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})\prime}_{\mathrm{z},t-1})'$ *and recalling* $\mathbf{M}^{(-1)}_{D,t-1} = \mathbf{M}^{-1}_{11}(\mathbf{I}_{t-2}, \mathbf{0}_{t-2\times T-t+1})$ *where* $\mathbf{M}_{11}$ *is the top-left block of size* $(t-2)\times(t-2)$ *of the matrix* $\mathbf{M}_D$*.*

Proof. For every $2 \leq s \leq t-1$,

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{is}\hat{m}_{i,t-1}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\iota}'_{s-1,T-1}\mathbb{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i\Big(m_{i,t-1} + (\mathbf{x}'_i - \hat{\mathbf{x}}'_i)\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + \hat{\mathbf{x}}'_i(\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}) + \mathbf{z}'_{i,t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\hat{\boldsymbol{\gamma}}}^{*(\mathrm{m})}_{\mathrm{z},t-1}) + \epsilon_{i,t}\Big)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\iota}'_{s-1,T-1}\mathbb{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i\Big(m_{i,t-1} + \epsilon_{i,t}\Big) + O_p(N^{-\frac{1}{2}}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\iota}'_{s-1,T-1}(\mathbb{M}_D - \mathbb{P}_{\tilde{Z}_i})\boldsymbol{\epsilon}_i\Big(m_{i,t-1} + \epsilon_{i,t}\Big) + O_p(N^{-\frac{1}{2}})$$
$$= \frac{1}{N}\boldsymbol{\iota}'_{s-1,T-1}\mathbb{M}_D\boldsymbol{\epsilon}\Big(\mathbf{m}_{t-1} + \boldsymbol{\epsilon}_t\Big) - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\iota}'_{s-1,T-1}\mathbb{P}_{\tilde{Z}_i}\boldsymbol{\epsilon}_i\Big(m_{i,t-1} + \epsilon_{i,t}\Big) + O_p(N^{-\frac{1}{2}})$$
$$\rightarrow_p \boldsymbol{\iota}'_{s-1,T-1}(\mathbb{M}_D\begin{bmatrix}\boldsymbol{\theta}_{t-1,m}\\ \mathbf{0}_{T-t+1}\end{bmatrix} + \sigma^2\mathbb{M}_D\boldsymbol{\iota}_{t-1,T-1}),$$

as $N^{-1}\sum_{i=1}^{N}\mathbb{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i(\mathbf{x}'_i - \hat{\mathbf{x}}'_i)\boldsymbol{\Gamma}_{t-1} = O_p(N^{-\frac{1}{2}})$, given $\mathbf{M}_{\tilde{D}_i}P = \mathbf{0}_{T-1\times K_f}$, recalling that $\mathbf{M}_{\tilde{D}_i} = \mathbb{M}_D - \mathbf{P}_{\tilde{Z}_i}$, implying $N^{-1}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i m_{i,t-1} = \mathbb{M}_D(N^{-1}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i m_{i,t-1}) + o_p(1)$. Therefore, recalling $\mathbf{M}_D = \begin{bmatrix}\mathbf{M}_{11} & \mathbf{M}_{12}\\ \mathbf{M}'_{12} & \mathbf{M}_{22}\end{bmatrix}$, where $\mathbf{M}_{11}$ is $t-2\times t-2$,

$$\begin{bmatrix}\frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{i,2}\hat{m}_{i,t-1}\\ \frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{i,3}\hat{m}_{i,t-1}\\ \vdots\\ \frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{i,t-1}\hat{m}_{i,t-1}\end{bmatrix} \rightarrow_p = \mathbf{M}_{11}\boldsymbol{\theta}_{t-1,m} + \sigma^2(\mathbf{I}_{t-2}, \mathbf{0}_{t-2\times T-t+1})\mathbf{M}_D\boldsymbol{\iota}_{t-1,T-1}).$$

Reorganizing terms, assuming without loss of generality that $\mathbf{M}_{11}$ is nonsingular, as $\mathbf{D}$ is full rank (for $t < T - K_f$), yields

$$\hat{\boldsymbol{\theta}}_{t-1,m} \equiv \mathbf{M}_{11}^{-1}\Big[\Big(\begin{bmatrix}\frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{i,2}\hat{m}_{i,t-1}\\ \frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{i,3}\hat{m}_{i,t-1}\\ \vdots\\ \frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_{i,t-1}\hat{m}_{i,t-1}\end{bmatrix}\Big) - \hat{\sigma}^2(\mathbf{I}_{t-2}, \mathbf{0}_{t-2\times T-t+1})\mathbf{M}_D\boldsymbol{\iota}_{t-1,T-1}\Big]$$
$$= \mathbf{S}_{t-1}(\frac{1}{N}\hat{\boldsymbol{\epsilon}}\hat{\mathbf{m}}_{t-1} - \hat{\sigma}^2\mathbf{M}_D\boldsymbol{\iota}_{t-1,T-1}) \rightarrow_p \boldsymbol{\theta}_{t-1,m}.$$

QED

Lemma 9. *Under Assumptions OA.1–OA.7 and OA.12,*

$$\frac{\mathbf{1}'_N\hat{\mathbf{m}}_{t-1}}{N} \rightarrow_p 0,$$

$$\hat{\sigma}_{t-1,mm} \equiv \frac{\hat{\mathbf{m}}'_{t-1}\hat{\mathbf{m}}_{t-1}}{N} - \hat{\sigma}^2\hat{\mathbf{Q}}'_{t-1}\hat{\mathbf{Q}}_{t-1} + 2\hat{\boldsymbol{\delta}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1}\mathbf{P}'\begin{bmatrix}\hat{\boldsymbol{\theta}}_{t-1,m}\\ \mathbf{0}_{T-t+1}\end{bmatrix} \rightarrow_p \sigma_{t-1,mm}.$$

Proof. Let us first establish the second statement. Consider

$$
\begin{aligned}
&\hat{\mathbf{m}}_{t-1}'\hat{\mathbf{m}}_{t-1} = \mathbf{m}_{t-1}'\mathbf{m}_{t-1} + \boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t + \tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}'(\mathbf{X}-\hat{\mathbf{X}})'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + (\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1})'\hat{\mathbf{X}}'\hat{\mathbf{X}}(\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1})\\
&+(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{m*}_{\mathrm{z},t-1})'\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{m*}_{\mathrm{z},t-1}) + 2\mathbf{m}_{t-1}'\boldsymbol{\epsilon}_t + 2\mathbf{m}_{t-1}'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{t-1} + 2\mathbf{m}_{t-1}'\hat{\mathbf{X}}(\tilde{\boldsymbol{\Gamma}}_{t-1} - \hat{\boldsymbol{\Gamma}}^{m*}_{t-1})\\
&+2\mathbf{m}_{t-1}'\mathbf{Z}_{t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1}) + 2\boldsymbol{\epsilon}_t'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + 2\boldsymbol{\epsilon}_t'\hat{\mathbf{X}}(\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}) + 2\boldsymbol{\epsilon}_t'\mathbf{Z}_{t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1})\\
&+2\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}'(\mathbf{X}-\hat{\mathbf{X}})'\hat{\mathbf{X}}(\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}) + 2\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}'(\mathbf{X}-\hat{\mathbf{X}})'\mathbf{Z}_{t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1})\\
&+2(\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1})'\hat{\mathbf{X}}'\mathbf{Z}_{t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1})\\
&= \mathbf{m}_{t-1}'\mathbf{m}_{t-1} + \boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t + \tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}'(\mathbf{X}-\hat{\mathbf{X}})'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + 2\mathbf{m}_{t-1}'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + 2\boldsymbol{\epsilon}_t'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + O_p(N^{\frac{1}{2}}).
\end{aligned}
$$

The result then follows, noticing that

$$
\begin{aligned}
&\hat{\sigma}^2 - \frac{\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t}{N} \to_p 0,\\
&\hat{\sigma}^2\hat{\boldsymbol{\delta}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1}\mathbf{P}'\mathbf{P}\hat{\boldsymbol{\delta}}^{*(\mathrm{m})}_{\mathrm{f},t-1} - \tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}'\frac{(\mathbf{X}-\hat{\mathbf{X}})'(\mathbf{X}-\hat{\mathbf{X}})}{N}\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \to_p 0,\\
&2(\hat{\boldsymbol{\theta}}_{t-1,m}', \mathbf{0}_{T-t+1}')\mathbf{P}\hat{\boldsymbol{\delta}}^{*(\mathrm{m})}_{\mathrm{f},t-1} - 2\frac{\mathbf{m}_{t-1}'(\mathbf{X}-\hat{\mathbf{X}})}{N}\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \to_p 0 \text{ and}\\
&2\hat{\sigma}^2\boldsymbol{\iota}_{t-1,T-1}'\mathbf{P}\hat{\boldsymbol{\delta}}^{*(\mathrm{m})}_{\mathrm{f},t-1} - 2\frac{\boldsymbol{\epsilon}_t'(\mathbf{X}-\hat{\mathbf{X}})}{N}\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \to_p 0,
\end{aligned}
$$

and collecting terms, recognizing that $\hat{\mathbf{Q}}_{t-1}'\hat{\mathbf{Q}}_{t-1} = 1 + \hat{\boldsymbol{\delta}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1}\mathbf{P}'\mathbf{P}\hat{\boldsymbol{\delta}}^{*(\mathrm{m})}_{\mathrm{f},t-1} - 2\boldsymbol{\iota}_{t-1,T-1}'\mathbf{P}\hat{\boldsymbol{\delta}}^{*(\mathrm{m})}_{\mathrm{f},t-1}$.

The first statement easily follows from $\mathbf{1}_N'\hat{\mathbf{m}}_{t-1} = \mathbf{1}_N'\mathbf{m}_{t-1} + \mathbf{1}_N'\boldsymbol{\epsilon}_t + \mathbf{1}_N'(\mathbf{X}-\hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} + \mathbf{1}_N'\hat{\mathbf{X}}(\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}) + \mathbf{1}_N'\mathbf{Z}_{t-1}(\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1}) = \mathbf{1}_N'\mathbf{m}_{t-1} + O_p(N^{\frac{1}{2}})$. QED

OA.4 Proofs of theorems

Proof of Theorem 1. First, let us define the quantities that make the asymptotic covariance matrix of the estimator:

$$\mathbf{L}_{t-1} \equiv \begin{bmatrix} \mathbf{\Sigma}_{\mathrm{X}} & \mathbf{\Sigma}'_{\mathrm{ZX,t-1}} \\ \mathbf{\Sigma}_{\mathrm{ZX,t-1}} & \mathbb{J}'_{t-1}\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1} \end{bmatrix}, \quad \text{and} \quad \mathbf{O}_{t-1} \equiv \begin{bmatrix} \mathbf{U}_{t-1} & \sigma^2\mathbf{G}_{t-1}\mathbf{H}'_{t-1} \\ \sigma^2\mathbf{H}_{t-1}\mathbf{G}'_{t-1} & \mathbf{H}_{t-1}\mathbf{\Sigma}_{\mathrm{U}}\mathbf{H}'_{t-1} \end{bmatrix}, \tag{OA.35}$$

with $\mathbf{L}_{t-1}$ is assumed to be positive definite, $\mathbf{U}_{t-1} \equiv \sigma^2\mathbf{Q}'_{t-1}\mathbf{Q}_{t-1}\mathbf{\Sigma}_{\mathrm{X}} + \begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} \\ \mathbf{0}_{K_{\mathrm{f}}} & \mathbf{V}'_{t-1}\mathbf{U}_{\epsilon}\mathbf{V}_{t-1} \end{bmatrix}$, $N^{-1}\sum_{i=1}^{N}\mathbb{M}_{\tilde{D}_i} \to_p \mathbb{M}_{\tilde{D}}$, $\mathbf{Z}'_{t-1}\hat{\mathbf{X}}/N \to_p \mathbf{\Sigma}_{\mathrm{ZX,t-1}}$, $\mathbf{V}_{t-1} \equiv (\mathbf{Q}_{t-1}\otimes\mathbf{P}) - \big(\mathrm{vec}(\mathbb{M}_{\tilde{D}})/(T-2-K)\big)\,\mathbf{Q}'_{t-1}\mathbf{P}$, $\mathbf{G}_{t-1} \equiv \left[\mathbf{Q}_{t-1}\otimes\boldsymbol{\mu}_{\mathrm{z},T-1}, \quad \mathbf{Q}_{t-1}\otimes\mathbf{\Sigma}_{\mathrm{ZB}}\right]'$, and $\mathbf{H}_{t-1} \equiv \mathbf{Q}'_{t-1}\otimes\mathbb{J}'_{t-1}$, with $\mathbf{Q}_{t-1} \equiv \boldsymbol{\imath}_{t-1,T-1} - \mathbf{P}\boldsymbol{\delta}_{\mathrm{f},t-1}$, and where $\mathbf{\Sigma}_{\mathrm{X}}$, $\mathbf{U}_{\epsilon}$, $\mathbf{\Sigma}_{\mathrm{ZB}}$, $\mathbf{\Sigma}_{\mathrm{Z}}$, $\mathbf{\Sigma}_{\mathrm{U}}$ and $\boldsymbol{\mu}_{\mathrm{z},T-1}$ are defined in Assumptions OA.3, OA.6, and OA.7.

Starting from (31), we rewrite

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{\Gamma}}^*_{\mathrm{f},t-1} \\ \hat{\gamma}^*_{\mathrm{z},t-1} \end{bmatrix} &= \begin{bmatrix} \mathbf{\Gamma}_{\mathrm{f},t-1} \\ \gamma_{\mathrm{z,t-1}} \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\mathbf{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \times \\ &\times \left(\begin{bmatrix} N\hat{\mathbf{\Lambda}}_1 \\ \mathbf{0}_{K_{\mathrm{z}}\times(K_{\mathrm{f}}+1)} \end{bmatrix} \mathbf{\Gamma}_{\mathrm{f},t-1} + \begin{bmatrix} \hat{\mathbf{X}}' \\ \mathbf{Z}'_{t-1} \end{bmatrix} \left(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\mathbf{\Gamma}_{\mathrm{f},t-1}\right) - \begin{bmatrix} N\hat{\mathbf{\Lambda}}_{2,t-1} \\ \mathbf{0}_{K_{\mathrm{z}}} \end{bmatrix} \right) \end{aligned} \tag{OA.36}$$

By Lemmas 1-3

$$\hat{\mathbf{\Lambda}}_1\mathbf{\Gamma}_{\mathrm{f},t-1} + \frac{1}{N}\hat{\mathbf{X}}'\left(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\mathbf{\Gamma}_{t-1}\right) - \hat{\mathbf{\Lambda}}_{2,t-1} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \text{and that}$$

$$\frac{1}{N}\mathbf{Z}'_{t-1}\left(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\mathbf{\Gamma}_{\mathrm{f},t-1}\right) \to_p \mathbf{0}_{K_{\mathrm{z}}}$$

Moreover, by Lemmas 1 and 2, $N^{-1}\begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\mathbf{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} = O_p(1)$.

To prove part (ii), noticing that $\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}'\boldsymbol{\imath}_{t-1,T-1}$, we get that

$$\frac{1}{\sqrt{N}}\left(N\hat{\mathbf{\Lambda}}_1\mathbf{\Gamma}_{\mathrm{f},t-1}+\hat{\mathbf{X}}'\left(\boldsymbol{\epsilon}_t+(\mathbf{X}-\hat{\mathbf{X}})\mathbf{\Gamma}_{\mathrm{f},t-1}\right)-N\hat{\mathbf{\Lambda}}_{2,t-1}\right)=$$

$$=\begin{bmatrix}\frac{\mathbf{1}_N'\epsilon'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\end{bmatrix}+\begin{bmatrix}0\\ \mathcal{P}'\frac{\boldsymbol{\epsilon\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}+\sqrt{N}\hat{\sigma}^2\mathbf{P}'\mathbf{P}\boldsymbol{\delta}_{\mathrm{f},t-1}\end{bmatrix}+\begin{bmatrix}0\\ \hat{\sigma}^2\mathbf{P}'\boldsymbol{\imath}_{t-1,T-1}\end{bmatrix} \tag{OA.37}$$

$$=\begin{bmatrix}\frac{\mathbf{1}_N'\epsilon'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\end{bmatrix}+\begin{bmatrix}0\\ \mathcal{P}'\frac{\boldsymbol{\epsilon\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}-\sqrt{N}\hat{\sigma}^2\mathbf{P}'\mathbf{Q}_{t-1}\end{bmatrix}.$$

Moreover, we have that

$$\frac{1}{\sqrt{N}}\mathbf{Z}_{t-1}'\left(\boldsymbol{\epsilon}_t+(\mathbf{X}-\hat{\mathbf{X}})\mathbf{\Gamma}_{\mathrm{f},t-1}\right)=\frac{\mathbf{Z}_{t-1}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \tag{OA.38}$$

Therefore, using (OA.37) and (OA.38) in (OA.36), it follows that

$$\begin{aligned}\sqrt{N}\begin{bmatrix}\hat{\mathbf{\Gamma}}^*_{\mathrm{f},t-1}-\mathbf{\Gamma}_{\mathrm{f},t-1}\\ \hat{\gamma}^*_{\mathrm{z},t-1}-\boldsymbol{\gamma}_{\mathrm{z,t-1}}\end{bmatrix} &= \begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-\hat{\mathbf{\Lambda}}_1 & \frac{\hat{\mathbf{X}}'\mathbf{Z}_{t-1}}{N}\\ \frac{\mathbf{Z}_{t-1}'\hat{\mathbf{X}}}{N} & \frac{\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}}{N}\end{bmatrix}^{-1}\times\\ &\quad\left(\begin{bmatrix}\frac{\mathbf{1}_N'\epsilon'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \mathbf{0}_{K_\mathrm{z}}\end{bmatrix}+\begin{bmatrix}0\\ \mathbf{P}'\frac{\boldsymbol{\epsilon\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}-\sqrt{N}\hat{\sigma}^2\mathbf{P}'\mathbf{Q}_{t-1}\\ \mathbf{0}_{K_\mathrm{z}}\end{bmatrix}+\begin{bmatrix}0\\ \mathbf{0}_{K_\mathrm{z}}\\ \frac{\mathbf{Z}_{t-1}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\end{bmatrix}\right)\\ &= \begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-\hat{\mathbf{\Lambda}}_1 & \frac{\hat{\mathbf{X}}'\mathbf{Z}_{t-1}}{N}\\ \frac{\mathbf{Z}_{t-1}'\hat{\mathbf{X}}}{N} & \frac{\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}}{N}\end{bmatrix}^{-1}(I_1+I_2+I_3)\,.\end{aligned} \tag{OA.39}$$

Now, by Lemmas 1(iv) and 2, we have that

$$\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}}-N\hat{\mathbf{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z}_{t-1}\\ \mathbf{Z}_{t-1}'\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}\end{bmatrix}\to_p\mathbf{L}_{t-1}. \tag{OA.40}$$

Regarding the variances of the terms I_1 and I_2, under Assumption OA.6, we get

$$\mathrm{Var}\left(\begin{bmatrix}\frac{\mathbf{1}_N'\epsilon'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\end{bmatrix}\right)\to_p\sigma^2\mathbf{Q}_{t-1}'\mathbf{Q}_{t-1}\mathbf{\Sigma}_\mathrm{X},\quad\text{and} \tag{OA.41}$$

$$\mathrm{Var}\left(\mathbf{P}'\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}-\sqrt{N}\hat{\sigma}^2\mathbf{P}'\mathbf{Q}_{t-1}\right)\to_p \mathbf{V}'_{t-1}\mathbf{U}_{\epsilon}\mathbf{V}_{t-1}. \tag{OA.42}$$

Notice also that, under Assumption OA.5, the two terms I_1 and I_2 are uncorrelated. Consider now the term I_3, and notice that $\mathbf{z}_{i,t}\equiv\mathbb{J}'_t\mathbf{Z}'\boldsymbol{\imath}_{i,N}$. Then, using the properties of the vec$(\cdot)$ operator, recalling that $\mathbf{u}_i\equiv\boldsymbol{\epsilon}_i\otimes\mathbf{z}_i$ and $\mathbf{H}_{t-1}\equiv\mathbf{Q}'_{t-1}\otimes\mathbb{J}'_{t-1}$, it follows that

$$\begin{aligned}
\frac{\mathbf{Z}'_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{z}_{i,t-1}\boldsymbol{\epsilon}'_i\mathbf{Q}_{t-1}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{J}'_{t-1}\mathbf{Z}'\boldsymbol{\imath}_{i,N}\boldsymbol{\epsilon}'_i\mathbf{Q}_{t-1}\\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{Q}'_{t-1}\otimes\mathbb{J}'_{t-1}\right)\mathrm{vec}\left(\mathbf{Z}'\boldsymbol{\imath}_{i,N}\boldsymbol{\epsilon}'_i\right)\\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{Q}'_{t-1}\otimes\mathbb{J}'_{t-1}\right)\left(\boldsymbol{\epsilon}_i\otimes\mathbf{z}_i\right)\\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{H}_{t-1}\mathbf{u}_i.
\end{aligned} \tag{OA.43}$$

Therefore, using (OA.43), and under Assumption OA.7(v), one obtains

$$\mathrm{Var}\left(\frac{\mathbf{Z}'_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\right)=\mathbf{H}_{t-1}\frac{1}{N}\sum_{i,j=1}^{N}\boldsymbol{\Sigma}_{\mathrm{u,ij}}\mathbf{H}'_{t-1}\to\mathbf{H}_{t-1}\boldsymbol{\Sigma}_{\mathrm{U}}\mathbf{H}'_{t-1} \tag{OA.44}$$

Finally, let us consider the covariance terms of I_3 with both I_1 and I_2. To derive the covariance between I_3 and I_2, we establish

$$\begin{aligned}
&\mathrm{Cov}\left(\mathbf{P}'\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}-\sqrt{N}\hat{\sigma}^2\mathbf{P}'\mathbf{Q}_{t-1},\mathbf{Q}'_{t-1}\frac{\boldsymbol{\epsilon}\mathbf{Z}_{t-1}}{\sqrt{N}}\right)=\\
=\ &\mathrm{E}\left[\left((\mathbf{Q}_{t-1}\otimes\mathbf{P})-\frac{\mathrm{vec}(\mathbf{M}_{\tilde{D}})}{T-2-K}\mathbf{Q}'_{t-1}\mathbf{P}\right)'\frac{1}{N}\sum_{i,j=1}^{N}\mathrm{vec}\left(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_i-\sigma_i^2\mathbf{I}_{T-1}\right)\mathbf{u}'_j\mathbf{H}'_{t-1}\right]+o(1)\\
=\ &\mathbf{V}'_{t-1}\frac{1}{N}\sum_{i,j=1}^{N}\boldsymbol{\Sigma}_{u\epsilon,ij}\mathbf{H}'_{t-1}\to\mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{z}}}
\end{aligned}$$

by Assumption OA.7(viii). Finally, to derive the Cov(I_1,I'_3), we need to calculate

$$\begin{aligned}
\mathrm{Cov}\left(\begin{bmatrix}\frac{\mathbf{1}'_N\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\end{bmatrix},\mathbf{Q}'_{t-1}\frac{\boldsymbol{\epsilon}\mathbf{Z}_{t-1}}{\sqrt{N}}\right) &= \frac{1}{N}\sum_{i,j=1}^{N}\sigma_{ij}\left(\mathbf{Q}'_{t-1}\otimes\mathrm{E}\left[\begin{bmatrix}1\\ \boldsymbol{\beta}_i\end{bmatrix}\mathbf{z}'_j\right]\right)\mathbf{H}'_{t-1}\\
&\to \sigma^2\mathbf{G}_{t-1}\mathbf{H}'_{t-1}.
\end{aligned}$$

by Assumption OA.7 and where we set $\mathbf{G}_{t-1} \equiv \left[\mathbf{Q}_{t-1} \otimes \boldsymbol{\mu}_{\mathrm{z},T-1}, \quad \mathbf{Q}_{t-1} \otimes \boldsymbol{\Sigma}_{\mathrm{ZB}}\right]' \left[\mathbf{Q}_{t-1} \otimes \boldsymbol{\mu}_{\mathrm{z},T-1}, \quad \mathbf{Q}_{t-1} \otimes \boldsymbol{\Sigma}_{\mathrm{ZB}}\right]'$. Therefore, putting all the above results together, and recalling that $\mathbf{U}_{t-1} \equiv \sigma^2 \mathbf{Q}'_{t-1}\mathbf{Q}_{t-1}\boldsymbol{\Sigma}_{\mathrm{X}} + \begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} \\ \mathbf{0}_{K_{\mathrm{f}}} & \mathbf{V}'_{t-1}\mathbf{U}_{\epsilon}\mathbf{V}_{t-1} \end{bmatrix}$, yield $\mathbf{O}_{t-1}$, concluding the proof. ■

Proof of Theorem 2. By Lemma 1 and Lemma 2(iii) and (iv), it follows that $\hat{\mathbf{L}}_{t-1} \to_p \mathbf{L}_{t-1}$. By part (i) of Theorem 1, then $\hat{\boldsymbol{\delta}}^*_{\mathrm{f},t-1} \to_p \boldsymbol{\delta}_{\mathrm{f},t-1}$, implying that $\hat{\mathbf{Q}}_{t-1}$ is a consistent estimator of $\mathbf{Q}_{t-1}$. Moreover, as $N \to \infty$, $\overline{\mathbb{M}}_{\tilde{\mathrm{D}}} \to_p \mathbb{M}_{\tilde{D}}$, $\hat{\boldsymbol{\mu}}_{T-1,z} \to_p \boldsymbol{\mu}_{T-1,z}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ZB}} \to_p \boldsymbol{\Sigma}_{\mathrm{ZB}}$, and $\mathbf{Z}'\mathbf{Z}/N \to_p \boldsymbol{\Sigma}_{\mathrm{Z}}$. It follows that $\hat{\mathbf{V}}_{t-1} \to_p \mathbf{V}_{t-1}$, $\hat{\mathbf{G}}_{t-1} \to_p \mathbf{G}_{t-1}$, and $\hat{\mathbf{H}}_{t-1} \to_p \mathbf{H}_{t-1}$. Finally, a consistent estimator of $\hat{\mathbf{U}}_{t-1}$ requires a consistent estimate of the matrix $\mathbf{U}_{\epsilon}$, which can be obtained using Lemma 1(ix). This concludes the proof of Theorem 2. ■

Proof of Theorem 3. First, let us define the quantities that make the asymptotic covariance matrix of the estimator. Let $\boldsymbol{\mu}_{\mathrm{z}} \equiv \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[\mathbf{z}_{i,t}]$, and $\boldsymbol{\mu}_{\mathrm{x}} \equiv (1, \boldsymbol{\mu}'_{\beta})'$. Set $\lambda_{t-1} \equiv \mathbf{Y}'_{t-1}\boldsymbol{\Sigma}_{\mathrm{V}}\mathbf{Y}_{t-1}$, with $\mathbf{Y}_{t-1} \equiv \mathbf{Q}_{t-1} \otimes \boldsymbol{\imath}_{t-1,T-1}$ and define $\mathbf{S}_{t-1} \equiv \boldsymbol{\mu}_{\mathrm{z}}\mathbf{Y}'_{t-1}(\sigma^2 \boldsymbol{I}_{T-1} \otimes \boldsymbol{\Sigma}'_{\mathrm{ZW}})\mathbf{H}'_{t-1}$.Then, set

$$\mathbf{O}^{(\mathrm{w})}_{t-1} \equiv \lambda_{t-1} \begin{bmatrix} \boldsymbol{\mu}_{\mathrm{x}}\boldsymbol{\mu}'_{\mathrm{x}} & \boldsymbol{\mu}_{\mathrm{x}}\boldsymbol{\mu}'_{\mathrm{z}} \\ \boldsymbol{\mu}_{\mathrm{z}}\boldsymbol{\mu}'_{\mathrm{x}} & \boldsymbol{\mu}_{\mathrm{z}}\boldsymbol{\mu}'_{\mathrm{z}} \end{bmatrix} + \mathbf{M}^{(\mathrm{w})}_{t-1} \tag{OA.45}$$

with

$$\mathbf{M}^{(\mathrm{w})}_{t-1} \equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} & \mathbf{0}'_{K_{\mathrm{z}}} \\ \mathbf{0}_{K_{\mathrm{f}}} & \mathbf{V}'_{t-1}\mathbf{U}_{\epsilon}\mathbf{V}_{t-1} & \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{z}}} \\ \mathbf{0}_{K_{\mathrm{z}}} & \mathbf{0}_{K_{\mathrm{z}}\times K_{\mathrm{f}}} & \mathbf{H}_{t-1}\boldsymbol{\Sigma}_{\mathrm{U}}\mathbf{H}'_{t-1} + \mathbf{S}_{t-1} + \mathbf{S}'_{t-1} \end{bmatrix}$$

where $\mathbf{Q}_{t-1}$, $\mathbf{V}_{t-1}$, and $\mathbf{H}_{t-1}$ are defined in the proof to Theorem 1, $\boldsymbol{\Sigma}_{\mathrm{ZW}}$, $\boldsymbol{\Sigma}_{\mathrm{Z},t-1}$, $\boldsymbol{\Sigma}_{\mathrm{V}}$, and $\boldsymbol{\mu}_{\mathrm{z},t-1}$ are defined in Assumption OA.11, and $\boldsymbol{\Sigma}_{\mathrm{X}}$, $\mathbf{U}_{\epsilon}$, and $\boldsymbol{\Sigma}_{\mathrm{U}}$ are defined in Assumptions OA.3, OA.6, and OA.7.

Next, let us start the proof from the definition in (42) and rewrite

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}^{*(\mathrm{w})}_{\mathrm{f},t-1}\\ \hat{\gamma}^{*(\mathrm{w})}_{\mathrm{z},t-1}\end{bmatrix}-\begin{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1}\\ \gamma_{\mathrm{z},t-1}\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1}\hat{\mathbf{X}}-N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\end{bmatrix}^{-1}\times$$

$$\times\left(\begin{bmatrix}N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1}\\ \mathbf{0}_{\mathrm{K}+1}\end{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1}+\begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1}\\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\end{bmatrix}(\boldsymbol{\epsilon}_t+(\mathbf{X}-\hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1})-\begin{bmatrix}N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{2,t-1}\\ \mathbf{0}_{K_\mathrm{z}}\end{bmatrix}\right). \quad \text{(OA.46)}$$

To shorten the proof, let us establish first part (ii) of the theorem and derive the asymptotic distribution of the above expression, since its $\sqrt{N}$-rate of convergence will then immediately follow.

First notice that, by Assumptions OA.8, OA.9 and OA.11,

$$\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1}\hat{\mathbf{X}}-N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\end{bmatrix}\to_p \mathbf{L}_{t-1},$$

where we set $\boldsymbol{\mu}_\mathrm{X}\equiv\begin{bmatrix}1\\ \boldsymbol{\mu}_\beta\end{bmatrix}$.

Now consider the next term in (OA.46) and notice that:

$$\begin{aligned}
&\hat{\mathbf{X}}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t+\hat{\mathbf{X}}'\mathbf{W}_{t-1}(\mathbf{X}-\hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1}-N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{2,t-1}+N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1}\boldsymbol{\Gamma}_{\mathrm{f},t-1}\\
= \;&\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t+\begin{bmatrix}0\\ \mathbf{P}'\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t-N\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{P}'\boldsymbol{\imath}_{t-1,T-1}\end{bmatrix}\\
- \;&\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{P}\boldsymbol{\delta}_{\mathrm{f},t-1}+\begin{bmatrix}0\\ -\mathbf{P}'\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{P}\boldsymbol{\delta}_{\mathrm{f},t-1}+N\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{P}'\mathbf{P}\boldsymbol{\delta}_{\mathrm{f},t-1}\end{bmatrix}\\
= \;&\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\boldsymbol{\imath}_{t-1,T-1}-\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{P}\boldsymbol{\delta}_{\mathrm{f},t-1}+\begin{bmatrix}0\\ \mathbf{P}'\left(\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'-N\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1}\end{bmatrix}\\
= \;&\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}+\begin{bmatrix}0\\ \mathbf{P}'\left(\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'-N\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1}\end{bmatrix}\\
= \;&\mathbf{a}_{11}+\mathbf{a}_{12},
\end{aligned}$$

setting $\mathbf{a}_{11} \equiv \mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}$ and $\mathbf{a}_{12} \equiv \begin{bmatrix} 0 \\ \mathbf{P}'\left(\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - N\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1}\end{bmatrix}$. Moreover,

$$\begin{aligned}
\mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\boldsymbol{\epsilon}_t + \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}(\mathbf{X}-\hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1} &= \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} \\
&= \left(\mathrm{E}[\mathbf{Z}_{t-1}]'\mathbf{W}_{t-1}\boldsymbol{\epsilon}' + (\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}])'\,\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\right)\mathbf{Q}_{t-1} \\
&= \mathrm{E}[\mathbf{Z}_{t-1}]'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} + (\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}])'\,\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} \\
&= \mathbf{a}_{21} + \mathbf{a}_{22},
\end{aligned}$$

setting $\mathbf{a}_{21} \equiv \mathrm{E}[\mathbf{Z}_{t-1}]'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}$ and $\mathbf{a}_{22} \equiv (\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}])'\,\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}$.

Let us start with the term $\mathbf{a}_{11}$. Using Assumption OA.11, and noticing that $\mathbf{W}_{t-1}\mathbf{1}_N \equiv \mathbf{W}\boldsymbol{\imath}_{t-1,T-1}$, we can write that

$$\begin{aligned}
\frac{1}{\sqrt{N}}\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} &= \frac{1}{\sqrt{N}}\left(\mathbf{X}'\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - \mathbf{X}'\frac{\mathbf{1}_N\mathbf{1}'_N}{N}\mathbf{W}_{t-1}\boldsymbol{\epsilon}' + \mathbf{X}'\frac{\mathbf{1}_N\mathbf{1}'_N}{N}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\right)\mathbf{Q}_{t-1} \\
&= \frac{1}{\sqrt{N}}\left(\frac{\mathbf{X}'\mathbf{1}_N\mathbf{1}'_N\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{N}\right)\mathbf{Q}_{t-1} + o_p(1) \\
&= \frac{\mathbf{X}'\mathbf{1}_N}{N}\frac{\mathbf{1}'_N\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}}{\sqrt{N}} + o_p(1) \\
&= \frac{\mathbf{X}'\mathbf{1}_N}{N}\frac{\boldsymbol{\imath}'_{t-1,T-1}\mathbf{W}'\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}}{\sqrt{N}} + o_p(1) \\
&= \frac{\mathbf{X}'\mathbf{1}_N}{N}\left(\mathbf{Q}'_{t-1}\otimes\boldsymbol{\imath}'_{t-1,T-1}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i\otimes\mathrm{w}_i + o_p(1) \\
&\to_d \mathcal{N}\left(\mathbf{0}_{K_{\mathrm{f}}+1}, \lambda_{t-1}\boldsymbol{\mu}_{\mathrm{X}}\boldsymbol{\mu}'_{\mathrm{X}}\right),
\end{aligned}$$

recalling $\boldsymbol{\mu}_{\mathrm{X}} = (1, \boldsymbol{\mu}'_{\beta})'$, $\lambda_{t-1} = \mathbf{Y}'_{t-1}\boldsymbol{\Sigma}_{\mathrm{V}}\mathbf{Y}_{t-1}$, with $\mathbf{Y}_{t-1} = (\mathbf{Q}_{t-1}\otimes\boldsymbol{\imath}_{t-1,T-1})$ and $\boldsymbol{\Sigma}_{\mathrm{V}}$ defined in Assumption OA.11.

Consider now the term in $\mathbf{a}_{12}$ and notice that, by Assumptions OA.10 and OA.11, and following the same steps adopted for (OA.42),

$$
\begin{aligned}
& \frac{1}{\sqrt{N}}\mathbf{P}'\left(\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - N\hat{\sigma}_{t-1}^{2(\mathrm{w})}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1} \\
= \quad & \sqrt{N}\mathbf{P}'\left(\frac{\boldsymbol{\epsilon}\left(\mathbf{W}_{t-1}-\mathrm{E}[\mathbf{W}_{t-1}]\right)\boldsymbol{\epsilon}'}{N} + \frac{\boldsymbol{\epsilon}\mathrm{E}[\mathbf{W}_{t-1}]\boldsymbol{\epsilon}'}{N} - \hat{\sigma}_{t-1}^{2(\mathrm{w})}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1} \\
= \quad & \sqrt{N}\mathbf{P}'\left(\frac{\boldsymbol{\epsilon}\mathrm{E}[\mathbf{W}_{t-1}]\boldsymbol{\epsilon}'}{N} - \hat{\sigma}_{t-1}^{2(\mathrm{w})}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1} + o_p(1) \\
= \quad & \sqrt{N}\left(\mathbf{Q}_{t-1}'\otimes\mathbf{P}'\right)\mathrm{vec}\left(\frac{\boldsymbol{\epsilon}\mathrm{E}[\mathbf{W}_{t-1}]\boldsymbol{\epsilon}'}{N} - \hat{\sigma}_{t-1}^{2(\mathrm{w})}\mathbf{I}_{T-1}\right) + o_p(1) \\
= \quad & \mathbf{V}_{t-1}'\mu_{\mathrm{w},t-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathrm{vec}\left(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}\right) + o_p(1) \\
\to_d \quad & \mathcal{N}\left(\mathbf{0}_{K_{\mathrm{f}}}, \mathbf{V}_{t-1}'\mathbf{U}_{\epsilon}\mathbf{V}_{t-1}\right)
\end{aligned}
$$

where $\mu_{\mathrm{w},t-1} \to_p 1$ by Assumption OA.8, with $\mathbf{U}_\epsilon$ defined in Assumption OA.6.

For the term $\mathbf{a}_{21}$, using Assumption OA.6 and following the same steps for $\mathbf{a}_{11}$, we get

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\mathrm{E}[\mathbf{Z}_{t-1}]'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} \quad = \quad & \frac{1}{\sqrt{N}}\left(\mathrm{E}[\mathbf{Z}_{t-1}]'\left(\mathbf{I}_N - \frac{\mathbf{1}_N\mathbf{1}_N'}{N}\right)\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} + \mathrm{E}[\mathbf{Z}_{t-1}]'\frac{\mathbf{1}_N\mathbf{1}_N'}{N}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}\right) \\
= \quad & \frac{1}{\sqrt{N}}\left(\mathrm{E}[\mathbf{Z}_{t-1}]'\frac{\mathbf{1}_N\mathbf{1}_N'}{N}\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}\right) + o_p(1) \\
= \quad & \mathrm{E}[\mathbf{Z}_{t-1}]'\frac{\mathbf{1}_N}{N}\ \frac{1}{\sqrt{N}}\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1} + o_p(1) \\
= \quad & \mathrm{E}[\mathbf{Z}_{t-1}]'\frac{\mathbf{1}_N}{N}\ \left(\mathbf{Q}_{t-1}'\otimes\boldsymbol{\imath}_{t-1,T-1}'\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i\otimes\mathrm{w}_i + o_p(1) \\
\to_d \quad & \mathcal{N}\left(\mathbf{0}_{K_{\mathrm{z}}}, \lambda_{t-1}\boldsymbol{\mu}_{\mathrm{z}}\boldsymbol{\mu}_{\mathrm{z}}'\right).
\end{aligned}
$$

Finally, for the term $\mathbf{a}_{22}$, the following holds

$$
\begin{aligned}
& \sqrt{N}\left(\frac{(\mathbf{Z}_{t-1}-\mathrm{E}[\mathbf{Z}_{t-1})]'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{N}\right)\mathbf{Q}_{t-1} = \\
& \sqrt{N}\left(\frac{(\mathbf{Z}_{t-1}-\mathrm{E}[\mathbf{Z}_{t-1}])'\boldsymbol{\epsilon}'}{N}\right)\mathbf{Q}_{t-1} + o_p(1) \\
& \to_d \mathcal{N}(\mathbf{0}_{K_{\mathrm{z}}}, \mathbf{H}_{t-1}\boldsymbol{\Sigma}_{\mathrm{U}}\mathbf{H}_{t-1}').
\end{aligned}
$$

We can now derive the asymptotic distribution of the estimator. Indeed, using all the results derived above, we have

$$\frac{1}{\sqrt{N}}\left(\begin{bmatrix} N\hat{\mathbf{\Lambda}}_{1,t-1}^{(\mathrm{w})} \\ \mathbf{0}_{\mathrm{K}+1}\end{bmatrix}\mathbf{\Gamma}_{\mathrm{f},t-1} + \begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1} \\ \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\end{bmatrix}(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\mathbf{\Gamma}_{\mathrm{f},t-1}) - \begin{bmatrix} N\hat{\mathbf{\Lambda}}_{2,t-1}^{(\mathrm{w})} \\ \mathbf{0}_{K_\mathrm{z}}\end{bmatrix}\right)$$

$$\begin{aligned} = & \begin{bmatrix} \frac{\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \\ \left(\frac{\mathbf{B}'\mathbf{1}_N}{N}\right)\frac{\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \\ \mathbf{0}_{K_\mathrm{z}} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{P}'\frac{1}{\sqrt{N}}\left(\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - N\hat{\sigma}_{t-1}^{2(\mathrm{w})}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1} \\ \mathbf{0}_{K_\mathrm{z}} \end{bmatrix} \\ + & \begin{bmatrix} 0 \\ \mathbf{0}_{K_\mathrm{f}} \\ \left(\frac{\mathrm{E}[\mathbf{Z}_{t-1}]'\mathbf{1}_N}{N}\right)\left(\frac{\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\right)\mathbf{Q}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{0}_{K_\mathrm{f}} \\ \frac{(\mathbf{Z}_{t-1}-\mathrm{E}[\mathbf{Z}_{t-1}])'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \end{bmatrix} \\ = & \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 + \mathbf{a}_4, \end{aligned}$$

setting $\mathbf{a}_1 \equiv \begin{bmatrix} \frac{\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \\ \left(\frac{\mathbf{B}'\mathbf{1}_N}{N}\right)\frac{\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \\ \mathbf{0}_{K_\mathrm{z}} \end{bmatrix}$, $\mathbf{a}_2 \equiv \begin{bmatrix} 0 \\ \mathbf{P}'\frac{1}{\sqrt{N}}\left(\boldsymbol{\epsilon}\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - N\hat{\sigma}_{t-1}^{2(\mathrm{w})}\mathbf{I}_{T-1}\right)\mathbf{Q}_{t-1} \\ \mathbf{0}_{K_\mathrm{z}} \end{bmatrix}$, $\mathbf{a}_3 \equiv \begin{bmatrix} 0 \\ \mathbf{0}_{K_\mathrm{f}} \\ \left(\frac{\mathrm{E}[\mathbf{Z}_{t-1}]'\mathbf{1}_N}{N}\right)\left(\frac{\mathbf{1}_N'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\right)\mathbf{Q}_{t-1} \end{bmatrix}$, and $\mathbf{a}_4 \equiv \begin{bmatrix} 0 \\ \mathbf{0}_{K_\mathrm{f}} \\ \frac{(\mathbf{Z}_{t-1}-\mathrm{E}[\mathbf{Z}_{t-1}])'\mathbf{W}_{t-1}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1} \end{bmatrix}$.

Notice that all the terms have zero mean. Therefore, using the results above:

$$\mathrm{Var}[\mathbf{a}_1] \to \lambda_{t-1}\begin{bmatrix} \boldsymbol{\mu}_\mathrm{X}\boldsymbol{\mu}_\mathrm{X}' & \mathbf{0}_{(K_\mathrm{f}+1)\times K_\mathrm{z}} \\ \mathbf{0}_{K_\mathrm{z}\times(K_\mathrm{f}+1)} & \mathbf{0}_{K_\mathrm{z}\times K_\mathrm{z}} \end{bmatrix}, \quad \mathrm{Var}[\mathbf{a}_2] \to \begin{bmatrix} 0 & \mathbf{0}_{K_\mathrm{f}}' & \mathbf{0}_{K_\mathrm{z}}' \\ \mathbf{0}_{K_\mathrm{f}} & \mathbf{V}_{t-1}'\mathbf{U}_\epsilon\mathbf{V}_{t-1} & \mathbf{0}_{K_\mathrm{f}\times K_\mathrm{z}} \\ \mathbf{0}_{K_\mathrm{z}} & \mathbf{0}_{K_\mathrm{z}\times K_\mathrm{f}} & \mathbf{0}_{K_\mathrm{z}\times K_\mathrm{z}} \end{bmatrix},$$

$$\mathrm{Var}[\mathbf{a}_3] \to \lambda_{t-1}\begin{bmatrix} 0 & \mathbf{0}_{K_\mathrm{f}}' & \mathbf{0}_{K_\mathrm{z}}' \\ \mathbf{0}_{K_\mathrm{f}} & \mathbf{0}_{K_\mathrm{f}\times K_\mathrm{f}} & \mathbf{0}_{K_\mathrm{f}\times K_\mathrm{z}} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \boldsymbol{\mu}_\mathrm{z}\boldsymbol{\mu}_\mathrm{z}' \end{bmatrix}, \quad \text{and} \quad \mathrm{Var}[\mathbf{a}_4] \to \begin{bmatrix} 0 & \mathbf{0}_{K_\mathrm{f}}' & \mathbf{0}_{K_\mathrm{z}}' \\ \mathbf{0}_{K_\mathrm{f}} & \mathbf{0}_{K_\mathrm{f}\times K_\mathrm{f}} & \mathbf{0}_{K_\mathrm{f}\times K_\mathrm{z}} \\ \mathbf{0}_{K_\mathrm{z}} & \mathbf{0}_{K_\mathrm{z}\times K_\mathrm{f}} & \mathbf{H}_{t-1}\mathbf{\Sigma}_\mathrm{U}\mathbf{H}_{t-1}' \end{bmatrix}.$$

It remains to evaluate the covariances terms. Under Assumption OA.5, then $\mathrm{Cov}[\mathbf{a}_1, \mathbf{a}_2'] \to \mathbf{0}_{(K+1)\times(K+1)}$, and $\mathrm{Cov}[\mathbf{a}_1, \mathbf{a}_4'] \to \mathbf{0}_{(K+1)\times(K+1)}$, while

$$\mathrm{Cov}[\mathbf{a}_1, \mathbf{a}_3'] \to \lambda_{t-1}\begin{bmatrix} \mathbf{0}_{(K_\mathrm{f}+1)\times(K_\mathrm{f}+1)} & \boldsymbol{\mu}_\mathrm{x}\boldsymbol{\mu}_z' \\ \mathbf{0}_{K_\mathrm{z}\times(K_\mathrm{f}+1)} & \mathbf{0}_{K_\mathrm{z}\times K_\mathrm{z}} \end{bmatrix}$$

Regarding $\mathbf{a}_2$, notice that, Assumption OA.5, $\text{Cov}[\mathbf{a}_2, \mathbf{a}_3'] \to \mathbf{0}_{(K+1)\times(K+1)}$ and $\text{Cov}(\mathbf{a}_2, \mathbf{a}_4') \to \mathbf{0}_{(K+1)\times(K+1)}$. Finally,

$$\text{Cov}[\mathbf{a}_3, \mathbf{a}_4'] \to \begin{bmatrix} 0 & \mathbf{0}_{K_\text{f}}' & \mathbf{0}_{K_\text{z}}' \\ \mathbf{0}_{K_\text{f}} & \mathbf{0}_{K_\text{f}\times K_\text{f}} & \mathbf{0}_{K_\text{f}\times K_\text{z}} \\ \mathbf{0}_{K_\text{z}} & \mathbf{0}_{K_\text{z}\times K_\text{f}} & \mathbf{S}_{t-1} \end{bmatrix}$$

setting $\mathbf{S}_{t-1} \equiv \boldsymbol{\mu}_\text{z}\mathbf{Y}_{t-1}'(\sigma^2\mathbf{I}_{T-1} \otimes \boldsymbol{\Sigma}_{\text{ZW}}')\mathbf{H}_{t-1}'$. In fact,

$$\begin{aligned} \text{Cov}[\mathbf{a}_3, \mathbf{a}_4'] \quad &= \text{E}\left(\boldsymbol{\mu}_\text{z}\boldsymbol{\imath}_{t-1,T-1}'\frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{i.}\boldsymbol{\epsilon}_i'\mathbf{Q}_{t-1}\frac{1}{N}\sum_{j=1}^{N}w_{j,t-1}vec'(\mathbf{z}_j\boldsymbol{\epsilon}_j')(\mathbf{Q}_{t-1}\otimes\mathbb{J}_{t-1})\right) \\ &= \boldsymbol{\mu}_\text{z}(\mathbf{Q}_{t-1}'\otimes\boldsymbol{\imath}_{t-1,T-1}')\text{E}\left(\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_i\otimes\mathbf{w}_{i.})\frac{1}{N}\sum_{j=1}^{N}w_{j,t-1}(\boldsymbol{\epsilon}_j'\otimes\mathbf{z}_j')\right)(\mathbf{Q}_{t-1}\otimes\mathbb{J}_{t-1}) \\ &= \boldsymbol{\mu}_\text{z}(\mathbf{Q}_{t-1}'\otimes\boldsymbol{\imath}_{t-1,T-1}')\text{E}\left(\frac{1}{N}\sum_{i=1}^{N}w_{i,t-1}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'\otimes\mathbf{w}_{i.}\mathbf{z}_j')\right)(\mathbf{Q}_{t-1}\otimes\mathbb{J}_{t-1}) + o(1). \end{aligned}$$

Collecting terms concludes the proof.

■

Proof of Theorem 5. Rewrite the premia estimator as

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}_{\text{f},t-1}^{*(\text{m})} \\ \hat{\gamma}_{\text{z},t-1}^{*(\text{m})}\end{bmatrix} =$$

$$\begin{bmatrix}\tilde{\boldsymbol{\Gamma}}_{\text{f},t-1} \\ \tilde{\gamma}_{\text{z},t-1}\end{bmatrix} + \begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}} - N(\hat{\boldsymbol{\Lambda}}_1 + \hat{\boldsymbol{\Lambda}}_{1,t-1}^{(\text{m})}) & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} - N\hat{\boldsymbol{\Lambda}}_{3,t-1}^{(\text{m})} \\ \mathbf{Z}_{t-1}'\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}\end{bmatrix}^{-1}$$

$$\times\left[\begin{bmatrix}N(\hat{\boldsymbol{\Lambda}}_1 + \hat{\boldsymbol{\Lambda}}_{1,t-1}^{(\text{m})}) \\ \mathbf{0}_{K_z\times K_f+1}\end{bmatrix}\tilde{\boldsymbol{\Gamma}}_{\text{f},t-1} + \begin{bmatrix}N\hat{\boldsymbol{\Lambda}}_{3,t-1}^{(\text{m})} \\ \mathbf{0}_{K_z\times K_z}\end{bmatrix}\tilde{\gamma}_{\text{z},t-1} + \begin{bmatrix}\hat{\mathbf{X}}' \\ \mathbf{Z}_{t-1}'\end{bmatrix}(\mathbf{m}_{t-1} + \boldsymbol{\epsilon}_t + (\mathbf{X} - \hat{\mathbf{X}})\tilde{\boldsymbol{\Gamma}}_{\text{f},t-1}) - \begin{bmatrix}N(\hat{\boldsymbol{\Lambda}}_{2,t-1} + \hat{\boldsymbol{\Lambda}}_{2,t-1}^{(\text{m})}) \\ \mathbf{0}_{K_z}\end{bmatrix}\right]$$

Concerning the bias terms, the proof follows the corresponding part of the proof to Theorem 1, except for the additional terms arising from $\frac{1}{N}\hat{\mathbf{X}}'\mathbf{m}_{t-1} - \hat{\boldsymbol{\Lambda}}_{2,t-1}^{(\text{m})} + \hat{\boldsymbol{\Lambda}}_{1,t-1}^{(\text{m})}\tilde{\boldsymbol{\Gamma}}_{\text{f},t-1} + \hat{\boldsymbol{\Lambda}}_{3,t-1}^{(\text{m})}\tilde{\gamma}_{\text{z},t-1}$, equal to (excluding the zero elements of the first row)

$$
\begin{aligned}
&\frac{1}{N}(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{m}_{t-1} - \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\mathbf{R}_t\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix} + \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\hat{\mathbf{X}}\Big)\\ \mathbf{0}_{T-t+1\times K_f+1}\end{bmatrix}\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}\\
&+\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\mathbf{Z}_{t-1}\Big)\\ \mathbf{0}_{T-t+1\times K_z}\end{bmatrix}\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} + \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big(\hat{\sigma}^2(\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\mathbb{M}_{\mathrm{D}}\boldsymbol{\iota}_{t-1,T-1}\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix}\\
&=\frac{1}{N}(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{m}_{t-1} - \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\mathbf{m}_{t-1}\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix}\\
&+\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\hat{\mathbf{X}}\Big)\\ \mathbf{0}_{T-t+1\times K_f+1}\end{bmatrix}\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} - \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\mathbf{X}\Big)\\ \mathbf{0}_{T-t+1\times K_f+1}\end{bmatrix}\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}\\
&+\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\mathbf{Z}_{t-1}\Big)\\ \mathbf{0}_{T-t+1\times K_z}\end{bmatrix}\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} - \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\mathbf{Z}_{t-1}\Big)\\ \mathbf{0}_{T-t+1\times K_z}\end{bmatrix}\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1}\\
&+\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big(\hat{\sigma}^2(\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\mathbb{M}_{\mathrm{D}}\boldsymbol{\iota}_{t-1,T-1}\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix} - \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\frac{1}{N}\hat{\boldsymbol{\epsilon}}\boldsymbol{\epsilon}_t\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix}\\
&=-\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}\\ \mathbf{I}_{T-t+1}\end{bmatrix}\frac{1}{N}\boldsymbol{\epsilon}_{t,T-1}\mathbf{m}_{t-1} + \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})(\frac{1}{N}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i')\mathbf{P}\Big)\\ \mathbf{0}_{T-t+1\times K_f}\end{bmatrix}\tilde{\boldsymbol{\delta}}_{\mathrm{f},t-1}\\
&+\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})(\hat{\sigma}^2\mathbb{M}_{\mathrm{D}} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i')\boldsymbol{\iota}_{t-1,T-1}\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix}\\
&=-\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}\\ \mathbf{I}_{T-t+1}\end{bmatrix}\frac{1}{N}\boldsymbol{\epsilon}_{t,T-1}\mathbf{m}_{t-1} - \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})(\frac{1}{N}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \hat{\sigma}^2\mathbb{M}_{\mathrm{D}})\mathbf{Q}_{t-1,T-1}\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix}\\
&=O_p(N^{-\frac{1}{2}}),
\end{aligned}
$$

recalling that $(\hat{\mathbf{B}}-\mathbf{B})'\mathbf{m}_{t-1} = \hat{\mathbf{B}}'\mathbf{m}_{t-1}$, $N^{-1}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i m_{i,t-1} = \mathbb{M}_D(N^{-1}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i m_{i,t-1}) + o_p(1)$, and $\boldsymbol{\epsilon} = \begin{bmatrix}\boldsymbol{\epsilon}_{2,t-1}\\ \boldsymbol{\epsilon}_{t,T-1}\end{bmatrix}$, with $\boldsymbol{\epsilon}_{2,t-1} \equiv (\boldsymbol{\epsilon}_2,\ldots,\boldsymbol{\epsilon}_{t-1})', \boldsymbol{\epsilon}_{t,T-1} \equiv (\boldsymbol{\epsilon}_t,\ldots,\boldsymbol{\epsilon}_{T-1})'$. Notice that the bias adjustment corresponding to the anomalies $\mathbf{Z}_{t-1}$ cancels out for every finite N.

Regarding part (ii), the limiting distribution of $\hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}$ and $\hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1}$, and their (joint) asymptotic covariance matrix, follow by

$$\sqrt{N}\begin{bmatrix}\hat{\mathbf{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}-\tilde{\mathbf{\Gamma}}_{\mathrm{f},t-1}\\ \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1}-\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1}\end{bmatrix}=\begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-(\hat{\mathbf{\Lambda}}_1+\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{1,t-1}) & \frac{\hat{\mathbf{X}}'\mathbf{Z}_{t-1}}{N}-\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{3,t-1}\\ \frac{\mathbf{Z}'_{t-1}\hat{\mathbf{X}}}{N} & \frac{\mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}}{N}\end{bmatrix}^{-1}\times\left(\begin{bmatrix}\frac{\mathbf{1}'_N\epsilon'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}\\ \mathbf{0}_{K_z}\end{bmatrix}\right.$$

$$+\begin{bmatrix}0\\ \mathbf{P}'\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}_{t-1}-\sqrt{N}\hat{\sigma}^2\mathbf{P}'\mathbf{Q}_{t-1}\\ \mathbf{0}_{K_z}\end{bmatrix}+\begin{bmatrix}0\\ \mathbf{0}_{K_f}\\ \sqrt{N}\left(\frac{\mathbf{Z}'_{t-1}\boldsymbol{\epsilon}'}{N}\right)\mathbf{Q}_{t-1}\end{bmatrix}$$

$$\left.+\begin{bmatrix}0\\ -\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}\\ \mathbf{I}_{T-t+1}\end{bmatrix}\frac{1}{\sqrt{N}}\boldsymbol{\epsilon}_{t,T-1}\mathbf{m}_{t-1}\\ \mathbf{0}_{K_z}\end{bmatrix}+\begin{bmatrix}0\\ -\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\Big((\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\sqrt{N}(\frac{1}{N}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_i-\hat{\sigma}^2\mathbf{M}_D)\mathbf{Q}_{t-1,T-1}\Big)\\ \mathbf{0}_{T-t+1}\end{bmatrix}\\ \mathbf{0}_{K_z}\end{bmatrix}\right)$$

$$=\begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-(\hat{\mathbf{\Lambda}}_1+\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{1,t-1}) & \frac{\hat{\mathbf{X}}'\mathbf{Z}_{t-1}}{N}-\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{3,t-1}\\ \frac{\mathbf{Z}'_{t-1}\hat{\mathbf{X}}}{N} & \frac{\mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}}{N}\end{bmatrix}^{-1}(\mathbf{a}_1+\mathbf{a}_2+\mathbf{a}_3+\mathbf{a}_4+\mathbf{a}_5)\,.$$

We follow the corresponding part of the proof to Theorem 1, and use their results, to which we need to add the derivation of $\mathrm{Var}[\mathbf{a}_4],\mathrm{Var}[\mathbf{a}_5]$, and of the covariances of $\mathbf{a}_4$ and $\mathbf{a}_5$ with $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3$, obtaining

$$
\begin{aligned}
\mathrm{Var}[\mathbf{a}_4] &\equiv \sigma^2\sigma_{t-1mm}\begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & \mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\mathbf{M}_{12} \\ \mathbf{I}_{T-t+1}\end{bmatrix}\begin{bmatrix}\mathbf{M}'_{12}\mathbf{M}_{11}^{-1} & \mathbf{I}_{T-t+1}\end{bmatrix}\mathbf{P} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix}, \\
\mathrm{Var}[\mathbf{a}_5] &\equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & \mathbf{A}_{t-1}\mathbf{U}_\epsilon\mathbf{A}'_{t-1} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix}, \\
\mathrm{Cov}[\mathbf{a}_4,\mathbf{a}'_1] &\equiv \mathbf{0}_{K_z+K_f+1\times K_z+K_f+1}, \\
\mathrm{Cov}[\mathbf{a}_4,\mathbf{a}'_2] &\equiv
\end{aligned}
$$

$$
\begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & -\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\mathbf{M}_{12} \\ \mathbf{I}_{T-t+1}\end{bmatrix}\begin{bmatrix}\mathbf{0}_{T-t+1\times(t-2)^2} & (\boldsymbol{\theta}'_{t-1,m}\otimes\sigma^2\mathbf{I}_{T-1}) & (\sigma^2\mathbf{I}_{T-1}\otimes\boldsymbol{\theta}'_{t-1,m}) & \mathbf{0}_{T-t+1\times(T-t+1)^2}\end{bmatrix}\mathbf{V}_{t-1} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix},
$$

$$
\begin{aligned}
\mathrm{Cov}[\mathbf{a}_4,\mathbf{a}'_3] &\to \mathbf{0}_{K_z+K_f+1\times K_z+K_f+1}, \\
\mathrm{Cov}[\mathbf{a}_4,\mathbf{a}'_5] &=
\end{aligned}
$$

$$
\begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & -\mathbf{P}'\begin{bmatrix}\mathbf{M}_{11}^{-1}\mathbf{M}_{12} \\ \mathbf{I}_{T-t+1}\end{bmatrix}\begin{bmatrix}\mathbf{0}_{T-t+1\times(t-2)^2} & (\boldsymbol{\theta}'_{t-1,m}\otimes\sigma^2\mathbf{I}_{T-1}) & (\sigma^2\mathbf{I}_{T-1}\otimes\boldsymbol{\theta}'_{t-1,m}) & \mathbf{0}_{T-t+1\times(T-t+1)^2}\end{bmatrix}\mathbf{A}_{t-1'} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix},
$$

$$
\begin{aligned}
\mathrm{Cov}[\mathbf{a}_5,\mathbf{a}'_1] &\equiv \mathbf{0}_{K_z+K_f+1\times K_z+K_f+1}, \\
\mathrm{Cov}[\mathbf{a}_5,\mathbf{a}'_2] &\equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & \mathbf{A}_{t-1}\mathbf{U}_\epsilon\mathbf{V}_{t-1} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix} \text{ and} \\
\mathrm{Cov}[\mathbf{a}_5,\mathbf{a}'_3] &\to_p \mathbf{0}_{K_f+K_z+1\times K_f+K_z+1},
\end{aligned}
$$

recalling that $\mathbf{m}'_{t-1}\mathbf{1}_N = 0$ by construction and $\frac{\mathbf{m}'_{t-1}\mathbf{m}_{t-1}}{N} \to_p \sigma_{t-1,mm}$ by our assumptions, and we set

$$
\mathbf{A}_{t-1} \equiv -\mathbf{P}'\begin{bmatrix}(\mathbf{Q}'_{t-1,T-1}\otimes\mathbf{M}_{11}^{-1}(\mathbf{I}_{t-2},\mathbf{0}_{t-2\times T-t+1})\mathbf{M}_D)(\mathbf{I}_{(T-1)^2} - vec(\mathbf{I}_{T-1})\frac{vec'(\mathbf{M}_{\tilde{D}})}{T-K_f-K_z2}) \\ \mathbf{0}_{T-t+1\times(T-1)^2}\end{bmatrix},
$$

and where for $\mathrm{Cov}[I_4, I'_3]$ we used the result $N^{-1}\sum_{i=1}^N m_{i,t-1}\boldsymbol{\epsilon}_{t,T-1}\boldsymbol{\epsilon}'\mathbf{Q}_{t-1}\mathbf{z}'_{i,t-1} \to_p \mathbf{0}_{T-t+1\times K_z}$. Collecting terms, one finally obtains

$$
\begin{aligned}
\mathbf{M}^{(\mathrm{m})}_{t-1} \equiv \quad & \mathrm{Var}[\mathbf{a}_1] + \mathrm{Var}[\mathbf{a}_2] + \mathrm{Var}[\mathbf{a}_3] + \mathrm{Var}[\mathbf{a}_4] + \mathrm{Var}[\mathbf{a}_5] + \qquad \text{(OA.47)} \\
& \mathrm{Cov}[\mathbf{a}_3,\mathbf{a}'_1] + \mathrm{Cov}[\mathbf{a}_1,\mathbf{a}'_3] + \mathrm{Cov}[\mathbf{a}_4,\mathbf{a}'_2] + \mathrm{Cov}[\mathbf{a}_2,\mathbf{a}'_4] + \mathrm{Cov}[\mathbf{a}_5,\mathbf{a}'_2] + \mathrm{Cov}[\mathbf{a}_2,\mathbf{a}'_5] + \mathrm{Cov}[\mathbf{a}_4,\mathbf{a}'_5] + \mathrm{Cov}[\mathbf{a}_5,\mathbf{a}'_4].
\end{aligned}
$$

The asymptotic covariance matrix for $\hat{\mathbf{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1}$ and $\hat{\gamma}^{*(\mathrm{m})}_{\mathrm{z},t-1}$ will then be $\mathbf{L}^{-1}_{t-1}(\mathbf{O}_{t-1}+\mathbf{M}^{(\mathrm{m})}_{t-1})\mathbf{L}^{-1}_{t-1}$, with $\mathbf{L}_{t-1}$ and $\mathbf{O}_{t-1}$ defined in (OA.35) and, given $N^{-1}\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{3,t-1}=o_p(1)$,

$$\hat{\mathbf{L}}^{(\mathrm{m})}_{t-1}=\begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-(\hat{\mathbf{\Lambda}}_1+\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{1,t-1}) & \frac{\hat{\mathbf{X}}'\mathbf{Z}_{t-1}}{N}-\hat{\mathbf{\Lambda}}^{(\mathrm{m})}_{3,t-1}\\ \frac{\mathbf{Z}'_{t-1}\hat{\mathbf{X}}}{N} & \frac{\mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}}{N}\end{bmatrix}\rightarrow_p \mathbf{L}_{t-1}. \tag{OA.48}$$

■

Proof of Theorem 6 We need to establish the limiting distribution of (61) under (i) and (ii). First, notice that, by using the definition in (60), and replacing it into (61), we get

$$\mathcal{T}^2_{\mathrm{z},t-1}\equiv N\left(\frac{\hat{\boldsymbol{\gamma}}^{*\prime}_{\mathrm{z},t-1}\mathbf{Z}'_{t-1}\mathbb{M}_{\hat{X}}\mathbf{Z}_{t-1}\hat{\boldsymbol{\gamma}}^{*}_{\mathrm{z},t-1}}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t}\right). \tag{OA.49}$$

Let us start from the denominator of (OA.49) and rewrite it as $\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t=\mathbf{R}'_t\mathbf{R}_t-\mathbf{R}'_t\mathbf{1}_N\mathbf{1}'_N\mathbf{R}_t/N$. Recalling that $\mathbf{R}_t=\gamma_{0,t-1}\mathbf{1}_N+\mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1}+\mathbf{B}\boldsymbol{\delta}_{\mathrm{f},t-1}+\boldsymbol{\epsilon}_t$, then:

$$\begin{aligned}\frac{\mathbf{R}'_t\mathbf{1}_N}{N} &= \gamma_{0,t-1}+\boldsymbol{\delta}'_{\mathrm{f},t-1}\frac{\mathbf{B}'\mathbf{1}_N}{N}+\boldsymbol{\gamma}'_{\mathrm{z},t-1}\frac{\mathbf{Z}'_{t-1}\mathbf{1}_N}{N}+\frac{\boldsymbol{\epsilon}_t{}'\mathbf{1}_N}{N}\\ &\rightarrow_p \gamma_{0,t-1}+\boldsymbol{\delta}'_{\mathrm{f},t-1}\boldsymbol{\mu}_\beta+\boldsymbol{\gamma}'_{\mathrm{z},t-1}\boldsymbol{\mu}_z.\end{aligned}$$

Moreover, using the same arguments,

$$\begin{aligned}\frac{\mathbf{R}'_t\mathbf{R}_t}{N} &= \gamma^2_{0,t-1}+\boldsymbol{\delta}'_{\mathrm{f},t-1}\frac{\mathbf{B}'\mathbf{B}}{N}\boldsymbol{\delta}_{\mathrm{f},t-1}+\boldsymbol{\gamma}'_{\mathrm{z},t-1}\frac{\mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}}{N}\boldsymbol{\gamma}_{\mathrm{z},t-1}+\frac{\boldsymbol{\epsilon}_t{}'\boldsymbol{\epsilon}_t}{N}\\ &+ 2\,\gamma_{0,t-1}\frac{\mathbf{1}'_N\mathbf{B}}{N}\boldsymbol{\delta}_{\mathrm{f},t-1}+2\,\gamma_{0,t-1}\frac{\mathbf{1}'_N\mathbf{Z}_{t-1}}{N}\boldsymbol{\gamma}_{\mathrm{z},t-1}+2\gamma_{0,t-1}\frac{\mathbf{1}'_N\boldsymbol{\epsilon}_t}{N}\\ &+ 2\boldsymbol{\delta}'_{\mathrm{f},t-1}\frac{\mathbf{B}'\mathbf{Z}_{t-1}}{N}\boldsymbol{\gamma}_{\mathrm{z},t-1}+2\boldsymbol{\delta}'_{\mathrm{f},t-1}\frac{\mathbf{B}'\boldsymbol{\epsilon}_t}{N}+2\boldsymbol{\gamma}'_{\mathrm{z},t-1}\frac{\mathbf{Z}'_{t-1}\boldsymbol{\epsilon}_t}{N}\\ &\rightarrow_p \gamma^2_{0,t-1}+\boldsymbol{\delta}'_{\mathrm{f},t-1}\mathbf{\Sigma}_\beta\boldsymbol{\delta}_{\mathrm{f},t-1}+\boldsymbol{\gamma}'_{\mathrm{z},t-1}\mathbb{J}'_{t-1}\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1}+\sigma^2\\ &+ 2\gamma_{0,t-1}\boldsymbol{\mu}_\beta{}'\boldsymbol{\delta}_{\mathrm{f},t-1}+2\gamma_{0,t-1}\boldsymbol{\mu}'_z\boldsymbol{\gamma}_{\mathrm{z},t-1}+2\boldsymbol{\delta}'_{\mathrm{f},t-1}\mathbf{\Sigma}'_{\mathrm{ZB}}\mathbb{J}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1}\end{aligned}$$

Therefore, combining terms

$$\begin{aligned}\frac{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t}{N} &= \frac{\mathbf{R}'_t\mathbf{R}_t}{N}-\frac{\mathbf{R}'_t\mathbf{1}_N\mathbf{1}'_N\mathbf{R}_t}{N}\\ &\rightarrow_p \boldsymbol{\delta}'_{\mathrm{f},t-1}\left(\mathbf{\Sigma}_\beta-\boldsymbol{\mu}_\beta\boldsymbol{\mu}'_\beta\right)\boldsymbol{\delta}_{\mathrm{f},t-1}+\boldsymbol{\gamma}'_{\mathrm{z},t-1}\left(\mathbb{J}'_{t-1}\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1}-\boldsymbol{\mu}_z\boldsymbol{\mu}'_z\right)\boldsymbol{\gamma}_{\mathrm{z},t-1}\\ &+ \sigma^2+2\boldsymbol{\delta}'_{\mathrm{f},t-1}\left(\mathbf{\Sigma}'_{\mathrm{ZB}}\mathbb{J}_{t-1}-\boldsymbol{\mu}_\beta\boldsymbol{\mu}'_z\right)\boldsymbol{\gamma}_{\mathrm{z},t-1}\\ &\equiv \sigma_{\tilde{\mathrm{R}},t},\end{aligned}$$

implying that $\sigma_{\tilde{\mathrm{R}},t} > \sigma^2 > 0$.

Consider now the numerator of (OA.49) and notice that, under Assumptions OA.3–OA.7 and using Lemma 2,

$$\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{\hat{X}}\mathbf{Z}_{t-1}}{N} \to_p \mathbb{J}'_{t-1}\boldsymbol{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1} - \boldsymbol{\Sigma}_{\mathrm{ZX,t-1}}\left(\boldsymbol{\Sigma}_{\mathrm{X}} + \sigma^2\mathbf{P}'\mathbf{P}\right)^{-1}\boldsymbol{\Sigma}'_{\mathrm{ZX,t-1}} \equiv \boldsymbol{\Sigma}_{\mathrm{Z\hat{X}Z},t-1}.$$

Moreover, using Theorems 1-(ii) and 2, we have that

$$\sqrt{N}\left(\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1} - \boldsymbol{\gamma}_{\mathrm{z},t-1}\right) \to_d \mathcal{N}\left(\mathbf{0}_{K_{\mathrm{z}}}, \mathbf{L}^{-1}_{\mathrm{z},t-1}\,\mathbf{O}_{t-1}\,\mathbf{L}^{-1\prime}_{\mathrm{z},t-1}\right),$$

with $\mathbf{L}_{\mathrm{z},t-1} \equiv \left[\mathbf{0}_{K_{\mathrm{z}}\times(K_{\mathrm{f}}+1)}, \mathbf{I}_{K_{\mathrm{z}}}\right]\mathbf{L}_{t-1}$. Therefore, under the null hypothesis of $\boldsymbol{\gamma}_{\mathrm{z},t-1} = \mathbf{0}_{K_{\mathrm{z}}}$, and denoting for brevity $\hat{\mathbf{V}}_{\mathrm{LOL}} \equiv \hat{\mathbf{L}}^{-1}_{\mathrm{z},t-1}\,\hat{\mathbf{O}}_{t-1}\,\hat{\mathbf{L}}^{-1\prime}_{\mathrm{z},t-1}$, and $\mathbf{V}_{\mathrm{LOL}} \equiv \mathbf{L}^{-1}_{\mathrm{z},t-1}\,\mathbf{O}_{t-1}\,\mathbf{L}^{-1\prime}_{\mathrm{z},t-1}$, with $\hat{\mathbf{L}}_{\mathrm{z},t-1} \equiv \left[\mathbf{0}_{K_{\mathrm{z}}\times(K_{\mathrm{f}}+1)}, \mathbf{I}_{K_{\mathrm{z}}}\right]\hat{\mathbf{L}}_{t-1}$, we have that $\sqrt{N}\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{-\frac{1}{2}}\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1} \to_d \mathcal{N}(\mathbf{0}_{K_{\mathrm{z}}}, \mathbf{I}_{K_{\mathrm{z}}})$, and

$$\frac{\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{\frac{1}{2}}\left(\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{\hat{X}}\mathbf{Z}_{t-1}}{N}\right)\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{\frac{1}{2}}}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t/N} \to_p \frac{\left(\mathbf{V}_{\mathrm{LOL}}\right)^{\frac{1}{2}}\boldsymbol{\Sigma}_{\mathrm{Z\hat{X}Z},t-1}\left(\mathbf{V}_{\mathrm{LOL}}\right)^{\frac{1}{2}}}{\sigma_{\tilde{\mathrm{R}},t}} \equiv \boldsymbol{\Theta}_{t-1},$$

with $\boldsymbol{\Theta}_{t-1}$ being a symmetric and positive definite matrix admitting the spectral decomposition $\boldsymbol{\Theta}_{t-1} = \boldsymbol{\Delta}_{1,t-1}\boldsymbol{\Delta}_{2,t-1}\boldsymbol{\Delta}'_{1,t-1}$, for an orthogonal matrix $\boldsymbol{\Delta}_{1,t-1}$ and a diagonal matrix $\boldsymbol{\Delta}_{2,t-1} = \mathrm{diag}(d_{1,t-1}, \cdots, d_{K_{\mathrm{z}},t-1})$, whose diagonal elements correspond to the eigenvalues of $\boldsymbol{\Theta}_{t-1}$. Therefore, combining these results into (OA.49), we have

$$\begin{aligned}
\mathcal{T}^2_{\mathrm{z},t-1} &\equiv \frac{\sqrt{N}\hat{\boldsymbol{\gamma}}^{*\prime}_{\mathrm{z},t-1}\left(\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{\hat{X}}\mathbf{Z}_{t-1}}{N}\right)\sqrt{N}\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1}}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t/N} \\
&= \frac{\sqrt{N}\hat{\boldsymbol{\gamma}}^{*\prime}_{\mathrm{z},t-1}\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{-\frac{1}{2}}\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{\frac{1}{2}}\left(\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{\hat{X}}\mathbf{Z}_{t-1}}{N}\right)\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{\frac{1}{2}}\left(\hat{\mathbf{V}}_{\mathrm{LOL}}\right)^{-\frac{1}{2}}\sqrt{N}\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1}}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t/N} \\
&\to_d \quad \xi'\boldsymbol{\Theta}_{t-1}\xi = \xi'\boldsymbol{\Delta}_{1,t-1}\boldsymbol{\Delta}_{2,t-1}\boldsymbol{\Delta}'_{1,t-1}\xi = \sum_{i=1}^{K_{\mathrm{z}}}\chi^2_{1,j}d_{j,t-1}
\end{aligned}$$

denoting $\xi \sim \mathcal{N}(\mathbf{0}'_{K_{\mathrm{z}}}, \mathbf{I}_{K_{\mathrm{z}}})$, and where the last equality follows by noticing that $\xi\boldsymbol{\Delta}_{1,t-1} \sim \mathcal{N}(\mathbf{0}_{K_{\mathrm{z}}}, \mathbf{I}_{K_{\mathrm{z}}})$, given $\boldsymbol{\Delta}'_{1,t-1}\mathbf{I}_{K_{\mathrm{z}}}\boldsymbol{\Delta}_{1,t-1} = \mathbf{I}_{K_{\mathrm{z}}}$. Finally, notice that, under the null hypothesis of part (i),

$$\sigma_{\tilde{\mathrm{R}},t} = \boldsymbol{\delta}'_{\mathrm{f},t-1}\left(\boldsymbol{\Sigma}_\beta - \boldsymbol{\mu}_\beta\boldsymbol{\mu}'_\beta\right)\boldsymbol{\delta}_{\mathrm{f},t-1} + \sigma^2.$$

This concludes the proof of part (i).

To prove part (ii), let us first define the following $d \times 1$ random vector

$$\boldsymbol{\theta} = \left(\text{vec}\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right), \frac{\mathbf{Z}'\mathbf{1}_N}{N}, \text{vec}\left(\frac{\mathbf{Z}'\mathbf{B}}{N}\right), \text{vec}\left(\frac{\mathbf{Z}'\boldsymbol{\epsilon}'}{N}\right), \text{vec}\left(\frac{\mathbf{B}'\boldsymbol{\epsilon}'}{N}\right), \text{vec}\left(\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{N}\right), \frac{\boldsymbol{\epsilon}'\mathbf{1}_N}{N} \right) \quad \text{(OA.50)}$$

with $d \equiv ((T-1)K_{\mathrm{z}})^2 + (T-1)K_{\mathrm{z}} + (T-1)K_{\mathrm{z}}K_{\mathrm{f}} + (T-1)^2K_{\mathrm{z}} + (T-1)K_{\mathrm{f}} + (T-1)^2 + (T-1)$. Then, consider $\hat{R}^2_{\mathrm{z},t-1}$ in (60) and notice that it can be written as a function of $\boldsymbol{\theta}$, such that

$$\hat{R}^2_{\mathrm{z},t-1} = g_{t-1}(\boldsymbol{\theta}),$$

with $g_{t-1}(\cdot)$ being an elementary and differentiable function, made by simple products and ratios of the arguments in $\boldsymbol{\theta}$.[8] All the random quantities in $\boldsymbol{\theta}$ admit a continuous second-order derivative, implying

$$\sqrt{N} \begin{bmatrix} \text{vec}\left(\frac{\mathbf{Z}'\mathbf{Z}}{N} - \boldsymbol{\Sigma}_{\mathrm{Z}}\right) \\ \frac{\mathbf{Z}'\mathbf{1}_N}{N} - \boldsymbol{\mu}_{\mathrm{z},T-1} \\ \text{vec}\left(\frac{\mathbf{Z}'\mathbf{B}}{N} - \boldsymbol{\Sigma}_{\mathrm{ZB}}\right) \\ \text{vec}\left(\frac{\mathbf{Z}'\boldsymbol{\epsilon}'}{N}\right) \\ \text{vec}\left(\frac{\mathbf{B}'\boldsymbol{\epsilon}'}{N}\right) \\ \text{vec}\left(\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{N} - \frac{\sigma^2}{T-1}\right) \\ \frac{\boldsymbol{\epsilon}'\mathbf{1}_N}{N} \end{bmatrix} \to_d \mathcal{N}(\mathbf{0}_d, \mathbf{V}_{\boldsymbol{\theta}}),$$

with the $d \times d$ covariance matrix $\mathbf{V}_{\boldsymbol{\theta}}$ having the following form

$$\mathbf{V}_{\boldsymbol{\theta}} \equiv \begin{bmatrix} \text{Var}[\boldsymbol{\theta}_1] & \text{Cov}[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2'] & \cdots & \text{Cov}[\boldsymbol{\theta}_1, \boldsymbol{\theta}_7'] \\ \text{Cov}[\boldsymbol{\theta}_2, \boldsymbol{\theta}_1'] & \text{Var}(\boldsymbol{\theta}_2] & \cdots & \text{Cov}[\boldsymbol{\theta}_2, \boldsymbol{\theta}_7'] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[\boldsymbol{\theta}_7, \boldsymbol{\theta}_1'] & \text{Cov}[\boldsymbol{\theta}_7, \boldsymbol{\theta}_2'] & \cdots & \text{Var}[\boldsymbol{\theta}_7] \end{bmatrix}, \quad \text{(OA.51)}$$

setting $\boldsymbol{\theta}_1 \equiv \sqrt{N}\text{vec}\left(\frac{\mathbf{Z}'\mathbf{Z}}{N} - \boldsymbol{\Sigma}_{\mathrm{Z}}\right)$, $\boldsymbol{\theta}_2 \equiv \sqrt{N}\left(\frac{\mathbf{Z}'\mathbf{1}_N}{N} - \boldsymbol{\mu}_{\mathrm{z},T-1}\right)$, $\boldsymbol{\theta}_3 \equiv \sqrt{N}\text{vec}\left(\frac{\mathbf{Z}'\mathbf{B}}{N} - \boldsymbol{\Sigma}_{\mathrm{ZB}}\right)$, $\boldsymbol{\theta}_4 \equiv \text{vec}\left(\frac{\mathbf{Z}'\boldsymbol{\epsilon}'}{\sqrt{N}}\right)$, $\boldsymbol{\theta}_5 \equiv \text{vec}\left(\frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\right)$, $\boldsymbol{\theta}_6 \equiv \sqrt{N}\text{vec}\left(\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{N} - \frac{\sigma^2}{T-1}\right)$, and $\boldsymbol{\theta}_7 \equiv \frac{\boldsymbol{\epsilon}'\mathbf{1}_N}{\sqrt{N}}$.

Now, by Assumptions OA.7 and OA.13, $\text{Var}(\boldsymbol{\theta}_1) = \text{Var}\left[\text{vec}\left(\frac{\mathbf{Z}'\mathbf{Z}}{\sqrt{N}}\right)\right] \to \mathbf{U}_{\mathrm{Z}}$, $\text{Var}(\boldsymbol{\theta}_2) = \text{Var}\left[\frac{\mathbf{Z}'\mathbf{1}_N}{\sqrt{N}}\right] \to \boldsymbol{\Sigma}_{\mathrm{Z}} - \boldsymbol{\mu}_{\mathrm{z},T-1}\boldsymbol{\mu}_{\mathrm{z},T-1}'$, $\text{Var}(\boldsymbol{\theta}_3) = \text{Var}\left[\text{vec}\left(\frac{\mathbf{Z}'\mathbf{B}}{\sqrt{N}}\right)\right] \to (\boldsymbol{\Sigma}_\beta \otimes \boldsymbol{\Sigma}_{\mathrm{Z}}) - \left(\boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta' \otimes \boldsymbol{\mu}_{\mathrm{z},T-1}\boldsymbol{\mu}_{\mathrm{z},T-1}'\right) \equiv \tilde{\boldsymbol{\Sigma}}_{\beta\otimes\mathrm{Z}}$, $\text{Var}(\boldsymbol{\theta}_4) = \text{Var}\left[\text{vec}\left(\frac{\mathbf{Z}'\boldsymbol{\epsilon}'}{\sqrt{N}}\right)\right] \to (\sigma^2\mathbf{I}_{T-1} \otimes \boldsymbol{\Sigma}_{\mathrm{Z}})$, $\text{Var}(\boldsymbol{\theta}_5) = \text{Var}\left[\text{vec}\left(\frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\right]\right) \to (\sigma^2\mathbf{I}_{T-1} \otimes \boldsymbol{\Sigma}_\beta)$, $\text{Var}(\boldsymbol{\theta}_6) = \text{Var}\left[\text{vec}\left(\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{\sqrt{N}}\right)\right] \to \mathbf{U}_\epsilon$, and $\text{Var}(\boldsymbol{\theta}_7) = \text{Var}\left[\frac{\boldsymbol{\epsilon}'\mathbf{1}_N}{\sqrt{N}}\right] \to \sigma^2$.

[8]To ease the exposition, we do not repor the $g_{t-1}(.)$ function, because it is elementary. Details are available upon request.

Consider now all the covariance terms. Under Assumptions OA.6, OA.7 and OA.13, we have $\mathrm{Cov}[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2'] \to \boldsymbol{\Sigma}_{\mathrm{z}\otimes\mathrm{z}}$, $\mathrm{Cov}[\boldsymbol{\theta}_1, \boldsymbol{\theta}_3'] \to \boldsymbol{\Sigma}_{\mathrm{z}\otimes\mathrm{z}}\left(\boldsymbol{\mu}_\beta' \otimes \mathbf{I}_{K_\mathrm{z}}\right)$, $\mathrm{Cov}[\boldsymbol{\theta}_2, \boldsymbol{\theta}_3'] \to \boldsymbol{\mu}_\beta' \otimes \left(\boldsymbol{\Sigma}_\mathrm{Z} - \boldsymbol{\mu}_{\mathrm{z},T-1}\boldsymbol{\mu}_{\mathrm{z},T-1}'\right)$, $\mathrm{Cov}[\boldsymbol{\theta}_4, \boldsymbol{\theta}_5'] \to \sigma^2 \mathbf{I}_{T-1} \otimes \boldsymbol{\mu}_{\mathrm{z},T-1}\boldsymbol{\mu}_\beta'$, $\mathrm{Cov}[\boldsymbol{\theta}_4, \boldsymbol{\theta}_7'] \to \sigma^2 \mathbf{I}_{T-1} \otimes \boldsymbol{\mu}_{\mathrm{z},T-1}'$, and $\mathrm{Cov}[\boldsymbol{\theta}_5, \boldsymbol{\theta}_7'] \to \sigma^2 \mathbf{I}_{T-1} \otimes \boldsymbol{\mu}_\beta'$. Moreover, under Assumption OA.5, it follows that all the remaining covariance terms are zero matrices. Putting all together gives $\mathbf{V}_{\boldsymbol{\theta}}$.

Therefore, by the mean-value theorem, it follows that

$$\sqrt{N}(\hat{R}^2_{\mathrm{z},t-1} - R^2_{\mathrm{z},t-1}) \to_d \mathcal{N}(0, \omega_{\mathrm{z},t-1}) \quad \text{with} \quad \omega_{\mathrm{z},t-1} \equiv \mathbf{G}_{t-1}' \mathbf{V}_{\boldsymbol{\theta}} \mathbf{G}_{t-1}, \tag{OA.52}$$

setting $\mathbf{G}_{t-1}(\mathbf{x}) \equiv \partial g_{t-1}(\mathbf{x})/\partial \mathbf{x}$ for a generic d-dimensional vector $\mathbf{x}$, and $\mathbf{G}_{t-1} \equiv \mathbf{G}_{t-1}(\mathbf{x}_0)$ with [9]

$$\mathbf{x}_0 \equiv \left(\mathrm{vec}'\left(\boldsymbol{\Sigma}_\mathrm{Z}\right), \boldsymbol{\mu}_{\mathrm{z},T-1}', \mathrm{vec}'(\boldsymbol{\Sigma}_{\mathrm{ZB}}), \mathbf{0}_{(T-1)^2K_\mathrm{z}}', \mathbf{0}_{(T-1)K_\mathrm{f}}', \mathrm{vec}'(\sigma^2 \mathbf{I}_{T-1}), \mathbf{0}_{(T-1)}'\right)'.$$

Finally, a consistent estimator $\hat{\omega}_{\mathrm{z},t-1}$ for $\omega_{\mathrm{z},t-1}$ is obtained by replacing $\mathbf{G}_{t-1}$ with

$$\hat{\mathbf{G}}_{t-1} \equiv \mathbf{G}_{t-1}\left((\mathrm{vec}'\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right), \mathrm{vec}'\left(\frac{\mathbf{Z}'\mathbf{1}_N}{N}\right), \mathrm{vec}'\left(\frac{\mathbf{Z}'\hat{\mathbf{B}}}{N}\right), \mathbf{0}_{(T-1)^2K_\mathrm{z}}', \mathbf{0}_{(T-1)K_\mathrm{f}}', \mathrm{vec}'(\hat{\sigma}^2 \mathbf{I}_{T-1}))', \mathbf{0}_{(T-1)}'\right),$$

and replacing the terms of $\mathbf{V}_{\boldsymbol{\theta}}$ with their sample counterparts, yielding

$$\hat{\omega}_{\mathrm{z},t-1} \equiv \hat{\mathbf{G}}_{t-1}' \hat{\mathbf{V}}_{\boldsymbol{\theta}} \hat{\mathbf{G}}_{t-1}. \tag{OA.53}$$

■

[9] The mapping from vectorized matrices to the original matrices is given by the linear function $\mathrm{vec}^{-1}_{m\times p}(\mathbf{x}) = (\mathrm{vec}'(\mathbf{I}_p) \otimes \mathbf{I}_m)(\mathbf{I}_p \otimes \mathbf{x})$ mapping the $mp\times$ vector $\mathbf{x}$ into a $m \times p$ matrix.

OA.5 Further Results on the CSR OLS and CSR WLS Estimators

This section is structured as follows. First, we focus on the traditional two-pass CSR OLS and CSR WLS in (30) in (41), respectively, showing the bias that arises, both in the fixed-T and large-T cases. Next, focusing for simplicity only on the CSR OLS estimator, we study the limiting properties of the locally-averaged estimator (40). Finally, we provide further results for the CSR OLS under global misspecification (53), by deriving the misspecification-robust estimator (53), together with the corresponding standard error, asymptotically valid when $N \to \infty$.

OA.5.1 The Augmented-Traditional CSR OLS and CSR WLS Eestimators

In this section, we investigate the limiting behaviour of the augmented traditional CSR OLS and CSR WLS estimators defined in (30) in (41), respectively, when $N \to \infty$ and T is kept fixed. These estimators generalize the conventional approach to the case when both anomalies and (estimated) loadings are used in the cross-sectional OLS regression, hence resolving the bias coming from the potential lack of orthogonality between the risk factors and the anomalies. However, as we show in Propositions OA.1 and OA.2 below, other sources of bias arise in the fixed-T setup, making them biased.

Proposition OA.1 (biases of CSR OLS — time-varying estimator)**.** *Let $K = K_{\mathrm{z}} + K_{\mathrm{f}}$, and define the two matrices*

$$\boldsymbol{\Lambda}_1 \equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} \\ \mathbf{0}_{K_{\mathrm{f}}} & \sigma^2 \mathbf{P}'\mathbf{P} \end{bmatrix}, \qquad \boldsymbol{\Lambda}_{2,t-1} \equiv \sigma^2 \begin{bmatrix} 0 \\ \mathbf{P}'\boldsymbol{\imath}_{t-1,T-1} \end{bmatrix}. \tag{OA.54}$$

Under Assumptions OA.1–OA.7, as $N \to \infty$,

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{f,t-1} \\ \hat{\boldsymbol{\gamma}}_{\mathrm{z},t-1} \end{bmatrix} \to_p \begin{bmatrix} \boldsymbol{\Gamma}_{\mathrm{f},t-1} \\ \boldsymbol{\gamma}_{\mathrm{z},\mathrm{t}-1} \end{bmatrix} + \mathbf{C}_{t-1}^{-1} \left(- \begin{bmatrix} \boldsymbol{\Lambda}_1 \\ \mathbf{0}_{K+1} \end{bmatrix} \boldsymbol{\Gamma}_{t-1} + \begin{bmatrix} \boldsymbol{\Lambda}_{2,t-1} \\ \mathbf{0}_{K_{\mathrm{z}}} \end{bmatrix} \right), \tag{OA.55}$$

where

$$\mathbf{C}_{t-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{X}} + \boldsymbol{\Lambda}_1 & \boldsymbol{\Sigma}'_{\mathrm{ZX,t-1}} \\ \boldsymbol{\Sigma}_{\mathrm{ZX,t-1}} & \mathbb{J}'_{t-1}\boldsymbol{\Sigma}_{\mathrm{Z}}\mathbb{J}_{t-1} \end{bmatrix}. \tag{OA.56}$$

Proof. Using (29), we can rewrite:

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \\ \hat{\boldsymbol{\gamma}}_{\mathrm{z},t-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_{\mathrm{f},t-1} \\ \boldsymbol{\gamma}_{\mathrm{z},t-1} \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}' \\ \mathbf{Z}'_{t-1} \end{bmatrix} (\boldsymbol{\epsilon}_t + (\mathbf{X} - \hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1}).$$

By Lemmas 1 and 2, then $N^{-1}\begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{Z}_{t-1}\\ \mathbf{Z}_{t-1}'\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}\end{bmatrix} \to_p \mathbf{C}_{t-1}$, and, by Lemma 3,

$$\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\\ \mathbf{Z}_{t-1}'\end{bmatrix}(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1}) \to_p \left(-\begin{bmatrix}\boldsymbol{\Lambda}_1\\ \mathbf{0}_{K+1}\end{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1} + \begin{bmatrix}\boldsymbol{\Lambda}_{2,t-1}\\ \mathbf{0}_{K_{\mathrm{z}}}\end{bmatrix}\right).$$

■

Proposition OA.2 (biases of CSR WLS — time-varying estimator—weighted)**.** *Under Assumptions OA.1–OA.11, as $N \to \infty$,*

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{(\mathrm{w})}\\ \hat{\gamma}_{\mathrm{z},t-1}^{(\mathrm{w})}\end{bmatrix} \to_p \begin{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1}\\ \gamma_{\mathrm{z},t-1}\end{bmatrix} + \mathbf{D}_{t-1}^{-1}\left(-\begin{bmatrix}\boldsymbol{\Lambda}_1\\ \mathbf{0}_{K_z\times(K_f+1)}\end{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1} + \begin{bmatrix}\boldsymbol{\Lambda}_{t-1,2}\\ \mathbf{0}_{K_z}\end{bmatrix}\right) \tag{OA.57}$$

where

$$\mathbf{D}_{t-1} \equiv \begin{bmatrix}1 & \boldsymbol{\mu}_\beta' & \boldsymbol{\mu}_{\mathrm{z},t-1}'\\ \boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta + \sigma^2\mathbf{P}'\mathbf{P} & \boldsymbol{\mu}_\beta\boldsymbol{\mu}_{\mathrm{z},t-1}'\\ \boldsymbol{\mu}_{\mathrm{z},t-1} & \boldsymbol{\mu}_{\mathrm{z},t-1}\boldsymbol{\mu}_\beta' & \boldsymbol{\Sigma}_{Z,t-1}\end{bmatrix},$$

with $\boldsymbol{\mu}_{\mathrm{z},t-1}, \boldsymbol{\Sigma}_{Z,t-1}$ are defined in Assumption OA.11.

Proof. Rewrite:

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{w}\\ \hat{\gamma}_{\mathrm{z},t-1}^{w}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1}\\ \gamma_{\mathrm{z},t-1}\end{bmatrix} + \begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1}\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\\ \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\end{bmatrix}^{-1}\begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1}\\ \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\end{bmatrix}(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\boldsymbol{\Gamma}_{\mathrm{f},t-1}).$$

By Lemmas 4–5,

$$\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\mathbf{W}_{t-1}\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\\ \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\end{bmatrix} \to_p \mathbf{D}_{t-1}. \tag{OA.58}$$

By Lemmas 5 and 6,

$$\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\\ \mathbf{Z}_{t-1}'\end{bmatrix}\mathbf{W}_{t-1}(\boldsymbol{\epsilon}_t + (\mathbf{X}-\hat{\mathbf{X}})\boldsymbol{\Gamma}_{t-1}) \to_p \left(-\begin{bmatrix}\boldsymbol{\Lambda}_1\\ \mathbf{0}_{K_z\times(K_f+1)}\end{bmatrix}\boldsymbol{\Gamma}_{t-1} + \begin{bmatrix}\boldsymbol{\Lambda}_{t-1,2}\\ \mathbf{0}_{K_z}\end{bmatrix}\right).$$

QED

Remark OA.19. *Proposition OA.1 shows that the conventional augmented CSR OLS estimator is biased whenever $N \to \infty$ and T is kept fixed. However, it is possible to show that bias also arises in the conventional large-T–fixed-N setting. Specifically, when $\mathbf{P}'\boldsymbol{\epsilon} \to_p \mathbf{0}_{K_{\mathrm{f}}\times N}$ and $(\mathbf{X}, \mathbf{Z}_{t-1})$ has full-column rank, then*

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}\\ \hat{\gamma}_{\mathrm{z},t-1}\end{bmatrix} \to_p \begin{bmatrix}\boldsymbol{\Gamma}_{\mathrm{f},t-1}\\ \gamma_{\mathrm{z},t-1}\end{bmatrix} + \begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_{t-1}\\ \mathbf{Z}_{t-1}'\mathbf{X} & \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\\ \mathbf{Z}_{t-1}'\end{bmatrix}\boldsymbol{\epsilon}_t. \tag{OA.59}$$

The result in (OA.59) *shows a bias term which is linear in* $\boldsymbol{\epsilon}_t$ *and, hence, random and not predictable, making the bias term impossible to estimate consistently.*

The same applies to the CSR WLS estimator. When $\mathbf{P}'\boldsymbol{\epsilon} \to_p \mathbf{0}_{K_{\mathrm{f}}\times N}$ *and* $\mathbf{W}_{t-1}^{\frac{1}{2}}(\mathbf{X}, \mathbf{Z}_{t-1})$ *has a full-column rank, then as* $T \to \infty$ *and* N *remains fixed,*

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{(\mathrm{w})} \\ \\ \hat{\boldsymbol{\gamma}}_{\mathrm{z},t-1}^{(\mathrm{w})} \end{bmatrix} \to_p \begin{bmatrix} \boldsymbol{\Gamma}_{\mathrm{f},t-1} \\ \\ \boldsymbol{\gamma}_{\mathrm{z},t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{X}'\mathbf{W}_{t-1}\mathbf{X} & \mathbf{X}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1} \\ \\ \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{X} & \mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{W}_{t-1} \\ \\ \mathbf{Z}_{t-1}'\mathbf{W}_{t-1} \end{bmatrix} \boldsymbol{\epsilon}_t, \tag{OA.60}$$

hence, also affected by a random bias.

OA.5.2 Anomalies with Time-Varying Premia: Locally-Averaged CSR OLS Estimation

In Theorems 1 we have shown that our time-varying estimator $\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1}$ and $\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1}$ accurately capture the true premia $\boldsymbol{\Gamma}_{\mathrm{f},t-1}$ and $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ at any given point in time. However, when premia's time-variation is sufficiently smooth and not too abrupt - something that seems not so hard to justify in our fixed-T environment - one could benefit from the time-series dimension of the panel and obtain more precise estimates of the premia parameters by means of rolling-windows average estimates. As explained in Section 5, this reasoning suggests to use the locally-averaged CSR OLS estimator (40):

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}^*_{\mathrm{f}} \\ \hat{\boldsymbol{\gamma}}^*_{\mathrm{z}} \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\bar{\mathbf{Z}} \\ \bar{\mathbf{Z}}'\hat{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\bar{\mathbf{R}} \\ \bar{\mathbf{Z}}'\bar{\mathbf{R}} \end{bmatrix},$$

where $\hat{\boldsymbol{\Lambda}}_1$ is defined in (32), and where $\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f}} \equiv \left(\hat{\gamma}^*_0, \, \hat{\boldsymbol{\delta}}^{*'}_{\mathrm{f}}\right)'$. The next theorem establishes its limiting statistical properties.

Theorem OA.1 (large-N–fixed-T - consistency and asymptotic normality of the locally-averaged bias-adjusted CSR OLS estimator)**.** *Under Assumptions OA.1—OA.7 and* $\widehat{\mathrm{Cov}}(\mathbf{Z}_{t-1}, \boldsymbol{\gamma}_{\mathrm{z},t-1}) = o_p(N^{-1/2})$*, as* $N \to \infty$*,*

(i)

$$\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f}} - \boldsymbol{\Gamma}_{\mathrm{f}} = O_p\left(\frac{1}{\sqrt{N}}\right) \quad \textit{and} \quad \hat{\boldsymbol{\gamma}}^*_{\mathrm{z}} - \boldsymbol{\gamma}_{\mathrm{z}} = O_p\left(\frac{1}{\sqrt{N}}\right), \tag{OA.61}$$

(ii)

$$\sqrt{N} \begin{bmatrix} \hat{\boldsymbol{\Gamma}}^*_{\mathrm{f}} - \boldsymbol{\Gamma}_{\mathrm{f}} \\ \hat{\boldsymbol{\gamma}}^*_{\mathrm{z}} - \boldsymbol{\gamma}_{\mathrm{z}} \end{bmatrix} \to_d \mathcal{N}\left(\mathbf{0}_{K+1}, \mathbf{L}^{-1}\mathbf{O}\mathbf{L}^{-1'}\right), \tag{OA.62}$$

where

$$\mathbf{L} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{X}} & \boldsymbol{\Sigma}'_{\mathrm{ZX}} \\ \boldsymbol{\Sigma}_{\mathrm{ZX}} & \mathbb{J}'\boldsymbol{\Sigma}_{\mathrm{Z}}\mathbb{J} \end{bmatrix} > 0, \quad \textit{and} \quad \mathbf{O} \equiv \begin{bmatrix} \mathbf{U} & \sigma^2\mathbf{G}\mathbf{H}' \\ \sigma^2\mathbf{H}\mathbf{G}' & \mathbf{H}\boldsymbol{\Sigma}_{\mathrm{U}}\mathbf{H}', \end{bmatrix} \tag{OA.63}$$

with $\mathbf{U} \equiv \frac{\sigma^2}{T-1}\left[1 + (T-1)\bar{\boldsymbol{\delta}}_{\mathrm{f}}{}'\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}\right]\boldsymbol{\Sigma}_{\mathrm{X}} + \begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} \\ \mathbf{0}_{K_{\mathrm{f}}} & \mathbf{V}'\mathbf{U}_{\epsilon}\mathbf{V} \end{bmatrix}$, $\mathbf{U}_{\epsilon}$, $\boldsymbol{\Sigma}_{\mathrm{ZB}}$, $\boldsymbol{\Sigma}_{\mathrm{ZX}}$, $\boldsymbol{\Sigma}_{\mathrm{U}}$ *and* $\boldsymbol{\mu}_{\mathrm{z},T-1}$

defined in Assumptions OA.3 and OA.7, and in Lemma 2, and where

$$\begin{aligned}
\mathbf{Q} &\equiv \frac{\mathbf{1}_{T-1}}{(T-1)} - \mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}, \\
\mathbf{V} &\equiv (\mathbf{Q}\otimes\mathbf{P}) - \frac{\mathrm{vec}(\mathbb{M}_{\tilde{D}})}{T-K-2}\mathbf{Q}'\mathbf{P}, \\
\mathbf{G} &\equiv \left[\mathbf{Q}\otimes\boldsymbol{\mu}_{\mathrm{z},T-1}, \quad \mathbf{Q}\otimes\boldsymbol{\Sigma}_{\mathrm{ZB}}\right]', \\
\mathbf{H} &\equiv \mathbf{Q}'\otimes\mathbb{J}'.
\end{aligned}$$

Proof. Starting from the definition in (40), we can rewrite

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f}} \\ \hat{\gamma}^*_{\mathrm{z}}\end{bmatrix} = \begin{bmatrix}\bar{\boldsymbol{\Gamma}}_{\mathrm{f}} \\ \bar{\gamma}_{\mathrm{z}}\end{bmatrix} + \begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\bar{\mathbf{Z}} \\ \bar{\mathbf{Z}}'\hat{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}}\end{bmatrix}^{-1}\left(\begin{bmatrix}N\hat{\boldsymbol{\Lambda}}_1\bar{\boldsymbol{\Gamma}}_{\mathrm{f}} \\ \mathbf{0}_{K_{\mathrm{z}}}\end{bmatrix} + \begin{bmatrix}\hat{\mathbf{X}}' \\ \bar{\mathbf{Z}}'\end{bmatrix}(\bar{\boldsymbol{\epsilon}} + (\mathbf{X}-\hat{\mathbf{X}})\bar{\boldsymbol{\Gamma}}_{\mathrm{f}})\right). \tag{OA.64}$$

By Lemma 1 and Assumption OA.7, $\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\bar{\mathbf{Z}} \\ \bar{\mathbf{Z}}'\hat{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}}\end{bmatrix}^{-1} = O_p(1)$. Moreover, notice that

$$\left[\frac{\hat{\mathbf{X}}'}{N}\bar{\boldsymbol{\epsilon}} - \frac{\hat{\mathbf{X}}'}{N}(\hat{\mathbf{X}}-\mathbf{X})\bar{\Gamma}_{\mathrm{f}} + \hat{\boldsymbol{\Lambda}}_1\bar{\Gamma}_{\mathrm{f}}\right] = \left[\frac{\hat{\mathbf{X}}'}{N}\bar{\boldsymbol{\epsilon}} - \begin{bmatrix} 1_N'\frac{\boldsymbol{\epsilon}'}{N}\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}} \\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{N}\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}} + \mathbf{P}'\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{N}\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}} - \hat{\sigma}^2\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}\end{bmatrix}\right],$$

where

$$\frac{1}{N}\hat{\mathbf{X}}'\bar{\boldsymbol{\epsilon}} = \frac{1}{N}(\hat{\mathbf{X}}-\mathbf{X})'\bar{\boldsymbol{\epsilon}} + \frac{1}{N}\mathbf{X}'\bar{\boldsymbol{\epsilon}} = O_p\left(N^{-\frac{1}{2}}\right)$$

by Lemma 1, whereas by Assumptions OA.6(i) and OA.6(iii)

$$\frac{1}{N}\mathbf{P}'\sum_{i=1}^{N}\boldsymbol{\epsilon}_i = O_p\left(N^{-\frac{1}{2}}\right), \qquad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_i\boldsymbol{\epsilon}_i'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}} = O_p\left(N^{-\frac{1}{2}}\right).$$

Next, note that the term $\mathbf{P}'\frac{1}{N}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{P}\bar{\boldsymbol{\delta}}_1 - \hat{\sigma}^2\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}$ can be rewritten as

$$\mathbf{P}'\left(\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{N} - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\mathbf{I}_{T-1}\right)\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}} - \left[(\hat{\sigma}^2-\sigma^2) - \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 - \sigma^2\right)\right]\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}, \tag{OA.65}$$

with $\mathbf{P}'\frac{1}{N}\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' - \sum_{i=1}^{N}\sigma_i^2\mathbf{I}_{T-1}\right)\mathbf{P}\bar{\boldsymbol{\delta}}_1 = O_p\left(N^{-\frac{1}{2}}\right)$ by Assumption OA.6(ii), $\hat{\sigma}^2 - \sigma^2 = O_p\left(N^{-\frac{1}{2}}\right)$, and $\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 - \sigma^2 = o\left(N^{-\frac{1}{2}}\right)$ by Lemma 1 and Assumption OA.5(i). It implies that the term in (OA.65) is $O_p\left(N^{-\frac{1}{2}}\right)$, which concludes the proof of part (i).

To prove part (ii), first notice that,

$$
\begin{aligned}
\sqrt{N}\begin{bmatrix}\hat{\boldsymbol{\Gamma}}_{\mathrm{f}}^{*}-\bar{\boldsymbol{\Gamma}}_{\mathrm{f}}\\ \hat{\gamma}_{\mathrm{z}}^{*}-\bar{\gamma}_{\mathrm{z}}\end{bmatrix} &= \begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-\hat{\boldsymbol{\Lambda}}_{1} & \frac{\hat{\mathbf{X}}'\bar{\mathbf{Z}}}{N}\\ \frac{\bar{\mathbf{Z}}'\hat{\mathbf{X}}}{N} & \frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{N}\end{bmatrix}^{-1}\times \\
&\times\left(\begin{bmatrix}\frac{\mathbf{1}_{N}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}\\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}\\ \mathbf{0}_{K_{\mathrm{z}}}\end{bmatrix}+\begin{bmatrix}0\\ \mathbf{P}'\frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}-\sqrt{N}\hat{\sigma}^{2}\mathbf{P}'\mathbf{Q}\\ \mathbf{0}_{K_{\mathrm{z}}}\end{bmatrix}+\begin{bmatrix}0\\ \mathbf{0}_{K_{\mathrm{f}}}\\ \sqrt{N}\left(\frac{\bar{\mathbf{Z}}'\boldsymbol{\epsilon}'}{N}\right)\mathbf{Q}\end{bmatrix}\right) \\
&= \begin{bmatrix}\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{N}-\hat{\boldsymbol{\Lambda}}_{1} & \frac{\hat{\mathbf{X}}'\bar{\mathbf{Z}}}{N}\\ \frac{\bar{\mathbf{Z}}'\hat{\mathbf{X}}}{N} & \frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{N}\end{bmatrix}^{-1}\left(\mathbf{a}_{1}+\mathbf{a}_{2}+\mathbf{a}_{3}\right), \qquad \text{(OA.66)}
\end{aligned}
$$

where we use the fact that $\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}=-\mathbf{P}'\mathbf{Q}$, with $\mathbf{Q}\equiv\frac{\mathbf{1}_{T-1}}{(T-1)}-\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}$. Now, using Lemmas 1 and 2, we have

$$
\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}}-N\hat{\boldsymbol{\Lambda}}_{1} & \hat{\mathbf{X}}'\bar{\mathbf{Z}}\\ \bar{\mathbf{Z}}'\hat{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}}\end{bmatrix}\to_{p}\begin{bmatrix}\boldsymbol{\Sigma}_{\mathrm{X}} & \boldsymbol{\Sigma}_{\mathrm{ZX}}'\\ \boldsymbol{\Sigma}_{\mathrm{ZX}} & \mathbb{J}'\boldsymbol{\Sigma}_{\mathrm{Z}}\mathbb{J}\end{bmatrix}\equiv\mathbf{L}. \qquad \text{(OA.67)}
$$

Next, consider the term $\mathbf{a}_1$. By Assumption OA.6, this term has zero mean with variance

$$
\begin{aligned}
\mathrm{Var}[I_1] &= \begin{bmatrix}\mathbf{Q}'\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i']\mathbf{Q} & \mathbf{Q}'\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'](\mathbf{Q}\otimes\boldsymbol{\beta}_i') & \mathbf{0}_{K_{\mathrm{z}}}'\\ \frac{1}{N}\sum_{i=1}^{N}(\mathbf{Q}'\otimes\boldsymbol{\beta}_i)\mathrm{E}[\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i']\mathbf{Q} & \frac{1}{N}\sum_{i=1}^{N}(\mathbf{Q}'\otimes\boldsymbol{\beta}_i)\mathrm{E}[\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'](\mathbf{Q}\otimes\boldsymbol{\beta}_i') & \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{z}}}\\ \mathbf{0}_{K_{\mathrm{z}}} & \mathbf{0}_{K_{\mathrm{z}}\times K_{\mathrm{f}}} & \mathbf{0}_{K_{\mathrm{z}}\times K_{\mathrm{z}}}\end{bmatrix}+o(1) \\
&\to \begin{bmatrix}\sigma^{2}\mathbf{Q}'\mathbf{Q}\boldsymbol{\Sigma}_{\mathrm{X}} & \mathbf{0}_{(K_{\mathrm{f}}+1)\times K_{\mathrm{z}}}\\ \mathbf{0}_{K_{\mathrm{z}}\times(K_{\mathrm{f}}+1)} & \mathbf{0}_{K_{\mathrm{z}}\times K_{\mathrm{z}}}\end{bmatrix}=\begin{bmatrix}\frac{\sigma^{2}}{T-1}\left[1+(T-1)\bar{\boldsymbol{\delta}}_{\mathrm{f}}'\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}\right]\boldsymbol{\Sigma}_{\mathrm{X}} & \mathbf{0}_{(K_{\mathrm{f}}+1)\times K_{\mathrm{z}}}\\ \mathbf{0}_{K_{\mathrm{z}}\times(K_{\mathrm{f}}+1)} & \mathbf{0}_{K_{\mathrm{z}}\times K_{\mathrm{z}}}\end{bmatrix} \qquad \text{(OA.68)}
\end{aligned}
$$

Consider now term $\mathbf{a}_2$. First notice that

$$
\mathbf{P}'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma_i^{2}\mathbf{Q}+\frac{1}{T-K-2}\mathrm{tr}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}\sigma_i^{2}\right)\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{f}}=\mathbf{0}_{K_{\mathrm{f}}}. \qquad \text{(OA.69)}
$$

Therefore, using the properties of the vec$(\cdot)$ operator and exploiting the result in (OA.65), it follows that

$$
\mathbf{a}_2 = \begin{bmatrix}0\\ \mathbf{a}_{22}\\ \mathbf{0}_{K_{\mathrm{z}}}\end{bmatrix}+o_p(1). \qquad \text{(OA.70)}
$$

setting

$$\mathbf{a}_{22} = (\mathbf{Q}' \otimes \mathbf{P}')\text{vec}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1})\right) - \frac{\text{tr}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{M}_{\tilde{D}_i}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1})\right)}{T-K-2}\mathbf{P}'\mathbf{Q}.$$

Under Assumptions OA.5(i) and OA.6(ii), using (OA.69), and recalling $\mathbf{V} = (\mathbf{Q}\otimes\mathbf{P}) - \frac{\text{vec}(\mathbb{M}_{\tilde{D}})}{T-K-2}\mathbf{Q}'\mathbf{P}$, the variance of $\mathbf{a}_{22}$ equals to

$$\begin{aligned}\text{Var}(\mathbf{a}_{22}) &= \text{E}\left[\mathbf{a}_{22}\mathbf{a}_{22}'\right] \\ &\to \left[(\mathbf{Q}' \otimes \mathbf{P}') + \mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\text{f}}\frac{\text{vec}(\mathbf{M}_{\tilde{D}})'}{T-K-2}\right]\mathbf{U}_\epsilon\left[(\mathbf{Q} \otimes \mathbf{P}) + \frac{\text{vec}(\mathbf{M}_{\tilde{D}})}{T-K-2}\bar{\boldsymbol{\delta}}_{\text{f}}'\mathbf{P}'\mathbf{P}\right] \equiv \mathbf{V}'\mathbf{U}_\epsilon\mathbf{V},\end{aligned}$$

implying that

$$\text{Var}(\mathbf{a}_2) \to \begin{bmatrix} 0 & \mathbf{0}_{K_{\text{f}}}' & \mathbf{0}_{K_{\text{z}}}' \\ \mathbf{0}_{K_{\text{f}}} & \mathbf{V}'\mathbf{U}_\epsilon\mathbf{V} & \mathbf{0}_{K_{\text{f}}\times K_{\text{z}}} \\ \mathbf{0}_{K_{\text{z}}} & \mathbf{0}_{K_{\text{z}}\times K_{\text{f}}} & \mathbf{0}_{K_{\text{z}}\times K_{\text{z}}} \end{bmatrix}.$$

Moreover, notice that $\mathbf{a}_1$ and $\mathbf{a}_2$ are (asymptotically) uncorrelated, therefore $\text{Cov}(\mathbf{a}_1, \mathbf{a}_2') \to \mathbf{0}_{(K+1)\times(K+1)}$. Consider now the term $\mathbf{a}_3$, and notice that

$$\frac{\bar{\mathbf{Z}}'\boldsymbol{\epsilon}'}{N}\mathbf{Q} = \left(\mathbf{Q}' \otimes \mathbf{J}'\right)\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_i \otimes \mathbf{z}_i) = \mathbf{H}\frac{1}{N}\sum_{i=1}^{N}\mathbf{u}_i,$$

where we set $\mathbf{u}_i \equiv \boldsymbol{\epsilon}_i \otimes \mathbf{z}_i$ and $\mathbf{H} \equiv \mathbf{Q}' \otimes \mathbb{J}'$. Under Assumptions OA.6 and OA.7, we have that

$$\text{Var}\left(\frac{\bar{\mathbf{Z}}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}\right) = \text{E}\left[\frac{\bar{\mathbf{Z}}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q}\mathbf{Q}'\frac{\boldsymbol{\epsilon}\bar{\mathbf{Z}}}{\sqrt{N}}\right] = \mathbf{H}\frac{1}{N}\sum_{i,j=1}^{N}\boldsymbol{\Sigma}_{\text{u},ij}\mathbf{H}' \to \mathbf{H}\Sigma_{\text{U}}\mathbf{H}',$$

implying that

$$\text{Var}(\mathbf{a}_3) \to \begin{bmatrix} 0 & \mathbf{0}_{K_{\text{f}}}' & \mathbf{0}_{K_{\text{z}}}' \\ \mathbf{0}_{K_{\text{f}}} & \mathbf{0}_{K_{\text{f}}\times K_{\text{f}}} & \mathbf{0}_{K_{\text{f}}\times K_{\text{z}}} \\ \mathbf{0}_{K_{\text{z}}} & \mathbf{0}_{K_{\text{z}}\times K_{\text{f}}} & \mathbf{H}\Sigma_{\text{U}}\mathbf{H}' \end{bmatrix}.$$

Finally, let us consider the covariance terms between $\mathbf{a}_2$ and $\mathbf{a}_3$, and $\mathbf{a}_1$ and $\mathbf{a}_3$. By Assumption OA.7(viii), it follows that

$$\text{Cov}\left(\mathbf{a}_{22}, \mathbf{Q}'\frac{\boldsymbol{\epsilon}\bar{\mathbf{Z}}}{\sqrt{N}}\right) = \mathbf{V}'\frac{1}{N}\sum_{i,j=1}^{N}\boldsymbol{\Sigma}_{\text{u}\epsilon,ij}\mathbf{H}' \to \mathbf{0}_{K_{\text{f}}\times K_{\text{z}}},$$

implying that $\text{Cov}(\mathbf{a}_2, \mathbf{a}_3') \to \mathbf{0}_{(K+1)\times(K+1)}$. Finally, note that

$$\text{Cov}\left(\begin{bmatrix} \frac{\mathbf{1}_N'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q} \\ \frac{\mathbf{B}'\boldsymbol{\epsilon}'}{\sqrt{N}}\mathbf{Q} \end{bmatrix}, \mathbf{Q}'\frac{\boldsymbol{\epsilon}\bar{\mathbf{Z}}}{\sqrt{N}}\right) = \frac{1}{N}\sum_{i,j=1}^{N}\sigma_{ij}\left(\mathbf{Q}'\otimes\begin{bmatrix}1\\ \boldsymbol{\beta}_i\end{bmatrix}\text{E}(\mathbf{z}_j)'\right)\mathbf{H}' \to \sigma^2\mathbf{G}\mathbf{H}'$$

setting $\mathbf{G} \equiv \left[\mathbf{Q}\otimes\boldsymbol{\mu}_{\text{z},T-1}, \quad \mathbf{Q}\otimes\boldsymbol{\Sigma}_{\text{ZB}}\right]'$, yielding

$$\text{Cov}(\mathbf{a}_1, \mathbf{a}_3') \quad \to \quad \begin{bmatrix} \mathbf{0}_{(K_\text{f}+1)\times(K_\text{f}+1)} & \sigma^2\mathbf{G}\mathbf{H}' \\ \mathbf{0}_{K_\text{z}\times(K_\text{f}+1)} & \mathbf{0}_{K_\text{z}\times K_\text{z}} \end{bmatrix}.$$

∎

The following theorem shows how to construct asymptotically valid standard errors.

Theorem OA.2 (standard errors of the locally-averaged bias-adjusted CSR OLS estimator)**.** *Under Assumptions OA.1—OA.7, $\widehat{\text{Cov}}(\mathbf{Z}_{t-1}, \boldsymbol{\gamma}_{\text{z},t-1}) = o_p(N^{-1/2})$, and the identification condition $\kappa_4 = 0$, as $N \to \infty$,*

$$\hat{\mathbf{L}}^{-1}\,\hat{\mathbf{O}}\,\hat{\mathbf{L}}^{-1'} \quad \to_p \quad \mathbf{L}^{-1}\,\mathbf{O}\,\mathbf{L}^{-1'} \tag{OA.71}$$

where

$$\hat{\mathbf{L}} \equiv \frac{1}{N}\begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{X}} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}, \quad \textit{and} \quad \hat{\mathbf{O}} \equiv \begin{bmatrix} \hat{\mathbf{U}} & \hat{\sigma}^2\hat{\mathbf{G}}\hat{\mathbf{H}}' \\ \hat{\sigma}^2\hat{\mathbf{H}}\hat{\mathbf{G}}' & \hat{\mathbf{H}}\hat{\boldsymbol{\Sigma}}_\text{U}\hat{\mathbf{H}}' \end{bmatrix}, \tag{OA.72}$$

with $\hat{\mathbf{U}} \equiv \frac{\hat{\sigma}^2}{T-1}\left[1 + (T-1)\hat{\boldsymbol{\delta}}_\text{f}^{*'}\mathbf{P}'\mathbf{P}\hat{\boldsymbol{\delta}}_\text{f}^*\right](\hat{\boldsymbol{\Sigma}}_\text{X} - \hat{\boldsymbol{\Lambda}}_1) + \begin{bmatrix} 0 & \mathbf{0}_{K_\text{f}}' \\ \mathbf{0}_{K_\text{f}} & \hat{\mathbf{V}}'\hat{\mathbf{U}}_\epsilon\hat{\mathbf{V}} \end{bmatrix}$ *and where* $\hat{\mathbf{U}}_\epsilon = \mathbf{U}_\epsilon(\kappa_4 = 0, \hat{\sigma}^4)$ *is a consistent plug-in estimator of* $\mathbf{U}_\epsilon = \mathbf{U}_\epsilon(\kappa_4, \sigma^4)$ *obtained by replacing* σ^4 *with*

$$\hat{\sigma}^4 = \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\hat{\epsilon}_{it}^4}{3\,\text{tr}\left(\overline{\mathbb{M}}_{\tilde{\text{D}}}^{(2)}\right)}, \qquad \textit{with} \quad \overline{\mathbb{M}}_{\tilde{\text{D}}}^{(2)} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{M}_{\tilde{\text{D}}_i}\odot\mathbb{M}_{\tilde{\text{D}}_i}\right), \tag{OA.73}$$

recalling $\overline{\mathbb{M}}_{\tilde{\text{D}}} = N^{-1}\sum_{i=1}^{N}\mathbb{M}_{\tilde{\text{D}}_i}$*, with* $\mathbb{M}_{\tilde{\mathbf{D}}_i} = \mathbf{I}_{T-1} - \tilde{\mathbf{D}}_i(\tilde{\mathbf{D}}_i'\tilde{\mathbf{D}}_i)^{-1}\tilde{\mathbf{D}}_i'$*,* $\tilde{\mathbf{D}}_i = (\mathbf{D}, \tilde{\mathbf{Z}}_i)$*, with* $\mathbf{D} = (\mathbf{1}_{T-1}, \mathbf{F})$*,* $\hat{\boldsymbol{\Sigma}}_\text{X} = N^{-1}\hat{\mathbf{X}}'\hat{\mathbf{X}}$*,* $\hat{\boldsymbol{\Sigma}}_\text{ZB} = N^{-1}\mathbf{Z}'\hat{\mathbf{B}}$*,* $\hat{\boldsymbol{\mu}}_{\text{z},T-1} = N^{-1}\mathbf{Z}'\mathbf{1}_N$*, and* $\hat{\boldsymbol{\Sigma}}_\text{U} \equiv \hat{\sigma}^2\boldsymbol{I}_{T-1}\otimes\mathbf{Z}'\mathbf{Z}/N$*, with* $\hat{\sigma}^2$ *defined in* (33)*, and defining*

$$\begin{aligned} \hat{\mathbf{H}} \quad &\equiv \hat{\mathbf{Q}}'\otimes\mathbb{J}', \hat{\mathbf{Q}} \equiv \frac{\mathbf{1}_{T-1}}{(T-1)} - \mathbf{P}\hat{\boldsymbol{\delta}}_\text{f}^*, \\ \hat{\mathbf{V}} \quad &\equiv (\hat{\mathbf{Q}}\otimes\mathbf{P}) - \frac{\text{vec}(\overline{\mathbb{M}}_{\tilde{\text{D}}})}{T-K-2}\hat{\mathbf{Q}}'\mathbf{P}, \\ \hat{\mathbf{G}} \quad &\equiv \left[\hat{\mathbf{Q}}\otimes\hat{\boldsymbol{\mu}}_{\text{z},T-1}, \quad \hat{\mathbf{Q}}\otimes\hat{\boldsymbol{\Sigma}}_\text{ZB}\right]'. \end{aligned}$$

Proof. By Lemma 1 and Lemma 2(i) and (ii), it follows that $\hat{\mathbf{L}} \to_p \mathbf{L}$. By part (i) of Theorem OA.1, then $\hat{\boldsymbol{\delta}}^*_{\mathrm{f}} \to_p \bar{\boldsymbol{\delta}}_{\mathrm{f}}$, implying that $\hat{\mathbf{Q}}$ is a consistent estimator of $\mathbf{Q}$. Moreover, as $N \to \infty$, $\overline{\mathbb{M}}_{\tilde{\mathrm{D}}} \to_p \mathbb{M}_{\tilde{D}}$, $\hat{\boldsymbol{\mu}}_{T-1,z} \to_p \boldsymbol{\mu}_{T-1,z}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ZB}} \to_p \boldsymbol{\Sigma}_{\mathrm{ZB}}$, and $\mathbf{Z}'\mathbf{Z}/N \to_p \boldsymbol{\Sigma}_{\mathrm{Z}}$. It follows that $\hat{\mathbf{V}} \to_p \mathbf{V}$, $\hat{\mathbf{G}} \to_p \mathbf{G}$, and $\hat{\mathbf{H}} \to_p \mathbf{H}$. Finally, a consistent estimator of $\hat{\mathbf{U}}$ requires a consistent estimate of the matrix $\mathbf{U}_\epsilon$, which can be obtained using Lemma 1(ix). This concludes the proof of Theorem OA.2. ■

OA.5.3 Anomalies with Time-Varying Premia: Global Misspecification - Asymptotics

We first establish the additional biases induced by global misspecification for the CSR OLS estimator, then construct the misspecification-robust bias-adjusted estimator of the premia (53). Finally, we show how to construct asymptotically-valid standard errors.

Proposition OA.3 (biases of CSR OLS — time-varying estimator with misspecification)**.** *Under Assumptions OA.1–OA.7 and OA.12 (listed in Appendix A.1), as $N \to \infty$,*

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{t-1} \\ \hat{\gamma}_{t-1,z} \end{bmatrix} \to_p \begin{bmatrix} \tilde{\boldsymbol{\Gamma}}_{t-1} \\ \tilde{\gamma}_{t-1,z} \end{bmatrix} + \mathbf{C}_{t-1}^{-1} \left(- \begin{bmatrix} \boldsymbol{\Lambda}_1 \\ \mathbf{0}_{K_z \times K_f+1} \end{bmatrix} \tilde{\boldsymbol{\Gamma}}_{t-1} + \begin{bmatrix} \boldsymbol{\Lambda}_{t-1,2} \\ \mathbf{0}_{K_z} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{P}' \begin{bmatrix} \boldsymbol{\theta}_{t-1,m} \\ \mathbf{0}_{T-t+1} \end{bmatrix} \\ \mathbf{0}_{K_z} \end{bmatrix} \right), \qquad \text{(OA.74)}$$

where $\mathbf{C}_{t-1}, \boldsymbol{\Lambda}_1,$ *and* $\boldsymbol{\Lambda}_{t-1,2}$ *are defined in* (OA.54) *and (OA.56), respectively, and* $\boldsymbol{\theta}_{t-1,m}$ *is defined in Assumption OA.12.*

Proof. Both parts (i) and (ii) follow by the steps adopted to proof Proposition OA.1. For part (i), one needs to consider the additional term

$$\frac{1}{N} \begin{bmatrix} \hat{\mathbf{X}}' \\ \mathbf{Z}'_{t-1} \end{bmatrix} \mathbf{m}_{t-1} = \frac{1}{N} \begin{bmatrix} 0 \\ (\hat{\mathbf{B}} - \mathbf{B})'\mathbf{m}_{t-1} \\ \mathbf{0}_{K_z} \end{bmatrix},$$

where

$$\frac{1}{N}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{m}_{t-1} = \frac{1}{N}\mathbf{P}'\boldsymbol{\epsilon}\mathbf{m}_{t-1} \to \mathbf{P}' \begin{bmatrix} \theta_{t-3,m} \\ \theta_{t-4,m} \\ \vdots \\ \theta_{0,m} \\ \mathbf{0}_{T-t+1} \end{bmatrix} \equiv \mathbf{P}' \begin{bmatrix} \boldsymbol{\theta}_{t-1,m} \\ \mathbf{0}_{T-t+1} \end{bmatrix},$$

using the property $(\mathbf{X}, \mathbf{Z}_{t-1})'\mathbf{m}_{t-1} = \mathbf{0}_{K_z+K_f+1}$. For part (ii), the result follows as $\hat{\mathbf{X}} \to_p \mathbf{X}$ as $T \to \infty$, by the assumed conditions $\mathbf{P}'\boldsymbol{\epsilon} \to_p \mathbf{0}_{K_f \times N}$ and $(\mathbf{X}, \mathbf{Z}_{t-1})$ being full-column rank. QED

To construct the bias-adjusted estimator, we proceed as follows. A natural estimator for $\boldsymbol{\theta}_{t-1,m}$ is $N^{-1}\hat{\boldsymbol{\epsilon}}_{2,t-1}'\hat{\mathbf{m}}_{t-1}^{prelim}$, with $\hat{\mathbf{m}}_{t-1}^{prelim} \equiv \mathbf{R}_t - (\hat{\mathbf{X}}, \mathbf{Z}_{t-1})(\hat{\boldsymbol{\Gamma}}_{t-1}^{prelim\prime}, \hat{\gamma}_{t-1z}^{prelim\prime})'$, setting $\boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_{2,t-1} \\ \boldsymbol{\epsilon}_{t,T-1} \end{bmatrix}$, with $\boldsymbol{\epsilon}_{2,t-1} \equiv (\boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_{t-1})', \boldsymbol{\epsilon}_{t,T-1} \equiv (\boldsymbol{\epsilon}_t, \ldots, \boldsymbol{\epsilon}_{T-1})'$, and *assuming* the availability of preliminary consistent estimators for the premia parameters $(\hat{\boldsymbol{\Gamma}}_{t-1}^{prelim\prime}, \hat{\gamma}_{t-1z}^{prelim\prime})$. Notice that $\boldsymbol{\epsilon}_{t,T-1}$ and $\mathbf{m}_{t-1}$ are mutually independent, and hence their covariance is zero. Lemma 8 shows that $N^{-1}\hat{\boldsymbol{\epsilon}}_{2,t-1}'\hat{\mathbf{m}}_{t-1}^{prelim}$ is biased but that, by adding a bias-correction term to it, one can construct the valid estimator for $\boldsymbol{\theta}_{t-1,m}$, yielding the overall bias-adjustment term for the CSR OLS estimator, as follows,

$$\begin{bmatrix} 0 \\ \mathbf{P}' \begin{bmatrix} \hat{\boldsymbol{\theta}}_{t-1,m}^{prelim} \\ \mathbf{0}_{T-t+1} \end{bmatrix} \\ \mathbf{0}_{K_z} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{2,t-1}^{(\mathrm{m})} \\ \mathbf{0}_{K_z} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{1,t-1}^{(\mathrm{m})} \\ \mathbf{0}_{K_z \times K_f+1} \end{bmatrix} \hat{\boldsymbol{\Gamma}}_{t-1}^{prelim} - \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{3,t-1}^{(\mathrm{m})} \\ \mathbf{0}_{K_z \times K_z} \end{bmatrix} \hat{\gamma}_{t-1z}^{prelim},$$

setting

$$\hat{\boldsymbol{\Lambda}}_{1,t-1}^{(\mathrm{m})} \equiv \frac{1}{N} \begin{bmatrix} \mathbf{0}_{K_\mathrm{f}+1}' \\ \mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{D}\hat{\mathrm{X}}} \end{bmatrix}, \ \hat{\boldsymbol{\Lambda}}_{2,t-1}^{(\mathrm{m})} \equiv \frac{1}{N} \begin{bmatrix} 0 \\ \mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{DR}} - \hat{\sigma}^2\mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{D}\tilde{\mathrm{D}}} \end{bmatrix}, \text{ and } \hat{\boldsymbol{\Lambda}}_{3,t-1}^{(\mathrm{m})} \equiv \frac{1}{N} \begin{bmatrix} \mathbf{0}_{K_z}' \\ \mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{DZ}} \end{bmatrix}, \tag{OA.75}$$

with $\hat{\boldsymbol{\Psi}}_{\mathrm{D}\hat{\mathrm{X}}} \equiv \begin{bmatrix} \mathbf{M}_{\mathrm{D},t-1}^{(-1)}\hat{\boldsymbol{\epsilon}}\hat{\mathbf{X}} \\ \mathbf{0}_{(T-t+1)\times(K_\mathrm{f}+1)} \end{bmatrix}$, $\hat{\boldsymbol{\Psi}}_{\mathrm{DZ}} \equiv \begin{bmatrix} \mathbf{M}_{\mathrm{D},t-1}^{(-1)}\hat{\boldsymbol{\epsilon}}\mathbf{Z}_{t-1} \\ \mathbf{0}_{(T-t+1)\times K_\mathrm{z}} \end{bmatrix}$, $\hat{\boldsymbol{\Psi}}_{\mathrm{DR}} \equiv \begin{bmatrix} \mathbf{M}_{\mathrm{D},t-1}^{(-1)}\hat{\boldsymbol{\epsilon}}\mathbf{R}_t \\ \mathbf{0}_{T-t+1} \end{bmatrix}$, and $\hat{\boldsymbol{\Psi}}_{\mathrm{D}\tilde{\mathrm{D}}} \equiv \begin{bmatrix} \mathbf{M}_{\mathrm{D},t-1}^{(-1)}\mathbb{M}_{\mathrm{D}}\boldsymbol{\imath}_{t-1,T-1} \\ \mathbf{0}_{T-t+1} \end{bmatrix}$, setting the $(t-2)\times(T-1)$ matrix $\mathbf{M}_{\mathrm{D},t-1}^{(-1)} \equiv \mathbb{M}_{11}^{-1}[\mathbf{I}_{t-2}, \mathbf{0}_{(t-2)\times(T-t+1)}]$, where $\mathbb{M}_{11}$ denotes the $(t-2)\times(t-2)$ top-left block of $\mathbb{M}_D = \mathbf{I}_{T-1} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$, where we use the partition $\mathbb{M}_D = \begin{bmatrix} \mathbb{M}_{11} & \mathbb{M}_{12} \\ \mathbb{M}_{21} & \mathbb{M}_{22} \end{bmatrix}$.

Estimator (53) is then obtained finding the 'fixed-point' solution to the system of equations

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{*(\mathrm{m})} \\ \hat{\gamma}_{\mathrm{z},t-1}^{*(\mathrm{m})} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{t-1} \\ \hat{\gamma}_{t-1,z} \end{bmatrix} - \hat{\mathbf{C}}_{t-1}^{-1} \left(- \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_1 \\ \mathbf{0}_{K_z \times K_f+1} \end{bmatrix} \hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{*(\mathrm{m})} + \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{t-1,2} \\ \mathbf{0}_{K_z} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{P}' \begin{bmatrix} \hat{\boldsymbol{\theta}}_{t-1,m}^{*} \\ \mathbf{0}_{T-t+1} \end{bmatrix} \\ \mathbf{0}_{K_z} \end{bmatrix} \right),$$

that is setting

$$\begin{bmatrix} 0 \\ \mathbf{P}' \begin{bmatrix} \hat{\boldsymbol{\theta}}_{t-1,m}^{*} \\ \mathbf{0}_{T-t+1} \end{bmatrix} \\ \mathbf{0}_{K_z} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{2,t-1}^{(\mathrm{m})} \\ \mathbf{0}_{K_z} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{1,t-1}^{(\mathrm{m})} \\ \mathbf{0}_{K_z \times K_f+1} \end{bmatrix} \hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{*(\mathrm{m})} - \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_{3,t-1}^{(\mathrm{m})} \\ \mathbf{0}_{K_z \times K_z} \end{bmatrix} \hat{\gamma}_{\mathrm{z},t-1}^{*(\mathrm{m})},$$

obtained replacing $\hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{prelim}, \hat{\gamma}_{\mathrm{z},t-1}^{prelim}$ with $\hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}^{*(\mathrm{m})}, \hat{\gamma}_{\mathrm{z},t-1}^{*(\mathrm{m})}$ into $\hat{\boldsymbol{\theta}}_{t-1,m}^{prelim}$ above.

The following theorem shows how to construct asymptotically valid standard errors.[10]

[10]We do not report the proof to Theorem OA.3, given that it follows closely the proof to Theorem 2.

Theorem OA.3 (standard errors of CSR OLS — time-varying estimator robust to misspecification)**.** *Under Assumptions OA.1–OA.7 and OA.12, and the identification condition $\kappa_4 = 0$, as $N \to \infty$,*

$$\hat{\mathbf{L}}^{(\mathrm{m})}_{t-1} \equiv \frac{1}{N}\begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N(\hat{\boldsymbol{\Lambda}}_1 + \hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{1,t-1}) & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} - N\hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{3,t-1} \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix} \to_p \mathbf{L}^{(\mathrm{m})}_{t-1} \text{ and}$$

$$\hat{\boldsymbol{\Omega}}^{(\mathrm{m})}_{t-1} \equiv \hat{\boldsymbol{\Omega}}_{1,t-1} + \hat{\boldsymbol{\Omega}}_{2,t-1} + \hat{\boldsymbol{\Omega}}_{3,t-1} + \hat{\boldsymbol{\Omega}}'_{3,t-1} + \hat{\boldsymbol{\Omega}}_{4,t-1} + \hat{\boldsymbol{\Omega}}'_{4,t-1} \to_p \boldsymbol{\Omega}^{(\mathrm{m})}_{t-1},$$

setting

$$\hat{\boldsymbol{\Omega}}_{1,t-1} \equiv \hat{\sigma}^2\hat{\sigma}_{t-1mm}\begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & \mathbf{P}'\begin{bmatrix}\mathbb{M}_{11}^{-1}\mathbb{M}_{12} \\ \mathbf{I}_{T-t+1}\end{bmatrix}\begin{bmatrix}\mathbb{M}'_{12}\mathbb{M}_{11}^{-1} & \mathbf{I}_{T-t+1}\end{bmatrix}\mathbf{P} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix},$$

$$\hat{\boldsymbol{\Omega}}_{2,t-1} \equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & \hat{\mathbf{A}}_{t-1}\hat{\mathbf{U}}_{\epsilon}\hat{\mathbf{A}}'_{t-1} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix},$$

$$\hat{\boldsymbol{\Omega}}_{3,t-1} \equiv$$

$$\begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & -\mathbf{P}'\begin{bmatrix}\mathbb{M}_{11}^{-1}\mathbb{M}_{12} \\ \mathbf{I}_{T-t+1}\end{bmatrix}\begin{bmatrix}\mathbf{0}_{T-t+1\times(t-2)^2} & (\hat{\boldsymbol{\theta}}'_{t-1,m}\otimes\hat{\sigma}^2\mathbf{I}_{T-1}) & (\hat{\sigma}^2\mathbf{I}_{T-1}\otimes\hat{\boldsymbol{\theta}}'_{t-1,m}) & \mathbf{0}_{T-t+1\times(T-t+1)^2}\end{bmatrix}\hat{\mathbf{V}}_{t-1} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix}, \text{ and}$$

$$\hat{\boldsymbol{\Omega}}_{4,t-1} =$$

$$\begin{bmatrix} 0 & \mathbf{0}'_{K_f} & \mathbf{0}'_{K_z} \\ \mathbf{0}_{K_f} & -\mathbf{P}'\begin{bmatrix}\mathbb{M}_{11}^{-1}\mathbb{M}_{12} \\ \mathbf{I}_{T-t+1}\end{bmatrix}\begin{bmatrix}\mathbf{0}_{T-t+1\times(t-2)^2} & (\hat{\boldsymbol{\theta}}'_{t-1,m}\otimes\hat{\sigma}^2\mathbf{I}_{T-1}) & (\hat{\sigma}^2\mathbf{I}_{T-1}\otimes\hat{\boldsymbol{\theta}}'_{t-1,m}) & \mathbf{0}_{T-t+1\times(T-t+1)^2}\end{bmatrix}\hat{\mathbf{A}}_{t-1'} & \mathbf{0}_{K_f\times K_z} \\ \mathbf{0}_{K_z} & \mathbf{0}_{K_z\times K_f} & \mathbf{0}_{K_z\times K_z}\end{bmatrix},$$

setting CHECK IF $\bar{\mathbf{M}}_{\tilde{D}}$ or $\mathbb{M}_D$ in $\hat{\boldsymbol{\theta}}_{t-1,m}$

$$\hat{\boldsymbol{\theta}}_{t-1,m} \equiv \mathbf{M}^{(-1)}_{\mathrm{D},t-1}\left(\frac{1}{N}\hat{\boldsymbol{\epsilon}}'\hat{\mathbf{m}}_{t-1} - \hat{\sigma}^2\mathbb{M}_D\boldsymbol{\iota}_{t-1,T-1}\right), \text{ where } \hat{\mathbf{m}}_{t-1} \equiv \mathbf{R}_t - (\hat{\mathbf{X}}, \mathbf{Z}_{t-1})(\hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1}, \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})\prime}_{\mathrm{z},t-1})', \tag{OA.76}$$

$$\hat{\sigma}_{t-1mm} \equiv \frac{\hat{\mathbf{m}}'_{t-1}\hat{\mathbf{m}}_{t-1}}{N} - \hat{\sigma}^2\hat{\mathbf{Q}}'_{t-1}\hat{\mathbf{Q}}_{t-1} + 2\hat{\boldsymbol{\delta}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1}\mathbf{P}'\begin{bmatrix}\hat{\boldsymbol{\theta}}_{t-1,m} \\ \mathbf{0}_{T-t+1}\end{bmatrix}, \tag{OA.77}$$

and

$$\hat{\mathbf{A}}_{t-1,T-1} = -\mathbf{P}'\begin{bmatrix}\left(\hat{\mathbf{Q}}'_{t-1}\otimes\mathbb{M}_{11}^{-1}(\mathbf{I}_{t-2}, \mathbf{0}_{t-2\times T-t+1})\overline{\mathbb{M}}_{\tilde{D}}\right)\left(\mathbf{I}^2_{T-1} - vec(\mathbf{I}_{T-1})\frac{vec'(\overline{\mathbb{M}}_{\tilde{D}})}{T-K_f-K_z-2}\right) \\ \mathbf{0}_{T-t+1\times(T-1)^2}\end{bmatrix},$$

where we recall $\boldsymbol{M}^{(-1)}_{D,t-1} = \mathbb{M}_{11}^{-1}[\mathbf{I}_{t-2}, \mathbf{0}_{(t-2)\times(T-t+1)}]$, and all the other quantities are defined in Theorem 2.

OA.6 Two-Pass Methodology and Anomalies: the Conventional Approach - Asymptotics

This section provides the formal derivations of the main results established in Section 4 of the paper. Particularly, we derive the limiting behavior of the *conventional* anomaly premia estimator (see Fama and French (2008)), which consists of running, for every time period, a CSR OLS regression of asset returns $\mathbf{R}_t$ on the anomaly variables $\mathbf{Z}_{t-1}$, yielding the time-varying estimator in (13). The T premia estimates are then averaged across time, resulting in the average premia estimator (14). This approach coincides exactly with the second step of the two-pass Fama and MacBeth (1973) procedure where, however, one *excludes* the betas from the model, to avoid EIV-related issues.

In the following, we analyze the limiting behavior of the conventional estimators in (13) and (14), together with the corresponding conventional t-ratio in (15), under three different sampling schemes: (i) when $T \to \infty$ and N is fixed, (ii) when $N \to \infty$ and T is fixed, and (iii) when both N and T are allowed to diverge. We first present the results in the univariate-regression setting, that is when one considers only one anomaly at the time ($K_{\mathrm{z}} = 1$) in the regression model. Extension to the multivariate case are presented in the final remarks.

Consider the following model

$$\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\gamma_{\mathrm{z},t-1} + \mathbf{B}\boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t, \tag{OA.78}$$

with the objective of estimating the anomaly premium $\gamma_{\mathrm{z},t-1}$. Following Fama and French (2008), at each point in time, we run a cross-sectional OLS regression of (OA.78), using the anomaly $\mathbf{Z}_{t-1}$, but *excluding* the betas $\mathbf{B}$ from the model. This gives the time-varying OLS estimator

$$\tilde{\gamma}_{\mathrm{z},t-1} \equiv \frac{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{R}_t}{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}, \tag{OA.79}$$

which satisfies

$$\tilde{\gamma}_{\mathrm{z},t-1} = \gamma_{\mathrm{z},t-1} + \frac{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\left(\mathbf{B}\boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t\right)}{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}.$$

Then, averaging the time-varying estimator across time yields the *average* OLS estimator

$$\bar{\tilde{\gamma}}_{\mathrm{z}} \equiv \frac{1}{T-1}\sum_{t=2}^{T}\tilde{\gamma}_{\mathrm{z},t-1} = \frac{1}{T}\sum_{t=1}^{T}\frac{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{R}_t}{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1}} \tag{OA.80}$$

which satisfies

$$\bar{\bar{\gamma}}_{\mathrm{z}} = \bar{\gamma}_{\mathrm{z}} + \frac{1}{T-1}\sum_{t=2}^{T}\left(\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}(\mathbf{B}\boldsymbol{\delta}_{\mathrm{f},t-1}+\boldsymbol{\epsilon}_t)}{\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}\right), \tag{OA.81}$$

with $\bar{\gamma}_{\mathrm{z}} = \frac{1}{T-1}\sum_{t=2}^{T}\gamma_{\mathrm{z},t-1}$. and with estimated variance (squared standard error)

$$\widehat{\mathrm{Var}}[\bar{\bar{\gamma}}_z] = \frac{1}{(T-1)^2}\sum_{t=2}^{T}(\tilde{\gamma}_{\mathrm{z},t-1} - \bar{\bar{\gamma}}_{\mathrm{z}})^2 = \frac{\tilde{\Sigma}_{\gamma_{\mathrm{z}}}}{(T-1)}, \tag{OA.82}$$

with $\tilde{\Sigma}_{\gamma_{\mathrm{z}}}$ defined in (22) denoting the sample average of the time-varying estimates.

In the following, we want to derive the limiting behaviour of the time-varying estimator (OA.79) and locally-averaged estimator (OA.80) under the three sampling schemes mentioned above. Before introducing our results, we state below the set of assumptions required (not necessarily at the same time) to derive our results.

Assumption OA.14 (Finite-N Orthogonality)**.** For a given fixed N, the $\boldsymbol{\beta}_i$ and $Z_{i,t-1}$ are cross-sectionally orthogonal in the sample:

$$\mathbf{B}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1} = \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{z}}}.$$

Assumption OA.15 (Large-N Orthogonality)**.** As $N \to \infty$, the $\boldsymbol{\beta}_i$ and $Z_{i,t-1}$ are asymptotically cross-sectionally orthogonal:

$$\frac{1}{N}\mathbf{B}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1} \to_p \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{z}}}.$$

Assumption OA.16 (Uncorrelatedness of risk factors and asset returns)**.** For every $t = 1, ..., T$, the risk factors $\mathbf{f}_t$ and the asset returns $\mathbf{R}_t$ are uncorrelated:

$$\mathbf{B} = \mathbf{0}_{N\times K_{\mathrm{f}}}.$$

Assumption OA.17 (Constant anomaly premia)**.** The anomaly premia $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ are constant over time

$$\boldsymbol{\gamma}_{\mathrm{z},t-1} = \boldsymbol{\gamma}_{\mathrm{z}}, \qquad \textit{for every } t = 2, .., T$$

OA.6.1 The large-T–fixed-N case

Under the large-T–fixed-N sampling scheme, clearly the time-varying OLS estimator in (OA.79) remains unchanged and no meaningful asymptotic property can be established. The results regarding the locally-averaged OLS estimator in (OA.80) are summarized in the following theorem.[11]

Theorem OA.4. *(Locally-Averaged OLS Estimator - large T–fixed N) Assume the* finite-N *orthogonality Assumption* (OA.14)*, and the following regularity conditions:* $E[\boldsymbol{\epsilon}_t|\mathbf{Z}_{t-1}] = \mathbf{0}_N$, $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathbf{Z}_{t-1}] = \boldsymbol{\Sigma}$*, and* $\frac{1}{T-1}\sum_{t=2}^{T}\frac{(\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\boldsymbol{\Sigma}\,\mathbb{M}_{1_N}\mathbf{Z}_{t-1})}{(\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1})^2} \to_p V_N$*, with* $0 < V_N < \infty$*. Then, as* $T \to \infty$ *and* N *is fixed,*

(i) Let $\bar{\gamma}_{\mathrm{z}}^0 = \lim_{T\to\infty}\bar{\gamma}_{\mathrm{z}}$, then

$$\bar{\tilde{\gamma}}_{\mathrm{z}} \to_p \bar{\gamma}_{\mathrm{z}}^0$$

(ii) When, in addition, $\frac{1}{\sqrt{T-1}}\sum_{t=2}^{T}\mathbf{c}_{t-1}'\boldsymbol{\epsilon}_t \to_d \mathcal{N}(0, V_N)$, setting $\mathbf{c}_{t-1} \equiv \frac{\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}{(\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1})}$, then

$$\sqrt{T}(\bar{\tilde{\gamma}}_{\mathrm{z}} - \bar{\gamma}_{\mathrm{z}}) \to_d \mathcal{N}(0, V_N).$$

(iii) When, in addition, $\frac{1}{T-1}\sum_{t=2}^{T}(\gamma_{\mathrm{z},t-1} - \bar{\gamma}_{\mathrm{z}})^2 \to \sigma_{\gamma_{\mathrm{z}}}^2$, then

$$\tilde{\Sigma}_{\gamma_{\mathrm{z}}} = \frac{1}{T-1}\sum_{t=2}^{T}(\tilde{\gamma}_{\mathrm{z},t-1} - \bar{\tilde{\gamma}}_{\mathrm{z}})^2 \to_p \sigma_{\gamma_{\mathrm{z}}}^2 + V_N.$$

Proof. Parts (i) and (ii) follow immediately from (OA.81) and the assumptions made above. To prove part (iii), notice that, using (OA.80) and (OA.81), we have that

$$\tilde{\gamma}_{\mathrm{z},t-1} - \bar{\tilde{\gamma}}_{\mathrm{z}} = \gamma_{\mathrm{z},t-1} - \bar{\gamma}_{\mathrm{z}} + \frac{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\boldsymbol{\epsilon}_t}{\mathbf{Z}_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}} - \frac{1}{T-1}\sum_{s=2}^{T}\frac{\mathbf{Z}_{s-1}'\mathbb{M}_{1_N}\boldsymbol{\epsilon}_s}{\mathbf{Z}_{s-1}\mathbb{M}_{1_N}\mathbf{Z}_{s-1}},$$

implying that

$$(\tilde{\gamma}_{\mathrm{z},t-1} - \bar{\tilde{\gamma}}_{\mathrm{z}})^2 = (\gamma_{\mathrm{z},t-1} - \bar{\gamma}_{\mathrm{z}})^2 + \left(\frac{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\boldsymbol{\epsilon}_t}{\mathbf{Z}_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}\right)^2 + o_p(1).$$

Therefore, using the definition in (OA.82) and the assumptions above, we get

$$\begin{aligned}
(T-1)\widehat{\mathrm{Var}}(\tilde{\gamma}_{\mathrm{z},t-1}) &= \frac{1}{T-1}\sum_{t=2}^{T}(\tilde{\gamma}_{\mathrm{z},t-1} - \bar{\tilde{\gamma}}_{\mathrm{z}})^2 \\
&= \frac{1}{T-1}\sum_{t=2}^{T}(\gamma_{\mathrm{z},t-1} - \bar{\gamma}_{\mathrm{z}})^2 + \frac{1}{T-1}\sum_{t=2}^{T}\frac{\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}{(\mathbf{Z}_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1})^2} + o_p(1) \\
&\to_p \sigma_{\gamma_{\mathrm{z}}}^2 + V_N
\end{aligned}$$

■

[11] We continue to define (OA.80) as locally-averaged for consistency with our definition even though, in this case, we let $T \to \infty$.

Remark OA.20. Unless Assumption OA.17 is satisfied, namely $\gamma_{z,t-1}$ is constant, the conventional standard error of the average OLS estimator from the Fama and MacBeth (1973) regression contains an upward bias given by $\sigma^2_{\gamma_z} > 0$. This implies that the associated t-ratio would be smaller than what it should, leading to potential under-rejections.

Remark OA.21. Theorem OA.4 easily extends to the multivariate case of $K_z > 1$. In this case:

(i) Let $\bar{\boldsymbol{\gamma}}^0_z = \lim_{T\to\infty} \bar{\boldsymbol{\gamma}}_z$, with $\bar{\boldsymbol{\gamma}}_z = \frac{1}{T-1}\sum_{t=2}^T \boldsymbol{\gamma}_{z,t-1}$, then

$$\bar{\tilde{\boldsymbol{\gamma}}}_z \to_p \bar{\boldsymbol{\gamma}}^0_z$$

(ii) Assuming that $\frac{1}{T-1}\sum_{t=2}^T \left(\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}\right)^{-1} \mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\boldsymbol{\Sigma}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}\left(\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}\right)^{-1} \to_p \mathbf{V}_N$, with $\mathbf{V}_N$ being a symmetric and positive-definite matrix, and

$\frac{1}{\sqrt{T-1}}\sum_{t=2}^T \mathbf{C}'_{t-1}\boldsymbol{\epsilon}_t \to_d \mathcal{N}(\mathbf{0}_{K_z}, \mathbf{V}_N)$, setting $\mathbf{C}_{t-1} \equiv \mathbb{M}_{1_N}\mathbf{Z}_{t-1}(\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1})^{-1}$, then

$$\sqrt{T}\left(\bar{\tilde{\boldsymbol{\gamma}}}_z - \bar{\boldsymbol{\gamma}}_z\right) \to_d \mathcal{N}\left(\mathbf{0}_{K_z}, \mathbf{V}_N\right).$$

(iii) When, in addition, $\frac{1}{T-1}\sum_{t=2}^T \left(\boldsymbol{\gamma}_{z,t-1} - \bar{\boldsymbol{\gamma}}_z\right)\left(\boldsymbol{\gamma}_{z,t-1} - \bar{\boldsymbol{\gamma}}_z\right)' \to_p \boldsymbol{\Sigma}_{\gamma_z}$, with $\boldsymbol{\Sigma}_{\gamma_z}$ being a symmetric and positive-definite matrix, then

$$(T-1)\widehat{\text{Var}}\left[\bar{\tilde{\boldsymbol{\gamma}}}_z\right] = \frac{1}{T-1}\sum_{t=2}^T \left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \bar{\tilde{\boldsymbol{\gamma}}}_z\right)\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \bar{\tilde{\boldsymbol{\gamma}}}_z\right)' \to_p \boldsymbol{\Sigma}_{\gamma_z} + \mathbf{V}_N$$

OA.6.2 The fixed-T–large-N case

We now establish the asymptotic properties of the conventional estimators under the fixed-T–large-N setting. Under this scheme, now the time-varying OLS estimator (OA.79) does change (with N) and one can study its limiting behaviour. Therefore, in the next theorem, we summarize the main results regarding both the conventional time-varying and average OLS estimators.

Theorem OA.5. *(Locally-Averaged and Time-Varying OLS Estimators - fixed T – large N) Assume that the* large-N *orthogonality condition in Assumption* (OA.15) *holds. Assume also that the following regularity conditions are satisfied:* $E[\boldsymbol{\epsilon}_t|\mathbf{Z}_{t-1}] = \mathbf{0}_N, E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t|\mathbf{Z}_{t-1}] = \boldsymbol{\Sigma}$, $N^{-1}\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1} \to_p a_{t-1} > 0$, *and* $N^{-1}(\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\boldsymbol{\Sigma}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}) \to_p \sigma^2 a_{t-1}$, *with* $\mathbf{1}'_N\boldsymbol{\Sigma}\mathbf{1}_N/N \to \sigma^2$. *Then, as* $N \to \infty$ *and* T *is fixed,*

(i)

$$\tilde{\gamma}_{\text{z},t-1} \quad \to_p \quad \gamma_{\text{z},t-1} \qquad \textit{and} \qquad \bar{\tilde{\gamma}}_{\text{z}} \quad \to_p \quad \bar{\gamma}^0_{\text{z}}.$$

(ii) Let $V_{t-1} = \sigma^2 / a_{t-1}$. *When, in addition, Assumption OA.15 is strengthened with* $\frac{1}{\sqrt{N}} \mathbf{B}' \mathbb{M}_{1_N} \mathbf{Z}_{t-1} \to_p \mathbf{0}_{K_\text{f} \times K_\text{z}}$, *and assuming that* $\frac{1}{\sqrt{N}} \mathbf{Z}'_{t-1} \mathbb{M}_{1_N} \boldsymbol{\epsilon}_t \to_d \mathcal{N}(0, \sigma^2 a_{t-1})$, *then*

$$\sqrt{N} \left(\tilde{\gamma}_{\text{z},t-1} - \gamma_{\text{z},t-1} \right) \to_d \mathcal{N} \left(0, V_{t-1} \right) \qquad \textit{and}$$

$$\sqrt{N} \left(\bar{\tilde{\gamma}}_{\text{z}} - \bar{\gamma}_{\text{z}} \right) \to_d \mathcal{N} \left(0, \bar{V} \right) \qquad \textit{with} \quad \bar{V} = \frac{1}{(T-1)^2} \sum_{t=2}^{T} V_{t-1}.$$

(iii)

$$\widehat{\text{Var}}[\bar{\tilde{\gamma}}_{\text{z}}] = \frac{1}{(T-1)^2} \sum_{t=2}^{T} (\tilde{\gamma}_{\text{z},t-1} - \bar{\tilde{\gamma}}_{\text{z}})^2 \quad \to_p \quad \frac{1}{(T-1)^2} \sum_{t=2}^{T} (\gamma_{\text{z},t-1} - \bar{\gamma}_{\text{z}})^2.$$

Proof. Parts (i) and (ii) follow immediately from (OA.80) and (OA.81), together with the assumptions made above. To prove part (iii), notice that, using (OA.80) and (OA.81), we have that

$$\tilde{\gamma}_{\text{z},t-1} - \bar{\tilde{\gamma}}_{\text{z}} = \gamma_{\text{z},t-1} - \bar{\gamma}_{\text{z}} + \frac{\mathbf{Z}'_{t-1} \mathbb{M}_{1_N} (\mathbf{B} \boldsymbol{\delta}_{\text{f},t-1} + \boldsymbol{\epsilon}_t)}{\mathbf{Z}_{t-1} \mathbb{M}_{1_N} \mathbf{Z}_{t-1}} - \frac{1}{T-1} \sum_{s=2}^{T} \frac{\mathbf{Z}'_{s-1} \mathbb{M}_{1_N} (\mathbf{B} \boldsymbol{\delta}_{\text{f},s-1} + \boldsymbol{\epsilon}_s)}{\mathbf{Z}_{s-1} \mathbb{M}_{1_N} \mathbf{Z}_{s-1}}$$

Then, under the assumptions stated above, as $N \to \infty$

$$\frac{1}{(T-1)^2} \sum_{t=2}^{T} \left(\tilde{\gamma}_{\text{z},t-1} - \bar{\tilde{\gamma}}_{\text{z}} \right)^2 \quad = \quad \frac{1}{(T-1)^2} \sum_{t=2}^{T} \left(\gamma_{\text{z},t-1} - \bar{\gamma}_{\text{z}} \right)^2 + o_p(1).$$

■

Remark OA.22. Theorem OA.5 (iii) shows that the sample variance of the time-varying OLS estimates converges to a positive constant, different from V_{t-1} or $\bar{V}$. However, one can still obtain a consistent estimation of the asymptotic variance of both $\tilde{\gamma}_{\text{z},t-1}$ and $\bar{\tilde{\gamma}}_{\text{z},t-1}$, by imposing further orthogonality conditions, as we show below. Assume that Assumption OA.16 holds, so that risk factors and returns are orthogonal to each other. Let

$$\tilde{\sigma}^2_{zt} = \frac{\tilde{\boldsymbol{\epsilon}}'_t \tilde{\boldsymbol{\epsilon}}_t}{N - K_\text{z} - 1} \tag{OA.83}$$

with $\tilde{\boldsymbol{\epsilon}}_t = \boldsymbol{R}_t - \mathbf{1}_N \bar{R}_t - (\boldsymbol{Z}_{t-1} - \mathbf{1}_N \bar{Z}_{t-1})\tilde{\gamma}_{z,t-1}$. Then, under the above assumptions, $\tilde{\sigma}_{zt}^2 - \sigma^2 \to_p 0$, implying that

$$\tilde{V}_{t-1} \equiv \frac{\tilde{\sigma}_{zt}^2}{\frac{1}{N}\left(\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1}\right)} \to_p V_{t-1} \qquad \text{and} \qquad \frac{1}{(T-1)^2}\sum_{t=2}^{T}\tilde{V}_{t-1} \to_p \bar{V}.$$

Remark OA.23. Theorem (OA.5) extends to the multivariate case ($K_z > 1$) as follows.

(i) Let $\bar{\boldsymbol{\gamma}}_z \equiv \frac{1}{T-1}\sum_{t=2}^{T}\boldsymbol{\gamma}_{z,t-1}$. Then, under the same assumptions of Theorem (OA.5),

$$\tilde{\boldsymbol{\gamma}}_{z,t-1} \;\to_p\; \boldsymbol{\gamma}_{z,t-1} \qquad \text{and} \qquad \bar{\tilde{\boldsymbol{\gamma}}}_z \;\to_p\; \bar{\boldsymbol{\gamma}}_z^0.$$

(ii) Let $\boldsymbol{V}_{t-1} = \sigma^2 \boldsymbol{A}_{t-1}^{-1}$, with $\frac{1}{N}\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1} \to_p \mathbf{A}_{t-1}$ as $N \to \infty$. Then, under the same assumptions of Theorem (OA.5),

$$\sqrt{N}\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \boldsymbol{\gamma}_{z,t-1}\right) \to_d \mathcal{N}\left(\mathbf{0}_{K_z}, \boldsymbol{V}_{t-1}\right) \qquad \text{and}$$

$$\sqrt{N}\left(\bar{\tilde{\boldsymbol{\gamma}}}_z - \bar{\boldsymbol{\gamma}}_z\right) \to_d \mathcal{N}\left(\mathbf{0}_{K_z}, \bar{\boldsymbol{V}}\right) \qquad \text{with} \quad \bar{\boldsymbol{V}} = \frac{1}{(T-1)^2}\sum_{t=2}^{T}\boldsymbol{V}_{t-1}.$$

(iii)

$$(T-1)\widehat{\text{Var}}\left[\bar{\tilde{\boldsymbol{\gamma}}}_z\right] \to_p \frac{1}{T-1}\sum_{t=2}^{T}\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \bar{\tilde{\boldsymbol{\gamma}}}_z\right)\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \bar{\tilde{\boldsymbol{\gamma}}}_z\right)'$$

OA.6.3 The large-T–large-N case

In this section we generalize all the above results to the case of both $N, T \to \infty$. We show below that the limiting properties of the time-varying OLS estimator $\tilde{\gamma}_{z,t-1}$ remain the same as the ones obtained in the fixed-T–large-N case described in Section OA.6.2. Similar results hold also for the locally-averaged estimator, even though it now benefits from the faster rate of convergence, given that both N and T are now allowed to diverge jointly. The main results are summarized in the following theorem.

Theorem OA.6. *(Locally-Averaged and Time-Varying OLS Estimators - large-N–large-T) Assume the all the conditions stated in Theorems OA.5 and OA.4 are satisfied. In addition, assume that $(T-1)^{-1}\sum_{t=2}^{T}\frac{\sigma^2}{a_{t-1}} = (T-1)^{-1}\sum_{t=2}^{T}V_{t-1} \to_p \bar{\mathcal{V}}$. Then, as $N, T \to \infty$,*

(i)

$$\tilde{\gamma}_{z,t-1} \to_p \gamma_{z,t-1} \qquad \textit{and} \qquad \bar{\tilde{\gamma}}_z \to_p \bar{\gamma}_z^0.$$

(ii)

$$\sqrt{N}\left(\tilde{\gamma}_{z,t-1} - \gamma_{z,t-1}\right) \to_d \mathcal{N}\left(0, V_{t-1}\right) \qquad \textit{and}$$

$$\sqrt{NT}\left(\bar{\tilde{\gamma}}_z - \bar{\gamma}_z\right) \to_d \mathcal{N}\left(0, \bar{\mathcal{V}}\right).$$

(iii)

$$T\widehat{\mathrm{Var}}\left[\bar{\tilde{\gamma}}_z\right] = \frac{1}{T-1}\sum_{t=2}^{T}(\tilde{\gamma}_{z,t-1} - \bar{\tilde{\gamma}}_z)^2 \to_p \boldsymbol{\Sigma}^2_{\gamma_z}.$$

Proof. The proof follows immediately by combining all the results obtained in the above theorems.

Remark OA.24. Theorem OA.6 (iii) shows that, even in the case of both $N, T \to \infty$, the sample variance of the time-varying OLS estimates converges to a positive quantity, different from $\bar{\mathcal{V}}$ and, therefore, it would not be appropriate to use it for inferential conclusions on $\tilde{\gamma}_{z,t-1}$ or $\bar{\tilde{\gamma}}_z$. However, it is still possible to obtain a consistent estimation of the asymptotic variance of the estimators. Let $\mathrm{tr}(\cdot)$ denote the trace operator and define, for $\tilde{\boldsymbol{\epsilon}} = (\tilde{\boldsymbol{\epsilon}}_2, \cdots, \tilde{\boldsymbol{\epsilon}}_T)'$,

$$\tilde{\sigma}^2 = \mathrm{tr}\frac{\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}'}{(T-1)(N-K_z-1)}$$

Then, under the assumptions made above, $\tilde{\sigma}^2 \to_p \sigma^2$, implying that

$$\tilde{\mathcal{V}}_{t-1} \equiv \frac{\tilde{\sigma}^2}{\frac{1}{N}\left(\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}\right)} \to_p V_{t-1} \qquad \text{and} \qquad \frac{1}{T-1}\sum_{t=2}^{T}\tilde{\mathcal{V}}_{t-1} \to_p \bar{\mathcal{V}}.$$

Remark OA.25. Theorem OA.6 extends to the multivariate case ($K_z > 1$) as follows.

(i)

$$\tilde{\boldsymbol{\gamma}}_{z,t-1} \to_p \boldsymbol{\gamma}_{z,t-1} \qquad \text{and} \qquad \bar{\tilde{\boldsymbol{\gamma}}}_z \to_p \bar{\boldsymbol{\gamma}}_z^0.$$

(ii) Let $\frac{\sigma^2}{(T-1)}\sum_{t=2}^{T}\left(\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}}{N}\right)^{-1} \to_p \bar{\mathbf{V}}$. Then

$$\sqrt{N}\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \boldsymbol{\gamma}_{z,t-1}\right) \to_d \mathcal{N}\left(\mathbf{0}_{K_z}, \boldsymbol{V}_{t-1}\right) \qquad \text{and}$$

$$\sqrt{NT}\left(\bar{\tilde{\boldsymbol{\gamma}}}_z - \bar{\boldsymbol{\gamma}}_z\right) \to_d \mathcal{N}\left(\mathbf{0}_{K_z}, \bar{\mathbf{V}}\right).$$

(iii)

$$(T-1)\widehat{\mathrm{Var}}\left[\bar{\tilde{\boldsymbol{\gamma}}}_z\right] = \frac{1}{T-1}\sum_{t=2}^{T}\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \bar{\tilde{\boldsymbol{\gamma}}}_z\right)\left(\tilde{\boldsymbol{\gamma}}_{z,t-1} - \bar{\tilde{\boldsymbol{\gamma}}}_z\right)' \to_p \boldsymbol{\Sigma}_{\gamma_z}$$

OA.7 Monte Carlo Experiments

VALENTINA - to replace-integrate the following.

OA.7.1 Premia Estimators: Finite-Sample Performance

In this section, we undertake a Monte Carlo simulation experiment to study the empirical performance of the locally-averaged bias-adjusted estimator (40)

The return-generating process is given by

$$R_{it} \quad = \quad \gamma_0 + Z_{1i,t-1}\gamma_{z_1} + Z_{2i,t-1}\gamma_{z_2} + \beta_i(\gamma_1 + f_t - E[f_t]) + \epsilon_{it}. \tag{OA.84}$$

We consider balanced panels with a time-series dimension of $T = 36$ and $T = 72$ observations. Specifically, f_t in (OA.84) is the excess market return (from Kenneth French's website) from January 2013 to December 2015 for T=36, and from January 2011 to December 2015 for T=72. In addition, $E[f_t]$ in (OA.84) is set equal to the time-series mean of f_t over the two sample periods 2013-2015 and 2011-2015, when performing the analysis for $T = 36$ and $T = 72$, respectively. To obtain representative values for γ_0, γ_1 and β_i in (OA.84), we employ a cross-section of 1,000 stocks from CRSP database in addition to the excess market return. Based on this balanced panel of 1,000 stock returns and the excess market return, for each time-series sample size, we compute the OLS estimates of β_i, γ_0, and γ_1 and we set them in (OA.84).

For the anomalies $Z_{1i,t-1}$ and $Z_{2i,t-1}$ in (OA.84), we first use data on two firms' characteristics, namely the book-to-market ratio ($Z^{\dagger}_{1i,t-1}$) and the asset growth ($Z^{\dagger}_{2i,t-1}$), over the two (lagged) sample periods from December 2012 to Novemeber 2015 (for $T = 36$) and from December 2010 to November 2015 (for $T = 72$). We then orthogonalize both $Z^{\dagger}_{1i,t-1}$ and $Z^{\dagger}_{2i,t-1}$ with the market factor f_t, that is we derive

$$\boldsymbol{Z}_{1i} \quad = \quad \boldsymbol{M}_D \boldsymbol{Z}^{\dagger}_{1i} + \boldsymbol{1}_{T-1}\bar{Z}^{\dagger}_{1i} \tag{OA.85}$$

$$\boldsymbol{Z}_{2i} \quad = \quad \boldsymbol{M}_D \boldsymbol{Z}^{\dagger}_{2i} + \boldsymbol{1}_{T-1}\bar{Z}^{\dagger}_{2i} \tag{OA.86}$$

where $\boldsymbol{Z}_{1i} = [Z_{1i,1}, ..., Z_{1i,T-1}]'$, $\boldsymbol{Z}_{2i} = [Z_{2i,1}, ..., Z_{2i,T-1}]'$, $\boldsymbol{M}_D = \boldsymbol{I}_{T-1} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$, with $\boldsymbol{D} = (\boldsymbol{1}_{T-1}, \boldsymbol{f})$ and where $\bar{Z}^{\dagger}_{ik} = \frac{1}{T-1}\sum_{t=1}^{T-1} Z^{\dagger}_{ki,t}$, with $k = 1, 2$. In this way we ensure that

$\text{Cov}(Z_{1i,t-1}, f_t) = \text{Cov}(Z_{2i,t-1}, f_t) = 0$ and that the sample averages of both $\boldsymbol{Z}_{1i}$ and $\boldsymbol{Z}_{2i}$ match the ones of the original data. We then use (OA.85) and (OA.86) in the data generating process in (OA.84). In our simulation design, both the factor and the anomalies realizations are taken as given and kept fixed throughout.

To set the values of γ_{z_1} and γ_{z_2} , we consider three different cases. In the first case, we set both the parameters equal to zero, i.e. $\gamma_{z_1} = \gamma_{z_2} = 0$. In the second case, we set $\gamma_{z_2} = 0$, while we compute the estimate of γ_{z_1} using our bias-adjusted estimator in Section **??**. The third case considers $\gamma_{z_1} \neq \gamma_{z_2} \neq 0$, where γ_{z_1} is the same of the previous case and γ_{z_2} has been calibrated by estimating the regression of the stock returns on the excess market factor and asset growth anomaly.

The calibration of the error term ϵ_{it} in (OA.84) is a more delicate task and is described in the next two subsections. In all the simulation experiments, we consider cross-sections of $N = 100, 500$ and $1,000$ stocks. All the results are based on 3,000 Monte Carlo replications.

Case (i): $\mathbf{Z}_t$ *and* ϵ_t *uncorrelated*

We start by considering the simplest case in which we assume that $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{T-1})$ and that it is also uncorrelated with both $Z_{1i,t-1}$ and $Z_{2i,t-1}$. We calibrate the parameter σ^2 using the estimator in (33) applied to our data of stock returns, excess market factor and the two anomalies. Tables II, III and IV report the percentage bias (Bias %) and root mean squared error (RMSE) of the bias-adjusted CSR OLS estimator (**??**) for the three cases $\gamma_{z_1} = \gamma_{z_2} = 0$ (Table II), $\gamma_{z_1} \neq 0, \gamma_{z_2} = 0$ (Table III), and $\gamma_{z_1} \neq \gamma_{z_2} \neq 0$ (Table IV) . Panels A and B are for the cases of $T = 36$ and $T = 72$, respectively.

In Tables V, IX and VII we report the empirical rejection rates of the t-test. The null hypothesis is that the parameter of interest is equal to its true value. The results are reported for different levels of significance (10%, 5% and 1%) and for different values of N and T. The t-statistics are derived using the asymptotic distribution of the bias-adjusted CSR OLS estimator in Theorem **??** and are compared with the critical values of a standard normal distribution.

Table II: Case 1. $\gamma^*_{z_1} = \gamma^*_{z_2} = 0$. Bias and RMSE

Panel A: T=36						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.291	-0.061	-0.011	0.343	0.142	0.106
$\hat{\gamma}^*_1$	0.405	0.066	0.029	0.166	0.067	0.055
$\hat{\gamma}^*_{z_1}$	0.011	0.000	0.000	0.484	0.230	0.161
$\hat{\gamma}^*_{z_2}$	0.011	0.000	0.000	1.052	0.509	0.315
Panel A: T=72						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.095	-0.007	-0.003	0.054	0.022	0.016
$\hat{\gamma}^*_1$	0.018	0.001	0.013	0.026	0.011	0.009
$\hat{\gamma}^*_{z_1}$	0.000	0.000	0.000	0.079	0.037	0.013
$\hat{\gamma}^*_{z_2}$	0.000	0.000	0.000	0.120	0.054	0.037

Table III: Case 2. $\gamma^*_{z_1} \neq 0, \gamma^*_{z_2} = 0$. Bias and RMSE

Panel A: T=36						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.552	-0.061	-0.051	0.172	0.069	0.052
$\hat{\gamma}^*_1$	0.115	0.025	0.015	0.085	0.035	0.029
$\hat{\gamma}^*_{z_1}$	0.523	-0.482	0.327	0.251	0.118	0.084
$\hat{\gamma}^*_{z_2}$	0.003	-0.002	0.000	0.379	0.170	0.118
Panel A: T=72						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.340	-0.010	-0.005	0.105	0.044	0.032
$\hat{\gamma}^*_1$	0.062	0.009	0.002	0.049	0.020	0.016
$\hat{\gamma}^*_{z_1}$	1.249	0.235	0.155	0.150	0.071	0.050
$\hat{\gamma}^*_{z_2}$	0.002	-0.001	0.000	0.326	0.158	0.098

Table IV: Case 3. $\gamma^*_{z_1} \neq \gamma^*_{z_2} \neq 0$. Bias and RMSE

Panel A: T=36						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-1.563	-0.613	-0.503	0.579	0.226	0.173
$\hat{\gamma}^*_1$	1.013	0.201	0.132	0.313	0.126	0.105
$\hat{\gamma}^*_{z_1}$	-1.875	-1.396	0.480	0.802	0.374	0.265
$\hat{\gamma}^*_{z_2}$	2.768	-1.496	-0.572	0.921	0.539	0.376
Panel A: T=72						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.080	0.001	0.001	0.032	0.014	0.010
$\hat{\gamma}^*_1$	0.013	0.002	0.000	0.015	0.006	0.005
$\hat{\gamma}^*_{z_1}$	0.904	0.178	0.171	0.047	0.022	0.016
$\hat{\gamma}^*_{z_2}$	1.026	-0.653	-0.396	0.101	0.049	0.030

Table V: Case 1. $\gamma^*_{z_1} = \gamma^*_{z_2} = 0$. Rejection rates

Panel A: T=36									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.100	0.056	0.014	0.097	0.049	0.009	0.103	0.050	0.011
$\hat{\gamma}^*_1$	0.098	0.047	0.013	0.104	0.055	0.011	0.104	0.053	0.010
$\hat{\gamma}^*_{z_1}$	0.101	0.051	0.011	0.101	0.056	0.010	0.101	0.053	0.010
$\hat{\gamma}^*_{z_2}$	0.099	0.049	0.011	0.097	0.043	0.009	0.101	0.048	0.008
Panel B: T=72									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.100	0.053	0.010	0.105	0.053	0.009	0.103	0.052	0.011
$\hat{\gamma}^*_1$	0.102	0.050	0.012	0.096	0.050	0.009	0.099	0.051	0.009
$\hat{\gamma}^*_{z_1}$	0.101	0.053	0.009	0.102	0.053	0.010	0.101	0.051	0.013
$\hat{\gamma}^*_{z_2}$	0.105	0.060	0.011	0.105	0.052	0.010	0.101	0.048	0.010

Table VI: Case 2. $\gamma^*_{z_1} \neq 0, \gamma^*_{z_2} = 0$. Rejection Rates

Panel A: T=36									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.103	0.055	0.016	0.098	0.050	0.010	0.103	0.051	0.011
$\hat{\gamma}^*_1$	0.098	0.054	0.013	0.105	0.053	0.010	0.104	0.053	0.008
$\hat{\gamma}^*_{z_1}$	0.101	0.051	0.011	0.100	0.053	0.011	0.102	0.052	0.010
$\hat{\gamma}^*_{z_2}$	0.098	0.049	0.011	0.098	0.043	0.009	0.101	0.048	0.009
Panel B: T=72									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.102	0.051	0.011	0.102	0.053	0.010	0.102	0.052	0.011
$\hat{\gamma}^*_1$	0.105	0.051	0.010	0.098	0.049	0.008	0.098	0.048	0.008
$\hat{\gamma}^*_{z_1}$	0.102	0.052	0.009	0.102	0.054	0.010	0.102	0.052	0.010
$\hat{\gamma}^*_{z_2}$	0.109	0.059	0.012	0.105	0.052	0.010	0.101	0.047	0.010

Table VII: Case 3. $\gamma^*_{z_1} \neq \gamma^*_{z_2} \neq 0$. Rejection Rates

PANEL A: T=36									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.106	0.057	0.015	0.098	0.050	0.011	0.100	0.052	0.010
$\hat{\gamma}^*_1$	0.106	0.053	0.014	0.102	0.056	0.010	0.095	0.052	0.010
$\hat{\gamma}^*_{z_1}$	0.105	0.047	0.01	0.098	0.054	0.011	0.101	0.055	0.010
$\hat{\gamma}^*_{z_2}$	0.099	0.054	0.013	0.098	0.044	0.009	0.099	0.048	0.008
Panel A: T=72									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.105	0.048	0.011	0.102	0.056	0.009	0.101	0.055	0.011
$\hat{\gamma}^*_1$	0.104	0.051	0.007	0.092	0.050	0.009	0.098	0.048	0.009
$\hat{\gamma}^*_{z_1}$	0.105	0.053	0.009	0.102	0.054	0.010	0.101	0.055	0.013
$\hat{\gamma}^*_{z_2}$	0.109	0.061	0.012	0.104	0.052	0.010	0.101	0.047	0.010

Case (ii): $\mathbf{Z}_t$ and ϵ_t weakly cross-sectionally correlated

In the second case, we allow the model disturbances to be weakly cross-sectionally correlated with the anomalies. In particular, we consider the following data-generating process for the error terms:

$$\epsilon_{it} = \frac{\sigma}{\tau}\left(u_{it} + \frac{\mathbf{1}'_{K_z}\boldsymbol{\eta}_{it}}{N^{\delta}}\right) \tag{OA.87}$$

where $\tau = 1 + \frac{K_z}{N^{2\delta}}$, u_{it} is generated from an i.i.d. standard normal random variable, σ^2 is calibrated as in the uncorrelated case, while the $K_z \times 1$ vector $\boldsymbol{\eta}_{it}$ is calibrated using the standardized residuals obtained by fitting a vector autoregressive (VAR) process of order 1 on the two anomalies. The parameter δ controls the strength of the cross-sectional correlation between the shocks and the anomalies: the higher the value of δ is, the weaker the cross-sectional correlation is. For our theoretical results to hold, we require $\delta \geq 0.5$.

Table VIII reports the percentage bias (Bias %) and root mean squared error (RMSE) of the bias-adjusted CSR OLS estimator derived in (**??**) for case $\gamma_{z_1} \neq 0, \gamma_{z_2} = 0$, where we use $\delta = 0.5$ in the data-generating process (OA.87). Panels A and B are for the cases of $T = 36$ and $T = 72$, respectively. The empirical rejection rates of the t-test, under the null hypothesis that the parameters of interest are equal to the true values are reported in Table IX. As before, the results are reported for the three levels of significance of 10%, 5% and 1% and for different values of N and T. The t-statistics are derived again using the standard errors the CSR OLS estimator in Theorem OA.2, and are compared with the critical values of a standard normal distribution.

The empirical distributions of the four parameters $\hat{\gamma}_0^*$, $\hat{\gamma}_1^*$, $\hat{\gamma}_{z_1}^*$, and $\hat{\gamma}_{z_2}^*$, for different values of δ (i.e., $\delta = 0.1, 0.25, 0.50, 1$) are depicted in Figure 1, where the black solid line represents the standard normal density. The results are obtained using 3,000 Monte Carlo replications, where we set $T = 72$ and $N = 1000$.

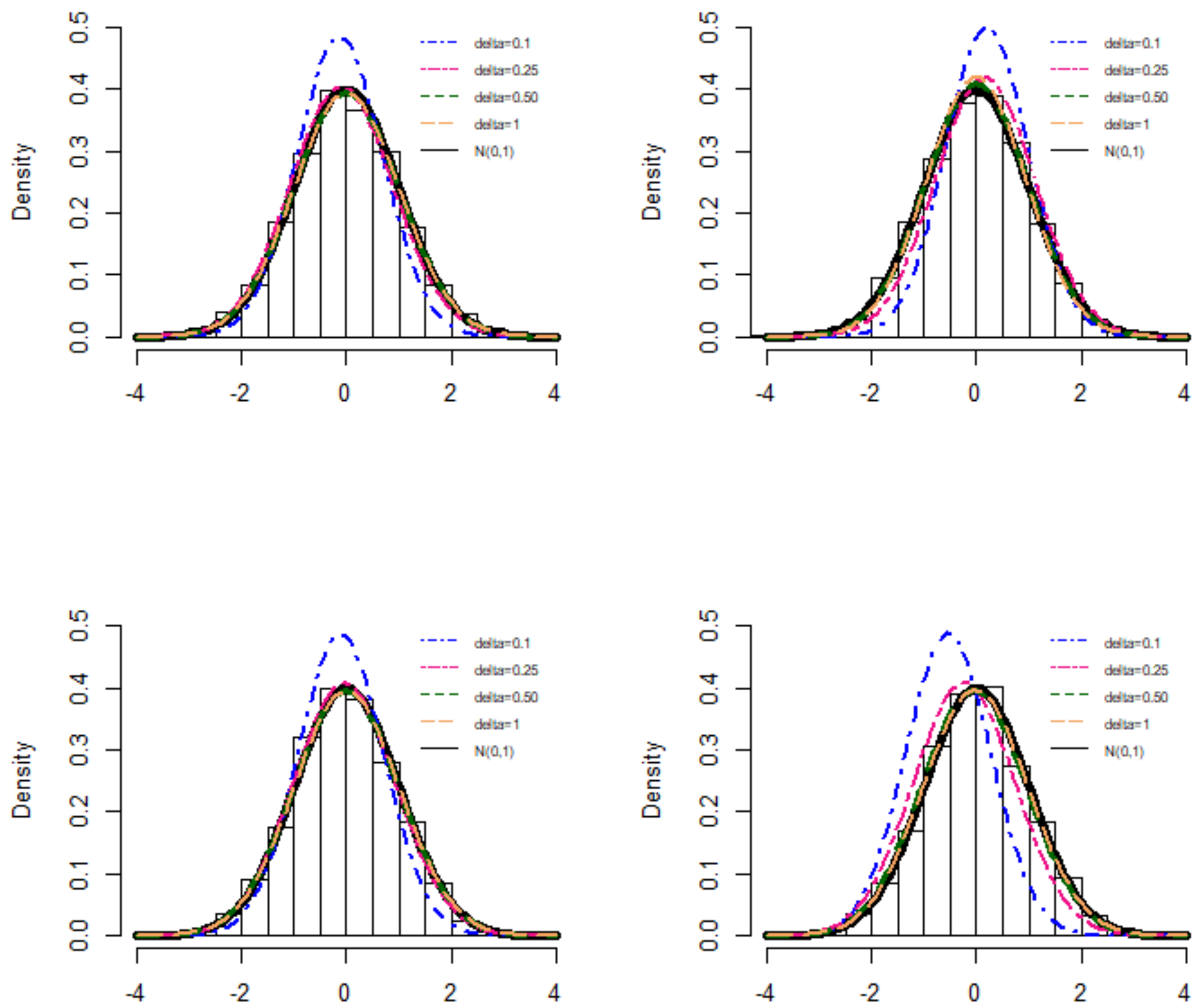
Table VIII: Case 2. $\gamma^*_{z_1} \neq 0, \gamma^*_{z_2} = 0$ and $\delta = 0.5$. Bias and RMSE

Panel A: T=36						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.974	-0.794	-0.529	0.576	0.227	0.174
$\hat{\gamma}^*_1$	0.934	0.243	0.219	0.305	0.125	0.105
$\hat{\gamma}^*_{z_1}$	0.144	0.025	0.026	0.799	0.373	0.266
$\hat{\gamma}^*_{z_2}$	0.004	0.004	0.003	1.190	0.538	0.376
Panel A: T=72						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.010	-0.008	0.000	0.005	0.002	0.002
$\hat{\gamma}^*_1$	0.001	0.000	0.000	0.003	0.001	0.001
$\hat{\gamma}^*_{z_1}$	0.559	0.172	-0.016	0.008	0.004	0.003
$\hat{\gamma}^*_{z_2}$	-0.001	0.000	0.000	0.017	0.008	0.005

Table IX: Case 2. $\gamma^*_{z_1} \neq 0, \gamma^*_{z_2} = 0$ and $\delta = 0.5$. Rejection Rates

Panel A: T=36									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.106	0.059	0.015	0.101	0.054	0.010	0.100	0.050	0.011
$\hat{\gamma}^*_1$	0.102	0.052	0.014	0.102	0.056	0.010	0.102	0.051	0.010
$\hat{\gamma}^*_{z_1}$	0.103	0.049	0.011	0.103	0.052	0.012	0.103	0.054	0.011
$\hat{\gamma}^*_{z_2}$	0.097	0.048	0.01	0.097	0.046	0.009	0.099	0.048	0.009
Panel B: T=72									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.097	0.044	0.010	0.106	0.050	0.010	0.102	0.056	0.011
$\hat{\gamma}^*_1$	0.104	0.048	0.009	0.091	0.049	0.007	0.097	0.049	0.008
$\hat{\gamma}^*_{z_1}$	0.098	0.049	0.008	0.102	0.053	0.010	0.102	0.052	0.012
$\hat{\gamma}^*_{z_2}$	0.105	0.058	0.011	0.101	0.051	0.010	0.101	0.048	0.010

Figure 1: Empirical distribution of $\hat{\gamma}_0^*$, $\hat{\gamma}_1^*$, $\hat{\gamma}_{z_1}^*$, and $\hat{\gamma}_{z_2}^*$ for different values of δ.



Case (iii): $\mathbf{Z}_t$ *and* ϵ_t *cross-sectionally correlated with cross-sectionally dependent error terms*

As a third experiment, in this Section we consider the more general case in which we allow also for a weak form of cross-sectional dependence among the model disturbances, besides the cross-sectional correlation between ϵ_{it} and the anomalies.

Particularly, we generate the disturbances using the following weak factor structure

$$\epsilon_{it} = \frac{\sigma_i}{\tau}\left(v_{it} + \frac{\mathbf{1}'_{K_z}\boldsymbol{\eta}_{it}}{N^{\delta}}\right) \tag{OA.88}$$

where

$$v_{it} = \frac{1}{\omega_i}\left(\nu_t \frac{\sqrt{\theta}}{N^{\kappa}} c_i + \sqrt{1-\theta}\,\xi_{it}\right)\sigma_i, \qquad \omega_i = \sqrt{\frac{\theta}{N^{2\kappa}} c_i^2 + (1-\theta)}$$

and where ν_t, c_i and ξ_{it} are generated from i.i.d. standard normal random variables. The parameter κ controls the strength of the cross-sectional dependence of the shocks (the bigger κ is, the weaker the dependence), while $0 \leq \theta \leq 1$ is a shrinkage parameter that controls the weight assigned to the diagonal end extra-diagonal elements of the covariance matrix Σ. To obtain representative values for each σ_i, we first estimate the residual variances from historical data. Then, at each Monte Carlo iteration, we generate a string of N values from a Uniform distribution, with parameters calibrated to the 10%-winsorized minimum and maximum value of the series of cross-sectional estimated variances $\hat{\sigma}_i^2$. This resampling procedure is used to minimize the impact of an ill-conditioned Σ on the simulation results. For our theoretical results to hold, we require $\kappa \geq 0.5$. Therefore, in Tables X and XI, we report the results for the case of $\theta = \kappa = \delta = 0.5$, setting the true values $\gamma_{z_1} \neq 0$ and $\gamma_{z_2} = 0$. Panels A and B of Table X report the bias and the RMSE of the parameter estimates for the case of $T = 36$ and $T = 72$, respectively. The empirical rejection rates of the t-test, under the null hypothesis that the parameters of interest are equal to the true values are reported in Table XI.

Finally, Figure 2 shows the empirical distributions of the four premia parameters, obtained with 3,000 Monte Carlo replications, with $T = 72$ and $N = 1000$. In the figure, we fix the parameters that regulate the cross-sectional dependence to $\theta = \kappa = 0.5$ and we show the results for different values of $\delta = 0.1, 0.25, 0.50, 1$.

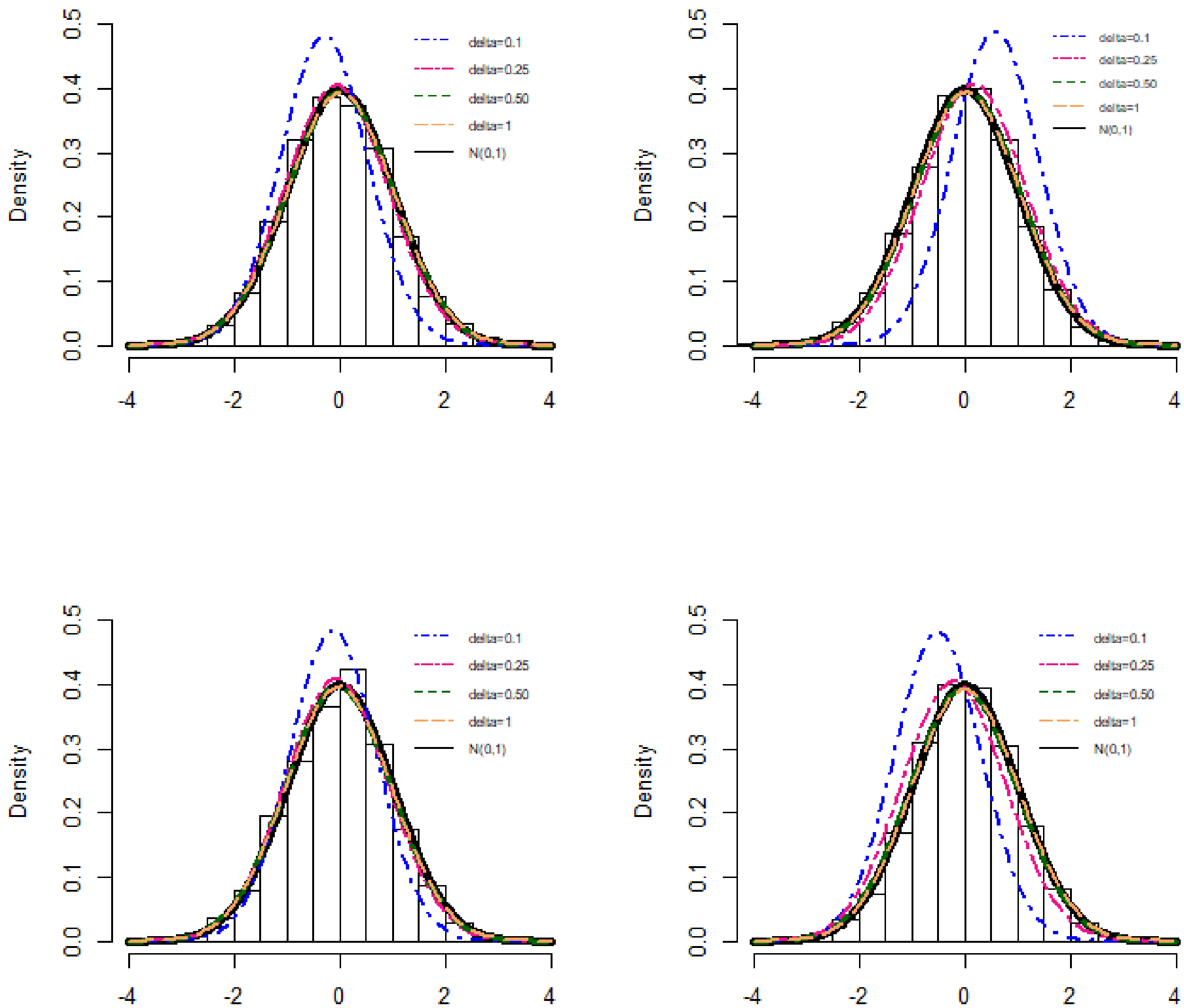
Table X: Case 3. $\gamma^*_{z_1} \neq 0, \gamma^*_{z_2} = 0$ and $\delta = \theta = \kappa = 0.5$. Bias and RMSE

Panel A: T=36						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-2.041	-1.079	-0.765	0.531	0.216	0.162
$\hat{\gamma}^*_1$	0.710	0.163	0.210	0.278	0.117	0.097
$\hat{\gamma}^*_{z_1}$	0.126	0.026	0.020	0.748	0.352	0.249
$\hat{\gamma}^*_{z_2}$	0.002	0.002	0.001	0.525	0.432	0.357
Panel A: T=72						
	Bias %			RMSE		
N	100	500	1000	100	500	1000
$\hat{\gamma}^*_0$	-0.178	-0.839	0.013	0.320	0.134	0.099
$\hat{\gamma}^*_1$	0.211	0.075	0.036	0.155	0.065	0.053
$\hat{\gamma}^*_{z_1}$	-0.997	-0.584	-0.333	0.446	0.211	0.151
$\hat{\gamma}^*_{z_2}$	-0.002	0.001	-0.001	0.502	0.418	0.220

Table XI: Case 3. $\gamma^*_{z_1} \neq 0, \gamma^*_{z_2} = 0$ and $\delta = \theta = \kappa = 0.5$. Rejection Rates

Panel A: T=36									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.093	0.049	0.008	0.103	0.053	0.010	0.102	0.051	0.010
$\hat{\gamma}^*_1$	0.089	0.046	0.011	0.099	0.053	0.011	0.097	0.048	0.010
$\hat{\gamma}^*_{z_1}$	0.103	0.050	0.009	0.095	0.053	0.009	0.100	0.052	0.009
$\hat{\gamma}^*_{z_2}$	0.099	0.050	0.011	0.103	0.053	0.010	0.101	0.052	0.009
Panel B: T=72									
	N=100			N=500			N=1000		
	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
$\hat{\gamma}^*_0$	0.100	0.048	0.009	0.100	0.050	0.012	0.100	0.052	0.012
$\hat{\gamma}^*_1$	0.106	0.050	0.011	0.103	0.052	0.013	0.101	0.053	0.012
$\hat{\gamma}^*_{z_1}$	0.099	0.049	0.010	0.093	0.052	0.008	0.100	0.049	0.011
$\hat{\gamma}^*_{z_2}$	0.103	0.055	0.013	0.103	0.052	0.010	0.100	0.051	0.011

Figure 2: Empirical distribution of $\hat{\gamma}_0^*$, $\hat{\gamma}_1^*$, $\hat{\gamma}_{z_1}^*$, and $\hat{\gamma}_{z_2}^*$ for different values of δ, with $\kappa = \theta = 0.5$.



OA.7.2 Cross-Sectional R^2 Test: Size and Power

In this Section, we investigate the size and power properties of the cross-sectional R2-test, based on the $\mathcal{T}_z$ statistics derived in Theorem 6. Specifically, we generate asset returns using the specification in (OA.84), with the error disturbances as in (OA.88), using different scenarios for the parameters δ, θ, κ. To evaluate the size of the test, we generate the returns as in (OA.84) and under the null hypothesis of no anomalies ($\gamma_{z_1} = \gamma_{z_2} = 0$). For the evaluation of the power we set ($\gamma_{z_1} \neq \gamma_{z_2} \neq 0$), where γ_{z_1} and γ_{z_2} have been calibrated using real data as in the previous cases. Then, at each Monte Carlo simulation we calculate the $\mathcal{T}_z$ statistics as in Theorem 6 and compare its empirical distribution (over the 3,000 Monte Carlo replications) with the linear combination of i.i.d chi-squared distributions defined in Theorem 6. Tables XII and XIII report the rejection rates for different levels of significance (10%, 5%, 1%) and for different values of N (100, 500, 1000) when $T = 36$ and $T = 72$, respectively. In both the tables we consider the case of cross-sectional dependence among $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$, setting $\delta = 0.5$ (Panels A and B) and $\delta = 0.25$ (Panels C and D), under both the assumptions of $\boldsymbol{\Sigma}$ diagonal (i.e. $\theta = 0$ and $\kappa = 0.5$) and $\boldsymbol{\Sigma}$ full (where we set $\theta = \kappa = 0.5$).

The results in the two tables suggest that the rejection rates of our test under the null of no anomalies are excellent for the diagonal and the full cases, when $\delta = 0.5$ When simulating with $\delta = 0.25$, consistently with our theory, the test starts to over-rejects as N increases, especially for the case of $T = 36$. The power properties of the test are fairly good when $N = 100$ and excellent when $N \geq 500$. As expected, power increases when the number of assets becomes large and the rejection rates are similar across time-series sample sizes. Overall, these simulation results suggest that our test $\mathcal{T}_z$ should be fairly reliable for the time-series and cross-sectional dimensions encountered in our empirical work.

Table XII: Size and power of the $\mathcal{T}_z$ -test in a one-factor model with two anomalies ($T = 36$)

	SIZE			POWER		
N	10%	5%	1%	10%	5%	1%
Panel A: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$ cross-sectionally correlated ($\delta = 0.5$) with $\boldsymbol{\Sigma}$ diagonal ($\theta = 0$, $\kappa = 0.5$)						
100	0.104	0.046	0.008	0.996	0.993	0.972
500	0.106	0.054	0.010	1.000	1.000	1.000
1000	0.104	0.050	0.010	1.000	1.000	1.000
Panel B: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$_i cross-sectionally correlated ($\delta = 0.5$) with $\boldsymbol{\Sigma}$ full ($\theta = \kappa = 0.5$)						
100	0.105	0.052	0.011	0.944	0.859	0.816
500	0.103	0.053	0.012	1.000	0.985	0.942
1000	0.102	0.052	0.012	1.000	1.000	1.000
Panel C: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$ cross-sectionally correlated ($\delta = 0.25$) with $\boldsymbol{\Sigma}$ diagonal ($\theta = 0$, $\kappa = 0.5$)						
100	0.103	0.042	0.006	1.000	0.998	0.984
500	0.114	0.057	0.012	1.000	1.000	1.000
1000	0.124	0.069	0.012	1.000	1.000	1.000
Panel D: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$ cross-sectionally correlated ($\delta = 0.25$) with $\boldsymbol{\Sigma}$ full ($\theta = \kappa = 0.5$)						
100	0.098	0.045	0.006	0.975	0.924	0.895
500	0.115	0.060	0.012	1.000	1.000	1.000
1000	0.122	0.070	0.012	1.000	1.000	1.000

The table presents the size and power properties of the $\mathcal{T}_z$ test. The null hypothesis is that there are no anomalies. The alternative hypothesis is that there is at least one anomaly. The results are reported for $T = 36$, for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 3,000 simulations. The rejection rates are based on the asymptotic distribution in Theorem 6.

OA.8 Granularity

The relevance of granularity can be understood from the following result, reported without proof, which extends Adler and Rosalsky (1991).

Proposition OA.4 (limiting behavior of weighted averages)**.** *For an iid sequence* Y_i *with* $E|Y|^2 < \infty$*, assume the following granularity conditions hold, for some finite constant* C*,*

$$\sum_{i=1}^{N} a_i^2 = o(b_N^2), \;\; \frac{1}{b_N}\sum_{i=1}^{N} a_i \to C.$$

(i) Then

$$\frac{1}{b_N}\sum_{i=1}^{N} a_i Y_i \to_p CEY < \infty.$$

(ii) If, in addition, for some finite $n < N$ *and some constant* $0 < C_i < \infty$, $1 \le i \le n$*,*

$$\frac{a_i}{b_N} \to C_i \textit{ for every } 1 \le i \le n < \infty,$$

then

$$\frac{1}{b_N}\sum_{i=1}^{N} a_j Y_i \to_p CEY + \sum_{i=1}^{n} C_i(Y_i - EY).$$

Case (i) is the granular case, which is implicit in our regularity assumptions, setting $a_i = w_{i,t-1}$ and $b_N = N$, with $C = 1$, such that $w_{i,t}/N = O_p(N^{-1})$. Note that only when $C = 1$, the simple and weighted average converge to the same limit, namely, EY. Case (ii) is the non-granular case, which leads to a random limit of the weighted average.

OA.9 No-Arbitrage with Anomalies

We show how the presence of anomalies does not necessarily rule out no-arbitrage with the following proposition, reported without proof.

Proposition OA.5 (no-arbitrage with anomalies)**.** *The asset pricing restriction (4) does* not *violate conditional no-arbitrage whenever (6) holds with*

$$\sup_N \boldsymbol{\gamma}'_{t-1,z}\mathbf{Z}'_{t-1}[E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t|I_{t-1}, \mathbf{\Pi})]^{-1}\mathbf{Z}_{t-1}\boldsymbol{\gamma}'_{t-1,z} \le C < \infty \textit{ almost surely,} \tag{OA.89}$$

for some constant C*, where we set the* $N \times K_z$ *matrix* $\mathbf{Z}_t \equiv (\mathbf{z}_{1,t}, \cdots, \mathbf{z}_{N,t})'$*.*

Table XIII: Size and power of the $\mathcal{T}_z$-test in a one-factor model with two anomalies ($T = 72$)

	SIZE			POWER		
N	10%	5%	1%	10%	5%	1%
Panel A: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$ cross-sectionally correlated ($\delta = 0.5$) with $\boldsymbol{\Sigma}$ diagonal ($\theta = 0$, $\kappa = 0.5$)						
100	0.107	0.051	0.013	1.000	1.000	1.000
500	0.096	0.049	0.012	1.000	1.000	1.000
1000	0.099	0.050	0.010	1.000	1.000	1.000
Panel B: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$_i cross-sectionally correlated ($\delta = 0.5$) with $\boldsymbol{\Sigma}$ full ($\theta = \kappa = 0.5$)						
100	0.107	0.049	0.013	1.000	1.000	1.000
500	0.096	0.048	0.012	1.000	1.000	1.000
1000	0.099	0.049	0.011	1.000	1.000	1.000
Panel C: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$ cross-sectionally correlated ($\delta = 0.25$) with $\boldsymbol{\Sigma}$ diagonal ($\theta = 0$, $\kappa = 0.5$)						
100	0.078	0.034	0.005	1.000	1.000	1.000
500	0.086	0.036	0.006	1.000	1.000	1.000
1000	0.100	0.044	0.011	1.000	1.000	1.000
Panel D: $\boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i$ cross-sectionally correlated ($\delta = 0.25$) with $\boldsymbol{\Sigma}$ full ($\theta = \kappa = 0.5$)						
100	0.082	0.036	0.006	1.000	1.000	1.000
500	0.087	0.037	0.006	1.000	1.000	1.000
1000	0.102	0.043	0.011	1.000	1.000	1.000

The table presents the size and power properties of the $\mathcal{T}_z$ test. The null hypothesis is that there are no anomalies. The alternative hypothesis is that there is at least one anomaly. The results are reported for $T = 72$, for different levels of significance (10%, 5%, and 1%) and for different values of the number of stocks (N) using 3,000 simulations. The rejection rates are based on the asymptotic distribution in Theorem 6.

As practical examples when Proposition OA.5 does and does not hold, consider the case when $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|I_{t-1}, \boldsymbol{\Pi})]$ exhibits a limited degree of cross-dependence of the $\boldsymbol{\epsilon}_t$, such as when equal to $\sigma^2\mathbf{I}_N$ for some constant scalar σ^2. Then, condition (OA.89) is satisfied when either $\boldsymbol{\gamma}_z = O(N^{-\frac{1}{2}})$ or $\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1} = O(1)$. These restrictions are needed because the $\mathbf{Z}_{t-1}$ affect the mean but not the variances and covariances of the returns. In contrast, when instead $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|I_{t-1}, \boldsymbol{\Pi})] = \mathbf{Z}_{t-1}\mathbf{C}_z\mathbf{Z}_{t-1}' + \sigma^2\mathbf{I}_N$, for some $K_z \times K_z$ constant non-singular matrix $\mathbf{C}_z$, then (OA.89) is redundant as it imposes no cross-sectional constraint. In fact, by the Sherman-Morrison decomposition, whenever $\mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}$ diverges as $N \to \infty$, one obtains $\mathbf{Z}_{t-1}'[E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|I_{t-1}, \boldsymbol{\Pi})]^{-1}\mathbf{Z}_{t-1} = \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1}(\sigma^2\mathbf{C}_z^{-1} + \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1})^{-1}\mathbf{C}_z^{-1} \to_p \mathbf{C}_z^{-1}$, that is, bounded even for large N. No restriction arises because the $\mathbf{Z}_{t-1}$ affect the mean, variance, and covariances of the returns in the same way, making their effect neutral in terms of the risk-return trade-off.

OA.10 Time-Varying Betas

This section establishes that, under Assumption 1, namely

$$\frac{(\mathbf{B}_s - \mathbf{B})'(\mathbf{B}_s - \mathbf{B})}{N} = o(N^{-\frac{1}{2}}), \tag{OA.90}$$

the whole asymptotic analysis remains unchanged as if the loadings were constant.

Our smoothing assumption is extremely general and accommodates a great variety of time-varying patterns of the loadings. Important examples include the case when

$$\boldsymbol{\beta}_{i,s} = \boldsymbol{\beta}_i + \mathbf{B}_{1i}\mathbf{g}_s + \mathbf{B}_{2i}\mathbf{z}_{is}, \tag{OA.91}$$

for matrices of coefficients $\boldsymbol{\beta}_{0i}$ ($K_{\text{f}} \times 1$), $\mathbf{B}_{1i}$ ($K_{\text{f}} \times K_{\text{g}}$), and $\mathbf{B}_2$ ($K_{\text{f}} \times K_{\text{z}}$) such that

$$\sum_{i=1}^{N}(\mathbf{B}_{1i} \otimes \mathbf{B}_{1i}) = o(N^{\frac{1}{2}}) \text{ and } \sum_{i=1}^{N}(\mathbf{B}_{2i} \otimes \mathbf{B}_{2i})(\mathbf{z}_{is} \otimes \mathbf{z}_{is}) = o(N^{\frac{1}{2}}).$$

In turn, the former conditions are implied when $\mathbf{B}_{1i} = \mathbf{B}_{1i}^*/(N^{\frac{1}{4}}log(N))$ and $\mathbf{B}_{2i} = \mathbf{B}_{2i}^*/(N^{\frac{1}{4}}log(N))$ for coefficients $\mathbf{B}_{1i}^*$ and $\mathbf{B}_{2i}^*$ satisfying $N^{-1}\sum_{i=1}^{N}(\mathbf{B}_{1i}^* \otimes \mathbf{B}_{1i}^*) = O(1)$ and $N^{-1}\sum_{i=1}^{N}(\mathbf{B}_{2i}^* \otimes \mathbf{B}_{2i}^*) = O(1)$.

SOME SIMULATIONS/NUMERICAL ILLUSTRATIONS

We will focus on our main estimator, the CSR OLS-type estimator of Section 5. Consider again the *conditional* asset pricing model (see (7))

$$\mathbf{R}_t = \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1} + \mathbf{X}_{t-1}\boldsymbol{\Gamma}_{\mathrm{f},t-1} + \boldsymbol{\epsilon}_t \tag{OA.92}$$

setting $\mathbf{X}_{t-1} \equiv (1_N, \mathbf{B}_{t-1})$, where $\boldsymbol{\Gamma}_{\mathrm{f},t-1} = (\gamma_{0,t-1}, \boldsymbol{\delta}'_{\mathrm{f},t-1})'$, with $\boldsymbol{\delta}_{\mathrm{f},t-1}$ defined in (8). Then

$$\mathbf{R}_t = \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1} + \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{f},t-1} + \boldsymbol{\epsilon}_t + (\mathbf{X}_{t-1} - \mathbf{X})\boldsymbol{\Gamma}_{\mathrm{f},t-1}, \tag{OA.93}$$

and in the matrix sense

$$\mathbf{R} = \gamma_0 1'_N + \boldsymbol{\Delta}_z + (\mathbf{I}_{T-1} \otimes \boldsymbol{\gamma}'_z) \begin{bmatrix} \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \\ \vdots \\ \mathbf{Z}'_{T-1} \end{bmatrix} + \boldsymbol{\Delta}_B + \boldsymbol{\delta}_{\mathrm{f}}\mathbf{B}' + \boldsymbol{\epsilon},$$

setting

$$\boldsymbol{\Delta}_B \equiv \begin{bmatrix} \boldsymbol{\delta}'_{\mathrm{f},1} & \mathbf{0}'_{K_\mathrm{f}} & \cdots & \mathbf{0}'_{K_\mathrm{f}} \\ \mathbf{0}'_{K_\mathrm{f}} & \boldsymbol{\delta}'_{\mathrm{f},2} & \cdots & \mathbf{0}'_{K_\mathrm{f}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}'_{K_\mathrm{f}} & \mathbf{0}'_{K_\mathrm{f}} & \cdots & \boldsymbol{\delta}'_{\mathrm{f},T-1} \end{bmatrix} \begin{bmatrix} (\mathbf{B}_1 - \mathbf{B})' \\ (\mathbf{B}_2 - \mathbf{B})' \\ \vdots \\ (\mathbf{B}_{T-1} - \mathbf{B})' \end{bmatrix}, \quad \boldsymbol{\Delta}_z = \begin{bmatrix} \boldsymbol{\gamma}'_{z,1} - \boldsymbol{\gamma}'_{\mathrm{z}} & \mathbf{0}'_{K_\mathrm{z}} & \cdots & \mathbf{0}'_{K_\mathrm{z}} \\ \mathbf{0}'_{K_\mathrm{z}} & \boldsymbol{\gamma}'_{z,2} - \boldsymbol{\gamma}'_{\mathrm{z}} & \cdots & \mathbf{0}'_{K_\mathrm{z}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}'_{K_\mathrm{z}} & \mathbf{0}'_{K_\mathrm{z}} & \cdots & \boldsymbol{\gamma}'_{\mathrm{z},t-1} - \boldsymbol{\gamma}'_{\mathrm{z}} \end{bmatrix} \begin{bmatrix} \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \\ \vdots \\ \mathbf{Z}'_{T-1} \end{bmatrix}, \boldsymbol{\delta}_{\mathrm{f}} \equiv \begin{bmatrix} \boldsymbol{\delta}'_{\mathrm{f},1} \\ \vdots \\ \boldsymbol{\delta}'_{\mathrm{f,T-1}} \end{bmatrix},$$

with $\boldsymbol{\Delta}_z$ defined in Assumption OA.1. Then

$$\begin{aligned} \hat{\mathbf{B}} &= \mathbf{R}'\mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F}(\mathbf{F}'\mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F})^{-1} = \mathbf{R}'\mathbf{P} = \left(1_N\boldsymbol{\gamma}'_0 + \boldsymbol{\Delta}'_z + (\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{T-1})(\mathbf{I}_{T-1} \otimes \boldsymbol{\gamma}_z) + \boldsymbol{\Delta}'_B + \mathbf{B}\boldsymbol{\delta}'_{\mathrm{f}} + \boldsymbol{\epsilon}'\right)\mathbf{P} \\ &= (\boldsymbol{\epsilon}' + \boldsymbol{\Delta}'_B)\mathbf{P} + \mathbf{B}, \end{aligned}$$

where the last equation follows from Assumption OA.1. The term $\boldsymbol{\Delta}'_B\mathbf{P}$ on the right-hand side of $\hat{\mathbf{B}}$ appears when considering the time-varying loadings but its effect will be shown to be asymptotically negligible (as $N \to \infty$).

We now show that asymptotically one obtains that, for our CSR OLS estimator, the results of Theorems 1-2 are obtained by replacing $\mathbf{B}_{t-1}$ with $\mathbf{B}$. The same applies to all our asymptotic analyses of the CSR WLS estimator, the R^2 test, and the misspecification-robust case.[12]

[12]Details are available upon request.

By inspecting our CSR OLS-type estimator, namely

$$\begin{bmatrix}\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1}\\ \hat{\gamma}^*_{\mathrm{z},t-1}\end{bmatrix} \equiv \begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z}_{t-1}\\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1}\end{bmatrix}^{-1}\begin{bmatrix}\hat{\mathbf{X}}'\mathbf{R}_t - N\hat{\boldsymbol{\Lambda}}_{2,t-1}\\ \mathbf{Z}'_{t-1}\mathbf{R}_t\end{bmatrix},$$

one sees that we need to study the following quantities (and show that are asymptotically equivalent when the locally-constant case is considered). First, considering the terms in the matrix inverse, namely

$$\begin{aligned}\frac{1}{N}\hat{\mathbf{B}}'\hat{\mathbf{B}} &= \frac{1}{N}[\mathbf{P}'(\boldsymbol{\Delta}_B+\boldsymbol{\epsilon})+\mathbf{B}'][(\boldsymbol{\Delta}_B+\boldsymbol{\epsilon})'\mathbf{P}+\mathbf{B}]\\ &= \underbrace{\frac{1}{N}[\mathbf{P}'\boldsymbol{\epsilon}+\mathbf{B}'][\boldsymbol{\epsilon}'\mathbf{P}+\mathbf{B}]}_{\text{(locally-constant term)}} + \underbrace{\frac{1}{N}\mathbf{P}'\boldsymbol{\Delta}_B\boldsymbol{\Delta}'_B\mathbf{P} + \frac{1}{N}\mathbf{P}'\boldsymbol{\Delta}_B(\mathbf{B}+\boldsymbol{\epsilon}'\mathbf{P}) + \frac{1}{N}(\mathbf{B}'+\mathbf{P}'\boldsymbol{\epsilon})\boldsymbol{\Delta}'_B\mathbf{P}}_{\text{(time-varying terms)}}.\end{aligned}$$

which requires

$$\frac{1}{N}\boldsymbol{\Delta}_B\boldsymbol{\Delta}'_B = o(1),\ \ \frac{1}{N}\boldsymbol{\Delta}_B\mathbf{B} = o(1),\quad \text{and } \boldsymbol{\Delta}_B\boldsymbol{\epsilon}' = o_p(1),$$

and

$$\begin{aligned}\frac{1}{N}\mathbf{Z}'_{t-1}\hat{\mathbf{B}} &= \frac{1}{N}\mathbf{Z}'_{t-1}[\mathbf{B}+(\boldsymbol{\Delta}_B+\boldsymbol{\epsilon})'\mathbf{P}]\\ &= \underbrace{\frac{1}{N}\mathbf{Z}'_{t-1}[\mathbf{B}+\boldsymbol{\epsilon}'\mathbf{P}]}_{\text{(locally-constant terms)}} + \underbrace{\frac{1}{N}\mathbf{Z}'_{t-1}\boldsymbol{\Delta}'_B P,}_{\text{(time-varying term)}}\end{aligned}$$

which requires

$$\frac{1}{N}\mathbf{Z}'_{t-1}\boldsymbol{\Delta}'_B = o_p(1).$$

Consider now the terms to the right hand of the matrix inverse. Then,

$$\begin{aligned}\frac{1}{N}\mathbf{R}'_t\hat{\mathbf{B}} &= \frac{1}{N}(\mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1}+\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{f},t-1}+\boldsymbol{\epsilon}_t+(\mathbf{X}_{t-1}-\mathbf{X})\boldsymbol{\Gamma}_{\mathrm{f},t-1})'[\mathbf{B}+(\boldsymbol{\Delta}_B+\boldsymbol{\epsilon})'\mathbf{P}]\\ &= \underbrace{\frac{1}{N}(\mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1}+\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{f},t-1}+\boldsymbol{\epsilon}_t)'[\mathbf{B}+\boldsymbol{\epsilon}'\mathbf{P}]}_{\text{(locally-constant terms)}}\\ &+ \underbrace{\frac{1}{N}(\mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1}+\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{f},t-1}+\boldsymbol{\epsilon}_t+(\mathbf{X}_{t-1}-\mathbf{X})\boldsymbol{\Gamma}_{\mathrm{f},t-1})'\boldsymbol{\Delta}'_B\mathbf{P} + \frac{1}{N}((\mathbf{X}_{t-1}-\mathbf{X})\boldsymbol{\Gamma}_{\mathrm{f},t-1})'[\mathbf{B}+\boldsymbol{\epsilon}'\mathbf{P}]}_{\text{(time-varying term)}},\ \text{and}\end{aligned}$$

$$\frac{1}{N}\mathbf{Z}'_{t-1}\hat{\mathbf{B}} = \mathbf{Z}'_{t-1}[\mathbf{B}+(\boldsymbol{\Delta}_B+\boldsymbol{\epsilon})'\mathbf{P}] = \underbrace{\frac{1}{N}\mathbf{Z}'_{t-1}[\mathbf{B}+\boldsymbol{\epsilon}'\mathbf{P}]}_{\text{(locally-constant terms)}} + \underbrace{\frac{1}{N}\mathbf{Z}'_{t-1}\boldsymbol{\Delta}'_B,}_{\text{(time-varying term)}}$$

which requires strengthening the above conditions to rate $o_p(N^{-1/2})$, namely

$$\frac{1}{N}\boldsymbol{\Delta}_B\boldsymbol{\Delta}_B' = o(N^{-1/2}), \ \frac{1}{N}\boldsymbol{\Delta}_B\mathbf{B} = o(N^{-1/2}), \quad \text{and } \boldsymbol{\Delta}_B\boldsymbol{\epsilon}' = o_p(N^{-1/2}), \text{ and } \frac{1}{N}\mathbf{Z}_{t-1}'\boldsymbol{\Delta}_B' = o_p(N^{-1/2}). \tag{OA.94}$$

It turns out that Assumption 1 delivers precisely the required sufficient condition. For example, for the following quantity

$$\frac{1}{N}\boldsymbol{\Delta}_B\boldsymbol{\Delta}_B' =$$
$$\frac{1}{N}\begin{bmatrix}\boldsymbol{\delta}_{\mathrm{f},1}' & \mathbf{0}_{K_\mathrm{f}}' & \cdots & \mathbf{0}_{K_\mathrm{f}}' \\ \mathbf{0}_{K_\mathrm{f}}' & \boldsymbol{\delta}_{\mathrm{f},2}' & \cdots & \mathbf{0}_{K_\mathrm{f}}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K_\mathrm{f}}' & \mathbf{0}_{K_\mathrm{f}}' & \cdots & \boldsymbol{\delta}_{\mathrm{f},T-1}'\end{bmatrix}\begin{bmatrix}(\mathbf{B}_1-\mathbf{B})'(\mathbf{B}_1-\mathbf{B}) & \cdots \\ \vdots & \vdots \\ \cdots & (\mathbf{B}_{T-1}-\mathbf{B})'(\mathbf{B}_{T-1}-\mathbf{B})\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_{\mathrm{f},1}' & \mathbf{0}_{K_\mathrm{f}}' & \cdots & \mathbf{0}_{K_\mathrm{f}}' \\ \mathbf{0}_{K_\mathrm{f}}' & \boldsymbol{\delta}_{\mathrm{f},2}' & \cdots & \mathbf{0}_{K_\mathrm{f}}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K_\mathrm{f}}' & \mathbf{0}_{K_\mathrm{f}}' & \cdots & \boldsymbol{\delta}_{\mathrm{f},T-1}'\end{bmatrix},$$

to be $o_p(N^{-\frac{1}{2}})$ one needs $N^{-1}(\mathbf{B}_s-\mathbf{B})'(\mathbf{B}_t-\mathbf{B}) = o(N^{-\frac{1}{2}})$ for every $t, s = 1, \cdots, T-1$ which, by Holder's inequality for matrices, is implied immediately by Assumption 1. Similar arguments apply to all the other terms in (OA.94).

OA.11 List of Variables

Table A.1: **List of variables.** The table shows the list of predictors used in the paper. A detailed description of the variables can be found in **?**. The variables have been grouped following the ex-ante categorization of Hou et al. (2020)

Variable name	Description	Category
ResidualMomentum	6 month residual momentum	Momentum
AnnouncementReturn	Earnings announcement return	Momentum
CustomerMomentum	Customer momentum	Momentum
retConglomerate	Conglomerate return	Momentum
EarningsSurprise	Earnings Surprise	Momentum
High52	52 week high	Momentum
IndMom	Industry Momentum	Momentum
AnalystRevision	Analysts revision	Momentum
EarnSupBig	Earnings surprise of big firms	Momentum
IndRetBig	Industry return of big firms	Momentum
RevenueSurprise	Revenue Surprise	Momentum
Mom12m	Momentum (12 month)	Momentum
Mom6m	Momentum (6 month)	Momentum
MomVol	Momentum and Volume	Momentum
IntMom	Intermediate Momentum	Momentum
EarningsConsistency	Earnings growth for consistent growers	Value Versus Growth
SP	Sales-to-price	Value Versus Growth
EP	Earnings-to-Price Ratio	Value Versus Growth
NetPayoutYield	Net Payout Yield	Value Versus Growth
PayoutYield	Payout Yield	Value Versus Growth
IntanBM	Intangible return using BM	Value Versus Growth
IntanCFP	Intangible return using CFtoP	Value Versus Growth
IntanEP	Intangible return using EP	Value Versus Growth
IntanSP	Intangible return using Sale2P	Value Versus Growth
LRreversal	Long-run reversal	Value Versus Growth
MRreversal	Momentum-Reversal	Value Versus Growth
ShortInterest	Short Interest	Value Versus Growth
EquityDuration	Equity Duration	Value Versus Growth
cfp	Operating Cash flows to price	Value Versus Growth
sfe	Earnings Forecast to price	Value Versus Growth
AM	Total assets to market	Value Versus Growth
BMdec	Book to market using December ME	Value Versus Growth
AnalystValue	Analyst Value	Value Versus Growth
DivSeason	Dividends	Value Versus Growth
ShareRepurchase	Share repurchases	Value Versus Growth
fgr5yrLag	Long-term EPS forecast	Value Versus Growth
CF	Cash flow to market	Value Versus Growth
MeanRankRevGrowth	Revenue Growth Rank	Value Versus Growth
DivYieldST	Dividend Yield	Value Versus Growth
EntMult	Enterprise Multiple	Value Versus Growth
BPEBM	Leverage component of BM	Value Versus Growth
EBM	Enterprise component of BM	Value Versus Growth
NetDebtPrice	Net debt to price	Value Versus Growth
BM	Book to market using most recent ME	Value Versus Growth

List of variables (continued)

Variable name	Description	Category
ChInvIA	Change in capital investment	Investment
grcapx	Change in capex (two years)	Investment
grcapx3y	Change in capex (three years)	Investment
InvGrowth	Inventory Growth	Investment
NetDebtFinance	Net debt financing	Investment
NetEquityFinance	Net equity financing	Investment
XFIN	Net external financing	Investment
AssetGrowth	Asset Growth	Investment
CompEquIss	Composite equity issuance	Investment
ShareIss5Y	Share issuance (5 year)	Investment
GrLTNOA	Growth in Long term net operating assets	Investment
PctAcc	Percent Operating Accruals	Investment
PctTotAcc	Percent Total Accruals	Investment
NOA	Net Operating Assets NOA	Investment
dNoa	change in net operating assets	Investment
CompositeDebtIssuance	Composite debtissuance	Investment
InvestPPEInv	change in ppe and inv/assets	Investment
ShareIss1Y	Share issuance (1 year)	Investment
DelCOA	Change in current operating assets	Investment
DelCOL	Change in currentoperating liabilities	Investment
DelEqu	Change in equity to assets	Investment
DelFINL	Change in financial liabilities	Investment
DelLTI	Change in long-term investment	Investment
TotalAccruals	Total accruals	Investment
Accruals	Accruals	Investment
DebtIssuance	Debt Issuance	Investment
ChInv	Inventory Growth	Investment
ChTax	Change in Taxes	Investment
Investment	Investment to revenue	Investment
AbnormalAccruals	Abnormal Accruals	Investment

List of variables (continued)

Variable name	Description	Category
roaq	Return on assets including extraordinary income	Profitability
CBOperProf	Cash-based operating profitability	Profitability
OperProfRD	Cash-based operating profitability	Profitability
CashProd	Cash Productivity	Profitability
OScore	O Score	Profitability
BookLeverage	Book leverage (annual)	Profitability
OperProf	operating profits / book equity	Profitability
RoE	net income / book equity	Profitability
VarCF	Cash-flow to price variance	Profitability
VolumeTrend	Volume Trend	Profitability
Tax	Taxable income to income	Profitability
ChEQ	Sustainable Growth	Profitability
MS	Mohanram G-score	Profitability
GP	gross profits / total assets	Profitability
PS	Piotroski F-score	Profitability
DelDRC	Deferred Revenue	Profitability
ChAssetTurnover	Change in Asset Turnover	Profitability
ChNNCOA	Change in Net Noncurrent Operating	Profitability
ChNWC	Change in Net Working Capital	Profitability
Mom6mJunk	Junk Stock Momentum	Profitability

List of variables (continued)

Variable name	Description	Category
GrSaleToGrInv	Gross Margin growth over sales growth	Intangibles
GrSaleToGrOverhead	Sales growth over overhead growth	Intangibles
OrderBacklogChg	Order backlog	Intangibles
hire	Employment growth	Intangibles
BrandInvest	Brand capital investment	Intangibles
Leverage	Market leverage	Intangibles
FEPS	Failure probability	Intangibles
AdExp	Advertising Expense	Intangibles
RD	R&D over market cap	Intangibles
RDAbility	R&D ability	Intangibles
Activism1	Shareholder activism 1	Intangibles
Activism2	Shareholder activism 2	Intangibles
ExclExp	Excluded Expenses	Intangibles
SurpriseRD	Unexpected R&D increase	Intangibles
OrgCap	Organizational Capital	Intangibles
AOP	Analyst Optimism	Intangibles
PredictedFE	Predicted Analyst forecast error	Intangibles
FR	Pension Funding Status	Intangibles
Governance	Governance Index	Intangibles
tang	Tangibility	Intangibles
Mom12mOffSeason	Returns in not-same month last year	Intangibles
MomOffSeason	Returns in not-same month	Intangibles
MomOffSeason06YrPlus	Returns in different months years 6 to 10	Intangibles
MomOffSeason11YrPlus	Returns in different months years 11 to 15	Intangibles
MomOffSeason16YrPlus	Returns in not-same month years 16 to 20	Intangibles
MomSeason	Return seasonality	Intangibles
MomSeason06YrPlus	Return seasonality years 6 to 10	Intangibles
MomSeason11YrPlus	Return seasonality years 11 to 15	Intangibles
MomSeason16YrPlus	Return seasonality years 16 to 20	Intangibles
MomSeasonShort	Return seasonality last year	Intangibles
PriceDelayRsq	Price delay r square	Intangibles
PriceDelaySlope	Price delay coeff	Intangibles
PriceDelayTstat	Price delay SE adjusted	Intangibles
Herf	Industry concentration (Herfindahl) sales	Intangibles
HerfAsset	Industry concentration (Herfindahl) assets	Intangibles
HerfBE	Industry concentration (Herfindahl) book	Intangibles
RDcap	R&D capital-to-assets	Intangibles
EarningsStreak	Earnings streak indicator	Intangibles
NumEarnIncrease	Number of consecutive earnings increases	Intangibles
GrAdExp	Growth in advertising expenses	Intangibles
RIO_Disp	Institutional Own and Forecast Dispersion	Intangibles
RIO_MB	Institutional Own and BM	Intangibles
RIO_Turnover	Institutional Own and Turnover	Intangibles
RIO_Volatility	Institutional Own and Idio Vol	Intangibles

List of variables (continued)

Variable name	Description	Category
OPLeverage	Operating Leverage	Intangibles
Cash	Cash to assets	Intangibles
OrderBacklog	Order backlog	Intangibles
realestate	Real estate holdings	Intangibles
ConvDebt	Convertible debt indicator	Intangibles
IdioVolAHT	Idiosyncratic risk	Trading Frictions
Illiquidity	Amihud (2002) illiquidity	Trading Frictions
BidAskSpread	Bid-ask spread	Trading Frictions
betaVIX	Systematic volatility	Trading Frictions
IdioRisk	Idiosyncratic risk	Trading Frictions
IdioVol3F	Idiosyncratic risk (3 factor)	Trading Frictions
CoskewACX	Coskewness	Trading Frictions
MaxRet	Maximum return over month	Trading Frictions
ReturnSkew	Skewness of daily returns	Trading Frictions
ReturnSkew3F	Skewness of daily idiosyncratic returns (3F)	Trading Frictions
DolVol	Past trading volume	Trading Frictions
std_turn	Share turnover volatility	Trading Frictions
VolSD	Volume Variance	Trading Frictions
ProbInformedTrading	Probability of Informed Trading	Trading Frictions
Beta	CAPM beta	Trading Frictions
BetaFP	Frazzini-Pedersen Beta	Trading Frictions
Coskewness	Coskewness	Trading Frictions
VolMkt	Volume to market equity	Trading Frictions
OptionVolume1	Option Volume to Stock Volume	Trading Frictions
OptionVolume2	Option Volume relative to recent average	Trading Frictions
BetaTailRisk	Tail risk beta	Trading Frictions
zerotrade	Days with zero trades	Trading Frictions
zerotradeAlt1	Days with zero trades	Trading Frictions
zerotradeAlt12	Days with zero trades	Trading Frictions
BetaLiquidityPS	Pastor-Stambaugh liquidity beta	Trading Frictions
skew1	Volatility smirk near the money	Trading Frictions
SmileSlope	Put volatility minus call volatility	Trading Frictions

References

Adler, A., and A. Rosalsky (1991): "On the weak law of large numbers for normed weighted sums of I.I.D. random variables," *International Journal of Mathematics and Mathematical Sciences*, 14, 191–202.

Fama, E. F., and K. R. French (2008): "Dissecting anomalies," *The Journal of Finance*, 63, 1653–1678.

Fama, E. F., and J. D. MacBeth (1973): "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 81, 607–636.

Raponi, V., C. Robotti, and P. Zaffaroni (2020): "Testing beta-pricing models using large cross-sections," *The Review of Financial Studies*, 33(6), 2796–2842.

Comments

1. I have removed the section on the notation and explained everything in the text. Double check if everything is properly defined.
2. Regarding assumptions: should we leave the regularity conditions in the Appendix and move the main ones in the text? For example assumption 7 on smooth time variation and Assumption 8 on granularity should be moved in the text?
3. The Appendix on U_ϵ is now a remark (See Remark OA.6). Does it work?
4. Do we really want to keep the misspecification section? It is really hard to write and follow.