Factor Models with Downside Risk*

Daniele Massacci[†] **Lucio Sarno**[‡] **Lorenzo Trapani**[§]

This Version: November 2022

*We are grateful for comments and suggestions to Tim Bollerslev, Adam Farago, Massimo Guidolin, Alexandros Kostakis, Alessandro Melone, Olivier Scaillet, Michael Weber, Dacheng Xiu and to participants at presentations held at: the 31st EC^2 Conference, Paris, 11-12 December 2020; the 14th International Conference on Computational and Financial Econometrics (CFE 2020); the 9th Italian Congress of Econometrics and Empirical Economics; the 2022 Annual SoFiE Conference; the University of Essex; the University of Liverpool; and the University of Verona. All errors are our own.

[†]Daniele Massacci is with King's College London and the University of Naples Federico II. Email: daniele.massacci@kcl.ac.uk.

[‡]Lucio Sarno is with the University of Cambridge and the Centre for Economic Policy Research (CEPR). Email: l.sarno@jbs.cam.ac.uk.

[§]Lorenzo Trapani is with the University of Nottingham. Email: lorenzo.trapani@nottingham.ac.uk.

Factor Models with Downside Risk

We propose a conditional model of asset returns in the presence of common factors and downside risk. Specifically, we generalize existing latent factor models in three ways: we show how to estimate the threshold which identifies the "disappointment" event triggering the bad state of the world; we permit different factor structures for asset returns in good and bad states; we show how to estimate consistently the conditional risk premia of observable factors from the estimated latent factors. The usefulness of the model is illustrated through two applications to cross-sections of asset returns in equity markets and other major asset classes.

1 Introduction

A growing body of research argues that the cross-section of asset returns should and does reflect a premium for bearing downside risk. This risk premium compensates investors for holding assets that covary strongly with the market when the market declines or is abnormally low: such assets are undesirable precisely because they offer particularly low returns in bad times. For example, Ang, Chen, and Xing (2006) show that there is a downsize risk premium in the cross-section of US stock returns, a result confirmed by Farago and Tédongap (2018) using a five-factor conditional model. Lettau, Maggiori, and Weber (2014) show that a conditional capital asset pricing model that allows for downside risk can explain not only US stocks, but also the cross section of returns in currency and other major asset classes. These results are consistent with the original conceptual framework of Markovitz (1959), who advocated using semivariance as a measure of risk, rather than variance, because semivariance measures downside losses rather than upside gains.

In a parallel development in the context of asset pricing models, the literature has also sought to tackle the vexing issue of potentially biased estimation of unconditional (i.e., linear, and hence free from downside risk) asset pricing models arising from omitting pricing factors. Since it is not at all clear which pricing factors should be used (the true factors), virtually any model which uses observable factors is at risk of omitting a relevant pricing factor, thereby suffering from omitted variable bias. This misspecification problem is addressed rigorously in a recent, seminal contribution by Giglio and Xiu (2021), who propose a three-pass procedure to conduct inference on the risk premia of any observable factor. The procedure solves both the omitted variable bias problem, and the much debated "factor zoo" issue (see Cochrane, 2011; and Harvey and Liu, 2020), by considering a latent factor model which allows to span the underlying true factor space.

The two research areas summarized above indicate that an *unconditional*, linear asset pricing model based on observable factors, which constitutes the workhorse model of empirical asset pricing, can suffer from at least two possible misspecification issues: the omission of relevant pricing factors, and the presence of a downside regime. Both these issues can be resolved by using a *conditional*, latent factor model which allows for the presence of downside risk in asset returns. However, despite the intuitive appeal of such a setup, inference for asset pricing models with downside risk and latent factors is a challenging task. In addition to the nonlinearity of the model, allowing for different regimes means that the number of common latent factors may be different between good and bad states of the world; the risk exposures on the factors may also change across states; the "disappointment event" that triggers the downside state is, in general, unknown; and, finally, the estimation of risk premia in this context, taking downside risk into account, needs to be fully developed. Our paper offers a methodological contribution, addressing all these issues, and presenting a nonlinear model of asset returns in the presence of common factors and downside risk; further, we consider a *latent* factor model approach, which allows to span the entire factor space as advocated by Giglio and Xiu (2021). This setup is general enough to encompass previous models with downside risk, while allowing for different common factors across states and estimation of the threshold determining the downside state.[1]

We make at least three contributions to the literature on factor models of asset returns. First, while we follow the literature in determining the disappointment event in terms of the relative position of the market return with respect to a threshold value, unlike previous studies we estimate the value of the threshold rather than setting it *a priori* equal to a pre-specified value. This is achieved by generalizing the inferential theory for panel threshold models proposed by Massacci (2017) to the case where the number of common factors and their loadings are allowed to differ across states; we also derive the relevant inferential theory for the estimated threshold, and propose a methodology to carry out hypothesis testing for the threshold value, e.g. to test whether it is equal to a pre-specified value such as the ones used in the literature. Second, we develop a procedure to estimate the number of common factors in a model with downside risk, by explicitly allowing for the case where such number may differ across regimes. This is done in a similar fashion to Trapani (2018), although it is not a trivial extension since it requires a preliminary round of estimation where the full factor model is estimated (thereby including the threshold,

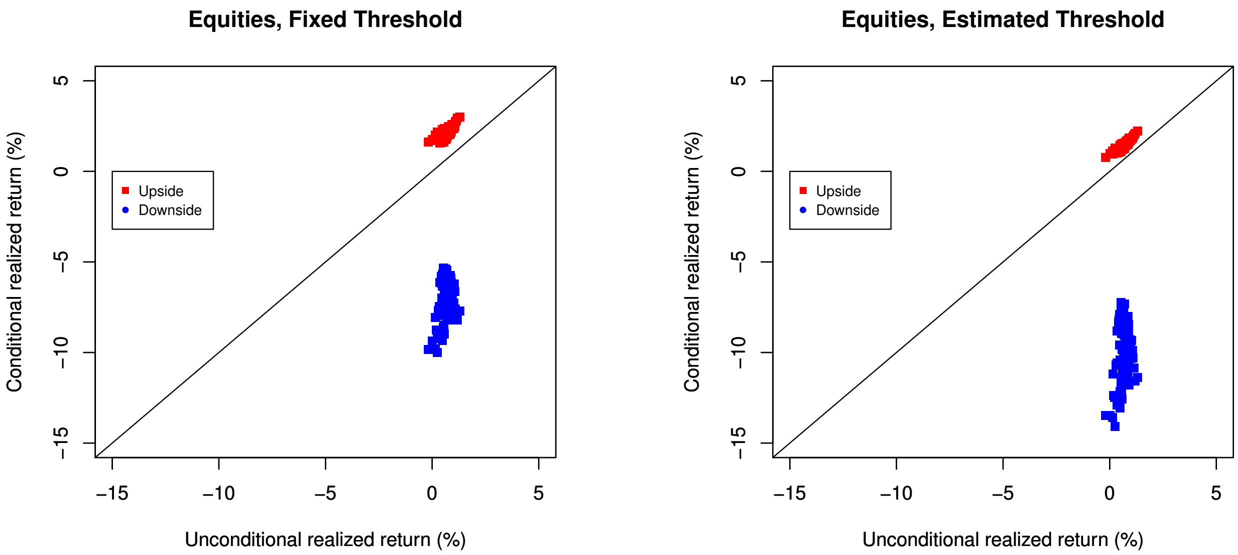
[1]The motivation for a downside risk premium arises from various theoretical perspectives. In some cases, the rationale is behavioral and linked to theories of loss aversion, as in the setups of Ang, Chen, and Xing (2006) and Farago and Tédongap (2018), whereas in other cases it is based on rational asset pricing models, as in Lettau, Maggiori, and Weber (2014). The specific nature of the mechanism capable of generating the downside risk premium is not a focus of this paper. We simply take seriously the finding that such premium can exist in equilibrium and ask how it should be identified in an empirical asset pricing model.

and the spaces spanned by factors and loadings), and the strong consistency of all estimates is required. Finally, we complete the inferential theory by developing a further step, to allow for the mapping of the estimated latent factors in each stage onto economically relevant observable ones. This step can be viewed as an extension of Giglio and Xiu (2021) to a nonlinear setting, delivering estimates of risk premia for any candidate risk factor that are robust to omitted variables and measurement error biases. Overall, the distilled essence of these results is a general conditional model of asset returns with downside risk that can be estimated and tested without imposing arbitrary restrictions on the stochastic process of returns. In our view, this setup constitutes at least a workhorse model for conducting empirical asset pricing in a downside risk framework.

We employ this methodology to study empirically the factor structure and estimate risk premia in the presence of downside risk using two datasets previously employed in other studies - a dataset composed of equity portfolios (the same as used by Farago and Tédongap, 2018) and a dataset spanning multiple asset classes (based on a variety of sources similar to Lettau et al., 2014). In both cases, our approach produces superior goodness of fit relative to both unconditional (linear) factor models and restricted conditional models that impose the number of factors or the threshold. The empirical estimation of the proposed latent factor model with downside risk proves illuminating when compared to an unconditional model of asset returns. Indeed, the unconditional model performs poorly in both states of the world, with systematic pricing errors that are positive in the downside state and negative in the satisfactory state - i.e., the unconditional model predicts much higher returns than the realized ones in the downside (bad) state, and lower returns than the realized ones in the satisfactory (good) state. Puzzingly, when analyzing the full sample performance of the unconditional model, its performance can appear satisfactory, but this is only because averaging across negative and positive pricing errors in different states gives the false impression that the model can capture the average excess returns accurately. In other words, assessing the performance of the unconditional factor model over the full sample can give the illusion that it performs well when in fact it performs poorly in both regimes. In contrast, the latent factor model with downside risk appears to have no systematic tendency to generate positive or negative pricing errors in either regime, and therefore also unconditionally, fitting the cross-sections of asset returns used here very well.

The inadequacy of the unconditional model is well illustrated by Figure 1, which plots unconditional portfolio equity returns against average portfolio equity returns conditional on either good or bad states (defined with a threshold of −3 and −6 percent on the left-hand-side and right-hand-side plots, respectively). The scatter plots make clear how a linear approximation underestimates returns in good times and overestimates returns in bad times. It is also clear that a linear relationship cannot possibly fit these returns data in the downside state, making apparent the need for a nonlinear model that changes across states.

Figure 1: Conditional and unconditional realized returns for equity portfolios



Notes. The scatter plot shows realized average returns in upside and downside states against unconditional average returns for the 130 equity portfolios described in Section 6.1.1 for fixed and estimated threshold values, which are discussed in Section 6.2.1.

Our empirical results highlight the importance of estimating the threshold (rather than imposing it *a priori*) and of allowing for different common factors across states. For example, with respect to the estimation of the threshold, we find a remarkable similarity in its point estimate across the two datasets employed, with the estimate being around −6 percent. This estimate is twice as large in absolute value as the most common value imposed in estimation by researchers, which is −3 percent (see Farago and Tédongap, 2018; and also, albeit implicitly, Lettau et al., 2014), thereby implying that the downside risk state is characterized by a smaller fraction of the return distribution with more extreme negative returns. We also find that, while up to six common factors are needed in the satisfactory state of the world to capture the factor structure of returns, at most three common factors are needed in the downside risk state. This is consistent with the widely held view, and much anecdotal evidence, that returns display a lower-dimensional factor structure in bad times than in good times, i.e. that diversification benefits diminish in bad times when they are needed most (see e.g. Cappiello et al., 2006). This result arises endogenously in the model, estimated using our procedure that explicitly allows the factors to change across states. Finally, the model uncovers the existence of highly asymmetric dynamics in risk premia across good and bad states of the world, both for widely used tradable factors (e.g. the Fama-French factors) and non-tradable risk factors (e.g. volatility and liquidity risk, intermediary capital, and others). In short, the empirical results provide a strong case for the full-blown, unconstrained estimation of latent factor models with downside risk in the context of asset pricing.

Beyond the aforementioned contributions, our paper relates to (at least) three additional strands of literature. The first one studies asset pricing with asymmetric pricing of "good" and "bad" covolatility: in a continuous time framework, Bollerslev, Li, and Todorov (2016) show that continuous and jump CAPM betas are priced in a different way; Bollerslev, Patton, and Quaedvlieg (2022) decompose the market beta into four semibetas, which depend on the covariation between the return on the market and that on individual assets; Bollerslev (2022) provides a comprehensive overview of contributions on pricing "good" and "bad" realized volatilities. Second, our paper is also linked to the literature on consumption-based asset pricing with disappointment aversion: Delikouras (2017) develops a model in which disappointment events take place when lifetime utility falls below its certainty equivalent; Delikouras and Kostakis (2019) consider the more general setup in which the disappointment threshold is a multiple of the certainty equivalent. Finally, our paper is related to the broader literature on conditional asset pricing that allows for time variation in risk exposures and/or the price of risk, and there are many examples of this approach for equity returns (e.g. Ferson and Harvey, 1991; Boguth et al., 2011), currency returns (Lustig et al., 2011; Colacito et al., 2020), and hedge fund returns (Patton and Ramadorai, 2013).

The remainder of the paper is organized as follows. Section 2 introduces the asset pricing model. Section 3 deals with identification of risk premia. Section 4 presents the methodology to determine the number of pricing factors. Section 5 discusses estimation and inference of pricing factors, risk exposures and risk premia. Section 6 presents the empirical analysis. Section 7 concludes. In the paper, we only present the main results for the sake of a concise discussion; assumptions, simulations, technical lemmas and proofs are all relegated to the Internet Appendix.

2 Model

In this section, we first present, for convenience, the setup of the unconditional asset pricing model that is most commonly used in the literature (Section 2.1). We then discuss the conditional asset pricing model with downside risk, whose estimation is the main contribution of this paper (Section 2.2).

At the outset, it is useful to establish some notation. We denote the ordinary limit as "$\to$". $\mathbb{I}(\cdot)$ denotes the indicator function. We use $\boldsymbol{\iota}_N$ to denote an N-dimensional vector of ones; I_k is an identity matrix of dimension k; finally, we set $C_{N,T} = \min\left\{N^{1/2}, T^{1/2}\right\}$.

2.1 Unconditional asset pricing model

Let $R_{i,t}$ be the return (in excess of the risk-free rate) on the test asset i at time t. The linear, or unconditional, asset pricing model can be written as

$$R_{i,t} = \gamma_0 + \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{\gamma}_1 + \boldsymbol{\beta}_i' \mathbf{u}_t + \epsilon_{i,t}, \tag{1}$$

with $1 \leq i \leq N$, $1 \leq t \leq T$, where N and T denote the cross section and time series dimensions of the available sample, respectively, and

$$\mathbf{u}_t = \mathbf{f}_t - \boldsymbol{\mu}_f. \tag{2}$$

In Eq. (2), $\mathbf{f}_t = (f_{1,t}, \ldots, f_{P,t})'$ is a $P \times 1$ vector of latent factors with $E(\mathbf{f}_t) = \boldsymbol{\mu}_f$, so that $\mathbf{u}_t$ in Eq. (1) represents the $P \times 1$ zero mean vector of factor innovations; further, in Eq. (1), γ_0 is the zero-beta rate; $\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \ldots, \gamma_{1,P})'$ is the $P \times 1$ vector of risk premia; $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,P})'$ is the

$P \times 1$ vector of risk exposures; α_i is the pricing error; and $\epsilon_{i,t}$ is the idiosyncratic component such that $E(\epsilon_{i,t}) = 0$ and $E(\mathbf{u}_t \epsilon_{i,t}) = 0$. In matrix notation, the model becomes

$$\mathbf{R}_t = \gamma_0 \boldsymbol{\iota}_N + \boldsymbol{\alpha} + \mathbf{B}\gamma_1 + \mathbf{B}\mathbf{u}_t + \boldsymbol{\epsilon}_t, \tag{3}$$

where $\mathbf{R}_t = (R_{1t}, \ldots, R_{Nt})'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$, and $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)'$.

Following Giglio and Xiu (2021), we let some of the true factors be omitted or measured with error (or both). In particular, the K observable factors $\mathbf{g}_t$ can be either tradable or nontradable (or both), and have the structure

$$\mathbf{g}_t = \mathbf{a} + \boldsymbol{\Lambda}\mathbf{u}_t + \mathbf{e}_t. \tag{4}$$

As shown in Giglio and Xiu (2021), given $\boldsymbol{\gamma}_1$, the $K \times 1$ vector of risk premia of $\mathbf{g}_t$ is $\gamma_{\mathbf{g}} = \boldsymbol{\Lambda}\gamma_1$.[2]

2.2 Conditional asset pricing model with downside risk

The conditional factor model of interest is:

$$\begin{aligned} R_{i,t} &= \mathbb{I}(\mathcal{D}_t)\left(\gamma_{\mathcal{D},0} + \alpha_{\mathcal{D},i} + \boldsymbol{\beta}'_{\mathcal{D},i}\boldsymbol{\gamma}_{\mathcal{D},1} + \boldsymbol{\beta}'_{\mathcal{D},i}\mathbf{u}_{\mathcal{D},t}\right) \\ &\quad + \mathbb{I}(\mathcal{S}_t)\left(\gamma_{\mathcal{S},0} + \alpha_{\mathcal{S},i} + \boldsymbol{\beta}'_{\mathcal{S},i}\boldsymbol{\gamma}_{\mathcal{S},1} + \boldsymbol{\beta}'_{\mathcal{S},i}\mathbf{u}_{\mathcal{S},t}\right) + \epsilon_{i,t}, \end{aligned} \tag{5}$$

where

$$\mathbf{u}_{j,t} = \mathbf{f}_{j,t} - \boldsymbol{\mu}_{j,f}, \quad j = \mathcal{D}, \mathcal{S}. \tag{6}$$

The model in Eq. (5) generalizes Eq. (1) to allow for downside risk, as captured by the disappointment event $\mathcal{D}_t$. The satisfactory event $\mathcal{S}_t$ occurs whenever $\mathcal{D}_t$ does not take place, i.e. $\mathbb{I}(\mathcal{D}_t) + \mathbb{I}(\mathcal{S}_t) = 1$ and $\mathbb{I}(\mathcal{D}_t)\,\mathbb{I}(\mathcal{S}_t) = 0$. Section 2.3 below discusses $\mathcal{D}_t$ and $\mathcal{S}_t$ in detail. In Eq. (6), $\mathbf{f}_{j,t} = \left(f_{j,1,t}, \ldots, f_{j,P_j,t}\right)'$ is the $P_j \times 1$ vector of latent factors satisfying $E\left[\mathbf{f}_{j,t} \,|\mathbb{I}(j_t) = 1\right] = \mu_{j,f}$; $\mathbf{u}_{j,t}$ is the $P_j \times 1$ vector of factor innovations such that $E\left[\mathbf{u}_{jt} \,|\mathbb{I}(j_t) = 1\right] = 0$. As for (5): $\gamma_{j,0}$ is the zero-beta rate and $\boldsymbol{\gamma}_{j,1} = \left(\gamma_{j,1,1}, \ldots, \gamma_{j,1,P_j}\right)'$ is the $P_j \times 1$ vector of risk premia; $\boldsymbol{\beta}_{j,i} = \left(\beta_{j,i,1}, \ldots, \beta_{j,i,P_j}\right)'$

[2]This is the same setup as Giglio and Xiu (2021), who impose the zero-beta rate $\gamma_0 = 0$ in their exposition, and illustrate the case for $\gamma_0 \neq 0$ in their Internet Appendix. While we allow for a zero-beta rate for generality in our exposition, we also do not allow for a constant in the estimation of risk premia in the empirical work below since the constant is difficult to interpret (just like a constant in the second step of the Fama-MacBeth procedure).

is the $P_j \times 1$ vector of risk exposures; $\alpha_{j,i}$ is the pricing error; the idiosyncratic components $\epsilon_{i,t}$ satisfy $E\left[\mathbf{u}_{j,t}\epsilon_{i,t}\left|\mathbb{I}\left(j_t\right)=1\right.\right]=0$. The model in matrix form is

$$\begin{aligned}\mathbf{R}_t &= \mathbb{I}\left(\mathcal{D}_t\right)\left(\gamma_{\mathcal{D},0}\boldsymbol{\iota}_N+\boldsymbol{\alpha}_{\mathcal{D}}+\mathbf{B}_{\mathcal{D}}\boldsymbol{\gamma}_{\mathcal{D},1}+\mathbf{B}_{\mathcal{D}}\mathbf{u}_{\mathcal{D},t}\right) \\ &\quad +\mathbb{I}\left(\mathcal{S}_t\right)\left(\gamma_{\mathcal{S},0}\boldsymbol{\iota}_N+\boldsymbol{\alpha}_{\mathcal{S}}+\mathbf{B}_{\mathcal{S}}\boldsymbol{\gamma}_{\mathcal{S},1}+\mathbf{B}_{\mathcal{S}}\mathbf{u}_{\mathcal{S},t}\right)+\boldsymbol{\epsilon}_t,\end{aligned} \tag{7}$$

where $\boldsymbol{\alpha}_j=(\alpha_{j,1},\ldots,\alpha_{j,N})'$, and $\mathbf{B}=(\boldsymbol{\beta}_{j,1},\ldots,\boldsymbol{\beta}_{j,N})'$. We generalize the mapping between the true latent factors and observable factors in Eq. (4) as

$$\mathbf{g}_t=\mathbb{I}\left(\mathcal{D}_t\right)\left(\mathbf{a}_{\mathcal{D}}+\boldsymbol{\Lambda}_{\mathcal{D}}\mathbf{u}_{\mathcal{D},t}\right)+\mathbb{I}\left(\mathcal{S}_t\right)\left(\mathbf{a}_{\mathcal{S}}+\boldsymbol{\Lambda}_{\mathcal{S}}\mathbf{u}_{\mathcal{S},t}\right)+\mathbf{e}_t. \tag{8}$$

The $K \times 1$ vector of risk premia of $\mathbf{g}_t$ becomes

$$\boldsymbol{\gamma}_{\mathbf{g}}=\mathbb{I}\left(\mathcal{D}_t\right)\boldsymbol{\gamma}_{\mathcal{D},\mathbf{g}}+\mathbb{I}\left(\mathcal{S}_t\right)\boldsymbol{\gamma}_{\mathcal{S},\mathbf{g}}, \quad \boldsymbol{\gamma}_{j,\mathbf{g}}=\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_{j,1}, \quad j=\mathcal{D},\mathcal{S}. \tag{9}$$

When the disappointment event occurs, we thus have $\boldsymbol{\gamma}_{\mathbf{g}}=\boldsymbol{\gamma}_{\mathcal{D},\mathbf{g}}=\boldsymbol{\Lambda}_{\mathcal{D}}\boldsymbol{\gamma}_{\mathcal{D},1}$.

2.3 Disappointment event

Following, *inter alia*, Farago and Tédongap (2018), we define the disappointment event as $\mathcal{D}_t=\{r_{W,t}\leq\theta\}$, where $r_{W,t}$ is the log-return on the market and θ is the corresponding threshold value that determines the occurrence of $\mathcal{D}_t$. The satisfactory event then is $\mathcal{S}_t=\{r_{W,t}>\theta\}$. Both $\mathcal{D}_t$ and $\mathcal{S}_t$ depend on θ: to stress this dependence, we write $d_{j,t}(\theta)=\mathbb{I}\left(j_t\right)$, for $j=\mathcal{D},\mathcal{S}$.[3]

[3]The conditional asset pricing model in Eq. (7) allows for unrestricted zero-beta rates $\gamma_{\mathcal{D},0}$ and $\gamma_{\mathcal{S},0}$, and for non-zero pricing errors $\boldsymbol{\alpha}_{\mathcal{D}}$ and $\boldsymbol{\alpha}_{\mathcal{S}}$. None of these invalidates our methodology. In particular, non-zero conditional pricing errors, which for instance may arise from failing no arbitrage conditions, do not affect estimation and inference on the risk premia.

3 Identification

3.1 Identification of risk premia

In the conditional factor model in Eq. (7), there are two kinds of identification issues: *(i)* the rotational indeterminacy typical of statistical factor models; and *(ii)* the unobserved heterogeneity problem induced by the pricing errors.

Rotational indeterminacy simply means that risk exposures, pricing factors and risk premia in the linear model in Eq. (3) are identified up to a linear transformation. Indeed, for any $P \times P$ positive definite matrix $\mathbf{L}$ we have

$$\begin{aligned} \mathbf{R}_t &= \gamma_0 \boldsymbol{\iota}_N + \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\gamma}_1 + \mathbf{B}\mathbf{u}_t + \boldsymbol{\epsilon}_t \\ &= \gamma_0 \boldsymbol{\iota}_N + \boldsymbol{\alpha} + \mathbf{B}\mathbf{L}\mathbf{L}^{-1}\boldsymbol{\gamma}_1 + \mathbf{B}\mathbf{L}\mathbf{L}^{-1}\mathbf{u}_t + \boldsymbol{\epsilon}_t. \end{aligned}$$

Hence, $\mathbf{B}$, $\boldsymbol{\gamma}_1$ and $\mathbf{u}_t$ are only identified up to $\mathbf{L}$, as they are all unobservable. However, as shown in Giglio and Xiu (2021), the risk premia for the observable factors $\mathbf{g}_t$ in Eq. (4) are identified, since it holds that $\gamma_{\mathbf{g}} = \mathbf{\Lambda}\gamma_1 = \mathbf{\Lambda}\mathbf{L}\mathbf{L}^{-1}\gamma_1$.

The conditional pricing model in Eq. (5) faces essentially the same identification features. For any $P_j \times P_j$ positive definite matrix $\mathbf{L}_j$, with $j = \mathcal{D}, \mathcal{S}$, we have

$$\begin{aligned} \mathbf{R}_t &= d_{\mathcal{D},t}(\theta)\left(\gamma_{\mathcal{D},0}\boldsymbol{\iota}_N + \boldsymbol{\alpha}_{\mathcal{D}} + \mathbf{B}_{\mathcal{D}}\mathbf{L}_{\mathcal{D}}\mathbf{L}_{\mathcal{D}}^{-1}\boldsymbol{\gamma}_{\mathcal{D},1} + \mathbf{B}_{\mathcal{D}}\mathbf{L}_{\mathcal{D}}\mathbf{L}_{\mathcal{D}}^{-1}\mathbf{u}_{\mathcal{D},t}\right) \\ &\quad + d_{\mathcal{S},t}(\theta)\left(\gamma_{\mathcal{S},0}\boldsymbol{\iota}_N + \boldsymbol{\alpha}_{\mathcal{S}} + \mathbf{B}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}^{-1}\boldsymbol{\gamma}_{\mathcal{S},1} + \mathbf{B}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}^{-1}\mathbf{u}_{\mathcal{S},t}\right) + \boldsymbol{\epsilon}_t. \end{aligned}$$

The risk premia $\boldsymbol{\gamma}_{\mathbf{g}}$ defined in Eq. (9), however, are identified, since

$$\begin{aligned} \boldsymbol{\gamma}_{\mathbf{g}} &= d_{\mathcal{D},t}(\theta)\mathbf{\Lambda}_{\mathcal{D}}\boldsymbol{\gamma}_{\mathcal{D},1} + d_{\mathcal{S},t}(\theta)\mathbf{\Lambda}_{\mathcal{S}}\boldsymbol{\gamma}_{\mathcal{S},1} \\ &= d_{\mathcal{S},t}(\theta)\mathbf{\Lambda}_{\mathcal{D}}\mathbf{L}_{\mathcal{D}}\mathbf{L}_{\mathcal{D}}^{-1}\boldsymbol{\gamma}_{\mathcal{D},1} + d_{\mathcal{S},t}(\theta)\mathbf{\Lambda}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}^{-1}\boldsymbol{\gamma}_{\mathcal{S},1}. \end{aligned}$$

As far as the second identification issue is concerned, we solve the unobserved heterogeneity problem induced by the pricing errors by expressing the model for $\mathbf{R}_t$ in terms of deviations from the conditional means. Formally, let $T_j(\theta) = \sum_{t=1}^{T} d_{j,t}(\theta)$ be the number of times that the event j occurs, for $j = \mathcal{D}, \mathcal{S}$. Define the conditional average returns $\overline{\mathbf{R}}_j(\theta) = T_j(\theta)^{-1} \sum_{t=1}^{T} d_{j,t}(\theta)\mathbf{R}_t$: $\overline{\mathbf{R}}_j(\theta)$ is the mean of $\mathbf{R}_t$ when j occurs. In order to estimate pricing factors and risk exposures, we consider the model

$$\widetilde{\mathbf{R}}_t(\theta) = \mathbf{R}_t - d_{\mathcal{D},t}(\theta)\overline{\mathbf{R}}_{\mathcal{D}}(\theta) - d_{\mathcal{S},t}(\theta)\overline{\mathbf{R}}_{\mathcal{S}}(\theta) = d_{\mathcal{D},t}(\theta)\mathbf{B}_{\mathcal{D}}\tilde{\mathbf{u}}_{\mathcal{D},t}(\theta) + d_{\mathcal{S},t}(\theta)\mathbf{B}_{\mathcal{S}}\tilde{\mathbf{u}}_{\mathcal{S}t}(\theta) + \tilde{\boldsymbol{\epsilon}}_t \tag{10}$$

where $\tilde{\mathbf{u}}_{j,t}(\theta) = \mathbf{u}_t - \overline{\mathbf{u}}_j(\theta)$ is the deviation of $\mathbf{u}_t$ from the conditional mean $\overline{\mathbf{u}}_j(\theta) = T_j(\theta)^{-1} \sum_{t=1}^{T} d_{j,t}(\theta)\mathbf{u}_t$; and $\tilde{\boldsymbol{\epsilon}}_t = d_{\mathcal{D},t}(\theta)[\boldsymbol{\epsilon}_t - \bar{\boldsymbol{\epsilon}}_{\mathcal{D}}(\theta)] + d_{\mathcal{S},t}(\theta)[\boldsymbol{\epsilon}_t - \bar{\boldsymbol{\epsilon}}_{\mathcal{S}}(\theta)]$, where $\bar{\boldsymbol{\epsilon}}_j(\theta) = T_j(\theta)^{-1} \sum_{t=1}^{T} d_{j,t}(\theta)\boldsymbol{\epsilon}_{j,t}$ is the conditional mean of ϵ_t when j occurs.

3.2 The disappointment threshold

As implicitly noted in Section 3.1, the identification of factors, exposures and risk premia requires knowledge of the disappointment event $\mathcal{D}_t$, and thus of the threshold θ. Typically, the literature pre-specifies a value for θ, and the subsequent estimation of the asset pricing model is then conditioned on that value being true.[4] In general, however, θ is not a structural parameter on which theory can provide guidance, which makes it difficult to assume a particular value of θ. In this paper, we study the Least Squares estimator of θ (Section 4.1).

4 Inference

In this section, we study inference on Eq. (5): we begin by presenting the estimation and hypothesis testing on θ (Section 4.1); we then discuss the estimation of the number of common factors in each regime (Section 4.2); and, finally, we present the Principal Component estimation of $\boldsymbol{\beta}_{j,i}$ and $\boldsymbol{\gamma}_{j,1}$ (Section 4.3).

We begin by introducing some further notation. The superscript "^{0}" denotes the true value of a parameter; when this is not used, we refer to a generic value. We also use the short-

[4]For example, as mentioned in the introduction, Farago and Tédongap (2018) *a priori* set $\theta = -0.03$ and run robustness checks for $\theta \in \{0, -0.015, -0.04\}$.

hand notation $T_j = T_j(\theta^0)$ and $\pi_j = \pi_j(\theta^0) = T_j(\theta^0)/T$; further, the true number of factors in each regime is denoted as $P_{\mathcal{D}}^0$ and $P_{\mathcal{S}}^0$ respectively, also using, when needed, $P^0 = P_{\mathcal{D}}^0 + P_{\mathcal{S}}^0$. Finally, here and throughout the whole paper, P^* denotes the probability conditional on $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$; we use "$\stackrel{D^*}{\rightarrow}$" and "$\stackrel{P^*}{\rightarrow}$" to indicate convergence in distribution and in probability according to P^*, respectively.

4.1 Estimating θ^0

We begin by studying the estimation of θ^0. Recall the $N \times 1$ vector of demeaned returns $\widetilde{\mathbf{R}}_t(\theta) = \left(\widetilde{R}_{1,t}, \ldots, \widetilde{R}_{N,t}\right)'$, for $1 \leq t \leq T$, defined in Eq. (10). For generic numbers of factors $P_{\mathcal{D}}$, $P_{\mathcal{S}}$ and $P = P_{\mathcal{D}} + P_{\mathcal{S}}$, define the $N \times P$ matrix of loadings $\mathbf{B}^P = \left(\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}}, \mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}}\right)$, and the associated $P_{\mathcal{D}} \times T$ and $P_{\mathcal{S}} \times T$ matrices of demeaned factor innovations $\mathbf{U}_{\mathcal{D}}^{P_{\mathcal{D}}} = \left(\mathbf{u}_{\mathcal{D},1}^{P_{\mathcal{D}}}, \ldots, \mathbf{u}_{\mathcal{D},T}^{P_{\mathcal{D}}}\right)$ and $\mathbf{U}_{\mathcal{S}}^{P_{\mathcal{S}}} = \left(\mathbf{u}_{\mathcal{S},1}^{P_{\mathcal{S}}}, \ldots, \mathbf{u}_{\mathcal{S},T}^{P_{\mathcal{S}}}\right)$, with $\mathbf{U}^P = \left(\mathbf{U}_{\mathcal{D}}^{P_{\mathcal{D}}\prime}, \mathbf{U}_{\mathcal{S}}^{P_{\mathcal{S}}\prime}\right)'$.

The objective function in terms of $\mathbf{B}^P$, $\mathbf{U}^P$ and θ, is given by the sum of squared residuals (divided by NT):

$$S\left(\mathbf{B}^P, \mathbf{U}^P, \theta\right) = \frac{1}{NT}\sum_{t=1}^{T}\left\|\widetilde{\mathbf{R}}_t(\theta) - d_{\mathcal{D},t}(\theta)\,\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}}\mathbf{u}_{\mathcal{D},t}^{P_{\mathcal{D}}} - d_{\mathcal{S},t}(\theta)\,\mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}}\mathbf{u}_{\mathcal{S},t}^{P_{\mathcal{S}}}\right\|^2. \tag{11}$$

For given P, the estimators of $\widehat{\mathbf{B}}^P = \left(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}, \widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\right)$ – with $\widehat{\mathbf{B}}_j^{P_j} = \left(\widehat{\beta}_{j,1}, ..., \widehat{\beta}_{j,N}\right)'$ for $j = \mathcal{D}, \mathcal{S}$ – and $\widehat{\mathbf{U}}^P$ and $\widehat{\theta}^P$ are defined as

$$\left\{\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \widehat{\theta}^P\right\} = \arg\min_{\mathbf{B}^P, \mathbf{U}^P, \theta} S\left(\mathbf{B}^P, \mathbf{U}^P, \theta\right).$$

In order to solve this minimization problem, we begin by concentrating the objective function $S\left(\mathbf{B}^P, \mathbf{U}^P, \theta\right)$ with respect to $\mathbf{B}^P$ and θ, subject to the constraints $N^{-1}\mathbf{B}_j^{P_j\prime}\mathbf{B}_j^{P_j} = I_{P_j}$ for $j = \mathcal{D}, \mathcal{S}$. It follows that the estimator of $\mathbf{U}^P$ is

$$\widehat{\mathbf{u}}_{\mathcal{D},t}^{P_{\mathcal{D}}}\left(\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}}, \theta\right) = N^{-1}\left(d_{\mathcal{D},t}(\theta)\,\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}}(\theta)\right)'\widetilde{\mathbf{R}}_t,$$
$$\widehat{\mathbf{u}}_{\mathcal{S},t}^{P_{\mathcal{S}}}\left(\mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}}, \theta\right) = N^{-1}\left(d_{\mathcal{S},t}(\theta)\,\mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}}(\theta)\right)'\widetilde{\mathbf{R}}_t.$$

Hence, we can define

$$S\left(\mathbf{B}^P,\theta\right)=\frac{1}{NT}\sum_{t=1}^{T}\left\|\widetilde{\mathbf{R}}_t(\theta)-d_{\mathcal{D},t}\left(\theta\right)\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}}\hat{\mathbf{u}}_{\mathcal{D},t}^{P_{\mathcal{D}}}\left(\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}},\theta\right)-d_{\mathcal{S},t}\left(\theta\right)\mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}}\hat{\mathbf{u}}_{\mathcal{S},t}^{P_{\mathcal{S}}}\left(\mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}},\theta\right)\right\|^2.$$

Letting

$$\widehat{\mathbf{\Sigma}}_{j,\widetilde{\mathbf{R}}}\left(\theta\right)=\frac{1}{NT}\sum_{t=1}^{T}d_{j,t}\left(\theta\right)\widetilde{\mathbf{R}}_t\left(\theta\right)\widetilde{\mathbf{R}}_t\left(\theta\right)',\tag{12}$$

for $j=\mathcal{D},\mathcal{S}$, the Principal Component estimator $\widehat{\mathbf{B}}_j^{P_j}\left(\theta\right)$ for $\mathbf{B}_j^{P_j}$ given θ is $\sqrt{N}$ times the $N\times P_j$ matrix of eigenvectors of $\widehat{\mathbf{\Sigma}}_{j,\widetilde{\mathbf{R}}}\left(\theta\right)$ corresponding to its largest P_j eigenvalues, under the orthonormalization restriction $N^{-1}\left(\widehat{\mathbf{B}}_j^{P_j}\left(\theta\right)'\widehat{\mathbf{B}}_j^{P_j}\left(\theta\right)\right)=\mathbf{I}_{P_j}$, for $j=\mathcal{D},\mathcal{S}$ (see e.g. Bai, 2003).

We now define

$$S\left(\theta\right)=\frac{1}{NT}\sum_{t=1}^{T}\left\|\widetilde{\mathbf{R}}_t(\theta)-d_{\mathcal{D},t}\left(\theta\right)\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta\right)\hat{\mathbf{u}}_{\mathcal{D},t}^{P_{\mathcal{D}}}\left(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta\right);\theta\right)-d_{\mathcal{S},t}\left(\theta\right)\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta\right)\hat{\mathbf{u}}_{\mathcal{S},t}^{P_{\mathcal{S}}}\left(\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta\right);\theta\right)\right\|^2.$$

Let $\widehat{\mathbf{B}}^P\left(\theta\right)=\left[\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta\right),\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta\right)\right]$ and $\widehat{\mathbf{U}}^P\left(\theta\right)=\left[\widehat{\mathbf{U}}_{\mathcal{D}}^{P_{\mathcal{D}}\prime}\left(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta\right);\theta\right),\widehat{\mathbf{U}}_{\mathcal{S}}^{P_{\mathcal{S}}\prime}\left(\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta\right);\theta\right)\right]'$. For $\overline{P}=\overline{P}_{\mathcal{D}}+\overline{P}_{\mathcal{S}}$ with $\overline{P}_{\mathcal{D}}\geq P_{\mathcal{D}}^0$ and $\overline{P}_{\mathcal{S}}\geq P_{\mathcal{S}}^0$, we finally have the estimator

$$\widehat{\theta}=\arg\min_{\theta}S\left[\widehat{\mathbf{B}}^{\overline{P}}\left(\theta\right),\widehat{\mathbf{U}}^{\overline{P}}\left(\theta\right),\theta\right];\tag{13}$$

$\widehat{\theta}$ can be obtained e.g. through a grid search. In Section B of the Internet Appendix, we complement the theory with a small-scale simulation exercise, reporting bias and mean squared error for the estimated θ^0. We find that the estimator works well in general, especially when the factor structure of the two regimes differs substantially.

In the Internet Appendix, we show that $\widehat{\theta}$ is a strongly consistent estimator of θ^0, deriving its rate. In principle, the limiting distribution of $\widehat{\theta}$ can be derived along similar steps as in Chan (1993), and can be employed to carry out inference on θ^0. However, this is bound to depend on several nuisance parameters, and therefore the asymptotic distribution of $\widehat{\theta}$ would be only of theoretical interest. Hence, in this paper we propose a different approach to test for null hypotheses of the type

$$H_0 : \theta^0 = c, \tag{14}$$

where c is a value of interest; we note that being able to test for (14) also offers the possibility of computing confidence intervals, by inverting the test. This procedure does not require the limiting distribution of $\widehat{\theta}$, and it relies only on the rate of convergence, using an approach similar to the one in Horváth and Trapani (2019). In particular, this approach is very similar to the one proposed in Section 4.2, and therefore, in order to avoid repetitions, the details are relegated to Section A.2 of the Internet Appendix, to which we refer the interested reader.

4.2 Determining the number of common factors

After estimating θ, it is possible to procede to the estimation of the factor structures in each regime. To this end, the first step of the analysis is the estimation of the dimensions of the factor spaces $P_{\mathcal{D}}^0$ and $P_{\mathcal{S}}^0$. In principle, after splitting the sample around the estimated threshold $\widehat{\theta}$, any of the existing techniques can be employed. Given that, in the presence of a factor structure, the sample second moment matrix of the data has as many spiked eigenvalues as there are common factors, the vast majority of the existing methodologies are based on thresholding such eigenvalues. A possible approach (see Bai and Ng, 2002) is to use a threshold based on the penalty function of some information criterion; however, the critique in Hallin and Liška (2007) suggests that such a threshold may be quite arbitrary, and the resulting estimate may not be robust to its tuning. A different approach (see, *inter alia*, Ahn and Horenstein, 2013, and Lam and Yao, 2012) is based on using the ratio of adjacent eigenvalues - thus, in essence, using each eigenvalue as a "natural" threshold for the previous eigenvalue. Whilst this technique ameliorates the arbitrariness connected with thresholding methods, existing contributions make - explicitly or implicitly - extensive use of (large) random matrix theory, which requires several constraints on the form and amount of serial and cross sectional dependence, and on the relative rate of divergence of N and T as they pass to infinity. These issues are overcome by the procedure proposed by Trapani (2018): being based on (sequential) testing, it avoids having to set an arbitrary threshold; also, the results in that paper hold for virtually any level of (weak) serial and cross-sectional dependence, and for any combination of N and T. Hence, in this section we extend the procedure suggested by Trapani (2018) to estimate the number of factors in each regime, $P^0_{\mathcal{D}}$ and $P^0_{\mathcal{S}}$. Such an extension is not trivial, since it requires controlling the regime-splitting errors coming from using an estimated value of θ.[5]

As in the previous section, here we only report the relevant algorithm; details and the theory are in Section A.3 of the Internet Appendix.

Let $\widehat{d}_{\mathcal{D},t} = \mathbb{I}\left(\widehat{\mathcal{D}}_t\right)$ and $\widehat{d}_{\mathcal{S},t} = 1 - \widehat{d}_{\mathcal{D},t}$, and define $\widehat{\pi}_{\mathcal{S}} = T^{-1}\sum_{t=1}^{T}\widehat{d}_{\mathcal{D},t}$ and $\widehat{\pi}_{\mathcal{D}} = 1 - \widehat{\pi}_{\mathcal{S}}$. Also, we henceforth use the short-hand notation $\widetilde{\mathbf{R}}_t = \widetilde{\mathbf{R}}_t(\theta)$. We use the sample covariance matrices

$$\widehat{\Sigma}_{\mathcal{D}} = \frac{1}{T\widehat{\pi}_{\mathcal{D}}}\sum_{t=1}^{T}\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\widehat{d}_{\mathcal{D},t}, \tag{15}$$

$$\widehat{\Sigma}_{\mathcal{S}} = \frac{1}{T\widehat{\pi}_{\mathcal{S}}}\sum_{t=1}^{T}\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\widehat{d}_{\mathcal{S},t}, \tag{16}$$

and denote the i-th largest eigenvalue of $\widehat{\Sigma}_j$ as $\widehat{g}_j^{(i)}$, with $j = \mathcal{D}$ or $\mathcal{S}$.

Using Lemma A.1 and Theorem A.2 of Section A.3 in the Internet Appendix, we can determine the number of common factors in each regime using the following method, based on two separate steps.

In the first step, we study a test for the individual eigenvalues, i.e.

$$\begin{cases} H_0: & g_j^{(i)} = c_j^{(i)}N \\ H_A: & g_j^{(i)} \leq c_j^{(i)} \end{cases}, \tag{17}$$

for some $0 < c_j^{(i)} < \infty$ and $j = \mathcal{D}, \mathcal{S}$. In essence, the null hypothesis is that the i-th largest eigenvalue of the covariance matrix of the data - in each regime - diverges, thus suggesting that there are at least i common factors. Conversely, upon rejecting the null hypothesis, the conclusion can be drawn that there are fewer than i common factors in regime j. Thus, the second step of the analysis is to determine $P^0_{\mathcal{D}}$ and $P^0_{\mathcal{S}}$ by carrying out a sequence of tests for $i = 1, ..., P_{\max}$, with

[5] We would like to point out that, in principle, all other existing techniques could be extended to our setup, with the same difficulties as mentioned above.

$P_{\max}$ a user-defined upper bound.

We begin with (17). Define

$$\psi\left(\widehat{g}_j^{(i)}\right) = \exp\left(\frac{N^{-\varrho_j}\widehat{g}_j^{(i)}}{\overline{g}_j\left(p\right)}\right), \tag{18}$$

where[6]

$$\overline{g}_j\left(p\right) = \frac{1}{N-p}\sum_{h=p+1}^{N}\widehat{g}_j^{(h)}.$$

In Eq. (18), we have defined

$$\varrho_j = \begin{cases} \varepsilon & \text{when } \frac{\ln T_j}{\ln N} \geq \frac{1}{2} \\ 1-\frac{1}{2}\frac{\ln T_j}{\ln N}+\varepsilon & \text{when } \frac{\ln T_j}{\ln N} < \frac{1}{2} \end{cases}, \tag{19}$$

where $\varepsilon > 0$ is a (small) user-chosen number. We now offer a brief, heuristic discussion of the role of the tuning parameter ϱ_j. As we show in Lemma A.1, the *population* eigenvalues $g_j^{(i)}$ diverge as fast as N if the number of common factors equals or exceeds j, whereas they are bounded otherwise. However, according to Theorem A.2, the *sample* eigenvalues have an estimation error which may also diverge, albeit at a slower rate than N. The value of ϱ_j (which, as can be verified, is strictly smaller than 1) is chosen in such a way that $N^{-\varrho_j}\widehat{g}_j^{(i)}$ still passes to positive infinity whenever $g_j^{(i)}$ does, whereas it drifts to zero whenever $g_j^{(i)}$ is bounded: in other words, pre-multiplying $\widehat{g}_j^{(i)}$ by $N^{-\varrho_j}$ cancels the estimation error when $g_j^{(i)}$ is bounded, and it does not cancel the signal in $g_j^{(i)}$ when this is not bounded. Naturally, if ϱ_j is large (i.e., if ε in equation (19) is chosen as a "large" number), this will attenuate the estimation error in $\widehat{g}_j^{(i)}$, but it will also attenuate the signal in $\widehat{g}_j^{(i)}$ when $g_j^{(i)}$ diverges, and therefore it can be expected that the number of common factors may be understated. Conversely, a small value of ϱ_j (corresponding to using a "small" value of ε, e.g. $\varepsilon = 10^{-4}$) is less likely to attenuate the estimation error, and therefore it can be expected that the number of common factors may be overstated. In our empirical exercises, we explore the

[6] The purpose of $\overline{g}_j\left(p\right)$ is to rescale $\widehat{g}_j^{(i)}$; in Trapani (2018), it is suggested to set $p = 0$ (that is, to use the trace to rescale $\widehat{g}_j^{(i)}$); other choices are also possible, such as $p = i$, which is recommended in Barigozzi and Trapani (2022).

robustness of our methodology to estimate $P_{\mathcal{D}}^{0}$ and $P_{\mathcal{S}}^{0}$ to the choice of $\varrho_{\mathcal{D}}$ and $\varrho_{\mathcal{S}}$, respectively.

We now describe the test for (17).

Step 1 Generate an $i.i.d.N(0,1)$ sequence $\left\{\xi_{j,m}^{(i)}, 1 \leq m \leq M\right\}$, where the $\xi_{j,m}^{(i)}$s are independent across i and j.

Step 2 Define the Bernoulli sequence $\zeta_{j,m}^{(i)}(s) = \mathbb{I}\left(\psi\left(\widehat{g}_{j}^{(i)}\right) \times \xi_{i,m}^{(j)} \leq s\right)$.

Step 3 Compute

$$\Upsilon_{j}^{(i)} = \int_{-\infty}^{+\infty} \left(\upsilon_{j}^{(i)}(s)\right)^{2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}s^{2}\right) ds,$$

where

$$\upsilon_{j}^{(i)}(s) = \frac{2}{M^{1/2}} \sum_{m=1}^{M} \left(\zeta_{j,m}^{(i)}(s) - \frac{1}{2}\right).$$

We show that, under the null, $\Upsilon_{j}^{(i)} \overset{D^{*}}{\to} \chi_{1}^{2}$, thus providing a rule to decide between H_0 and H_A in (17). On account of this, it is possible to propose a sequential procedure to determine $P_{\mathcal{D}}^{0}$ and $P_{\mathcal{S}}^{0}$.

This is based on the following two-step algorithm, for $j = \mathcal{D}, \mathcal{S}$.

Step 1 Run the test for $H_0 : g_{j}^{(i)} = \infty$ based on $\Upsilon_{j}^{(1)}$. If the null is rejected, set $\widehat{P}_{j} = 0$ and stop, otherwise go to the next step.

Step 2 Starting from $i = 1$, run the test for $H_0 : g_{j}^{(i+1)} = \infty$ based on $\Upsilon_{j}^{(i+1)}$. If the null is rejected, set $\widehat{P}_{j} = i$ and stop; otherwise repeat the step until the null is rejected, or until the pre-specified maximum number $P_{\max}$ is reached.

Let $\alpha = \alpha(N,T)$ denote the level of the individual tests and $c_{\alpha} = c_{\alpha}(N,T)$ be the corresponding critical value. We show that, as long as the level of the individual tests goes to zero, $\widehat{P}_{j}$ is a consistent estimator.

Theorem 1. *We assume that Assumptions 1-6 in Section A.1 of the Internet Appenendix are satisfied. As* $\min(N,T_j) \to \infty$ *under (A.11), if* $c_{\alpha}(N,T) \to \infty$ *with* $c_{\alpha}(N,T) = o(M)$*, then it holds that*

$$\widehat{P}_{j} = P_{j}^{0} + o_{P^{*}}(1),$$

for almost all realizations of $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$*, and* $j = \mathcal{D}, \mathcal{S}$*.*

In the implementation of the algorithm, it is necessary to have an upper bound $P_{\max}$; this is chosen as

$$P_{\max} = O\left(C_{N,T}^{1-c}\right), \tag{20}$$

for some user-defined $c > 0$. The selection rule in (20) is as an extension of the otherwise customarily employed Schwert's rule; see Schwert (1989), and also the comment in Bai and Ng (2002). In the empirical analysis, we use and recommend $P_{\max} = \left\lfloor \min\left\{N^{1/3}, T^{1/3}\right\}\right\rfloor$. We note that, in Section B of the Internet Appendix, we report simulation evidence on the performance of the estimator of the number of common factors in finite samples, and results from trying some variants of (20) as robustness checks, but results are virtually unaffected. The main takeway from the simulation experiments is that the test for the number of factors is reliable in detecting the true number of common factors in finite sample setups of the data generating process that match the samples used in the empirical analysis for N and T.

4.3 Estimating common factors and loadings

Once $\widehat{\theta}$ and $\widehat{P}_j$ (for $j = \mathcal{D}, \mathcal{S}$) have been obtained, estimation of factors and loadings in each regime follows by standard arguments, e.g. applying Principal Components (see e.g. Bai, 2003), as also mentioned in Section 4.1. In particular, let

$$\widehat{\mathbf{\Sigma}}_{j,\widetilde{\mathbf{R}}}\left(\widehat{\theta}\right) = \frac{1}{NT}\sum_{t=1}^{T} d_{j,t}\left(\widehat{\theta}\right)\widetilde{\mathbf{R}}_t\left(\widehat{\theta}\right)\widetilde{\mathbf{R}}_t\left(\widehat{\theta}\right)',$$

for $j = \mathcal{D}, \mathcal{S}$, be an estimator of $\widehat{\mathbf{\Sigma}}_{j,\widetilde{\mathbf{R}}}(\theta)$ defined in Eq. (12). The estimator $\widehat{\mathbf{B}}_j^{\widehat{P}_j}\left(\widehat{\theta}\right)$ for $\mathbf{B}_j^{P_j}$ is given by $\sqrt{N}$ times the $N \times \widehat{P}_j$ matrix of eigenvectors of $\widehat{\mathbf{\Sigma}}_{j,\widetilde{\mathbf{R}}}\left(\widehat{\theta}\right)$ corresponding to its largest $\widehat{P}_j$ eigenvalues, under the orthonormalization restriction $N^{-1}\left(\widehat{\mathbf{B}}_j^{\widehat{P}_j}\left(\widehat{\theta}\right)'\widehat{\mathbf{B}}_j^{\widehat{P}_j}\left(\widehat{\theta}\right)\right) = \mathbf{I}_{\widehat{P}_j}$, for $j = \mathcal{D}, \mathcal{S}$. We then estimate $\{\mathbf{u}_{j,t}\}_{t=1}^{T}$ by OLS as

$$\widehat{\mathbf{u}}_{j,t}^{\widehat{P}_j}\left(\widehat{\mathbf{B}}_j^{\widehat{P}_j}\left(\widehat{\theta}\right), \widehat{\theta}\right) = N^{-1}\left(d_{\mathcal{D},t}\left(\widehat{\theta}\right)\widehat{\mathbf{B}}_j^{\widehat{P}_j}\left(\widehat{\theta}\right)\right)'\widetilde{\mathbf{R}}_t,$$

for $j = \mathcal{D}, \mathcal{S}$.

5 Three-pass estimation of risk premia

We now show how to estimate conditional risk premia in good and bad states of the world for the observable factors $\mathbf{g}_t$ in the spirit of Giglio and Xiu (2021). In particular, we propose the following three-pass procedure:

First-pass: Based on the results in Section 4, estimate θ^0, P_j^0, $\{\mathbf{u}_{j,t}\}_{t=1}^T$, and $\{\beta_{j,i}\}_{i=1}^N$ for $j = \mathcal{D}, \mathcal{S}$: recall that the notation for the corresponding estimators is $\widehat{\theta}$, $\widehat{P}_j$, $\left\{\widehat{\mathbf{u}}_{j,t}^{\widehat{P}_j}\right\}_{t=1}^T$, and $\left\{\widehat{\beta}_{j,i}^{\widehat{P}_j}\right\}_{i=1}^N$ for $j = \mathcal{D}, \mathcal{S}$, respectively.

Second-pass: Estimate the risk premia of $\mathbf{f}_{j,t}$ by running regime-specific cross-sectional regressions (Eq. (21) below).

Third-pass: Estimate $\mathbf{\Lambda}_j$ within each regime by running a regression of $\mathbf{g}_t$ on $\widehat{\mathbf{u}}_{j,t}^{\widehat{P}_j}$ based on (8) (Eq. (22) below) and rotate the risk premia of $\mathbf{f}_{j,t}$ (Eq. (23) below).

We have discussed the first pass in Section 4.2, where consistency of $\widehat{\theta}$ and $\widehat{P}_j$ are derived. We now turn to the second and third one. In this section, we report only the main results on the risk premia; further results are in Section A.4 of the Internet Appendix.

5.1 Regime-specific cross-sectional regression

Given the $N \times \widehat{P}_j$ matrices of beta estimates $\widehat{\mathbf{B}}_j^{\widehat{P}_j} = \left(\widehat{\beta}_{j,1}^{\widehat{P}_j}, \ldots, \widehat{\beta}_{j,N}^{\widehat{P}_j}\right)'$, define the $N \times \left(1 + \widehat{P}_j\right)$ matrices $\widehat{\mathbf{X}}_j = \left(\iota_N, \widehat{\mathbf{B}}_j^{\widehat{P}_j}\right)$, for $j = \mathcal{D}, \mathcal{S}$. Let $\mathbf{\Gamma}_j = \left(\gamma_{j,0}, \gamma_{j,1}'\right)'$, for $j = \mathcal{D}, \mathcal{S}$, and $\mathbf{R}_t = (R_{1,t}, \ldots, R_{N,t})'$, for $t = 1, \ldots, T$.

The regime specific cross-sectional regression of $\mathbf{R}_t$ on $\widehat{\mathbf{X}}_j$ yields

$$\widehat{\mathbf{\Gamma}}_{j,t} = \left(\widehat{\mathbf{X}}_j'\widehat{\mathbf{X}}_j\right)^{-1} \left(\widehat{\mathbf{X}}_j'\mathbf{R}_t\right) \hat{d}_{j,t}, \tag{21}$$

for $j = \mathcal{D}, \mathcal{S}$ and $1 \leq t \leq T$. Let now $\widehat{T}_j = T\widehat{\pi}_j$ for $j = \mathcal{D}, \mathcal{S}$. Then, $\mathbf{\Gamma}_j$ is estimated as

$$\widehat{\mathbf{\Gamma}}_j = \begin{pmatrix} \widehat{\gamma}_{j,0} \\ \\ \widehat{\gamma}_{j,1} \end{pmatrix} = \frac{1}{\widehat{T}_j} \sum_{t=1}^{T} \widehat{\mathbf{\Gamma}}_{j,t} = \left(\widehat{\mathbf{X}}_j' \widehat{\mathbf{X}}_j\right)^{-1} \left[\widehat{\mathbf{X}}_j' \overline{\mathbf{R}}_j \left(\widehat{\theta}\right)\right],$$

for $j = \mathcal{D}, \mathcal{S}$, where $\overline{\mathbf{R}}_j \left(\widehat{\theta}\right) = \widehat{T}_j^{-1} \sum_{t=1}^{T} \widehat{d}_{j,t} \mathbf{R}_t$.

5.2 Estimation of regime-specific risk premia

Let $\widetilde{\mathbf{g}}_{j,t} = \mathbf{g}_t \widehat{d}_{j,t} \left(\widehat{\theta}\right) - \overline{\mathbf{g}}_j$, with $\overline{\mathbf{g}}_j = \widehat{T}_j^{-1} \sum_{t=1}^{T} \mathbf{g}_t \widehat{d}_{j,t} \left(\widehat{\theta}\right)$, for $j = \mathcal{D}, \mathcal{S}$.

The estimator for $\mathbf{\Lambda}_j$ within each regime is

$$\widehat{\mathbf{\Lambda}}_j = \left[\sum_{t=1}^{T} \widehat{d}_{j,t} \left(\widehat{\theta}\right) \widetilde{\mathbf{g}}_{j,t} \widehat{\mathbf{u}}_{j,t}'\right] \left[\sum_{t=1}^{T} \widehat{d}_{j,t} \left(\widehat{\theta}\right) \widehat{\mathbf{u}}_{j,t} \widehat{\mathbf{u}}_{j,t}'\right]^{-1}, \tag{22}$$

for $j = \mathcal{D}, \mathcal{S}$. Similarly, the estimator for $\gamma_{j,\mathbf{g}}$ is

$$\widehat{\gamma}_{j,\mathbf{g}} = \widehat{\mathbf{\Lambda}}_j \widehat{\gamma}_{j,1}, \tag{23}$$

for $j = \mathcal{D}, \mathcal{S}$.

5.3 Asymptotics: consistency and limiting distribution of $\widehat{\gamma}_{j,\mathbf{g}}$

We report here the asymptotics for $\widehat{\gamma}_{j,\mathbf{g}}$; results for $\widehat{\mathbf{\Gamma}}_j$ and $\widehat{\mathbf{\Lambda}}_j$ are in Section A.4.3 of the Internet Appendix. We begin by showing the consistency of our estimators.

Theorem 2. *We assume that Assumptions 1-8 in Section A.1 of the Internet Appendix are satisfied. Then, as* $\min(N, T_j) \to \infty$, *it holds that*

$$\widehat{\gamma}_{j,\mathbf{g}} - \gamma_{j,\mathbf{g}}^0 = o_P(1) + o_{P^*}(1),$$

for $j = \mathcal{D}, \mathcal{S}$, *for almost all realizations of* $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$.

Theorem 2 states that $\widehat{\gamma}_{j,\mathbf{g}}$ is consistent; no restrictions are required on the relative rate of divergence of N and T as they pass to infinity.

Next, we present the limiting distribution of the estimation error.

Theorem 3. *We assume that Assumptions 1-9 in Section A.1 are satisfied, and that* $\min(N, T_j) \to \infty$*, with*

$$\pi_j \in (0,1), \tag{24}$$

$$\frac{T^{1/2}}{N} \to 0. \tag{25}$$

Then, it holds that

$$\left(\frac{1}{T}\Sigma_{\gamma,j} + \frac{1}{N}\Sigma_{\alpha\gamma,j}\right)^{-1/2}\left(\widehat{\gamma}_{j,\mathbf{g}} - \gamma^0_{j,\mathbf{g}}\right) \stackrel{D^*}{\to} N(0, I_K), \tag{26}$$

for $j = \mathcal{D}, \mathcal{S}$*, for almost all realizations of* $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$*, where* $\Sigma_{\gamma,j}$ *and* $\Sigma_{\alpha\gamma,j}$ *are defined in Section A.4 of the Internet Appendix.*[7]

6 Empirical analysis

We now apply the conditional factor model and the methods described above to two cross-sections of asset returns.

6.1 Data

We use two sets of test assets: a cross-section of equity portfolio returns (Section 6.1.1), and a set of returns for multiple asset classes (Section 6.1.2).

[7] A comment on (25) may be in order. By Theorem 2, all estimators are consistent, irrespective of the relative rate of divergence between N and T as they pass to infinity. Theorem 3, conversely, requires a restriction, namely that $T = o(N^2)$. However, it can be envisaged that most datasets (unless T is very "short", which is typically not the case in empirical work) should satisfy this requirement.

6.1.1 Equity panel

We use returns (in excess of the risk-free rate) from the following set of $N = 130$ US equity portfolios: 25 portfolios sorted by size and book-to-market; 25 portfolios sorted by size and momentum; 10 size-sorted portfolios; 10 book-to-market portfolios; 10 momentum portfolios; 25 portfolios sorted by size and operating profitability; 25 portfolios sorted by size and investment. The risk-free rate is the one-month US Treasury bill rate. The market return used to determine the disappointment event (as explained in Section 2.3) is the value-weighted average log-return computed with respect to all stocks available from CRSP.

The data are obtained from Kenneth French's website.[8] We consider the sample period from July 1963 to December 2018, a total of $T = 666$ monthly observations. Farago and Tédongap (2018) analyze each of these sets of portfolios individually.[9] Equipped with our conditional model, we can consider the entire cross-section of returns, which is aligned with Lewellen, Nagel, and Shanken (2010), who advocate the use of sizeable cross-sections of asset returns.

6.1.2 Data for multiple asset classes

We also consider a cross section of portfolio excess returns that includes equity, government bond, corporate bond and currency returns, with $N = 57$ portfolios in total. This allows us to analyze the effect of downside risk on multiple asset classes.[10]

The data for these test assets are obtained from He, Kelly, and Manela (2017).[11] In particular, we employ 25 equity portfolios sorted by size and book-to-market ratio; 10 maturity-sorted government bond portfolios from the CRSP "Fama Bond Portfolios" file with maturities in six month intervals up to five years; 10 corporate bond portfolios sorted on yield spreads from Nozawa (2017); 6 currency portfolios sorted on interest rate differential (carry) from Lettau, Maggiori, and Weber (2014); 6 currency portfolios sorted on momentum from Menkhoff, Sarno, Schmeling, and Schrimpf (2012b). As for the equity panel, the market return to determine the disappointment

[8]See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[9]See Section 3.2 and Table 1 in Farago and Tédongap (2018).

[10]Cochrane (2011) suggests investigating the factor structure across multiple asset classes to study the corresponding underlying discount factors.

[11]The data are kindly made available on Asaf Manela's website at http://apps.olin.wustl.edu/faculty/manela/data.html.

event as detailed in Section 2.3 is the value-weighted average log-return computed with respect to all stocks available from CRSP. The data sample runs from October 1983 to January 2010, with $T = 316$ time series observations.

6.2 Estimation and inference

We begin by reporting results on estimation and inference on the threshold level θ (Section 6.2.1) and on the number of factors in each regime (Section 6.2.2).

6.2.1 Estimation and inference on θ

Results on estimation are in Panel A of Table 1. In the case of the equity sample, the estimated disappointment threshold $\widehat{\theta}$ is equal to -0.061: this generates $\widehat{T}_{\mathcal{D}} = 45$ downside observations, with a relative frequency of approximately 6.8%. Notice that $\widehat{\theta}$ is lower than the fixed threshold $\overline{\theta} = -0.03$ used, e.g., by Farago and Tédongap (2018), which would generate 105 downside periods with a relative frequency approximately equal to 15.8% in these data. An analogous result is obtained in the case of the sample of multi-asset portfolios, where the estimated disappointment threshold $\widehat{\theta} = -0.062$ and gives $\widehat{T}_{\mathcal{D}} = 21$ downside observations, with frequency approximately equal to 6.6%. In this case, the fixed threshold $\overline{\theta} = -0.03$ would generate 49 downside observations with a relative frequency approximately equal to 15.5%. Therefore, for both sets of test assets, the estimated downside state is less frequent but more severe than implied by the fixed threshold typically assumed in the literature on downside risk.

To enhance our understanding of the downside regime, we evaluate the correlation between the downside indicators, as resulting from using the estimated threshold $\widehat{\theta}$ and the value $\overline{\theta} = -0.03$, and various measures of economic and financial conditions that may be related to the downside regime. We consider the following set of variables: the NBER US recession indicator (REC); the year-on-year log-change in industrial production (ΔIP); the CBOE volatility index based on S&P 500 index options (VIX); the CBOE volatility index based on S&P 100 index options (VXO); the economic policy uncertainty index of Baker, Bloom, and Davis (2016) (EPU); the Equity Market Volatility tracker of Baker, Bloom, and Davis (2016) (EMV); the Policy-Related Equity Market

Volatility tracker of Baker, Bloom, Davis, and Kost (2019) (*PREMV*); the Geopolitical Risk index of Caldara and Iacoviello (2022) (*GPR*); the TED spread (*TED*); the disaster probability for the US of Barro and Liao (2021) (*SPX*); the spread for the T-bond between 10 and 2 year maturities (*T10Y2Y*). Broadly speaking, the correlations point towards a connection between the downside indicators and the measures of economic and financial conditions which we consider, and they appear slightly stronger when using the estimate $\widehat{\theta}$ compared to when using the fixed value $\overline{\theta}$. This finding is consistent with a vast literature on asset pricing, which shows that volatility risk is a priced risk factor (see, *inter alia*, Ang et al., 2006, Ang et al., 2009, and Menkhoff et al., 2012a). Note, however, that the bilateral correlations are far from perfect and they often revolve around 0.5; this indicates that the determinants of the threshold go beyond volatility. For example, in the case of the equity panel, the correlation with the NBER recession indicator is weakly significant and positive.[12]

We then run inference on the threshold parameter, and the results are shown in Panel B of Table 1. We first conduct a test, whose details are spelt out in Section A.2, for the null hypothesis $H_0 : \theta = \overline{\theta}$: this allows us to formally assess whether the choice of the fixed threshold $\overline{\theta} = -0.03$ is supported by the data. As can be seen, the null is strongly rejected, which unequivocally indicates that the estimated threshold $\widehat{\theta}$ is significantly different from $\overline{\theta}$: this provides evidence in favor of our approach, which allows to estimate the threshold value θ for the downside regime. In addition to testing for the significance of the difference between $\widehat{\theta}$ and $\overline{\theta}$, we use the same testing approach to test for the null that the estimated thresholds are the same across the two datasets of equity and multi-asset portfolios (last row of Table 1): these results show that the estimated threshold is the same across the two datasets we consider, as the null hypothesis that the threshold is the same in the two samples cannot be rejected at conventional levels of significance.

6.2.2 Estimating the number of common factors

We estimate the number of common factors for three cases: the linear unconditional model described in Section 2.1 as applied to the whole available sample (this case is denoted henceforth

[12] An interesting question is what drives the value of θ, and the model could allow for θ to be endogenously determined by some state variables and to vary over time. We leave this further generalization for future research.

without using any subscripts), the satisfactory regime $j = \mathcal{S}$, and the downside counterpart $j = \mathcal{D}$. In light of the results in Section 6.2.1, we only report results using the estimated threshold value $\widehat{\theta}$. In order to assess the robustness of the estimates to the choice of the tuning parameters $\varrho_{\mathcal{D}}$ and $\varrho_{\mathcal{S}}$, we try several values of ε in Eq. (19) in the range between 0.02 and 0.12.[13] By construction, implementing the sequential tests with lower values of ε is bound to result in the detection of more common factors (or, equivalently, more statistically significant eigenvalues) than with higher values of ε. On the one hand, this provides us with a range of common factors to employ in the estimation of the model; on the other hand, we note that the estimated number of common factors stabilizes after a certain level, indicating the robustness of our results.

Alongside the estimated number of common factors, we evaluate the proportion of the total variance explained by each factor. Letting $g_j^{(i)}$ denote the i-th largest eigenvalue for each of the two regimes $j = \mathcal{D}, \mathcal{S}$ (and, similarly, using $g^{(i)}$ for the whole sample), the proportion of the total variance associated with the i-th common factor is measured as

$$\nu_j^{(i)} = \frac{g_j^{(i)}}{\sum_{i=1}^{N} g_j^{(i)}}, \tag{27}$$

with $\nu^{(i)}$ defined similarly for the whole sample.

Results are in Table 2.[14] As mentioned above, we have tried different values of ε in the construction of ϱ_j according to Eq. (19); as predicted, when ε is very close to 0, the number of estimated common factors is higher, and it decreases as ε increases. Importantly, after a certain value of ε, the estimated number of common factors stabilizes, indicating that the estimated values of $P_{\mathcal{D}}^0$ and $P_{\mathcal{S}}^0$ have become robust to the choice of ε. Based on the estimates $\widehat{P}_{\mathcal{D}}$ and $\widehat{P}_{\mathcal{S}}$, the most striking finding in Table 2 is the discrepancy between the number of estimated common factors in the two regimes. We note two main points. First, for each of the two cross-sections of returns examined, the number of common factors in the downside state, $\widehat{P}_{\mathcal{D}}$ is found to be in the range

[13] Recall that, according to Eq. (19), $\varepsilon > 0$ is required. We have tried positive values of ε smaller than 0.02, and results are the same as obtained with $\varepsilon = 0.02$.

[14] We have used the algorithm described in Section 4.2. We have carried out the individual tests with $M = N$ in Step 1; similarly to our findings in the simulations (reported in the Internet Appendix), we assessed the robustness of the results by using also $M = \lfloor N/2 \rfloor$ and $M = 2N$, but recorded virtually no changes. In the determination of the number of factors, we have carried out the individual tests using a nominal level $\alpha = \frac{0.05}{T}$.

between one and three, depending on the tuning parameter. The number of common factors in the satisfactory regime, $\widehat{P}_{\mathcal{S}}$ is always greater than in the downside state, being equal to six and four for the equity and multi-asset cross-sections respectively. This means that the factor structure of asset returns becomes lower dimensional in the downside state, i.e. the data tend to be driven by few(er) common factors during bad times, suggesting that dependence among asset returns increases in a downturn. In turn, this result is intuitively clear and consistent with the anecdotal evidence that factor diversification tends to decline when needed most (see e.g. Page and Panariello, 2018 and the references therein).[15] Secondly, applying the sequential testing procedure to the unconditional, linear model would lead to detecting a number of factors equal to the highest number of factors (detected for the satisfactory regime), finding six and four respectively in the two datasets. Indeed, the number of factors and the factors themselves are assumed to be identical for all observations in the linear model: this leads to overestimation of the number of factors required to fit the data in the downside state, where one would expect the linear model to perform particularly poorly.

These findings clearly indicate that it is important to allow for a different number of factors in each regime. From a statistical point of view, it is natural that removing a restriction leads to better inference - in this case, the number of common factors is neither understated (which would correspond to an omitted variable problem) nor overstated (which would be tantamount to including irrelevant variables). From an empirical viewpoint, the results clearly imply that restricting the number of common factors to be the same across regimes is not an adequate choice.

6.3 Model fit

In this section, we compare the performance of our latent factor model with downside risk against a linear specification, in a similar vein as Lettau, Maggiori, and Weber (2014). We report three measures of goodness of fit: the R^2; the root mean squared pricing error ($\overline{RMS_\alpha}$); and the average pricing error ($\overline{AVG_\alpha}$). For the $\overline{AVG_\alpha}$, we also report a p-value for the null hypothesis that $\overline{AVG_\alpha}$ is equal to zero, obtained using the wild bootstrap described in Appendix E, adapted from Giglio et al. (2021). We report results from estimating both the unconditional and the conditional

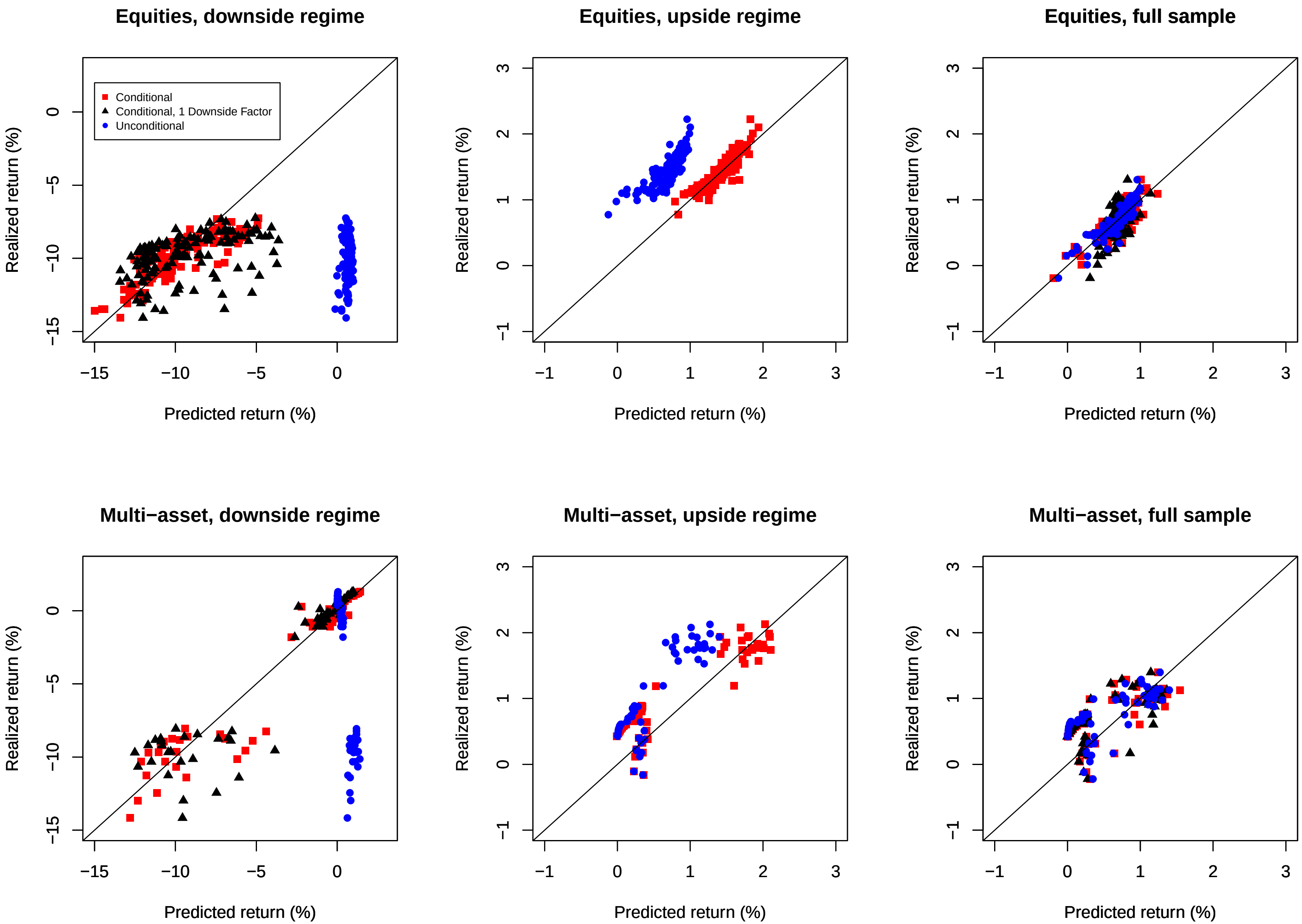
[15] This result is also consistent with the literature that shows that, in crisis times, the macroeconomy is driven by a smaller number of common factors (we refer, *inter alia*, to Stock and Watson, 2009, Cheng et al., 2016 and Li et al., 2019).

model with factors ranging from one to six. Recall that the tests in the previous section indicate that the number of common factors for the unconditional model is six for the equity sample and four for the multi-asset sample. The same number of factors is suggested by the tests for the conditional model in Regime $\mathcal{S}$, whereas in Regime $\mathcal{D}$ the tests suggest a range between one and three factors. While we present results for all possible number of factors from one to six also in regime $\mathcal{D}$, we expect that one factor is most likely insufficient, given the existing evidence in the asset pricing literature that a one-factor model cannot price the cross-section of asset returns. Hence we favor the use of two or three factors in regime $\mathcal{D}$, and comment primarily on those models when discussing the estimation results below.

The results in Panel A of Table 3 show that the conditional model performs well, with a very high R^2 both in the $\mathcal{S}$ regime and in the $\mathcal{D}$ regime. The results also suggest that the performance of the linear, unconditional model over the full sample is satisfactory with an R^2 of 0.98 and 0.76 (when the number of common factors P is set to its estimate $\hat{P}$ of six and four) for the equity and multi-asset samples, respectively. However, we know from earlier evidence that the factor structure of the cross sections of asset returns analyzed here is very different in the two regimes, and that the number of latent factors that matter for pricing the cross-section of asset returns in the downside regime is much smaller (namely, between one and three) than the number of factors required in the $\mathcal{S}$ regime. Given that the data experience such strong regime dependence, it seems surprising that the unconditional, regime-independent pricing model performs as well as it does. To better understand the gain from using a conditional asset pricing model that allows for downside risk, it is therefore useful to check the performance of the unconditional model for observations in the $\mathcal{S}$ and $\mathcal{D}$ regimes separately. The results in Panel A of Table 3 show that the performance of the unconditional model is very poor in *both* the $\mathcal{D}$ and $\mathcal{S}$ regimes, with very low explanatory power. In fact, the R^2 is negative for observations in the $\mathcal{D}$ regime and below 0.3 in the $\mathcal{S}$ regime. This suggests that the seemingly good performance of the unconditional model over the full sample is illusory, in the sense that it comes from somehow combining large pricing errors from two regimes which generate small pricing errors on aggregate, over the full sample.

The latter point can be seen clearly in Figure 2, where (for both sets of test assets) we report scatter plots of realized versus model-implied returns, for both the unconditional model, and the conditional model under three different cases: the $\mathcal{D}$ regime, the $\mathcal{S}$ regime, and the full sample. This figure makes it apparent how the unconditional model (blue circles) performs very poorly in the downside regime. Note the large positive errors: the unconditional model predicts much higher returns than the realized ones, which tend to be very negative in this regime. Hence, the unconditional model is not able to capture such extreme, bad events. The conditional model performs much better (especially when allowing for three factors in the downside state rather than just one, as conjectured earlier), and there appears to be no systematic tendency to generate positive or negative errors in the downside regime (black triangles and red squares for the conditional models with one and three factors, respectively). Turning to Regime $\mathcal{S}$, the conditional model performs very well (red squares), with the majority of test asset returns lying close to, or on, the 45 degree line. In contrast, the returns implied by the unconditional model display a monotonic pattern (hence are highly positively correlated) with realized returns in Regime $\mathcal{S}$, but do not lie on the 45 degree line for any of the test asset returns. This means that, even in Regime $\mathcal{S}$, the unconditional model has a tendency to generate systematic (non-random) pricing errors; however, these pricing errors have the opposite sign than in the $\mathcal{D}$ regime, since the unconditional model predicts lower returns than realized in the $\mathcal{S}$ regime. Over the full sample, the averaging across the errors of these two regimes - large positive errors for a small number of observations in the $\mathcal{D}$ regimes and more moderate negative pricing errors for a large number of observations in the $\mathcal{S}$ regime - gives the appearance that the unconditional model performs well, but this is of course fallacious: the unconditional model performs poorly in *both* regimes.

Figure 2: Predicted and realized returns



Notes. The scatter plot shows realized average returns against predicted average returns in downside, upside and over the entire sample for the 130 equity portfolios and the 57 multi-asset portfolios described in Sections 6.1.1 and 6.1.2, respectively. Predicted returns in downside are obtained using one factor and the number of factors estimated using the test detailed in Section 4.2.

Panel B of Table 3 provides corroborative numerical evidence on the differences in pricing errors illustrated graphically in Figure 2. The $\overline{RMS_\alpha}$ for the unconditional model appears to be low (0.105 for the equity sample with six factors, and 0.370 for the multi-asset sample with four factors); yet the $\overline{RMS_\alpha}$ is extremely large in the $\mathcal{D}$ regime (over 10 and 7 for the two sets of test assets considered) and quite large also in the $\mathcal{S}$ regime (0.774 and 0.633 respectively). The conditional model produces a $\overline{RMS_\alpha}$ that is at least an order of magnitude smaller in both states of the world. Specifically, for the equity sample, the conditional model with three factors in the $\mathcal{D}$ regime and six factors in the $\mathcal{S}$ regime generates a $\overline{RMS_\alpha}$ of 1.250 in regime $\mathcal{D}$, and 0.105 in regime $\mathcal{S}$; for the multi-asset sample, the conditional model with three factors in the $\mathcal{D}$ regime and four factors in the $\mathcal{S}$ regime generates a $\overline{RMS_\alpha}$ of 1.388 in the $\mathcal{D}$ regime, and 0.354 in the $\mathcal{S}$ regime.

Panel C of Table 3 sheds further light on the sign of the pricing errors by reporting the average α across the test assets, $\overline{AVG_\alpha}$ alongside the bootstrap p-value for the test of the null that $\overline{AVG_\alpha}$ is equal to zero (in parenthesis).[16] Starting from the equity sample, again the unconditional model with six factors displays a very low $\overline{AVG_\alpha}$, which is not statistically significantly different from zero over the full sample. However, it produces very large and statistically significant pricing errors in both Regimes $\mathcal{D}$ and $\mathcal{S}$ for any number of factors considered. In the multi-asset sample, the unconditional model also produces low $\overline{AVG_\alpha}$, although four factors are not sufficient to deliver an $\overline{AVG_\alpha}$ that is statistically insignificantly different from zero (at least five factors are required instead). Again, this seemingly good pricing performance is illusory in that the unconditional model produces very large and statistically significant average pricing errors in both Regimes $\mathcal{D}$ and $\mathcal{S}$ for any number of factors considered. The sign of the $\overline{AVG_\alpha}$ is clearly negative in Regime $\mathcal{D}$ and positive in Regime $\mathcal{S}$, as argued earlier on the basis of the scatter plots in Figure 2. In contrast, the conditional model delivers, for both cross sections, average pricing errors in the two regimes which, although they have the same sign as the unconditional model, are at least one order of magnitude lower and generally statistically insignificantly different from zero. We note, however, that five factors appear to be needed in Regime $\mathcal{S}$ (as opposed to the four factors

[16]It is worth noting that zero average pricing errors across portfolios is a requirement in our estimation as it is (part of) our Assumption 7 – which is also required in the three-pass method of Giglio and Xiu (2021); see their Assumption I.1. In this sense the tests presented below are also a test of the validity of Assumption 7.

suggested by our tests) in order for the average pricing error to become statistically insignificantly different from zero in the multi-asset sample.[17]

The analysis above clarifies the gain from using a conditional latent factor model of asset pricing that allows for different regimes, and hence different sets of factors in good and bad times. The pricing model is very different in these two regimes: this captures the stark difference in the time-varying factor structure of asset returns, which is lower dimensional in bad times, and requires more factors in good times. A model that does not allow for this feature, which is prominent across both datasets examined here, can spuriously generate a strong correlation between model-implied and realized returns over the full sample, but in fact it performs poorly in both good and bad times. At an applied level, the conditional model is consistent with the notion that asset returns become more highly correlated in bad times, with their factor structure reducing in dimensionality to a smaller set of factors, and this feature is key for the model to generate expected returns that capture downside risk.

6.4 Risk premia

Tables 4 and 5 report the results from the estimation of the risk premia for tradable and non-tradable factors, respectively, obtained using the three-pass procedure detailed in Section 5. We are clearly more interested in results relating to non-tradable factors, where the three-pass procedure is needed to obtain robust risk premia estimates (taking into account omitted variables and measurement error), while for tradable factors a model-free estimate of the risk premium is readily available (i.e., the average of the factor itself). However, it is still interesting to examine the risk premia estimates for tradable factors to assess how the risk premium for standard risk factors developed in the equity asset pricing literature vary across good and bad states in the two different cross-sections of test assets studied here.

Tradable factors. Starting from the results on tradable factors reported in Table 4, we consider the five Fama and French (2015) factors: the return on the market in excess of the risk-free rate ($RmRf$), size (SMB), value (HML), operating profitability (RMW), and investment

[17] This suggests that the tests for the number of factors are better seen as an indication of the minimum number of factors required, and that more factors may be allowed for in practical applications if they are required to fully span the information in the test asset returns.

(*CMA*), obtained from Kenneth French website. The first row associated with each factor in Table 4 reports the observed risk premium for the factor, as captured by the mean (both unconditional and conditional on the state of the world). For both equity and multi-asset portfolios, we perform the analysis using the latent factors from both conditional and unconditional specifications with varying number of factors from one to six; the unconditional specification is essentially the three-pass procedure of Giglio and Xiu (2021), for which results are reported in the column denoted "Unconditional". The columns labelled "Conditional" in Table 4 report estimated risk premia from the conditional model for each of the two regimes, while the final column reports the implied average risk premium from the conditional model, calculated as the weighted average of the conditional, regime-specific estimates with weights equal to the number of observations in each regime, with equal number of factors varying from one to six. We will focus primarily on the risk premia estimates obtained with the largest number of (six) latent factors, which are more likely to span the information in the test assets' space.[18]

Starting from equity portfolios, in the case of the unconditional model all risk premia are correctly signed and of reasonable magnitude. They are statistically significantly different from zero at the five percent level for *RmRf* and *HML*, while for *RMW* and *CMA* they are statistically insignificant, and the risk premium for *SMB* is only statistically significant at the ten percent level. These results are fairly comparable to the ones reported by Giglio and Xiu (2021) on a different equity cross section. The unconditional model, however, hides highly asymmetric dynamics between the regimes: *RmRf* and *SMB* generate procyclical excess returns that are negative and statistically significant in bad times, while being positive and significant in good times. However, *HML*, *RMW* and *CMA* appear to generate positive excess returns in the downside state that are larger than the excess returns in the good state of the world; this feature, which is also present in the conditional mean of the factors (i.e. the model-free estimate of these risk premia), is well captured by the conditional model. The average implied risk premium from both the conditional and unconditional model is precisely estimated (statistically significant at least at

[18]Giglio and Xiu (2021) suggest that, while it is inappropriate to use less latent factors than indicated by testing procedures in the third pass (as it will lead to inconsistent estimates of risk premia), it is desirable to use more factors than indicated by the tests since risk premia estimates are expected to stabilize as one increases the number of latent factors.

the five percent level) only for *RmRf* and *HML*.

Moving to multi-asset portfolios, we observe less statistical significance in the risk premia estimates. Specifically, we find that the only clearly statistically significant (at the one percent significance level) risk premium estimate is the one associated with the market excess return *RmRf*. This is the case for both unconditional and conditional models. However, again, the conditional model is able to capture well the strong conditional variation in the risk premium of the stock market excess return. The other four factors display risk premia with strong conditional variation, but none of them is clearly statistically significant in both regimes, either in terms of the raw conditional mean, or on the basis of the estimates produced by the unconditional and conditional models. This suggests that these risk factors, which are often priced in the cross-section of equity returns, are not priced in the multi-asset cross section, indicating some degree of market segmentation between equity and other asset markets.

Non-tradable factors. Turning to the non-tradable risk factors, in Table 5 we report risk premia estimates for seven non-tradable factors: expected industrial production growth (IP), measured as the AR(1) innovation in the year-on-year log-change in industrial production; the liquidity factor (LIQ) of Pástor and Stambaugh (2003); the change in the VXO index, capturing expected volatility in the S&P100 embedded in option prices (ΔVXO); the intermediary capital factor HKM of He, Kelly, and Manela (2017); changes in the economic uncertainty index of Baker, Bloom, and Davis (2016) (ΔEPU); changes in the geopolitical risk index (ΔGPR) of Caldara and Iacoviello (2022); and changes in equity market volatility (ΔEMV) of Baker, Bloom, and Davis (2016). The setup of Table 5 is the same as Table 4 except that the first row associated with each factor in Table 5 reports the expected sign of the risk premium for the factor, rather than the factor mean (since the latter has no clear meaning in terms of risk premia for non-tradable factors). The expected sign for procyclical factors (IP, LIQ and HKM) is positive unconditionally and in Regime S, and negative in Regime D. The expected sign for countercyclical factors (ΔVXO, ΔEPU and ΔGPR) is negative unconditionally and in Regime S, and positive in Regime D.

Starting from the equity portfolios, the unconditional model gives estimates of the risk premia that are correctly signed for each non-tradable factor using the three-pass procedure of Giglio and Xiu (2021). However, the risk premia for IP and ΔGPR are not statistically significantly different from zero, and ΔEPU is only marginally significant at the ten percent level. Just like for tradable factors, the unconditional risk premium masks highly asymmetric dynamics between the regimes: procyclical (countercyclical) factors generate excess returns that are negative (positive) in bad (good) times. The conditional model is able to capture these conditional risk premia adequately. The estimation results are particularly clear-cut for ΔVXO and HKM, which are priced both conditionally and unconditionally. For LIQ and ΔEMV, the excess returns are clearly strongly statistically significant in good times, while in bad times they are less precisely estimated, so that the implied average risk premium from the conditional model is not significant for these two factors. This result raises the possibility that LIQ and ΔEMV are time series factors that matter for conditional pricing by not for pricing the cross-section of average returns.

Moving to multi-asset portfolios, just like for tradable factors, we observe less statistical significance in the risk premia estimates. Specifically, we find that the only clearly statistically significant (at the one percent significance level) risk premium estimates are the ones associated with ΔVXO and HKM. This result is consistent using both the unconditional and the conditional asset pricing model. However, the conditional model is also able to capture well the strong conditional variation in their risk premia. The other candidate risk factors considered display risk premia with strong conditional variation, but none of them is priced in terms of the implied average risk premium.

Overall, the key piece of evidence uncovered in this section is that ΔVXO and HKM are priced factors in both equity and multi-asset cross sections of returns, both conditionally and unconditionally. There is strong conditional variation in excess returns for all factors considered which is hidden when employing a standard unconditional, linear asset pricing model both for ΔVXO and HKM and for other non-tradable factors. However, differences in terms of the performance of pricing factors and their point estimates remain important across equity and multi-asset cross-sections, potentially indicating some degree of market segmentation across financial markets.

7 Conclusions

In this paper, we propose a methodology to model and price the cross section of asset returns in the presence of common factors and downside risk. Both features have been shown to offer superior explanatory power in several empirical papers in asset pricing (e.g. Lettau et al., 2014). We extend the framework used in the extant literature in three ways: we allow for downside risk via a threshold specification which allows for the estimation of the (usually set *a priori*) threshold level; we consider different factor structures in the two regimes, allowing also for different numbers of factors; we adapt the methodology developed by Giglio and Xiu (2021) to recover the observable factors' risk premia from the estimated latent ones in both regimes.

The proposed model is illustrated through two empirical applications to a large cross-section of equity returns and a smaller, but more diverse, cross-section of returns from different asset classes. The results show that: estimating, rather than setting *a priori*, the threshold level determines substantial differences in the identification of good and bad states of the world relative to using an exogenously fixed level; estimating the number of common factors separately yields very different estimates across regimes, with the downside risk state having a small number of common factors (at most three in our data), whilst returns in the good state of the world are driven by more common factors; estimating the risk premia of popular observable factors yields estimates consistent with strong conditional variation of risk premia in the two regimes.

The empirical analysis shows clearly the limitations of unconditional asset pricing models that assume a constant relationship between risk and return, while illustrating that the proposed conditional model is capable to adequately characterize the key properties of different cross-sections of asset returns. Yet, we merely view our model as a tentative adequate characterization of asset returns data which is superior to the unconditional model in a number of respects, but it is only one of many alternative nonlinear specifications. Future work could focus, for example, on endogeneizing the threshold that identifies the downside state as a function of state variables, on allowing explicitly for weak pricing factors, and on out-of-sample testing. We leave these developments to future research.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

Ang, A., J. Chen, and Y. Xing (2006). Downside risk. *Review of Financial Studies 19*, 1191–1239.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*, 259–299.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics 91*, 1–23.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, Z. and J. Yao (2008). Central limit theorems for eigenvalues in a spiked population model. In *Annales de l'IHP Probabilités et Statistiques*, Volume 44, pp. 447–474.

Bai, Z. and J. Yao (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis 106*, 167–177.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*, 1593–1636.

Baker, S. R., N. Bloom, S. J. Davis, and K. J. Kost (2019). Policy news and stock market volatility. NBER Working Paper 25720.

Barigozzi, M. and L. Trapani (2022). Testing for common trends in nonstationary large datasets. *Journal of Business & Economic Statistics 40*(3), 1107–1122.

Barro, R. J. and G. Y. Liao (2021). Rare disaster probability and options pricing. *Journal of Financial Economics 139*, 750–769.

Bates, B. J., M. Plagborg-Møller, J. H. Stock, and M. W. Watson (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics 177*, 289–304.

Boguth, O., M. Carlson, A. Fisher, and M. Simutin (2011). Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics 102*(2), 363–389.

Bollerslev, T. (2022). Realized semi (co) variation: Signs that all volatilities are not created equal. *Journal of Financial Econometrics 20*(2), 219–252.

Bollerslev, T., S. Z. Li, and V. Todorov (2016). Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns. *Journal of Financial Economics 120*, 464–490.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2022). Realized semibetas: Disentangling "good" and "bad" downside risks. *Journal of Financial Economics 144*(1), 227–246.

Caldara, D. and M. Iacoviello (2022). Measuring geopolitical risk. *American Economic Review 112*(4), 1194–1225.

Cappiello, L., R. F. Engle, and K. Sheppard (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics 4*, 537–572.

Chan, K.-S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics 21*, 520–533.

Cheng, X., Z. Liao, and F. Schorfheide (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies 83*, 1511–1543.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*, 1047–1108.

Colacito, R., S. J. Riddiough, and L. Sarno (2020). Business cycles and currency returns. *Journal of Financial Economics 137*(3), 659–678.

Delikouras, S. (2017). Where's the kink? Disappointment events in consumption growth and equilibrium asset prices. *Review of Financial Studies 30*, 2851–2889.

Delikouras, S. and A. Kostakis (2019). A single-factor consumption-based asset pricing model. *Journal of Financial and Quantitative Analysis 54*, 789–827.

Fama, E. F. and K. D. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*, 1–22.

Farago, A. and R. Tédongap (2018). Downside risks and the cross-section of asset returns. *Journal of Financial Economics 129*, 69–86.

Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of political economy 99*(2), 385–415.

Giglio, S., Y. Liao, and D. Xiu (2021). Thousands of alpha tests. *The Review of Financial Studies 34*(7), 3456–3496.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*, 1947–1990.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*, 603–617.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*, 575–603.

Harvey, C. R. and Y. Liu (2020). A census of the factor zoo. Working paper, Duke University and Purdue University.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: new evidence from many asset classes. *Journal of Financial Economics 126*, 1–35.

Horváth, L. and L. Trapani (2019). Testing for randomness in a random coefficient autoregression model. *Journal of Econometrics 209*, 338–352.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics 40*, 694–726.

Lettau, M., M. Maggiori, and M. Weber (2014). Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics 114*, 197–225.

Lettau, M. and M. Pelger (2020). Estimating latent asset-pricing factors. *Journal of Econometrics 218*, 1–31.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*, 175–194.

Li, J., V. Todorov, G. Tauchen, and H. Lin (2019). Rank tests at jump events. *Journal of Business & Economic Statistics 37*(2), 312–321.

Lustig, H., N. Roussanov, and A. Verdelhan (2011). Common risk factors in currency markets. *The Review of Financial Studies 24*(11), 3731–3777.

Markovitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments.* John Wiley.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics 197*, 101–129.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Annals of Probability 2*, 620–628.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012a). Carry trades and global foreign exchange volatility. *Journal of Finance 67*, 681–718.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012b). Currency momentum strategies. *Journal of Financial Economics 106*, 660–684.

Merikoski, J. K. and R. Kumar (2004). Inequalities for spreads of matrix sums and products. *Applied Mathematics E-Notes 4*, 150–159.

Moricz, F. (1983). A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Mathematica Hungarica 41*, 337–346.

Nozawa, Y. (2017). What drives the cross-section of credit spreads?: A variance decomposition approach. *Journal of Finance 72*, 2045–2072.

Page, S. and R. A. Panariello (2018). When diversification fails. *Financial Analysts Journal 74*, 19–32.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 642–685.

Patton, A. J. and T. Ramadorai (2013). On the high-frequency dynamics of hedge fund risk exposures. *The Journal of Finance 68*(2), 597–635.

Peligrad, M. (1987). On the central limit theorem for ρ-mixing sequences of random variables. *Annals of Probability 15*, 1387–1394.

Peligrad, M., S. Utev, and W. Wu (2007). A maximal l_p-inequality for stationary sequences and its applications. *Proceedings of the American Mathematical Society 135*, 541–550.

Salzer, H. E., R. Zucker, and R. Capuano (1952). Table of the zeros and weight factors of the first twenty hermite polynomials. *Journal of Research of the National Bureau of Standards 48*, 111–116.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance 44*, 1115–1153.

Shao, Q.-M. (1995). Maximal inequalities for partial sums of ρ-mixing sequences. *Annals of Probability 23*, 948–965.

Stock, J. H. and M. W. Watson (2009). Forecasting in dynamic factor models subject to structural instability. In J. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford University Press.

Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *Journal of the American Statistical Association 113*, 1341–1349.

Uematsu, Y. and T. Yamagata (2021). Estimation of weak factor models. *Journal of Business and Economic Statistics*. Forthcoming.

Wang, W. and J. Fan (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Annals of Statistics 45*, 1342.

Table 1: Estimation and inference on θ

Panel A: Estimation				
	Equity		Multi-asset	
	fixed	estimated	fixed	estimated
θ	-0.030	-0.061	-0.030	-0.062
$T_{\mathcal{D}}$	105	45	49	21
$\pi_{\mathcal{D}}$	0.158	0.068	0.155	0.066
$Corr\,(\mathcal{D}, REC)$	0.166*	0.244*	0.197	0.338
$Corr\,(\mathcal{D}, \Delta IP)$	-0.054	-0.084	-0.032	-0.169
$Corr\,(\mathcal{D}, VIX)$	0.508***	0.538***	0.475***	0.541**
$Corr\,(\mathcal{D}, VXO)$	0.503***	0.545***	0.486***	0.572***
$Corr\,(\mathcal{D}, EPU)$	0.166*	0.293**	0.200	0.354
$Corr\,(\mathcal{D}, EMV)$	0.450***	0.525***	0.414***	0.494**
$Corr\,(\mathcal{D}, PREMV)$	0.426***	0.516***	0.399***	0.493**
$Corr\,(\mathcal{D}, GPR)$	-0.003	0.045	0.030	0.096
$Corr\,(\mathcal{D}, TED)$	0.163*	0.210	0.185	0.247
$Corr\,(\mathcal{D}, SPX)$	0.431***	0.493***	0.388***	0.483**
$Corr\,(\mathcal{D}, T10Y2Y)$	-0.059	-0.010	0.060	0.150

Panel B: Inference		
	Equity	Multi-asset
$H_0 : \theta^0 = -0.03$	0.000***	0.000***
$H_0 : \theta^0_E = \theta^0_{MA}$	1.000	

Notes. The table contains all the relevant inference on θ in the datasets considered. In Panel A, for each dataset we report the estimated value, $\widehat{\theta}$, and, as a term of comparison, the exogenously fixed value $\overline{\theta} = -0.030$, which is proposed in Farago and Tédongap (2018). For both values of θ, we also report the proportion of time periods in the downside regime $\mathcal{D}$, denoted as $\pi_{\mathcal{D}}$, and the corresponding number of time periods, denoted as $T_{\mathcal{D}}$. In Panel A we also report the correlation between the downside indicator functions implicitly defined by both $\widehat{\theta}$ and $\overline{\theta}$, and the following variables: *REC* is the NBER U.S. recession indicator; ΔIP is the log difference of industrial production; *VIX* is the CBOE volatility index based on S&P 500 index options; *VXO* is the CBOE volatility index based on S&P 100 index options; *EPU* is the economic policy uncertainty index of Baker, Bloom, and Davis (2016); *EMV* is the Equity Market Volatility tracker of Baker, Bloom, and Davis (2016); *PREMV* is the Policy-Related Equity Market Volatility tracker of Baker, Bloom, Davis, and Kost (2019); *GPR* is the Geopolitical Risk index of Caldara and Iacoviello (2022); *TED* is the TED spread; *SPX* is the disaster probability for the U.S. of Barro and Liao (2021); *T10Y2Y* is the spread for the T-bill between 10 and 2 year maturities. In Panel B we report p-values for tests based on Section A.2. The first test is for the null hypothesis that θ^0 is equal to -0.03. The test at the bottom of the table is for the null that θ^0 is the same across the two datasets: we refer to the values of θ^0 in the equity and multi-asset sample as to θ^0_E and θ^0_{MA} respectively. ***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 2: Number of estimated common factors

Equity							
	Number of factors				% of variance explained		
ε	All data	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Eigenvalue	All data	Regime $\mathcal{S}$	Regime $\mathcal{D}$
0.02	6	6	3	1	0.833	0.781	0.720
0.04	6	6	3	2	0.050	0.068	0.075
0.06	6	6	1	3	0.028	0.037	0.060
0.08	6	6	1	4	0.019	0.021	0.035
0.10	6	6	1	5	0.009	0.012	0.017
0.12	6	6	1	6	0.007	0.009	0.015

Multi-asset							
	Number of factors				% of variance explained		
ε	All data	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Eigenvalue	All data	Regime $\mathcal{S}$	Regime $\mathcal{D}$
0.02	4	4	3	1	0.712	0.621	0.679
0.04	4	4	3	2	0.085	0.115	0.117
0.06	4	4	3	3	0.053	0.071	0.079
0.08	4	4	1	4	0.031	0.036	0.039
0.10	4	4	1	5	0.021	0.028	0.020
0.12	4	4	1	6	0.013	0.018	0.013

Notes. For both equity and multi-asset portfolios, the table contains the number of estimated common factors for different values of ε defined in equation (19). For each dataset considered, we report inference on the number of common factors in the two regimes $\mathcal{S}$ and $\mathcal{D}$, as well as over the full sample period, which is tantamount to estimating the number of common factors in a linear specification ("All data"). The table also reports the percentage of the variance explained by each individual common factor through the corresponding eigenvalue, as defined according to (27).

Table 3: Goodness of fit measures

Panel A: R^2

	Equity					Multi-asset				
	Unconditional Model			Conditional Model		Unconditional Model			Conditional Model	
	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Regime $\mathcal{S}$
f_{1t}	0.865	-0.044	0.170	0.940	0.957	0.637	-0.106	0.186	0.916	0.812
$f_{1t}-f_{2t}$	0.865	-0.044	0.170	0.967	0.959	0.751	-0.105	0.232	0.939	0.904
$f_{1t}-f_{3t}$	0.905	-0.043	0.179	0.983	0.974	0.762	-0.105	0.236	0.951	0.909
$f_{1t}-f_{4t}$	0.977	-0.043	0.196	0.987	0.993	0.765	-0.105	0.237	0.956	0.911
$f_{1t}-f_{5t}$	0.977	-0.043	0.196	0.990	0.993	0.901	-0.103	0.292	0.958	0.961
$f_{1t}-f_{6t}$	0.977	-0.043	0.197	0.993	0.994	0.907	-0.102	0.294	0.960	0.961

Panel B: $\overline{RMS}_\alpha$

	Equity					Multi-asset				
	Unconditional Model			Conditional Model		Unconditional Model			Conditional Model	
	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Regime $\mathcal{S}$
f_{1t}	0.255	10.570	0.810	2.373	0.290	0.460	7.304	0.703	1.825	0.519
$f_{1t}-f_{2t}$	0.255	10.570	0.810	1.751	0.284	0.381	7.318	0.639	1.553	0.369
$f_{1t}-f_{3t}$	0.213	10.573	0.794	1.250	0.222	0.373	7.318	0.634	1.388	0.359
$f_{1t}-f_{4t}$	0.106	10.570	0.774	1.086	0.113	0.370	7.316	0.633	1.321	0.354
$f_{1t}-f_{5t}$	0.106	10.570	0.774	0.954	0.106	0.240	7.306	0.571	1.283	0.229
$f_{1t}-f_{6t}$	0.105	10.570	0.774	0.805	0.105	0.233	7.304	0.570	1.254	0.229

Panel C: $\overline{AVG}_\alpha$

	Equity					Multi-asset				
	Unconditional Model			Conditional Model		Unconditional Model			Conditional Model	
	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$	Regime $\mathcal{D}$	Regime $\mathcal{S}$
f_{1t}	0.020	-10.450	0.779	-0.526	0.037	0.266	-4.807	0.629	-0.010	0.299
	(0.138)	(0.000)	(0.000)	(0.766)	(0.062)	(0.000)	(0.000)	(0.000)	(0.964)	(0.000)
$f_{1t}-f_{2t}$	0.020	-10.449	0.780	-0.239	0.027	0.198	-4.875	0.561	-0.359	0.195
	(0.138)	(0.000)	(0.000)	(0.974)	(0.118)	(0.000)	(0.000)	(0.000)	(0.101)	(0.000)
$f_{1t}-f_{3t}$	0.013	-10.457	0.772	-0.218	0.016	0.163	-4.910	0.526	-0.107	0.156
	(0.360)	(0.000)	(0.000)	(0.959)	(0.360)	(0.000)	(0.000)	(0.000)	(0.519)	(0.000)
$f_{1t}-f_{4t}$	0.001	-10.468	0.761	-0.125	0.004	0.164	-4.910	0.526	-0.089	0.155
	(0.717)	(0.000)	(0.000)	(0.997)	(0.724)	(0.000)	(0.000)	(0.000)	(0.486)	(0.000)
$f_{1t}-f_{5t}$	0.001	-10.468	0.761	-0.094	0.002	0.048	-5.025	0.411	-0.024	0.042
	(0.706)	(0.000)	(0.000)	(0.999)	(0.749)	(0.280)	(0.000)	(0.000)	(0.793)	(0.315)
$f_{1t}-f_{6t}$	0.001	-10.468	0.760	-0.062	0.001	0.045	-5.028	0.408	-0.071	0.042
	(0.745)	(0.000)	(0.000)	(1.000)	(0.824)	(0.360)	(0.000)	(0.000)	(0.552)	(0.368)

Notes. This table contains the following goodness of fit measures for the equity and the multi-asset portfolios detailed in Section 6.1: the R-squared (R^2); the root-mean-squared pricing error ($\overline{RMS}_\alpha$); the average pricing error ($\overline{AVG}_\alpha$). In the case of $\overline{AVG}_\alpha$, between brackets the table reports the p-value of the bootstrap test for the null hypothesis that $\overline{AVG}_\alpha$ is equal to zero detailed in Appendix E. The measures are computed as a function of the number of latent factors included in the model. They are obtained for the unconditional model over the full sample, and in regime $j = \mathcal{D}, \mathcal{S}$; and for the conditional model in regime $j = \mathcal{D}, \mathcal{S}$. For the unconditional model, the pricing error over the full sample is $\alpha = \bar{\mathbf{R}} - \hat{\mathbf{X}}\hat{\mathbf{\Gamma}}$ where $\bar{\mathbf{R}} = \hat{T}^{-1}\sum_{t=1}^{T}\mathbf{R}_t$, whereas the pricing error in regime $j = \mathcal{D}, \mathcal{S}$ is $\alpha_{j\mathcal{U}} = \bar{\mathbf{R}}_j - \hat{\mathbf{X}}\hat{\mathbf{\Gamma}}$ where $\bar{\mathbf{R}}_j = \hat{T}_j^{-1}\sum_{t=1}^{T}\hat{d}_{j,t}\mathbf{R}_t$. For the conditional model, the pricing error in regime $j = \mathcal{D}, \mathcal{S}$ is $\alpha_j = \bar{\mathbf{R}}_j - \hat{\mathbf{X}}_j\hat{\mathbf{\Gamma}}_j$. R^2 is computed as: $1 - \left(\alpha'\alpha \,/ \bar{\mathbf{R}}'\bar{\mathbf{R}}\right)\left[(N-1)\Big/\Big(N-\hat{P}\Big)\right]$ for the unconditional model over the full sample; $1 - \left(\alpha'_{j\mathcal{U}}\alpha_{j\mathcal{U}} \,/ \bar{\mathbf{R}}'\bar{\mathbf{R}}\right)\left[(N-1)\Big/\Big(N-\hat{P}\Big)\right]$ for the unconditional model in regime $j = \mathcal{D}, \mathcal{S}$; $1 - \left(\alpha'_j\alpha_j \,/ \bar{\mathbf{R}}'_j\bar{\mathbf{R}}_j\right)\left[(N-1)\Big/\Big(N-\hat{P}_j\Big)\right]$ for the conditional model in regime $j = \mathcal{D}, \mathcal{S}$. $\overline{RMS}_\alpha$ is computed as: $\sqrt{\alpha'\alpha/N}$ for the unconditional model over the full sample; $\sqrt{\alpha'_{j\mathcal{U}}\alpha_{j\mathcal{U}}/N}$ for the unconditional model in regime $j = \mathcal{D}, \mathcal{S}$; $\sqrt{\alpha'_j\alpha_j/N}$ for the conditional model in regime $j = \mathcal{D}, \mathcal{S}$. $\overline{AVG}_\alpha$ is the corresponding average pricing error.

Table 4: Risk premia, tradable factors

	Equity				Multi-Asset			
	Unconditional	Conditional			Unconditional	Conditional		
		Regime $\mathcal{D}$	Regime $\mathcal{S}$	Avg		Regime $\mathcal{D}$	Regime $\mathcal{S}$	Avg
RmRf	0.513***	-9.133***	1.212***	0.513***	0.510**	-10.025***	1.260***	0.510***
f_{1t}	0.545***	-6.104***	1.122***	0.634***	0.848***	-5.781***	1.406***	0.928***
$f_{1t} - f_{2t}$	0.539**	-6.850***	1.175***	0.633***	0.996***	-6.556***	1.693***	1.145***
$f_{1t} - f_{3t}$	0.623**	-7.501***	1.277***	0.684***	1.002***	-6.984***	1.710***	1.132***
$f_{1t} - f_{4t}$	0.516**	-8.118***	1.170***	0.542***	0.924***	-7.044***	1.612***	1.037***
$f_{1t} - f_{5t}$	0.519**	-8.351***	1.196***	0.551***	0.881***	-7.521***	1.586***	0.981***
$f_{1t} - f_{6t}$	0.520**	-8.767***	1.199***	0.525**	0.830***	-7.672***	1.577***	0.962***
SMB	0.239**	-1.921***	0.396***	0.239**	0.054	-0.981	0.128	0.054
f_{1t}	0.217***	-5.619***	0.529***	0.113	0.295***	-4.788	0.695***	0.331**
$f_{1t} - f_{2t}$	0.230*	-3.713***	0.416***	0.137	-0.145	-4.920***	0.039	-0.291
$f_{1t} - f_{3t}$	0.174	-3.802***	0.359**	0.078	-0.143	-4.839***	0.027	-0.296
$f_{1t} - f_{4t}$	0.254*	-2.932***	0.450***	0.221	-0.043	-4.376***	0.153	-0.148
$f_{1t} - f_{5t}$	0.249*	-2.481***	0.406***	0.211	-0.002	-3.864***	0.166	-0.102
$f_{1t} - f_{6t}$	0.250*	-2.166***	0.409***	0.235	0.091	-3.546***	0.182	-0.066
HML	0.325***	1.707**	0.225**	0.325***	0.363**	1.396	0.289*	0.363**
f_{1t}	-0.062	-1.704	-0.135*	-0.242**	-0.140	-2.862	-0.408**	-0.571***
$f_{1t} - f_{2t}$	-0.064	-2.106	-0.124*	-0.257**	0.284*	-0.373	0.098	0.067
$f_{1t} - f_{3t}$	-0.236***	1.506	-0.358***	-0.232*	0.234	1.146	0.033	0.107
$f_{1t} - f_{4t}$	0.303***	2.460**	0.198	0.351**	0.363	1.497	0.205	0.291
$f_{1t} - f_{5t}$	0.307**	1.635	0.240*	0.334**	0.478**	1.687	0.348	0.437*
$f_{1t} - f_{6t}$	0.298**	1.348	0.208*	0.285**	0.429*	1.525	0.341	0.420*
RMW	0.258***	1.754***	0.149*	0.258***	0.470***	3.202***	0.276*	0.470***
f_{1t}	-0.086**	-0.352	-0.194***	-0.205**	-0.233**	-0.710	-0.478**	-0.493***
$f_{1t} - f_{2t}$	-0.090*	-0.038	-0.159**	-0.151	0.076	-0.080	-0.054	-0.055
$f_{1t} - f_{3t}$	-0.046	1.659*	-0.120	0.000	0.044	0.917	-0.076	-0.010
$f_{1t} - f_{4t}$	0.113*	0.723	0.004	0.052	0.049	0.366	-0.078	-0.049
$f_{1t} - f_{5t}$	0.110*	2.076**	-0.025	0.117	-0.049	0.022	-0.183	-0.169
$f_{1t} - f_{6t}$	0.126	2.260**	0.035	0.185*	0.032	0.085	-0.171	-0.154
CMA	0.282***	2.224***	0.142*	0.282***	0.323***	2.575***	0.163	0.323***
f_{1t}	-0.083**	-0.822	-0.159***	-0.204***	-0.156**	-1.197	-0.290***	-0.350***
$f_{1t} - f_{2t}$	-0.083**	-0.571	-0.158***	-0.186**	-0.008	-0.060	-0.139	-0.134
$f_{1t} - f_{3t}$	-0.155***	1.397**	-0.259***	-0.147*	-0.034	0.882	-0.169	-0.099
$f_{1t} - f_{4t}$	0.147	1.684**	0.048	0.159	0.040	0.625	-0.068	-0.022
$f_{1t} - f_{5t}$	0.150*	1.466*	0.077	0.171*	0.074	0.666	-0.027	0.020
$f_{1t} - f_{6t}$	0.143	1.674**	0.051	0.161*	0.047	0.560	-0.031	0.008

Notes. For equity and multi-asset portfolios, the table displays estimated risk premia for *RmRf*, *SMB*, *HML*, *RMW* and *CMA*, described in Section 6.4. Risk premia for the unconditional model are estimated over the full sample period. Risk premia for the conditional model are estimated within the regimes $\mathcal{D}$ and $\mathcal{S}$, and as a weighted average of the two according to (9) (Avg). ***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 5: Risk premia, non-tradable factors

	Equity				Multi-Asset			
	Unconditional	Conditional			Unconditional	Conditional		
		Regime $\mathcal{D}$	Regime $\mathcal{S}$	Avg		Regime $\mathcal{D}$	Regime $\mathcal{S}$	Avg
IP	+	-	+	+	+	-	+	+
f_{1t}	0.014**	-0.444*	0.047***	0.014	0.027*	-0.340	0.078***	0.050*
$f_{1t} - f_{2t}$	0.014*	-0.224	0.049***	0.031	0.026	-0.654**	0.092***	0.042
$f_{1t} - f_{3t}$	0.006	-0.431	0.038**	0.007	0.028	-0.679**	0.097***	0.046
$f_{1t} - f_{4t}$	-0.002	-0.400	0.028	-0.001	0.026	-0.846***	0.099***	0.036
$f_{1t} - f_{5t}$	-0.002	-0.362	0.026	-0.001	0.046*	-0.866***	0.124***	0.058
$f_{1t} - f_{6t}$	-0.002	-0.355	0.027	0.001	0.048*	-0.886***	0.124***	0.057
Liq	+	-	+	+	+	-	+	+
f_{1t}	0.241***	-7.311*	0.282***	-0.231	0.365**	-7.550	0.325**	-0.199
$f_{1t} - f_{2t}$	0.239**	-5.634	0.294***	-0.106	0.506***	-9.352*	0.521***	-0.135
$f_{1t} - f_{3t}$	0.266**	-7.527	0.253***	-0.273	0.616***	-9.269*	0.484**	-0.164
$f_{1t} - f_{4t}$	0.296**	-8.850*	0.308***	-0.311	0.460**	-9.582*	0.509**	-0.161
$f_{1t} - f_{5t}$	0.294**	-8.207*	0.284***	-0.290	0.161	-11.456**	0.324	-0.459
$f_{1t} - f_{6t}$	0.300**	-8.660*	0.300***	-0.305	0.148	-10.193	0.321	-0.378
ΔVXO	-	+	-	-	-	+	-	-
f_{1t}	-0.471***	2.748***	-0.811***	-0.540***	-0.411***	2.351**	-0.666***	-0.446***
$f_{1t} - f_{2t}$	-0.509***	2.772***	-0.883***	-0.606***	-0.466***	2.395***	-0.771***	-0.540***
$f_{1t} - f_{3t}$	-0.523***	3.299***	-0.892***	-0.573***	-0.471***	3.040***	-0.767***	-0.489***
$f_{1t} - f_{4t}$	-0.476***	3.450***	-0.854***	-0.527***	-0.429***	3.269***	-0.736***	-0.444***
$f_{1t} - f_{5t}$	-0.472***	2.518***	-0.846***	-0.590***	-0.396***	3.463***	-0.673***	-0.371**
$f_{1t} - f_{6t}$	-0.457**	2.524***	-0.828***	-0.573***	-0.346**	2.658***	-0.650***	-0.408**
HKM	+	-	+	+	+	-	+	+
f_{1t}	0.614***	-7.365***	1.332***	0.725***	0.981***	-9.279***	1.565***	0.845***
$f_{1t} - f_{2t}$	0.657**	-8.501***	1.490***	0.793**	1.449***	-10.143***	2.344***	1.514***
$f_{1t} - f_{3t}$	0.594*	-7.755***	1.382***	0.744**	1.387***	-9.697***	2.294***	1.497***
$f_{1t} - f_{4t}$	0.660**	-8.042***	1.424***	0.763**	1.340***	-9.538***	2.244***	1.461***
$f_{1t} - f_{5t}$	0.677**	-8.751***	1.514***	0.797**	1.335***	-9.446***	2.343***	1.559***
$f_{1t} - f_{6t}$	0.675**	-9.358***	1.509***	0.750**	1.194***	-10.597***	2.314***	1.456***
ΔEPU	-	+	-	-	-	+	-	-
f_{1t}	-0.849***	8.067	-1.092**	-0.396	-0.833***	8.418	-1.100**	-0.406
$f_{1t} - f_{2t}$	-0.843***	7.394	-1.076*	-0.433	-0.780***	10.896**	-1.026*	-0.157
$f_{1t} - f_{3t}$	-0.743**	7.603	-0.757	-0.122	-0.671**	10.883*	-0.914	-0.054
$f_{1t} - f_{4t}$	-0.683**	8.261	-0.704	-0.023	-0.717**	10.741**	-1.028*	-0.170
$f_{1t} - f_{5t}$	-0.693**	7.119	-0.692	-0.099	-0.119	11.802***	-0.376	0.512
$f_{1t} - f_{6t}$	-0.657*	7.116	-0.599	-0.013	0.011	7.277	-0.306	0.247
ΔGPR	-	+	-	-	-	+	-	-
f_{1t}	-0.383	6.294	-0.393	0.115	-0.411	2.005	-0.451	-0.272
$f_{1t} - f_{2t}$	-0.426	8.130	-0.380	0.267	-0.520	4.438	-0.596	-0.229
$f_{1t} - f_{3t}$	-0.326	14.518	-0.466	0.672	-0.100	3.806	-0.151	0.138
$f_{1t} - f_{4t}$	-0.297	13.902	-0.389	0.696	-0.043	7.483	-0.152	0.405
$f_{1t} - f_{5t}$	-0.467	7.646	-1.236	-0.562	-0.952	7.008	-1.308	-0.702
$f_{1t} - f_{6t}$	-0.253	7.446	-0.911	-0.277	-0.089	5.741	-0.914	-0.428
ΔEMV	-	+	-	-	-	+	-	-
f_{1t}	-0.375***	4.484***	-0.600***	-0.214	-0.416***	6.262***	-0.733***	-0.223
$f_{1t} - f_{2t}$	-0.351***	4.122***	-0.568***	-0.212	-0.331***	6.302***	-0.592***	-0.089
$f_{1t} - f_{3t}$	-0.340***	3.773**	-0.505***	-0.180	-0.361***	7.308***	-0.612***	-0.034
$f_{1t} - f_{4t}$	-0.300***	3.822*	-0.453***	-0.128	-0.348***	8.116***	-0.592***	0.043
$f_{1t} - f_{5t}$	-0.261***	2.098	-0.296**	-0.114	-0.222	8.018***	-0.389**	0.224
$f_{1t} - f_{6t}$	-0.268***	2.085	-0.291**	-0.110	-0.274	6.410***	-0.397**	0.099

Notes. This table provides results analogous to those in Table 4 for *IP*, *Liq*, ΔVXO, *HKM*, ΔEPU, ΔGPR, and ΔEMV, described in Section 6.4. "+ " and "-" are the expected sign of the corresponding risk premium.

Internet Appendix
(not for publication)

Factor models with downside risk

A Complements to Sections 4 and 5

In this section of the Internet Appendix, we complement the theory spelt out in Sections 4 and 5, adding further details and explanations. In particular, in Section A.1, we list and discuss all the relevant assumptions used in the paper. In Section A.2, we discuss in greater depth the test presented in Section 4.1. In Section A.3, we present all the relevant theory for the estimation of the number of common factors in each regime, discussed in Section 4.2. Finally, in Section A.4, we complete the asymptotic theory developed in Section 5.3.

Here and in the rest of the appendix, we use the notation c_0, c_1,... to denote positive and finite constants, which do not depend on the sample sizes and whose value can change from line to line, and we use the expression "a.s." as short-hand for "almost surely".

A.1 Assumptions and further notation

Let

$$\underline{\mathbf{u}}_t^0 = \mathbf{u}_{\mathcal{D},t}^0 \cup \mathbf{u}_{\mathcal{S},t}^0, \tag{A.1}$$

where $\underline{\mathbf{u}}_t^0$ has dimension P_u^0 such that $\max\left(P_{\mathcal{D}}^0, P_{\mathcal{S}}^0\right) \leq P_u^0 \leq P_{\mathcal{D}}^0 + P_{\mathcal{S}}^0$, and let $\mathbf{B}_{u,j}^0 = \left(\beta_{u,j,1}^0, ..., \beta_{u,j,N}^0\right)'$ be, for $j = \mathcal{D}, \mathcal{S}$, the $N \times P_u^0$ matrix suitably filled with zeros so that $\mathbf{B}_{u,j}^0 \underline{\mathbf{u}}_t^0 = \mathbf{B}_j^0 \mathbf{u}_{j,t}^0$ for $j = \mathcal{D}, \mathcal{S}$. Finally, let

$$\delta_i^0 = \beta_{u,\mathcal{S},i}^0 - \beta_{u,\mathcal{D},i}^0. \tag{A.2}$$

Assumption 1. *There exists an* $\eta \in \left(\frac{1}{2}, 1\right]$ *such that* $\delta_i^0 \neq 0$ *for* $i = 1, ..., \lfloor N^\eta \rfloor$ *and* $\sum_{i=\lfloor N^\eta \rfloor + 1}^N \|\delta_i^0\| < \infty$.

Assumption 1 is related to the results in Bates, Plagborg-Møller, Stock, and Watson (2013), who show that, if no more than a fraction $O\left(N^{1/2}\right)$ of the cross-sectional units in a large dimensional factor model have a structural break in the loadings, then the principal components estimator applied to a (misspecified) linear model has the same convergence rate as in the no break case - namely, $O_p\left(C_{N,T}^{-1}\right)$ (see Bai and Ng, 2002). Thus, Assumption 1 requires that at least as many as $O\left(N^\eta\right)$ of the N assets have regime specific factor loadings, for $\frac{1}{2} < \eta \leq 1$.

Assumption 2. *It holds that, for all N: (i) (a) the largest eigenvalue of $T_j^{-1}\sum_{t=1}^{T} E\left(\epsilon_t\epsilon_t' d_{j,t}\left(\theta^0\right)\right)$ is finite for $j = \mathcal{D}, \mathcal{S}$, and (b) the smallest eigenvalue of $T_j^{-1}\sum_{t=1}^{T} E\left(\epsilon_t\epsilon_t' d_{j,t}\left(\theta^0\right)\right)$ is positive for $j = \mathcal{D}, \mathcal{S}$; (ii)*

$$\max_{1\leq i\leq N}\sum_{k=1}^{N}\frac{1}{T_j}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{i,t}\epsilon_{k,s}d_{j,t}\left(\theta^0\right)d_{j,s}\left(\theta^0\right)\right)\right| < \infty,$$

for $j = \mathcal{D}, \mathcal{S}$.

Assumption 3. *It holds that (i) $E\left\|\mathbf{f}_{j,t}\right\|^{4+2r} < \infty$ for some $r > 0$, with $j = \mathcal{D}, \mathcal{S}$ and $1 \leq t \leq T$; (ii) it holds that, as $T \to \infty$*

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\mathbf{f}_{j,t}^0 d_{j,t}\left(\theta\right)\mathbf{f}_{j,t}' d_{j,t}\left(\theta^0\right)\right) \to \Sigma_{\mathbf{f},j}\left(\theta, \theta^0\right),$$

where $\Sigma_{\mathbf{f},j}\left(\theta, \theta^0\right)$ is positive definite for $j = \mathcal{D}, \mathcal{S}$ and all θ; (iii) $\left\|\beta_{j,i}^0\right\| < \infty$ for $j = \mathcal{D}, \mathcal{S}$ and $1 \leq i \leq N$; (iv) $N^{-1}\mathbf{B}_j'\mathbf{B}_j \to \Sigma_{\mathbf{B},j}$ as $N \to \infty$, where $\Sigma_{\mathbf{B},j}$ are positive definite matrices for $j = \mathcal{D}, \mathcal{S}$; (v) the eigenvalues of $\Sigma_{\mathbf{B},j}\Sigma_{\mathbf{f},j}\left(\theta^0, \theta^0\right)$ are distinct for $j = \mathcal{D}, \mathcal{S}$.

Assumption 2 stipulates that the idiosyncratic part of the model (i.e., the errors $\epsilon_{i,t}$) can be cross-sectionally correlated; however, by part *(i)* of the assumption such cross-sectional correlation is weak. Conversely, by Assumption 3, the "signal" part of the model (i.e., the one related to the common factor structure) introduces strong cross-sectional correlation between the $\widetilde{\mathbf{R}}_{i,t}$s. In particular, as we show in Lemma A.1 below, Assumption 3 entails that the eigenvalues of the covariance matrix of the $\widetilde{\mathbf{R}}_{i,t}$s have a spiked structure, whereby the largest eigenvalues diverge proportionally to N whereas the other ones are bounded.

Assumption 4. *It holds that (i) $E\left(\epsilon_{i,t}\right) = 0$ and $E\left|\epsilon_{i,t}\right|^8 < \infty$; (ii) $E\left(d_{j,t}\left(\theta\right)d_{j,v}\left(\theta\right)\epsilon_{i,t}\epsilon_{i,v}\right) = \tau_{j,i,t,v}\left(\theta\right)$ with $\left|\tau_{j,i,t,v}\left(\theta\right)\right| < \left|\tau_{j,t,v}\right|$ for $1 \leq i \leq N$, and $T_j^{-1}\sum_{t=1}^{T}\sum_{v=1}^{T}\left|\tau_{j,t,v}\right| < \infty$ for $j = \mathcal{D}, \mathcal{S}$; (iii) $E\left(T_j^{-1}\sum_{t=1}^{T} d_{j,t}\left(\theta\right)\epsilon_{i,t}\epsilon_{l,t}\right) = \sigma_{j,i,l,t}\left(\theta\right)$ with $\left|\sigma_{j,i,l,t}\left(\theta\right)\right| \leq \left|\sigma_{j,i,l}\right|$ for $1 \leq t \leq T$ and $N^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\left|\sigma_{j,i,l}\right| < \infty$ for $j = \mathcal{D}, \mathcal{S}$; (iv)*

$$E\left|T_j^{-1/2}\sum_{t=1}^{T}\left(d_{j,t}\left(\theta\right)\epsilon_{i,t}\epsilon_{l,t} - E\left(d_{j,t}\left(\theta\right)\epsilon_{i,t}\epsilon_{l,t}\right)\right)\right|^4 < \infty,$$

for $j = \mathcal{D}, \mathcal{S}$ *and* $1 \leq i, l \leq N$*; (v) it holds that* $E\left(\mathbf{f}_{j,s}^{0} \epsilon_{i,t} | z_t\right) = 0$ *for* $j = \mathcal{D}, \mathcal{S}$ *and* $1 \leq t, s \leq T$*, with*

$$E\left[N^{-1} \sum_{i=1}^{N} \left\| T_j^{-1/2} \sum_{t=1}^{T} d_{j,t}\left(\theta\right) \mathbf{f}_{j,t}^{0} \epsilon_{i,t} \right\|^2 \right] < \infty,$$

for all θ *and* $j = \mathcal{D}, \mathcal{S}$*.*

Assumption 4 mimics similar assumptions in the literature (see e.g. Assumptions C and D in Bai, 2003), and essentially it states that the idiosyncratic errors are weakly dependent.

Assumption 5. *It holds that (i)* $\left\{\mathbf{f}_{j,t}^{0}, z_t, \epsilon_t\right\}$ *is a strictly stationary, ergodic,* ρ*-mixing sequence with mixing numbers* $\rho_{j,m}$ *satisfying* $\sum_{m=1}^{\infty} \rho_{j,m}^{1/2} < \infty$ *for* $j = \mathcal{D}, \mathcal{S}$ *(ii)* $E\left(\left\|\mathbf{f}_{j,t}^{0} \epsilon_{i,t}\right\|^4 | z_t\right) < \infty$ *a.s. and* $E\left(\left\|\mathbf{f}_{j,t}^{0}\right\|^4 | z_t\right) < \infty$ *a.s. for* $j = \mathcal{D}, \mathcal{S}$ *and* $1 \leq i \leq N$*; (iii)* z_t *has bounded and strictly positive density; (iv) the density of* z_t *is continuous at* θ^0*; (v)* $E\left(\mathbf{f}_{j,t}^{0} \mathbf{f}_{j,t}^{0\prime} | z_t = \theta\right)$ *is a.s. positive definite and continuous at* $\theta = \theta^0$ *for* $j = \mathcal{D}, \mathcal{S}$ *with* $\sum_{i=\lfloor N^{\eta} \rfloor + 1}^{N} \delta_i^{0\prime} E\left(\mathbf{f}_{j,t}^{0} \mathbf{f}_{j,t}^{0\prime} | z_t = \theta^0\right) \delta_i^0 < \infty$ *a.s.; (vi)* $\left\{\mathbf{f}_{j,t}^{0}\right\}$ *and* $\{\epsilon_{i,t}\}$ *are two mutually independent groups conditionally on* z_t *for* $j = \mathcal{D}, \mathcal{S}$ *and* $1 \leq i \leq N$*.*

Assumption 5 draws upon Assumption 1 in Hansen (2000) (see also Massacci, 2017), to which we refer to for further discussion. In our context, it is used in order to obtain the convergence rate of the estimators for factor innovations and loadings. We point out that some of the other assumptions above - chiefly the ones that contain "time series results"(e.g., Assumptions 4*(iv)*-*(v)*) - would follow from Assumption 5, although they could also be derived from other dependence assumptions.

A final note on the assumptions above. As mentioned above, Assumption 3*(iv)* is typical in the literature on inference on factor models (we refer, inter alia, to the contributions by Bai and Ng, 2002 and Bai, 2003) and, in essence, it requires that the common factors be "strong" or "pervasive". A well-known consequence of the assumption is that - as we show in Lemma A.1 - the eigenvalues of the second moment matrix of the data diverge at a rate exactly equal to N, which we exploit in the construction of our tests for the determination of the number of common factors. On the other hand, Assumption 3*(iv)* rules out the potentially interesting case of having "weak" common factors. As pointed out in Uematsu and Yamagata (2021), a "weak" common factor can be defined as a common factor whose corresponding eigenvalue in the second moment of the matrix diverges, but at a rate slower than N. Such a case could be of interest in general, and especially in the context of a threshold factor model, where e.g. a pervasive factor affects only a fraction of the units in one regime, and is absent in the other regime.

Assumption 6. *As* $\min(N,T)\to\infty$*, it holds that*

$$\frac{(\ln N)^{1+\epsilon}(\ln T)^{1/2+\epsilon}}{T^{1/2}\pi_j}=0,$$

for $j=\mathcal{D},\mathcal{S}$*.*

By Assumption 6, $\widehat{g}_j^{(i)}-g_j^{(i)}$ is of a smaller order of magnitude than N. As mentioned above, this ensures the spiked structure of the spectrum of the sample second moment matrices.[i]

Assumption 7. [ii]*It holds that, for* $j=\mathcal{D},\mathcal{S}$*: (i)* $E\left(\sum_{i=1}^{N}w_i\alpha_{j,i}\right)^2\leq c_0N$ *for all* $\{w_i\}_{i=1}^{N}$ *such that* $\sum_{i=1}^{N}w_i^2<\infty$*; (ii)* $\{\alpha_{j,i}\}_{i=1}^{N}$ *is independent of all other quantities; (iii) it holds that, as* $N\to\infty$*,* $\left[Var\left(\sum_{i=1}^{N}v_i\alpha_{j,i}\right)\right]^{-1/2}\sum_{i=1}^{N}v_i\alpha_i\xrightarrow{D}N(0,I_\nu)$ *for every* ν*-dimensional sequence* v_i*,* $1\leq i\leq N$*, such that* $\left\|N^{-1}\sum_{i=1}^{N}v_iv_i'\right\|<\infty$*.*

Assumption 8. *It holds that: (i) in (4), (a)* $\|\Lambda\|=O(1)$ *and (b)* $E(\mathbf{e}_t)=0$ *with* $E\|\mathbf{e}_t\|^4<\infty$*; (ii)* $T^{-1}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{u}_{j,t}^{0\prime}=O_P(1)$ *for* $j=\mathcal{D},\mathcal{S}$*.*

Assumption 9. *For* $j=\mathcal{D},\mathcal{S}$*, it holds that (i)* $E\left(\mathbf{e}_t|\mathbf{u}_{j,t}^{0}\right)=0$ *a.s. for all* t *and (ii) as* $T\to\infty$

$$T^{-1/2}\begin{pmatrix}\sum_{t=1}^{T}vec\left(\mathbf{e}_t\mathbf{u}_{j,t}^{0\prime}\right)d_{j,t}\left(\theta^0\right)\\ \sum_{t=1}^{T}\mathbf{u}_{j,t}^{0}d_{j,t}\left(\theta^0\right)\end{pmatrix}\xrightarrow{D}N\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}V_{\mathbf{eu},j}&\Pi_{\mathbf{ue},j}'\\ \Pi_{\mathbf{ue},j}&V_{\mathbf{u},j}\end{pmatrix}\right],$$

[i]In essence, the assumption is a restriction on the rates at which π_j can drift to zero: in particular, the assumption is satisfied as long as $\pi_j^{-1}=O\left(T^{1/2-\epsilon}\right)$ for any $\epsilon>0$. Note that the assumption also (very mildly) restricts the relative rate of divergence between N and T as they pass to infinity - in essence, however, it allows for N to grow at an arbitrarily high polynomial order with T, i.e. it allows for $N=O\left(T^{\kappa}\right)$ for any value of κ.

[ii]Assumption 7 contains some regularity conditions on the pricing errors α_i in (1). By part *(i)*, we require that the second moment of the α_is grows linearly with N; a possible case in which this would happen is when α_i is *i.i.d.* with zero mean and finite variance. However, the assumption is more general than this, and it allows e.g. for the α_is to have nonzero mean. Indeed, if $E(\alpha_i)\neq 0$ for $1\leq i\leq N_\alpha$ with $N_\alpha=O\left(N^{1/2}\right)$, and $E\left(\sum_{i=N_\alpha+1}^{N}w_i\alpha_i\right)^2\leq c_0N$, the assumption would hold anyway. Note also that part *(iii)* of the assumption could be derived using more primitive assumptions on the α_is (see e.g. the very general CLT in McLeish, 1974).

where

$$
\begin{aligned}
V_{\mathbf{eu},j} &= \lim_{T\to\infty} E\left(T^{-1}\sum_{t,s=1}^{T} vec\left(\mathbf{e}_t \mathbf{u}_{j,t}^{0\prime}\right)\left(vec\left(\mathbf{e}_s \mathbf{u}_{j,s}^{0\prime}\right)\right)' d_{j,t}\left(\theta^0\right) d_{j,s}\left(\theta^0\right)\right), \\
V_{\mathbf{u},j} &= \lim_{T\to\infty} E\left(T^{-1}\sum_{t,s=1}^{T} \mathbf{u}_{j,t}^{0} \mathbf{u}_{j,s}^{\prime 0} d_{j,t}\left(\theta^0\right) d_{j,s}\left(\theta^0\right)\right), \\
\Pi_{\mathbf{ue},j} &= \lim_{T\to\infty} E\left(T^{-1}\sum_{t,s=1}^{T} \mathbf{u}_{j,t}^{0} \left(vec\left(\mathbf{e}_s \mathbf{u}_{j,s}^{0\prime}\right)\right)' d_{j,t}\left(\theta^0\right) d_{j,s}\left(\theta^0\right)\right).
\end{aligned}
$$

Assumption 10. *It holds that $\alpha_{j,i}$ is i.i.d. across i with $E\left(\alpha_{j,1}^2\right) = \sigma_{\alpha,j}^2 < \infty$ for $j = \mathcal{D}, \mathcal{S}$.*

Assumption 10 is a simplifying assumption, which complements Assumption 7 by requiring the homoskedasticity of the $\alpha_{j,i}$ across i.

Assumption 11. *It holds that: (i) $\alpha_{\mathcal{D}}$ and $\alpha_{\mathcal{D}}$ are independent; (ii) as $T \to \infty$*

$$
T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} vec\left(e_t u_{\mathcal{D},t} d_{\mathcal{D},t}\right) \\ u_{\mathcal{D},t} d_{\mathcal{D},t} \\ vec\left(e_t u_{\mathcal{S},t} d_{\mathcal{S},t}\right) \\ u_{\mathcal{D},t} d_{\mathcal{S},t} \end{pmatrix} \xrightarrow{d} N\left(0, V\right)
$$

with

$$
V = \begin{pmatrix} V_{ue,\mathcal{D}} & & & \\ \Pi_{ue,\mathcal{D}} & V_{u,\mathcal{D}} & & \\ \Pi_{ue,\mathcal{DS}} & \Pi_{ue,\mathcal{S};u,\mathcal{D}} & V_{ue,\mathcal{S}} & \\ \Pi_{ue,\mathcal{D};u,\mathcal{S}} & \Pi_{u,\mathcal{D};u,\mathcal{S}} & \Pi_{ue,\mathcal{S}} & V_{u,\mathcal{S}} \end{pmatrix},
$$

where $V_{ue,j}$, $V_{u,j}$, and $\Pi_{ue,j}$ are defined in Assumption 9 (for $j = \mathcal{D}, \mathcal{S}$) and

$$\begin{aligned}
\Pi'_{ue,\mathcal{S};u,\mathcal{D}} &= \frac{1}{T} E\left(\left(\sum_{t=1}^{T} u_{\mathcal{D},t} d_{\mathcal{D},t}\right)\left(\sum_{t=1}^{T} vec\left(e_t u_{\mathcal{S},t} d_{\mathcal{S},t}\right)\right)'\right), \\
\Pi'_{ue,\mathcal{D};u,\mathcal{S}} &= \frac{1}{T} E\left(\left(\sum_{t=1}^{T} u_{\mathcal{S},t} d_{\mathcal{S},t}\right)\left(\sum_{t=1}^{T} vec\left(e_t u_{\mathcal{D},t} d_{\mathcal{D},t}\right)\right)'\right), \\
\Pi'_{ue,\mathcal{D}\mathcal{S}} &= \frac{1}{T} E\left(\left(\sum_{t=1}^{T} vec\left(e_t u_{\mathcal{D},t} d_{\mathcal{D},t}\right)\right)\left(\sum_{t=1}^{T} vec\left(e_t u_{\mathcal{S},t} d_{\mathcal{S},t}\right)\right)'\right), \\
\Pi'_{u,\mathcal{D};u,\mathcal{S}} &= \frac{1}{T} E\left(\left(\sum_{t=1}^{T} u_{\mathcal{D},t} d_{\mathcal{D},t}\right)\left(\sum_{t=1}^{T} u_{\mathcal{S},t} d_{\mathcal{S},t}\right)'\right).
\end{aligned}$$

A.2 Complements to Section 4.1: the test for $\theta^0 = c$

We discuss our test for $H_0 : \theta^0 = c$. The test is based on similar passages (albeit in a very different context) to Horváth and Trapani (2019); some arguments are very similar to the ones used in Section A.3, and we therefore present them only once, to avoid repetitions, in the next section.

We know that, by Lemma C.7

$$\widehat{\theta} - \theta^0 = o_{a.s.}\left(T^{-1}\overline{v}_{N,T}(\epsilon)\right),$$

where $\overline{v}_{N,T}(\epsilon)$ is defined in (C.4). Define now the deterministic sequence $s_{N,T}$ such that

$$\lim_{\min(N,T)\to\infty} s_{N,T} = \infty,$$

$$\lim_{\min(N,T)\to\infty} s_{N,T} T^{-1}\overline{v}_{N,T}(\epsilon) = 0,$$

for all $\epsilon > 0$. By construction, we have $s_{N,T}\left(\widehat{\theta} - c\right) = s_{N,T}\left(\theta^0 - c\right) + o_{a.s.}(1)$. This, and Lemma C.7, entail that, under H_0

$$P\left(\omega : \lim_{\min(N,T)\to\infty} s_{N,T}\left(\widehat{\theta} - c\right) = 0\right) = 1,$$

and therefore we can assume that, under the null

$$\lim_{\min(N,T)\to\infty} s_{N,T}\left(\widehat{\theta}-c\right)=0.$$

Conversely, whenever $\theta^0 \neq c$, it follows that

$$P\left(\omega : \lim_{\min(N,T)\to\infty} s_{N,T}\left(\widehat{\theta}-c\right)=\infty\right)=1,$$

and therefore we can assume that, under H_A

$$\lim_{\min(N,T)\to\infty} s_{N,T}\left(\widehat{\theta}-c\right)=\infty.$$

This dichotomous behaviour can be reversed by choosing a transformation $h(\cdot)$ such that $\lim_{x\to\infty} h(x) = 0$ and $\lim_{x\to 0} h(x)=\infty$, thus defining

$$f_{NT}=h\left(s_{N,T}\left(\widehat{\theta}-c\right)\right).$$

Exactly as above, we can assume, on account of continuity, that

$$\lim_{\min(N,T)\to\infty} h\left(s_{N,T}\left(\widehat{\theta}-c\right)\right)=\infty,$$
$$\lim_{\min(N,T)\to\infty} h\left(s_{N,T}\left(\widehat{\theta}-c\right)\right)=0,$$

under H_0 and H_A respectively.

We can now construct the following test (also similar to the one in Section 4.2 below):

Step 1 Generate an *i.i.d.* sequence $\left\{\xi_m^\theta, 1\leq m\leq M_\theta\right\}$, with common distribution $N(0,1)$;

Step 2 Generate the Bernoulli sequence $\zeta_m^\theta(s)=I\left(f_{N,T}^{1/2}\xi_m^\theta\leq s\right)$;

Step 3 Define

$$S^\theta=\int_{-\infty}^{+\infty}\left|\sigma^\theta(s)\right|^2\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}s^2\right)ds,$$

where

$$\sigma^{\theta}(s) = \frac{2}{M_{\theta}^{1/2}} \sum_{m=1}^{M_{\theta}} \left(\zeta_m^{\theta}(s) - \frac{1}{2} \right).$$

Theorem A.1. *We assume that the assumptions of Lemma C.7 are satisfied. Then, as* $\min(N, M_{\theta}, T) \to \infty$ *with*

$$\frac{M_{\theta}}{h^{1/2}\left(s_{N,T} T^{-1} \overline{\upsilon}_{N,T}(\epsilon)\right)} \to 0, \tag{A.3}$$

it holds that $S^{\theta} \overset{D^*}{\to} \chi_1^2$*, under* H_0*, and that* $R^{-1} S^{\theta} \overset{P^*}{\to} c_0 > 0$ *under* H_A*, for almost all realizations of* $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$.

The proof of Theorem A.1 is very similar to the proofs in Horváth and Trapani (2019), and it is therefore omitted to save space. Letting χ_1^2 denote a chi-squared distribution with one degree of freedom, and defining the critical value c_{α} as $P(\chi_1^2 \geq c_{\alpha}) = \alpha$ for a nominal level α, Theorem A.1 entails

$$\lim_{\min(N,T,M_{\theta}) \to \infty} P^* \left(S^{\theta} \geq c_{\alpha} \right) = 1, \tag{A.4}$$

under H_A, and

$$\lim_{\min(N,T,M_{\theta}) \to \infty} P^* \left(S^{\theta} \geq c_{\alpha} \right) = \alpha, \tag{A.5}$$

under H_0. Equation (A.4) has the "traditional" interpretation: when the null is false, the test - asymptotically - rejects with probability 1. Conversely, equation (A.5) does not have a straightforward meaning. The test is constructed using a randomization based on $\left\{\xi_m^{\theta}\right\}$ which does not vanish asymptotically, and therefore the asymptotics of S^{θ} is driven by the added randomness. Thus, different researchers using the same data will obtain different values of S^{θ}, and, consequently, different p-values. In order to ameliorate this, we - along the lines of Horváth and Trapani (2019) - propose to generate the test statistic S_b^{θ} for $1 \leq b \leq B$ times, using, at each iteration, an *i.i.d.* sequence $\left\{\xi_{m,b}^{\theta}\right\}$, $1 \leq m \leq M_{\theta}$, with common distribution $N(0,1)$, independent across b. Define

$$Q_{\alpha} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(S_b^{\theta} \leq c_{\alpha} \right). \tag{A.6}$$

The Law of the Iterated Logarithm entails that[iii]

$$\lim\inf_{B\to\infty}\lim_{\min(N,T,M_\theta)\to\infty}\sqrt{\frac{B}{2\ln\ln B}}\frac{Q_\alpha-(1-\alpha)}{\sqrt{\alpha(1-\alpha)}}=-1.$$

Hence, we propose the following "strong rule" to decide in favour of H_0:

$$Q_\alpha \geq (1-\alpha)-\sqrt{\alpha(1-\alpha)}\sqrt{\frac{2\ln\ln B}{B}}. \tag{A.7}$$

Finally, we point out that, in our empirical exercises, we have used $f_{NT}=\left(\sqrt{T}\left|\widehat{\theta}-c\right|\right)^{-1}$ and $M_\theta = T^{1/4}$.

The test for $H_0 : \theta^0 = c$ discussed above can be generalized in order to compare estimates of θ from different datasets. Indeed, let the subscripts 1 and 2 denote each datatset, and define the sample sizes as (N_1, T_1) and (N_2, T_2), and the relevant estimates as $\widehat{\theta}_1$ and $\widehat{\theta}_2$. Then, upon defining a sequence $s^*_{N,T} = s^*_{N,T}(T_1, T_2)$ such that

$$\lim_{\min(N_1,T_1,N_2,T_2)\to\infty} s^*_{N,T}=\infty,$$
$$\lim_{\min(N_1,T_1,N_2,T_2)\to\infty}\frac{s^*_{N,T}}{\min\{T_1,T_2\}}\max\left\{\overline{v}_{N_1,T_1}(\epsilon),\overline{v}_{N_2,T_2}(\epsilon)\right\}=0,$$

the test described above can be applied readily using

$$f^*_{N,T}=\frac{1}{\sqrt{\min\{T_1,T_2\}}\left|\widehat{\theta}_1-\widehat{\theta}_2\right|},$$

e.g. using, in Step 1 of the algorithm, $M_\theta = \min\left\{T_1^{1/4}, T_2^{1/4}\right\}$.

[iii]Details on this argument are in Horváth and Trapani (2019).

A.3 Complements to Section 4.2: theory

Equations (15) and (16) in Section 4.2 contain two estimators of the population covariance matrices, defined as

$$\Sigma_{\mathcal{D}} = \frac{1}{T\pi_{\mathcal{S}}} \sum_{t=1}^{T} E\left(\widetilde{\mathbf{R}}_t \widetilde{\mathbf{R}}_t' d_{\mathcal{D},t}\left(\theta^0\right)\right), \tag{A.8}$$

$$\Sigma_{\mathcal{S}} = \frac{1}{T\pi_{\mathcal{S}}} \sum_{t=1}^{T} E\left(\widetilde{\mathbf{R}}_t \widetilde{\mathbf{R}}_t' d_{\mathcal{S},t}\left(\theta^0\right)\right). \tag{A.9}$$

Recall that we denote as $g_j^{(i)}$ the i-th largest eigenvalue of $\Sigma_{\mathcal{D}}$ and $\Sigma_{\mathcal{S}}$ (according as $j = \mathcal{D}$ or $\mathcal{S}$).

Lemma A.1. *We assume that Assumptions 1-3 are satisfied. Then it holds that, for* $j = \mathcal{D}, \mathcal{S}$

$$\underline{c}_j^{(i)} N \leq g_j^{(i)} \leq \overline{c}_j^{(i)} N, \ \textit{for } 0 \leq i \leq P_j^0,$$

for some $0 < \underline{c}_j^{(i)} \leq \overline{c}_j^{(i)} < \infty$*, and*

$$g_j^{(i)} \leq c_j^{(i)}, \ \textit{for } i \geq P_j^0 + 1,$$

for some $c_j^{(i)} < \infty$*.*

Theorem A.2. *We assume that Assumptions 1-5 are satisfied. Then it holds that, for all* $1 \leq i \leq N$ *and* $j = \mathcal{D}, \mathcal{S}$

$$\widehat{g}_j^{(i)} = g_j^{(i)} + o_{a.s.}\left(\frac{NT^{1/2}}{T\pi_j^0} (\ln N)^{1+\epsilon} (\ln T)^{1/2+\epsilon}\right),$$

for every $\epsilon > 0$*.*

Lemma A.1 and Theorem A.2 can be read together. According to Lemma A.1, the spectrum of each second moment matrix has a spiked structure; more specifically, the largest P_j^0 eigenvalues diverge to positive infinity as $N \to \infty$, whereas the others are bounded. The fact that the divergence rate is exactly N is a direct consequence of Assumption 3*(iv)* and also of having assumed $\eta = 1$. However, as long as the largest P_j^0 eigenvalues diverge, even at a slower rate, the arguments below are still valid although the construction of the tests may be more convoluted.

Theorem A.2 offers the sample counterpart to Lemma A.1, giving a bound for the estimation error $\widehat{g}_j^{(i)} - g_j^{(i)}$.[iv] [v]

Hereafter, we provide some comments, and the relevant asymptotics, on the individual tests for (17), viz.

$$\begin{cases} H_0: & g_j^{(i)} = c_j^{(i)} N \\ H_A: & g_j^{(i)} \leq c_j^{(i)} \end{cases}.$$

Since Theorem A.2 only contains rates, we propose a randomized version of the estimated eigenvalues $\widehat{g}_j^{(i)}$, based on Trapani (2018). Based on Lemma A.1, each $g_j^{(i)}$ should diverge or not according to whether there are at least i common factors or fewer. Thus, we would expect the same separation result to also hold for $\widehat{g}_j^{(i)}$. Indeed, whenever $g_j^{(i)}$ diverges as fast as N (which, in (17), represents the null hypothesis), $\widehat{g}_j^{(i)}$ also does, at the same rate: this is an immediate consequence of Theorem A.2. Conversely, there is no guarantee that $\widehat{g}_j^{(i)}$ converges when $g_j^{(i)}$ does so: in this case (which corresponds to the alternative hypothesis in (17)), from Theorem A.2, it is still possible that the estimation error $\widehat{g}_j^{(i)} - g_j^{(i)}$ will diverge at a rate $NT^{-1/2}$, modulo some logarithmic terms.

Therefore, in order to ensure that $\widehat{g}_j^{(i)}$ has the same separation result as its population counterpart $g_i^{(j)}$, we propose to use the statistic $\psi\left(\widehat{g}_j^{(i)}\right)$ defined in (18). In (19), we have defined

$$\varrho_j = \begin{cases} \varepsilon & \text{when } \frac{\ln T_j}{\ln N} \geq \frac{1}{2} \\ 1 - \frac{1}{2}\frac{\ln T_j}{\ln N} + \varepsilon & \text{when } \frac{\ln T_j}{\ln N} < \frac{1}{2} \end{cases}.$$

[iv] As is typical in (large) Random Matrix Theory, there is no guarantee that $\widehat{g}_j^{(i)} - g_j^{(i)}$ will drift to zero as $\min(N, T) \to \infty$. See for example the related contributions by Bai and Yao (2008), Bai and Yao (2012) and Wang and Fan (2017), and also Lemma 2 in Trapani (2018), where a similar result is derived, under different assumptions and for a linear model. Indeed, $\widehat{g}_j^{(i)} - g_j^{(i)}$ may even diverge to infinity. However, as long as $\widehat{g}_j^{(i)} - g_j^{(i)}$ is of smaller magnitude than $g_j^{(i)}$ for $j \leq P_j^0$, the spiked structure of the spectrum of the second moment matrices is preserved, in that $\widehat{g}_j^{(i)}$ is of smaller magnitude when $j > P_j^0$ compared to the case $j \leq P_j^0$.

[v] Note that we are not imposing that $0 < \pi_{\mathcal{D}}, \pi_{\mathcal{S}} < 1$; we are therefore entertaining the possibility that e.g. $\pi_{\mathcal{D}} = \pi_{\mathcal{D}}(T) \to 0$, which may be well suited to a situation in which one of the two regimes occurs very rarely.

The rationale for this choice is that we need to choose ϱ_j such that

$$\lim_{\min(N,T)\to\infty} \frac{N^{1-\varrho_j}T^{1/2}}{T\pi_j} (\ln N)^{1+\epsilon} (\ln T)^{1/2+\epsilon} = 0; \tag{A.10}$$

(19) is therefore proposed under the condition $0 < \pi_{\mathcal{D}}, \pi_2 < 1$, although it is valid for any values of $\pi_{\mathcal{D}}$ and $\pi_{\mathcal{S}}$. If this is violated, in principle (A.10) can be employed to construct a refined proposal for ϱ_j. Based on (19) and on the strong rates in Theorem A.2, it is easy to see that

$$\lim_{\min(N,T)\to\infty} \psi\left(\widehat{g}_j^{(i)}\right) = \infty \text{ under } H_0,$$
$$\lim_{\min(N,T)\to\infty} \psi\left(\widehat{g}_j^{(i)}\right) = 1 \text{ under } H_A.$$

Let $\varkappa$ be defined such that

$$0 < \varkappa < \frac{\underline{c}_j^{(i)}}{\frac{1}{N-p}\sum_{h=p+1}^{N} \overline{c}_j^{(h)}},$$

where $\underline{c}_j^{(i)}$ and the $\overline{c}_j^{(h)}$s are defined in Lemma A.1. It holds that

Theorem A.3. *We assume that Assumptions 1-6 are satisfied with $\eta = 1$ in Assumption 1. Let $M = M(N)$, so that $\lim_{N\to\infty} M(N) = \infty$. Then, as $\min(N, T_j) \to \infty$ with*

$$M \exp\left(-\varkappa N^{1-\varrho_j}\right) \to 0, \tag{A.11}$$

it holds that under H_0

$$\Upsilon_j^{(i)} \overset{D^*}{\to} \chi_1^2, \tag{A.12}$$

for almost all realizations of $\{R_{i,t}, 1 \le i \le N, 1 \le t \le T\}$, $j = \mathcal{D}, \mathcal{S}$ and all $1 \le i \le N$.

Under H_A, it holds that there exists a positive, finite constant c_0 such that

$$\frac{1}{4M}\Upsilon_j^{(i)} \overset{P^*}{\to} c_0, \tag{A.13}$$

for almost all realizations of $\{R_{i,t}, 1 \le i \le N, 1 \le t \le T\}$, $j = \mathcal{D}, \mathcal{S}$ and all $1 \le i \le N$.

The theorem, and the algorithm above, suggest that some tuning is needed prior to implementing the test. To begin with, note that the power of tests based on $\Upsilon_j^{(i)}$ increases with M by (A.13); conversely, in the proof of (A.12) it is shown that the test statistic has a non-centrality parameter which is controlled by (A.11), and which vanishes faster the smaller M. Thus, the choice of M reflects the well-known size-power trade-off. In Section B and in the empirics, we use $M = N$, which satisfies (A.11) and is our recommended choice.

In Step 2, we recommend to draw the values of s from a standard normal. By integrating s out in Step 4 of the algorithm, the statistic $\Upsilon_j^{(i)}$ becomes scale-invariant, in the sense that $\Upsilon_j^{(i)}$ does not depend on the support of s.[vi] This form of scale invariance is clearly desirable. From a practical point of view, we propose to use a Gauss-Hermite quadrature to approximate the integral that defines $\Upsilon_j^{(i)}$, viz.

$$\widehat{\Upsilon}_j^{(i)} = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n_S} w_i \left(\upsilon_j^{(i)} \left(\sqrt{2} x_i \right) \right)^2, \tag{A.14}$$

where the x_i, $1 \leq i \leq n_S$, are the zeros of the Hermite polynomial $H_{n_S}(x)$ and

$$w_i = \frac{\sqrt{\pi} 2^{n_S - 1} (n_S - 1)!}{n_S \left[H_{n_S - 1} \left(\sqrt{2} x_i \right) \right]^2}. \tag{A.15}$$

Thus, when constructing $\upsilon_j^{(i)}(s)$, we construct n_S of these statistics, each with $s = \sqrt{2} x_i$.[vii]

[vi] We point out that, in the original paper by Trapani (2018), this is not the case since the methodology is implemented by extracting several values of s from a uniform distribution with support, say, S: doing this makes the test statistics dependent on S, thus making them not scale invariant.

[vii] The values of the roots x_i, and of the corresponding weights w_i, are tabulated e.g. in Salzer, Zucker, and Capuano (1952).

A.4 Complements to Section 5.2

A.4.1 Assumptions

For $j = \mathcal{D}, \mathcal{S}$, consider the matrices $M_{\mathbf{u}}^{(j)} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbf{u}_{j,t}^{0} \mathbf{u}_{j,t}^{0\prime} d_{j,t}\left(\theta^{0}\right)$. We then define

$$\begin{aligned}
\Sigma_{\gamma,j} &= c_j^{-2} \mathbf{\Lambda}_j^0 V_{\mathbf{u},j} \mathbf{\Lambda}_j^{0\prime} + \left(\left(\gamma_{j,1}^{0\prime} \left(M_{\mathbf{u}}^{(j)}\right)^{-1}\right) \otimes I_K\right) V_{\mathbf{eu},j} \left(\left(\left(M_{\mathbf{u}}^{(j)}\right)^{-1} \gamma_{j,1}^{0}\right) \otimes I_K\right) \\
&\quad + c_j^{-1} \mathbf{\Lambda}_j^0 \Pi_{\mathbf{ue},j} \left(\left(\left(M_{\mathbf{u}}^{(j)}\right)^{-1} \gamma_{j,1}^{0}\right) \otimes I_K\right) + c_j^{-1} \left(\left(\gamma_{j,1}^{0\prime} \left(M_{\mathbf{u}}^{(j)}\right)^{-1}\right) \otimes I_K\right) \Pi_{\mathbf{ue},j}' \mathbf{\Lambda}_j^{0\prime}, \\
\Sigma_{\alpha\gamma,j} &= \sigma_{\alpha,j}^2 \mathbf{\Lambda}_j^0 \left[\frac{1}{N} \mathbf{B}_j^{0\prime} \mathbf{B}_j^0 - \left(\frac{1}{N} \mathbf{B}_j^{0\prime} i_N\right) \left(\frac{1}{N} \mathbf{B}_j^{0\prime} i_N\right)'\right]^{-1} \mathbf{\Lambda}_j^{0\prime},
\end{aligned}$$

where i_N is an $N \times 1$ vector of ones.

A.4.2 Estimation of long-run covariance matrices

We propose the following estimators

$$\begin{aligned}
\widehat{\Sigma}_{\gamma,j} &= c_j^{-2} \widehat{\mathbf{\Lambda}}_j \widehat{V}_{\mathbf{u},j} \widehat{\mathbf{\Lambda}}_j' + \left(\left(\widehat{\gamma}_{j,1}' \left(\widehat{M}_{\mathbf{u}}^{(j)}\right)^{-1}\right) \otimes I_K\right) \widehat{V}_{\mathbf{eu},j} \left(\left(\left(\widehat{M}_{\mathbf{u}}^{(j)}\right)^{-1} \widehat{\gamma}_{j,1}\right) \otimes I_K\right) \qquad \text{(A.16)} \\
&\quad + c_j^{-1} \widehat{\mathbf{\Lambda}}_j \widehat{\Pi}_{\mathbf{ue},j} \left(\left(\left(\widehat{M}_{\mathbf{u}}^{(j)}\right)^{-1} \widehat{\gamma}_{j,1}\right) \otimes I_K\right) + c_j^{-1} \left(\left(\widehat{\gamma}_{j,1}' \left(\widehat{M}_{\mathbf{u}}^{(j)}\right)^{-1}\right) \otimes I_K\right) \widehat{\Pi}_{\mathbf{ue},j}' \widehat{\mathbf{\Lambda}}_j', \\
\widehat{\Sigma}_{\alpha\gamma,j} &= \widehat{\sigma}_{\alpha,j}^2 \widehat{\mathbf{\Lambda}}_j \left[\frac{1}{N} \widehat{\mathbf{B}}_j^{\widehat{P}_j\prime} \widehat{\mathbf{B}}_j^{\widehat{P}_j} - \left(\frac{1}{N} \widehat{\mathbf{B}}_j^{\widehat{P}_j\prime} i_N\right) \left(\frac{1}{N} \widehat{\mathbf{B}}_j^{\widehat{P}_j\prime} i_N\right)'\right]^{-1} \widehat{\mathbf{\Lambda}}_j', \qquad \text{(A.17)}
\end{aligned}$$

In (A.16), recall that $\widehat{\mathbf{X}}_j = \left(\iota_N, \widehat{\mathbf{B}}_j^{\widehat{P}_j}\right)$. Also, we have used

$$\begin{aligned}
\widehat{M}_{\mathbf{u}}^{(j)} &= \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{u}}_{j,t} \widehat{\mathbf{u}}_{j,t}' \widehat{d}_{j,t}, \qquad \text{(A.18)} \\
\widehat{V}_{\mathbf{u},j} &= \widehat{m}_{\mathbf{u},0}^{(j)} + \sum_{k=1}^{h_m} \left(1 - \frac{k}{h_m + 1}\right) \left(\widehat{m}_{\mathbf{u},k}^{(j)} + \widehat{m}_{\mathbf{u},k}^{(j)\prime}\right), \qquad \text{(A.19)} \\
\widehat{V}_{\mathbf{eu},j} &= \widehat{m}_{\mathbf{eu},0}^{(j)} + \sum_{k=1}^{h_m} \left(1 - \frac{k}{h_m + 1}\right) \left(\widehat{m}_{\mathbf{eu},k}^{(j)} + \widehat{m}_{\mathbf{eu},k}^{(j)\prime}\right), \qquad \text{(A.20)} \\
\widehat{\Pi}_{\mathbf{ue},j}' &= \frac{1}{T} \sum_{t=1}^{T} \left(vec\left(\widehat{\mathbf{e}}_t \widehat{\mathbf{u}}_{j,t}'\right)\right) \widehat{\mathbf{u}}_{j,t}' \widehat{d}_{j,t} + \sum_{k=1}^{h_m} \left(1 - \frac{k}{h_m + 1}\right) \left(\widehat{m}_{\mathbf{\Pi},k}^{(j)} + \widetilde{m}_{\mathbf{\Pi},k}^{(j)}\right), \qquad \text{(A.21)}
\end{aligned}$$

with

$$\widehat{m}_{\mathbf{u},k}^{(j)} = \frac{1}{T}\sum_{t=k+1}^{T}\left(\widehat{\mathbf{u}}_{j,t}\widehat{d}_{j,t}\right)\left(\widehat{\mathbf{u}}_{j,t-k}\widehat{d}_{j,t-k}\right)',$$

$$\widehat{m}_{\mathbf{eu},k}^{(j)} = \frac{1}{T}\sum_{t=k+1}^{T}\left(vec\left(\widehat{\mathbf{e}}_{t}\widehat{\mathbf{u}}_{j,t}'\right)\widehat{d}_{j,t}\right)\left(vec\left(\widehat{\mathbf{e}}_{t-k}\widehat{\mathbf{u}}_{j,t-k}'\right)\widehat{d}_{j,t-k}\right)',$$

$$\widehat{m}_{\mathbf{\Pi},k}^{(j)} = \frac{1}{T}\sum_{t=k+1}^{T}\left(vec\left(\widehat{\mathbf{e}}_{t-k}\widehat{\mathbf{u}}_{j,t-k}'\right)\widehat{d}_{j,t-k}\right)\left(\widehat{\mathbf{u}}_{j,t}\widehat{d}_{j,t}\right)',$$

$$\widetilde{m}_{\mathbf{\Pi},k}^{(j)} = \frac{1}{T}\sum_{t=k+1}^{T}\left(vec\left(\widehat{\mathbf{e}}_{t}\widehat{\mathbf{u}}_{j,t}'\right)\widehat{d}_{j,t}\right)\left(\widehat{\mathbf{u}}_{j,t-k}\widehat{d}_{j,t-k}\right)',$$

for $k = 0, 1, ...$ In (A.19)-(A.20), we use $h_m = O\left(T^{1/3}\right)$, and $\widehat{\mathbf{e}}_t = \mathbf{g}_t - \left(\widehat{\mathbf{a}} + \widehat{\mathbf{\Lambda}}\widehat{\mathbf{u}}_t\right)$. Finally, in order to compute $\widehat{\sigma}_{\alpha,j}^2$, let

$$\widehat{e}_{j,i,t} = R_{i,t}\widehat{d}_{j,t} - \left[\widehat{\gamma}_{j,0} + \widehat{\boldsymbol{\beta}}_{j,i}'\left(\widehat{\boldsymbol{\gamma}}_{j,1} + \widehat{\mathbf{u}}_{j,t}\right)\right]\widehat{d}_{j,t},$$

and define $\widehat{e}_{j,i} = \widehat{T}_j^{-1}\sum_{t=1}^{T}\widehat{e}_{j,i,t}$ and

$$\widehat{\sigma}_{\alpha,j}^2 = \frac{1}{N}\sum_{i=1}^{N}\widehat{e}_{j,i}^2.$$

All the estimators defined above are consistent - in order to save space, we omit the proofs, which is based on standard, if tedious, calculations based on using the results reported in the above. The only result that does not follow immediately is the consistency of $\widehat{\sigma}_{\alpha,j}^2$, which we state as a lemma.

Lemma A.2. *We assume that Assumptions 1-10 hold. Then, as* $\min(N, T_j) \to \infty$ *for* $j = \mathcal{D}, \mathcal{S}$ *under (24) and (25), it holds that* $\widehat{\sigma}_{\alpha,j}^2 = \sigma_{\alpha,j}^2 + o_{P^*}(1)$*, for almost all realizations of* $\{R_{i,t}, 1 \le i \le N, 1 \le t \le T\}$.

A.4.3 Estimation of $\mathbf{\Gamma}_j$ and $\mathbf{\Lambda}_j$: asymptotics

Define the $P_j^0 \times T$ matrices of regime-specific factor innovations $\mathbf{U}_j(\theta) = [\mathbf{u}_{j,1}(\theta), \ldots, \mathbf{u}_{j,T}(\theta)]$, for $j = \mathcal{D}, \mathcal{S}$, such that $\mathbf{U}_{\mathcal{D}}(\theta) + \mathbf{U}_{\mathcal{S}}(\theta) = (\mathbf{u}_1, \ldots, \mathbf{u}_T) = \mathbf{U}$.

As is well known, $\mathbf{\Gamma}_j$ (and $\mathbf{\Lambda}_j$) cannot be estimated consistently, but only up to a rotation.

Define the $P_j^0 \times \widehat{P}_j$ rotation matrix

$$\widehat{\mathbf{H}}_{jj}^{\widehat{P}_j}(\theta) = \frac{\mathbf{U}_j(\theta^0)\,\mathbf{U}_j(\theta)'}{T}\frac{\mathbf{B}_j^{0\prime}\widehat{\mathbf{B}}_j(\theta)}{N}\widehat{\mathbf{V}}_j(\theta)^{-1}, \tag{A.22}$$

for $j = \mathcal{D}, \mathcal{S}$, where $\widehat{\mathbf{V}}_j(\theta)$ is the $\widehat{P}_j \times \widehat{P}_j$ diagonal matrix of the first $\widehat{P}_j$ largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{j,\widetilde{\mathbf{R}}}(\theta)$ defined in (12). Define also

$$\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}\widehat{P}_j}(\theta) = \begin{bmatrix} 1 & \mathbf{0}'_{\widehat{P}_j} \\ \mathbf{0}_{\widehat{P}_j} & \widehat{\mathbf{H}}_{jj}^{\widehat{P}_j}(\theta) \end{bmatrix}, \tag{A.23}$$

for $j = \mathcal{D}, \mathcal{S}$.

For short, we define $\widehat{\mathbf{H}}_{jj}(\theta)$ as $\widehat{\mathbf{H}}_{jj}^{\widehat{P}_j}(\theta)$ when calculated with P_j^0 common factors, and $\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}(\theta)$ similarly. Finally, we define $\widehat{\mathbf{H}}_{jj}(\theta^0) = \widehat{\mathbf{H}}_{jj}$ and $\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}(\theta^0) = \widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}$.

Theorem A.4. *We assume that Assumptions 1-8 are satisfied. Then, as* $\min(N, T_j) \to \infty$, *it holds that*

$$\widehat{\boldsymbol{\Gamma}}_j - \left(\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}\left(\theta^0\right)\right)^{-1}\boldsymbol{\Gamma}_j^0 = o_P(1) + o_{P^*}(1), \tag{A.24}$$

$$\widehat{\boldsymbol{\Lambda}}_j - \boldsymbol{\Lambda}_j^0\widehat{\mathbf{H}}_{jj}\left(\theta^0\right) = o_P(1) + o_{P^*}(1), \tag{A.25}$$

for $j = \mathcal{D}, \mathcal{S}$, *for almost all realizations of* $\{R_{i,t}, 1 \le i \le N, 1 \le t \le T\}$.

For $j = \mathcal{D}, \mathcal{S}$, consider the matrices: $M_{\mathbf{H},\mathbf{X}}^{(j)} = \lim_{N\to\infty} N^{-1}\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}\prime}(\theta^0)\,\mathbf{X}_j^{0\prime}\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}(\theta^0)$; $M_{\mathbf{H},\mathbf{X},\mathbf{B}}^{(j)} = \lim_{N\to\infty} N^{-1}\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}\prime}(\theta^0)\,\mathbf{X}_j^{0\prime}\mathbf{B}_j^0$; and $M_{\mathbf{H},\mathbf{u}}^{(j)} = \left(M_{\mathbf{u}}^{(j)}\right)^{-1}\widehat{\mathbf{H}}_{jj}(\theta^0)$. We then define

$$\begin{aligned}
\Sigma_{\Gamma,j} &= c_j^{-2}\left(M_{\mathbf{H},\mathbf{X}}^{(j)}\right)^{-1} M_{\mathbf{H},\mathbf{X},\mathbf{B}}^{(j)} V_{\mathbf{u},j} M_{\mathbf{H},\mathbf{X},\mathbf{B}}^{(j)\prime}\left(M_{\mathbf{H},\mathbf{X}}^{(j)}\right)^{-1}, \\
\Sigma_{\alpha\Gamma,j} &= \sigma_{\alpha,j}^2\left(M_{\mathbf{H},\mathbf{X}}^{(j)}\right)^{-1}, \\
\Sigma_{\Lambda,j} &= \left(M_{\mathbf{H},\mathbf{u}}^{(j)\prime} \otimes I_K\right) V_{\mathbf{eu},j}\left(M_{\mathbf{H},\mathbf{u}}^{(j)} \otimes I_K\right).
\end{aligned}$$

Theorem A.5. *We assume that Assumptions 1-8 are satisfied, and that* $\min(N, T_j) \to \infty$*, with (24) and (25). Then, under Assumptions 7 and 9, it holds that*

$$\left(\frac{1}{T}\Sigma_{\Gamma,j} + \frac{1}{N}\Sigma_{\alpha\Gamma,j}\right)^{-1/2}\left(\widehat{\mathbf{\Gamma}}_j - \left(\widehat{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\left(\theta^0\right)\right)^{-1}\mathbf{\Gamma}^0_j\right) \stackrel{D^*}{\to} N\left(0, I_{P_j+1}\right). \tag{A.26}$$

Under Assumption 9(ii), if $\frac{T^{1/2}}{N} \to 0$*, it holds that*

$$\left(\frac{1}{T}\Sigma_{\Lambda,j}\right)^{-1/2} vec\left(\widehat{\mathbf{\Lambda}}_j - \mathbf{\Lambda}^0_j\widehat{\mathbf{H}}_{jj}\left(\theta^0\right)\right) \stackrel{D^*}{\to} N\left(0, I_{P_jK}\right), \tag{A.27}$$

for $j = \mathcal{D}, \mathcal{S}$*, for almost all realizations of* $\{R_{i,t}, 1 \le i \le N, 1 \le t \le T\}$*.*

In order to use the results in Theorem A.5, we propose the following estimators

$$\begin{aligned}
\widehat{\Sigma}_{\Gamma,j} =& c_j^{-2}\left(\frac{\widehat{\mathbf{X}}_j'\widehat{\mathbf{X}}_j}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{X}}_j'\widehat{\mathbf{B}}_j^{\widehat{P}_j}}{N}\right)\widehat{V}_{\mathbf{u},j}\left(\frac{\widehat{\mathbf{X}}_j'\widehat{\mathbf{B}}_j^{\widehat{P}_j}}{N}\right)'\left(\frac{\widehat{\mathbf{X}}_j'\widehat{\mathbf{X}}_j}{N}\right)^{-1}, \\
\widehat{\Sigma}_{\alpha\Gamma,j} =& \widehat{\sigma}^2_{\alpha,j}\left(\frac{\widehat{\mathbf{X}}_j'\widehat{\mathbf{X}}_j}{N}\right)^{-1}, \\
\widehat{\Sigma}_{\Lambda,j} =& \left(\left(\widehat{M}^{(j)}_{\mathbf{u}}\right)^{-1} \otimes I_K\right)\widehat{V}_{\mathbf{eu},j}\left(\left(\widehat{M}^{(j)}_{\mathbf{u}}\right)^{-1} \otimes I_K\right),
\end{aligned}$$

whose consistency can be shown along the same lines as for the other estimators.

A.4.4 Asymptotics for the zero beta rate

The following lemma (which can be compared with Theorem 3 in Giglio and Xiu, 2021) characterizes the distribution of $\widehat{\gamma}_{j,0}$ and the test for (**??**).

Theorem A.6. *We assume that Assumptions 1-10 hold, and that* $E\left(\alpha_{i,\mathcal{D}}\alpha_{k,\mathcal{S}}\right) = 0$ *for all* $1 \le i, k \le N$*. Then, as* $\min(N, T_j) \to \infty$ *with* $\frac{N}{T_j^2} \to 0$*, it holds that*

$$\frac{\widehat{\gamma}_{j,0} - \gamma_{j,0}}{\left(\sigma^2_{\alpha,j}\left(i_N'\mathbf{M}^0_{B,j}i_N\right)^{-1}\right)^{1/2}} \stackrel{D^*}{\to} N\left(0, 1\right), \tag{A.28}$$

for $j = \mathcal{D}, \mathcal{S}$, *and*

$$\frac{(\widehat{\pi}_{\mathcal{D}}\widehat{\gamma}_{\mathcal{D},0} + \widehat{\pi}_{\mathcal{S}}\widehat{\gamma}_{\mathcal{S},0}) - (\pi_{\mathcal{D}}\gamma_{\mathcal{D},0} + \pi_{\mathcal{S}}\gamma_{\mathcal{S},0})}{\left(\pi_{\mathcal{D}}^2\sigma_{\alpha,\mathcal{D}}^2\left(i_N'\mathbf{M}_{B,\mathcal{D}}^0 i_N\right)^{-1} + \pi_{\mathcal{S}}^2\sigma_{\alpha,\mathcal{S}}^2\left(i_N'\mathbf{M}_{B,\mathcal{S}}^0 i_N\right)^{-1}\right)^{1/2}} \stackrel{D^*}{\to} N(0,1), \tag{A.29}$$

for almost all realizations of $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$, *where* $\mathbf{M}_{B,j}^0 = \mathbf{B}_j^0\left(\mathbf{B}_j^{0\prime}\mathbf{B}_j^0\right)^{-1}\mathbf{B}_j^{0\prime}$ *for* $j = \mathcal{D}, \mathcal{S}$.

A.4.5 Testing for zero weighted intercept

We consider the following result, which can be used e.g. to test for

$$H_0 : \pi_{\mathcal{D}}\gamma_{\mathcal{D},g} + \pi_{\mathcal{S}}\gamma_{\mathcal{S},g} = 0.$$

The result is a direct application of Assumption 11, and we therefore report it without proof. Define

$$\widetilde{\Sigma}_{\alpha,\gamma} = \pi_{\mathcal{D}}^2\sigma_{\alpha,\mathcal{D}}^2\Lambda_{\mathcal{D}}\left(\frac{B_{\mathcal{D}}'M_{iN}B_{\mathcal{D}}}{N}\right)^{-1}\Lambda_{\mathcal{D}}' + \pi_{\mathcal{S}}^2\sigma_{\alpha,\mathcal{S}}^2\Lambda_{\mathcal{S}}\left(\frac{B_{\mathcal{S}}'M_{iN}B_{\mathcal{S}}}{N}\right)^{-1}\Lambda_{\mathcal{S}}'$$

and

$$\begin{aligned}
\widetilde{\Sigma} \;=\; & \pi_{\mathcal{D}}^2\Sigma_{\gamma,\mathcal{D}} + \pi_{\mathcal{S}}^2\Sigma_{\gamma,\mathcal{S}} + \pi_{\mathcal{D}}\pi_{\mathcal{S}}c_{\mathcal{D}}^{-1}c_{\mathcal{S}}^{-1}\left(\Lambda_{\mathcal{D}}\Pi_{u,\mathcal{D};u,\mathcal{S}}'\Lambda_{\mathcal{D}}' + \Lambda_{\mathcal{S}}\Pi_{u,\mathcal{D};u,\mathcal{S}}\Lambda_{\mathcal{S}}'\right) \\
& + \pi_{\mathcal{D}}\pi_{\mathcal{S}}\left[\left(\gamma_{\mathcal{D},1}' \cdot M_{Hu}^{(\mathcal{D})\prime}\right) \otimes I_K\right]\Pi_{ue,\mathcal{D}\mathcal{S}}'\left[\left(\gamma_{\mathcal{S},1} \cdot M_{Hu}^{(\mathcal{D})}\right) \otimes I_K\right] \\
& + \left(\pi_{\mathcal{D}}\pi_{\mathcal{S}}\left[\left(\gamma_{\mathcal{D},1}' \cdot M_{Hu}^{(\mathcal{D})\prime}\right) \otimes I_K\right]\Pi_{ue,\mathcal{D}\mathcal{S}}'\left[\left(\gamma_{\mathcal{S},1} \cdot M_{Hu}^{(\mathcal{D})}\right) \otimes I_K\right]\right)' \\
& + \pi_{\mathcal{D}}\pi_{\mathcal{S}}c_{\mathcal{D}}^{-1}\Lambda_{\mathcal{D}}\Pi_{ue,\mathcal{S};u,\mathcal{D}}'\left[\left(\gamma_{\mathcal{S},1} \cdot M_{Hu}^{(\mathcal{D})}\right) \otimes I_K\right] + \left(\pi_{\mathcal{D}}\pi_{\mathcal{S}}c_{\mathcal{D}}^{-1}\Lambda_{\mathcal{D}}\Pi_{ue,\mathcal{S};u,\mathcal{D}}'\left[\left(\gamma_{\mathcal{S},1} \cdot M_{Hu}^{(\mathcal{D})}\right) \otimes I_K\right]\right)' \\
& + \pi_{\mathcal{D}}\pi_{\mathcal{S}}c_{\mathcal{S}}^{-1}\left[\left(\gamma_{\mathcal{D},1} \cdot M_{Hu}^{(\mathcal{S})}\right) \otimes I_K\right]'\Pi_{ue,\mathcal{D};u,\mathcal{S}}'\Lambda_{\mathcal{S}}' + \left(\pi_{\mathcal{D}}\pi_{\mathcal{S}}c_{\mathcal{S}}^{-1}\left[\left(\gamma_{\mathcal{D},1} \cdot M_{Hu}^{(\mathcal{S})}\right) \otimes I_K\right]'\Pi_{ue,\mathcal{D};u,\mathcal{S}}'\Lambda_{\mathcal{S}}'\right)'.
\end{aligned}$$

Theorem A.7. *We assume that Assumptions 1-11 hold, and that* $E(\alpha_{i,\mathcal{D}}\alpha_{k,\mathcal{S}}) = 0$ *for all* $1 \leq i, k \leq N$. *Then, as* $\min(N, T_j) \to \infty$ *with* $\frac{N}{T_j^2} \to 0$, *it holds that*

$$\left(\frac{1}{N}\widetilde{\Sigma}_{\alpha,\gamma} + \frac{1}{T}\widetilde{\Sigma}\right)^{-1/2}\left[(\widehat{\pi}_{\mathcal{D}}\widehat{\gamma}_{\mathcal{D},g} + \widehat{\pi}_{\mathcal{S}}\widehat{\gamma}_{\mathcal{S},g}) - (\pi_{\mathcal{D}}\gamma_{\mathcal{D},g} + \pi_{\mathcal{S}}\gamma_{\mathcal{S},g})\right] \stackrel{d}{\to} N(0, I_K).$$

The long-run covariance matrices that appear in the theorem can be estimated as

$$
\begin{aligned}
\widehat{\Pi}'_{u,\mathcal{D};u,S} &= \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_{\mathcal{D},t}\widehat{u}'_{\mathcal{S},t}\widehat{d}_{\mathcal{D},t}\widehat{d}_{\mathcal{S},t}\right)+\sum_{k=1}^{h_m}\left(1-\frac{k}{h_m+1}\right)\left(\widehat{m}^{(\mathcal{DS})}_{\Pi,k}+\widetilde{m}^{(\mathcal{DS})}_{\Pi,k}\right)\\
&= \sum_{k=1}^{h_m}\left(1-\frac{k}{h_m+1}\right)\left(\widehat{m}^{(\mathcal{DS})}_{\Pi,k}+\widetilde{m}^{(\mathcal{DS})}_{\Pi,k}\right),
\end{aligned}
$$

with

$$
\begin{aligned}
\widehat{m}^{(\mathcal{DS})}_{\Pi,k} &= \frac{1}{T}\sum_{t=k+1}^{T}\widehat{u}_{\mathcal{S},t}\widehat{u}'_{\mathcal{S},t-k}\widehat{d}_{\mathcal{S},t}\widehat{d}_{\mathcal{S},t-k},\\
\widetilde{m}^{(\mathcal{DS})}_{\Pi,k} &= \frac{1}{T}\sum_{t=k+1}^{T}\widehat{u}_{\mathcal{S},t-k}\widehat{u}'_{\mathcal{S},t}\widehat{d}_{\mathcal{S},t-k}\widehat{d}_{\mathcal{S},t},
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{\Pi}'_{ue,\mathcal{S};u,\mathcal{D}} &= \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{\mathcal{D},t}\widehat{d}_{\mathcal{D},t}vec\left(\widehat{e}_t\widehat{u}_{\mathcal{S},t}\widehat{d}_{\mathcal{S},t}\right)'+\sum_{k=1}^{h_m}\left(1-\frac{k}{h_m+1}\right)\left(\widehat{m}^{(\mathcal{DS})}_{\Pi,ue,\mathcal{SD},k}+\widetilde{m}^{(\mathcal{DS})}_{\Pi,ue,\mathcal{SD},k}\right)\\
&= \sum_{k=1}^{h_m}\left(1-\frac{k}{h_m+1}\right)\left(\widehat{m}^{(\mathcal{DS})}_{\Pi,ue,\mathcal{SD},k}+\widetilde{m}^{(\mathcal{DS})}_{\Pi,ue,\mathcal{SD},k}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{m}^{(\mathcal{DS})}_{\Pi,ue,\mathcal{SD},k} &= \frac{1}{T}\sum_{t=k+1}^{T}\widehat{u}_{\mathcal{D},t}vec\left(\widehat{e}_{t-k}\widehat{u}_{\mathcal{S},t-k}\right)'\widehat{d}_{\mathcal{D},t}\widehat{d}_{\mathcal{S},t-k},\\
\widetilde{m}^{(\mathcal{DS})}_{\Pi,ue,\mathcal{SD},k} &= \frac{1}{T}\sum_{t=k+1}^{T}\widehat{u}_{\mathcal{D},t-k}vec\left(\widehat{e}_{t}\widehat{u}_{\mathcal{S},t}\right)'\widehat{d}_{\mathcal{D},t-k}\widehat{d}_{\mathcal{S},t},
\end{aligned}
$$

etc.

B Simulations

We evaluate the performance of *(i)* the Least Squares estimator of θ, and *(ii)* the sequential procedure to determine the number of common factors.[i]

Data are generated according to (5), that is

$$R_{i,t} = \mathbb{I}\left(\mathcal{D}_t\right)\left(\gamma_{\mathcal{D},0} + \alpha_{\mathcal{D},i} + \boldsymbol{\beta}'_{\mathcal{D},i}\boldsymbol{\gamma}_{\mathcal{D},1} + \boldsymbol{\beta}'_{\mathcal{D},i}\mathbf{u}_{\mathcal{D},t}\right) + \mathbb{I}\left(\mathcal{S}_t\right)\left(\gamma_{\mathcal{S},0} + \alpha_{\mathcal{S},i} + \boldsymbol{\beta}'_{\mathcal{S},i}\boldsymbol{\gamma}_{\mathcal{S},1} + \boldsymbol{\beta}'_{\mathcal{S},i}\mathbf{u}_{\mathcal{S},t}\right) + \sqrt{0.5}\epsilon_{i,t}. \tag{B.1}$$

In this data generating process (DGP), we generate all variables as having zero mean for simplicity. Where possible and relevant, we calibrate the values of parameters to values that arise from empirical analysis. In particular, as far as the downside risk is concerned, we set $\theta^0 = -0.03$, which is the value set in Farago and Tédongap (2018) and Lettau, Maggiori, and Weber (2014). Similarly, we generate z_t as $i.i.d.N\left(\mu_z, \sigma_z^2\right)$. Based on our data, we set $\sigma_z = 0.044$; also, in Section 6 we found that $P\left(z_t \leq \theta^0\right) \simeq 0.15$, which corresponds to a value $\mu_z = -0.015$. The rest of the DGP is specified as follows. For $j = \mathcal{D}, \mathcal{S}$, we generate $\mathbf{u}_{j,t}$ as $i.i.d.$ $N(0,1)$ for $1 \leq t \leq T$; similarly, we generate $\beta_{i,j}$ as $i.i.d.$ $N(0,1)$ for $1 \leq i \leq N$; finally, $\alpha_{j,i}$ is also $i.i.d.$ $N(0,1)$ for $1 \leq i \leq N$.[ii] The number of replications is set to $1,000$.

The specifications of our estimators and tests are as follows. In the estimation of θ, we use a grid search on the interval $[-0.08, -0.04]$, with steps of size 0.001. In the estimation of P_j, we use the algorithm described in Section 4.2, with: $M = N$, $\varepsilon = 0.01$ in (19), $n_S = 4$, and the nominal level of the individual tests set to $\alpha = \frac{0.05}{T}$.[iii]

---

[i] As far as the latter is concerned, the results complement the ones in Trapani (2018), which considers the case of a linear model.

[ii] We note that we have carried out some limited experiments where, e.g., a vanishing proportion of the $\alpha_{j,i}$s had nonzero mean, but results did not change. As far as the idiosyncratic innovation $\epsilon_{i,t}$ is concerned, we only consider the case of no cross sectional independence in our simulations; thus, we generate $\epsilon_{i,t}$ as $i.i.d.$ $N(0, 0.5)$.

[iii] We point out that altering the choice of M within a reasonable range does not have any impact on the final results. Indeed, we tried $M = N/2$ and $M = 2N$, without noting any changes. This robustness is in line with the simulations in Trapani (2018).

Similarly, we have set ϱ_j according to (19), with $\varepsilon = 0.01$; based on Theorem A.3, ε does not need to be too large, since its purpose is to smooth away a slowly varying sequence. Even in this case, as the theory prescribes, the impact of ϱ_j is minimal; we have tried, as robustness check, $\varepsilon = 0.05$ and $\varepsilon = 0.1$, and results virtually do not change.

Results are reported for $(N, T) = (57, 316)$ and $(130, 666)$, which correspond to the sample sizes in our empirical applications. Further simulations, which reinforce the findings in this section, are available upon request. Similarly, we consider various number of common factors which mimic the findings in our empirical applications: $P^0_{\mathcal{D}} \in \{1, 2\}$ and $P^0_{\mathcal{S}} \in \{1, 2, 3, 4, 5, 6\}$.[iv]

Consider the estimation of θ, whose performance, in terms of bias and RMSE, we report in Table B.1. Results show that both indicators decrease with T, as expected, and also with N, which is due to a second order improvement in the estimation of the factor structure. We note that $\widehat{\theta}$ tends to be biased towards zero, with this vanishing as $(N, T) \to \infty$.

As far as estimation of the number of common factors is concerned, the results are in Table B.2, where we have reported two indicators: *(i)* the average estimated number of factors

$$\overline{\widehat{P}_j} = \frac{1}{1000} \sum_{r=1}^{1000} \widehat{P}_{j,r},$$

$j = \mathcal{D}, \mathcal{S}$, and *(ii)* the percentage of times that $\widehat{P}_{j,r} \neq P^0_j$, $j = \mathcal{D}, \mathcal{S}$. Results are very satisfactory, with the exception of the case $(P^0_{\mathcal{D}}, P^0_{\mathcal{S}}) = (2, 1)$, when $(N, T) = (57, 316)$. This case is however unlikely to occur, since our experience indicates that the number of common factors in the downside regime tends to be lower than the number of common factors in the other regime. In all other cases, the estimated numbers of factors is always close to the true number of factors, and the simulation results provide comfort that the number of common factors is precisely identified by our procedure even in moderately-sized cross-sections of asset returns.

[iv]Therefore we do not consider the case where there is no factor at all (i.e., either $P^0_{\mathcal{D}} = 0$ or $P^0_{\mathcal{S}} = 0$, or both), requiring that at least one factor should always be present in asset returns, which seems reasonable (see Lettau and Pelger, 2020 for a related discussion).

Table B.1: Bias and RMSE for $\widehat{\theta}$

Panel A: $\theta = -0.03$

Bias

$(N,T) = (57,316)$			$(N,T) = (130,666)$		
$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2	$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2
1	0.001	0.000	1	0.000	0.000
2	0.000	0.000	2	0.000	0.000
3	0.000	0.000	3	0.000	0.000
4	0.000	0.000	4	0.000	0.000
5	0.000	0.000	5	0.000	0.000
6	0.000	0.000	6	0.000	0.000

RMSE

$(N,T) = (57,316)$			$(N,T) = (130,666)$		
$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2	$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2
1	0.003	0.002	1	0.003	0.000
2	0.000	0.000	2	0.000	0.000
3	0.000	0.000	3	0.000	0.000
4	0.000	0.000	4	0.000	0.000
5	0.000	0.000	5	0.000	0.000
6	0.000	0.000	6	0.000	0.000

Panel B: $\theta = -0.06$

Bias

$(N,T) = (57,316)$			$(N,T) = (130,666)$		
	1	2		1	2
1	0.008	0.006	1	0.014	0.010
2	0.005	0.003	2	0.004	0.001
3	0.002	0.000	3	0.000	0.000
4	0.000	0.001	4	0.000	0.000
5	0.000	0.000	5	0.000	0.000
6	0.000	0.000	6	0.000	0.000

RMSE

$(N,T) = (57,316)$			$(N,T) = (130,666)$		
	1	2		1	2
1	0.009	0.008	1	0.014	0.012
2	0.007	0.005	2	0.007	0.004
3	0.004	0.002	3	0.001	0.000
4	0.002	0.001	4	0.000	0.000
5	0.001	0.000	5	0.000	0.000
6	0.001	0.001	6	0.000	0.000

The table contains bias and RMSE for $\widehat{\theta}$.

Table B.2: Number of estimated common factors across regimes

Mean					
	$(N,T) = (57,316)$			$(N,T) = (130,666)$	
$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2	$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2
1	0.996, 0.999	1.000, 1.599	1	0.998, 0.996	0.996, 1.984
2	1.997, 0.999	1.992, 1.996	2	1.980, 0.996	1.990, 1.984
3	3.000, 0.998	2.985, 1.951	3	2.976, 1.000	2.950, 1.980
4	3.996.0.999	3.996, 1.980	4	3.966, 0.992	3.958, 1.980
5	4.978, 0.999	4.985, 1.992	5	4.946, 0.994	4.924, 1.984
6	5.975, 1.000	5.929, 1.883	6	5.924, 0.994	5.912, 1.980

Missed true value (%)					
	$(N,T) = (57,316)$			$(N,T) = (130,666)$	
$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2	$P_{\mathcal{S}} \backslash P_{\mathcal{D}}$	1	2
1	0.400, 0.100	0.000, 40.00	1	0.200, 0.800	0.400, 1.000
2	0.200, 0.100	0.200, 0.700	2	1.400, 0.400	0.600, 1.000
3	0.000, 0.200	0.700, 4.800	3	1.000, 0.000	2.000, 1.400
4	0.200, 0.100	0.200, 1.800	4	1.200, 0.800	2.000, 1.600
5	0.700, 0.100	0.500, 0.700	5	2.400, 0.600	2.000, 1.000
6	1.000, 0.000	4.400, 11.50	6	2.000, 0.600	2.200, 1.200

The table contains the average numbers of estimated factors and the percentage of times the estimator is wrong.

C Technical Lemmas

We begin with a Borel-Cantelli type result, whose proof can be found in Barigozzi and Trapani (2022).

Lemma C.1. *Consider a multi-index partial sums process* $U_{i_1,\ldots,i_h}$, *with* $1 \leq i_1 \leq S_1$, $1 \leq i_2 \leq S_2$, *etc... Assume that*

$$\sum_{S_1} \cdot \cdot \sum_{S_h} \frac{1}{S_1 \cdot \ldots \cdot S_h} P\left(\max_{1 \leq i_1 \leq S_1, \ldots, 1 \leq i_h \leq S_h} |U_{i_1,\ldots,i_h}| > \epsilon L_{S_1,\ldots,S_h} \right) < \infty, \tag{C.2}$$

for some $\epsilon > 0$ *and a sequence* $L_{S_1,\ldots,S_h}$ *defined as*

$$L_{S_1,\ldots,S_h} = S_1^{d_1} \cdot \ldots \cdot S_h^{d_h} l_1 \left(S_1\right) \cdot \ldots l_h \left(S_h\right),$$

where d_1, d_2, *etc. are non-negative numbers and* $l_1(\cdot)$, $l_2(\cdot)$, *etc. are slowly varying functions in the sense of Karamata. Then it holds that*

$$\lim \sup_{(S_1,\ldots,S_h) \to \infty} \frac{|U_{S_1,\ldots,S_h}|}{L_{S_1,\ldots,S_h}} = 0 \ a.s. \tag{C.3}$$

Henceforth, we will extensively use the following notation:

$$\overline{v}_{N,T}(\epsilon) = (\ln N \ln T)^{2\frac{3+r+\epsilon}{r}}, \tag{C.4}$$

where $\epsilon > 0$ and r is defined in Assumption 3*(i)*. Recall (11)

$$S\left(\mathbf{B}^P, \mathbf{U}^P, \theta\right) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\widetilde{R}_{i,t} - \left(\beta_{\mathcal{D},i}^{P_{\mathcal{D}}\prime} \mathbf{u}_{\mathcal{D},t}^{P_{\mathcal{D}}} d_{\mathcal{D},t}(\theta) + \beta_{\mathcal{S},i}^{P_{\mathcal{S}}\prime} \mathbf{u}_{\mathcal{S},t}^{P_{\mathcal{S}}} d_{\mathcal{S},t}(\theta) \right) \right]^2,$$

which is the least squares loss function calculated at the triplet $\left(\mathbf{B}^P, \mathbf{U}^P, \theta\right)$; similarly, we define

$$S_{\mathbf{UB}}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\widetilde{R}_{i,t} - \left(\beta_{\mathcal{D},i}^{P_{\mathcal{D}}^0\prime} \mathbf{u}_{\mathcal{D},t}^{P_{\mathcal{D}}^0} d_{\mathcal{D},t}(\theta) + \beta_{\mathcal{S},i}^{P_{\mathcal{S}}^0\prime} \mathbf{u}_{\mathcal{S},t}^{P_{\mathcal{S}}^0} d_{\mathcal{S},t}(\theta) \right) \right]^2, \tag{C.5}$$

which denotes the concentrated version of $S\left(\mathbf{B}^{P},\mathbf{U}^{P},\theta\right)$ when using the correct number of factors $P_{\mathcal{D}}^{0}$ and $P_{\mathcal{S}}^{0}$. Finally, when using $P_{\mathcal{D}}$ and $P_{\mathcal{S}}$ factors, with $P_{\mathcal{D}}^{0}\leq P_{\mathcal{D}}\leq P_{\max}$ and $P_{\mathcal{S}}^{0}\leq P_{\mathcal{S}}\leq P_{\max}$, we write

$$S_{\mathbf{uB}}^{(P)}\left(\theta\right)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\widetilde{R}_{i,t}-\left(\widehat{\beta}_{\mathcal{D},i}^{P_{\mathcal{D}}\prime}\widehat{\mathbf{u}}_{\mathcal{D},t}^{P_{\mathcal{D}}}d_{\mathcal{D},t}\left(\theta\right)+\widehat{\beta}_{\mathcal{S},i}^{P_{\mathcal{S}}\prime}\widehat{\mathbf{u}}_{\mathcal{S},t}^{P_{\mathcal{S}}}d_{\mathcal{S},t}\left(\theta\right)\right)\right]^{2}, \tag{C.6}$$

to define the concentrated version of $S\left(\mathbf{B}^{P},\mathbf{U}^{P},\theta\right)$ in that case. We will also use the $P_{m}\times P_{j}$ projection matrices

$$\widehat{H}_{jm}^{P_{j,m}}\left(\theta\right)=\frac{\mathbf{U}_{m,m}^{0}\left(\theta^{0}\right)\mathbf{U}_{m,j}^{0}\left(\theta\right)^{\prime}}{T}\frac{\mathbf{B}_{m}^{0\prime}\widehat{\mathbf{B}}_{j}^{P_{j}}\left(\theta\right)}{N}\left(\widehat{V}_{j}^{P_{j}}\left(\theta\right)\right)^{-1}, \tag{C.7}$$

where $j,m=\mathcal{D},\mathcal{S}$, $\mathbf{U}_{m,j}^{0}\left(\theta\right)$ is defined as the $P_{m}\times T$ matrix of regime-specific factors, $\mathbf{U}_{m}\left(\theta\right)$, multiplied by a $T\times T$ diagonal matrix whose entries are $d_{j,t}\left(\theta\right)$, $1\leq t\leq T$; finally, $\widehat{V}_{j}^{P_{j}}\left(\theta\right)$ is a $P_{j}\times P_{j}$ matrix containing the largest P_{j} eigenvalues of

$$\frac{1}{NT}\sum_{t=1}^{T}\widetilde{\mathbf{R}}_{t}\widetilde{\mathbf{R}}_{t}^{\prime}d_{j,t}\left(\theta\right).$$

Note that, for $j\neq m$, $\widehat{H}_{jm}^{P_{jm}}\left(\theta^{0}\right)$ reduces to a matrix of zeros. In order to make the notation lighter, the superscript $P_{j,m}$ is set equal to P_{j} when $j=m$.

We now present a few lemmas to show the strong consistency of $\widehat{\theta}$.

Lemma C.2. *We assume that Assumptions 1-4 are satisfied. Then it holds that*

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{j,i}^{P_{j}}\left(\theta^{0}\right)-\widehat{H}_{jj}^{P_{j}}\left(\theta^{0}\right)^{\prime}\beta_{j,i}^{0}\right\|^{2}=o_{a.s.}\left(P_{j}C_{N,T}^{-2}\overline{v}_{N,T}\left(\epsilon\right)\right),$$

for every $\epsilon>0$ *and* $j=\mathcal{D},\mathcal{S}$.

Proof. This follows by adapting the passages in the proof of Theorem 3.2 in Massacci (2017).

Then we have

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{j,i}^{P_j}\left(\theta^0\right)-\widehat{H}_{jj}^{P_j}\left(\theta^0\right)'\beta_{j,i}^0\right\|^2 \qquad\qquad (C.8)\\
&\leq\left[\frac{4}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}\sigma_{jil}^2\left(\theta^0\right)+\frac{1}{N}\left(\frac{1}{N^2}\sum_{l=1}^{N}\sum_{q=1}^{N}\left|\sum_{i=1}^{N}\varkappa_{jil}\left(\theta^0\right)\varkappa_{jiq}\left(\theta^0\right)\right|^2\right)^{1/2}\right.\\
&\left.+\frac{2}{N}\sum_{i=1}^{N}\left(\left\|\beta_{\mathcal{D},i}^0\right\|^2+\left\|\beta_{\mathcal{S},i}^0\right\|^2\right)\frac{1}{NT^2}\sum_{l=1}^{N}\left\|\sum_{t=1}^{T}d_{j,t}\left(\theta^0\right)\mathbf{u}_{j,t}^0e_{l,t}\right\|^2\right]\\
&\times\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{j,i}^{P_j}\left(\theta^0\right)\right\|^2\right)\\
&=A_{N,T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{j,i}^{P_j}\left(\theta^0\right)\right\|^2\right),
\end{aligned}
$$

having defined

$$
\begin{aligned}
\sigma_{jil}\left(\theta^0\right)&=\frac{1}{T}\sum_{t=1}^{T}E\left(d_{j,t}\left(\theta^0\right)e_{i,t}e_{l,t}\right),\\
\varkappa_{jil}\left(\theta^0\right)&=\frac{1}{T}\sum_{t=1}^{T}d_{j,t}\left(\theta^0\right)e_{i,t}e_{l,t}-\sigma_{jil}\left(\theta^0\right).
\end{aligned}
$$

By the proof of Theorem 3.2 in Massacci (2017), it follows immediately that $E\left|A_{N,T}\right|\leq c_0C_{N,T}^{-2}$. Thus, using the maximal inequality for multi-parameter processes in Corollary 4 of it follows that

$$
E\max_{1\leq i\leq N,1\leq t\leq T}\left|A_{i,t}\right|\leq c_0C_{N,T}^{-2}\ln N\ln T.
$$

Lemma C.1 and Markov inequality immediately yield $A_{N,T}=o_{a.s.}\left(C_{N,T}^{-2}\left(\ln N\right)^{2+\epsilon}\left(\ln T\right)^{2+\epsilon}\right)$ for every $\epsilon>0$. Also,

$$
\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{j,i}^{P_j}\left(\theta^0\right)\right\|^2=P_j,
$$

by construction, since $\widehat{\mathbf{B}}_j^{P_j}\left(\theta^0\right)'\widehat{\mathbf{B}}_j^{P_j}\left(\theta^0\right)=NI_{P_j}$. The desired result follows by putting everything together. □

Lemma C.3. *We assume that Assumptions 1-5 are satisfied. Then it holds that*

$$\left|\widehat{\theta}-\theta^{0}\right|=o_{a.s.}\left(1\right),$$

for all $P_j^0 \leq P_j \leq P_{\max}$*,* $j=\mathcal{D},\mathcal{S}$*.*

Proof. The desired result follows upon showing

$$S_{\mathbf{UB}}^{(P)}\left(\theta\right)-S_{\mathbf{UB}}^{(P_0)}\left(\theta^{0}\right)=c_0\left|\theta-\theta^{0}\right|+o_{a.s.}\left(1\right), \tag{C.9}$$

with $c_0>0$, for every $\theta\neq\theta^0$ - note that $S_{\mathbf{UB}}^{(P)}\left(\theta\right)$ is defined in (C.6). We being by considering the identity

$$\begin{aligned}
&S_{\mathbf{UB}}^{(P)}\left(\theta\right)-S_{\mathbf{UB}}^{(P)}\left(\theta^{0}\right)\\
=&S_{\mathbf{UB}}^{(P)}\left(\theta\right)-S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}\left(\theta\right)+\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SD}}^{P_{\mathcal{S},\mathcal{D}}}\left(\theta\right),\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DS}}^{P_{\mathcal{D},\mathcal{S}}}\left(\theta\right)+\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}\left(\theta\right),\theta^{0}\right)\\
&+S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}\left(\theta\right)+\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SD}}^{P_{\mathcal{S},\mathcal{D}}}\left(\theta\right),\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DS}}^{P_{\mathcal{D},\mathcal{S}}}\left(\theta\right)+\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}\left(\theta\right),\theta^{0}\right)\\
&-S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}\left(\theta\right),\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}\left(\theta\right),\theta^{0}\right)+S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}\left(\theta\right),\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}\left(\theta\right),\theta^{0}\right)-S_{\mathbf{UB}}^{(P)}\left(\theta^{0}\right),
\end{aligned}$$

where $S_{\mathbf{B}}^{(P)}$ denotes the loss function $S\left(\mathbf{B}^P,\mathbf{U}^P,\theta\right)$ concentrated at $\mathbf{B}^P$. By construction, it holds that

$$S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}\left(\theta\right),\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}\left(\theta\right),\theta^{0}\right)=S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0},\mathbf{B}_{\mathcal{S}}^{0},\theta^{0}\right).$$

Following the proof of Lemma A.2 and Theorem 3.3 in Massacci (2017), it can be shown by standard arguments that

$$\begin{aligned}
&E\left|S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0},\mathbf{B}_{\mathcal{S}}^{0},\theta^{0}\right)-S_{\mathbf{UB}}^{(P)}\left(\theta^{0}\right)\right|^{2}\leq c_0C_{N,T}^{-2},\\
&E\left|S_{\mathbf{UB}}^{(P)}\left(\theta\right)-S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}\left(\theta\right)+\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SD}}^{P_{\mathcal{S},\mathcal{D}}}\left(\theta\right),\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{DS}}^{P_{\mathcal{D},\mathcal{S}}}\left(\theta\right)+\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}\left(\theta\right),\theta^{0}\right)\right|^{2}\leq c_0C_{N,T}^{-2};
\end{aligned}$$

then, using again Corollary 4 in Moricz (1983), it follows that

$$E \max_{1\leq t\leq T, 1\leq i\leq N} \left| S^{(P)}_{\mathbf{B},it}\left(\mathbf{B}^0_{\mathcal{D}}, \mathbf{B}^0_{\mathcal{S}}, \theta^0\right) - S^{(P)}_{\mathbf{UB},it}\left(\theta^0\right)\right|^2 \leq c_0 C^{-2}_{N,T} (\ln N)(\ln T),$$

and

$$E \max_{1\leq t\leq T, 1\leq i\leq N} \left| S^{(P)}_{\mathbf{UB},it}(\theta) - S^{(P)}_{\mathbf{B},it}\left(\mathbf{B}^0_{\mathcal{D}}\widehat{H}^{P_{\mathcal{D}}}_{\mathcal{DD}}(\theta) + \mathbf{B}^0_{\mathcal{S}}\widehat{H}^{P_{\mathcal{S}},\mathcal{D}}_{\mathcal{SD}}(\theta), \mathbf{B}^0_1\widehat{H}^{P_{\mathcal{D}},\mathcal{S}}_{\mathcal{DS}}(\theta) + \mathbf{B}^0_{\mathcal{S}}\widehat{H}^{P_{\mathcal{S}}}_{\mathcal{SS}}(\theta), \theta^0\right)\right|^2$$
$$\leq c_0 C^{-2}_{N,T} (\ln N)(\ln T),$$

where

$$S^{(P)}_{\mathbf{UB},it}(\theta) = \frac{1}{NT}\sum_{i'=1}^{i}\sum_{t'=1}^{t}\left[\widetilde{R}_{i,t} - \left(\widehat{\beta}'_{\mathcal{D},i}\widehat{\mathbf{u}}_{\mathcal{D},t} d_{\mathcal{D},t}(\theta) + \widehat{\beta}'_{\mathcal{S},i}\widehat{\mathbf{u}}_{\mathcal{S},t} d_{\mathcal{S},t}(\theta)\right)\right]^2,$$

and $S_{\mathbf{B},it}$ defined similarly. Hence, Lemma C.1 yields

$$\begin{aligned} & S^{(P)}_{\mathbf{UB}}(\theta) - S^{(P)}_{\mathbf{B}}\left(\mathbf{B}^0_{\mathcal{D}}\widehat{H}^{P_{\mathcal{D}}}_{\mathcal{DD}}(\theta) + \mathbf{B}^0_{\mathcal{S}}\widehat{H}^{P_{\mathcal{S}},\mathcal{D}}_{\mathcal{SD}}(\theta), \mathbf{B}^0_{\mathcal{D}}\widehat{H}^{P_{\mathcal{D}},\mathcal{S}}_{\mathcal{DS}}(\theta) + \mathbf{B}^0_{\mathcal{S}}\widehat{H}^{P_{\mathcal{S}}}_{\mathcal{SS}}(\theta), \theta^0\right) \\ = \; & o_{a.s.}\left(C^{-1}_{N,T}(\ln N)^{1+\epsilon}(\ln T)^{1+\epsilon}\right). \end{aligned}$$

Similarly, by using Lemma A.3 in Massacci (2017) and the same arguments as above, it follows that there exists a $c_0 > 0$ such that

$$\begin{aligned} & S^{(P)}_{\mathbf{B}}\left(\mathbf{B}^0_{\mathcal{D}}\widehat{H}^{P_{\mathcal{D}}}_{\mathcal{DD}}(\theta) + \mathbf{B}^0_{\mathcal{S}}\widehat{H}^{P_{2},\mathcal{D}}_{\mathcal{SD}}(\theta), \mathbf{B}^0_{\mathcal{D}}\widehat{H}^{P_{\mathcal{D}},\mathcal{S}}_{\mathcal{DS}}(\theta) + \mathbf{B}^0_{\mathcal{S}}\widehat{H}^{P_{\mathcal{S}}}_{\mathcal{SS}}(\theta), \theta^0\right) - S^{(P)}_{\mathbf{B}}\left(\mathbf{B}^0_{\mathcal{D}}, \mathbf{B}^0_{\mathcal{S}}, \theta^0\right) \\ = \; & c_0\left|\theta - \theta^0\right| + o_{a.s.}(1). \end{aligned}$$

Putting everything together, we have

$$S^{(P)}_{\mathbf{UB}}(\theta) - S^{(P)}_{\mathbf{UB}}\left(\theta^0\right) = c_0 + o_{a.s.}(1).$$

Then (C.9) follows if we show that

$$S^{(P)}_{\mathbf{UB}}\left(\theta^0\right) - S^{(P_0)}_{\mathbf{UB}}\left(\theta^0\right) = o_{a.s.}(1). \tag{C.10}$$

It holds that

$$\left|S_{\mathbf{UB}}^{(P)}\left(\theta^{0}\right)-S_{\mathbf{UB}}^{(P_0)}\left(\theta^{0}\right)\right|\leq\left|S_{\mathbf{UB}}^{1}\left(\theta^{0}\right)\right|+\left|S_{\mathbf{UB}}^{2}\left(\theta^{0}\right)\right|,$$

where

$$\begin{aligned}S_{\mathbf{UB}}^{1}\left(\theta^{0}\right)=&\frac{2}{NT}\sum_{t=1}^{T}\Big(d_{\mathcal{D},t}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{D},t}^{0\prime}\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}+}\left(\theta^{0}\right)^{\prime}\left(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)-\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)\right)^{\prime}\\&+d_{\mathcal{S},t}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{S},t}^{0\prime}\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}+}\left(\theta^{0}\right)^{\prime}\left(\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)-\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)\right)^{\prime}\Big)\mathbf{e}_{t},\end{aligned}$$

$$\begin{aligned}S_{\mathbf{UB}}^{2}\left(\theta^{0}\right)=&\frac{1}{NT}\sum_{t=1}^{T}\Big(d_{\mathcal{D},t}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{D},t}^{0\prime}\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}+}\left(\theta^{0}\right)^{\prime}\left(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)-\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)\right)^{\prime}\\&\times\left(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)-\mathbf{B}_{\mathcal{D}}^{0}\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)\right)\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}+}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{D},t}^{0}\\&+d_{\mathcal{S},t}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{S},t}^{0\prime}\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}+}\left(\theta^{0}\right)^{\prime}\left(\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)-\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)\right)^{\prime}\\&+\left(\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)-\mathbf{B}_{\mathcal{S}}^{0}\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)\right)\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}+}\left(\theta^{0}\right)\mathbf{u}_{2,t}^{0}\Big).\end{aligned}$$

We now estimate the order of magnitude of $S_{\mathbf{UB}}^{1}\left(\theta^{0}\right)$ and $S_{\mathbf{UB}}^{2}\left(\theta^{0}\right)$. We have

$$\begin{aligned}\left|S_{\mathbf{UB}}^{1}\left(\theta^{0}\right)\right|\leq&\,2T^{-1/2}P_{\mathcal{D}}^{0}\left\|\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}+}\left(\theta^{0}\right)\right\|\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{\mathcal{D},i}^{P_{\mathcal{D}}}\left(\theta^{0}\right)-\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)^{\prime}\beta_{\mathcal{D},i}^{0}\right\|^{2}\right)^{1/2}\\&\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_{\mathcal{D},t}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{D},t}^{0}e_{i,t}\right\|^{2}\right)^{1/2}\\&+2T^{-1/2}P_{\mathcal{S}}^{0}\left\|\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}+}\left(\theta^{0}\right)\right\|\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{\mathcal{S},i}^{P_{\mathcal{S}}}\left(\theta^{0}\right)-\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)^{\prime}\beta_{\mathcal{S},i}^{0}\right\|^{2}\right)^{1/2}\\&\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_{\mathcal{S},t}\left(\theta^{0}\right)\mathbf{u}_{\mathcal{S},t}^{0}e_{i,t}\right\|^{2}\right)^{1/2}.\end{aligned}$$

By construction, note that

$$\left\|\widehat{H}_{jj}^{P_{j}+}\left(\theta^{0}\right)\right\|\leq c_{0}P_{j}^{1/2}\tag{C.11}$$

for $j=\mathcal{D},\mathcal{S}$. Also, Assumption 4*(v)* entails that, by the same logic as in the previous passages that

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_{j,t}\left(\theta^{0}\right)\mathbf{u}_{j,t}^{0}e_{i,t}\right\|^{2}=o_{a.s.}\left(\overline{v}_{N,T}\left(\epsilon\right)\right),$$

for every $\epsilon>0$ and $j=\mathcal{D},\mathcal{S}$. Thus, using Lemma C.2, we finally have

$$\left|S_{\mathbf{UB}}^{1}\left(\theta^{0}\right)\right|=o_{a.s.}\left(\overline{P}C_{N,T}^{-1}T^{-1/2}\overline{v}_{N,T}\left(\epsilon\right)\right),$$

having set $\overline{P}=\max\left\{P_{\mathcal{D}},P_{\mathcal{S}}\right\}$; this term is therefore $o_{a.s.}\left(1\right)$ by (20). Similarly, it holds that (see Massacci, 2017)

$$\begin{aligned}\left|S_{\mathbf{UB}}^{2}\left(\theta^{0}\right)\right|\leq&\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{u}_{\mathcal{D},t}^{0}\right\|^{2}\left\|\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}+}\left(\theta^{0}\right)\right\|^{2}\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{\mathcal{D},i}^{P_{\mathcal{D}}}\left(\theta^{0}\right)-\widehat{H}_{\mathcal{D}\mathcal{D}}^{P_{\mathcal{D}}}\left(\theta^{0}\right)^{\prime}\beta_{\mathcal{D},i}^{0}\right\|^{2}\\&+\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{u}_{\mathcal{S},t}^{0}\right\|^{2}\left\|\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}+}\left(\theta^{0}\right)\right\|^{2}\frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_{\mathcal{S},i}^{P_{\mathcal{S}}}\left(\theta^{0}\right)-\widehat{H}_{\mathcal{S}\mathcal{S}}^{P_{\mathcal{S}}}\left(\theta^{0}\right)^{\prime}\beta_{\mathcal{S},i}^{0}\right\|^{2}.\end{aligned}$$

Using (C.11) and Lemma C.2, and noting that Assumptions 3*(i)* and 5*(i)* entail

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{u}_{j,t}^{0}\right\|^{2}=O_{a.s.}\left(1\right),$$

for $j=\mathcal{D},\mathcal{S}$, we have

$$\left|S_{\mathbf{UB}}^{2}\left(\theta^{0}\right)\right|=o_{a.s.}\left(\overline{P}^{2}C_{N,T}^{-2}\overline{v}_{N,T}\left(\epsilon\right)\right),$$

which again is $o_{a.s.}\left(1\right)$ by (20). Thus, (C.10) follows. Putting all together, the desired result obtains. □

In the next two lemmas, we define the set

$$B_{N,T}=\left(\frac{\overline{v}_{N,T}\left(\epsilon\right)}{T},c_{B}\right),\tag{C.12}$$

where $0 < c_0 < \infty$ and $\overline{v}_{N,T}$ is defined in (C.4), and the functions

$$\omega^0(\eta,\theta) = \frac{1}{N^\eta T}\sum_{i=1}^N\sum_{t=1}^T \left| d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}\left(\theta^0\right)\right| \left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2, \tag{C.13}$$

$$h^0(\eta,\theta) = \frac{1}{N^\eta T}\sum_{i=1}^N\sum_{t=1}^T d_{\mathcal{S},t}(\theta)\,\delta_i^{0\prime}\underline{\mathbf{u}}_t^0 e_{i,t}, \tag{C.14}$$

where $\underline{\mathbf{u}}_t^0$ and δ_i^0 are defined in (A.1) and (A.2) respectively.

Lemma C.4. *We assume that Assumptions 3 and 5 are satisfied. Then there exist two random variables N_0 and T_0, and a positive, finite constant c_0, such that for $N \geq N_0$ and $T \geq T_0$, it holds that*

$$\inf_{|\theta-\theta^0|\in B_{N,T}} \frac{\omega^0(\eta,\theta)}{|\theta-\theta^0|} \geq c_0.$$

Proof. When possible, we let $\eta = 1$ to avoid a burdensome notation, and we consider the case $\theta - \theta^0 > 0$ only (the opposite case can be shown by symmetry). Define

$$\overline{\omega}_{N,T}^0 = \omega^0(\eta,\theta) - E\omega^0(\eta,\theta),$$

and note, to start with, that, by Assumption 5*(iii)*, there exist two positive, finite constants c_0 and c_1 such that

$$c_0\left|\theta-\theta^0\right| \leq E\omega^0(\eta,\theta) \leq c_1\left|\theta-\theta^0\right|,$$

so that

$$\inf_{|\theta-\theta^0|\in B_{N,T}} \frac{E\omega^0(\eta,\theta)}{|\theta-\theta^0|} \geq c_0. \tag{C.15}$$

We now show that

$$\sup_{|\theta-\theta^0|\in B_{N,T}} \frac{\left|\overline{\omega}_{N,T}^0\right|}{|\theta-\theta^0|} = o_{a.s.}(1). \tag{C.16}$$

Consider the construction $\theta_j = \theta^0 + c^j\overline{v}_{N,T}(\epsilon)/T$, where $1 < c < \infty$, $j = 0, ..., Q$ and Q is defined such that $c^{Q-1}\overline{v}_{N,T}(\epsilon)/T \leq c_B$ and $c^Q\overline{v}_{N,T}(\epsilon)/T > c_B$ where c_B is defined in (C.12), and let

$$\overline{\omega}_{\widetilde{i},\widetilde{t}}^0(\theta_j) = \sum_{i=1}^{\widetilde{i}}\sum_{t=1}^{\widetilde{t}}\left(\left|d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 - E\left(\left|d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2\right)\right).$$

Then, standard calculations based on the modulus of continuity yield

$$\begin{aligned}
&\sup_{|\theta-\theta^0|\in B_{N,T}} \frac{\overline{\omega}_{N,T}^0}{|\theta-\theta^0|} \qquad\qquad \text{(C.17)}\\
\leq\ & 2\sup_{0\leq j\leq Q} \frac{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2\left(\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|-E\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\right)}{|\theta_j-\theta^0|}\\
&+2\sup_{0\leq j\leq Q}\sup_{\theta_j\leq\theta\leq\theta_{j+1}} \frac{\frac{\ln\ln T}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2\left(\left|d_{\mathcal{S},t}\left(\theta\right)-d_{\mathcal{S},t}\left(\theta_j\right)\right|-E\left|d_{\mathcal{S},t}\left(\theta\right)-d_{\mathcal{S},t}\left(\theta_j\right)\right|\right)}{|\theta_j-\theta^0|}\\
=\ & 2\sup_{0\leq j\leq Q}\frac{\overline{\omega}^0\left(\theta_j,\theta^0\right)}{|\theta_j-\theta^0|}+2\sup_{0\leq j\leq Q}\sup_{\theta_j\leq\theta\leq\theta_{j+1}}\frac{\overline{\omega}^0\left(\theta,\theta_j\right)}{|\theta_j-\theta^0|}=I+II.
\end{aligned}$$

We begin with I. Let $\delta_{N,T}=NT$; it holds that

$$\begin{aligned}
&\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left[\max_{1\leq\widetilde{i}\leq N,1\leq\widetilde{t}\leq T}\sup_{0\leq j\leq Q-1}\frac{\overline{\omega}_{\widetilde{i},\widetilde{t}}^0(\theta_j)}{|\theta_j-\theta^0|}>\delta_{N,T}\right]\\
\leq\ & \sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left[\max_{1\leq\widetilde{i}\leq N,1\leq\widetilde{t}\leq T}\frac{\overline{\omega}_{\widetilde{i},\widetilde{t}}^0(\theta_j)}{|\theta_j-\theta^0|}>\delta_{N,T}\right]\\
\leq\ & \sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{\delta_{N,T}^{-2}}{NT}E\left(\max_{1\leq\widetilde{i}\leq N,1\leq\widetilde{t}\leq T}\left|\frac{\overline{\omega}_{\widetilde{i},\widetilde{t}}^0(\theta_j)}{\theta_j-\theta^0}\right|^2\right).
\end{aligned}$$

Also, consider the scalar case for simplicity and note that

$$\begin{aligned}
&E\left(NT\overline{\omega}_{N,T}^0\left(\theta_j\right)\right)^2\\
=\ & E\left[\sum_{i=1}^N\left(\delta_i^0\right)^2\sum_{t=1}^T\left(\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2-E\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right)\right]^2\\
\leq\ & c_0\left(\sum_{i=1}^N\left(\delta_i^0\right)^2\right)^2E\left[\sum_{t=1}^T\left(\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]-E\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]\right)\right]^2\\
\leq\ & c_0N^2E\left[\sum_{t=1}^T\left(\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]-E\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]\right)\right]^2.
\end{aligned}$$

By Assumption 5*(i)*, $\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2$ is a strictly stationary, ρ-mixing sequence with mixing numbers ρ_m such that $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$ - we refer to the proof of Lemma C.11 for details on how to show this. Thus, Lemma 1 in Peligrad (1987) entails

$$
\begin{aligned}
& E\left[\sum_{t=1}^{T}\left(\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]-E\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]\right)\right]^2 \\
\leq\ & c_0 T E\left[\left(\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]-E\left[\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|\left(\underline{\mathbf{u}}_t^0\right)^2\right]\right)\right]^2 \\
\leq\ & c_0 T E\left(\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|^2\left(\underline{\mathbf{u}}_t^0\right)^4\right).
\end{aligned}
$$

Also, note that, by Assumption 5*(ii)*

$$
\begin{aligned}
& E\left(\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|^2\left(\underline{\mathbf{u}}_t^0\right)^4\right) \\
=\ & E\left(\left|d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right|^2 E\left[\left(\underline{\mathbf{u}}_t^0\right)^4|z_t\right]\right)\leq c_0\left|\theta_j-\theta^0\right|.
\end{aligned}
$$

Thus,

$$
E\left(NT\overline{\omega}_{N,T}^0\left(\theta_j\right)\right)^2\leq c_0\left|\theta_j-\theta^0\right|N^2T;
$$

using again Corollary 4 in Moricz (1983), this entails that

$$
E\left(\max_{1\leq\widetilde{i}\leq N,1\leq\widetilde{t}\leq T}\left|\frac{\overline{\omega}_{\widetilde{i},\widetilde{t}}^0(\theta_j)}{\theta_j-\theta^0}\right|^2\right)\leq c_0\left|\theta_j-\theta^0\right|^{-1}N^2T\left(\ln N\right)\left(\ln T\right).
$$

Therefore

$$
\begin{aligned}
& \sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq\widetilde{i}\leq N,1\leq\widetilde{t}\leq T}\sup_{0\leq j\leq Q-1}\frac{\overline{\omega}_{\widetilde{i},\widetilde{t}}^0(\theta_j)}{|\theta_j-\theta^0|}>\delta_{N,T}\right) \\
\leq\ & c_0\left(\sum_{j=0}^{\infty}c^{-j}\right)\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{\delta_{N,T}^{-2}}{NT}\left(\overline{v}_{N,T}\left(\epsilon\right)\right)^{-1}N^2T^2\left(\ln N\right)\left(\ln T\right)<\infty,
\end{aligned}
\tag{C.18}
$$

so that finally this entails

$$
\sup_{0\leq j\leq Q-1}\frac{\overline{\omega}_{NT}^0\left(\theta_j\right)}{|\theta_j-\theta^0|}=o_{a.s.}\left(1\right).
$$

We now turn to II. We now turn to II; as in Hansen (2000), we consider the construction $\theta_{j+1}=\theta_j+\delta_j$, where $\delta_j=T^{-1}c^{j-1}\left(c-1\right)\overline{v}_{NT}$; letting $m_j=T\delta_j=c^{j-1}\left(c-1\right)\overline{v}_{NT}$, we also define

$\theta_k = \theta_j + T^{-1}(k-1)$, $\leq 1 \leq k \leq m_j$. We rely on the inequality

$$\begin{aligned}
&\sup_{\theta_j \leq \theta \leq \theta_{j+1}} \frac{1}{NT} \sum_{i=1}^{\widetilde{i}} \sum_{t=1}^{\widetilde{t}} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 \left(\left|d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)\right| - E\left|d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)\right|\right) \qquad \text{(C.19)} \\
\leq\ & \max_{2 \leq k \leq m_j+1} \frac{1}{NT} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_k, \theta_j) \right| + \max_{1 \leq k \leq m_j} \frac{1}{NT} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_{k+1}, \theta_k) \right| = a + b,
\end{aligned}$$

where

$$\begin{aligned}
d_{2t}(\theta_k, \theta_j) &= \left|d_{\mathcal{S},t}(\theta_k) - d_{\mathcal{S},t}(\theta_j)\right| - E\left|d_{\mathcal{S},t}(\theta_k) - d_{\mathcal{S},t}(\theta_j)\right|, \\
d_{2t}(\theta_{k+1}, \theta_k) &= \left|d_{\mathcal{S},t}(\theta_{k+1}) - d_{\mathcal{S},t}(\theta_k)\right| - E\left|d_{\mathcal{S},t}(\theta_{k+1}) - d_{\mathcal{S},t}(\theta_k)\right|.
\end{aligned}$$

We begin by considering a

$$\begin{aligned}
&\sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{1}{NT} P\left(\max_{1 \leq n' \leq N, 1 \leq t' \leq T} \sup_{0 \leq j \leq Q} \max_{2 \leq k \leq m_j+1} \frac{\frac{1}{NT}\left|\sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_k, \theta_j)\right|}{\left|\theta_j - \theta^0\right|} > \varepsilon \right) \\
\leq\ & \sum_{j=0}^{\infty} \sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{1}{NT} P\left(\max_{1 \leq n' \leq N, 1 \leq t' \leq T} \max_{2 \leq k \leq m_j+1} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_k, \theta_j) \right| > \varepsilon NT \left|\theta_j - \theta^0\right| \right) \\
\leq\ & \sum_{j=0}^{\infty} \sum_{k=2}^{m_j+1} \sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{1}{NT} P\left(\max_{1 \leq n' \leq N, 1 \leq t' \leq T} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_k, \theta_j) \right| > \varepsilon NT \left|\theta_j - \theta^0\right| \right) \\
\leq\ & \sum_{j=0}^{\infty} \sum_{k=2}^{m_j+1} \sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{\varepsilon^{-(2+r)} (NT)^{-(2+r)} T^{2+r} c^{-(2+r)j}}{\overline{v}_{N,T}^{2+r}(\epsilon) NT} \\
& \times E\left(\max_{1 \leq n' \leq N, 1 \leq t' \leq T} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_k, \theta_j) \right|^{2+r} \right)
\end{aligned}$$

for some $r > 0$ such that $E\left\|\underline{\mathbf{u}}_t^0\right\|^{4+2r} \leq c_0 T^{2+r}$. Note now that, as in the above, $\left(\delta_i^{0\prime} \underline{\mathbf{u}}_t^0\right)^2 d_{2t}(\theta_k, \theta_j)$ is a strictly stationary, ρ-mixing sequence with mixing numbers ρ_m such that $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$;

thus, using Burkholder's inequality (see Theorem 1.1 in Shao, 1995), we obtain

$$
\begin{aligned}
& E\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 d_{2t}\left(\theta_k,\theta_j\right)\right|^{2+r} \\
\leq \;& c_0 N^{2+r}T^{1+r/2}\left(E\left|\left\|\underline{\mathbf{u}}_t^0\right\|^2 d_{2t}\left(\theta_k,\theta_j\right)\right|^2\right)^{1+r/2} \leq c_1 N^{2+r}T^{1+r/2}\left|E\left\|\underline{\mathbf{u}}_t^0\right\|^4 d_{2t}^2\left(\theta_k,\theta_j\right)\right|^{1+r/2} \\
\leq \;& c_0 N^{2+r}T^{1+r/2}\left|\theta_k-\theta_j\right|^{1+r/2} \leq c_1 N^{2+r}\overline{v}_{NT}^{1+r/2}\left(\epsilon\right)c^{j(1+r/2)}.
\end{aligned}
$$

Applying Corollary 4 in Moricz (1983), it follows that

$$
\begin{aligned}
& \sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\sup_{0\leq j\leq Q}\max_{2\leq k\leq m_j+1}\frac{\frac{1}{NT}\left|\sum_{i=1}^{n'}\sum_{t=1}^{t'}\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 d_{2t}\left(\theta_k,\theta_j\right)\right|}{\left|\theta_j-\theta^0\right|}>\varepsilon\right) \\
\leq \;& c_0\sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{\varepsilon^{-(2+r)}\left(NT\right)^{-(2+r)}T^{2+r}c^{-(2+r)j}}{\overline{v}_{N,T}^{2+r}\left(\epsilon\right)NT}N^{2+r}\overline{v}_{NT}^{1+r/2}\left(\epsilon\right)c^{j(1+r/2)}\left(\ln N\ln T\right)^{2+r} \\
\leq \;& c_0\sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{c^{-(1+r/2)j}}{\overline{v}_{N,T}^{1+r/2}\left(\epsilon\right)NT}\left(\ln N\ln T\right)^{2+r}m_j \\
\leq \;& c_0\left(\sum_{j=0}^{\infty}c^{-rj/2}\right)\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{\left(\ln N\ln T\right)^{2+r}}{\overline{v}_{NT}^{r/2}\left(\epsilon\right)NT}<\infty.
\end{aligned}
$$

This entails that $a=o_{a.s.}\left(1\right)$. Similarly, considering b

$$
\begin{aligned}
& \sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\sup_{0\leq j\leq Q}\max_{1\leq k\leq m_j}\frac{\frac{1}{NT}\left|\sum_{i=1}^{n'}\sum_{t=1}^{t'}\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 d_{2t}\left(\theta_{k+1},\theta_k\right)\right|}{\left|\theta_j-\theta^0\right|}>\varepsilon\right) \\
\leq \;& c_0\sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\left|\sum_{i=1}^{n'}\sum_{t=1}^{t'}\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 d_{2t}\left(\theta_{k+1},\theta_k\right)\right|>NT\left|\theta_j-\theta^0\right|\varepsilon\right) \\
\leq \;& c_0\sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}c^{-2j}\frac{\varepsilon^{-2}\left(NT\right)^{-2}T^2}{\overline{v}_{N,T}^2\left(\epsilon\right)NT}E\max_{1\leq n'\leq N,1\leq t'\leq T}\left|\sum_{i=1}^{n'}\sum_{t=1}^{t'}\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 d_{2t}\left(\theta_{k+1},\theta_k\right)\right|^2.
\end{aligned}
$$

Also, using again Theorem 1.1 in Shao (1995)

$$
\begin{aligned}
& E\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\right)^2 d_{2t}\left(\theta_{k+1},\theta_k\right)\right|^2 \\
\leq\ & c_0 N^2 T E\left\|\underline{\mathbf{u}}_t^0\right\|^4 d_{2t}^2\left(\theta_{k+1},\theta_k\right)\leq c_1 N^2 T\left|\theta_{k+1}-\theta_k\right| \\
\leq\ & c_1 N^2 T\frac{\delta_j}{m_j}=c_1 N^2,
\end{aligned}
$$

so that Corollary 4 in Moricz (1983) yields

$$
\begin{aligned}
& \sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\sup_{0\leq j\leq Q}\max_{1\leq k\leq m_j}\frac{\frac{\ln\ln T}{NT^2}\left|\sum_{i=1}^{n'}\sum_{t=1}^{t'}\left(\Delta_i f_t\right)^2 d_{2t}\left(\theta_{k+1},\theta_k\right)\right|}{\left|\theta_j-\theta^0\right|}>\varepsilon\right) \\
\leq\ & c_0\sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}c^{-2j}\frac{\varepsilon^{-2}}{\overline{v}_{N,T}^2\left(\epsilon\right)NT}\left(\ln N\ln T\right)^2 \\
\leq\ & c_0\left(\sum_{j=0}^{\infty}c^{-j}\right)\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{\varepsilon^{-2}}{\overline{v}_{NT}NT}\left(\ln N\ln T\right)^2<\infty,
\end{aligned}
$$

so that $b=o_{a.s.}(1)$. Putting all together, the final result now obtains. □

Lemma C.5. *We assume that Assumptions 3-5 are satisfied. Then there exist two random variables N_0 and T_0, and a positive, finite constant c_0, such that for $N\geq N_0$ and $T\geq T_0$, it holds that*

$$
\sup_{\left|\theta-\theta^0\right|\in B_{N,T}}\frac{\left|h^0\left(\eta,\theta\right)-h^0\left(\eta,\theta^0\right)\right|}{\left|\theta-\theta^0\right|}=o_{a.s.}\left(1\right).
$$

Proof. The proof is very similar to the one of the previous lemma, and we only report the main passages to save space. Again, set $\eta=1$ and write $h^0\left(\eta,\theta\right)=h^0\left(\theta\right)$ (and $h^0\left(\eta,\theta^0\right)=h^0\left(\theta^0\right)$) for short; we have

$$
\begin{aligned}
\sup_{\theta\in V_c\left(\overline{v}_{NT}\right)}\frac{h^0\left(\theta\right)}{\left|\theta-\theta^0\right|} &\leq 2\sup_{0\leq j\leq Q}\frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\left[d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right]}{\left|\theta_j-\theta^0\right|} \\
&\quad +2\sup_{0\leq j\leq Q}\sup_{\theta_j\leq\theta\leq\theta_{j+1}}\frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\left[d_{\mathcal{S},t}\left(\theta\right)-d_{\mathcal{S},t}\left(\theta_j\right)\right]}{\left|\theta_j-\theta^0\right|} \\
&= 2\sup_{0\leq j\leq Q}\frac{h^0\left(\theta_j,\theta^0\right)}{\left|\theta_j-\theta^0\right|}+2\sup_{0\leq j\leq Q}\sup_{\theta_j\leq\theta\leq\theta_{j+1}}\frac{h^0\left(\theta,\theta_j\right)}{\left|\theta_j-\theta^0\right|}=I+II.
\end{aligned}
\tag{C.20}
$$

Let

$$h^0_{n',t'}\left(\theta_j,\theta^0\right)=\frac{1}{NT}\sum_{i=1}^{n'}\sum_{t=1}^{t'}e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\left[d_{\mathcal{S},t}\left(\theta_j\right)-d_{\mathcal{S},t}\left(\theta^0\right)\right].$$

It holds that

$$\begin{aligned}&\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\sup_{0\leq j\leq Q}\frac{h^0_{n',t'}\left(\theta_j,\theta^0\right)}{\left|\theta_j-\theta^0\right|}>N^{-1/2}\right)\\ \leq\ &\sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}h^0_{n',t'}\left(\theta_j,\theta^0\right)>N^{-1/2}\left|\theta_j-\theta^0\right|\right)\\ \leq\ &\sum_{j=0}^{\infty}c^{-2j}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{N}{NT}\frac{T^2}{\overline{v}^2_{NT}\left(\epsilon\right)}E\left|\max_{1\leq n'\leq N,1\leq t'\leq T}h^0_{n',t'}\left(\theta_j,\theta^0\right)\right|^2.\end{aligned}$$

Also, using Assumption 4*(v)* it can be shown after some algebra that $E\left|h^0\left(\theta_j,\theta^0\right)\right|^2\leq c_0\left(NT\right)^{-1}\left|\theta_j-\theta^0\right|$; hence

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\sup_{0\leq j\leq Q}\frac{h^0_{n',t'}\left(\theta_j,\theta^0\right)}{\left|\theta_j-\theta^0\right|}>\varepsilon\right)\leq c_0\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{\overline{v}_{NT}\left(\epsilon\right)NT}\left(\ln N\ln T\right)^2<\infty.$$

Thus, in (D.12), $I=O_{a.s.}\left(N^{-1/2}\right)$. We now turn to II. We have

$$\begin{aligned}&\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\sup_{0\leq j\leq Q}\sup_{\theta_j\leq\theta\leq\theta_{j+1}}\frac{h^0_{n',t'}\left(\theta,\theta_j\right)}{\left|\theta_j-\theta^0\right|}\geq NT\right)\\ \leq\ &\sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\max_{2\leq k\leq m_j+1}h^0_{n',t'}\left(\theta_k,\theta_j\right)\geq NT\left|\theta_j-\theta^0\right|\right)\\ &+\sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\max_{1\leq k\leq m_j}h^0_{n',t'}\left(\theta_{k+1},\theta_k\right)\geq NT\left|\theta_j-\theta^0\right|\right)\end{aligned}$$

having defined

$$\begin{aligned}h^0_{n',t'}\left(\theta,\theta_j\right)&=\sum_{i=1}^{n'}\sum_{t=1}^{t'}e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\widetilde{d}'_{2t}\left(\theta,\theta_j\right),\\ h^0_{n',t'}\left(\theta_k,\theta_j\right)&=\sum_{i=1}^{n'}\sum_{t=1}^{t'}e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\widetilde{d}'_{2t}\left(\theta_k,\theta_j\right),\\ h^0_{n',t'}\left(\theta_{k+1},\theta_k\right)&=\sum_{i=1}^{n'}\sum_{t=1}^{t'}e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\widetilde{d}'_{2t}\left(\theta_{k+1},\theta_k\right),\end{aligned}$$

with $\widetilde{d}'_{2t}(\theta,\theta_j) = d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)$, and $\widetilde{d}'_{2t}(\theta_k,\theta_j)$ and $\widetilde{d}'_{2t}(\theta_{k+1},\theta_k)$ defined similarly. We have

$$\begin{aligned}
& \sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\le n'\le N,1\le t'\le T}\max_{2\le k\le m_j+1} h^0_{n',t'}(\theta_k,\theta_j) \ge NT\left|\theta_j-\theta^0\right|\right) \\
\le\ & \sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\le n'\le N,1\le t'\le T} h^0_{n',t'}(\theta_k,\theta_j) \ge NT\left|\theta_j-\theta^0\right|\right) \\
\le\ & \sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{(NT)^{-(2+r)}}{NT}\frac{c^{-(2+r)j}T^{2+r}}{\overline{v}^{2+r}_{NT}(\epsilon)}E\left|\max_{1\le n'\le N,1\le t'\le T} h^0_{n',t'}(\theta_k,\theta_j)\right|^{2+r}.
\end{aligned}$$

Also, using again Theorem 1.1 in Shao (1995), it can be shown that

$$E\left|\sum_{i=1}^{N}\sum_{t=1}^{T} e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\widetilde{d}'_{2t}(\theta_k,\theta_j)\right|^{2+r} \le c_0 N^{2+r}c^{j(1+r/2)}\overline{v}^{1+r/2}_{NT}.$$

Thus, by Corollary 4 in Moricz (1983)

$$\begin{aligned}
& \sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\le n'\le N,1\le t'\le T}\max_{2\le k\le m_j+1} h^0_{n',t'}(\theta_k,\theta_j) \ge NT\left|\theta_j-\theta^0\right|\right) \\
\le\ & c_0\sum_{j=0}^{\infty}\sum_{k=2}^{m_j+1}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}\frac{c^{-jr/2}}{\overline{v}^{r/2}_{NT}(\epsilon)}(\ln N\ln T)^{2+r}
\end{aligned}$$

Similarly

$$\begin{aligned}
& \sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\le n'\le N,1\le t'\le T}\max_{1\le k\le m_j} h^0_{n',t'}(\theta_{k+1},\theta_k) \ge NT\left|\theta_j-\theta^0\right|\right) \\
\le\ & \sum_{j=0}^{\infty}\sum_{k=1}^{m_j}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\le n'\le N,1\le t'\le T} h^0_{n',t'}(\theta_{k+1},\theta_k) \ge NT\left|\theta_j-\theta^0\right|\right) \\
\le\ & \sum_{j=0}^{\infty}\sum_{k=1}^{m_j}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{(NT)^{-2}}{NT}\frac{c^{-2j}T^{2}}{\overline{v}^{2}_{NT}(\epsilon)}E\left|\max_{1\le n'\le N,1\le t'\le T} h^0_{n',t'}(\theta_{k+1},\theta_k)\right|^{2},
\end{aligned}$$

with, after some algebra

$$E\left|\sum_{i=1}^{N}\sum_{t=1}^{T} e_{i,t}\delta_i^{0\prime}\underline{\mathbf{u}}_t^0\widetilde{d}'_{2t}(\theta_{k+1},\theta_k)\right|^{2} \le c_0 N.$$

Thus

$$\sum_{j=0}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left(\max_{1\leq n'\leq N,1\leq t'\leq T}\max_{2\leq k\leq m_j+1}h_{n',t'}^{0}\left(\theta_k,\theta_j\right)\geq N^{1/2}T\left|\theta_j-\theta^0\right|\right)$$
$$\leq \left(\sum_{j=0}^{\infty}c^{-j}\right)\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{\left(NT^2\right)^{-1}}{NT}\frac{T^2}{\overline{v}_{NT}^2\left(\epsilon\right)}N\left(\ln N\ln T\right)^2<\infty.$$

The desired result now follows. □

Lemma C.6. *We assume that Assumptions 1-4 are satisfied. Then it holds that, for every* $P_j \geq P_j^0$

$$\left\|\widehat{\beta}_{j,i}^{P_j\prime}\widehat{\mathbf{u}}_{j,t}\widehat{d}_{j,t}-\beta_{j,i}^{0\prime}\mathbf{u}_{j,t}^{0}d_{j,t}\left(\theta^0\right)\right\|=o_{a.s.}\left(1\right), \tag{C.21}$$

for $j=\mathcal{D},\mathcal{S}$ *and all* $1\leq i\leq N$, $1\leq t\leq T$. *Also,* $S\left(\mathbf{B}^P,\mathbf{U}^P,\theta\right)$ *is continuous in* $\left(\mathbf{B},\mathbf{U}\right)$.

Proof. Using the results in the proof of Lemma A.8 in Massacci (2017), it can be shown that $E\left\|\widehat{\beta}_{j,i}^{P_j}-\widehat{H}_{jj}^{P_j}\left(\theta^0\right)\beta_{j,i}^{0}\right\|^2\leq c_0C_{N,T}^{-2}$. Thus, using the same logic as above, Lemma C.1 entails that for all $1\leq i\leq N$, $\left\|\widehat{\beta}_{j,i}^{P_j}-\widehat{H}_{jj}^{P_j}\left(\theta^0\right)\beta_{j,i}^{0}\right\|=o_{a.s.}\left(C_{N,T}^{-1}\left(\ln N\right)^{1+\epsilon}\left(\ln T\right)^{1+\epsilon}\right)$. With the same logic, by Corollary 3.2 in Massacci (2017), it follows that, for all $1\leq t\leq T$, $\left\|\widehat{\mathbf{u}}_{j,t}\widehat{d}_{j,t}-\left(\widehat{H}_{jj}^{P_j}\left(\theta^0\right)'\right)^{-1}\mathbf{u}_{j,t}^{0}d_{j,t}\left(\theta^0\right)\right\|=o_{a.s.}\left(C_{N,T}^{-1}\left(\ln N\right)^{1+\epsilon}\left(\ln T\right)^{1+\epsilon}\right)$. Equation (C.21) now follows immediately. As a consequence, we can assume that

$$\lim_{N,T\to\infty}\widehat{\beta}_{j,i}^{P_j\prime}\widehat{\mathbf{u}}_{j,t}\widehat{d}_{j,t}=\beta_{j,i}^{0\prime}\mathbf{u}_{j,t}^{0}d_{j,t}\left(\theta^0\right). \tag{C.22}$$

Equation (C.22) and elementary algebra now give $\lim_{N,T\to\infty}S\left(\widehat{\mathbf{B}}^P,\widehat{\mathbf{U}}^P,\theta\right)=S\left(\mathbf{B}^0,\mathbf{U}^0,\theta\right)$, which proves the last statement of the lemma. □

We are now ready to derive a strong rate for $\widehat{\theta}$.

Lemma C.7. *We assume that Assumptions 1-5 are satisfied. Then it holds that*

$$\widehat{\theta}-\theta^0=o_{a.s.}\left(T^{-1}\overline{v}_{N,T}\left(\epsilon\right)\right),$$

for every $\epsilon>0$.

Proof. Lemma C.3 entails that we can focus on the event $\left|\widehat{\theta}-\theta^0\right| < c_B$. Further, after some algebra, Lemma C.6 entails that there exists two random variables N_0 and T_0 such that, for $N \geq N_0$ and $T \geq T_0$, we have

$$\frac{S\left(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \theta\right) - S\left(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \theta^0\right)}{\left|\theta-\theta^0\right|}$$
$$\geq \inf_{|\theta-\theta_0|\in B_{N,T}} \frac{E\omega^0\left(\eta,\theta\right)}{\left|\theta-\theta^0\right|} - \sup_{|\theta-\theta_0|\in B_{N,T}} \frac{\left|\overline{\omega}^0_{N,T}\right|}{\left|\theta-\theta^0\right|} - 2\sup_{|\theta-\theta_0|\in B_{N,T}} \frac{\left|h^0\left(\eta,\theta\right)-h^0\left(\eta,\theta^0\right)\right|}{\left|\theta-\theta^0\right|}.$$

By using Lemmas C.4 and C.5, it follows that, for all θ such that $|\theta-\theta_0| \in B_{N,T}$

$$S\left(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \theta\right) - S\left(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \theta^0\right) \geq c_0 > 0 \text{ a.s.}$$

Since, by definition, $S\left(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \widehat{\theta}\right) \leq S\left(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \theta^0\right)$, this entails that $\left|\widehat{\theta}-\theta^0\right| \notin B_{N,T}$, whence the desired result. □

Lemma C.8. *We assume that Assumptions 1-5 are satisfied. Then it holds that*

$$\frac{T_j - \widehat{T}_j}{\widehat{T}_j} = o_{a.s.}\left(T_j^{-1}\left(\ln T\right)^{1+\epsilon'}\overline{v}_{N,T}\left(\epsilon\right)\right),$$

for $j = \mathcal{D}, \mathcal{S}$ *and every* $\epsilon, \epsilon' > 0$.

Proof. We show the lemma for $j = \mathcal{D}$. Recall that $T_{\mathcal{D}} = \sum_{t=1}^T d_{\mathcal{D},t}\left(\theta^0\right)$ and $\widehat{T}_{\mathcal{D}} = \sum_{t=1}^T \widehat{d}_{\mathcal{D},t}$. It holds that

$$T_{\mathcal{D}} - \widehat{T}_{\mathcal{D}} \leq \sum_{t=1}^T \left|d_{\mathcal{D},t}\left(\theta^0\right) - \widehat{d}_{\mathcal{D},t}\right|$$

Denoting the density of z_t by $f_t\left(z\right) = f\left(z\right)$, Assumption 5*(iii)* entails

$$E\left|\widehat{d}_{\mathcal{D},t} - d_{\mathcal{D},t}\left(\theta^0\right)\right| = \left|\int_{\widehat{\theta}}^{\theta^0} f\left(z\right)dz\right| \leq c_0\left|\widehat{\theta}-\theta^0\right|, \tag{C.23}$$

so that

$$T_{\mathcal{D}} - \widehat{T}_{\mathcal{D}} \leq c_0 T\left|\widehat{\theta}-\theta^0\right|\left(\ln T\right)^{1+\epsilon},$$

having used Lemma C.1. The desired result follows immediately from Lemma C.7.

□

Lemma C.9. *We assume that Assumptions 3-5 are satisfied. Then it holds that*

$$E\left|\widetilde{R}_{i,t}\right|^{4+\epsilon}<\infty \tag{C.24}$$

for some $\epsilon>0$*, and*

$$E\max_{1\leq t\leq T}\left|\sum_{s=1}^{t}\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{\mathcal{D},s}\left(\theta^{0}\right)-E\left[\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{\mathcal{D},s}\left(\theta^{0}\right)\right]\right|^{2} \leq c_{1}T, \tag{C.25}$$

$$E\max_{1\leq t\leq T}\left|\sum_{s=1}^{t}\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{\mathcal{S},s}\left(\theta^{0}\right)-E\left[\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{\mathcal{S},s}\left(\theta^{0}\right)\right]\right|^{2} \leq c_{2}T, \tag{C.26}$$

for all $1\leq h,l\leq N$*.*

Proof. Equation (C.24) is an immediate consequence of Assumptions 3*(i)* and 4*(i)*. We show only (C.25); (C.26) follows from exactly the same arguments. To begin with, note that Assumption 5*(i)* entails that $\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{1,s}\left(\theta^{0}\right)$ is also a strictly stationary, ρ-mixing sequence with mixing numbers ρ_m such that $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$ - we refer to the proof of Lemma C.11 for details. Using (C.24), it follows that

$$E\left|\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{\mathcal{D},s}\left(\theta^{0}\right)\right|^{2}\leq\left(E\left|\widetilde{R}_{h,s}\right|^{4}\right)^{1/2}\left(E\left|\widetilde{R}_{l,s}\right|^{4}\right)^{1/2}<\infty.$$

Note also that

$$\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty\Rightarrow\sum_{m=1}^{\infty}\frac{\rho_m}{m}<\infty\Rightarrow\sum_{m=1}^{\infty}\rho_m\left(2^{m}\right)<\infty.$$

Hence, Corollary 1.1 in Shao (1995) and the strict stationarity of $\widetilde{R}_{h,s}\widetilde{R}_{l,s}d_{\mathcal{D},s}\left(\theta^{0}\right)$ immediately yield (C.25). □

Lemma C.10. *We assume that Assumptions 1-6 are satisfied. Then it holds that*

$$\lim \sup_{\min(N,T)\to\infty} \frac{1}{N-p} \sum_{h=p+1}^{N} \widehat{g}_j^{(h)} = \overline{g}_j < \infty,$$
$$\lim \inf_{\min(N,T)\to\infty} \frac{1}{N-p} \sum_{h=p+1}^{N} \widehat{g}_j^{(h)} = \underline{g}_j > 0,$$

for $j = \mathcal{D}, \mathcal{S}$ *and every* $0 \leq p \leq P_{\max}$.

Proof. Note that

$$\begin{aligned} & \sum_{h=p+1}^{N} \widehat{g}_j^{(h)} - \sum_{h=p+1}^{N} g_j^{(h)} \\ = & \sum_{h=1}^{N} \widehat{g}_j^{(h)} - \sum_{h=1}^{N} g_j^{(h)} - \sum_{h=1}^{p} \widehat{g}_j^{(h)} + \sum_{h=1}^{p} g_j^{(h)} \\ = & \; tr\left(\widehat{\Sigma}_j\right) - tr\left(\Sigma_j\right) - \left(\sum_{h=1}^{p} \left(\widehat{g}_j^{(h)} - g_j^{(h)}\right)\right). \end{aligned}$$

Using Lemma A.2, it is easy to see that

$$\sum_{h=1}^{p} \left(\widehat{g}_j^{(h)} - g_j^{(h)}\right) = o_{a.s.}\left(\frac{pNT^{1/2}}{T\pi_j} \left(\ln N\right)^{1+\epsilon} \left(\ln T\right)^{1/2+\epsilon}\right). \tag{C.27}$$

Also, the same passages as in the proof of Lemma A.1 in Trapani (2018) yield

$$tr\left(\widehat{\Sigma}_j\right) - tr\left(\Sigma_j\right) = o_{a.s.}\left(\frac{N\left(\ln N \ln T\right)^{\frac{1+\epsilon}{2}}}{T^{1/2}\pi_j}\right). \tag{C.28}$$

Thus, recalling that $P_{\max} = O\left(\min\left\{T^{1/2-c}, N^{1/2-c}\right\}\right)$, it follows that

$$\frac{1}{N-p} \sum_{h=p+1}^{N} \widehat{g}_j^{(h)} = \frac{1}{N-p} \sum_{h=p+1}^{N} g_j^{(h)} + o_{a.s.}\left(\frac{\left(\ln N \ln T\right)^{\frac{1+\epsilon}{2}}}{T^{1/2}\pi_j}\right) = \frac{1}{N-p} \sum_{h=p+1}^{N} g_j^{(h)} + o_{a.s.}\left(1\right),$$

under Assumption 6.

Let now $g_j^{u,(h)}$ and $g_j^{e,(h)}$ denote the h-th largest eigenvalues of $\frac{1}{T\pi_j} \sum_{t-1}^{T} E\left(\mathbf{B}_j^0 \mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0\prime} \mathbf{B}_j^{0\prime} d_{j,t}\left(\theta^0\right)\right)$

and $\frac{1}{T\pi_j}\sum_{t-1}^{T} E\left(\mathbf{e}_t\mathbf{e}_t' d_{j,t}\left(\theta^0\right)\right)$ respectively. By Weyl's inequality

$$g_j^{e,(N)} + g_j^{u,(h)} \leq g_j^{(h)} \leq g_j^{u,(h)} + g_j^{e,(1)},$$

so that

$$g_j^{e,(N)} + \frac{1}{N-p}\sum_{h=p+1}^{N} g_j^{u,(h)} \leq \frac{1}{N-p}\sum_{h=p+1}^{N} g_j^{(h)} \leq \frac{1}{N-p}\sum_{h=p+1}^{N} g_j^{u,(h)} + g_j^{e,(1)}.$$

By Assumption 2*(i)(b)*, it holds that

$$g_j^{e,(N)} + \frac{1}{N-p}\sum_{h=p+1}^{N} g_j^{u,(h)} > 0;$$

also, Assumption 2*(i)(a)* and (D.1) entail

$$\frac{1}{N-p}\sum_{h=p+1}^{N} g_j^{u,(h)} + g_j^{e,(1)} < \infty.$$

Hence, as $N \to \infty$

$$0 < \lim\inf_{N\to\infty} < \frac{1}{N-p}\sum_{h=p+1}^{N} g_j^{(h)} < \lim\sup_{N\to\infty} < \infty.$$

Putting all together, the lemma follows. □

Lemma C.11. *We assume that Assumptions 1-5 are satisfied. Then it holds that, for* $j = \mathcal{D}, \mathcal{S}$

$$E\left\|T^{-1}\sum_{t=1}^{T}\mathbf{u}_{j,t}^0 d_{j,t}\left(\theta^0\right)\right\|^2 \leq c_0 T^{-1}. \tag{C.29}$$

Further, under Assumption 9(ii), as $T \to \infty$

$$T^{-1/2}\sum_{t=1}^{T}\mathbf{u}_{j,t}^0 d_{j,t}\left(\theta^0\right) \xrightarrow{D} N\left(0, V_{\mathbf{u},j}\right), \tag{C.30}$$

where $V_{\mathbf{u},j} = \lim_{T\to\infty} E\left(T^{-1}\sum_{t,s=1}^{T}\mathbf{u}_{j,t}^0\mathbf{u}_{j,s}'^0 d_{j,t}\left(\theta^0\right) d_{j,s}\left(\theta^0\right)\right)$.

Proof. We begin by showing that $\left\{\mathbf{f}_{j,t}^0 d_{j,t}\left(\theta^0\right)\right\}$ is a ρ-mixing sequence with mixing number $\widetilde{\rho}_{j,m}$

such that $\sum_{m=1}^{\infty} \widetilde{\rho}_{j,m}^{1/2} < \infty$. Note that $\mathbf{f}_{j,t}^{0} d_{j,t}(\theta^0)$ is a measurable transformation of $\left\{\mathbf{f}_{j,t}^{0}, z_t, \epsilon_t\right\}$. Let now $\Im_{-\infty}^{t}$ and $\Im_{t+m}^{\infty}$ be the σ-fields generated by $\left\{\mathbf{f}_{j,t'}^{0}, z_{t'}, \epsilon_{t'}\right\}_{t'=-\infty}^{t}$ and $\left\{\mathbf{f}_{j,t'}^{0}, z_{t'}, \epsilon_{t'}\right\}_{t'=t+m}^{\infty}$ respectively, and similarly $\widetilde{\Im}_{-\infty}^{t}$ and $\widetilde{\Im}_{t+m}^{\infty}$ be the σ-fields generated by $\left\{\mathbf{f}_{j,t'}^{0} d_{j,t'}(\theta^0)\right\}_{t'=-\infty}^{t}$ and $\left\{\mathbf{f}_{j,t'}^{0} d_{j,t'}(\theta^0)\right\}_{t'=t+m}^{\infty}$ respectively. By measurability, $\widetilde{\Im}_{-\infty}^{t} \subseteq \Im_{-\infty}^{t}$ and $\widetilde{\Im}_{t+m}^{\infty} \subseteq \Im_{t+m}^{\infty}$; hence, $\widetilde{\rho}_{j,m} = \widetilde{\rho}_{j,m}\left(\widetilde{\Im}_{-\infty}^{t}, \widetilde{\Im}_{t+m}^{\infty}\right) \leq \rho_{j,m} = \rho_{j,m}\left(\Im_{-\infty}^{t}, \Im_{t+m}^{\infty}\right)$. This therefore entails that $\sum_{m=1}^{\infty} \widetilde{\rho}_{j,m}^{1/2} < \sum_{m=1}^{\infty} \rho_{j,m}^{1/2} < \infty$ by Assumption 5*(i)*, which in turn entails that $\sum_{k=1}^{\infty} \widetilde{\rho}_{j}\left(2^k\right) < \infty$.

Note now that $E\left\|\mathbf{u}_{j,t}^{0} d_{j,t}(\theta^0)\right\|^2 \leq E\left\|\mathbf{u}_{j,t}^{0}\right\|^2$, which is finite by Assumption 3*(i)*. Thus

$$E\left\|\sum_{t=1}^{T} \mathbf{u}_{j,t}^{0} d_{j,t}\left(\theta^0\right)\right\|^2 \leq E \max_{1 \leq k \leq T}\left\|\sum_{t=1}^{k} \mathbf{u}_{j,t}^{0} d_{j,t}\left(\theta^0\right)\right\|^2 \leq c_0 T,$$

where the last inequality follows from Lemma 1 in Peligrad, Utev, and Wu (2007). This proves (C.29). As far as (C.30) is concerned, it follows readily upon checking the assumptions of Theorem 0 in Peligrad (1987). □

D Proofs

Henceforth, we use E^* and V^* to denote, respectively, the expected value and the variance with respect to P^*.

Proof of Lemma A.1. The lemma is an adaptation of Lemma 1 in Trapani (2018). In order to prove it, note that Assumption 4*(v)* entails that, for $j = \mathcal{D}, \mathcal{S}$

$$\begin{aligned} & \frac{1}{T\pi_j}\sum_{t=1}^{T} E\left[\left(\mathbf{B}_j^0 \mathbf{u}_{j,t}^0 + \mathbf{e}_t\right)\left(\mathbf{B}_j^0 \mathbf{u}_{j,t}^0 + \mathbf{e}_t\right)' d_{j,t}\left(\theta^0\right)\right] \\ = \;& \mathbf{B}_j^0 \frac{1}{T\pi_j}\sum_{t=1}^{T} E\left(\mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0\prime} d_{j,t}\left(\theta^0\right)\right)\mathbf{B}_j^{0\prime} + \frac{1}{T\pi_j}\sum_{t=1}^{T} E\left(\mathbf{e}_t \mathbf{e}_t' d_{j,t}\left(\theta^0\right)\right). \end{aligned}$$

We begin by showing that

$$g_j^{(i)}\left(\mathbf{B}_j^0 \frac{1}{T\pi_j}\sum_{t=1}^{T} E\left(\mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0\prime} d_{j,t}\left(\theta^0\right)\right)\mathbf{B}_j^{0\prime}\right)\begin{cases} \in \left(\underline{c}_j^{(i)} N, \overline{c}_j^{(i)} N\right) & \text{for } 0 \leq i \leq P_j^0 \\ = 0 & \text{for } i \geq P_j^0 + 1 \end{cases}, \tag{D.1}$$

for $j=\mathcal{D},\mathcal{S}$, where $0<\underline{c}_j^{(i)}\leq \overline{c}_j^{(i)}<\infty$. We show this result for the case $j=\mathcal{D}$ only - the case $j=\mathcal{S}$ is repetitive. By the multiplicative Weyl's inequality (Merikoski and Kumar, 2004) we have

$$g_{\mathcal{D}}^{(i)}\left(\mathbf{B}_{\mathcal{D}}^{0}\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\mathbf{u}_{\mathcal{D},t}^{0}\mathbf{u}_{\mathcal{D},t}^{0\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\mathbf{B}_{\mathcal{D}}^{0\prime}\right)\geq g_{\mathcal{D}}^{(\min)}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\mathbf{u}_{\mathcal{D},t}^{0}\mathbf{u}_{\mathcal{D},t}^{0\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)g_{\mathcal{D}}^{(i)}\left(\mathbf{B}_{\mathcal{D}}^{0\prime}\mathbf{B}_{\mathcal{D}}^{0}\right),$$
$$g_{\mathcal{D}}^{(i)}\left(\mathbf{B}_{\mathcal{D}}^{0}\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\mathbf{u}_{\mathcal{D},t}^{0}\mathbf{u}_{\mathcal{D},t}^{0\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\mathbf{B}_{\mathcal{D}}^{0\prime}\right)\leq g_{\mathcal{D}}^{(\max)}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\mathbf{u}_{\mathcal{D},t}^{0}\mathbf{u}_{\mathcal{D},t}^{0\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)g_{\mathcal{D}}^{(i)}\left(\mathbf{B}_{\mathcal{D}}^{0\prime}\mathbf{B}_{\mathcal{D}}^{0}\right),$$

by Assumption 3*(ii)* it holds that

$$g_{\mathcal{D}}^{(\min)}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\mathbf{u}_{\mathcal{D},t}^{0}\mathbf{u}_{\mathcal{D},t}^{0\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)>0.$$

Moreover, by Assumption 3*(iv)* we have

$$g_{\mathcal{D}}^{(i)}\left(\mathbf{B}_{\mathcal{D}}^{0}\mathbf{B}_{\mathcal{D}}^{0\prime}\right)\begin{cases}\geq c_1^{(i)}N & \text{for } 1\leq i\leq P_{\mathcal{D}}^{0}\\ =0 & \text{for } i\geq P_{\mathcal{D}}^{0}+1\end{cases}.$$

Equation (D.1) now follows immediately. Consider now

$$\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t^{\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)=\frac{1}{T\pi_{j}}\sum_{t=1}^{T}\epsilon_t\epsilon_t^{\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)-\overline{\epsilon}_{\mathcal{D}}\overline{\epsilon}_{\mathcal{D}}^{\prime},$$

so that

$$g_{\mathcal{D}}^{(1)}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\mathbf{e}_t\mathbf{e}_t^{\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)\leq g_{\mathcal{D}}^{(1)}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}E\left(\epsilon_t\epsilon_t^{\prime}d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)+g_{\mathcal{D}}^{(1)}\left(\overline{\epsilon}_{\mathcal{D}}\overline{\epsilon}_{\mathcal{D}}^{\prime}\right).$$

The first term is bounded by Assumption 2*(i)*. Also, after some algebra we have

$$g_{\mathcal{D}}^{(1)}\left(\overline{\epsilon}_{\mathcal{D}}\overline{\epsilon}_{\mathcal{D}}^{\prime}\right)\leq\max_{1\leq i\leq N}\sum_{k=1}^{N}\frac{1}{\left(T\pi_{\mathcal{D}}\right)^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{i,t}\epsilon_{k,s}d_{\mathcal{D},t}\left(\theta^{0}\right)d_{\mathcal{D},s}\left(\theta^{0}\right)\right)\right|,$$

which is bounded by Assumption 2*(ii)*. The proof of the lemma now follows immediately along the same lines as the proof of Lemma 1 in Trapani (2018). □

Proof of Theorem A.2. We prove the theorem for $j = 1$ - again, the proof for the case $j = 2$ is just a repetition of the same arguments. Note that

$$\left|\widehat{g}_1^{(i)} - g_1^{(i)}\right| \leq \left\|\widehat{\Sigma}_{\mathcal{S}} - \Sigma_{\mathcal{D}}\right\|_{op},$$

so that, by symmetry

$$\left|\widehat{g}_1^{(i)} - g_1^{(i)}\right| \leq \left|\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\left(\widetilde{R}_{h,t}\widetilde{R}_{l,t}\widehat{d}_{\mathcal{D},t} - E\left(\widetilde{R}_{h,t}\widetilde{R}_{l,t}\widehat{d}_{1,t}\right)\right)\right)^2\right|^{1/2}. \tag{D.2}$$

We define the short-hand notation $\delta_{h,l,t} = X_{h,t}X_{l,t}d_{\mathcal{D},t}(\theta^0) - E\left(X_{h,t}X_{l,t}d_{\mathcal{D},t}(\theta^0)\right)$; based on this, (D.2) can be rewritten as

$$\left|\widehat{g}_1^{(i)} - g_1^{(i)}\right| \leq \left|\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\left(\delta_{h,l,t} + \widetilde{R}_{h,t}\widetilde{R}_{l,t}\left(\widehat{d}_{\mathcal{D},t} - d_{\mathcal{D},t}\left(\theta^0\right)\right)\right)\right)^2\right|^{1/2}. \tag{D.3}$$

By repeated use of the C_r-inequality we have

$$\begin{aligned}\left|\widehat{g}_1^{(i)} - g_1^{(i)}\right| &\leq c_0\left|\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\delta_{h,l,t}\right)^2 + \sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\widetilde{R}_{h,t}\widetilde{R}_{l,t}\left(\widehat{d}_{\mathcal{D},t} - d_{\mathcal{D},t}\left(\theta^0\right)\right)\right)^2\right|^{1/2}\\ &\leq c_1\left|\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\delta_{h,l,t}\right)^2\right|^{1/2} + c_2\left|\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\widetilde{R}_{h,t}\widetilde{R}_{l,t}\left(\widehat{d}_{\mathcal{D},t} - d_{\mathcal{D},t}\left(\theta^0\right)\right)\right)^2\right|^{1/2}.\end{aligned} \tag{D.4}$$

Using Lemma C.9, we have

$$E\max_{1\leq h\leq N, 1\leq l\leq N, 1\leq t\leq T}\sum_{h'=1}^{h}\sum_{l'=1}^{l}\left(\sum_{t'=1}^{t}\delta_{h',l',t'}\right)^2 \leq c_0N^2T;$$

thus, by Lemma C.1 and Markov inequality, we have

$$\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\delta_{h,l,t}\right)^2 = o_{a.s.}\left(\frac{N^2T}{T^2\pi_{\mathcal{D}}^2}(\ln N)^{2+\epsilon}(\ln T)^{1+\epsilon}\right),$$

for every $\epsilon > 0$. Considering the second term in (D.4), convexity implies that

$$\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\widetilde{R}_{h,t}\widetilde{R}_{l,t}\left(\widehat{d}_{\mathcal{D},t}-d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)^{2}\leq\frac{1}{T\pi_{\mathcal{D}}^{2}}\sum_{h=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\widetilde{R}_{h,t}^{2}\widetilde{R}_{l,t}^{2}\left(\widehat{d}_{\mathcal{D},t}-d_{\mathcal{D},t}\left(\theta^{0}\right)\right)^{2}. \tag{D.5}$$

Hence, applying (C.23) to (D.5) it follows that

$$\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\widetilde{R}_{h,t}\widetilde{R}_{l,t}\left(\widehat{d}_{\mathcal{S},t}-d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)^{2}\leq c_{0}\left|\widehat{\theta}-\theta^{0}\right|^{2}\frac{1}{T\pi_{\mathcal{D}}^{2}}\sum_{h=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\widetilde{R}_{h,t}^{2}\widetilde{R}_{l,t}^{2}.$$

Equation (C.24) entails

$$E\left(\frac{1}{N^{2}T}\sum_{h=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\widetilde{R}_{h,t}^{2}\widetilde{R}_{l,t}^{2}\right)\leq c_{0}.$$

The maximal inequality for rectangular sums (Moricz, 1983) now yields

$$E\max_{1\leq h\leq N,1\leq l\leq N,1\leq t\leq T}\sum_{h'=1}^{h}\sum_{l'=1}^{l}\sum_{t'=1}^{t}\widetilde{R}_{h',t'}^{2}\widetilde{R}_{l',t'}^{2}\leq c_{0}N^{2}T\left(\ln N\right)^{2}\ln T,$$

which, by Lemma C.1 and Markov inequality, yields

$$\sum_{h=1}^{N}\sum_{l=1}^{N}\sum_{t=1}^{T}\widetilde{R}_{h,t}^{2}\widetilde{R}_{l,t}^{2}=o_{a.s.}\left(N^{2}T\left(\ln N\right)^{3+\epsilon}\left(\ln T\right)^{2+\epsilon}\right),$$

for every $\epsilon > 0$. Then, by Lemma C.7 we obtain the general result

$$\left|\sum_{h=1}^{N}\sum_{l=1}^{N}\left(\frac{1}{T\pi_{\mathcal{D}}}\sum_{t=1}^{T}\widetilde{R}_{h,t}\widetilde{R}_{l,t}\left(\widehat{d}_{1,t}-d_{\mathcal{D},t}\left(\theta^{0}\right)\right)\right)^{2}\right|^{1/2}=o_{a.s.}\left(\frac{N}{T\pi_{\mathcal{D}}}\left(\ln N\right)^{\frac{3+\epsilon}{2}}\left(\ln T\right)^{\frac{2+\epsilon}{2}}\overline{v}_{N,T}\left(\epsilon\right)\right);$$

recalling that we have assumed $\eta = 1$, the desired result follows. □

Proof of Theorem A.3. The proofs are similar to those of Theorems 3 and 4 in Horváth and Trapani (2019), and therefore we only report the main arguments to save space. We begin by showing (A.12). It follows from Lemma A.1 and Theorem A.2 that, for all i and $j = \mathcal{D}, \mathcal{S}$

$$P\left\{\omega : \lim_{\min(N,T)\to\infty}\left(\frac{N^{-\varrho_j}\widehat{g}_j^{(i)}}{\overline{g}_j\left(p\right)} - \varkappa N^{1-\varrho_j}\right) = \infty\right\} = 1.$$

Thus, by continuity, we can assume henceforth that

$$\lim_{\min(N,T)\to\infty} \exp\left(-\varkappa N^{1-\varrho_j}\right)\psi\left(\widehat{g}_j^{(i)}\right) = \infty. \tag{D.6}$$

Let $\Phi\left(\cdot\right)$ denote the standard normal distribution; it holds that

$$\begin{aligned} M^{-1/2}\sum_{m=1}^{M}\left(\zeta_{j,m}^{(i)}\left(s\right) - \frac{1}{2}\right) &= M^{-1/2}\sum_{m=1}^{M}\left(I\left\{\xi_{i,m}^{(j)} \leq 0\right\} - \frac{1}{2}\right) + M^{-1/2}\sum_{m=1}^{M}\left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right) - \frac{1}{2}\right) \\ &+ M^{-1/2}\sum_{m=1}^{M}\left[I\left\{\xi_{i,m}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - I\left\{\xi_{i,m}^{(j)} \leq 0\right\} - \left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right) - \frac{1}{2}\right)\right]. \end{aligned}$$

By definition

$$\begin{aligned} E^*\zeta_{j,m}^{(i)}\left(s\right) &= \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right), \\ V^*\zeta_{j,m}^{(i)}\left(s\right) &= \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right)\left[1 - \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right)\right]. \end{aligned}$$

Thus

$$\begin{aligned} &E^*\int_{-\infty}^{\infty}\left|M^{-1/2}\sum_{m=1}^{M}\left[I\left\{\xi_{i,m}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - I\left\{\xi_{i,m}^{(j)} \leq 0\right\} - \left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right) - \frac{1}{2}\right)\right]\right|^2 d\Phi(s) \\ &= \int_{-\infty}^{\infty} E^*\left|\left[I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - I\left\{\xi_{i,1}^{(j)} \leq 0\right\} - \left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right) - \frac{1}{2}\right)\right]\right|^2 d\Phi(s), \end{aligned}$$

on account of the independence of the $\xi_{i,m}^{(j)}$s across m. Also

$$\begin{aligned} E^{*}\left(I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right\}-I\left\{\xi_{i,1}^{(j)} \leq 0\right\}\right) &= \Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)-\frac{1}{2} \\ V^{*}\left(I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right\}-I\left\{\xi_{i,1}^{(j)} \leq 0\right\}\right) &= \left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)-\frac{1}{2}\right)\left(\frac{3}{2}-\Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)\right) \\ &\leq \Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)-\frac{1}{2}. \end{aligned}$$

Hence, we have

$$\begin{aligned} &\int_{-\infty}^{\infty} E^{*}\left|\left[I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right\}-I\left\{\xi_{i,1}^{(j)} \leq 0\right\}-\left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)-\frac{1}{2}\right)\right]\right|^{2} d\Phi(s) \\ &\leq \int_{-\infty}^{\infty}\left|\Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)-\frac{1}{2}\right| d\Phi(s) \leq \frac{1}{\sqrt{2\pi}\psi\left(\widehat{g}_{j}^{(i)}\right)} \int_{-\infty}^{\infty}|s|\, d\Phi(s)=\frac{1}{\pi\psi\left(\widehat{g}_{j}^{(i)}\right)}, \end{aligned}$$

which drifts to zero by (D.6). Also

$$\int_{-\infty}^{\infty} M^{-1/2} \sum_{m=1}^{M}\left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_{j}^{(i)}\right)}\right)-\frac{1}{2}\right) d\Phi(s) \leq \frac{M}{2\pi\left|\psi\left(\widehat{g}_{j}^{(i)}\right)\right|^{2}} \int_{-\infty}^{\infty} s^{2} d\Phi(s)=\frac{M}{2\pi\left|\psi\left(\widehat{g}_{j}^{(i)}\right)\right|^{2}}.$$

Hence, using (A.11), we conclude via Markov's inequality that

$$\begin{aligned} \Upsilon_{j}^{(i)} &= \int_{-\infty}^{\infty}\left\{\frac{2}{\sqrt{M}} \sum_{m=1}^{M}\left(I\left\{\xi_{i,m}^{(j)} \leq 0\right\}-\frac{1}{2}\right)\right\}^{2} d\Phi(s)+o_{P^{*}}(1) \\ &= \left\{\frac{2}{\sqrt{M}} \sum_{m=1}^{M}\left(I\left\{\xi_{i,m}^{(j)} \leq 0\right\}-\frac{1}{2}\right)\right\}^{2}+o_{P^{*}}(1), \end{aligned}$$

and therefore the desired result follows from the Central Limit Theorem for Bernoulli random variables.

We now turn to showing (A.13). Again Lemma A.1 and Theorem A.2 entail that

$$P\left\{\omega : \lim_{\min(N,T)\to\infty} \frac{N^{-\varrho_j}\widehat{g}_j^{(i)}}{\overline{g}_j\left(p\right)} = 0\right\} = 1.$$

By continuity, this means that we can assume from now on that

$$\lim_{\min(N,T)\to\infty} \psi\left(\widehat{g}_j^{(i)}\right) = 1. \tag{D.7}$$

Consider

$$\zeta_{j,m}^{(i)}\left(s\right) - \frac{1}{2} = I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} \pm \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right) - \frac{1}{2}.$$

Then we have

$$\begin{aligned} & E^* \int_{-\infty}^{\infty} \left| M^{-1/2} \sum_{m=1}^{M} \left(I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - \frac{1}{2}\right)\right|^2 d\Phi\left(s\right) \\ = \; & E^* \left(I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right)\right)^2 + M \int_{-\infty}^{\infty} \left(\Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right) - \frac{1}{2}\right)^2 d\Phi\left(s\right). \end{aligned}$$

Given that

$$E^* \left(I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right)\right)^2 < \infty,$$

by Markov's inequality it follows that

$$\int_{-\infty}^{\infty} \left[M^{-1/2} \sum_{m=1}^{M} \left(I\left\{\xi_{i,1}^{(j)} \leq \frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right\} - \Phi\left(\frac{s}{\psi\left(\widehat{g}_j^{(i)}\right)}\right)\right)\right]^2 d\Phi\left(s\right) = O_{P^*}(1),$$

for almost all realizations of $\{e_j, b_j, -\infty < j < \infty\}$. Thus, as $\min\left(M, N, T_j\right) \to \infty$ for $j = \mathcal{D}, \mathcal{S}$

$$\frac{1}{4M}\Upsilon_j^{(i)} = \int_{-\infty}^{\infty} \left(\Phi\left(s\right) - \frac{1}{2}\right)^2 d\Phi\left(s\right) + O_{P^*}(1). \tag{D.8}$$

Hence the proof of (A.13) is complete. □

Proof of Theorem 1. See the proof of (the identical) Theorem 3 in Trapani (2018). □

Proof of Theorem 2. The result follows from combining (A.24) and (A.25), with

$$\widehat{\gamma}_{j,\mathbf{g}} - \gamma_{j,\mathbf{g}}^{0} = O_P\left(\frac{1}{N^{1/2}}\right) + O_P\left(\frac{T^{1/2}}{T_j}\right) + O_P\left(C_{N,T}^{-2}\right) + O_P\left(T^{-1}\right). \tag{D.9}$$

□

Proof of Theorem 3. The result readily follows by combining the results in Theorem A.5 (see also Theorem 3 in Giglio and Xiu, 2021). □

Proof of Lemma A.2. We report the proof using the assumption that the true $d_{j,t}$ has been used (also in the computation of $\widehat{T}_j$); extending to the case where $\widehat{d}_{j,t}$ is employed is straighforward in light of the results derived above. It holds that

$$\widehat{e}_{j,i} = \alpha_{j,i} + \overline{\epsilon}_{j,i} + \left(\frac{1}{T_j}\sum_{t=1}^{T} d_{j,t}\right)(\gamma_{j,0} - \widehat{\gamma}_{j,0}) + \frac{1}{T_j}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}'(\boldsymbol{\gamma}_{j,1} + \mathbf{u}_{j,t}) - \widehat{\boldsymbol{\beta}}_{j,i}'(\widehat{\boldsymbol{\gamma}}_{j,1} + \widehat{\mathbf{u}}_{j,t})\right) d_{j,t},$$

where

$$\overline{\epsilon}_{j,i} = \frac{1}{T_j}\sum_{t=1}^{T}\epsilon_{i,t} d_{j,t}.$$

Also, applying Theorem 3.4 and Corollary 3.2 in Massacci (2017) (recalling also (24)) yields

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{\beta}_{j,i} - \widehat{\boldsymbol{\beta}}_{j,i}\right\|^2 = o_P(1), \tag{D.10}$$

$$\left\|\frac{1}{T_j}\sum_{t=1}^{T}(\mathbf{u}_{j,t} - \widehat{\mathbf{u}}_{j,t})\right\|^2 \leq \frac{T}{T_j}\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{u}_{j,t} - \widehat{\mathbf{u}}_{j,t}\right\|^2 = o_P(1), \tag{D.11}$$

having omitted rotation matrices for the sake of the notation. Note now that

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\widehat{e}_{j,i}^2 &= \frac{1}{N}\sum_{i=1}^{N}\alpha_{j,i}^2 \\ &+ \frac{1}{N}\sum_{i=1}^{N}\left(\overline{\epsilon}_{j,i} + (\gamma_{j,0} - \widehat{\gamma}_{j,0}) + \frac{1}{T_j}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}'(\boldsymbol{\gamma}_{j,1} + \mathbf{u}_{j,t}) - \widehat{\boldsymbol{\beta}}_{j,i}'(\widehat{\boldsymbol{\gamma}}_{j,1} + \widehat{\mathbf{u}}_{j,t})\right) d_{j,t}\right)^2 \\ &+ \frac{2}{N}\sum_{i=1}^{N}\alpha_{j,i}\left(\overline{\epsilon}_{j,i} + (\gamma_{j,0} - \widehat{\gamma}_{j,0}) + \frac{1}{T_j}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}'(\boldsymbol{\gamma}_{j,1} + \mathbf{u}_{j,t}) - \widehat{\boldsymbol{\beta}}_{j,i}'(\widehat{\boldsymbol{\gamma}}_{j,1} + \widehat{\mathbf{u}}_{j,t})\right) d_{j,t}\right).\end{aligned}$$

The LLN entails

$$\frac{1}{N}\sum_{i=1}^{N}\alpha_{j,i}^{2}=\sigma_{\alpha,j}^{2}+o_{P}\left(1\right);$$

upon showing that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\overline{\epsilon}_{j,i}+\left(\gamma_{j,0}-\widehat{\gamma}_{j}\right)+\frac{1}{T_{j}}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}^{\prime}\left(\boldsymbol{\gamma}_{j,1}+\mathbf{u}_{j,t}\right)-\widehat{\boldsymbol{\beta}}_{j,i}^{\prime}\left(\widehat{\boldsymbol{\gamma}}_{j,1}+\widehat{\mathbf{u}}_{j,t}\right)\right)d_{j,t}\right)^{2}=o_{P}\left(1\right),$$

and using (repeatedly) the Cauchy-Schwartz inequality, the lemma follows. Now

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\left(\overline{\epsilon}_{j,i}+\left(\gamma_{j,0}-\widehat{\gamma}_{j}\right)+\frac{1}{T_{j}}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}^{\prime}\left(\boldsymbol{\gamma}_{j,1}+\mathbf{u}_{j,t}\right)-\widehat{\boldsymbol{\beta}}_{j,i}^{\prime}\left(\widehat{\boldsymbol{\gamma}}_{j,1}+\widehat{\mathbf{u}}_{j,t}\right)\right)d_{j,t}\right)^{2}\\ \leq\ & c_{0}\left(\frac{1}{N}\sum_{i=1}^{N}\overline{\epsilon}_{j,i}^{2}+\left(\gamma_{j,0}-\widehat{\gamma}_{j}\right)^{2}+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T_{j}}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}^{\prime}\left(\boldsymbol{\gamma}_{j,1}+\mathbf{u}_{j,t}\right)-\widehat{\boldsymbol{\beta}}_{j,i}^{\prime}\left(\widehat{\boldsymbol{\gamma}}_{j,1}+\widehat{\mathbf{u}}_{j,t}\right)\right)\right)^{2}d_{j,t}\right).\end{aligned}$$

By (A.24), $\gamma_{j,0}-\widehat{\gamma}_{j}=o_{P}\left(1\right)$. Also, combining Theorem 2 and (D.10)-(D.11), it is easy to see that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T_{j}}\sum_{t=1}^{T}\left(\boldsymbol{\beta}_{j,i}^{\prime}\left(\boldsymbol{\gamma}_{j,1}+\mathbf{u}_{j,t}\right)-\widehat{\boldsymbol{\beta}}_{j,i}^{\prime}\left(\widehat{\boldsymbol{\gamma}}_{j,1}+\widehat{\mathbf{u}}_{j,t}\right)\right)\right)^{2}d_{j,t}=o_{P}\left(1\right).$$

Finally we have

$$\max_{1\leq i\leq N}E\left(\overline{\epsilon}_{j,i}^{2}\right)=\max_{1\leq i\leq N}\frac{1}{T_{j}^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\epsilon_{i,t}\epsilon_{i,s}d_{j,t}d_{j,s}\right)\leq c_{0}T_{j}^{-1},$$

using Assumption 4*(ii)*. Hence

$$\frac{1}{N}\sum_{i=1}^{N}\overline{\epsilon}_{j,i}^{2}=o_{P}\left(1\right).$$

Putting all together, the desired result follows. □

Proof of Theorem A.4. We present the full version of the proof of (A.24) only; the other two results follow from similar arguments and, where possible, we omit the details to avoid repetition. We begin by showing that the estimation error $\widehat{P}_{j}-P_{j}^{0}$ is negligible. To this end, we assume without loss of generality that $\widehat{P}_{j}\geq P_{j}^{0}$, and we omit dependence on θ or θ^{0} whenever possible.

Let

$$\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\widehat{P}_j}\widehat{\mathbf{\Gamma}}_j = \widehat{\mathbf{H}}_{jj}^{*\mathbf{\Gamma}}\widehat{\mathbf{\Gamma}}_j^* + \widetilde{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}\widetilde{\mathbf{\Gamma}}_j,$$

where $\widehat{\mathbf{H}}_{jj}^{*\mathbf{\Gamma}}$ is $\left(P_j^0+1\right)\times\left(P_j^0+1\right)$, $\widehat{\mathbf{\Gamma}}_j^*$ is $\left(P_j^0+1\right)\times 1$, $\widetilde{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}$ is $\left(P_j^0+1\right)\times\left(\widehat{P}_j-P_j^0\right)$ and $\widetilde{\mathbf{\Gamma}}_j$ is $\left(\widehat{P}_j-P_j^0\right)\times 1$, defined such that $\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\widehat{P}_j}=\left[\widehat{\mathbf{H}}_{jj}^{*\mathbf{\Gamma}}|\widetilde{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}\right]$ and $\widehat{\mathbf{\Gamma}}_j=\left(\widehat{\mathbf{\Gamma}}_j^{*\prime},\widetilde{\mathbf{\Gamma}}_j^{\prime}\right)^{\prime}$. We begin by showing that

$$\left\|\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\widehat{P}_j}\right\|_F = O_{P^*}(1). \tag{D.12}$$

Note that

$$\left\|\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\widehat{P}_j}\right\|_F \leq \left\|\frac{\mathbf{U}_j(\theta^0)\mathbf{U}_j(\theta)'}{T}\right\|_F \left\|\frac{\mathbf{B}_j'(\theta^0)\widehat{\mathbf{B}}_j(\theta)}{N}\right\|_F \left\|\widehat{\mathbf{V}}_j(\theta)^{-1}\right\|_F;$$

the first term is bounded by Assumption 3*(i)*; also note that, using the triangular inequality

$$\begin{aligned}
\left\|\frac{\mathbf{B}_j'(\theta^0)\widehat{\mathbf{B}}_j(\theta)}{N}\right\|_F &\leq \left\|\frac{\mathbf{B}_j'(\theta^0)\mathbf{B}_j(\theta^0)}{N}\right\|_F + \left\|\frac{\mathbf{B}_j'\left(\widehat{\mathbf{B}}_j(\theta)-\mathbf{B}_j(\theta^0)\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\widehat{P}_j}\right)}{N}\right\|_F \\
&\leq \left\|\frac{\mathbf{B}_j'(\theta^0)\mathbf{B}_j(\theta^0)}{N}\right\|_F + \left\|\frac{\mathbf{B}_j}{N^{1/2}}\right\|_F \left\|\frac{\widehat{\mathbf{B}}_j(\theta)-\mathbf{B}_j(\theta^0)\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\widehat{P}_j}}{N^{1/2}}\right\|_F.
\end{aligned}$$

The first term is bounded by virtue of Assumption 3*(iv)*, whereas the second one can be shown to be dominated by using the same arguments as in footnote 5 in Bai (2003). Indeed, Finally, note that

$$\begin{aligned}
& P^*\left(\left\|\widehat{\mathbf{V}}_j(\theta)^{-1}\right\|_F \geq M\right) \\
= \; & P^*\left(\left\|\widehat{\mathbf{V}}_j(\theta)^{-1}\right\|_F \geq M|\widehat{P}_j=P_j^0\right) + P^*\left(\left\|\widehat{\mathbf{V}}_j(\theta)^{-1}\right\|_F \geq M|\widehat{P}_j\neq P_j^0\right) \\
\leq \; & P^*\left(\left\|\widehat{\mathbf{V}}_j(\theta)^{-1}\right\|_F \geq M|\widehat{P}_j=P_j^0\right) + P^*\left(\widehat{P}_j\neq P_j^0\right) \\
= \; & o_{P^*}(1),
\end{aligned}$$

by Theorem 1, with $g^{\left(P_j^0\right)}\left(\widehat{\mathbf{\Sigma}}_{j,\widetilde{\mathbf{R}}}(\theta)\right)>0$ implied by Lemma C.10. By similar passages, it is easy to see that

$$\left\|\widehat{\mathbf{\Gamma}}_j\right\|_F = O_{P^*}(1). \tag{D.13}$$

We now show that

$$\left\|\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\widetilde{\mathbf{\Gamma}}_{j}\right\|_{F} = o_{P^{*}}\left(1\right), \tag{D.14}$$

for almost all realizations of the sample. Indeed, $\left\|\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\widetilde{\mathbf{\Gamma}}_{j}\right\|_{F} \leq \left\|\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\right\|_{F}\left\|\widetilde{\mathbf{\Gamma}}_{j}\right\|_{F}$, and $\left\|\widetilde{\mathbf{\Gamma}}_{j}\right\|_{F}$ is bounded by (D.13). Also

$$\left\|\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\right\|_{F} = tr\left(\widetilde{\mathbf{H}}^{\mathbf{\Gamma}\prime}_{jj}\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\right) \leq \left(\widehat{P}_{j} - P^{0}_{j}\right) g^{(1)}\left(\widetilde{\mathbf{H}}^{\mathbf{\Gamma}\prime}_{jj}\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\right),$$

with $\widehat{P}_{j} - P^{0}_{j} = o_{P^{*}}\left(1\right)$ by Theorem 1 and $g^{(1)}\left(\widetilde{\mathbf{H}}^{\mathbf{\Gamma}\prime}_{jj}\widetilde{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\right)$ bounded by (D.12). Therefore, $\widehat{\mathbf{H}}^{\mathbf{\Gamma}\widehat{P}_{j}}_{jj}\widehat{\mathbf{\Gamma}}_{j} = \widehat{\mathbf{H}}^{*\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*}_{j} + o_{P^{*}}\left(1\right)$. Turning to $\widehat{\mathbf{H}}^{*\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*}_{j}$, let $\widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j}$ be the estimate obtained when using P^{0}_{j}, and consider

$$\widehat{\mathbf{H}}^{*\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*}_{j} = \left(\widehat{\mathbf{H}}^{*\mathbf{\Gamma}}_{jj} \pm \widehat{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\right)\left(\widehat{\mathbf{\Gamma}}^{*}_{j} \pm \widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j}\right).$$

By the same arguments as in footnote 5 in Bai (2003), it is easy to see that $\left\|\widehat{\mathbf{\Gamma}}^{*}_{j} - \widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j}\right\|_{F} = o_{P^{*}}\left(1\right)$; then it follows immediately that

$$\widehat{\mathbf{H}}^{*\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*}_{j} = \widehat{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j} + o_{P^{*}}\left(1\right),$$

which, combined with (D.14), yields

$$\widehat{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}_{j} = \widehat{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j} + o_{P^{*}}\left(1\right). \tag{D.15}$$

We now conclude the proof of (A.24) by showing that

$$\widehat{\mathbf{H}}^{\mathbf{\Gamma}}_{jj}\widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j} = \mathbf{\Gamma}_{j} + o_{P}\left(1\right).$$

Let $\widehat{\mathbf{X}}^{P^{0}_{j}}_{j}(\theta_{0}) = \left[\iota_{N}, \widehat{\mathbf{B}}^{P^{0}_{j}}_{j}(\theta_{0})\right]$. By the same logic as in the previous passages, and using Theorem 3.4 in Massacci (2017), we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{\Gamma}}^{*P^{0}_{j}}_{j,t}\widehat{d}_{j,t} &= \left(\widehat{\mathbf{X}}^{\widehat{P}_{j}\prime}_{j}(\widehat{\theta})\widehat{\mathbf{X}}^{\widehat{P}_{j}}_{j}(\widehat{\theta})\right)^{-1}\widehat{\mathbf{X}}^{\widehat{P}_{j}\prime}_{j}(\widehat{\theta})\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{t}\widehat{d}_{j,t} \\
&= \left(\widehat{\mathbf{X}}^{P^{0}_{j}\prime}_{j}(\theta_{0})\widehat{\mathbf{X}}^{P^{0}_{j}}_{j}(\theta_{0})\right)^{-1}\widehat{\mathbf{X}}^{P^{0}_{j}\prime}_{j}(\theta_{0})\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_{t}\widehat{d}_{j,t} + o_{P^{*}}\left(1\right) + o_{P}\left(1\right).
\end{aligned}$$

Also

$$\left(\widehat{\mathbf{X}}_j^{P_j^0\prime}(\theta_0)\widehat{\mathbf{X}}_j^{P_j^0}(\theta_0)\right)^{-1}\widehat{\mathbf{X}}_j^{P_j^0\prime}(\theta_0)\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\widehat{d}_{j,t} \tag{D.16}$$

$$= \left(\widehat{\mathbf{X}}_j^{P_j^0\prime}(\theta_0)\widehat{\mathbf{X}}_j^{P_j^0}(\theta_0)\right)^{-1}\widehat{\mathbf{X}}_j^{P_j^0\prime}(\theta_0)\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t d_{j,t}\left(\theta^0\right)$$

$$+\left(\widehat{\mathbf{X}}_j^{P_j^0\prime}(\theta_0)\widehat{\mathbf{X}}_j^{P_j^0}(\theta_0)\right)^{-1}\widehat{\mathbf{X}}_j^{P_j^0\prime}(\theta_0)\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\left(\widehat{d}_{j,t}-d_{j,t}\left(\theta^0\right)\right)$$

$$= I + II.$$

Noting that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\left(\widehat{d}_{j,t}-d_{j,t}\left(\theta^0\right)\right) \leq \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{R}_t\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{d}_{j,t}-d_{j,t}\left(\theta^0\right)\right)^2\right)^{1/2},$$

and using Theorem 3.4 in Massacci (2017), it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t\left(\widehat{d}_{j,t}-d_{j,t}\left(\theta^0\right)\right) = O_P\left(T^{-1}\right),$$

so that, in equation (D.16), we have $II = O_P\left(T^{-1}\right)$. Also, by Corollary 3.1(a) in Massacci (2017), $\left\|\widehat{\mathbf{B}}_j^{P_j^0}-\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\right\|^2 = O_P(NC_{N,T}^{-2})$, which in turn implies

$$\widehat{\mathbf{X}}_j^{P_j^0}(\theta_0)-(\boldsymbol{\iota}_N,\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}) = \widehat{\mathbf{X}}_j^{P_j^0}(\theta_0)-\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}} = O_P(NC_{N,T}^{-2});$$

thus

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\Gamma}}_{j,t}^{*P_j^0}\widehat{d}_{j,t} = \left(\frac{\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}\prime}\mathbf{X}_j^{0\prime}}{N}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t d_{j,t}\left(\theta^0\right)\right)$$

$$+O_P(C_{N,T}^{-2})+O_P\left(T^{-1}\right)+o_{P^*}(1).$$

Let $\alpha = (\alpha_1, ..., \alpha_N)'$. Recalling (4), it follows that

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{\Gamma}}_{j,t}^{*P_j^0}\widehat{d}_{j,t} - \frac{T_j}{T}\left(\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}\right)^{-1}\mathbf{\Gamma}_j^0 \\
= & \frac{T_j}{T}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\alpha_j}{N}\right) \\
& + \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}}{N}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{B}_j^0\mathbf{u}_{j,t}^0 d_{j,t}\left(\theta^0\right)\right) \\
& + \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}}{N}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t d_{j,t}\left(\theta^0\right)\right) \\
& + O_P\left(C_{N,T}^{-2}\right) + O_P\left(T^{-1}\right) + o_{P^*}(1) \\
= & I_a + I_b + I_c + O_P\left(C_{N,T}^{-2}\right) + O_P\left(T^{-1}\right) + o_{P^*}(1)
\end{aligned}$$

Assumptions 3*(iv)* and 7*(i)* immediately yields $I_a = O_P\left(\frac{T_j}{N^{1/2}T}\right)$. Consider I_c; it holds that

$$\begin{aligned}
& E\left(\frac{1}{NT}\sum_{t=1}^{T}\mathbf{X}_j^{0\prime}\epsilon_t d_{j,t}\left(\theta^0\right)\right)^2 \\
= & \frac{1}{(NT)^2}\sum_{i,k=1}^{N}\sum_{t,s=1}^{T}X_{i,j}^0 X_{k,j}^{0\prime}E\left(\epsilon_{i,t}\epsilon_{j,s}d_{j,t}\left(\theta^0\right)d_{j,s}\left(\theta^0\right)\right) \\
\leq & \frac{1}{(NT)^2}\max_{1\leq i\leq N}\left\|X_{i,j}^0\right\|^2\sum_{i,k=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\epsilon_{i,t}\epsilon_{j,s}d_{j,t}\left(\theta^0\right)d_{j,s}\left(\theta^0\right)\right)\right| \\
\leq & c_0\frac{N}{(NT)^2}\max_{1\leq i\leq N}\sum_{k=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\epsilon_{i,t}\epsilon_{j,s}d_{j,t}\left(\theta^0\right)d_{j,s}\left(\theta^0\right)\right)\right| \leq \frac{c_0}{NT},
\end{aligned}$$

by Assumptions 3*(iii)* and 2*(iv)*; so, putting all together and using Assumption 3*(iv)*, it follows that $I_c = O_P\left(\frac{1}{\sqrt{NT}}\right)$. Turning to I_b, (C.29) yields

$$\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{X}_j^0\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{B}_j^0}{N}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_{j,t}^0 d_{j,t}\left(\theta^0\right)\right) = O_P\left(\frac{1}{\sqrt{T}}\right) \quad \text{(D.17)}$$

Putting all together, it follows that

$$\frac{1}{T_j}\sum_{t=1}^{T}\widehat{\mathbf{\Gamma}}_{j,t}^{*P_j^0}\widehat{d}_{j,t}-\left(\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}\right)^{-1}\mathbf{\Gamma}_j^0=O_P\left(\frac{1}{N^{1/2}}\right)+O_P\left(\frac{T^{1/2}}{T_j}\right)+O_P\left(\frac{T}{T_jN}\right)+O_P\left(T^{-1}\right)+o_{P^*}\left(1\right); \tag{D.18}$$

the desired result now follows immediately. We now turn to showing (A.25). We assume that P_j^0 is known; the estimation error $\widehat{P}_j-P_j^0$ can be shown to be negligible on account of the same arguments as above. We have

$$\widehat{\Lambda}_j-\Lambda_j^0\widehat{\mathbf{H}}_{jj}=\left(\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}\widehat{d}_{j,t}\right)\left(\sum_{t=1}^{T}\widehat{\mathbf{u}}_{j,t}^{P_j^0}\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}\widehat{d}_{j,t}\right)^{-1}.$$

Note that, by Corollary 3.2 in Massacci (2017), it is easy to see that

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{u}}_{j,t}^{P_j^0}\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}\widehat{d}_{j,t}=\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{H}}_{jj}^{-1}\mathbf{u}_{j,t}^{P_j^0}\mathbf{u}_{j,t}^{P_j^0\prime}\left(\widehat{\mathbf{H}}_{jj}'\right)^{-1}d_{j,t}\left(\theta^0\right)+o_P\left(1\right). \tag{D.19}$$

Also

$$\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}\widehat{d}_{j,t}=\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}d_{j,t}\left(\theta^0\right)+\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}\left(\widehat{d}_{j,t}-d_{j,t}\left(\theta^0\right)\right)=I+II. \tag{D.20}$$

Consider II; we have

$$II\leq\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\widetilde{\mathbf{g}}_t\right\|^4\right)^{1/4}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{\mathbf{u}}_{j,t}^{P_j^0}\right\|^4\right)^{1/4}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{d}_{j,t}-d_{j,t}\left(\theta^0\right)\right)^2\right)^{1/2}; \tag{D.21}$$

using Assumptions 3*(i)* and 8*(i)*, and using Theorem 3.4 in Massacci (2017), it follows that $II=O_P\left(N^{-\eta}T^{-1}\right)$. Turning to I, by the definition of $\widetilde{\mathbf{g}}_t\left(\widehat{\theta}\right)$ it holds that

$$I=\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{g}_t-\mathbf{a}\right)\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}d_{j,t}\left(\theta^0\right)+\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{g}_t-\mathbf{a}\right)\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{u}}_{j,t}^{P_j^0\prime}d_{j,t}\left(\theta^0\right)=I_a+I_b.$$

Using (C.29) and Assumption 8*(ii)*, it follows that $I_b = O_P\left(T^{-1}\right)$.

$$I_a = \frac{1}{T}\sum_{t=1}^{T}\Lambda_j^0 \mathbf{u}_{j,t}^0 \hat{\mathbf{u}}_{j,t}^{P_j^0\prime} d_{j,t}\left(\theta^0\right) + \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t \hat{\mathbf{u}}_{j,t}^{P_j^0\prime} d_{j,t}\left(\theta^0\right) = I_{a,1} + I_{a,2}.$$

Now

$$\begin{aligned} I_{a,1} &= \Lambda_j^0 \widehat{\mathbf{H}}_{jj} \widehat{\mathbf{H}}_{jj}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0\prime}\right)\left(\widehat{\mathbf{H}}_{jj}^{\prime}\right)^{-1} d_{j,t}\left(\theta^0\right) \\ &\quad + \Lambda_j^0 \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_{j,t}^0\left(\hat{\mathbf{u}}_{j,t}^{P_j^0} - \widehat{\mathbf{H}}_{jj}^{-1}\mathbf{u}_{j,t}^0\right)^{\prime} d_{j,t}\left(\theta^0\right) \\ &= I_{a,1,1} + I_{a,1,2}. \end{aligned}$$

Following the same passages as in the proof of Lemma B.2 in Bai (2003), it can be shown that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_{j,t}^0\left(\hat{\mathbf{u}}_{j,t}^{P_j^0} - \widehat{\mathbf{H}}_{jj}^{-1}\mathbf{u}_{j,t}^0\right)^{\prime} d_{j,t}\left(\theta^0\right) = O_P\left(C_{N,T}^{-2}\right), \tag{D.22}$$

so that $I_{a,1,2} = O_P\left(C_{N,T}^{-2}\right)$. Also

$$I_{a,2} = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t \mathbf{u}_{j,t}^{0\prime}\right)\left(\widehat{\mathbf{H}}_{jj}^{\prime}\right)^{-1} d_{j,t}\left(\theta^0\right) + \frac{1}{T}\sum_{t=1}^{T}\mathbf{e}_t\left(\hat{\mathbf{u}}_{j,t}^{P_j^0} - \widehat{\mathbf{H}}_{jj}^{-1}\mathbf{u}_{j,t}^0\right)^{\prime} d_{j,t}\left(\theta^0\right) = I_{a,2,1} + I_{a,2,2}.$$

Assumption 8*(ii)* entails that $I_{a,2,1} = O_P\left(T^{-1/2}\right)$. Finally, by the same token as for (D.22), $I_{a,2,2} = O_P\left(C_{N,T}^{-2}\right)$, whence $I_{a,2} = O_P\left(T^{-1/2}\right) + O_P\left(C_{N,T}^{-2}\right)$. Putting all together, we obtain

$$\widehat{\Lambda}_j - \Lambda_j^0 \widehat{\mathbf{H}}_{jj} = O_P\left(\frac{1}{T^{1/2}}\right) + O_P\left(C_{N,T}^{-2}\right) + O_P\left(T^{-1}\right). \tag{D.23}$$

□

Proof of Theorem A.5. Consider (A.26). Upon inspecting (D.18), under (24) and (25), and using Lemma C.8, it is immediate to see that the limiting behaviour of $T^{1/2}\left(\widehat{\boldsymbol{\Gamma}}_j^{*P_j^0} - \left(\widehat{\mathbf{H}}_{jj}^{\boldsymbol{\Gamma}}\right)^{-1}\boldsymbol{\Gamma}_j^0\right)$ is driven by

$$\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{X}_j^{0}\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N}\right)^{-1}\left[\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\mathbf{B}_j^{0}}{N}\right)\left(\frac{T}{T_j}\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{u}_{j,t}^{0}d_{j,t}\left(\theta^0\right)\right)+\frac{T_j}{T}\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}\prime}\mathbf{X}_j^{0\prime}\alpha_j}{N}\right)\right].$$

The desired result now follows immediately from (C.30) and Assumption 7. Turning to (A.27), combining (D.23) and (25), it follows that the limiting law of $T^{1/2}\left(\widehat{\mathbf{\Lambda}}_j-\mathbf{\Lambda}_j^0\widehat{\mathbf{H}}_{jj}\right)$ is driven by

$$\begin{aligned}
&\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{u}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}\right)^{-1}d_{j,t}\left(\theta^0\right)\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{H}}_{jj}\right)^{-1}\mathbf{u}_{j,t}^{0}\mathbf{u}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}'\right)^{-1}d_{j,t}\left(\theta^0\right)\right)^{-1}\\
=\;&\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{u}_{j,t}^{0\prime}d_{j,t}\left(\theta^0\right)\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_{j,t}^{0}\mathbf{u}_{j,t}^{0\prime}d_{j,t}\left(\theta^0\right)\right)^{-1}\widehat{\mathbf{H}}_{jj},
\end{aligned}$$

and using Assumption 9*(ii)* the desired result follows immediately. □

Proof of Theorem A.6. The proof makes appeal to several arguments which have already been used in the paper and therefore, for the sake of a concise discussion, we omit passages when this does not give rise to ambiguity.

We begin by noting that, by the Frisch-Waugh-Lovell theorem, it holds that

$$\widehat{\gamma}_{j,0}=\left(i_N'\widehat{\mathbf{M}}_{B,j}i_N\right)^{-1}\left(i_N'\widehat{\mathbf{M}}_{B,j}\overline{\mathbf{R}}_j\left(\widehat{\theta}\right)\right),$$

having defined

$$\widehat{\mathbf{M}}_{B,j}=\widehat{\mathbf{B}}_j\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1}\widehat{\mathbf{B}}_j',$$

for $j=\mathcal{D},\mathcal{S}$ - note that we use $\widehat{\mathbf{B}}_j$ in lieu of $\widehat{\mathbf{B}}_j^{P_j}$ to make the notation less burdensome. Thus, on account of equation (5), it holds that

$$\widehat{\gamma}_{j,0}=\gamma_{j,0}+\left(i_N'\widehat{\mathbf{M}}_{B,j}i_N\right)^{-1}\left(i_N'\widehat{\mathbf{M}}_{B,j}\alpha_j\right)+\left(i_N'\widehat{\mathbf{M}}_{B,j}i_N\right)^{-1}\left(i_N'\widehat{\mathbf{M}}_{B,j}\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t}\right). \tag{D.24}$$

We begin by noting that, using Corollary 3.1(a) in Massacci (2017), it is easy to see that

$$\left\| \widehat{\mathbf{M}}_{B,j} - \mathbf{M}^0_{B,j} \right\|^2 = O_P\left(N C^{-2}_{N,T} \right),$$

so that

$$\frac{i'_N \widehat{\mathbf{M}}_{B,j} i_N}{N} = \frac{i'_N \mathbf{M}^0_{B,j} i_N}{N} + O_P\left(C^{-2}_{N,T} \right) = O_P(1). \tag{D.25}$$

Also, it holds that

$$E\left(\widehat{\mathbf{B}}_j - \mathbf{B}^0_j \widehat{\mathbf{H}}_{jj} \right)' \alpha_j \alpha'_j \left(\widehat{\mathbf{B}}_j - \mathbf{B}^0_j \widehat{\mathbf{H}}_{jj} \right) = \sigma^2_{\alpha,j} E\left(\widehat{\mathbf{B}}_j - \mathbf{B}^0_j \widehat{\mathbf{H}}_{jj} \right)' \left(\widehat{\mathbf{B}}_j - \mathbf{B}^0_j \widehat{\mathbf{H}}_{jj} \right) = O\left(N C^{-2}_{N,T} \right),$$

again by Corollary 3.1(a) in Massacci (2017); this yields the (non sharp) bound

$$i'_N \left(\widehat{\mathbf{M}}_{B,j} - \mathbf{M}^0_{B,j} \right) \alpha_j = O_P\left(N^{1/2} C^{-1}_{N,T} \right). \tag{D.26}$$

Finally, note that Assumption 7*(i)* entails that

$$N^{1/2} \left(i'_N \mathbf{M}^0_{B,j} i_N \right)^{-1} \left(i'_N \mathbf{M}^0_{B,j} \alpha_j \right) = O_P(1). \tag{D.27}$$

Putting (D.25)-(D.27) together, we obtain

$$N^{1/2} \left(i'_N \widehat{\mathbf{M}}_{B,j} i_N \right)^{-1} \left(i'_N \widehat{\mathbf{M}}_{B,j} \alpha_j \right) = N^{1/2} \left(i'_N \mathbf{M}^0_{B,j} i_N \right)^{-1} \left(i'_N \mathbf{M}^0_{B,j} \alpha_j \right) + o_P(1). \tag{D.28}$$

Consider now

$$i'_N \widehat{\mathbf{M}}_{B,j} \frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} = i'_N \frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} + i'_N \widehat{\mathbf{B}}_j \left(\widehat{\mathbf{B}}'_j \widehat{\mathbf{B}}_j \right)^{-1} \widehat{\mathbf{B}}'_j \frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} = I + II. \tag{D.29}$$

Using Assumption 2*(ii)* and equation (C.23), it is easy to see that

$$\begin{aligned} E\left(i_N'\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t}\right)^2 &= T_j^{-2}E\left(\sum_{i,k=1}^{N}\sum_{t,s=1}^{T}\epsilon_{i,t}\epsilon_{k,s}\widehat{d}_{j,t}\widehat{d}_{j,s}\right) \\ &\leq NT_j^{-2}\max_{1\leq k\leq N}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\left|E\left(\epsilon_{i,t}\epsilon_{k,s}\widehat{d}_{j,t}\widehat{d}_{j,s}\right)\right| \leq c_0\frac{N}{T_j}, \end{aligned}$$

so that

$$i_N'\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} = O_P\left(N^{1/2}T_j^{-1/2}\right). \tag{D.30}$$

In order to study II in (D.29), note that

$$\begin{aligned} & i_N'\widehat{\mathbf{B}}_j\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1}\widehat{\mathbf{B}}_j'\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} \qquad \text{(D.31)} \\ &= i_N'\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1}\widehat{\mathbf{H}}_{jj}'\mathbf{B}_j^{0\prime}\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} + i_N'\left(\widehat{\mathbf{B}}_j-\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\right)\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1}\widehat{\mathbf{H}}_{jj}'\mathbf{B}_j^{0\prime}\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} \\ &+ i_N'\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1}\left(\widehat{\mathbf{B}}_j-\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\right)'\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} \\ &+ i_N'\left(\widehat{\mathbf{B}}_j-\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\right)\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1}\left(\widehat{\mathbf{B}}_j-\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\right)'\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} \\ &= I+II+III+IV. \end{aligned}$$

By construction, $\left(\widehat{\mathbf{B}}_j'\widehat{\mathbf{B}}_j\right)^{-1} = O_p\left(N^{-1}\right)$, and $i_N'\mathbf{B}_j^0 = O_P\left(N\right)$. Using Assumption 2*(ii)* and equation (C.23), similarly to (D.30), it also follows that

$$\mathbf{B}_j^{0\prime}T_j^{-1}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} = O_P\left(N^{1/2}T_j^{-1/2}\right), \tag{D.32}$$

so that $I = O_P\left(N^{-1/2}T_j^{-1/2}\right)$. Consider now

$$\begin{aligned}
&\mathbf{B}_j^0 - \widehat{\mathbf{B}}_j\widehat{\mathbf{H}}_{jj}^{-1} \\
&= T^{-1}\mathbf{B}_j^0\widehat{\mathbf{H}}_{jj}\left(\widehat{\mathbf{H}}_{jj}^{-1}\mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t}\right)\widehat{\mathbf{U}}_{j,t}'\widehat{\mathbf{H}}_{jj}^{-1} + T^{-1}\epsilon_j\left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right)\widehat{\mathbf{H}}_{jj}^{-1} \\
&+ T^{-1}\epsilon_j\mathbf{U}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\widehat{\mathbf{H}}_{jj}^{-1}, \qquad \text{(D.33)}
\end{aligned}$$

where ϵ_j is an $N \times T$ matrix with columns $\epsilon_t\widehat{d}_{j,t}$. Using exactly the same arguments as in the proofs of Lemmas B.1 and B.3 in Bai (2003), it holds that $\left\|T^{-1}\left(\widehat{\mathbf{H}}_{jj}^{-1}\mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t}\right)\widehat{\mathbf{U}}_{j,t}'\widehat{\mathbf{H}}\right\| = O_P\left(C_{N,T}^{-2}\right)$ and $\left\|T^{-1}\epsilon_j\left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right)\widehat{\mathbf{H}}\right\| = O_P\left(C_{N,T}^{-2}\right)$. Hence, it follows that

$$i_N'\mathbf{B}_j^0\left(T^{-1}\widehat{\mathbf{H}}_{jj}\left(\widehat{\mathbf{H}}_{jj}^{-1}\mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t}\right)\widehat{\mathbf{U}}_{j,t}'\right)\widehat{\mathbf{H}}_{jj}^{-1} = O_P\left(NC_{N,T}^{-2}\right). \qquad \text{(D.34)}$$

Also

$$\begin{aligned}
i_N'T^{-1}\epsilon_j\left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right) &= \sum_{i=1}^N T^{-1}\sum_{t=1}^T \epsilon_{j,i,t}\widehat{d}_{j,t}\left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right) \\
&\leq \sum_{i=1}^N \left\|T^{-1}\sum_{t=1}^T \epsilon_{j,i,t}\widehat{d}_{j,t}\left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right)\right\|;
\end{aligned}$$

again by using the same arguments as in the proof of Lemma B.1 in Bai (2003), it can be shown that

$$E\left\|T^{-1}\sum_{t=1}^T \epsilon_{j,i,t}\widehat{d}_{j,t}\left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right)\right\| \leq c_0C_{N,T}^{-2},$$

which yields that

$$i_N'T^{-1}\epsilon_j\left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0\prime}\left(\widehat{\mathbf{H}}_{jj}^{-1}\right)'\right) = O_P\left(NC_{N,T}^{-2}\right). \qquad \text{(D.35)}$$

Finally, note that

$$\begin{aligned}
E\left\|i_N'T^{-1}\epsilon_j\mathbf{U}_{j,t}^{0\prime}\right\|^2 &= \sum_{i=1}^N\sum_{j=1}^N T^{-2}\sum_{t=1}^T\sum_{s=1}^T E\left(\epsilon_{i,t}\epsilon_{j,s}\widehat{d}_{j,t}\widehat{d}_{j,s}\right)E\left\|\mathbf{u}_{j,t}^{0\prime}\mathbf{u}_{j,s}^0\right\| \\
&\leq \left(\max_{1\leq t\leq T} E\left\|\mathbf{u}_{j,t}^0\right\|^2\right)\frac{N}{T}\max_{1\leq i\leq N}\sum_{k=1}^N T^{-1}\sum_{t=1}^T\sum_{s=1}^T\left|E\left(\epsilon_{i,t}\epsilon_{k,s}\widehat{d}_{j,t}\widehat{d}_{j,s}\right)\right| \leq c_0\frac{N}{T},
\end{aligned}$$

after Assumptions 2*(ii)* and 3*(i)*, and using (C.23), which entails that

$$i_N' T^{-1} \epsilon_j \mathbf{U}_{j,t}^{0\prime} = O_P \left(N^{1/2} T^{-1/2} \right). \tag{D.36}$$

Putting (D.34)-(D.36) together, it follows that

$$i_N' \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right) = O_P \left(N C_{N,T}^{-2} \right). \tag{D.37}$$

Thus, considering (D.31), (D.32) and (D.37) immediately entail that $II = O_P \left(N^{1/2} T_j^{-1/2} C_{N,T}^{-2} \right)$. Turning to III, note that, using (D.33)

$$\begin{aligned}
&\left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right)' \left(\frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} \right) \\
&= \left(\mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} T^{-1} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}_{j,t}' \widehat{\mathbf{H}}_{jj}^{-1} \right)' \left(\frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} \right) \\
&+ \left(T^{-1} \epsilon_j \left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0\prime} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \widehat{\mathbf{H}}_{jj}^{-1} \right)' \left(\frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} \right) \\
&+ \left(T^{-1} \epsilon_j \mathbf{U}_{j,t}^{0\prime} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \widehat{\mathbf{H}}_{jj}^{-1} \right)' \left(\frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} \right) \\
&= I + II + III.
\end{aligned} \tag{D.38}$$

Using (D.32), it follows that

$$I = \left(\widehat{\mathbf{H}}_{jj} T^{-1} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}_{j,t}' \widehat{\mathbf{H}}_{jj}^{-1} \right) \left({}' \mathbf{B}_j^{0\prime} \frac{1}{T_j} \sum_{t=1}^{T} \epsilon_t \widehat{d}_{j,t} \right) = O_P \left(N^{1/2} T_j^{-1/2} C_{N,T}^{-2} \right);$$

further, we have

$$\sum_{i=1}^{N} \left(\frac{1}{T_j} \sum_{t=1}^{T} \epsilon_{i,t} \widehat{d}_{j,t} \right) T^{-1} \sum_{t=1}^{T} \epsilon_{i,t} \widehat{d}_{j,t} \left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0\prime} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) = O_P \left(N T_j^{-1/2} C_{N,T}^{-2} \right)$$

$$\sum_{i=1}^{N} \left(\frac{1}{T_j} \sum_{t=1}^{T} \epsilon_{i,t} \widehat{d}_{j,t} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} \mathbf{u}_{j,t}' \widehat{d}_{j,t} \right) = O_P \left(N T^{-1} \right),$$

whence $II = O_P\left(NT_j^{-1/2}C_{N,T}^{-2}\right)$ and $III = O_P\left(NT^{-1}\right)$. Finally, using the same arguments as above, it holds that IV is dominated. Putting all together, we conclude that, in (D.29), $II = O_P\left(NT^{-1}\right)$, whence

$$i_N'\widehat{\mathbf{M}}_{B,j}\frac{1}{T_j}\sum_{t=1}^{T}\epsilon_t\widehat{d}_{j,t} = O_P\left(N^{1/2}T^{-1/2}\right) + O_P\left(NT^{-1}\right). \quad \text{(D.39)}$$

Putting (D.25), (D.27) and (D.39) together, it follows that, in (D.24)

$$N^{1/2}\left(\widehat{\gamma}_{j,0} - \gamma_{j,0}\right) = \left(\frac{i_N'\mathbf{M}_{B,j}^0 i_N}{N}\right)^{-1}\left(\frac{i_N'\mathbf{M}_{B,j}^0\alpha_j}{N^{1/2}}\right) + O_P\left(N^{1/2}T^{-1}\right) + o_P\left(1\right). \quad \text{(D.40)}$$

Recalling that, by Lemma C.8, $\widehat{\pi}_j - \pi_j = O_P\left(T^{-1}\right)$ with $\eta > \frac{1}{2}$, the desired result follows from Assumption 7*(iii)* and the Cramer-Wold device. □

Proof of Theorem A.1. The proof is essentially the same as the proof of Theorems 3 and 4 in Horváth and Trapani (2019) and most of it is therefore omitted. In order to understand the role of (A.3), note that

$$\begin{aligned}
&M_\theta^{-1/2}\sum_{m=1}^{M_\theta}\left(\zeta_m^\theta\left(s\right) - \frac{1}{2}\right) = M_\theta^{-1/2}\sum_{m=1}^{M_\theta}\left(\zeta_m^\theta\left(0\right) - \frac{1}{2}\right) \\
&+ M_\theta^{-1/2}\sum_{m=1}^{M_\theta}\left(\zeta_m^\theta\left(s\right) - \zeta_m^\theta\left(0\right) - \int_0^{s/f_{N,T}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)dx\right) \\
&+ M_\theta^{-1/2}\sum_{m=1}^{M_\theta}\int_0^{s/f_{N,T}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)dx.
\end{aligned}$$

By the same passages as in the proof of Theorem 3 in Horváth and Trapani (2019), it is easy to see that

$$M_\theta^{-1/2}\sum_{m=1}^{M_\theta}\left(\zeta_m^\theta\left(0\right) - \frac{1}{2}\right) = O_{P^*}\left(1\right);$$

also

$$E^*\int_{-\infty}^{\infty}\left|M_\theta^{-1/2}\sum_{m=1}^{M_\theta}\left(\zeta_m^\theta\left(s\right) - \zeta_m^\theta\left(0\right) - \int_0^{s/f_{N,T}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)dx\right)\right|^2 ds = o\left(1\right),$$

and

$$\int_{-\infty}^{\infty}\left|M_{\theta}^{-1/2}\sum_{m=1}^{M_{\theta}}\int_{0}^{s/f_{N,T}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^{2}\right)dx\right|^{2}ds=c_{0}\frac{M_{\theta}}{f_{N,T}^{1/2}}=o\left(1\right),$$

after (A.3). The rest of the proof follows immediately. □

E Bootstrap algorithms

We develop three bootstrap algorithms for the null hypothesis that the average risk premium is equal to zero against the two-sided alternative that the average risk premium is different from zero: Algorithm 1 in Section E.1 is for the unconditional model over the full sample; Algorithm 2 in Section E.2 is for the unconditional model within regimes $\mathcal{D}$ and $\mathcal{S}$; Algorithm 3 in Section E.3 is for the conditional model within regimes $\mathcal{D}$ and $\mathcal{S}$.[v]

E.1 Algorithm 1: unconditional model, full sample

Step 1. For $i=1,\ldots,N$ and $t=1,\ldots,T$, generate a boostrap sample of $R_{i,t}^{*}$ by resampling weighted residuals as

$$R_{i,t}^{*}=\widehat{\beta}_{i}^{\prime}\widehat{\gamma}_{1}+\widehat{\beta}_{i}^{\prime}\widehat{\mathbf{u}}_{t}+\hat{\epsilon}_{i,t}^{*},\quad \hat{\epsilon}_{i,t}^{*}=\hat{\epsilon}_{i,t}^{*}w_{it},$$

where $\widehat{\mathbf{u}}_{t}$ is the principal components estimator for $\mathbf{u}_{t}$ in (1), and $\{w_{it}\}$ is a sequence of *i.i.d.* random variables such that $E\left(w_{it}\right)=0$ and $Var\left(w_{it}\right)=1$.

Step 2. For $t=1,\ldots,T$, given $\mathbf{R}^{*}=(\mathbf{R}_{1}^{*},\ldots,\mathbf{R}_{T}^{*})$, where $\mathbf{R}_{t}^{*}=\left(R_{1,t}^{*},\ldots,R_{N,t}^{*}\right)^{\prime}$, obtain

$$\widehat{\mathbf{B}}^{*}=\left(\mathbf{R}^{*}\mathbf{M}_{T}\widehat{\mathbf{U}}^{\prime}\right)\left(\widehat{\mathbf{U}}\mathbf{M}_{T}\widehat{\mathbf{U}}^{\prime}\right)^{-1},$$

where $\widehat{\mathbf{U}}=(\widehat{\mathbf{u}}_{1},\ldots,\widehat{\mathbf{u}}_{T})$, $\mathbf{M}_{T}=I_{T}-T^{-1}\iota_{T}\iota_{T}^{\prime}$ and $\widehat{\mathbf{B}}^{*}=\left(\widehat{\beta}_{1}^{*},\ldots,\widehat{\beta}_{N}^{*}\right)^{\prime}$.

[v] Consistently with the discussion in Footnote 2, we develop the algorithm under the maintained assumption that the zero-beta rate is equal to zero both in the unconditional and in the conditional model

Step 3. Given $\overline{\mathbf{R}}^* = \left(\overline{R}_1^*, \ldots, \overline{R}_N^*\right)' = T^{-1} \sum_{t=1}^{T} \mathbf{R}_t^*$, estimate the bootstrap risk premia as

$$\widehat{\gamma}_1^* = \left(\widehat{\mathbf{B}}^{*\prime} \widehat{\mathbf{B}}^*\right)^{-1} \left(\widehat{\mathbf{B}}^{*\prime} \overline{\mathbf{R}}^*\right).$$

Step 4. For $i = 1, \ldots, N$, estimate the average pricing error as

$$\overline{\widehat{\alpha}}^* = N^{-1} \sum_{i=1}^{N} \widehat{\alpha}_i^*,$$

where the estimated bootstrap pricing errors are

$$\widehat{\alpha}_i^* = \overline{R}_i^* - \widehat{\beta}_i^{*\prime} \widehat{\gamma}_1^*.$$

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\overline{\widehat{\alpha}}^{*,b}$, for $b = 1, \ldots, B$. Compute the bootstrap *p-value* as

$$\widehat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\left|\overline{\widehat{\alpha}}^{*,b}\right| > \left|\overline{\widehat{\alpha}}\right|\right),$$

where $\overline{\widehat{\alpha}}$ is the average value of the vector $\widehat{\alpha} = \overline{\mathbf{R}} - \widehat{\mathbf{B}}\widehat{\mathbf{\Gamma}}$.

E.2 Algorithm 2: unconditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1 - Step 4. Step 1 through Step 4 are the same as in Algorithm 1.

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\overline{\widehat{\alpha}}^{*,b}$, for $b = 1, \ldots, B$. For $j = \mathcal{D}, \mathcal{S}$, compute the bootstrap *p-value* as

$$\widehat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\left(\left|\overline{\widehat{\alpha}}^{*,b}\right| > \left|\overline{\widehat{\alpha}}_{j\mathcal{U}}\right|\right),$$

where $\overline{\widehat{\alpha}}_{j\mathcal{U}}$ is the average value of the vector $\widehat{\alpha}_{j\mathcal{U}} = \overline{\mathbf{R}}_j - \widehat{\mathbf{B}}\widehat{\mathbf{\Gamma}}$.

E.3 Algorithm 3: conditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, generate a boostrap sample of $R^*_{i,t}$ by resampling weighted residuals as

$$\begin{aligned} R^*_{i,t} &= \hat{d}_{\mathcal{D},t}\left(\widehat{\beta}'_{\mathcal{D},i}\widehat{\gamma}_{\mathcal{D},1} + \widehat{\beta}'_{\mathcal{D},i}\widehat{\mathbf{u}}_{\mathcal{D},t}\right) \\ &+ \hat{d}_{\mathcal{S},t}\left(\widehat{\beta}'_{\mathcal{S},i}\widehat{\gamma}_{\mathcal{S},1} + \widehat{\beta}'_{\mathcal{S},i}\widehat{\mathbf{u}}_{\mathcal{S},t}\right) + \hat{\epsilon}^*_{i,t}, \end{aligned} \qquad \hat{\epsilon}^*_{i,t} = \hat{\epsilon}^*_{i,t} w_{it},$$

with $\hat{\epsilon}^*_{i,t} = \hat{\epsilon}^*_{i,t} w_{it}$ where $\{w_{it}\}$ is a sequence of $i.i.d.$ random variables such that $E(w_{it}) = 0$ and $Var(w_{it}) = 1$.

Step 2. For $j = 1, 2$ and $t = 1, \ldots, T$, obtain

$$\widehat{\mathbf{B}}^*_j = \left(\mathbf{R}^*\mathbf{M}_{T_j}\widehat{\mathbf{U}}'_j\right)\left(\widehat{\mathbf{U}}_j\mathbf{M}_{T_j}\widehat{\mathbf{U}}'_j\right)^{-1},$$

where $\widehat{\mathbf{B}}^*_j = \left(\widehat{\beta}^*_{j,1}, \ldots, \widehat{\beta}^*_{j,N}\right)'$ and

$$\mathbf{M}_{T_j} = \begin{pmatrix} \hat{d}_{j,1} & 0 & \ldots & 0 \\ 0 & \hat{d}_{j,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{d}_{j,T} \end{pmatrix} - \frac{1}{T_j}\begin{pmatrix} \hat{d}_{j,1} \\ \hat{d}_{j,2} \\ \vdots \\ \hat{d}_{j,T} \end{pmatrix}\begin{pmatrix} \hat{d}_{j,1} \\ \hat{d}_{j,2} \\ \vdots \\ \hat{d}_{j,T} \end{pmatrix}'.$$

Step 3. Given $\overline{\mathbf{R}}^*_j = \left(\overline{R}^*_{j,1}, \ldots, \overline{R}^*_{j,N}\right)' = T_j^{-1}\sum_{t=1}^{T}\hat{d}_{j,t}\mathbf{R}^*_{j,t}$, estimate the bootstrap risk premia as

$$\widehat{\gamma}^*_{j,1} = \left(\widehat{\mathbf{B}}^{*\prime}_j\widehat{\mathbf{B}}^*_j\right)^{-1}\left(\widehat{\mathbf{B}}^{*\prime}_j\overline{\mathbf{R}}^*_j\right).$$

Step 4. For $j = \mathcal{D}, \mathcal{S}$ and $i = 1, \ldots, N$, estimate the average pricing error as

$$\overline{\widehat{\alpha}}^*_j = N^{-1}\sum_{i=1}^{N}\widehat{\alpha}^*_{j,i},$$

where the estimated bootstrap pricing errors are

$$\widehat{\alpha}^*_{j,i} = \overline{R}^*_{j,i} - \widehat{\beta}^{*\prime}_{j,i}\widehat{\gamma}^*_{j,1}.$$

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\overline{\widehat{\alpha}}^{*,b}_j$, for $b = 1, \ldots, B$. Compute the bootstrap *p-value* as

$$\widehat{p}^{*,B} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(\left|\overline{\widehat{\alpha}}^{*,b}_j\right| > \left|\overline{\widehat{\alpha}}_j\right|\right),$$

where $\overline{\widehat{\alpha}}_j$ is the average value of the vector $\widehat{\alpha}_j = \overline{\mathbf{R}}_j - \widehat{\mathbf{B}}_j\widehat{\boldsymbol{\Gamma}}_j$.