



Emerging Markets Group

EMG Working Paper Series

WP-EMG-01-2024

‘Factor Models with Downside Risk’

Daniele Massacci Lucio Sarno Lorenzo Trapani

July 2024

Emerging Markets Group Bayes
Business School City University
106 Bunhill Row London EC1Y 8TZ UK
www.bayes.city.ac.uk/emg/

Factor Models with Downside Risk^{*}

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

This Version: July 2024

^{*}We are grateful for comments and suggestions to Kay Giesecke (Editor), an anonymous Associate Editor, two anonymous Referees, 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. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

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

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

[§]Lorenzo Trapani is with the University of Leicester and the University of Pavia. Email: lt285@leicester.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 with an empirical application to a broad cross-section of stock portfolio returns.

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) find a downside 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 the presence of a downside risk premium in asset returns, the *unconditional*, linear asset pricing model, which constitutes the workhorse model of empirical asset pricing, is misspecified. A *conditional* model is required to characterize the different dynamics of asset returns and properties of risk premia in good and bad states of the world.¹ However, estimation and inference for asset pricing models with downside risk is a challenging task. This is because allowing for different regimes means that: the number of risk factors in the model 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 conditional model of asset returns in the presence of common factors and downside

¹To be clear, the existence of downside risk premia does not inherently render any linear asset pricing model misspecified. However, the workhorse model in asset pricing is based on a linear specification that imposes symmetry in the exposure to risk and is, therefore, unable to capture the downside risk premium. A conditional linear model that allows for asymmetric exposures to risk in the pricing equation is required, while the standard linear model that assumes symmetric exposures displays nonzero pricing errors when the true data generating process is a conditional linear model. Moreover, in the presence of a downside risk premium, the unconditional linear model also displays higher mean squared prediction errors. See Section A.1 in the Internet Appendix for a formalization of the notion of conditional versus unconditional specification, and the nature of misspecification in the unconditional linear model in the presence of a downside risk premium.

risk; further, we consider a *latent* factor model approach, which allows to span the entire factor space as advocated by Giglio and Xiu (2021).² This setup is general enough to also encompass previous models with downside risk, while allowing for different common factors across states and estimation of the threshold determining the downside state.

We make at least three contributions to the literature on factor models of asset returns. First, we develop a model where the state of the world is conditional within the model (i.e. the state at time $t + 1$ is predicted on the basis of information at time t), and estimate the value of the threshold that triggers the switching from one state of the world to the other. This is different from the existing literature cited above, which determines the disappointment event in terms of the relative position of the *contemporaneous* market return with respect to a threshold value that is set *a priori* equal to a pre-specified value. In essence, existing models of downside risk assume that the state of the world is known within the model, whereas in our model the switching between states is driven by a conditional expectation of a variable (or set of variables) crossing an unknown threshold 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.³ 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, 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.⁴ Overall, the distilled

²Since it is not 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 by Giglio and Xiu (2021), who propose a three-pass procedure to tackle the omitted variable bias problem by considering a latent factor model which allows to span the underlying true factor space.

³We 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.

⁴This step follows from extending the results of Giglio and Xiu (2021) to a conditional setting, delivering estimates of risk premia for any candidate risk factor that are robust to omitted variables and measurement error biases.

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 illustrate the usefulness of this methodology by studying empirically the factor structure of equity portfolio returns (the same cross-section used by Farago and Tédongap, 2018). The switching mechanism is driven by an indicator function that relies on a single conditioning variable, which captures the expectation of a bear market state, as embedded in the Arrow-Debreu (AD) portfolio recently proposed by Lu and Murray (2019) in a different context. Specifically, using traded option prices on S&P500 stocks, the price of the AD Bear portfolio – that pays one dollar when the market at expiration is in a bear state – is a forward-looking measure of the (risk-neutral) probability of future bear market states. The AD Bear portfolio return reflects the change in this probability, i.e., bear market risk. In short, we use the AD Bear return as the conditioning variable in the switching mechanism to predict a downside (upside) state. This setup makes the model conditional, in the sense that the state of the world is no longer assumed to be known when fitting a model for returns at time $t + 1$, which is the simplifying assumption in the literature on downside risk (Ang et al., 2006; Lettau et al., 2014; Farago and Tédongap, 2018). Rather, the state of the world at time $t + 1$ is now an *expected* state, with the expectation conditional on information available up to time t . We estimate the threshold of the AD portfolio return that drives the two states of the world using our procedure, and then estimate the full conditional model, producing results both in- and out-of-sample.

Starting from the in-sample results, 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 - 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, this 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

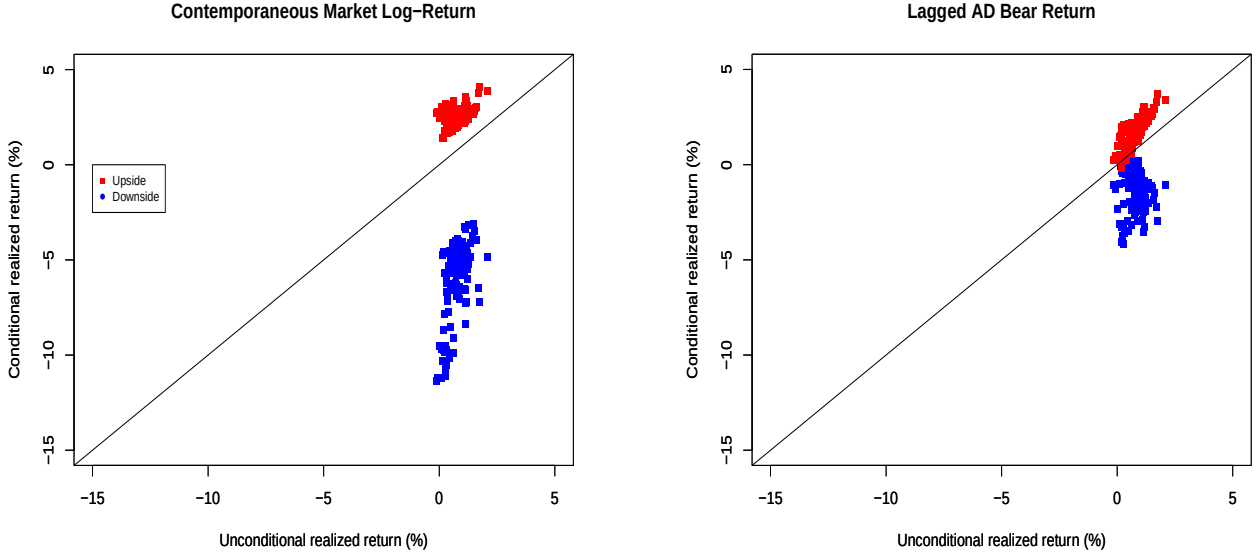
excess equity returns accurately. In other words, assessing the performance of the unconditional factor model over the full sample can give the illusion that it performs well when in fact it performs poorly in both regimes. In contrast, the latent factor model with downside risk appears to have no systematic tendency to generate positive or negative pricing errors in either regime, and therefore also unconditionally, fitting the cross-section of equity returns used here very well. The conditional model performs significantly better than the unconditional benchmark in terms of the goodness-of-fit criteria we consider, namely the R^2 , the root mean square error (RMSE) and the average pricing error. Moreover, we also find that the conditional model can price the test asset portfolio returns fully in-sample, with the average pricing error (alpha) being statistically insignificantly different from zero.⁵

The inadequacy of the unconditional model is well illustrated by Figure 1 below, which plots unconditional portfolio equity returns against average portfolio equity returns conditional on either good or bad states. Good and bad states are defined in two different ways: *(i)* depending on whether the contemporaneous aggregate stock market return is below or above -3% (a common definition of the state of the world in the literature) in a given month (left plot); and *(ii)* depending on whether the conditional expectation based on the AD Bear portfolio return predicts a bear market or not next month (right plot). Both 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 conditional model that changes across states.

Turning to the out-of-sample analysis, we produce results for the last ten years of the sample keeping constant the estimate of the threshold that drives the switching mechanism (which is set at the value estimated in the preceding sample), and forecast portfolio returns at time $t + 1$ conditional only on information available at time t , recursively. The outcome of the empirical analysis is qualitatively the same as for the in-sample analysis, in that the unconditional model has many shortcomings and statistically significant average pricing errors, while the conditional model fits the data much better. The conditional model outperforms the unconditional asset

⁵The results are qualitatively identical when estimating the model for the shorter sample period that reserves the last ten years for out-of-sample analysis as well as when using the full sample of data at our disposal.

Figure 1: Conditional and unconditional realized returns for equity portfolios



Notes. Over the period February 1997 - August 2005, the scatter plot shows realized average returns in upside (red dots) and downside (blue dots) states against unconditional average returns for the 130 equity portfolios described in Section 6.1 in the following two scenarios: (i) the state variable is the contemporaneous market log-return and the threshold value is fixed at -3% (left panel); (ii) the state variable is the one-month lagged value of the average AD Bear portfolio return and the threshold value is estimated using the procedure described in Section 4.1, and as discussed in Section 6.2.1 (right panel).

pricing model in terms of every measure of goodness-of-fit examined, displaying higher out-of-sample R^2 , lower RMSEs, and lower average pricing errors across each regime. However, we also find that the average pricing errors are not statistically equal to zero even for the conditional model, although they are one order of magnitude smaller than the average pricing errors from the unconditional model. This suggests that, as one might expect, there is some deterioration in performance when taking the conditional model out of sample, notwithstanding its superiority relative to the unconditional benchmark. We would like to point out that, both in- and out-of-sample, we test formally the hypothesis that the average pricing errors are the same for the conditional and unconditional models against the alternative that the average pricing errors from the conditional model are lower. In each case (i.e. both in-sample and out-of-sample) we strongly reject the null hypothesis, implying superior pricing performance of the conditional model.

These empirical results also highlight the importance of allowing for different common factors across states. We find that more common factors are needed in the satisfactory state of the world

than in the downside state to capture the factor structure of returns. 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 within the model, estimated using our procedure that explicitly allows the factors to change across states. 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 Quaadvlieg (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 (see e.g. Gagliardini et al., 2016, Gagliardini et al., 2019, and Zaffaroni, 2019, *inter alia*), 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 “ $\rightarrow$ ”; $\mathbb{I}(\cdot)$ 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 $C_{N,T} = \min \{N^{1/2}, T^{1/2}\}$.

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 unconditional asset pricing model can be written as

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

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

$$\mathbf{u}_t = \mathbf{f}_t - \boldsymbol{\mu}_f. \quad (2)$$

In Eq. (2), $\mathbf{f}_t = (f_{1,t}, \dots, f_{P,t})'$ 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; $\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \dots, \gamma_{1,P})'$ is the $P \times 1$ vector of risk premia; $\boldsymbol{\beta}_i = (\beta_{i,1}, \dots, \beta_{i,P})'$ 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

$$\mathbf{R}_t = \gamma_0 \boldsymbol{\iota}_N + \boldsymbol{\alpha} + \mathbf{B} \boldsymbol{\gamma}_1 + \mathbf{B} \mathbf{u}_t + \boldsymbol{\epsilon}_t, \quad (3)$$

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

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

$$\mathbf{g}_t = \mathbf{a} + \mathbf{\Lambda} \mathbf{u}_t + \mathbf{e}_t. \quad (4)$$

Given γ_1 , the $K \times 1$ vector of risk premia of $\mathbf{g}_t$ is $\gamma_{\mathbf{g}} = \mathbf{\Lambda} \gamma_1$.

2.2 Conditional asset pricing model with downside risk

The conditional factor model of interest is:

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

where

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

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 $\mathbb{I}(\mathcal{D}_t) \mathbb{I}(\mathcal{S}_t) = 0$. We define the disappointment event as $\mathcal{D}_t = \{E_{t-1}(Z_t) \leq \theta\}$, where Z_t is a state variable and θ is a threshold.⁶ Under perfect foresight, $E_{t-1}(Z_t) = Z_t$, which is the case typically studied in the literature on downside risk at least since Ang, Chen, and Xing (2006). The theoretical results derived below hold regardless of whether one works under perfect foresight (i.e., using contemporaneous information about the state) or with the conditional expectation (i.e., using lagged information to predict the state); see also the comments after Assumption 5 in the Internet Appendix. However, our focus is primarily on the

⁶Alternatively, Z_t could also be a function (e.g., a linear combination) of several variables. Furthermore, we would like to point out that the direction of the inequality in the definition of the state depends upon the specific variable Z_t chosen. For example, the definition $\mathcal{D}_t = \{E_{t-1}(Z_t) \leq \theta\}$ is a natural one when Z_t is a variable like the market excess return, which is associated with a bad state of the world when it takes abnormally low values. Conversely, in the case of a variable that is associated with a bad state of the world when it takes abnormally high values, the downside state would be defined as $\mathcal{D}_t = \{E_{t-1}(Z_t) \geq \theta\}$. This is the case, for example, if Z_t is a volatility or an illiquidity indicator, and is also the case for the AD Bear portfolio return we use in our later empirical work.

model that uses the conditional expectation, and in the empirical work we focus on this more realistic case, where the state of the world at time t is unknown and needs to be predicted using only information available up to time $t - 1$. Section 2.3 below discusses $\mathcal{D}_t$ and $\mathcal{S}_t$ in detail. In Eq. (6), $\mathbf{f}_{j,t} = (f_{j,1,t}, \dots, f_{j,P_j,t})'$ is the $P_j \times 1$ vector of latent factors satisfying $E[\mathbf{f}_{j,t} | \mathbb{I}(j_t) = 1] = \mu_{j,f}$; $\mathbf{u}_{j,t}$ is the $P_j \times 1$ vector of factor innovations such that $E[\mathbf{u}_{j,t} | \mathbb{I}(j_t) = 1] = 0$. As for (5): $\gamma_{j,0}$ is the zero-beta rate and $\boldsymbol{\gamma}_{j,1} = (\gamma_{j,1,1}, \dots, \gamma_{j,1,P_j})'$ is the $P_j \times 1$ vector of risk premia; $\boldsymbol{\beta}_{j,i} = (\beta_{j,i,1}, \dots, \beta_{j,i,P_j})'$ is the $P_j \times 1$ vector of risk exposures; $\alpha_{j,i}$ is the pricing error; the idiosyncratic components $\epsilon_{i,t}$ satisfy $E[\mathbf{u}_{j,t} \epsilon_{i,t} | \mathbb{I}(j_t) = 1] = 0$. The model in matrix form is

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

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

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

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

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

When the disappointment event occurs, we thus have $\boldsymbol{\gamma}_{\mathbf{g},t} = \boldsymbol{\gamma}_{\mathcal{D},\mathbf{g}} = \boldsymbol{\Lambda}_{\mathcal{D}} \boldsymbol{\gamma}_{\mathcal{D},1}$.⁷

2.3 The disappointment event

The motivation for a downside risk premium arises from various theoretical perspectives. In some setups, the rationale is behavioral and linked to theories of loss aversion, as in the papers by Ang, Chen, and Xing (2006) and Farago and Tédongap (2018); in other setups, it is based

⁷The 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.

on rational asset pricing models, as in Lettau, Maggiori, and Weber (2014). However, all these different frameworks are based, qualitatively, on the same definition of disappointment event. As a prime example, we focus on Farago and Tédongap (2018). The theory therein considers an economy where the representative investor has Epstein-Zin preferences over consumption, and embeds generalized disappointment aversion into this framework as in Routledge and Zin (2010). Maximization of the investor's lifetime utility results in a Stochastic Discount Factor (SDF) that takes the form

$$M_{t-1,t}^{GDA} = M_{t-1,t} \times f(E_{t-1}(Z_t)),$$

where $M_{t-1,t}$ is the SDF without disappointment aversion, and $f(E_{t-1}(Z_t))$ is a term that depends, *inter alia*, on whether the disappointment event has occurred or not; note that such a disappointment event $\mathcal{D}_t$ is defined in terms of the conditional expectation of a state variable Z_t , and on whether such expectation exceeds (or falls below) a threshold. Indeed, under further assumptions and algebraic manipulations, Farago and Tédongap (2018) show that the disappointment event $\mathcal{D}_t$ takes the following form

$$\mathcal{D}_t = \{r_{W,t} - a\Delta\sigma_{W,t}^2 \leq \theta\},$$

where: $r_{W,t}$ is the log-return on the market; $\Delta\sigma_{W,t}^2$ is a measure of the return on the volatility of the market (standardized to have the same standard deviation as $r_{W,t}$); a is a non-negative parameter which captures the relative weight of $r_{W,t}$ and $\Delta\sigma_{W,t}^2$ in determining the disappointment event; and, finally, θ is the threshold value that determines the occurrence of $\mathcal{D}_t$. Thus, the disappointment event refers to times when the return on the market is particularly low and/or changes in volatility are particularly large relative to the threshold θ . In their analysis, Farago and Tédongap (2018) experiment with various positive values of a , and find that there is no qualitative difference in the determination of the state of the world as a varies; hence, setting $a = 0$, they define the disappointment event as

$$\mathcal{D}_t = \{r_{W,t} \leq \theta\}.$$

Ang, Chen, and Xing (2006) and Lettau, Maggiori, and Weber (2014) also employ the same definition for $\mathcal{D}_t$.⁸ The satisfactory event is then defined as

$$\mathcal{S}_t = \{r_{W,t} > \theta\}.$$

Therefore, Farago and Tédongap (2018), and all the papers cited above, use *contemporaneous* information on the realized log-excess return in the aggregate stock market in order to empirically determine the state – which means the state of the world is determined assuming $E_{t-1}(Z_t) = Z_t = r_{W,t}$. Naturally, this setup prevents the model from being *fully* conditional.

We depart from this approach along two dimensions. First, the switching mechanism that determines the state of the world is conditional only on lagged information, which makes the model genuinely conditional: thus, the state is predicted within the model rather than assumed known. Second, such conditioning information does not include the realized stock market excess return. Defining the state on account of the conditional expectation $E_{t-1}(Z_t)$ entails that no contemporaneous information is used. The conditioning information set for the switching mechanism does not identify the disappointment state directly as a large negative return on the stock market. Rather, it is related to the market-based prediction of a bear market as priced in the stock options market: the AD Bear portfolio provides by construction a forward-looking measure of the (risk-neutral) probability of future bear market states. Specifically we define the disappointment state as $\mathcal{D}_t = \{E_{t-1}(Z_t) \geq \theta\}$ where $E_{t-1}(Z_t)$ is the AD Bear portfolio return at time $t - 1$.

Summing up, we build on the existing literature on downside risk by generalizing the standard setup to make the model conditional in the sense that the state of the world is no longer assumed to be known when fitting a model for returns. Rather, the state of the world is now an expected state, with the expectation conditional only on available information.

As a final remark on notation, both $\mathcal{D}_t$ and $\mathcal{S}_t$ depend on θ : to stress this dependence, we write $d_{j,t}(\theta) \equiv \mathbb{I}(j_t)$, for $j = \mathcal{D}, \mathcal{S}$.

⁸In Ang, Chen, and Xing (2006), θ is set equal to the sample mean return, whereas in Farago and Tédongap (2018) it is set to -3% in the core results, with some robustness checks for alternative values. A similar approach is followed also by Lettau, Maggiori, and Weber (2014).

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 only 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 $\boldsymbol{\gamma}_{\mathbf{g}} = \boldsymbol{\Lambda}\boldsymbol{\gamma}_1 = \boldsymbol{\Lambda}\mathbf{L}\mathbf{L}^{-1}\boldsymbol{\gamma}_1$.

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

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

As far as the second identification issue is concerned, in each regime returns are characterized by a factor model which allows the error terms $\epsilon_{i,t}$ to be correlated across time and across units (thus being an “approximate factor model,” using the terminology of Chamberlain and Rothschild,

1983). This setup requires such correlations to be weak, e.g. the serial correlation between $\epsilon_{i,t}$ and $\epsilon_{i,t-k}$ must vanish as the lag k increases; similarly, cross-correlation between $\epsilon_{i,t}$ and $\epsilon_{j,t-k}$ must be “small” for $i \neq j$. Under this assumption, one can estimate factors and loadings, e.g. using a “classical” estimator such as Principal Components, and ensure that loadings and the common factors are estimated consistently (up to a linear transformation). The presence of the pricing errors α_i , in each unit, introduces what we term “unobserved heterogeneity,” which adds onto the error term. The new error $\alpha_i + \epsilon_{i,t}$ is strongly correlated across t by construction (due to the presence of α_i) and therefore consistent estimation of factors and loadings would no longer be possible - in essence, due to a failure of the “approximate factor structure” mentioned above. Hence we pre-process the data by demeaning across time, which essentially smooths away α_i from each unit, ensuring that factors and loadings can be estimated consistently. This means that we express 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 $\bar{\mathbf{R}}_j(\theta) = T_j(\theta)^{-1} \sum_{t=1}^T d_{j,t}(\theta) \mathbf{R}_t$: $\bar{\mathbf{R}}_j(\theta)$ is the mean of $\mathbf{R}_t$ when j occurs. In order to estimate pricing factors and risk exposures, we consider the model

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

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

3.2 The disappointment threshold θ

As 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. In this paper, we provide an estimator of θ and derive its limiting distribution (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 $\beta_{j,i}$ and $\gamma_{j,1}$ (Section 4.3). The assumptions we make in the subsequent estimation procedure are listed and discussed in Internet Appendix A.2. However, we note here that the assumptions in each regime are essentially the same as the assumptions in Giglio and Xiu (2021).

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 shorthand 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, $\mathbb{P}^*$ denotes the probability conditional on $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$; we use “ $\xrightarrow{D^*}$ ” and “ $\xrightarrow{\mathbb{P}^*}$ ” to indicate convergence in distribution and in probability according to $\mathbb{P}^*$, respectively.

4.1 Estimating θ^0

We begin by studying the estimation of θ^0 . Recall the $N \times 1$ vector of demeaned returns $\widetilde{\mathbf{R}}_t(\theta) = (\widetilde{R}_{1,t}, \dots, \widetilde{R}_{N,t})'$, 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 $\mathbf{B}^P = (\mathbf{B}_{\mathcal{D}}^{P_{\mathcal{D}}}, \mathbf{B}_{\mathcal{S}}^{P_{\mathcal{S}}})$, and the associated $P_{\mathcal{D}} \times T$ and $P_{\mathcal{S}} \times T$ matrices of demeaned factor innovations $\mathbf{U}_{\mathcal{D}}^{P_{\mathcal{D}}} = (\mathbf{u}_{\mathcal{D},1}^{P_{\mathcal{D}}}, \dots, \mathbf{u}_{\mathcal{D},T}^{P_{\mathcal{D}}})$ and $\mathbf{U}_{\mathcal{S}}^{P_{\mathcal{S}}} = (\mathbf{u}_{\mathcal{S},1}^{P_{\mathcal{S}}}, \dots, \mathbf{u}_{\mathcal{S},T}^{P_{\mathcal{S}}})$, with $\mathbf{U}^P = (\mathbf{U}_{\mathcal{D}}^{P_{\mathcal{D}}}, \mathbf{U}_{\mathcal{S}}^{P_{\mathcal{S}}})'$.

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):

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

For given P , the estimators of $\widehat{\mathbf{B}}^P = (\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}, \widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}})$ – with $\widehat{\mathbf{B}}_j^{P_j} = (\widehat{\beta}_{j,1}, \dots, \widehat{\beta}_{j,N})'$ for $j = \mathcal{D}, \mathcal{S}$ – and

$\widehat{\mathbf{U}}^P$ and $\widehat{\theta}^P$ are defined as

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

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

It follows that the estimator of $\mathbf{U}^P$ is

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

Hence, we can define

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

Letting

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

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

We now define

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

Let $\widehat{\mathbf{B}}^P(\theta) = [\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}(\theta), \widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}(\theta)]$ and $\widehat{\mathbf{U}}^P(\theta) = [\widehat{\mathbf{U}}_{\mathcal{D}}^{P_{\mathcal{D}}'}(\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}(\theta); \theta), \widehat{\mathbf{U}}_{\mathcal{S}}^{P_{\mathcal{S}}'}(\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}(\theta); \theta)]'$. 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

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

$\hat{\theta}$ 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 $\hat{\theta}$ is a strongly consistent estimator of θ^0 , deriving its rate (see Lemma C.7). In principle, the limiting distribution of $\hat{\theta}$ 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 $\hat{\theta}$ 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 $\hat{\theta}$, 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.3 of the Internet Appendix, to which we refer the interested reader.

4.2 Determining the number of common factors

After estimating θ^0 , it is possible to proceed to the estimation of the factor structures in each regime. We would like to emphasize that the estimation error $\hat{\theta} - \theta^0$ does not have an impact on estimation and inference in the subsequent steps of our procedure – intuitively, this is due to the fast rate of convergence of $\hat{\theta} - \theta^0$ (see Lemma C.7). The next step of the analysis is the estimation of the dimensions of the factor spaces P_D^0 and P_S^0 . In principle, after splitting the sample around the estimated threshold $\hat{\theta}$, any of the techniques currently available for linear models can be employed – indeed, it can be shown 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, which is the building block for all existing methodologies.⁹ In this paper, we extend the procedure proposed by Trapani (2018) to estimate the number of factors in each regime, $P_{\mathcal{D}}^0$ and $P_{\mathcal{S}}^0$. Such an extension is not trivial, since it requires controlling the regime-splitting errors coming from using an estimated value of θ^0 .¹⁰

As in the previous section, here we only report the relevant algorithm; details and the theory are in Section A.4 of the Internet Appendix.

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

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

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

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

Using Lemma A.1 and Theorem A.3 of Section A.4 in the Internet Appendix, we can determine the number of common factors in each regime using the following method, based on two separate steps.

⁹For example, a possible approach (see Bai and Ng, 2002) would be to use a threshold, for the eigenvalues of the sample covariance matrix of the data, on the penalty function of some information criterion computed in both regimes; 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. Alternatively, a different approach based on using the ratio of adjacent eigenvalues (see, *inter alia*, Ahn and Horenstein, 2013, and Lam and Yao, 2012) can be used. 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. The procedure proposed herein, which extends the work by Trapani (2018), overcomes these issues: 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 .

¹⁰We 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.

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

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

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 $P_{\mathcal{D}}^0$ and $P_{\mathcal{S}}^0$ by carrying out a sequence of tests for $i = 1, \dots, P_{\max}$, with $P_{\max}$ a user-defined upper bound.

We begin with (17). Define

$$\psi(\hat{g}_j^{(i)}) = \exp\left(\frac{N^{-\varrho_j} \hat{g}_j^{(i)}}{\bar{g}_j(p)}\right), \quad (18)$$

where¹¹

$$\bar{g}_j(p) = \frac{1}{N-p} \sum_{h=p+1}^N \hat{g}_j^{(h)}.$$

In Eq. (18), we have defined

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

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.3, the *sample* eigenvalues have an estimation error which may

¹¹The purpose of $\bar{g}_j(p)$ is to rescale $\hat{g}_j^{(i)}$; in Trapani (2018), it is suggested to set $p = 0$ (that is, to use the trace to rescale $\hat{g}_j^{(i)}$); other choices are also possible, such as $p = i$, which is recommended in Barigozzi and Trapani (2022).

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 $N^{-\varrho_j} \hat{g}_j^{(i)}$ 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 $\hat{g}_j^{(i)}$ by $N^{-\varrho_j}$ cancels the estimation error when $g_j^{(i)}$ is bounded, whilst not cancelling the signal in $g_j^{(i)}$ when this is not bounded. Naturally, if ϱ_j is large (i.e., if ε in Eq. (19) is chosen as a “large” number), this will attenuate the estimation error in $\hat{g}_j^{(i)}$, but it will also attenuate the signal in $\hat{g}_j^{(i)}$ 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.¹²

We now describe the test for (17).

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

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

Step 3 Compute

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

where

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

We show that, under the null, $\Upsilon_j^{(i)} \xrightarrow{D^*} \chi_1^2$, 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 $\hat{P}_j = 0$ and stop, otherwise go to the next step.

¹²In our empirical exercises in Section 6, we explore the 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.

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 $\hat{P}_j = i$ 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, $\hat{P}_j$ is a consistent estimator.

Theorem 1. *We assume that Assumptions 1-6 in Section A.2 of the Internet Appendix are satisfied. As $\min(N, T_j) \rightarrow \infty$ under (A.22), if $c_\alpha(N, T) \rightarrow \infty$ with $c_\alpha(N, T) = o(M)$, then it holds that*

$$\hat{P}_j = P_j^0 + o_{\mathbb{P}^*}(1),$$

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 on p. 203 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 takeaway 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 are similar to the samples used in the empirical analysis for N and T , when T is sufficiently large, whereas the estimated number of common factors tends to overstate the true number, especially in the downside regime, when T is small. This can be viewed as conceptually similar to having a set of irrelevant variables in a regression model, and may be considered preferable to understating the number of common factors, which would correspond to an omitted variable scenario.

4.3 Estimating common factors and loadings

Once $\hat{\theta}$ and $\hat{P}_j$ (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

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

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

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

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 $\hat{\theta}$, $\hat{P}_j$, $\{\hat{\mathbf{u}}_{j,t}^{\hat{P}_j}\}_{t=1}^T$, and $\{\hat{\beta}_{j,i}^{\hat{P}_j}\}_{i=1}^N$ 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 Λ_j within each regime by running a regression of $\mathbf{g}_t$ on $\hat{\mathbf{u}}_{j,t}^{\hat{P}_j}$ 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 $\hat{\theta}$ and $\hat{P}_j$ 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.5 of the Internet Appendix.

5.1 Regime-specific cross-sectional regression

Given the $N \times \hat{P}_j$ matrices of beta estimates $\hat{\mathbf{B}}_j^{\hat{P}_j} = \left(\hat{\beta}_{j,1}^{\hat{P}_j}, \dots, \hat{\beta}_{j,N}^{\hat{P}_j} \right)'$, define the $N \times (1 + \hat{P}_j)$ matrices $\hat{\mathbf{X}}_j = \left(\iota_N, \hat{\mathbf{B}}_j^{\hat{P}_j} \right)$, 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 $\mathbf{R}_t = (R_{1,t}, \dots, R_{N,t})'$, for $t = 1, \dots, T$.

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

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

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

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

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

5.2 Estimation of regime-specific risk premia

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

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

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

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

$$\hat{\gamma}_{j,\mathbf{g}} = \hat{\mathbf{\Lambda}}_j \hat{\gamma}_{j,1}, \quad (23)$$

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

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

We report here the asymptotics for $\hat{\gamma}_{j,\mathbf{g}}$; results for $\hat{\Gamma}_j$ and $\hat{\Lambda}_j$ are in Section A.5.3 of the Internet Appendix. We begin by showing the consistency of our estimators.

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

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

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 $\hat{\gamma}_{j,\mathbf{g}}$ 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.¹³

Theorem 3. *We assume that Assumptions 1-9 in Section A.2 are satisfied, and that $\min(N, T_j) \rightarrow \infty$, with*

$$\begin{aligned} \pi_j &\in (0, 1), \\ \frac{T^{1/2}}{N} &\rightarrow 0. \end{aligned} \tag{24}$$

$$\tag{25}$$

Then, it holds that

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

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.5 of the Internet Appendix.¹⁴

¹³Finally, we note that, when estimating Γ_j , we also obtain an estimator of the zero-beta rate $\gamma_{j,0}$ for $j = \mathcal{D}, \mathcal{S}$.

¹⁴A comment on (25) in the theorem 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”,

6 Empirical analysis

We now apply the conditional factor model and the methods described above to a cross-section of equity returns. We assess the model’s performance both in- and out-of-sample.

6.1 Data

We use equity 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 data are obtained from Kenneth French’s website.¹⁵ We consider the sample period from February 1997 to September 2015, a total of $T = 224$ monthly observations, which is dictated by the state variable in the switching mechanism, as described below. Farago and Tédongap (2018) analyze each of these sets of portfolios individually.¹⁶ 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.

The state variable used to determine the disappointment event is the AD Bear portfolio return constructed by Lu and Murray (2019). This variable measures the time variation in the ex ante probability of future bear market states (i.e., states in which the market portfolio suffers a large loss), which makes it a well-suited candidate variable for the switching mechanism in our model. The construction of this variable is based on an Arrow (1964) and Debreu (1959) portfolio (hence the expression “AD Bear” portfolio) from traded S&P500 index options. The AD Bear portfolio pays off \$1 when the market at expiration is in a bear state. Therefore, the price of the AD Bear portfolio is a forward-looking measure of the (risk-neutral) probability of future bear market states, and the short-term AD Bear return reflects the change in this probability. In practice, the

which is typically not the case in empirical work) should satisfy this requirement. We also refer to the discussion after Assumption 6 in the Internet Appendix for a taxonomy, and further comments, on the restrictions between N and T which are required for some of our results.

¹⁵See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

¹⁶See Section 3.2 and Table 1 in Farago and Tédongap (2018).

AD Bear portfolio is a bear spread position in S&P500 index put options, which is long an out-of-the-money (OTM) put and short another OTM put. Bear spreads are used as hedges against downside risk by institutional investors. The Chicago Mercantile Exchange (CME) maintains a separate order book for S&P500 bear spread trades, which account for a substantial portion of S&P500 option trading volume on the CME. This variable is available from February 1997 to September 2015, and therefore this is the sample period which we employ in the empirical work.¹⁷

6.2 In-sample analysis

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). We reserve the last ten years of the sample period (October 2005-September 2015) for out-of-sample analysis, and hence perform in-sample estimation and inference on the basis of the sample from February 1997 to September 2005, although we also validate the results over the full sample from February 1997 to September 2015 later in this section.

6.2.1 Estimation of θ

Results on estimation of θ are in Table 1. The estimated disappointment threshold $\hat{\theta}$ is equal to 0.210: this generates $\hat{T}_D = 30$ downside observations, with a relative frequency of approximately 29%. Note that the AD Bear return has a mean of -0.256 , a median of -0.446 and a standard deviation of about 1, also displaying positive skewness of about 0.7. The negative mean reflects the compensation investors pay for the insurance provided by the AD Bear portfolio against downside risk (which is the mirror image of the premium required to provide such insurance in stock option markets), while the positive skewness reflects the fact that downside events are infrequent. The

¹⁷Also, note that Lu and Murray (2019) employ the AD Bear portfolio return to construct a beta-sorted portfolio investment strategy in stocks. The excess return from this strategy is shown to be statistically significant on average, reflecting compensation for bear risk as priced in equity markets. Therefore, an alternative state variable to determine the disappointment event is the excess return from this strategy, which is available at the monthly frequency. Further details are provided in Section 4.2 in Lu and Murray (2019). We employ the AD Bear return directly in this paper since it is more directly linked to our economic motivation for the state variable that drives the disappointment state. However, we also employed the excess return from the strategy and found qualitatively identical results (not reported). As the AD Bear portfolio return is available at the daily frequency, we aggregate it to monthly frequency by taking the average of the series within each month. These data are obtained from Zhongjin Lu's website at <https://sites.google.com/view/genelu/data>.

positive estimate of $\hat{\theta}$, 0.210, being higher than the mean of the AD Bear portfolio return and much higher than the median, identifies the downside state of the world where the AD Bear portfolio yields excess returns larger than 0.210.¹⁸ It is also instructive to notice that the stock market excess return is, on average in each regime, just under 1% per month in the satisfactory state and about -0.9% per month in the downside state, meaning that the regimes identified by the AD Bear return and the estimated $\hat{\theta}$ are consistent with sizable stock market losses in downside, and healthy returns in the satisfactory regime.

To enhance our understanding of the downside regime, we evaluate the correlation between the downside indicator, as resulting from using the estimated threshold $\hat{\theta}$, and various measures of economic and financial conditions that may be related to the downside regime. We consider the following set of variables: the stock market excess return ($RmRf$); 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 ($T10Y2Y$). The correlation with the market excess return is negative (-0.171) and statistically significant at the 10% level. Broadly speaking, the correlations point towards a connection between the downside indicator and the measures of economic and financial conditions which we consider, especially those related to volatility. This finding is consistent with a vast literature in 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., 2012). Note, however, that the bilateral correlations are far from perfect and are always lower than 0.5; this indicates that the determinants of the threshold go beyond volatility and uncertainty. Yet, the downside indicator displays positive and statistically significant correlations with the NBER recession indicator, REC ; with the VIX , VXO , EPU , and EMV ; with the Policy-Related Equity Market Volatility

¹⁸Descriptive statistics for the AD Bear return are provided in Table F.1 in the Internet Appendix.

tracker, PREMV; and with the disaster probability, SPX.¹⁹

6.2.2 Estimating the number of common factors

We estimate the number of common factors - using the sequential testing procedure described in Section 4.2 (see also Section A.4 in the Internet Appendix for further details) - for three cases: (1) the unconditional model described in Section 2.1 (this case is denoted henceforth without using any subscripts); (2) the satisfactory regime $j = \mathcal{S}$; and (3) the downside counterpart $j = \mathcal{D}$.

Whilst the details on the implementation of the procedure are in Sections 4.2 and A.4, here we note that 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 0.01 and 0.12.²⁰ 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, which indicates robustness of the results.²¹ 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

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

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

Results are in Table 2. As mentioned above, we have tried different values of ε in the construc-

¹⁹An 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.

²⁰Recall that, according to Eq. (19), $\varepsilon > 0$ is required. We have tried positive values of ε smaller than 0.01, and results are the same as obtained with $\varepsilon = 0.01$.

²¹For the sake of completeness, we carry 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 carry out the individual tests using a nominal level $\alpha = \frac{0.05}{T}$.

tion of ϱ_j according to Eq. (19). As predicted by the theory, when ε is very close to 0, the number of estimated common factors is higher, and it tends to decrease as ε increases. Based on the estimates $\hat{P}_{\mathcal{D}}$ and $\hat{P}_{\mathcal{S}}$, 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, the number of common factors in the downside state, $\hat{P}_{\mathcal{D}}$ is found to be in the range between one and four, depending on the tuning parameter; as we know from the simulations in the Internet Appendix, $\hat{P}_{\mathcal{D}}$ has, if anything, a tendency to overstate $P_{\mathcal{D}}$, so an estimate of four common factors in regime $\mathcal{D}$ may be overly conservative. The number of common factors in the satisfactory regime, $\hat{P}_{\mathcal{S}}$ is generally greater than in the downside state (except for the smallest values of ε considered, when the number of factors is four for both regimes), being equal to three even for the largest value of ε considered. This means that the factor structure of asset returns tends to be 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).²² Second, applying the sequential testing procedure to the unconditional, linear model would lead to detecting a number of factors equal to or even larger than the highest number of factors (detected for the satisfactory regime), suggesting up to five factors. 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.

²²This 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).

6.3 Model fit

In this section, we compare the performance of the latent factor model with downside risk against a linear specification. 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, and a p-value for the null hypothesis that $\overline{AVG}_\alpha$ is the same for the unconditional and the conditional models. These p-values are 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 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 up to five. For the conditional model, the tests suggest three or four factors in Regime $\mathcal{S}$, whereas in Regime $\mathcal{D}$ the tests suggest a broad range between one and four factors.

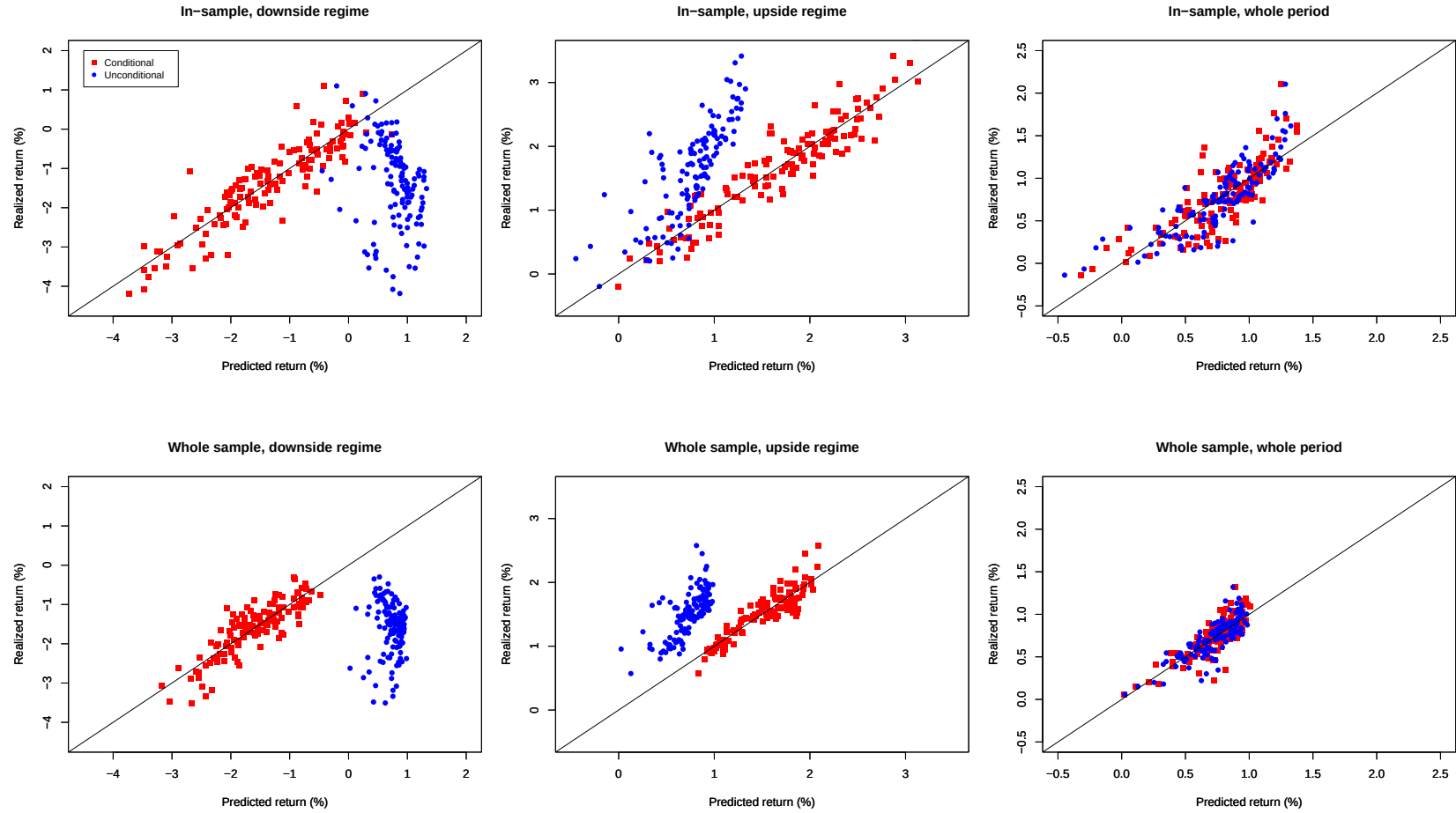
The results in Panel A of Table 3 show that the conditional model performs well, with a high R^2 both in the $\mathcal{S}$ regime and in the $\mathcal{D}$ regime. The results also suggest that the performance of the linear, unconditional model over the full sample is satisfactory with an R^2 of 0.928 (when the number of common factors P is set to its estimate $\hat{P}$ of five). However, we know from earlier evidence that the factor structure of the cross section 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 likely to be smaller 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 the $\mathcal{D}$ regime, and inferior to the conditional model also in the $\mathcal{S}$ regime.²³ 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

²³In fact, the R^2 turns negative for observations in the $\mathcal{D}$ regime, which can happen given that the unconditional model is fitted using all observations available and then assessed only on a subset of these observations.

errors with opposite sign from two regimes which generate small pricing errors on aggregate, over the full sample.

The relative performance of the unconditional and conditional models can be seen clearly in Figure 2, which reports scatter plots of realized versus model-implied returns for the unconditional model and for the conditional model (with three factors in downside, and four factors in the upside regime) under three different cases: the $\mathcal{D}$ regime, the $\mathcal{S}$ regime, and the full sample. This figure shows how the unconditional model (blue circles) performs very poorly in downside. Note the large errors in the top row of scatter plots in Figure 2: 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 (red squares) performs much better, and there is no systematic tendency to generate positive or negative errors in the downside regime. Turning to Regime $\mathcal{S}$, the conditional model again performs very well, 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 have a clear tendency to lie above the 45 degree line. This means that, even in Regime $\mathcal{S}$, the unconditional model generates 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 errors for a small number of observations in the $\mathcal{D}$ regime and more moderate pricing errors with the opposite sign for a large number of observations in the $\mathcal{S}$ regime – gives the appearance that the unconditional model performs well, but this is fallacious: the unconditional model performs poorly in *both* regimes.

Figure 2: Predicted and realized returns



Notes. For both the in-sample period February 1997 - August 2005, and the whole sample period February 1997 - August 2015, the scatter plot shows realized average returns against predicted average returns in downside, upside and over the whole period for the 130 equity portfolios described in Section 6.1 for both the conditional model (red dots) and the unconditional model (blue dots). Predicted returns for the conditional model are obtained with three factors in downside and with four factors in upside, and for the unconditional model with five factors.

Panel B of Table 3 provides corroborative numerical evidence on the differences in pricing errors illustrated graphically in Figure 2 for the sample 1997-2005. The $\overline{RMS}_\alpha$ for the unconditional model appears to be low (0.230 for five factors); yet the $\overline{RMS}_\alpha$ is very large in the $\mathcal{D}$ regime (over 2) and quite large also in the $\mathcal{S}$ regime (larger than 1). The conditional model produces a $\overline{RMS}_\alpha$ that is at least an order of magnitude smaller in both states of the world. Specifically, over the sample 1997-2005 the conditional model with three factors in Regime $\mathcal{D}$ and four factors in Regime $\mathcal{S}$ generates a $\overline{RMS}_\alpha$ of 0.53 in Regime $\mathcal{D}$ (more than four times smaller than the unconditional model), and 0.313 in Regime $\mathcal{S}$ (about one third the $\overline{RMS}_\alpha$ of the unconditional model).

Panel C of Table 3 sheds further light on the sign of the pricing errors by reporting the average α across the test assets, namely $\overline{AVG}_\alpha$, alongside its statistical significance based on the bootstrap test of the null that $\overline{AVG}_\alpha$ is equal to zero.²⁴ Again the unconditional model with five 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. The sign of $\overline{AVG}_\alpha$ 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 are at least one order of magnitude lower and statistically insignificantly different from zero. The average pricing errors of the conditional model become particularly small in magnitude when using three factors in Regime $\mathcal{D}$ and four factors in Regime $\mathcal{S}$ (consistent with the tests performed earlier).²⁵

Finally, Panel D of Table 3 reports results for the bootstrap tests of the null hypothesis that the difference in the average pricing errors between the conditional and the unconditional model is greater than or equal to zero against the alternative that it is lower than zero. Under the null hypothesis, the unconditional model performs at least as well as the conditional model at pricing the cross-section of equity returns, whereas under the alternative hypothesis the conditional model

²⁴It is worth noting that zero average pricing errors across portfolios is a requirement in our in-sample 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.

²⁵However, we view the tests for the number of factors as an indication of the number of factors required, and more factors may be allowed for in practical applications if they are required to fully span the information in the test asset returns.

performs better. In essence, this is the same null hypothesis of the Diebold-Mariano test for equal model forecast accuracy commonly applied in time series forecasting (see Diebold and Mariano (1995)), adapted for our purposes. The test results are clear-cut, in that the null is strongly rejected for each number of factors and each regime. This result again corroborates the superior performance of the conditional model.

To check robustness of these results, we also perform the full in-sample estimation of the unconditional and conditional models for the whole period from February 1997 to September 2015. The results, reported in Table 4, are consistent with the earlier evidence. The estimate of the disappointment threshold is 0.318, therefore just slightly higher than its estimate using data up to September 2005 (which was 0.210). In Table 4, we report the same three goodness-of-fit measures provided in Table 3, which all confirm the superiority of the conditional model under all three goodness-of-fit measures. The pricing accuracy of the model over the whole sample period is shown graphically in the bottom row of scatter plots in Figure 2, which confirm the deficiency of the unconditional model (especially in downside) and the accuracy of the conditional model in both states of the world. Overall, the results from the estimation carried out over the full sample 1997-2015 provide results that are qualitatively identical to the results obtained over the shorter 1997-2005 sample period.

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 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 characterize adequately returns in downside.²⁶

²⁶In a previous draft we also report the results from the estimation of the risk premia for various tradable and non-tradable factors, obtained using the three-pass procedure detailed in Section 5. We omit these results here for brevity, but recall that the estimation of conditional risk premia is straightforward as it follows the formulae given

6.4 Out-of-sample analysis

Next, we assess the *out-of-sample* performance of the conditional model relative to the unconditional model, using the last ten years of data (sample from October 2005 to September 2015) as the “testing sample” on which we carry out our forecasting exercise. In this application, we set the estimated disappointment threshold $\hat{\theta}$ equal to the in-sample estimate (obtained on the 1997-2005 sample) of 0.210, and keep it constant throughout the out-of-sample analysis. We estimate the factor structure recursively, i.e. conditioning on information available at time $t - 1$ to produce forecasts of excess returns at time t . The same recursive procedure is applied to the unconditional asset pricing model.

Formally, given the in-sample window of length T , let the out-of-sample window be equal to T_{OoS} . For $t_{OoS} = T, \dots, T + T_{OoS} - 1$, the forecast of $\mathbf{R}_{t_{OoS}+1}$ from the conditional model is

$$\widehat{\mathbf{R}}_{\mathcal{C}, t_{OoS}+1} = \hat{d}_{\mathcal{D}, t_{OoS}+1} \widehat{\mathbf{X}}_{\mathcal{D}, t_{OoS}} \widehat{\mathbf{\Gamma}}_{\mathcal{D}, t_{OoS}} + \hat{d}_{\mathcal{S}, t_{OoS}+1} \widehat{\mathbf{X}}_{\mathcal{S}, t_{OoS}} \widehat{\mathbf{\Gamma}}_{\mathcal{S}, t_{OoS}}, \quad (28)$$

where $\widehat{\mathbf{X}}_{j, t_{OoS}}$ and $\widehat{\mathbf{\Gamma}}_{j, t_{OoS}}$ are the recursive estimates of $\mathbf{X}_{j, t_{OoS}}$ and $\mathbf{\Gamma}_{j, t_{OoS}}$, respectively, obtained using information only available for $1 \leq t \leq t_{OoS}$. Note that, given the estimated disappointment threshold $\hat{\theta} = 0.21$, the indicators $\hat{d}_{\mathcal{D}, t_{OoS}+1}$ and $\hat{d}_{\mathcal{S}, t_{OoS}+1}$ are known at time t_{OoS} as they depend only on the value of the AD Bear portfolio return at time t_{OoS} - thus, our forecasting exercise is genuinely, fully out-of-sample. Similarly to (28), the forecast of $\mathbf{R}_{t_{OoS}+1}$ from the unconditional model is

$$\widehat{\mathbf{R}}_{\mathcal{U}, t_{OoS}+1} = \widehat{\mathbf{X}}_{t_{OoS}} \widehat{\mathbf{\Gamma}}_{t_{OoS}}, \quad (29)$$

where $\widehat{\mathbf{X}}_{t_{OoS}}$ and $\widehat{\mathbf{\Gamma}}_{t_{OoS}}$ are the recursive estimate for $\mathbf{X}_{t_{OoS}}$ and $\mathbf{\Gamma}_{t_{OoS}}$, respectively, obtained using only information available for $t = 1, \dots, t_{OoS}$. The average forecast values from the conditional model within each regime and from the unconditional model (namely, $\widehat{\mathbf{R}}_{\mathcal{C}, j, OoS}$ and $\widehat{\mathbf{R}}_{\mathcal{U}, OoS}$, respectively) are computed as the averages of $\widehat{\mathbf{R}}_{\mathcal{C}, t_{OoS}+1}$ (in both Regimes $\mathcal{D}$ and $\mathcal{S}$) and $\widehat{\mathbf{R}}_{\mathcal{U}, t_{OoS}+1}$, respectively. These average forecasts are then compared against the corresponding realized average values of $\mathbf{R}_{t_{OoS}+1}$ for the out-of-sample period, calculated both over the whole period and

in Section 5.

within each regime (namely, $\overline{\mathbf{R}}_{j,OoS}$ and $\overline{\mathbf{R}}_{OoS}$, respectively).

The results are reported in Table 5, where we present the same statistics conducted for the in-sample analysis, except that they are computed over the out-of-sample period using the recursive procedure described above. The out-of-sample R^2 (R_{OoS}^2) follows a similar pattern to the in-sample results, since the conditional model generates high R_{OoS}^2 in both regimes, whereas the unconditional model performs very poorly in Regime $\mathcal{D}$, where it displays negative values of R_{OoS}^2 . Again, over the full set of data of the out-of-sample period, the unconditional model gives the false impression to fit the data well, with the R_{OoS}^2 being always above 0.8 for each number of factors considered, but this masks its poor performance in the downside regime. These results are qualitatively identical to the in-sample results. In terms of the out-of-sample root-mean-squared pricing error ($\overline{RMS}_{\alpha,OoS}$), again the results are qualitatively identical to the in-sample counterpart in that the $\overline{RMS}_{\alpha,OoS}$ is substantially larger for the unconditional model than for the conditional model in both regimes, confirming the superior performance of the conditional model. For example, the unconditional model with five factors (as indicated by the tests for the number of factors) produces a $\overline{RMS}_{\alpha,OoS}$ of 1.848 in Regime $\mathcal{D}$ and 0.782 in Regime $\mathcal{S}$, whereas the conditional model produces an $\overline{RMS}_{\alpha,OoS}$ of 0.868 in Regime $\mathcal{D}$ and 0.533 in Regime $\mathcal{S}$ (with three and four factors, respectively).

Turning to the average pricing error ($\overline{AVG}_{\alpha,OoS}$) we notice again the same pattern observed in-sample in terms of their relative magnitude. However, the size of the average pricing errors is higher out-of-sample for both the conditional model and the unconditional model. Specifically, the unconditional model with five factors displays a low $\overline{AVG}_{\alpha}$, although it is statistically significantly different from zero. 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 contrast, the conditional model delivers much smaller average pricing errors in the two regimes. This evidence is consistent with the results obtained in-sample.

However, the key difference between the in-sample results and the out-of-sample results is that the average pricing errors are statistically different from zero in both regimes not only for the unconditional model, but also for the conditional model, on the basis of our bootstrap tests (see Panel C of Table 5). Therefore, while the conditional model produces much lower average pricing

errors (in absolute terms) than the unconditional benchmark, some alpha remains to be explained. Hence, we also formally test the null hypothesis that the difference in the average pricing error between the conditional and the unconditional model is statistically significantly different from zero. These further bootstrap tests are reported in the bottom part, Panel D of Table 5, and suggest clearly that the null hypothesis that the unconditional model performs at least as well as the conditional model is strongly rejected, thus corroborating the superior performance of the conditional model.

Overall, the empirical evidence reported above suggests that, as one might expect, there is some deterioration in performance when taking the conditional model out-of-sample, but its superiority relative to the unconditional benchmark continues to be as evident out-of-sample as it is in-sample.

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. We extend the framework used in the extant literature in three ways: we allow for downside risk via a conditional 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 provide analytical solutions to recover conditional observable factors' risk premia from the estimated latent ones in both regimes.

The proposed conditional model is illustrated through an empirical application to a large cross-section of equity returns. The results show that estimating the number of common factors separately yields very different estimates across regimes, with the downside risk state having a smaller number of common factors, whilst returns in the good state of the world are driven by more common factors. More generally, the empirical analysis shows clearly the limitations of unconditional asset pricing models that ignore state dependence, while illustrating that the proposed conditional model is capable to adequately characterize the key properties of the cross-

section of equity returns.

Yet, we merely view our model as a tentatively 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 conditional specifications. Future work could focus, for example, on endogeneizing the threshold that identifies the downside state as a function of state variables, on alternative variables to include in the switching mechanism that captures the state of the world, on allowing explicitly for weak pricing factors, and on richer characterizations of out-of-sample forecasting of the conditional factor structure. Also, it seems plausible that the conditional model proposed here could be employed in different applications in various risk management contexts, for example to examine CoVaR or marginal expected shortfall in order to better understand how market crashes can co-arise through different economic structures. 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.
- Arrow, K. J. (1964). The Role of Securities in the Optimal Allocation of Risk-bearing. *Review of Economic Studies* 31(2), 91–96.
- 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. Quaadvlieg (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.
- Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* 51(5), 1281–1304.
- 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.
- Colacito, R., S. J. Riddiough, and L. Sarno (2020). Business cycles and currency returns. *Journal of Financial Economics* 137(3), 659–678.

- Davies, N. and P. Newbold (1980). Forecasting with misspecified models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29(1), 87–92.
- Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*. Yale University Press.
- 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.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & economic statistics* 13(3), 253–263.
- 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.
- Fortin, A.-P., P. Gagliardini, and O. Scaillet (2023). Latent factor analysis in short panels. *Journal of Financial Econometrics*. forthcoming.
- Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica* 84(3), 985–1046.
- Gagliardini, P., E. Ossola, and O. Scaillet (2019). A diagnostic criterion for approximate factor structure. *Journal of Econometrics* 212(2), 503–521.
- Giglio, S., Y. Liao, and D. Xiu (2021). Thousands of alpha tests. *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.
- 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.
- Lu, Z. and S. Murray (2019). Bear beta. *Journal of Financial Economics* 131(3), 736–760.
- 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 (2012). Carry trades and global foreign exchange volatility. *Journal of Finance* 67, 681–718.
- 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.
- Page, S. and R. A. Panariello (2018). When diversification fails. *Financial Analysts Journal* 74, 19–32.
- Patton, A. J. and T. Ramadorai (2013). On the high-frequency dynamics of hedge fund risk exposures. *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.
- Phillips, P. C. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
- Routledge, B. R. and S. E. Zin (2010). Generalized disappointment aversion and asset prices. *Journal of Finance* 65(4), 1303–1332.
- 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.
- Zaffaroni, P. (2019). Factor models for asset pricing. *Available at SSRN 3398169*.

Table 1: Estimation of θ^0

$\hat{\theta}$	0.210
T	103
$\hat{T}_{\mathcal{D}}$	30
$\hat{\pi}_{\mathcal{D}}$	0.291
$\overline{RmRf}$	0.463
$\overline{RmRf}_{\mathcal{D}}$	-0.883
$\overline{RmRf}_{\mathcal{S}}$	0.955
$Corr\left(\hat{\mathcal{D}}_t, RmRf\right)$	-0.171*
$Corr\left(\hat{\mathcal{D}}_t, REC\right)$	0.180*
$Corr\left(\hat{\mathcal{D}}_t, \Delta IP\right)$	-0.059
$Corr\left(\hat{\mathcal{D}}_t, VIX\right)$	0.308***
$Corr\left(\hat{\mathcal{D}}_t, VXO\right)$	0.309***
$Corr\left(\hat{\mathcal{D}}_t, EPU\right)$	0.245**
$Corr\left(\hat{\mathcal{D}}_t, EMV\right)$	0.410***
$Corr\left(\hat{\mathcal{D}}_t, PREM V\right)$	0.416***
$Corr\left(\hat{\mathcal{D}}_t, GPR\right)$	0.116
$Corr\left(\hat{\mathcal{D}}_t, TED\right)$	0.014
$Corr\left(\hat{\mathcal{D}}_t, SPX\right)$	0.230**
$Corr\left(\hat{\mathcal{D}}_t, T10Y2Y\right)$	0.010

Notes. For the period February 1997 - August 2005, the table reports the estimated value of θ^0 ($\hat{\theta}$), the total number of time series observations (T), the estimated number of periods in the downside regime ($\hat{T}_{\mathcal{D}}$), the estimated frequency of downside events ($\hat{\pi}_{\mathcal{D}}$), the average value of the stock market return in excess of the risk-free rate over the whole period, as well as in downside and upside ($\overline{RmRf}$, $\overline{RmRf}_{\mathcal{D}}$ and $\overline{RmRf}_{\mathcal{S}}$, respectively). The table also reports the correlation between the estimated downside indicator function $\mathbb{I}\left(\hat{\mathcal{D}}_t\right)$ and the following variables: $RmRf$ is the market return in excess of the risk-free rate; REC is the NBER U.S. recession indicator; ΔIP is the year-on-year log-change in 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. ***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

Table 2: Number of estimated common factors

ε	Number of factors			Eigenvalue	% of variance explained		
	Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$		Full Sample	Regime $\mathcal{D}$	Regime $\mathcal{S}$
0.01	5	4	4	1	0.734	0.765	0.703
0.02	5	4	4	2	0.106	0.098	0.123
0.04	4	3	4	3	0.043	0.044	0.046
0.06	4	2	4	4	0.035	0.030	0.034
0.08	4	2	4	5	0.015	0.016	0.017
0.10	4	2	4	6	0.008	0.008	0.011
0.12	4	1	3				

Notes. For the period February 1997 - August 2005, the table contains the number of estimated common factors for different values of ε defined in Eq. (19). It reports inference on the number of common factors over the whole period, as well as in the two regimes $\mathcal{S}$ and $\mathcal{D}$. 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: In-sample goodness of fit measures

Panel A: R^2					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.689	-1.019	0.560	0.720	0.851
$f_{1t} - f_{2t}$	0.691	-1.035	0.567	0.855	0.870
$f_{1t} - f_{3t}$	0.895	-1.108	0.661	0.905	0.938
$f_{1t} - f_{4t}$	0.896	-1.107	0.661	0.928	0.969
$f_{1t} - f_{5t}$	0.928	-1.091	0.665	0.944	0.979
$f_{1t} - f_{6t}$	0.929	-1.091	0.666	0.946	0.981

Panel B: RMS_α					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.478	2.441	1.184	0.909	0.690
$f_{1t} - f_{2t}$	0.476	2.451	1.174	0.655	0.642
$f_{1t} - f_{3t}$	0.278	2.494	1.039	0.530	0.446
$f_{1t} - f_{4t}$	0.277	2.494	1.039	0.462	0.313
$f_{1t} - f_{5t}$	0.230	2.484	1.033	0.408	0.256
$f_{1t} - f_{6t}$	0.229	2.484	1.032	0.401	0.246

Panel C: AVG_α					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.074	-2.080***	0.960***	0.072	0.054
$f_{1t} - f_{2t}$	0.082	-2.072***	0.968***	-0.038	0.108
$f_{1t} - f_{3t}$	0.019	-2.136***	0.905***	0.028	0.072
$f_{1t} - f_{4t}$	0.020	-2.135***	0.906***	0.030	0.024
$f_{1t} - f_{5t}$	0.004	-2.150***	0.890***	0.014	0.003
$f_{1t} - f_{6t}$	0.005	-2.150***	0.891***	0.015	0.004

Panel D: $H_0 : \Delta \overline{AVG}_\alpha \geq 0$		
	Regime D	Regime S
f_{1t}	-2.008***	-0.906***
$f_{1t} - f_{2t}$	-2.034***	-0.860***
$f_{1t} - f_{3t}$	-2.108***	-0.832***
$f_{1t} - f_{4t}$	-2.105***	-0.882***
$f_{1t} - f_{5t}$	-2.136***	-0.888***
$f_{1t} - f_{6t}$	-2.135***	-0.887***

Notes. For the period February 1997 - August 2005, this table shows in-sample goodness of fit measures as a function of the number of factors: R-squared (R^2); root-mean-squared pricing error (RMS_α); average pricing error (AVG_α); test statistic for the null hypothesis that $|\overline{AVG}_\alpha|$ of the conditional model is greater than or equal to $|\overline{AVG}_\alpha|$ of the unconditional model within each regime ($H_0 : \Delta |\overline{AVG}_\alpha| \geq 0$). They are obtained for the unconditional model over the whole period, and in $j = \mathcal{D}, \mathcal{S}$; and for the conditional model in $j = \mathcal{D}, \mathcal{S}$. For the unconditional model, the pricing error over the whole period is $\hat{\alpha}_U = \bar{\mathbf{R}} - \hat{\mathbf{X}}\hat{\mathbf{\Gamma}}$ where $\bar{\mathbf{R}} = \hat{T}^{-1} \sum_{t=1}^T \mathbf{R}_t$, and the pricing error in $j = \mathcal{D}, \mathcal{S}$ is $\hat{\alpha}_{j,U} = \bar{\mathbf{R}}_j - \hat{\mathbf{X}}_j\hat{\mathbf{\Gamma}}$ 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 $j = \mathcal{D}, \mathcal{S}$ is $\hat{\alpha}_j = \bar{\mathbf{R}}_j - \hat{\mathbf{X}}_j\hat{\mathbf{\Gamma}}_j$. R^2 is: $1 - (\hat{\alpha}'_U \hat{\alpha}_U / \bar{\mathbf{R}}'\bar{\mathbf{R}}) \left[(N-1) / (N - \hat{P}) \right]$ for the unconditional model over the whole period; $1 - (\hat{\alpha}'_{j,U} \hat{\alpha}_{j,U} / \bar{\mathbf{R}}'\bar{\mathbf{R}}) \left[(N-1) / (N - \hat{P}) \right]$ for the unconditional model in $j = \mathcal{D}, \mathcal{S}$; $1 - (\hat{\alpha}'_j \hat{\alpha}_j / \bar{\mathbf{R}}'_j \bar{\mathbf{R}}_j) \left[(N-1) / (N - \hat{P}_j) \right]$ for the conditional model in $j = \mathcal{D}, \mathcal{S}$. RMS_α is: $\sqrt{\hat{\alpha}'_U \hat{\alpha}_U / N}$ for the unconditional model over the whole period; $\sqrt{\hat{\alpha}'_{j,U} \hat{\alpha}_{j,U} / N}$ for the unconditional model in $j = \mathcal{D}, \mathcal{S}$; $\sqrt{\hat{\alpha}'_j \hat{\alpha}_j / N}$ for the conditional model in $j = \mathcal{D}, \mathcal{S}$. *** indicates significance at 1% level. p-values are obtained with the bootstrap in Appendix E.1.

Table 4: Whole sample estimation

Panel A: θ^0					
$\hat{\theta}$	0.318				
T	223				
$\hat{T}_{\mathcal{D}}$	56				
$\hat{\pi}_{\mathcal{D}}$	0.318				
$\overline{RmRf}$	0.536				
$\overline{RmRf}_{\mathcal{D}}$	-1.241				
$\overline{RmRf}_{\mathcal{S}}$	1.105				
Panel B: R^2					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.860	-1.059	0.667	0.933	0.946
$f_{1t} - f_{2t}$	0.861	-1.063	0.670	0.940	0.951
$f_{1t} - f_{3t}$	0.943	-1.066	0.697	0.953	0.953
$f_{1t} - f_{4t}$	0.963	-1.069	0.705	0.960	0.988
$f_{1t} - f_{5t}$	0.964	-1.069	0.705	0.960	0.990
$f_{1t} - f_{6t}$	0.964	-1.067	0.704	0.960	0.990
Panel C: $\overline{RMS}_{\alpha}$					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.279	2.431	0.875	0.439	0.353
$f_{1t} - f_{2t}$	0.278	2.433	0.872	0.413	0.336
$f_{1t} - f_{3t}$	0.179	2.435	0.835	0.367	0.329
$f_{1t} - f_{4t}$	0.143	2.437	0.824	0.341	0.168
$f_{1t} - f_{5t}$	0.142	2.437	0.824	0.341	0.154
$f_{1t} - f_{6t}$	0.141	2.436	0.825	0.340	0.153
Panel D: $\overline{AVG}_{\alpha}$					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.035	-2.296***	0.817***	0.044	0.036
$f_{1t} - f_{2t}$	0.038	-2.294***	0.820***	0.019	0.047
$f_{1t} - f_{3t}$	0.005	-2.326***	0.787***	0.014	0.039
$f_{1t} - f_{4t}$	0.005	-2.326***	0.788***	0.011	0.007
$f_{1t} - f_{5t}$	0.004	-2.327***	0.787***	0.012	0.003
$f_{1t} - f_{6t}$	0.004	-2.328***	0.786***	0.011	0.003
Panel E: $H_0 : \Delta \overline{AVG}_{\alpha} \geq 0$					
	Regime D		Regime S		
f_{1t}	-2.253***		-0.781***		
$f_{1t} - f_{2t}$	-2.275***		-0.773***		
$f_{1t} - f_{3t}$	-2.313***		-0.748***		
$f_{1t} - f_{4t}$	-2.314***		-0.780***		
$f_{1t} - f_{5t}$	-2.315***		-0.784***		
$f_{1t} - f_{6t}$	-2.316***		-0.783***		

Notes. For the period February 1997 - August 2015, this table shows estimation results. $\hat{T}_{\mathcal{D}}$, $\hat{\pi}_{\mathcal{D}}$, $\overline{RmRf}$, $\overline{RmRf}_{\mathcal{D}}$ and $\overline{RmRf}_{\mathcal{S}}$ are as in Table 1. R^2 , $\overline{RMS}_{\alpha}$, $\overline{AVG}_{\alpha}$, and $H_0 : \Delta |\overline{AVG}_{\alpha}| \geq 0$ are as in Table 3. *** indicates significance at 1% level. p-values are obtained with the bootstrap in Appendix E.1.

Table 5: Out-of-sample goodness of fit measures

Panel A: R^2_{OoS}					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.860	-1.214	0.675	0.708	0.908
$f_{1t} - f_{2t}$	0.843	-1.226	0.665	0.581	0.851
$f_{1t} - f_{3t}$	0.883	-1.247	0.678	0.497	0.842
$f_{1t} - f_{4t}$	0.857	-1.282	0.672	0.446	0.847
$f_{1t} - f_{5t}$	0.841	-1.316	0.669	0.421	0.846
$f_{1t} - f_{6t}$	0.842	-1.333	0.666	0.414	0.842

Panel B: $\overline{RMS}_{\alpha, OoS}$					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.266	1.835	0.788	0.667	0.419
$f_{1t} - f_{2t}$	0.281	1.833	0.796	0.796	0.532
$f_{1t} - f_{3t}$	0.241	1.835	0.777	0.868	0.544
$f_{1t} - f_{4t}$	0.266	1.841	0.782	0.907	0.533
$f_{1t} - f_{5t}$	0.279	1.848	0.782	0.924	0.534
$f_{1t} - f_{6t}$	0.277	1.847	0.782	0.926	0.538

Panel C: $\overline{AVG}_{\alpha, OoS}$					
	Unconditional Model			Conditional Model	
	Full Sample	Regime D	Regime S	Regime D	Regime S
f_{1t}	0.081***	-1.720***	0.764***	0.523***	-0.177 ***
$f_{1t} - f_{2t}$	0.086***	-1.716***	0.769***	0.459***	-0.149***
$f_{1t} - f_{3t}$	0.045***	-1.756***	0.728***	0.496***	-0.155***
$f_{1t} - f_{4t}$	0.046***	-1.755***	0.730***	0.498***	-0.206***
$f_{1t} - f_{5t}$	0.042***	-1.760***	0.725***	0.492***	-0.214***
$f_{1t} - f_{6t}$	0.041***	-1.760***	0.724***	0.492***	-0.213***

Panel D: $H_0 : \Delta \overline{AVG}_{\alpha} \geq 0$		
	Regime D	Regime S
f_{1t}	-1.198***	-0.587***
$f_{1t} - f_{2t}$	-1.257***	-0.620***
$f_{1t} - f_{3t}$	-1.260***	-0.573***
$f_{1t} - f_{4t}$	-1.257***	-0.524***
$f_{1t} - f_{5t}$	-1.267***	-0.510***
$f_{1t} - f_{6t}$	-1.268***	-0.511***

Notes. For the period September 2005 - August 2015, this table shows out-of-sample goodness of fit measures as a function of the number of factors: R-squared (R^2_{OoS}); root-mean-squared pricing error ($\overline{RMS}_{\alpha, OoS}$); average pricing error ($\overline{AVG}_{\alpha, OoS}$); test statistic for the null hypothesis that $|\overline{AVG}_{\alpha, OoS}|$ of the conditional model is greater than or equal to $|\overline{AVG}_{\alpha, OoS}|$ of the unconditional model within each regime ($H_0 : \Delta |\overline{AVG}_{\alpha, OoS}| \geq 0$). They are obtained for the unconditional model over the whole period, and in $j = \mathcal{D}, \mathcal{S}$; and for the conditional model in $j = \mathcal{D}, \mathcal{S}$. Given the quantities in Section 6.4, for the unconditional model the pricing error over the whole period is $\hat{\alpha}_{\mathcal{U}, OoS} = \overline{\mathbf{R}}_{OoS} - \overline{\mathbf{R}}_{\mathcal{U}, OoS}$, and is $\hat{\alpha}_{j, \mathcal{U}, OoS} = \overline{\mathbf{R}}_{j, OoS} - \overline{\mathbf{R}}_{\mathcal{U}, OoS}$ in $j = \mathcal{D}, \mathcal{S}$. The pricing error in $j = \mathcal{D}, \mathcal{S}$ for the conditional model is $\hat{\alpha}_{j, OoS} = \overline{\mathbf{R}}_{j, OoS} - \overline{\mathbf{R}}_{\mathcal{C}, j, OoS}$. R^2_{OoS} is: $1 - \left(\hat{\alpha}'_{\mathcal{U}, OoS} \hat{\alpha}_{\mathcal{U}, OoS} / \overline{\mathbf{R}}'_{OoS} \overline{\mathbf{R}}_{OoS} \right) \left[(N-1) / (N - \hat{P}) \right]$ for the unconditional model over the whole period; $1 - \left(\hat{\alpha}'_{j, \mathcal{U}, OoS} \hat{\alpha}_{j, \mathcal{U}, OoS} / \overline{\mathbf{R}}'_{OoS} \overline{\mathbf{R}}_{OoS} \right) \left[(N-1) / (N - \hat{P}) \right]$ for the unconditional model in $j = \mathcal{D}, \mathcal{S}$; $1 - \left(\hat{\alpha}'_{j, OoS} \hat{\alpha}_{j, OoS} / \overline{\mathbf{R}}'_{j, OoS} \overline{\mathbf{R}}_{j, OoS} \right) \left[(N-1) / (N - \hat{P}_j) \right]$ for the conditional model in $j = \mathcal{D}, \mathcal{S}$. $\overline{RMS}_{\alpha, OoS}$ is: $\sqrt{\hat{\alpha}'_{\mathcal{U}, OoS} \hat{\alpha}_{\mathcal{U}, OoS} / N}$ for the unconditional model over the whole period; $\sqrt{\hat{\alpha}'_{j, \mathcal{U}, OoS} \hat{\alpha}_{j, \mathcal{U}, OoS} / N}$ for the unconditional model in $j = \mathcal{D}, \mathcal{S}$; $\sqrt{\hat{\alpha}'_{j, OoS} \hat{\alpha}_{j, OoS} / N}$ for the conditional model in $j = \mathcal{D}, \mathcal{S}$. *** indicates significance at 1% level. p-values are obtained as in Appendix E.2.

Internet Appendix

(not for publication)

Factor models with downside risk

A Complements

In this section of the Internet Appendix, we add a more in-depth analysis of our model, and complement the theory spelt out in Sections 4 and 5, adding further details and explanations. In particular, in Section A.1, we describe some properties of the conditional and unconditional model specifications. In Section A.2, we list and discuss all the relevant assumptions used in the paper. In Section A.3, we discuss in greater depth the test presented in Section 4.1. In Section A.4, 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.5, we complete the asymptotic theory developed in Section 5.3.

Here and in the rest of the appendix, we use the notation $c_0, c_1, \dots$ 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 Model analysis

In this section, we analyze the structure of the econometric specifications used in our paper - that is, the structure of both the *conditional* and the *unconditional* specifications.

In particular, we formally illustrate the nature of the misspecification in the standard, unconditional model used in asset pricing, when the true model is asymmetric due to the presence of a downside risk premium. We primarily focus on showing that, when using the unconditional model, in the presence of a downside risk premium, this results in nonzero *cross-sectional pricing* errors. We also briefly touch on the predictive ability of the two specifications, considering - in addition to the previously considered cross-sectional pricing errors - the *time series predictive* errors, thus analyzing the impact of choosing a conditional versus an unconditional specification along both dimensions.

In order to study both types of errors, consider the following simplified version of (1), with only one factor common to both states

$$R_{i,t} = \beta_{\mathcal{S},i} f_t \mathbb{I}(\mathcal{S}_t) + \beta_{\mathcal{D},i} f_t \mathbb{I}(\mathcal{D}_t) + \epsilon_{i,t}, \quad (\text{A.1})$$

with $1 \leq t \leq T$ and $1 \leq i \leq N$. To ensure that the two states $\mathcal{S}_t$ and $\mathcal{D}_t$ are well-defined, we remark that $\beta_{\mathcal{S},i} \neq \beta_{\mathcal{D},i}$ and that $0 < \mathbb{P}(\mathcal{D}_t) = \pi_{\mathcal{D}} < 1$, with $\mathbb{P}(\mathcal{S}_t) = \pi_{\mathcal{S}} = 1 - \pi_{\mathcal{D}}$. We also define

$$\begin{aligned} E_{\mathcal{C}}(f_t|\mathcal{D}_t) &= \mu_{\mathcal{D}}, \\ E_{\mathcal{C}}(f_t|\mathcal{S}_t) &= \mu_{\mathcal{S}}, \end{aligned}$$

where we use the notation “ $E_{\mathcal{C}}$ ” to emphasize that we refer to the conditional specification of (A.1). Hence, it follows that the expected excess returns conditional upon either state $\mathcal{S}_t$ and $\mathcal{D}_t$ are given by

$$\begin{aligned} E_{\mathcal{C}}(R_{i,t}|\mathcal{D}_t) &= \beta_{\mathcal{D},i}\mu_{\mathcal{D}}, \\ E_{\mathcal{C}}(R_{i,t}|\mathcal{S}_t) &= \beta_{\mathcal{S},i}\mu_{\mathcal{S}}. \end{aligned}$$

Conversely, the unconditional excess returns corresponding to (A.1) are equal to

$$E_{\mathcal{C}}(R_{i,t}) = \beta_{\mathcal{D},i}\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \beta_{\mathcal{S},i}\mu_{\mathcal{S}}\pi_{\mathcal{S}}.$$

Consider now the case where the researcher uses an unconditional specification in lieu of (A.1), viz.

$$R_{i,t} = \beta_i f_t + u_{i,t}, \tag{A.2}$$

obtained by constructing β_i so that $E(u_{i,t}) = 0$. In such a case, using the notation “ $E_{\mathcal{U}}$ ” to refer to the unconditional specification of (A.2), it holds that

$$\begin{aligned} E_{\mathcal{U}}(R_{i,t}) &= \beta_i E(f_t) = \beta_i [E(f_t|\mathcal{D}_t)\mathbb{P}(\mathcal{D}_t) + E(f_t|\mathcal{S}_t)\mathbb{P}(\mathcal{S}_t)] \\ &= \beta_i (\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}). \end{aligned}$$

Hence, using (A.2) introduces pricing errors in both states, seeing as

$$\begin{aligned}\alpha_{\mathcal{D},i} &= E_{\mathcal{C}}(R_{i,t}|\mathcal{D}_t) - E_{\mathcal{U}}(R_{i,t}) \\ &= \beta_{\mathcal{D},i}\mu_{\mathcal{D}} - \beta_i(\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}),\end{aligned}$$

and

$$\begin{aligned}\alpha_{\mathcal{S},i} &= E_{\mathcal{C}}(R_{i,t}|\mathcal{S}_t) - E_{\mathcal{U}}(R_{i,t}) \\ &= \beta_{\mathcal{S},i}\mu_{\mathcal{S}} - \beta_i(\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}).\end{aligned}$$

Similarly, the unconditional pricing model makes a pricing error even “on average”, i.e. unconditionally

$$\begin{aligned}\alpha_i &= E_{\mathcal{C}}(R_{i,t}) - E_{\mathcal{U}}(R_{i,t}) \\ &= \beta_{\mathcal{D},i}\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \beta_{\mathcal{S},i}\mu_{\mathcal{S}}\pi_{\mathcal{S}} - \beta_i(\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}),\end{aligned}$$

which is in general nonzero. Conversely, when only one regime is present, i.e. when $\beta_{\mathcal{S},i} = \beta_{\mathcal{D},i} = \beta_i$, it holds that

$$\begin{aligned}\alpha_i &= \beta_{\mathcal{D},i}\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \beta_{\mathcal{S},i}\mu_{\mathcal{S}}\pi_{\mathcal{S}} - \beta_i(\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}) \\ &= \beta_i(\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}) - \beta_i(\mu_{\mathcal{D}}\pi_{\mathcal{D}} + \mu_{\mathcal{S}}\pi_{\mathcal{S}}) = 0.\end{aligned}$$

We now turn to considering (suppressing, for simplicity, the subscript i) the predictive error of both models

$$R_t = \beta_{\mathcal{S}}f_t\mathbb{I}(\mathcal{S}_t) + \beta_{\mathcal{D}}f_t\mathbb{I}(\mathcal{D}_t) + \epsilon_t, \tag{A.3}$$

$$R_t = \beta f_t + u_t, \tag{A.4}$$

with again β constructed such that $E(u_t) = 0$. The one-step-ahead predictions are defined as

follows: the conditional model (A.3) predicts *asymmetrically*, on the grounds of the predicted state, viz.

$$\widehat{R}_{t+1|t}^{(c)} = \beta_S E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1}) + \beta_D E_t(f_{t+1}) \mathbb{I}(\mathcal{D}_{t+1}); \quad (\text{A.5})$$

conversely, the unconditional model in (A.4) predicts *symmetrically*, with

$$\widehat{R}_{t+1|t}^{(u)} = \beta E_t(f_{t+1}), \quad (\text{A.6})$$

respectively.ⁱ

Let

$$\begin{aligned} MSPE^{(c)} &= E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(c)} \right)^2, \\ MSPE^{(u)} &= E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(u)} \right)^2, \end{aligned}$$

denote the Mean Squared Predictive Errors of both models.

Theorem A.1. *We assume that $\{\epsilon_t\}$ and $\{f_t, Z_{t-1}\}$ are mutually independent, with $E(\epsilon_t) = 0$. Then it holds that*

$$MSPE^{(u)} \geq MSPE^{(c)},$$

with equality if and only if $\beta_S = \beta_D = \beta$.

Theorem A.1 states that, in the genuine presence of different regimes, the unconditional specification is “mis-specified”, in that it incurs a higher cost (expressed in terms of expected squared error of prediction) than a conditional specification (see Davies and Newbold, 1980). The “strong exogeneity” assumption that $\{\epsilon_t\}$ and $\{f_t, Z_{t-1}\}$ are mutually independent is stronger than necessary, and we make it only to simplify the main arguments in the proof, reported below.

ⁱNote that this (and the following ones) is an *infeasible* prediction, being based on the true parameter values β_S and β_D . In an asymptotic set-up, where parameters are estimated consistently, this simplification does not impact on the results reported in this section.

Note also that (A.5) corresponds to

$$\widehat{R}_{t+1|t}^{(c)} = \begin{cases} \beta_S E_t(f_{t+1}), & \text{if } E_t(Z_{t+1}) > \theta, \\ \beta_D E_t(f_{t+1}), & \text{if } E_t(Z_{t+1}) \leq \theta. \end{cases}$$

Proof of Theorem A.1. We begin by computing $MSPE^{(C)}$. It holds that

$$\begin{aligned} & E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(C)} \right)^2 \\ &= E \left[\epsilon_{t+1} + \beta_S (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{S}_{t+1}) + \beta_D (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{D}_{t+1}) \right]^2. \end{aligned}$$

We note that $E(\epsilon_{t+1} (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{S}_{t+1})) = 0$ by the independence between $\{\epsilon_t\}$ and $\{f_t, Z_{t-1}\}$.

Hence the expression above simplifies into

$$\begin{aligned} & E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(C)} \right)^2 \tag{A.7} \\ &= E \left(\epsilon_{t+1}^2 \right) + E \left[\beta_S (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{S}_{t+1}) + \beta_D (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{D}_{t+1}) \right]^2 \\ &= E \left(\epsilon_{t+1}^2 \right) + \beta_S^2 E \left[(f_{t+1} - E_t(f_{t+1}))^2 \mathbb{I}(\mathcal{S}_{t+1}) \right] + \beta_D^2 E \left[(f_{t+1} - E_t(f_{t+1}))^2 \mathbb{I}(\mathcal{D}_{t+1}) \right]. \end{aligned}$$

We now study $MSPE^{(U)}$; by construction we have

$$\begin{aligned} & E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(U)} \right)^2 \\ &= E \left[\epsilon_{t+1} + \beta_S f_{t+1} \mathbb{I}(\mathcal{S}_{t+1}) + \beta_D f_{t+1} \mathbb{I}(\mathcal{D}_{t+1}) - \beta E_t(f_{t+1}) \right]^2 \\ &= E \left[\epsilon_{t+1} + \beta_S f_{t+1} \mathbb{I}(\mathcal{S}_{t+1}) + \beta_D f_{t+1} \mathbb{I}(\mathcal{D}_{t+1}) - \beta E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1}) - \beta E_t(f_{t+1}) \mathbb{I}(\mathcal{D}_{t+1}) \right]^2 \\ &= E \left[\epsilon_{t+1} + \beta_S (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{S}_{t+1}) + \beta_D (f_{t+1} - E_t(f_{t+1})) \mathbb{I}(\mathcal{D}_{t+1}) \right. \\ &\quad \left. - (\beta - \beta_S) E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1}) - (\beta - \beta_D) E_t(f_{t+1}) \mathbb{I}(\mathcal{D}_{t+1}) \right]^2. \end{aligned}$$

We now note that, as above, $E[\epsilon_{t+1} E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1})] = 0$; further, consider

$$\begin{aligned} & E \left[(f_{t+1} - E_t(f_{t+1})) E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1}) \right] \\ &= E \left\{ E_t \left[(f_{t+1} - E_t(f_{t+1})) E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1}) \right] \right\} \\ &= E \left\{ E_t \left[(f_{t+1} - E_t(f_{t+1})) E_t(f_{t+1}) \right] \mathbb{I}(\mathcal{S}_{t+1}) \right\}, \end{aligned}$$

because $\mathbb{I}(\mathcal{S}_{t+1})$ is measurable with respect to the information set available until t by construction.

Also, it is easy to see that

$$\begin{aligned} & E_t [(f_{t+1} - E_t(f_{t+1})) E_t(f_{t+1})] \\ = & E_t [f_{t+1} E_t(f_{t+1})] - E_t [E_t(f_{t+1}) E_t(f_{t+1})] = 0, \end{aligned}$$

a.s., seeing as

$$E_t [E_t(f_{t+1}) E_t(f_{t+1})] = [E_t(f_{t+1})]^2,$$

and, by measurability

$$E_t [f_{t+1} E_t(f_{t+1})] = E_t(f_{t+1}) E_t(f_{t+1}).$$

This immediately entails that

$$\begin{aligned} & E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(\mathcal{U})} \right)^2 \\ = & E \left(R_{t+1} - \widehat{R}_{t+1|t}^{(\mathcal{C})} \right)^2 + (\beta - \beta_{\mathcal{S}})^2 E [E_t(f_{t+1}) \mathbb{I}(\mathcal{S}_{t+1})]^2 \\ & + (\beta - \beta_{\mathcal{D}})^2 E [E_t(f_{t+1}) \mathbb{I}(\mathcal{D}_{t+1})]^2. \end{aligned}$$

Note that equality holds only when $\beta_{\mathcal{S}} = \beta_{\mathcal{D}} = \beta$. □

A.2 Assumptions and further notation

Let

$$\underline{\mathbf{u}}_t^0 = \mathbf{u}_{\mathcal{D},t}^0 \cup \mathbf{u}_{\mathcal{S},t}^0, \tag{A.8}$$

where $\underline{\mathbf{u}}_t^0$ has dimension P_u^0 such that $\max(P_{\mathcal{D}}^0, P_{\mathcal{S}}^0) \leq P_u^0 \leq P_{\mathcal{D}}^0 + P_{\mathcal{S}}^0$, and let $\mathbf{B}_{u,j}^0 = (\beta_{u,j,1}^0, \dots, \beta_{u,j,N}^0)'$ 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

$$\delta_i^0 = \beta_{u,\mathcal{S},i}^0 - \beta_{u,\mathcal{D},i}^0. \tag{A.9}$$

Assumption 1. *There exists an $\eta \in \left(\frac{1}{2}, 1\right]$ such that $\delta_i^0 \neq 0$ for $i = 1, \dots, \lfloor N^\eta \rfloor$ 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 $O(N^{1/2})$ 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, $O_p(C_{N,T}^{-1})$ (see Bai and Ng, 2002). Thus, Assumption 1 requires that at least as many as $O(N^\eta)$ 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 $T_j^{-1} \sum_{t=1}^T E(\epsilon_t \epsilon_t' d_{j,t}(\theta^0))$ is finite for $j = \mathcal{D}, \mathcal{S}$, and (b) the smallest eigenvalue of $T_j^{-1} \sum_{t=1}^T E(\epsilon_t \epsilon_t' d_{j,t}(\theta^0))$ is positive for $j = \mathcal{D}, \mathcal{S}$; (ii)*

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

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

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

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

where $\Sigma_{\mathbf{f},j}(\theta, \theta^0)$ is positive definite for $j = \mathcal{D}, \mathcal{S}$ and all θ ; (iii) $\|\beta_{j,i}^0\| < \infty$ for $j = \mathcal{D}, \mathcal{S}$ and $1 \leq i \leq N$; (iv) $N^{-1} \mathbf{B}_j' \mathbf{B}_j \rightarrow \Sigma_{\mathbf{B},j}$ as $N \rightarrow \infty$, where $\Sigma_{\mathbf{B},j}$ are positive definite matrices for $j = \mathcal{D}, \mathcal{S}$; (v) the eigenvalues of $\Sigma_{\mathbf{B},j} \Sigma_{\mathbf{f},j}(\theta^0, \theta^0)$ 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) $E(\epsilon_{i,t}) = 0$ and $E|\epsilon_{i,t}|^8 < \infty$; (ii) $E(d_{j,t}(\theta) d_{j,v}(\theta) \epsilon_{i,t} \epsilon_{i,v}) = \tau_{j,i,t,v}(\theta)$ with $|\tau_{j,i,t,v}(\theta)| < |\tau_{j,t,v}|$ for $1 \leq i \leq N$, and $T_j^{-1} \sum_{t=1}^T \sum_{v=1}^T |\tau_{j,t,v}| < \infty$ for $j = \mathcal{D}, \mathcal{S}$; (iii) $E\left(T_j^{-1} \sum_{t=1}^T d_{j,t}(\theta) \epsilon_{i,t} \epsilon_{l,t}\right) = \sigma_{j,i,l,t}(\theta)$ with $|\sigma_{j,i,l,t}(\theta)| \leq |\sigma_{j,i,l}|$ for $1 \leq t \leq T$ and $N^{-1} \sum_{i=1}^N \sum_{l=1}^N |\sigma_{j,i,l}| < \infty$ for $j = \mathcal{D}, \mathcal{S}$; (iv)

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

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

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

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) $\{\mathbf{f}_{j,t}^0, z_t, \epsilon_t\}$ 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) $E(\|\mathbf{f}_{j,t}^0 \epsilon_{i,t}\|^4 | z_t) < \infty$ a.s. and $E(\|\mathbf{f}_{j,t}^0\|^4 | z_t) < \infty$ 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) $E(\mathbf{f}_{j,t}^0 \mathbf{f}_{j,t}^{0'} | z_t = \theta)$ is a.s. positive definite and continuous at $\theta = \theta^0$ for $j = \mathcal{D}, \mathcal{S}$ with $\sum_{i=\lfloor N^\eta \rfloor + 1}^N \delta_i^{0'} E(\mathbf{f}_{j,t}^0 \mathbf{f}_{j,t}^{0'} | z_t = \theta^0) \delta_i^0 < \infty$ a.s.; (vi) $\{\mathbf{f}_{j,t}^0\}$ 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.

We would like to point out that Assumptions 4 and 5 does not necessarily require z_t to be a contemporaneous series, i.e. a series observed at time t . Indeed, z_t can also be defined as the

prediction, based on information available up to a point in time in the past $t - h$, viz.

$$z_t = E_{t-h}(w_t) = E(w_t | \mathcal{F}_{w,-\infty}^{t-h}), \quad (\text{A.10})$$

where $\mathcal{F}_{w,-\infty}^{t-h}$ is the σ -field defined by the past information on w_t which is used for the prediction of w_t , i.e. $\mathcal{F}_{w,-\infty}^{t-h} = \sigma(w_{t-h}, w_{t-h-1}, \dots)$. This also naturally nests, as a special case, the case of z_t being genuinely observable, i.e. $z_t = w_t$ itself, which is typically used in the downside risk literature (see e.g. (Ang et al., 2006; Lettau et al., 2014; and Farago and Tédongap, 2018)). In the case where z_t is a conditional expectation of a variable w_t , it can be shown that z_t satisfies Assumption 5 as long as w_t does - that is, it is possible to spell out our assumption on the observable quantity w_t rather than on the more artificial z_t . Indeed, let $\mathcal{F}_{z,-\infty}^t = \sigma(z_t, z_{t-1}, \dots)$; then of course

$$\mathcal{F}_{z,-\infty}^t \subseteq \mathcal{F}_{w,-\infty}^{t-h} \subseteq \mathcal{F}_{w,-\infty}^t. \quad (\text{A.11})$$

Hence, for any $\mathcal{F}_{w,-\infty}^t$ -measurable function $\varphi(\cdot)$, it holds that

$$\begin{aligned} & E[\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{z,-\infty}^t] \\ &= E\{E[\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{w,-\infty}^{t-h}] | \mathcal{F}_{z,-\infty}^t\} \\ &= E\{E[E(\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{w,-\infty}^t) | \mathcal{F}_{w,-\infty}^{t-h}] | \mathcal{F}_{z,-\infty}^t\}. \end{aligned}$$

Hence, if $E[\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{w,-\infty}^t] = 0$ a.s., as prescribed by Assumption 4(v), it follows that

$$E[E(\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{w,-\infty}^t) | \mathcal{F}_{w,-\infty}^{t-h}] = 0 \quad \text{a.s.},$$

and therefore $E[\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{z,-\infty}^t] = 0$ a.s. also follows. If there exists a finite constant c_0 such that $E[\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{w,-\infty}^{t-h}] = 0 \leq c_0$ a.s., as in Assumption 5, then

$$E[E(\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{w,-\infty}^t) | \mathcal{F}_{w,-\infty}^{t-h}] \leq c_1,$$

a.s., for some finite constant c_1 ; hence, $E \left[\varphi(\mathbf{f}_t, \epsilon_t) | \mathcal{F}_{z, -\infty}^t \right] \leq c_2$ a.s. Finally, letting ρ_m^w be the rho-mixing coefficients of w_t , and ρ_m^z those of z_t , (A.11) immediately yields $\rho_m^z \leq \rho_m^w$ by definition.

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 $\min(N, T) \rightarrow \infty$, 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, $\hat{g}_j^{(i)} - g_j^{(i)}$ 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.ⁱⁱ

Some further comments on the asymptotic regimes - i.e., on the restrictions which are required between N and T - are in order; hereafter, we also provide a summary of the restrictions required by each result. According to Assumption 6, our general asymptotic regime is based on $(N, T) \rightarrow \infty$ jointly (we refer to Phillips and Moon, 1999, for a definition of joint versus sequential limits). The

ⁱⁱIn 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 $\pi_j^{-1} = O(T^{1/2-\epsilon})$ 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 $N = O(T^\kappa)$ for any value of κ .

restriction on the relative rate of divergence as N and T required by the assumption is, essentially, equivalent to

$$N = O\left(\exp\left(\sqrt{T}\right)\right). \quad (\text{A.12})$$

This restriction is satisfied in the vast majority of datasets; based on (A.12), N can be smaller than T , but also much bigger than T , even greater than any polynomial rate in T . Indeed, this (very mild) restriction is a consequence of our method of proof (see in particular Lemma C.1) to derive almost sure rates, and we conjecture that it could be further relaxed.

As far as the restrictions needed by each set of results, in general all our results involving rates and consistency do not require any further restrictions between N and T as they pass to infinity. For example, the rate of convergence for the estimated threshold, $\hat{\theta}$ (see Lemma C.7), only requires Assumption 6; hence, our test in Section A.3 for $H_0 : \theta^0 = c$ also only requires Assumption 6, and if the test were inverted, to construct confidence intervals, these also would not require any further restrictions between N and T . Similarly, the consistent estimation of the number of common factors (Section 4.2); the (consistent up to a linear transformation) estimation of factors and loadings (Section 4.3); and the consistent estimation of common components (Section 4.3) and risk premia (Section 5.3), do not require any further restrictions between N and T . Conversely, any result involving limiting distributions does require further restrictions between N and T , in addition to Assumption 6, although this is typical in general in the context of large factor models (see, as a prime example, Bai, 2003). For example, a restriction like

$$\frac{T^{1/2}}{N} \rightarrow 0, \quad (\text{A.13})$$

is required to have the limiting distribution of the estimated loadings (Theorem A.6); of the estimated risk premia, $\hat{\gamma}_{j,\mathbf{g}}$ (Theorem 3). This restriction, per se, does not require T to be smaller than N , and it is satisfied when N and T are of comparable dimension, or even if T is “a bit larger” than N . Similarly, the asymptotic distribution for the zero-beta rates (Theorem A.7) requires the restriction

$$\frac{N^{1/2}}{T} \rightarrow 0,$$

which is in a dual relationship with (A.13). Even in this case, it is not required that N be smaller than T , and it is satisfied when N and T are of comparable dimension, or even if N is “a bit larger” than T .

Assumption 7. ⁱⁱⁱ It holds that, for $j = \mathcal{D}, \mathcal{S}$: (i) $E \left(\sum_{i=1}^N w_i \alpha_{j,i} \right)^2 \leq c_0 N$ 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 $N \rightarrow \infty$, $\left[\text{Var} \left(\sum_{i=1}^N v_i \alpha_{j,i} \right) \right]^{-1/2} \sum_{i=1}^N v_i \alpha_i \xrightarrow{D} N(0, I_\nu)$ for every ν -dimensional sequence v_i , $1 \leq i \leq N$, such that $\left\| N^{-1} \sum_{i=1}^N v_i v'_i \right\| < \infty$.

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) $T^{-1} \sum_{t=1}^T \mathbf{e}_t \mathbf{u}_{j,t}^{0'} = O_{\mathbb{P}}(1)$ for $j = \mathcal{D}, \mathcal{S}$.

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

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

where

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

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

ⁱⁱⁱ 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_i s 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_i s to have nonzero mean. Indeed, if $E(\alpha_i) \neq 0$ for $1 \leq i \leq N_\alpha$ with $N_\alpha = O(N^{1/2})$, and $E \left(\sum_{i=N_\alpha+1}^N w_i \alpha_i \right)^2 \leq c_0 N$, the assumption would hold anyway. Note also that part (iii) of the assumption could be derived using more primitive assumptions on the α_i s (see e.g. the very general CLT in McLeish, 1974).

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 $\alpha_{\mathcal{S}}$ are independent; (ii) as $T \rightarrow \infty$*

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

with

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

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

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

A.3 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.4, and we therefore present them only once, to avoid repetitions, in the next section.

We know that, by Lemma C.7

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

where $\bar{v}_{N,T}(\epsilon)$ is defined in (C.3). Define now the deterministic sequence $s_{N,T}$ such that

$$\begin{aligned} \lim_{\min(N,T) \rightarrow \infty} s_{N,T} &= \infty, \\ \lim_{\min(N,T) \rightarrow \infty} s_{N,T} T^{-1} \bar{v}_{N,T}(\epsilon) &= 0, \end{aligned}$$

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

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

and therefore we can assume that, under the null

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

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

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

and therefore we can assume that, under H_A

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

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

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

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

$$\begin{aligned}\lim_{\min(N,T) \rightarrow \infty} h(s_{N,T}(\hat{\theta} - c)) &= \infty, \\ \lim_{\min(N,T) \rightarrow \infty} h(s_{N,T}(\hat{\theta} - c)) &= 0,\end{aligned}$$

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 $\{\xi_m^\theta, 1 \leq m \leq M_\theta\}$, with common distribution $N(0, 1)$;

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

Step 3 Define

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

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.2. *We assume that the assumptions of Lemma C.7 are satisfied. Then, as $\min(N, M_\theta, T) \rightarrow \infty$ with*

$$\frac{M_\theta}{h^{1/2}(s_{N,T} T^{-1} \bar{v}_{N,T}(\epsilon))} \rightarrow 0, \tag{A.14}$$

it holds that $S^\theta \xrightarrow{D^*} \chi_1^2$, under H_0 , and that $R^{-1} S^\theta \xrightarrow{\mathbb{P}^*} c_0 > 0$ 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.2 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 $\mathbb{P}(\chi_1^2 \geq c_\alpha) = \alpha$ for a nominal level α , Theorem A.2 entails

$$\lim_{\min(N,T,M_\theta) \rightarrow \infty} \mathbb{P}^*(S^\theta \geq c_\alpha) = 1, \tag{A.15}$$

under H_A , and

$$\lim_{\min(N,T,M_\theta) \rightarrow \infty} \mathbb{P}^*(S^\theta \geq c_\alpha) = \alpha, \tag{A.16}$$

under H_0 . Equation (A.15) has the “traditional” interpretation: when the null is false, the test - asymptotically - rejects with probability 1. Conversely, equation (A.16) does not have a straightforward meaning. The test is constructed using a randomization based on $\{\xi_m^\theta\}$ 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 $\{\xi_{m,b}^\theta\}$, $1 \leq m \leq M_\theta$, with common distribution $N(0, 1)$, independent across b . Define

$$Q_\alpha = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(S_b^\theta \leq c_\alpha). \quad (\text{A.17})$$

The Law of the Iterated Logarithm entails that^{iv}

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

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

$$Q_\alpha \geq (1 - \alpha) - \sqrt{\alpha(1 - \alpha)} \sqrt{\frac{2 \ln \ln B}{B}}. \quad (\text{A.18})$$

Finally, we point out that, in our empirical exercises, we have used $f_{NT} = \left(\sqrt{T} |\hat{\theta} - c|\right)^{-1}$ 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 dataset, and define the sample sizes as (N_1, T_1) and (N_2, T_2) , and the relevant estimates as $\hat{\theta}_1$ and $\hat{\theta}_2$. Then, upon

^{iv}Details on this argument are in Horváth and Trapani (2019).

defining a sequence $s_{N,T}^* = s_{N,T}^*(T_1, T_2)$ such that

$$\lim_{\min(N_1, T_1, N_2, T_2) \rightarrow \infty} s_{N,T}^* = \infty,$$

$$\lim_{\min(N_1, T_1, N_2, T_2) \rightarrow \infty} \frac{s_{N,T}^*}{\min\{T_1, T_2\}} \max\{\bar{v}_{N_1, T_1}(\epsilon), \bar{v}_{N_2, T_2}(\epsilon)\} = 0,$$

the test described above can be applied readily using

$$f_{N,T}^* = \frac{1}{\sqrt{\min\{T_1, T_2\} |\hat{\theta}_1 - \hat{\theta}_2|}},$$

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

A.4 Complements to Section 4.2: theory

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

$$\Sigma_{\mathcal{D}} = \frac{1}{T\pi_{\mathcal{S}}} \sum_{t=1}^T E\left(\widetilde{\mathbf{R}}_t \widetilde{\mathbf{R}}_t' d_{\mathcal{D},t}(\theta^0)\right), \quad (\text{A.19})$$

$$\Sigma_{\mathcal{S}} = \frac{1}{T\pi_{\mathcal{S}}} \sum_{t=1}^T E\left(\widetilde{\mathbf{R}}_t \widetilde{\mathbf{R}}_t' d_{\mathcal{S},t}(\theta^0)\right). \quad (\text{A.20})$$

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

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

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

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

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

Theorem A.3. *We assume that Assumptions 1-5 are satisfied. Then it holds that, for all $1 \leq i \leq N$ and $j = \mathcal{D}, \mathcal{S}$*

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

for every $\epsilon > 0$.

Lemma A.1 and Theorem A.3 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 $N \rightarrow \infty$, 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.3 offers the sample counterpart to Lemma A.1, giving a bound for the estimation error $\hat{g}_j^{(i)} - g_j^{(i)}$.^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.3 only contains rates, we propose a randomized version of the estimated eigenvalues $\hat{g}_j^{(i)}$, 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 $\hat{g}_j^{(i)}$. Indeed, whenever $g_j^{(i)}$ diverges as fast as N (which, in (17), represents the null hypothesis), $\hat{g}_j^{(i)}$ also does, at the same rate: this is an immediate consequence of Theorem A.3. Conversely, there is no guarantee that $\hat{g}_j^{(i)}$ converges when $g_j^{(i)}$ does so: in this

^vAs is typical in (large) Random Matrix Theory, there is no guarantee that $\hat{g}_j^{(i)} - g_j^{(i)}$ will drift to zero as $\min(N, T) \rightarrow \infty$. 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, $\hat{g}_j^{(i)} - g_j^{(i)}$ may even diverge to infinity. However, as long as $\hat{g}_j^{(i)} - g_j^{(i)}$ 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 $\hat{g}_j^{(i)}$ is of smaller magnitude when $j > P_j^0$ compared to the case $j \leq P_j^0$.

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

case (which corresponds to the alternative hypothesis in (17)), from Theorem A.3, it is still possible that the estimation error $\hat{g}_j^{(i)} - g_j^{(i)}$ will diverge at a rate $NT^{-1/2}$, modulo some logarithmic terms.

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

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

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

$$\lim_{\min(N,T) \rightarrow \infty} \frac{N^{1-\varrho_j} T^{1/2}}{T \pi_j} (\ln N)^{1+\epsilon} (\ln T)^{1/2+\epsilon} = 0; \quad (\text{A.21})$$

(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.21) can be employed to construct a refined proposal for ϱ_j . Based on (19) and on the strong rates in Theorem A.3, it is easy to see that

$$\begin{aligned} \lim_{\min(N,T) \rightarrow \infty} \psi(\hat{g}_j^{(i)}) &= \infty \text{ under } H_0, \\ \lim_{\min(N,T) \rightarrow \infty} \psi(\hat{g}_j^{(i)}) &= 1 \text{ under } H_A. \end{aligned}$$

Let $\varkappa$ be defined such that

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

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

Theorem A.4. *We assume that Assumptions 1-6 are satisfied with $\eta = 1$ in Assumption 1. Let $M = M(N)$, so that $\lim_{N \rightarrow \infty} M(N) = \infty$. Then, as $\min(N, T_j) \rightarrow \infty$ with*

$$M \exp(-\varkappa N^{1-\varrho_j}) \rightarrow 0, \quad (\text{A.22})$$

it holds that under H_0

$$\Upsilon_j^{(i)} \xrightarrow{D^*} \chi_1^2, \quad (\text{A.23})$$

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

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

$$\frac{1}{4M} \Upsilon_j^{(i)} \xrightarrow{\mathbb{P}^*} c_0, \quad (\text{A.24})$$

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

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.24); conversely, in the proof of (A.23) it is shown that the test statistic has a non-centrality parameter which is controlled by (A.22), 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.22) 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 .^{vii} 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.

$$\hat{\Upsilon}_j^{(i)} = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n_S} w_i \left(v_j^{(i)} \left(\sqrt{2} x_i \right) \right)^2, \quad (\text{A.25})$$

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

$$w_i = \frac{\sqrt{\pi} 2^{n_S-1} (n_S - 1)!}{n_S \left[H_{n_S-1} \left(\sqrt{2} x_i \right) \right]^2}. \quad (\text{A.26})$$

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

^{vii}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.

^{viii}The values of the roots x_i , and of the corresponding weights w_i , are tabulated e.g. in Salzer, Zucker, and

A.5 Complements to Section 5.2

A.5.1 Assumptions

For $j = \mathcal{D}, \mathcal{S}$, consider the matrices $M_{\mathbf{u}}^{(j)} = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0'} d_{j,t}(\theta^0)$. We then define

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

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

A.5.2 Estimation of long-run covariance matrices

We propose the following estimators

$$\begin{aligned}\widehat{\Sigma}_{\gamma,j} &= c_j^{-2} \widehat{\mathbf{\Lambda}}_j \widehat{V}_{\mathbf{u},j} \widehat{\mathbf{\Lambda}}_j' + \left(\left(\widehat{\gamma}_{j,1}' \left(\widehat{M}_{\mathbf{u}}^{(j)} \right)^{-1} \right) \otimes I_K \right) \widehat{V}_{\mathbf{eu},j} \left(\left(\left(\widehat{M}_{\mathbf{u}}^{(j)} \right)^{-1} \widehat{\gamma}_{j,1} \right) \otimes I_K \right) \\ &\quad + c_j^{-1} \widehat{\mathbf{\Lambda}}_j \widehat{\Pi}_{\mathbf{ue},j} \left(\left(\left(\widehat{M}_{\mathbf{u}}^{(j)} \right)^{-1} \widehat{\gamma}_{j,1} \right) \otimes I_K \right) + c_j^{-1} \left(\left(\widehat{\gamma}_{j,1}' \left(\widehat{M}_{\mathbf{u}}^{(j)} \right)^{-1} \right) \otimes I_K \right) \widehat{\Pi}_{\mathbf{ue},j}' \widehat{\mathbf{\Lambda}}_j',\end{aligned}\tag{A.27}$$

$$\widehat{\Sigma}_{\alpha\gamma,j} = \widehat{\sigma}_{\alpha,j}^2 \widehat{\mathbf{\Lambda}}_j \left[\frac{1}{N} \widehat{\mathbf{B}}_j^{\widehat{P}_j'} \widehat{\mathbf{B}}_j^{\widehat{P}_j} - \left(\frac{1}{N} \widehat{\mathbf{B}}_j^{\widehat{P}_j'} i_N \right) \left(\frac{1}{N} \widehat{\mathbf{B}}_j^{\widehat{P}_j'} i_N \right)' \right]^{-1} \widehat{\mathbf{\Lambda}}_j',\tag{A.28}$$

In (A.27), recall that $\widehat{\mathbf{X}}_j = \left(i_N, \widehat{\mathbf{B}}_j^{\widehat{P}_j} \right)$. Also, we have used

$$\widehat{M}_{\mathbf{u}}^{(j)} = \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{u}}_{j,t} \widehat{\mathbf{u}}_{j,t}' \widehat{d}_{j,t},\tag{A.29}$$

$$\widehat{V}_{\mathbf{u},j} = \widehat{m}_{\mathbf{u},0}^{(j)} + \sum_{k=1}^{h_m} \left(1 - \frac{k}{h_m + 1} \right) \left(\widehat{m}_{\mathbf{u},k}^{(j)} + \widehat{m}_{\mathbf{u},k}^{(j)'} \right),\tag{A.30}$$

$$\widehat{V}_{\mathbf{eu},j} = \widehat{m}_{\mathbf{eu},0}^{(j)} + \sum_{k=1}^{h_m} \left(1 - \frac{k}{h_m + 1} \right) \left(\widehat{m}_{\mathbf{eu},k}^{(j)} + \widehat{m}_{\mathbf{eu},k}^{(j)'} \right),\tag{A.31}$$

$$\widehat{\Pi}_{\mathbf{ue},j}' = \frac{1}{T} \sum_{t=1}^T \left(\text{vec} \left(\widehat{\mathbf{e}}_t \widehat{\mathbf{u}}_{j,t}' \right) \right) \widehat{\mathbf{u}}_{j,t}' \widehat{d}_{j,t} + \sum_{k=1}^{h_m} \left(1 - \frac{k}{h_m + 1} \right) \left(\widehat{m}_{\mathbf{\Pi},k}^{(j)} + \widehat{m}_{\mathbf{\Pi},k}^{(j)'} \right),\tag{A.32}$$

Capuano (1952).

with

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

for $k = 0, 1, \dots$. In (A.30)-(A.31), we use $h_m = O(T^{1/3})$, and $\widehat{\mathbf{e}}_t = \mathbf{g}_t - (\widehat{\mathbf{a}} + \widehat{\mathbf{\Lambda}} \widehat{\mathbf{u}}_t)$. Finally, in order to compute $\widehat{\sigma}_{\alpha,j}^2$, let

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

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 $\min(N, T_j) \rightarrow \infty$ for $j = \mathcal{D}, \mathcal{S}$ under (24) and (25), it holds that $\widehat{\sigma}_{\alpha,j}^2 = \sigma_{\alpha,j}^2 + o_{\mathbb{P}^*}(1)$, for almost all realizations of $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$.*

A.5.3 Estimation of $\mathbf{\Gamma}_j$ and $\mathbf{\Lambda}_j$: asymptotics

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

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

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

$$\widehat{\mathbf{H}}_{jj}^{\hat{P}_j}(\theta) = \frac{\mathbf{U}_j(\theta^0) \mathbf{U}_j(\theta)' \mathbf{B}_j^{0'} \widehat{\mathbf{B}}_j(\theta)}{T} \widehat{\mathbf{V}}_j(\theta)^{-1}, \quad (\text{A.33})$$

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

$$\widehat{\mathbf{H}}_{jj}^{\Gamma \hat{P}_j}(\theta) = \begin{bmatrix} 1 & \mathbf{0}'_{\hat{P}_j} \\ \mathbf{0}_{\hat{P}_j} & \widehat{\mathbf{H}}_{jj}^{\hat{P}_j}(\theta) \end{bmatrix}, \quad (\text{A.34})$$

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

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

Theorem A.5. *We assume that Assumptions 1-8 are satisfied. Then, as $\min(N, T_j) \rightarrow \infty$, it holds that*

$$\widehat{\Gamma}_j - \left(\widehat{\mathbf{H}}_{jj}^{\Gamma}(\theta^0) \right)^{-1} \Gamma_j^0 = o_{\mathbb{P}}(1) + o_{\mathbb{P}^*}(1), \quad (\text{A.35})$$

$$\widehat{\Lambda}_j - \Lambda_j^0 \widehat{\mathbf{H}}_{jj}(\theta^0) = o_{\mathbb{P}}(1) + o_{\mathbb{P}^*}(1), \quad (\text{A.36})$$

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

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

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

Theorem A.6. *We assume that Assumptions 1-8 are satisfied, and that $\min(N, T_j) \rightarrow \infty$, with (24) and (25). Then, under Assumptions 7 and 9, it holds that*

$$\left(\frac{1}{T}\Sigma_{\Gamma,j} + \frac{1}{N}\Sigma_{\alpha\Gamma,j}\right)^{-1/2} \left(\widehat{\Gamma}_j - \left(\widehat{\mathbf{H}}_{jj}^{\Gamma}(\theta^0)\right)^{-1} \Gamma_j^0\right) \xrightarrow{D^*} N(0, I_{P_j+1}). \quad (\text{A.37})$$

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

$$\left(\frac{1}{T}\Sigma_{\Lambda,j}\right)^{-1/2} \text{vec}\left(\widehat{\Lambda}_j - \Lambda_j^0 \widehat{\mathbf{H}}_{jj}(\theta^0)\right) \xrightarrow{D^*} N(0, I_{P_j K}), \quad (\text{A.38})$$

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

In order to use the results in Theorem A.6, we propose the following estimators

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

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

A.5.4 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

$$H_0 : \pi_{\mathcal{D}} \gamma_{\mathcal{D},0} + \pi_{\mathcal{S}} \gamma_{\mathcal{S},0} = 0. \quad (\text{A.39})$$

Theorem A.7. *We assume that Assumptions 1-10 hold, and that $E(\alpha_{i,\mathcal{D}} \alpha_{k,\mathcal{S}}) = 0$ for all $1 \leq$*

$i, k \leq N$. Then, as $\min(N, T_j) \rightarrow \infty$ with $\frac{N}{T_j^2} \rightarrow 0$, it holds that

$$\frac{\hat{\gamma}_{j,0} - \gamma_{j,0}}{\left(\sigma_{\alpha,j}^2 \left(i'_N \mathbf{M}_{B,j}^0 i_N\right)^{-1}\right)^{1/2}} \xrightarrow{D^*} N(0, 1), \quad (\text{A.40})$$

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

$$\frac{(\hat{\pi}_{\mathcal{D}} \hat{\gamma}_{\mathcal{D},0} + \hat{\pi}_{\mathcal{S}} \hat{\gamma}_{\mathcal{S},0}) - (\pi_{\mathcal{D}} \gamma_{\mathcal{D},0} + \pi_{\mathcal{S}} \gamma_{\mathcal{S},0})}{\left(\pi_{\mathcal{D}}^2 \sigma_{\alpha,\mathcal{D}}^2 \left(i'_N \mathbf{M}_{B,\mathcal{D}}^0 i_N\right)^{-1} + \pi_{\mathcal{S}}^2 \sigma_{\alpha,\mathcal{S}}^2 \left(i'_N \mathbf{M}_{B,\mathcal{S}}^0 i_N\right)^{-1}\right)^{1/2}} \xrightarrow{D^*} N(0, 1), \quad (\text{A.41})$$

for almost all realizations of $\{R_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$, where $\mathbf{M}_{B,j}^0 = \mathbf{B}_j^0 \left(\mathbf{B}_j^{0'} \mathbf{B}_j^0\right)^{-1} \mathbf{B}_j^{0'}$ for $j = \mathcal{D}, \mathcal{S}$.

A.5.5 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

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

and

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

Theorem A.8. *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 $\min(N, T_j) \rightarrow \infty$ with $\frac{N}{T_j^2} \rightarrow 0$, it holds that*

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

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

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

with

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

and

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

where

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

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

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

In this data generating process (DGP), we generate all variables - except z_t - 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 $\theta^0 = -6$, which is the value we found in our empirical application; by way of robustness check, we also consider the case $\theta^0 = -3$. Similarly, we generate z_t as *i.i.d.* $N(\mu_z, \sigma_z^2)$. Based on our data, we set $\mu_z = -1.134$, and $\sigma_z = 7.346$. In Section 6 we found that $\mathbb{P}(z_t \leq \theta^0) \simeq 0.23$; with our set-up, we have $\mathbb{P}(z_t \leq \theta^0) \simeq 0.254$, which ensures an almost like for like set of synthetic data. 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 500.

The specifications of our estimators and tests are as follows. In the estimation of θ , we use a grid search on the interval $[-8, -4]$, with steps of size 0.01. 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 $\alpha = 0.05/T$.ⁱⁱⁱ

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

ⁱⁱWe 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.* $N(0, 0.5)$.

ⁱⁱⁱWe 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.4, 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

Results are reported for $(N, T) = (130, 104)$ and $(130, 224)$, which correspond to the sample sizes in our empirical application. 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: $P_{\mathcal{D}}^0 \in \{1, 2, 3\}$ and $P_{\mathcal{S}}^0 \in \{1, 2, 3, 4, 5, 6\}$.^{iv}

Consider the estimation of θ , whose performance, in terms of bias and RMSE, we report in Table B.1; the results are fully in line with the theory, which predicts that the RMSE (in particular) declines as T increases; results are better, as can be expected, when the threshold θ^0 is closer to the mean μ_z .

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

$$\overline{\hat{P}}_j = \frac{1}{500} \sum_{r=1}^{500} \hat{P}_{j,r},$$

$j = \mathcal{D}, \mathcal{S}$, and (ii) the percentage of times that $\hat{P}_{j,r} \neq P_j^0$, $j = \mathcal{D}, \mathcal{S}$. Results are very satisfactory with large samples: when $(N, T) = (130, 224)$, the estimated numbers of factors is always close to the true number of factors in all cases considered, with the possible exception of the case $(P_{\mathcal{D}}^0, P_{\mathcal{S}}^0) = (1, 1)$. In this case, however, our estimator has the tendency to *overstate* $P_{\mathcal{D}}^0$, which may be regarded as a better outcome compared to understating it (in essence, due to the same reasons why including an irrelevant variable in a regression model is preferable to omitting a relevant one). When $(N, T) = (130, 104)$, the relatively low value of T worsens results, as may be expected; this is also compounded by the fact that, in the downside regime, the number of observations is rather small. Still, when the estimated number of common factors is wrong, this is due to an overstatement rather than to an understatement. Hence, empirical findings whereby $P_{\mathcal{D}}^0$ tends to be smaller than $P_{\mathcal{S}}^0$ are not a consequence of our procedure understating the true number of common factors due to the sample size T being small; indeed, if anything, our procedure is overly conservative, and we refer to a recent paper by Fortin et al. (2023) for a discussion of

change.

^{iv}Therefore we do not consider the case where there is no factor at all (i.e., either $P_{\mathcal{D}}^0 = 0$ or $P_{\mathcal{S}}^0 = 0$, 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).

possible problems arising in the presence of short panels when estimating the number of common factors.

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

Panel A: $\theta^0 = -3$															
$(N, T) = (130, 104)$				$(N, T) = (130, 224)$				$(N, T) = (130, 104)$				$(N, T) = (130, 224)$			
Bias								RMSE							
$P_S \backslash P_D$	1	2	3	$P_S \backslash P_D$	1	2	3	$P_S \backslash P_D$	1	2	3	$P_S \backslash P_D$	1	2	3
1	-0.001	-0.001	-0.002	1	-0.001	-0.001	-0.001	1	0.004	0.003	0.003	1	0.002	0.001	0.001
2	-0.002	-0.001	-0.002	2	-0.001	-0.001	-0.001	2	0.003	0.003	0.003	2	0.001	0.001	0.001
3	-0.002	-0.002	-0.002	3	-0.001	-0.001	-0.001	3	0.003	0.003	0.002	3	0.001	0.001	0.001
4	-0.002	-0.002	-0.002	4	-0.001	-0.001	-0.001	4	0.003	0.003	0.003	4	0.001	0.001	0.001
5	-0.002	-0.002	-0.002	5	-0.001	-0.001	-0.001	5	0.002	0.002	0.003	5	0.001	0.001	0.001
6	-0.002	-0.002	-0.002	6	-0.001	-0.001	-0.001	6	0.002	0.003	0.003	6	0.001	0.001	0.001

Panel B: $\theta^0 = -6$															
$(N, T) = (130, 104)$				$(N, T) = (130, 224)$				$(N, T) = (130, 104)$				$(N, T) = (130, 224)$			
Bias								RMSE							
$P_S \backslash P_D$	1	2	3	$P_S \backslash P_D$	1	2	3	$P_S \backslash P_D$	1	2	3	$P_S \backslash P_D$	1	2	3
1	0.006	0.003	0.006	1	0.004	0.005	-0.001	1	0.011	0.008	0.007	1	0.009	0.005	0.002
2	0.001	0.000	0.003	2	-0.001	-0.001	-0.001	2	0.007	0.006	0.003	2	0.002	0.002	0.001
3	0.000	-0.001	-0.002	3	-0.001	-0.001	-0.001	3	0.005	0.003	0.003	3	0.002	0.001	0.001
4	0.001	-0.002	-0.002	4	-0.001	-0.001	-0.001	4	0.003	0.003	0.003	4	0.001	0.001	0.001
5	0.002	-0.002	-0.002	5	-0.001	-0.001	-0.001	5	0.003	0.003	0.003	5	0.001	0.001	0.001
6	0.002	-0.002	-0.002	6	-0.001	-0.001	-0.001	6	0.003	0.003	0.003	6	0.001	0.001	0.001

Notes. The table contains bias and RMSE for $\hat{\theta}$, as defined in the main text. The estimation of θ is based on a grid search with steps of size 0.01, over the interval $[-2 + \theta^0, 2 + \theta^0]$ - hence, when $\theta = -3$, the estimator is computed over the grid $[-5, -1]$; when $\theta = -6$, the estimator is computed over the grid $[-8, -4]$.

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

Mean							
$(N, T) = (130, 104)$				$(N, T) = (130, 224)$			
$P_S \setminus P_D$	1	2	3	$P_S \setminus P_D$	1	2	3
1	(1.468, 1.000)	(2.228, 0.994)	(3.032, 0.996)	1	(1.198, 0.996)	(2.052, 0.998)	(2.988, 0.998)
2	(1.264, 1.986)	(2.090, 2.000)	(2.890, 1.970)	2	(1.010, 1.980)	(1.994, 1.984)	(2.986, 1.986)
3	(1.104, 2.992)	(2.014, 2.976)	(2.932, 2.964)	3	(1.000, 2.978)	(1.986, 2.986)	(2.954, 2.978)
4	(1.054, 3.938)	(2.006, 3.956)	(2.922, 3.984)	4	(0.992, 3.980)	(1.992, 3.986)	(2.978, 3.976)
5	(1.036, 4.916)	(2.006, 4.918)	(2.934, 4.924)	5	(0.996, 4.936)	(1.988, 4.956)	(2.950, 4.960)
6	(1.056, 5.896)	(2.048, 5.910)	(2.932, 5.884)	6	(0.994, 5.950)	(1.990, 5.942)	(2.974, 5.922)

Missed true value (%)							
$(N, T) = (130, 104)$				$(N, T) = (130, 224)$			
$P_S \setminus P_D$	1	2	3	$P_S \setminus P_D$	1	2	3
1	(47.60, 0.000)	(26.20, 0.600)	(17.40, 0.400)	1	(20.20, 0.400)	(7.200, 0.200)	(3.000, 0.200)
2	(15.00, 0.800)	(7.800, 0.000)	(5.200, 1.800)	2	(1.400, 1.200)	(0.400, 1.200)	(0.600, 1.000)
3	(6.400, 0.600)	(4.200, 0.800)	(4.000, 1.400)	3	(0.000, 1.200)	(1.000, 0.600)	(1.800, 1.400)
4	(6.000, 2.400)	(1.600, 1.600)	(4.200, 0.800)	4	(0.800, 0.800)	(0.400, 0.600)	(1.200, 0.800)
5	(3.600, 2.800)	(2.600, 2.600)	(3.600, 2.200)	5	(0.400, 2.200)	(0.800, 1.200)	(2.600, 1.000)
6	(4.800, 3.000)	(4.600, 2.400)	(5.800, 2.600)	6	(0.600, 1.400)	(0.800, 1.600)	(1.200, 2.000)

Notes. The table contains the average numbers of estimated factors (top panel) and the percentage of times the estimator is wrong (bottom panel). The specifications of the individual tests are as indicated in the main body of the paper, namely we have set: $M = N$, $\epsilon = 0.01$ in (19), and $n_S = 4$. The nominal level of the individual tests set to $\alpha = 0.05/T$. Altering these parameters does not change the final result in any significant way; robustness checks are available upon request.

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 $U_{i_1, \dots, i_h}$, with $1 \leq i_1 \leq S_1$, $1 \leq i_2 \leq S_2$, etc... Assume that*

$$\sum_{S_1} \cdots \sum_{S_h} \frac{1}{S_1 \cdots S_h} \mathbb{P} \left(\max_{1 \leq i_1 \leq S_1, \dots, 1 \leq i_h \leq S_h} |U_{i_1, \dots, i_h}| > \epsilon L_{S_1, \dots, S_h} \right) < \infty, \quad (\text{C.1})$$

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

$$L_{S_1, \dots, S_h} = S_1^{d_1} \cdots S_h^{d_h} l_1(S_1) \cdots l_h(S_h),$$

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

$$\lim_{(S_1, \dots, S_h) \rightarrow \infty} \sup \frac{|U_{S_1, \dots, S_h}|}{L_{S_1, \dots, S_h}} = 0 \text{ a.s.} \quad (\text{C.2})$$

Henceforth, we will extensively use the following notation:

$$\bar{v}_{N,T}(\epsilon) = (\ln N \ln T)^{2 \frac{3+r+\epsilon}{r}}, \quad (\text{C.3})$$

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

$$S(\mathbf{B}^P, \mathbf{U}^P, \theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\tilde{R}_{i,t} - \left(\beta_{\mathcal{D},i}^{P_{\mathcal{D}'}} \mathbf{u}_{\mathcal{D},t}^{P_{\mathcal{D}}} d_{\mathcal{D},t}(\theta) + \beta_{\mathcal{S},i}^{P_{\mathcal{S}'}} \mathbf{u}_{\mathcal{S},t}^{P_{\mathcal{S}}} d_{\mathcal{S},t}(\theta) \right) \right]^2,$$

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

$$S_{\mathbf{UB}}(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\tilde{R}_{i,t} - \left(\beta_{\mathcal{D},i}^{P_{\mathcal{D}'}}^0 \mathbf{u}_{\mathcal{D},t}^{P_{\mathcal{D}}}^0 d_{\mathcal{D},t}(\theta) + \beta_{\mathcal{S},i}^{P_{\mathcal{S}'}}^0 \mathbf{u}_{\mathcal{S},t}^{P_{\mathcal{S}}}^0 d_{\mathcal{S},t}(\theta) \right) \right]^2, \quad (\text{C.4})$$

which denotes the concentrated version of $S(\mathbf{B}^P, \mathbf{U}^P, \theta)$ 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

$$S_{\mathbf{uB}}^{(P)}(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\tilde{R}_{i,t} - \left(\hat{\beta}_{\mathcal{D},i}^{P_{\mathcal{D}}'} \hat{\mathbf{u}}_{\mathcal{D},t}^{P_{\mathcal{D}}} d_{\mathcal{D},t}(\theta) + \hat{\beta}_{\mathcal{S},i}^{P_{\mathcal{S}}'} \hat{\mathbf{u}}_{\mathcal{S},t}^{P_{\mathcal{S}}} d_{\mathcal{S},t}(\theta) \right) \right]^2, \quad (\text{C.5})$$

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

$$\widehat{H}_{jm}^{P_{j,m}}(\theta) = \frac{\mathbf{U}_{m,m}^0(\theta^0) \mathbf{U}_{m,j}^0(\theta)' \mathbf{B}_m^{0'} \widehat{\mathbf{B}}_j^{P_j}(\theta)}{T} \left(\widehat{V}_j^{P_j}(\theta) \right)^{-1}, \quad (\text{C.6})$$

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

$$\frac{1}{NT} \sum_{t=1}^T \widetilde{\mathbf{R}}_t \widetilde{\mathbf{R}}_t' d_{j,t}(\theta).$$

Note that, for $j \neq m$, $\widehat{H}_{jm}^{P_{j,m}}(\theta^0)$ 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*

$$\frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{j,i}^{P_j}(\theta^0) - \widehat{H}_{jj}^{P_j}(\theta^0)' \beta_{j,i}^0 \right\|^2 = o_{a.s.} \left(P_j C_{N,T}^{-2} \bar{v}_{N,T}(\epsilon) \right),$$

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

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{j,i}^{P_j}(\theta^0) - \widehat{H}_{jj}^{P_j}(\theta^0)' \beta_{j,i}^0 \right\|^2 \\
& \leq \left[\frac{4}{N^2} \sum_{i=1}^N \sum_{l=1}^N \sigma_{jil}^2(\theta^0) + \frac{1}{N} \left(\frac{1}{N^2} \sum_{l=1}^N \sum_{q=1}^N \left| \sum_{i=1}^N \kappa_{jil}(\theta^0) \kappa_{jiq}(\theta^0) \right|^2 \right)^{1/2} \right. \\
& \quad \left. + \frac{2}{N} \sum_{i=1}^N \left(\left\| \beta_{D,i}^0 \right\|^2 + \left\| \beta_{S,i}^0 \right\|^2 \right) \frac{1}{NT^2} \sum_{l=1}^N \left\| \sum_{t=1}^T d_{j,t}(\theta^0) \mathbf{u}_{j,t}^0 e_{l,t} \right\|^2 \right] \\
& \quad \times \left(\frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{j,i}^{P_j}(\theta^0) \right\|^2 \right) \\
& = A_{N,T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{j,i}^{P_j}(\theta^0) \right\|^2 \right),
\end{aligned} \tag{C.7}$$

having defined

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

By the proof of Theorem 3.2 in Massacci (2017), it follows immediately that $E|A_{N,T}| \leq c_0 C_{N,T}^{-2}$. Thus, using the maximal inequality for multi-parameter processes in Corollary 4 of it follows that

$$E \max_{1 \leq i \leq N, 1 \leq t \leq T} |A_{i,t}| \leq c_0 C_{N,T}^{-2} \ln N \ln T.$$

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

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

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

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

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

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

Proof. The desired result follows upon showing

$$S_{\mathbf{UB}}^{(P)}(\theta) - S_{\mathbf{UB}}^{(P_0)}(\theta^0) = c_0 \left| \theta - \theta^0 \right| + o_{a.s.}(1), \quad (\text{C.8})$$

with $c_0 > 0$, for every $\theta \neq \theta^0$ - note that $S_{\mathbf{UB}}^{(P)}(\theta)$ is defined in (C.5). We begin by considering the identity

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

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

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

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

$$\begin{aligned} E \left| S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^0, \mathbf{B}_{\mathcal{S}}^0, \theta^0\right) - S_{\mathbf{UB}}^{(P)}(\theta^0) \right|^2 &\leq c_0 C_{N,T}^{-2}, \\ E \left| S_{\mathbf{UB}}^{(P)}(\theta) - S_{\mathbf{B}}^{(P)}\left(\mathbf{B}_{\mathcal{D}}^0 \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta) + \mathbf{B}_{\mathcal{S}}^0 \widehat{H}_{\mathcal{SD}}^{P_{\mathcal{S}, \mathcal{D}}}(\theta), \mathbf{B}_{\mathcal{D}}^0 \widehat{H}_{\mathcal{DS}}^{P_{\mathcal{D}, \mathcal{S}}}(\theta) + \mathbf{B}_{\mathcal{S}}^0 \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta), \theta^0\right) \right|^2 &\leq c_0 C_{N,T}^{-2}; \end{aligned}$$

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

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

and

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

where

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

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

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

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

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

Putting everything together, we have

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

Then (C.8) follows if we show that

$$S_{\mathbf{UB}}^{(P)} (\theta^0) - S_{\mathbf{UB}}^{(P_0)} (\theta^0) = o_{a.s.} (1). \quad (\text{C.9})$$

It holds that

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

where

$$\begin{aligned} S_{\mathbf{UB}}^1(\theta^0) &= \frac{2}{NT} \sum_{t=1}^T \left(d_{\mathcal{D},t}(\theta^0) \mathbf{u}_{\mathcal{D},t}^{0'} \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}+}(\theta^0)' (\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}(\theta^0) - \mathbf{B}_{\mathcal{D}}^0 \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta^0))' \right. \\ &\quad \left. + d_{\mathcal{S},t}(\theta^0) \mathbf{u}_{\mathcal{S},t}^{0'} \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}+}(\theta^0)' (\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}(\theta^0) - \mathbf{B}_{\mathcal{S}}^0 \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta^0))' \right) \mathbf{e}_t, \end{aligned}$$

$$\begin{aligned} S_{\mathbf{UB}}^2(\theta^0) &= \frac{1}{NT} \sum_{t=1}^T \left(d_{\mathcal{D},t}(\theta^0) \mathbf{u}_{\mathcal{D},t}^{0'} \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}+}(\theta^0)' (\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}(\theta^0) - \mathbf{B}_{\mathcal{D}}^0 \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta^0))' \right. \\ &\quad \times (\widehat{\mathbf{B}}_{\mathcal{D}}^{P_{\mathcal{D}}}(\theta^0) - \mathbf{B}_{\mathcal{D}}^0 \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta^0)) \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}+}(\theta^0) \mathbf{u}_{\mathcal{D},t}^0 \\ &\quad + d_{\mathcal{S},t}(\theta^0) \mathbf{u}_{\mathcal{S},t}^{0'} \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}+}(\theta^0)' (\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}(\theta^0) - \mathbf{B}_{\mathcal{S}}^0 \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta^0))' \\ &\quad \left. + (\widehat{\mathbf{B}}_{\mathcal{S}}^{P_{\mathcal{S}}}(\theta^0) - \mathbf{B}_{\mathcal{S}}^0 \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta^0)) \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}+}(\theta^0) \mathbf{u}_{\mathcal{S},t}^0 \right). \end{aligned}$$

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

$$\begin{aligned} |S_{\mathbf{UB}}^1(\theta^0)| &\leq 2T^{-1/2} P_{\mathcal{D}}^0 \|\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}+}(\theta^0)\| \left(\frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{\mathcal{D},i}^{P_{\mathcal{D}}}(\theta^0) - \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta^0)' \beta_{\mathcal{D},i}^0 \right\|^2 \right)^{1/2} \\ &\quad \left(\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T d_{\mathcal{D},t}(\theta^0) \mathbf{u}_{\mathcal{D},t}^0 e_{i,t} \right\|^2 \right)^{1/2} \\ &\quad + 2T^{-1/2} P_{\mathcal{S}}^0 \|\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}+}(\theta^0)\| \left(\frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{\mathcal{S},i}^{P_{\mathcal{S}}}(\theta^0) - \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta^0)' \beta_{\mathcal{S},i}^0 \right\|^2 \right)^{1/2} \\ &\quad \left(\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T d_{\mathcal{S},t}(\theta^0) \mathbf{u}_{\mathcal{S},t}^0 e_{i,t} \right\|^2 \right)^{1/2}. \end{aligned}$$

By construction, note that

$$\|\widehat{H}_{jj}^{P_j+}(\theta^0)\| \leq c_0 P_j^{1/2} \tag{C.10}$$

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

that

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

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

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

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

$$\begin{aligned} |S_{\mathbf{UB}}^2(\theta^0)| &\leq \frac{1}{T} \sum_{t=1}^T \|\mathbf{u}_{\mathcal{D},t}^0\|^2 \|\widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta^0)\|^2 \frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{\mathcal{D},i}^{P_{\mathcal{D}}}(\theta^0) - \widehat{H}_{\mathcal{DD}}^{P_{\mathcal{D}}}(\theta^0)' \beta_{\mathcal{D},i}^0 \right\|^2 \\ &\quad + \frac{1}{T} \sum_{t=1}^T \|\mathbf{u}_{\mathcal{S},t}^0\|^2 \|\widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta^0)\|^2 \frac{1}{N} \sum_{i=1}^N \left\| \widehat{\beta}_{\mathcal{S},i}^{P_{\mathcal{S}}}(\theta^0) - \widehat{H}_{\mathcal{SS}}^{P_{\mathcal{S}}}(\theta^0)' \beta_{\mathcal{S},i}^0 \right\|^2. \end{aligned}$$

Using (C.10) and Lemma C.2, and noting that Assumptions 3(i) and 5(i) entail

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

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

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

which again is $o_{a.s.}(1)$ by (20). Thus, (C.9) follows. Putting all together, the desired result obtains. $\square$

In the next two lemmas, we define the set

$$B_{N,T} = \left(\frac{\bar{v}_{N,T}(\epsilon)}{T}, c_B \right), \tag{C.11}$$

where $0 < c_0 < \infty$ and $\bar{v}_{N,T}$ is defined in (C.3), and the functions

$$\omega^0(\eta, \theta) = \frac{1}{N\eta T} \sum_{i=1}^N \sum_{t=1}^T \left| d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta^0) \right| \left(\delta_i^{0'} \underline{\mathbf{u}}_t^0 \right)^2, \quad (\text{C.12})$$

$$h^0(\eta, \theta) = \frac{1}{N\eta T} \sum_{i=1}^N \sum_{t=1}^T d_{\mathcal{S},t}(\theta) \delta_i^{0'} \underline{\mathbf{u}}_t^0 e_{i,t}, \quad (\text{C.13})$$

where $\underline{\mathbf{u}}_t^0$ and δ_i^0 are defined in (A.8) and (A.9) 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

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

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

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

so that

$$\inf_{|\theta - \theta^0| \in B_{N,T}} \frac{E\omega^0(\eta, \theta)}{|\theta - \theta^0|} \geq c_0. \quad (\text{C.14})$$

We now show that

$$\sup_{|\theta - \theta^0| \in B_{N,T}} \frac{|\bar{\omega}_{N,T}^0|}{|\theta - \theta^0|} = o_{a.s.}(1). \quad (\text{C.15})$$

Consider the construction $\theta_j = \theta^0 + c^j \bar{v}_{N,T}(\epsilon) / T$, where $1 < c < \infty$, $j = 0, \dots, Q$ and Q is defined such that $c^{Q-1} \bar{v}_{N,T}(\epsilon) / T \leq c_B$ and $c^Q \bar{v}_{N,T}(\epsilon) / T > c_B$ where c_B is defined in (C.11), and let

$$\bar{\omega}_{i,t}^0(\theta_j) = \sum_{i=1}^{\tilde{i}} \sum_{t=1}^{\tilde{t}} \left(\left| d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}(\theta^0) \right| \left(\delta_i^{0'} \underline{\mathbf{u}}_t^0 \right)^2 - E \left(\left| d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}(\theta^0) \right| \left(\delta_i^{0'} \underline{\mathbf{u}}_t^0 \right)^2 \right) \right).$$

Then, standard calculations based on the modulus of continuity yield

$$\begin{aligned}
& \sup_{|\theta - \theta^0| \in B_{N,T}} \frac{\bar{\omega}_{N,T}^0}{|\theta - \theta^0|} \tag{C.16} \\
& \leq 2 \sup_{0 \leq j \leq Q} \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\delta_i^{0'} \mathbf{u}_t^0)^2 (|d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}(\theta^0)| - E |d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}(\theta^0)|)}{|\theta_j - \theta^0|} \\
& \quad + 2 \sup_{0 \leq j \leq Q} \sup_{\theta_j \leq \theta \leq \theta_{j+1}} \frac{\frac{\ln \ln T}{NT^2} \sum_{i=1}^N \sum_{t=1}^T (\delta_i^{0'} \mathbf{u}_t^0)^2 (|d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)| - E |d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)|)}{|\theta_j - \theta^0|} \\
& = 2 \sup_{0 \leq j \leq Q} \frac{\bar{\omega}^0(\theta_j, \theta^0)}{|\theta_j - \theta^0|} + 2 \sup_{0 \leq j \leq Q} \sup_{\theta_j \leq \theta \leq \theta_{j+1}} \frac{\bar{\omega}^0(\theta, \theta_j)}{|\theta_j - \theta^0|} = I + II.
\end{aligned}$$

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

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

Also, consider the scalar case for simplicity and note that

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

By Assumption 5(i), $|d_{\mathcal{S},t}(\theta_j) - d_{\mathcal{S},t}(\theta^0)| (\mathbf{u}_t^0)^2$ 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

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

Also, note that, by Assumption 5(ii)

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

Thus,

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

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

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

Therefore

$$\begin{aligned}
& \sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{1}{NT} P \left(\max_{1 \leq \tilde{i} \leq N, 1 \leq \tilde{t} \leq T} \sup_{0 \leq j \leq Q-1} \frac{\bar{\omega}_{\tilde{i},\tilde{t}}^0(\theta_j)}{|\theta_j - \theta^0|} > \delta_{N,T} \right) \\
& \leq c_0 \left(\sum_{j=0}^{\infty} c^{-j} \right) \sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{\delta_{N,T}^{-2}}{NT} (\bar{v}_{N,T}(\epsilon))^{-1} N^2 T^2 (\ln N) (\ln T) < \infty,
\end{aligned} \tag{C.17}$$

so that finally this entails

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

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 $\delta_j = T^{-1} c^{j-1} (c-1) \bar{v}_{NT}$; letting $m_j = T \delta_j = c^{j-1} (c-1) \bar{v}_{NT}$, we also define

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

$$\begin{aligned} & \sup_{\theta_j \leq \theta \leq \theta_{j+1}} \frac{1}{NT} \sum_{i=1}^{\tilde{i}} \sum_{t=1}^{\tilde{t}} \left(\delta_i^{0'} \underline{\mathbf{u}}_t^0 \right)^2 (|d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)| - E|d_{\mathcal{S},t}(\theta) - d_{\mathcal{S},t}(\theta_j)|) \\ & \leq \max_{2 \leq k \leq m_j+1} \frac{1}{NT} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0'} \underline{\mathbf{u}}_t^0 \right)^2 d_{2t}(\theta_k, \theta_j) \right| + \max_{1 \leq k \leq m_j} \frac{1}{NT} \left| \sum_{i=1}^{n'} \sum_{t=1}^{t'} \left(\delta_i^{0'} \underline{\mathbf{u}}_t^0 \right)^2 d_{2t}(\theta_{k+1}, \theta_k) \right| = a + b, \end{aligned} \quad (\text{C.18})$$

where

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

We begin by considering a

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

for some $r > 0$ such that $E \|\underline{\mathbf{u}}_t^0\|^{4+2r} \leq c_0 T^{2+r}$. Note now that, as in the above, $(\delta_i^{0'} \underline{\mathbf{u}}_t^0)^2 d_{2t}(\theta_k, \theta_j)$ 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

$$\begin{aligned}
& E \left| \sum_{i=1}^N \sum_{t=1}^T \left(\delta_i^{0'} \mathbf{u}_t^0 \right)^2 d_{2t}(\theta_k, \theta_j) \right|^{2+r} \\
& \leq c_0 N^{2+r} T^{1+r/2} \left(E \left| \left\| \mathbf{u}_t^0 \right\|^2 d_{2t}(\theta_k, \theta_j) \right|^2 \right)^{1+r/2} \leq c_1 N^{2+r} T^{1+r/2} \left| E \left\| \mathbf{u}_t^0 \right\|^4 d_{2t}^2(\theta_k, \theta_j) \right|^{1+r/2} \\
& \leq c_0 N^{2+r} T^{1+r/2} |\theta_k - \theta_j|^{1+r/2} \leq c_1 N^{2+r} \bar{v}_{NT}^{1+r/2}(\epsilon) c^{j(1+r/2)}.
\end{aligned}$$

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

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

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

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

Also, using again Theorem 1.1 in Shao (1995)

$$\begin{aligned}
& E \left| \sum_{i=1}^N \sum_{t=1}^T \left(\delta_i^{0'} \mathbf{u}_t^0 \right)^2 d_{2t}(\theta_{k+1}, \theta_k) \right|^2 \\
& \leq c_0 N^2 T E \left\| \mathbf{u}_t^0 \right\|^4 d_{2t}^2(\theta_{k+1}, \theta_k) \leq c_1 N^2 T |\theta_{k+1} - \theta_k| \\
& \leq c_1 N^2 T \frac{\delta_j}{m_j} = c_1 N^2,
\end{aligned}$$

so that Corollary 4 in Moricz (1983) yields

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

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

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*

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

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 $h^0(\eta, \theta) = h^0(\theta)$ (and $h^0(\eta, \theta^0) = h^0(\theta^0)$) for short; we have

$$\begin{aligned}
\sup_{\theta \in V_c(\bar{v}_{NT})} \frac{h^0(\theta)}{|\theta - \theta^0|} & \leq 2 \sup_{0 \leq j \leq Q} \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{i,t} \delta_i^{0'} \mathbf{u}_t^0 [d_{S,t}(\theta_j) - d_{S,t}(\theta^0)]}{|\theta_j - \theta^0|} \\
& \quad + 2 \sup_{0 \leq j \leq Q} \sup_{\theta_j \leq \theta \leq \theta_{j+1}} \frac{\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{i,t} \delta_i^{0'} \mathbf{u}_t^0 [d_{S,t}(\theta) - d_{S,t}(\theta_j)]}{|\theta_j - \theta^0|} \\
& = 2 \sup_{0 \leq j \leq Q} \frac{h^0(\theta_j, \theta^0)}{|\theta_j - \theta^0|} + 2 \sup_{0 \leq j \leq Q} \sup_{\theta_j \leq \theta \leq \theta_{j+1}} \frac{h^0(\theta, \theta_j)}{|\theta_j - \theta^0|} = I + II.
\end{aligned} \tag{C.19}$$

Let

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

It holds that

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

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

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

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

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

having defined

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

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

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

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

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

Thus, by Corollary 4 in Moricz (1983)

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

Similarly

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

with, after some algebra

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

Thus

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

The desired result now follows. $\square$

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

$$\left\| \hat{\beta}_{j,i}^{P_j'} \hat{\mathbf{u}}_{j,t} \hat{d}_{j,t} - \beta_{j,i}^{0'} \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \right\| = o_{a.s.}(1), \quad (\text{C.20})$$

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

Proof. Using the results in the proof of Lemma A.8 in Massacci (2017), it can be shown that $E \left\| \hat{\beta}_{j,i}^{P_j} - \widehat{H}_{jj}^{P_j}(\theta^0) \beta_{j,i}^0 \right\|^2 \leq c_0 C_{N,T}^{-2}$. Thus, using the same logic as above, Lemma C.1 entails that for all $1 \leq i \leq N$, $\left\| \hat{\beta}_{j,i}^{P_j} - \widehat{H}_{jj}^{P_j}(\theta^0) \beta_{j,i}^0 \right\| = o_{a.s.} \left(C_{N,T}^{-1} (\ln N)^{1+\epsilon} (\ln T)^{1+\epsilon} \right)$. With the same logic, by Corollary 3.2 in Massacci (2017), it follows that, for all $1 \leq t \leq T$, $\left\| \hat{\mathbf{u}}_{j,t} \hat{d}_{j,t} - \left(\widehat{H}_{jj}^{P_j}(\theta^0)' \right)^{-1} \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \right\| = o_{a.s.} \left(C_{N,T}^{-1} (\ln N)^{1+\epsilon} (\ln T)^{1+\epsilon} \right)$. Equation (C.20) now follows immediately. As a consequence, we can assume that

$$\lim_{N,T \rightarrow \infty} \hat{\beta}_{j,i}^{P_j'} \hat{\mathbf{u}}_{j,t} \hat{d}_{j,t} = \beta_{j,i}^{0'} \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0). \quad (\text{C.21})$$

Equation (C.21) and elementary algebra now give $\lim_{N,T \rightarrow \infty} S(\widehat{\mathbf{B}}^P, \widehat{\mathbf{U}}^P, \theta) = S(\mathbf{B}^0, \mathbf{U}^0, \theta)$, which proves the last statement of the lemma. $\square$

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

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

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

for every $\epsilon > 0$.

Proof. Lemma C.3 entails that we can focus on the event $|\hat{\theta} - \theta^0| < c_B$. 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

$$\begin{aligned} & \frac{S(\hat{\mathbf{B}}^P, \hat{\mathbf{U}}^P, \theta) - S(\hat{\mathbf{B}}^P, \hat{\mathbf{U}}^P, \theta^0)}{|\theta - \theta^0|} \\ & \geq \inf_{|\theta - \theta_0| \in B_{N,T}} \frac{E\omega^0(\eta, \theta)}{|\theta - \theta^0|} - \sup_{|\theta - \theta_0| \in B_{N,T}} \frac{|\bar{\omega}_{N,T}^0|}{|\theta - \theta^0|} - 2 \sup_{|\theta - \theta_0| \in B_{N,T}} \frac{|h^0(\eta, \theta) - h^0(\eta, \theta^0)|}{|\theta - \theta^0|}. \end{aligned}$$

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

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

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

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

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

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

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

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

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

$$E |\hat{d}_{\mathcal{D},t} - d_{\mathcal{D},t}(\theta^0)| = \left| \int_{\hat{\theta}}^{\theta^0} f(z) dz \right| \leq c_0 |\hat{\theta} - \theta^0|, \quad (\text{C.22})$$

so that

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

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*

$$E \left| \tilde{R}_{i,t} \right|^{4+\epsilon} < \infty \quad (\text{C.23})$$

for some $\epsilon > 0$, and

$$E \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{R}_{h,s} \tilde{R}_{l,s} d_{\mathcal{D},s}(\theta^0) - E \left[\tilde{R}_{h,s} \tilde{R}_{l,s} d_{\mathcal{D},s}(\theta^0) \right] \right|^2 \leq c_1 T, \quad (\text{C.24})$$

$$E \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{R}_{h,s} \tilde{R}_{l,s} d_{\mathcal{S},s}(\theta^0) - E \left[\tilde{R}_{h,s} \tilde{R}_{l,s} d_{\mathcal{S},s}(\theta^0) \right] \right|^2 \leq c_2 T, \quad (\text{C.25})$$

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

Proof. Equation (C.23) is an immediate consequence of Assumptions 3(i) and 4(i). We show only (C.24); (C.25) follows from exactly the same arguments. To begin with, note that Assumption 5(i) entails that $\tilde{R}_{h,s} \tilde{R}_{l,s} d_{1,s}(\theta^0)$ 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.23), it follows that

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

Note also that

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

Hence, Corollary 1.1 in Shao (1995) and the strict stationarity of $\tilde{R}_{h,s} \tilde{R}_{l,s} d_{\mathcal{D},s}(\theta^0)$ immediately yield (C.24). □

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

$$\begin{aligned} \lim_{\min(N,T) \rightarrow \infty} \sup \frac{1}{N-p} \sum_{h=p+1}^N \hat{g}_j^{(h)} &= \bar{g}_j < \infty, \\ \lim_{\min(N,T) \rightarrow \infty} \inf \frac{1}{N-p} \sum_{h=p+1}^N \hat{g}_j^{(h)} &= \underline{g}_j > 0, \end{aligned}$$

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

Proof. Note that

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

Using Lemma A.3, it is easy to see that

$$\sum_{h=1}^p (\hat{g}_j^{(h)} - g_j^{(h)}) = o_{a.s.} \left(\frac{pNT^{1/2}}{T\pi_j} (\ln N)^{1+\epsilon} (\ln T)^{1/2+\epsilon} \right). \quad (\text{C.26})$$

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

$$\text{tr}(\hat{\Sigma}_j) - \text{tr}(\Sigma_j) = o_{a.s.} \left(\frac{N(\ln N \ln T)^{\frac{1+\epsilon}{2}}}{T^{1/2}\pi_j} \right). \quad (\text{C.27})$$

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

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

under Assumption 6.

Let now $g_j^{u,(h)}$ and $g_j^{e,(h)}$ denote the h -th largest eigenvalues of $\frac{1}{T\pi_j} \sum_{t=1}^T E \left(\mathbf{B}_j^0 \mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0'} \mathbf{B}_j^{0'} d_{j,t}(\theta^0) \right)$

and $\frac{1}{T\pi_j} \sum_{t=1}^T E(\mathbf{e}_t \mathbf{e}_t' d_{j,t}(\theta^0))$ 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 $N \rightarrow \infty$

$$0 < \liminf_{N \rightarrow \infty} < \frac{1}{N-p} \sum_{h=p+1}^N g_j^{(h)} < \limsup_{N \rightarrow \infty} < \infty.$$

Putting all together, the lemma follows. $\square$

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

$$E \left\| T^{-1} \sum_{t=1}^T \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \right\|^2 \leq c_0 T^{-1}. \quad (\text{C.28})$$

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

$$T^{-1/2} \sum_{t=1}^T \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \xrightarrow{D} N(0, V_{\mathbf{u},j}), \quad (\text{C.29})$$

where $V_{\mathbf{u},j} = \lim_{T \rightarrow \infty} E \left(T^{-1} \sum_{t,s=1}^T \mathbf{u}_{j,t}^0 \mathbf{u}_{j,s}^0 d_{j,t}(\theta^0) d_{j,s}(\theta^0) \right)$.

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

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

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

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

where the last inequality follows from Lemma 1 in Peligrad, Utev, and Wu (2007). This proves (C.28). As far as (C.29) is concerned, it follows readily upon checking the assumptions of Theorem 0 in Peligrad (1987). $\square$

D Proofs

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

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

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

We begin by showing that

$$g_j^{(i)} \left(\mathbf{B}_j^0 \frac{1}{T\pi_j} \sum_{t=1}^T E \left(\mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0'} d_{j,t}(\theta^0) \right) \mathbf{B}_j^{0'} \right) \begin{cases} \in (\underline{c}_j^{(i)} N, \bar{c}_j^{(i)} N) & \text{for } 0 \leq i \leq P_j^0 \\ = 0 & \text{for } i \geq P_j^0 + 1 \end{cases}, \quad (\text{D.1})$$

for $j = \mathcal{D}, \mathcal{S}$, where $0 < \underline{c}_j^{(i)} \leq \bar{c}_j^{(i)} < \infty$. 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

$$\begin{aligned} g_{\mathcal{D}}^{(i)} \left(\mathbf{B}_{\mathcal{D}}^0 \frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\mathbf{u}_{\mathcal{D},t}^0 \mathbf{u}_{\mathcal{D},t}^{0'} d_{\mathcal{D},t}(\theta^0) \right) \mathbf{B}_{\mathcal{D}}^{0'} \right) &\geq g_{\mathcal{D}}^{(\min)} \left(\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\mathbf{u}_{\mathcal{D},t}^0 \mathbf{u}_{\mathcal{D},t}^{0'} d_{\mathcal{D},t}(\theta^0) \right) \right) g_{\mathcal{D}}^{(i)} \left(\mathbf{B}_{\mathcal{D}}^{0'} \mathbf{B}_{\mathcal{D}}^0 \right), \\ g_{\mathcal{D}}^{(i)} \left(\mathbf{B}_{\mathcal{D}}^0 \frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\mathbf{u}_{\mathcal{D},t}^0 \mathbf{u}_{\mathcal{D},t}^{0'} d_{\mathcal{D},t}(\theta^0) \right) \mathbf{B}_{\mathcal{D}}^{0'} \right) &\leq g_{\mathcal{D}}^{(\max)} \left(\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\mathbf{u}_{\mathcal{D},t}^0 \mathbf{u}_{\mathcal{D},t}^{0'} d_{\mathcal{D},t}(\theta^0) \right) \right) g_{\mathcal{D}}^{(i)} \left(\mathbf{B}_{\mathcal{D}}^{0'} \mathbf{B}_{\mathcal{D}}^0 \right), \end{aligned}$$

by Assumption 3(ii) it holds that

$$g_{\mathcal{D}}^{(\min)} \left(\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\mathbf{u}_{\mathcal{D},t}^0 \mathbf{u}_{\mathcal{D},t}^{0'} d_{\mathcal{D},t}(\theta^0) \right) \right) > 0.$$

Moreover, by Assumption 3(iv) we have

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

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

$$\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t' d_{\mathcal{D},t}(\theta^0) = \frac{1}{T\pi_j} \sum_{t=1}^T \epsilon_t \epsilon_t' d_{\mathcal{D},t}(\theta^0) - \bar{\epsilon}_{\mathcal{D}} \bar{\epsilon}_{\mathcal{D}}',$$

so that

$$g_{\mathcal{D}}^{(1)} \left(\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\mathbf{e}_t \mathbf{e}_t' d_{\mathcal{D},t}(\theta^0) \right) \right) \leq g_{\mathcal{D}}^{(1)} \left(\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T E \left(\epsilon_t \epsilon_t' d_{\mathcal{D},t}(\theta^0) \right) \right) + g_{\mathcal{D}}^{(1)} \left(\bar{\epsilon}_{\mathcal{D}} \bar{\epsilon}_{\mathcal{D}}' \right).$$

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

$$g_{\mathcal{D}}^{(1)} \left(\bar{\epsilon}_{\mathcal{D}} \bar{\epsilon}_{\mathcal{D}}' \right) \leq \max_{1 \leq i \leq N} \sum_{k=1}^N \frac{1}{(T\pi_{\mathcal{D}})^2} \sum_{t=1}^T \sum_{s=1}^T \left| E \left(\epsilon_{i,t} \epsilon_{k,s} d_{\mathcal{D},t}(\theta^0) d_{\mathcal{D},s}(\theta^0) \right) \right|,$$

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). $\square$

Proof of Theorem A.3. 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

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

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

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

By repeated use of the C_r -inequality we have

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

Using Lemma C.9, we have

$$E \max_{1 \leq h \leq N, 1 \leq l \leq N, 1 \leq t \leq T} \sum_{h'=1}^h \sum_{l'=1}^l \left(\sum_{t'=1}^t \delta_{h',l',t'} \right)^2 \leq c_0 N^2 T;$$

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

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

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

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

Hence, applying (C.22) to (D.5) it follows that

$$\sum_{h=1}^N \sum_{l=1}^N \left(\frac{1}{T\pi_{\mathcal{D}}} \sum_{t=1}^T \tilde{R}_{h,t} \tilde{R}_{l,t} (\hat{d}_{\mathcal{S},t} - d_{\mathcal{D},t}(\theta^0)) \right)^2 \leq c_0 |\hat{\theta} - \theta^0|^2 \frac{1}{T\pi_{\mathcal{D}}^2} \sum_{h=1}^N \sum_{l=1}^N \sum_{t=1}^T \tilde{R}_{h,t}^2 \tilde{R}_{l,t}^2.$$

Equation (C.23) entails

$$E \left(\frac{1}{N^2 T} \sum_{h=1}^N \sum_{l=1}^N \sum_{t=1}^T \tilde{R}_{h,t}^2 \tilde{R}_{l,t}^2 \right) \leq c_0.$$

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

$$E \max_{1 \leq h \leq N, 1 \leq l \leq N, 1 \leq t \leq T} \sum_{h'=1}^h \sum_{l'=1}^l \sum_{t'=1}^t \tilde{R}_{h',t'}^2 \tilde{R}_{l',t'}^2 \leq c_0 N^2 T (\ln N)^2 \ln T,$$

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

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

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

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

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

Proof of Theorem A.4. 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.23). It follows from Lemma A.1 and Theorem A.3 that, for all i and $j = \mathcal{D}, \mathcal{S}$

$$P \left\{ \omega : \lim_{\min(N,T) \rightarrow \infty} \left(\frac{N^{-\ell_j} \hat{g}_j^{(i)}}{\bar{g}_j(p)} - \varkappa N^{1-\ell_j} \right) = \infty \right\} = 1.$$

Thus, by continuity, we can assume henceforth that

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

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

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

By definition

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

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

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

Hence, we have

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

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

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

Hence, using (A.22), we conclude via Markov's inequality that

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

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

We now turn to showing (A.24). Again Lemma A.1 and Theorem A.3 entail that

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

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

$$\lim_{\min(N,T) \rightarrow \infty} \psi \left(\hat{g}_j^{(i)} \right) = 1. \quad (\text{D.7})$$

Consider

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

Then we have

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

Given that

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

by Markov's inequality it follows that

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

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

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

Hence the proof of (A.24) is complete. $\square$

Proof of Theorem 1. See the proof of (the identical) Theorem 3 in Trapani (2018). $\square$

Proof of Theorem 2. The result follows from combining (A.35) and (A.36), with

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

$\square$

Proof of Theorem 3. The result readily follows by combining the results in Theorem A.6 (see also Theorem 3 in Giglio and Xiu, 2021). $\square$

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 $\hat{T}_j$); extending to the case where $\hat{d}_{j,t}$ is employed is straightforward in light of the results derived above. It holds that

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

where

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

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

$$\frac{1}{N} \sum_{i=1}^N \left\| \beta_{j,i} - \hat{\beta}_{j,i} \right\|^2 = o_P(1), \quad (\text{D.10})$$

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

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

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

The LLN entails

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

upon showing that

$$\frac{1}{N} \sum_{i=1}^N \left(\bar{\epsilon}_{j,i} + (\gamma_{j,0} - \hat{\gamma}_j) + \frac{1}{T_j} \sum_{t=1}^T \left(\beta'_{j,i} (\gamma_{j,1} + \mathbf{u}_{j,t}) - \hat{\beta}'_{j,i} (\hat{\gamma}_{j,1} + \hat{\mathbf{u}}_{j,t}) \right) d_{j,t} \right)^2 = o_P(1),$$

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

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \left(\bar{\epsilon}_{j,i} + (\gamma_{j,0} - \hat{\gamma}_j) + \frac{1}{T_j} \sum_{t=1}^T \left(\beta'_{j,i} (\gamma_{j,1} + \mathbf{u}_{j,t}) - \hat{\beta}'_{j,i} (\hat{\gamma}_{j,1} + \hat{\mathbf{u}}_{j,t}) \right) d_{j,t} \right)^2 \\ & \leq c_0 \left(\frac{1}{N} \sum_{i=1}^N \bar{\epsilon}_{j,i}^2 + (\gamma_{j,0} - \hat{\gamma}_j)^2 + \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T_j} \sum_{t=1}^T \left(\beta'_{j,i} (\gamma_{j,1} + \mathbf{u}_{j,t}) - \hat{\beta}'_{j,i} (\hat{\gamma}_{j,1} + \hat{\mathbf{u}}_{j,t}) \right) \right)^2 d_{j,t} \right). \end{aligned}$$

By (A.35), $\gamma_{j,0} - \hat{\gamma}_j = o_P(1)$. Also, combining Theorem 2 and (D.10)-(D.11), it is easy to see that

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T_j} \sum_{t=1}^T \left(\beta'_{j,i} (\gamma_{j,1} + \mathbf{u}_{j,t}) - \hat{\beta}'_{j,i} (\hat{\gamma}_{j,1} + \hat{\mathbf{u}}_{j,t}) \right) \right)^2 d_{j,t} = o_P(1).$$

Finally we have

$$\max_{1 \leq i \leq N} E(\bar{\epsilon}_{j,i}^2) = \max_{1 \leq i \leq N} \frac{1}{T_j^2} \sum_{t=1}^T \sum_{s=1}^T E(\epsilon_{i,t} \epsilon_{i,s} d_{j,t} d_{j,s}) \leq c_0 T_j^{-1},$$

using Assumption 4(ii). Hence

$$\frac{1}{N} \sum_{i=1}^N \bar{\epsilon}_{j,i}^2 = o_P(1).$$

Putting all together, the desired result follows. $\square$

Proof of Theorem A.5. We present the full version of the proof of (A.35) 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 $\hat{P}_j - P_j^0$ is negligible. To this end, we assume without loss of generality that $\hat{P}_j \geq P_j^0$, and we omit dependence on θ or θ^0 whenever possible.

Let

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

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

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

Note that

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

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

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

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

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

We now show that

$$\left\| \widetilde{\mathbf{H}}_{jj}^{\mathbf{r}} \widetilde{\mathbf{\Gamma}}_j \right\|_F = o_{P^*}(1), \quad (\text{D.14})$$

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

$$\left\| \widetilde{\mathbf{H}}_{jj}^{\mathbf{r}} \right\|_F = \text{tr} \left(\widetilde{\mathbf{H}}_{jj}^{\mathbf{r}} \widetilde{\mathbf{H}}_{jj}^{\mathbf{r}} \right) \leq \left(\widehat{P}_j - P_j^0 \right) g^{(1)} \left(\widetilde{\mathbf{H}}_{jj}^{\mathbf{r}} \widetilde{\mathbf{H}}_{jj}^{\mathbf{r}} \right),$$

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

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

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

$$\widehat{\mathbf{H}}_{jj}^{*\mathbf{r}} \widehat{\mathbf{\Gamma}}_j^* = \widehat{\mathbf{H}}_{jj}^{\mathbf{r}} \widehat{\mathbf{\Gamma}}_j^{*P_j^0} + o_{P^*}(1),$$

which, combined with (D.14), yields

$$\widehat{\mathbf{H}}_{jj}^{\mathbf{r}} \widehat{\mathbf{\Gamma}}_j = \widehat{\mathbf{H}}_{jj}^{\mathbf{r}} \widehat{\mathbf{\Gamma}}_j^{*P_j^0} + o_{P^*}(1). \quad (\text{D.15})$$

We now conclude the proof of (A.35) by showing that

$$\widehat{\mathbf{H}}_{jj}^{\mathbf{r}} \widehat{\mathbf{\Gamma}}_j^{*P_j^0} = \mathbf{\Gamma}_j + o_P(1).$$

Let $\widehat{\mathbf{X}}_j^{P_j^0}(\theta_0) = \left[\boldsymbol{\iota}_N, \widehat{\mathbf{B}}_j^{P_j^0}(\theta_0) \right]$. By the same logic as in the previous passages, and using Theorem 3.4 in Massacci (2017), we have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{\Gamma}}_{j,t}^{*P_j^0} \widehat{d}_{j,t} &= \left(\widehat{\mathbf{X}}_j^{\widehat{P}_j'}(\widehat{\theta}) \widehat{\mathbf{X}}_j^{\widehat{P}_j}(\widehat{\theta}) \right)^{-1} \widehat{\mathbf{X}}_j^{\widehat{P}_j'}(\widehat{\theta}) \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t \widehat{d}_{j,t} \\ &= \left(\widehat{\mathbf{X}}_j^{P_j^0'}(\theta_0) \widehat{\mathbf{X}}_j^{P_j^0}(\theta_0) \right)^{-1} \widehat{\mathbf{X}}_j^{P_j^0'}(\theta_0) \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t \widehat{d}_{j,t} + o_{P^*}(1) + o_P(1). \end{aligned}$$

Also

$$\begin{aligned}
& \left(\widehat{\mathbf{X}}_j^{P_0'}(\theta_0) \widehat{\mathbf{X}}_j^{P_0}(\theta_0) \right)^{-1} \widehat{\mathbf{X}}_j^{P_0'}(\theta_0) \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t \widehat{d}_{j,t} \\
&= \left(\widehat{\mathbf{X}}_j^{P_0'}(\theta_0) \widehat{\mathbf{X}}_j^{P_0}(\theta_0) \right)^{-1} \widehat{\mathbf{X}}_j^{P_0'}(\theta_0) \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t d_{j,t}(\theta^0) \\
&\quad + \left(\widehat{\mathbf{X}}_j^{P_0'}(\theta_0) \widehat{\mathbf{X}}_j^{P_0}(\theta_0) \right)^{-1} \widehat{\mathbf{X}}_j^{P_0'}(\theta_0) \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t (\widehat{d}_{j,t} - d_{j,t}(\theta^0)) \\
&= I + II.
\end{aligned} \tag{D.16}$$

Noting that

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

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

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

so that, in equation (D.16), we have $II = O_P(T^{-1})$. Also, by Corollary 3.1(a) in Massacci (2017),

$$\left\| \widehat{\mathbf{B}}_j^{P_0} - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right\|^2 = O_P(NC_{N,T}^{-2}), \text{ which in turn implies}$$

$$\widehat{\mathbf{X}}_j^{P_0}(\theta_0) - (\boldsymbol{\iota}_N, \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj}) = \widehat{\mathbf{X}}_j^{P_0}(\theta_0) - \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^\Gamma = O_P(NC_{N,T}^{-2});$$

thus

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{\Gamma}}_{j,t}^{*P_0} \widehat{d}_{j,t} &= \left(\frac{\widehat{\mathbf{H}}_{jj}^{\Gamma'} \mathbf{X}_j^0 \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^\Gamma}{N} \right)^{-1} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\Gamma'} \mathbf{X}_j^0}{N} \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{R}_t d_{j,t}(\theta^0) \right) \\
&\quad + O_P(C_{N,T}^{-2}) + O_P(T^{-1}) + o_{P^*}(1).
\end{aligned}$$

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

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{\Gamma}}_{j,t}^{*P_j^0} \widehat{d}_{j,t} - \frac{T_j}{T} \left(\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}} \right)^{-1} \mathbf{\Gamma}_j^0 \\
&= \frac{T_j}{T} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'} \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N} \right)^{-1} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'} \alpha_j}{N} \right) \\
&\quad + \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'} \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N} \right)^{-1} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'}}{N} \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{B}_j^0 \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \right) \\
&\quad + \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'} \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N} \right)^{-1} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'}}{N} \right) \left(\frac{1}{T} \sum_{t=1}^T \epsilon_t d_{j,t}(\theta^0) \right) \\
&\quad + O_P(C_{N,T}^{-2}) + O_P(T^{-1}) + o_{P^*}(1) \\
&= I_a + I_b + I_c + O_P(C_{N,T}^{-2}) + O_P(T^{-1}) + o_{P^*}(1)
\end{aligned}$$

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

$$\begin{aligned}
& E \left(\frac{1}{NT} \sum_{t=1}^T \mathbf{X}_j^{0'} \epsilon_t d_{j,t}(\theta^0) \right)^2 \\
&= \frac{1}{(NT)^2} \sum_{i,k=1}^N \sum_{t,s=1}^T X_{i,j}^0 X_{k,j}^{0'} E \left(\epsilon_{i,t} \epsilon_{j,s} d_{j,t}(\theta^0) d_{j,s}(\theta^0) \right) \\
&\leq \frac{1}{(NT)^2} \max_{1 \leq i \leq N} \|X_{i,j}^0\|^2 \sum_{i,k=1}^N \sum_{t,s=1}^T |E \left(\epsilon_{i,t} \epsilon_{j,s} d_{j,t}(\theta^0) d_{j,s}(\theta^0) \right)| \\
&\leq c_0 \frac{N}{(NT)^2} \max_{1 \leq i \leq N} \sum_{k=1}^N \sum_{t,s=1}^T |E \left(\epsilon_{i,t} \epsilon_{j,s} d_{j,t}(\theta^0) d_{j,s}(\theta^0) \right)| \leq \frac{c_0}{NT},
\end{aligned}$$

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.28) yields

$$\left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'} \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}}}{N} \right)^{-1} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}'} \mathbf{X}_j^{0'} \mathbf{B}_j^0}{N} \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \right) = O_P\left(\frac{1}{\sqrt{T}}\right) \quad (\text{D.17})$$

Putting all together, it follows that

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

the desired result now follows immediately. We now turn to showing (A.36). 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

$$\widehat{\Lambda}_j - \Lambda_j^0 \widehat{\mathbf{H}}_{jj} = \left(\sum_{t=1}^T \widetilde{\mathbf{g}}_t \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} \widehat{d}_{j,t} \right) \left(\sum_{t=1}^T \widehat{\mathbf{u}}_{j,t}^{P_j^0} \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} \widehat{d}_{j,t} \right)^{-1}.$$

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

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

Also

$$\frac{1}{T} \sum_{t=1}^T \widetilde{\mathbf{g}}_t \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} \widehat{d}_{j,t} = \frac{1}{T} \sum_{t=1}^T \widetilde{\mathbf{g}}_t \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} d_{j,t}(\theta^0) + \frac{1}{T} \sum_{t=1}^T \widetilde{\mathbf{g}}_t \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} (\widehat{d}_{j,t} - d_{j,t}(\theta^0)) = I + II. \quad (\text{D.20})$$

Consider II ; we have

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

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

$$I = \frac{1}{T} \sum_{t=1}^T (\mathbf{g}_t - \mathbf{a}) \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} d_{j,t}(\theta^0) + \frac{1}{T} \sum_{t=1}^T (\mathbf{g}_t - \mathbf{a}) \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{u}}_{j,t}^{P_j^{0'}} d_{j,t}(\theta^0) = I_a + I_b.$$

Using (C.28) and Assumption 8(ii), it follows that $I_b = O_P(T^{-1})$.

$$I_a = \frac{1}{T} \sum_{t=1}^T \Lambda_j^0 \mathbf{u}_{j,t}^0 \hat{\mathbf{u}}_{j,t}^{P_j^{0'}} d_{j,t}(\theta^0) + \frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \hat{\mathbf{u}}_{j,t}^{P_j^{0'}} d_{j,t}(\theta^0) = I_{a,1} + I_{a,2}.$$

Now

$$\begin{aligned} I_{a,1} &= \Lambda_j^0 \widehat{\mathbf{H}}_{jj} \widehat{\mathbf{H}}_{jj}^{-1} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0'} \right) (\widehat{\mathbf{H}}_{jj}')^{-1} d_{j,t}(\theta^0) \\ &\quad + \Lambda_j^0 \frac{1}{T} \sum_{t=1}^T \mathbf{u}_{j,t}^0 \left(\hat{\mathbf{u}}_{j,t}^{P_j^0} - \widehat{\mathbf{H}}_{jj}^{-1} \mathbf{u}_{j,t}^0 \right)' d_{j,t}(\theta^0) \\ &= I_{a,1,1} + I_{a,1,2}. \end{aligned}$$

Following the same passages as in the proof of Lemma B.2 in Bai (2003), it can be shown that

$$\frac{1}{T} \sum_{t=1}^T \mathbf{u}_{j,t}^0 \left(\hat{\mathbf{u}}_{j,t}^{P_j^0} - \widehat{\mathbf{H}}_{jj}^{-1} \mathbf{u}_{j,t}^0 \right)' d_{j,t}(\theta^0) = O_P(C_{N,T}^{-2}), \quad (\text{D.22})$$

so that $I_{a,1,2} = O_P(C_{N,T}^{-2})$. Also

$$I_{a,2} = \left(\frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \mathbf{u}_{j,t}^{0'} \right) (\widehat{\mathbf{H}}_{jj}')^{-1} d_{j,t}(\theta^0) + \frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \left(\hat{\mathbf{u}}_{j,t}^{P_j^0} - \widehat{\mathbf{H}}_{jj}^{-1} \mathbf{u}_{j,t}^0 \right)' d_{j,t}(\theta^0) = I_{a,2,1} + I_{a,2,2}.$$

Assumption 8(ii) entails that $I_{a,2,1} = O_P(T^{-1/2})$. Finally, by the same token as for (D.22), $I_{a,2,2} = O_P(C_{N,T}^{-2})$, whence $I_{a,2} = O_P(T^{-1/2}) + O_P(C_{N,T}^{-2})$. Putting all together, we obtain

$$\widehat{\Lambda}_j - \Lambda_j^0 \widehat{\mathbf{H}}_{jj} = O_P\left(\frac{1}{T^{1/2}}\right) + O_P(C_{N,T}^{-2}) + O_P(T^{-1}). \quad (\text{D.23})$$

□

Proof of Theorem A.6. Consider (A.37). Upon inspecting (D.18), under (24) and (25), and using Lemma C.8, it is immediate to see that the limiting behaviour of $T^{1/2} \left(\widehat{\mathbf{\Gamma}}_j^{*P_j^0} - (\widehat{\mathbf{H}}_{jj}^{\mathbf{\Gamma}})^{-1} \mathbf{\Gamma}_j^0 \right)$ is

driven by

$$\left(\frac{\widehat{\mathbf{H}}_{jj}^{\Gamma'} \mathbf{X}_j^0 \mathbf{X}_j^0 \widehat{\mathbf{H}}_{jj}^{\Gamma}}{N} \right)^{-1} \left[\left(\frac{\widehat{\mathbf{H}}_{jj}^{\Gamma'} \mathbf{X}_j^0 \mathbf{B}_j^0}{N} \right) \left(\frac{T}{T_j} \frac{1}{T^{1/2}} \sum_{t=1}^T \mathbf{u}_{j,t}^0 d_{j,t}(\theta^0) \right) + \frac{T_j}{T} \left(\frac{\widehat{\mathbf{H}}_{jj}^{\Gamma'} \mathbf{X}_j^0 \alpha_j}{N} \right) \right].$$

The desired result now follows immediately from (C.29) and Assumption 7. Turning to (A.38), combining (D.23) and (25), it follows that the limiting law of $T^{1/2}(\widehat{\Lambda}_j - \Lambda_j^0 \widehat{\mathbf{H}}_{jj})$ is driven by

$$\begin{aligned} & \left(\frac{1}{T^{1/2}} \sum_{t=1}^T \mathbf{e}_t \mathbf{u}_{j,t}^{0'} (\widehat{\mathbf{H}}_{jj})^{-1} d_{j,t}(\theta^0) \right) \left(\frac{1}{T} \sum_{t=1}^T (\widehat{\mathbf{H}}_{jj})^{-1} \mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0'} (\widehat{\mathbf{H}}_{jj}')^{-1} d_{j,t}(\theta^0) \right)^{-1} \\ &= \left(\frac{1}{T^{1/2}} \sum_{t=1}^T \mathbf{e}_t \mathbf{u}_{j,t}^{0'} d_{j,t}(\theta^0) \right) \left(\frac{1}{T} \sum_{t=1}^T \mathbf{u}_{j,t}^0 \mathbf{u}_{j,t}^{0'} d_{j,t}(\theta^0) \right)^{-1} \widehat{\mathbf{H}}_{jj}, \end{aligned}$$

and using Assumption 9(ii) the desired result follows immediately. $\square$

Proof of Theorem A.7. 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

$$\widehat{\gamma}_{j,0} = \left(i_N' \widehat{\mathbf{M}}_{B,j} i_N \right)^{-1} \left(i_N' \widehat{\mathbf{M}}_{B,j} \mathbf{R}_j(\widehat{\theta}) \right),$$

having defined

$$\widehat{\mathbf{M}}_{B,j} = \widehat{\mathbf{B}}_j (\widehat{\mathbf{B}}_j' \widehat{\mathbf{B}}_j)^{-1} \widehat{\mathbf{B}}_j',$$

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

$$\widehat{\gamma}_{j,0} = \gamma_{j,0} + \left(i_N' \widehat{\mathbf{M}}_{B,j} i_N \right)^{-1} \left(i_N' \widehat{\mathbf{M}}_{B,j} \alpha_j \right) + \left(i_N' \widehat{\mathbf{M}}_{B,j} i_N \right)^{-1} \left(i_N' \widehat{\mathbf{M}}_{B,j} \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} \right). \quad (\text{D.24})$$

We begin by noting that, using Corollary 3.1(a) in Massacci (2017), it is easy to see that

$$\left\| \widehat{\mathbf{M}}_{B,j} - \mathbf{M}_{B,j}^0 \right\|^2 = O_P \left(N C_{N,T}^{-2} \right),$$

so that

$$\frac{i'_N \widehat{\mathbf{M}}_{B,j} i_N}{N} = \frac{i'_N \mathbf{M}_{B,j}^0 i_N}{N} + O_P \left(C_{N,T}^{-2} \right) = O_P(1). \quad (\text{D.25})$$

Also, it holds that

$$E \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right)' \alpha_j \alpha_j' \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right) = \sigma_{\alpha,j}^2 E \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right)' \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right) = O \left(N C_{N,T}^{-2} \right),$$

again by Corollary 3.1(a) in Massacci (2017); this yields the (non sharp) bound

$$i'_N \left(\widehat{\mathbf{M}}_{B,j} - \mathbf{M}_{B,j}^0 \right) \alpha_j = O_P \left(N^{1/2} C_{N,T}^{-1} \right). \quad (\text{D.26})$$

Finally, note that Assumption 7(i) entails that

$$N^{1/2} \left(i'_N \mathbf{M}_{B,j}^0 i_N \right)^{-1} \left(i'_N \mathbf{M}_{B,j}^0 \alpha_j \right) = O_P(1). \quad (\text{D.27})$$

Putting (D.25)-(D.27) together, we obtain

$$N^{1/2} \left(i'_N \widehat{\mathbf{M}}_{B,j} i_N \right)^{-1} \left(i'_N \widehat{\mathbf{M}}_{B,j} \alpha_j \right) = N^{1/2} \left(i'_N \mathbf{M}_{B,j}^0 i_N \right)^{-1} \left(i'_N \mathbf{M}_{B,j}^0 \alpha_j \right) + o_P(1). \quad (\text{D.28})$$

Consider now

$$i'_N \widehat{\mathbf{M}}_{B,j} \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} = i'_N \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} + i'_N \widehat{\mathbf{B}}_j \left(\widehat{\mathbf{B}}_j' \widehat{\mathbf{B}}_j \right)^{-1} \widehat{\mathbf{B}}_j' \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} = I + II. \quad (\text{D.29})$$

Using Assumption 2(ii) and equation (C.22), it is easy to see that

$$\begin{aligned} E \left(i'_N \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} \right)^2 &= T_j^{-2} E \left(\sum_{i,k=1}^N \sum_{t,s=1}^T \epsilon_{i,t} \epsilon_{k,s} \hat{d}_{j,t} \hat{d}_{j,s} \right) \\ &\leq NT_j^{-2} \max_{1 \leq k \leq N} \sum_{i=1}^N \sum_{t,s=1}^T \left| E \left(\epsilon_{i,t} \epsilon_{k,s} \hat{d}_{j,t} \hat{d}_{j,s} \right) \right| \leq c_0 \frac{N}{T_j}, \end{aligned}$$

so that

$$i'_N \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} = O_P \left(N^{1/2} T_j^{-1/2} \right). \quad (\text{D.30})$$

In order to study II in (D.29), note that

$$\begin{aligned} &i'_N \hat{\mathbf{B}}_j \left(\hat{\mathbf{B}}_j' \hat{\mathbf{B}}_j \right)^{-1} \hat{\mathbf{B}}_j' \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} \\ &= i'_N \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \left(\hat{\mathbf{B}}_j' \hat{\mathbf{B}}_j \right)^{-1} \widehat{\mathbf{H}}_{jj}' \mathbf{B}_j^{0'} \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} + i'_N \left(\hat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right) \left(\hat{\mathbf{B}}_j' \hat{\mathbf{B}}_j \right)^{-1} \widehat{\mathbf{H}}_{jj}' \mathbf{B}_j^{0'} \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} \\ &+ i'_N \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \left(\hat{\mathbf{B}}_j' \hat{\mathbf{B}}_j \right)^{-1} \left(\hat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right)' \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} \\ &+ i'_N \left(\hat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right) \left(\hat{\mathbf{B}}_j' \hat{\mathbf{B}}_j \right)^{-1} \left(\hat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right)' \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} \\ &= I + II + III + IV. \end{aligned} \quad (\text{D.31})$$

By construction, $\left(\hat{\mathbf{B}}_j' \hat{\mathbf{B}}_j \right)^{-1} = O_p(N^{-1})$, and $i'_N \mathbf{B}_j^0 = O_P(N)$. Using Assumption 2(ii) and equation (C.22), similarly to (D.30), it also follows that

$$\mathbf{B}_j^{0'} T_j^{-1} \sum_{t=1}^T \epsilon_t \hat{d}_{j,t} = O_P \left(N^{1/2} T_j^{-1/2} \right), \quad (\text{D.32})$$

so that $I = O_P(N^{-1/2}T_j^{-1/2})$. Consider now

$$\begin{aligned}
& \mathbf{B}_j^0 - \widehat{\mathbf{B}}_j \widehat{\mathbf{H}}_{jj}^{-1} \\
&= T^{-1} \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}_{j,t}' \widehat{\mathbf{H}}_{jj}^{-1} + T^{-1} \epsilon_j \left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \widehat{\mathbf{H}}_{jj}^{-1} \\
&+ T^{-1} \epsilon_j \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \widehat{\mathbf{H}}_{jj}^{-1},
\end{aligned} \tag{D.33}$$

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 $\|T^{-1} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}_{j,t}' \widehat{\mathbf{H}}\| = O_P(C_{N,T}^{-2})$ and $\|T^{-1} \epsilon_j \left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \widehat{\mathbf{H}}\| = O_P(C_{N,T}^{-2})$. Hence, it follows that

$$i_N' \mathbf{B}_j^0 \left(T^{-1} \widehat{\mathbf{H}}_{jj} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}_{j,t}' \right) \widehat{\mathbf{H}}_{jj}^{-1} = O_P(N C_{N,T}^{-2}). \tag{D.34}$$

Also

$$\begin{aligned}
i_N' T^{-1} \epsilon_j \left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) &= \sum_{i=1}^N T^{-1} \sum_{t=1}^T \epsilon_{j,i,t} \widehat{d}_{j,t} \left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \\
&\leq \sum_{i=1}^N \left\| T^{-1} \sum_{t=1}^T \epsilon_{j,i,t} \widehat{d}_{j,t} \left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \right\|;
\end{aligned}$$

again by using the same arguments as in the proof of Lemma B.1 in Bai (2003), it can be shown that

$$E \left\| T^{-1} \sum_{t=1}^T \epsilon_{j,i,t} \widehat{d}_{j,t} \left(\widehat{\mathbf{u}}_{j,t}' - \mathbf{u}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \right\| \leq c_0 C_{N,T}^{-2},$$

which yields that

$$i_N' T^{-1} \epsilon_j \left(\widehat{\mathbf{U}}_{j,t}' - \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) = O_P(N C_{N,T}^{-2}). \tag{D.35}$$

Finally, note that

$$\begin{aligned}
E \left\| i_N' T^{-1} \epsilon_j \mathbf{U}_{j,t}^{0'} \right\|^2 &= \sum_{i=1}^N \sum_{j=1}^N T^{-2} \sum_{t=1}^T \sum_{s=1}^T E \left(\epsilon_{i,t} \epsilon_{j,s} \widehat{d}_{j,t} \widehat{d}_{j,s} \right) E \left\| \mathbf{u}_{j,t}^{0'} \mathbf{u}_{j,s}^0 \right\| \\
&\leq \left(\max_{1 \leq t \leq T} E \left\| \mathbf{u}_{j,t}^0 \right\|^2 \right) \frac{N}{T} \max_{1 \leq i \leq N} \sum_{k=1}^N T^{-1} \sum_{t=1}^T \sum_{s=1}^T \left| E \left(\epsilon_{i,t} \epsilon_{k,s} \widehat{d}_{j,t} \widehat{d}_{j,s} \right) \right| \leq c_0 \frac{N}{T},
\end{aligned}$$

after Assumptions 2(ii) and 3(i), and using (C.22), which entails that

$$i'_N T^{-1} \epsilon_j \mathbf{U}_{j,t}^{0'} = O_P \left(N^{1/2} T^{-1/2} \right). \quad (\text{D.36})$$

Putting (D.34)-(D.36) together, it follows that

$$i'_N \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right) = O_P \left(N C_{N,T}^{-2} \right). \quad (\text{D.37})$$

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)

$$\begin{aligned} & \left(\widehat{\mathbf{B}}_j - \mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} \right)' \left(\frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} \right) \\ &= \left(\mathbf{B}_j^0 \widehat{\mathbf{H}}_{jj} T^{-1} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}'_{j,t} \widehat{\mathbf{H}}_{jj}^{-1} \right)' \left(\frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} \right) \\ &+ \left(T^{-1} \epsilon_j \left(\widehat{\mathbf{U}}'_{j,t} - \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) \widehat{\mathbf{H}}_{jj}^{-1} \right)' \left(\frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} \right) \\ &+ \left(T^{-1} \epsilon_j \mathbf{U}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \widehat{\mathbf{H}}_{jj}^{-1} \right)' \left(\frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} \right) \\ &= I + II + III. \end{aligned} \quad (\text{D.38})$$

Using (D.32), it follows that

$$I = \left(\widehat{\mathbf{H}}_{jj} T^{-1} \left(\widehat{\mathbf{H}}_{jj}^{-1} \mathbf{U}_{j,t}^0 - \widehat{\mathbf{U}}_{j,t} \right) \widehat{\mathbf{U}}'_{j,t} \widehat{\mathbf{H}}_{jj}^{-1} \right) \left(\mathbf{B}_j^{0'} \frac{1}{T_j} \sum_{t=1}^T \epsilon_t \widehat{d}_{j,t} \right) = O_P \left(N^{1/2} T_j^{-1/2} C_{N,T}^{-2} \right);$$

further, we have

$$\sum_{i=1}^N \left(\frac{1}{T_j} \sum_{t=1}^T \epsilon_{i,t} \widehat{d}_{j,t} \right) T^{-1} \sum_{t=1}^T \epsilon_{i,t} \widehat{d}_{j,t} \left(\widehat{\mathbf{u}}'_{j,t} - \mathbf{u}_{j,t}^{0'} \left(\widehat{\mathbf{H}}_{jj}^{-1} \right)' \right) = O_P \left(N T_j^{-1/2} C_{N,T}^{-2} \right)$$

$$\sum_{i=1}^N \left(\frac{1}{T_j} \sum_{t=1}^T \epsilon_{i,t} \widehat{d}_{j,t} \right) \left(\frac{1}{T} \sum_{t=1}^T \epsilon_{i,t} \mathbf{u}'_{j,t} \widehat{d}_{j,t} \right) = O_P \left(N T^{-1} \right),$$

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)

$$N^{1/2}(\widehat{\gamma}_{j,0} - \gamma_{j,0}) = \left(\frac{i'_N \mathbf{M}_{B,j}^0 i_N}{N}\right)^{-1} \left(\frac{i'_N \mathbf{M}_{B,j}^0 \alpha_j}{N^{1/2}}\right) + O_P\left(N^{1/2}T^{-1}\right) + o_P(1). \quad (\text{D.40})$$

Recalling that, by Lemma C.8, $\widehat{\pi}_j - \pi_j = O_P(T^{-1})$ with $\eta > \frac{1}{2}$, the desired result follows from Assumption 7(iii) and the Cramer-Wold device. $\square$

Proof of Theorem A.2. 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.14), note that

$$\begin{aligned} M_\theta^{-1/2} \sum_{m=1}^{M_\theta} \left(\zeta_m^\theta(s) - \frac{1}{2} \right) &= M_\theta^{-1/2} \sum_{m=1}^{M_\theta} \left(\zeta_m^\theta(0) - \frac{1}{2} \right) \\ &+ M_\theta^{-1/2} \sum_{m=1}^{M_\theta} \left(\zeta_m^\theta(s) - \zeta_m^\theta(0) - \int_0^{s/f_{N,T}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \right) \\ &+ M_\theta^{-1/2} \sum_{m=1}^{M_\theta} \int_0^{s/f_{N,T}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx. \end{aligned}$$

By the same passages as in the proof of Theorem 3 in Horváth and Trapani (2019), it is easy to see that

$$M_\theta^{-1/2} \sum_{m=1}^{M_\theta} \left(\zeta_m^\theta(0) - \frac{1}{2} \right) = O_{P^*}(1);$$

also

$$E^* \int_{-\infty}^{\infty} \left| M_\theta^{-1/2} \sum_{m=1}^{M_\theta} \left(\zeta_m^\theta(s) - \zeta_m^\theta(0) - \int_0^{s/f_{N,T}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \right) \right|^2 ds = o(1),$$

and

$$\int_{-\infty}^{\infty} \left| M_{\theta}^{-1/2} \sum_{m=1}^{M_{\theta}} \int_0^{s/f_{N,T}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \right|^2 ds = c_0 \frac{M_{\theta}}{f_{N,T}^{1/2}} = o(1),$$

after (A.14). The rest of the proof follows immediately. $\square$

E Bootstrap algorithms

In this section we detail the bootstrap algorithms used in Tables 3 and 5: this is done in Sections E.1 and E.2, respectively. As in Section 6.4, T is the length of the in-sample period, so that $T_{OoS} - T$ is the length of the out-of-sample period.

E.1 In-sample bootstrap algorithms

In this section we provide details on the bootstrap algorithm employed in Table 3 in relation to the in-sample analysis. Algorithm 1.1 in Section E.1.1 is for the unconditional model over the full sample. Algorithm 1.2 in Section E.1.2 is for the unconditional model within regimes $\mathcal{D}$ and $\mathcal{S}$. Algorithm 1.3 in Section E.1.3 is for the conditional model within regimes $\mathcal{D}$ and $\mathcal{S}$. Finally, Algorithm 1.4 in Section E.1.4 is for the comparison between conditional and unconditional model within regimes $\mathcal{D}$ and $\mathcal{S}$.^v

E.1.1 Algorithm 1.1: unconditional model, whole period

Step 1. For $i = 1, \dots, N$ and $t = 1, \dots, T$, generate a bootstrap sample of $R_{i,t}^*$ by resampling weighted residuals as

$$R_{i,t}^* = \hat{\beta}_i' \hat{\gamma}_1 + \bar{\alpha}_{\mathcal{U},i} + \hat{\beta}_i' \hat{\mathbf{u}}_t + \hat{\epsilon}_{i,t}^*, \quad \hat{\epsilon}_{i,t}^* = \hat{\epsilon}_{i,t}^* w_{it}, \quad (\text{D.41})$$

where: $\bar{\alpha}_{\mathcal{U},i} = \hat{\alpha}_{\mathcal{U},i} - \bar{\hat{\alpha}}_{\mathcal{U}}$; $\bar{\hat{\alpha}}_{\mathcal{U}} = N^{-1} \sum_{i=1}^N \hat{\alpha}_{\mathcal{U},i}$ is the average value of the vector $\hat{\alpha}_{\mathcal{U}} = \bar{\mathbf{R}} - \hat{\mathbf{B}}\hat{\Gamma}$, whose i -th element $\hat{\alpha}_{\mathcal{U},i}$ is the estimator for α_i in (1); $\hat{\mathbf{u}}_t$ is the principal components

^vWe 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.

estimator for $\mathbf{u}_t$; $\{w_{it}\}$ is a sequence of *i.i.d.* random variables such that $E(w_{it}) = 0$ and $Var(w_{it}) = 1$. Note that, in line with the null hypothesis being tested, $N^{-1} \sum_{i=1}^N \bar{\alpha}_{\mathcal{U},i} = 0$ by construction.

Step 2. Given $\mathbf{R}^* = (\mathbf{R}_1^*, \dots, \mathbf{R}_T^*)$, where $\mathbf{R}_t^* = (R_{1,t}^*, \dots, R_{N,t}^*)'$, obtain

$$\hat{\mathbf{B}}^* = (\mathbf{R}^* \mathbf{M}_T \widehat{\mathbf{U}}') (\widehat{\mathbf{U}} \mathbf{M}_T \widehat{\mathbf{U}}')^{-1},$$

where $\widehat{\mathbf{U}} = (\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_N)$, $\mathbf{M}_T = I_T - T^{-1} \iota_T \iota_T'$ and $\hat{\mathbf{B}}^* = (\hat{\beta}_1^*, \dots, \hat{\beta}_N^*)'$.

Step 3. Given $\bar{\mathbf{R}}^* = (\bar{R}_1^*, \dots, \bar{R}_N^*)' = T^{-1} \sum_{t=1}^T \mathbf{R}_t^*$, estimate the bootstrap risk premia as

$$\hat{\gamma}_1^* = (\hat{\mathbf{B}}^{*'} \hat{\mathbf{B}}^*)^{-1} (\hat{\mathbf{B}}^{*'} \bar{\mathbf{R}}^*).$$

Step 4. For $i = 1, \dots, N$, estimate the average pricing error as

$$\bar{\hat{\alpha}}_{\mathcal{U}}^* = N^{-1} \sum_{i=1}^N \hat{\alpha}_{\mathcal{U},i}^*,$$

where the estimated bootstrap pricing errors are

$$\hat{\alpha}_{\mathcal{U},i}^* = \bar{R}_i^* - \hat{\beta}_i^{*'} \hat{\gamma}_1^*.$$

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\bar{\hat{\alpha}}_{\mathcal{U}}^{*,b}$, for $b = 1, \dots, B$. Compute the bootstrap *p-value* as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(|\bar{\hat{\alpha}}_{\mathcal{U}}^{*,b}| > |\bar{\hat{\alpha}}_{\mathcal{U}}|).$$

E.1.2 Algorithm 1.2: unconditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1 - Step 4. Step 1 through Step 4 are the same as in Algorithm 1.1.

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\bar{\alpha}_{\mathcal{U}}^{*,b}$, for $b = 1, \dots, B$. For $j = \mathcal{D}, \mathcal{S}$, compute the bootstrap p -value as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I} \left(\left| \bar{\alpha}_{\mathcal{U}}^{*,b} \right| > \left| \bar{\alpha}_{j,\mathcal{U}} \right| \right),$$

where $\bar{\alpha}_{j,\mathcal{U}}$ is the average value of the vector $\hat{\alpha}_{j,\mathcal{U}} = \bar{\mathbf{R}}_j - \hat{\mathbf{B}}\hat{\Gamma}$.

E.1.3 Algorithm 1.3: conditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1. For $i = 1, \dots, N$ and $t = 1, \dots, T$, generate a bootstrap sample of $R_{i,t}^*$ by resampling weighted residuals as

$$\begin{aligned} R_{i,t}^* &= \hat{d}_{\mathcal{D},t} \left(\bar{\alpha}_{\mathcal{D},i} + \hat{\beta}'_{\mathcal{D},i} \hat{\gamma}_{\mathcal{D},1} + \hat{\beta}'_{\mathcal{D},i} \hat{\mathbf{u}}_{\mathcal{D},t} \right) \\ &\quad + \hat{d}_{\mathcal{S},t} \left(\bar{\alpha}_{\mathcal{S},i} + \hat{\beta}'_{\mathcal{S},i} \hat{\gamma}_{\mathcal{S},1} + \hat{\beta}'_{\mathcal{S},i} \hat{\mathbf{u}}_{\mathcal{S},t} \right) + \hat{\epsilon}_{i,t}^*, \end{aligned} \quad \hat{\epsilon}_{i,t}^* = \hat{\epsilon}_{i,t}^* w_{it}, \quad (\text{D.42})$$

where: $\bar{\alpha}_{j,i} = \hat{\alpha}_{j,i} - \bar{\alpha}_j$; $\bar{\alpha}_j = N^{-1} \sum_{i=1}^N \hat{\alpha}_{j,i}$ is the average value of the vector $\hat{\alpha}_j = \bar{\mathbf{R}}_j - \hat{\mathbf{B}}_j \hat{\Gamma}_j$, whose i -th element $\hat{\alpha}_{j,i}$ is the estimator for $\alpha_{j,i}$ in (5); $\hat{\mathbf{u}}_{j,t}$ is the principal components estimator for $\mathbf{u}_{j,t}$; $\{w_{it}\}$ is a sequence of *i.i.d.* random variables such that $E(w_{it}) = 0$ and $\text{Var}(w_{it}) = 1$. Note that, in line with the null hypothesis being tested, $N^{-1} \sum_{i=1}^N \bar{\alpha}_{j,i} = 0$ by construction.

Step 2. For $j = 1, 2$, obtain

$$\hat{\mathbf{B}}_j^* = \left(\mathbf{R}^* \mathbf{M}_{T_j} \hat{\mathbf{U}}_j' \right) \left(\hat{\mathbf{U}}_j \mathbf{M}_{T_j} \hat{\mathbf{U}}_j' \right)^{-1},$$

where $\hat{\mathbf{B}}_j^* = \left(\hat{\beta}_{j,1}^*, \dots, \hat{\beta}_{j,N}^* \right)'$ and

$$\mathbf{M}_{T_j} = \begin{pmatrix} \hat{d}_{j,1} & 0 & \dots & 0 \\ 0 & \hat{d}_{j,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{d}_{j,T} \end{pmatrix} - \frac{1}{T_j} \begin{pmatrix} \hat{d}_{j,1} \\ \hat{d}_{j,2} \\ \vdots \\ \hat{d}_{j,T} \end{pmatrix} \begin{pmatrix} \hat{d}_{j,1} \\ \hat{d}_{j,2} \\ \vdots \\ \hat{d}_{j,T} \end{pmatrix}'.$$

Step 3. Given $\overline{\mathbf{R}}_j^* = (\overline{R}_{j,1}^*, \dots, \overline{R}_{j,N}^*)' = T_j^{-1} \sum_{t=1}^T \hat{d}_{j,t} \mathbf{R}_{j,t}^*$, estimate the bootstrap risk premia as

$$\hat{\gamma}_{j,1}^* = (\hat{\mathbf{B}}_j^{*'} \hat{\mathbf{B}}_j^*)^{-1} (\hat{\mathbf{B}}_j^{*'} \overline{\mathbf{R}}_j^*).$$

Step 4. For $j = \mathcal{D}, \mathcal{S}$ and $i = 1, \dots, N$, estimate the average pricing error as

$$\widehat{\alpha}_j^* = N^{-1} \sum_{i=1}^N \hat{\alpha}_{j,i}^*,$$

where the estimated bootstrap pricing errors are

$$\hat{\alpha}_{j,i}^* = \overline{R}_{j,i}^* - \widehat{\beta}_{j,i}^{*'} \hat{\gamma}_{j,1}^*.$$

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\widehat{\alpha}_j^{*,b}$, for $b = 1, \dots, B$. Compute the bootstrap *p-value* as

$$\widehat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(|\widehat{\alpha}_j^{*,b}| > |\widehat{\alpha}_j|),$$

where $\widehat{\alpha}_j$ is the average value of the vector $\widehat{\alpha}_j = \overline{\mathbf{R}}_j - \widehat{\mathbf{B}}_j \widehat{\Gamma}_j$.

E.1.4 Algorithm 1.4: conditional vs unconditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1. For $i = 1, \dots, N$ and $t = 1, \dots, T$, generate a bootstrap sample of $R_{i,t}^*$ as in (D.41) and (D.42) for the unconditional and the conditional model respectively, by replacing the pricing errors $\overline{\alpha}_i$ and $\overline{\alpha}_{j,i}$, respectively, with $\hat{\alpha}_{j,i}$.

Step 2 - Step 4. Step 2 through Step 4 are the same as Algorithm 1.2 and Algorithm 1.3 for the unconditional and the conditional model, respectively.

Step 5. Repeat Step 1 through 4 B times. Denote the estimates for the average pricing error from Step 4 from the unconditional and the conditional model as $\widehat{\alpha}_{j,\mathcal{U}}^{*,b}$ and $\widehat{\alpha}_j^{*,b}$, respectively,

for $b = 1, \dots, B$ for $b = 1, \dots, B$. Compute the bootstrap p -value as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I} \left[\left(\left| \bar{\alpha}_j^{*,b} \right| - \left| \bar{\alpha}_{j,\mathcal{U}}^{*,b} \right| \right) + \left(\left| \bar{\alpha}_j \right| - \left| \bar{\alpha}_{j,\mathcal{U}} \right| \right) \geq 0 \right],$$

where $\bar{\alpha}_j$ and $\bar{\alpha}_{j,\mathcal{U}}$ are the average values of the vectors $\hat{\alpha}_j = \bar{\mathbf{R}}_j - \hat{\mathbf{B}}_j \hat{\Gamma}_j$ and $\hat{\alpha}_{j,\mathcal{U}} = \bar{\mathbf{R}}_j - \hat{\mathbf{B}} \hat{\Gamma}$, respectively.

E.2 Out-of-sample bootstrap algorithms

E.2.1 Algorithm 2.1: unconditional model, whole period

Step 1 - Step 4. For $t = T + 1, \dots, T_{OoS}$, Step 1 through Step 4 are as in Algorithm 1.1.

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\bar{\alpha}_{\mathcal{U}}^{*,b}$, for $b = 1, \dots, B$. Compute the bootstrap p -value as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I} \left(\left| \bar{\alpha}_{\mathcal{U}}^{*,b} \right| > \left| \bar{\alpha}_{\mathcal{U},OoS} \right| \right),$$

where $\bar{\alpha}_{\mathcal{U},OoS} = N^{-1} \sum_{i=1}^N \hat{\alpha}_{\mathcal{U},OoS,i}$ is the average value of the vector of out-of-sample pricing errors $\hat{\alpha}_{\mathcal{U},OoS} = \bar{\mathbf{R}}_{OoS} - \widehat{\mathbf{R}}_{\mathcal{U},OoS}$.

E.2.2 Algorithm 2.2: unconditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1 - Step 4. Step 1 through Step 4 are the same as in Algorithm 2.1.

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\bar{\alpha}_{\mathcal{U}}^{*,b}$, for $b = 1, \dots, B$. For $j = \mathcal{D}, \mathcal{S}$, compute the bootstrap p -value as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I} \left(\left| \bar{\alpha}_{\mathcal{U}}^{*,b} \right| > \left| \bar{\alpha}_{j,\mathcal{U},OoS} \right| \right),$$

where $\bar{\alpha}_{j,\mathcal{U},OoS}$ is the average value of the vector $\hat{\alpha}_{j,\mathcal{U},OoS} = \bar{\mathbf{R}}_{j,OoS} - \widehat{\mathbf{R}}_{\mathcal{U},OoS}$, and regime j is identified using the threshold value $\hat{\theta}$ estimated from the first T in-sample observations.

E.2.3 Algorithm 2.3: conditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1 - Step 4. For $t = T + 1, \dots, T_{OoS}$, Step 1 through Step 4 are as in Algorithm 1.3, where the threshold value $\hat{\theta}$ is estimated from the first T in-sample observations.

Step 5. Repeat Step 1 through Step 4 for B times. Denote the estimates for the average pricing error from Step 4 as $\bar{\alpha}_j^{*,b}$, for $b = 1, \dots, B$. Compute the bootstrap p -value as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I} \left(\left| \bar{\alpha}_j^{*,b} \right| > \left| \bar{\alpha}_{j,OoS} \right| \right),$$

where $\bar{\alpha}_{j,OoS}$ is the average value of the vector $\hat{\alpha}_{j,OoS} = \bar{\mathbf{R}}_{j,OoS} - \widehat{\bar{\mathbf{R}}}_{\mathcal{C},j,OoS}$, and regime j is identified using the threshold value $\hat{\theta}$ estimated from the first T in-sample observations.

E.2.4 Algorithm 2.4: conditional vs unconditional model, regimes $\mathcal{D}$ and $\mathcal{S}$

Step 1 - Step 4. For $t = T + 1, \dots, T_{OoS}$, Step 1 through Step 4 are as in Algorithm 1.4, where the threshold value $\hat{\theta}$ is estimated from the first T in-sample observations.

Step 5. Repeat Step 1 through 4 B times. Denote the estimates for the average pricing error from Step 4 from the unconditional and the conditional model as $\bar{\alpha}_{j,\mathcal{U}}^{*,b}$ and $\bar{\alpha}_j^{*,b}$, respectively, for $b = 1, \dots, B$. Compute the bootstrap p -value as

$$\hat{p}^{*,B} = \frac{1}{B} \sum_{b=1}^B \mathbb{I} \left[\left(\left| \bar{\alpha}_j^{*,b} \right| - \left| \bar{\alpha}_{j,\mathcal{U}}^{*,b} \right| \right) + \left(\left| \bar{\alpha}_{j,OoS} \right| - \left| \bar{\alpha}_{j,\mathcal{U},OoS} \right| \right) \geq 0 \right],$$

where $\bar{\alpha}_{j,OoS}$ and $\bar{\alpha}_{j,\mathcal{U},OoS}$ are the average values of the vectors $\hat{\alpha}_{j,OoS} = \bar{\mathbf{R}}_{j,OoS} - \widehat{\bar{\mathbf{R}}}_{\mathcal{C},j,OoS}$ and $\hat{\alpha}_{j,\mathcal{U},OoS} = \bar{\mathbf{R}}_{j,OoS} - \widehat{\bar{\mathbf{R}}}_{\mathcal{U},OoS}$, respectively, and regime j is identified using the threshold value $\hat{\theta}$ estimated from the first T in-sample observations.

F Descriptive statistics

In Table F.1, this section provides descriptive statistics for the average (within the month) daily AD Bear returns of Lu and Murray (2019).

Table F.1: Descriptive statistics, average (within the month) AD Bear daily portfolio returns

	199701 - 200507	199701 - 201507
Mean	-0.256	-0.287
Std. dev.	1.015	1.105
Median	-0.446	-0.498
Maximum	2.577	5.110
Minimum	-1.793	-1.804
Skewness	0.743	1.162
Kurtosis	-0.173	2.011
Jarque-Bera	52.671***	59.281***

Notes. This table presents descriptive statistics for the empirical distribution of average (within the month) daily AD Bear returns over the periods from January 1997 to July 2005, and from January 1997 to July 2015, a total of 103 and 223 observations, respectively. “Jarque–Bera” denotes the Jarque–Bera goodness of fit test for normality; the joint null hypothesis is that skewness and kurtosis are equal to 0 and 3, respectively. ***indicates significance at the 1% level.