Fundamental properties of linear factor models

Damir Filipović* Paul Schneider†

3 September 2024

Abstract

We study conditional linear factor models in the context of asset pricing panels. Our analysis focuses on conditional means and covariances to characterize the cross-sectional and inter-temporal properties of returns and factors as well as their interrelationships. We also review the conditions outlined in Kozak and Nagel (2024) and show how the conditional mean-variance efficient portfolio of an unbalanced panel can be spanned by low-dimensional factor portfolios, even without assuming invertibility of the conditional covariance matrices. Our analysis provides a comprehensive foundation for the specification and estimation of conditional linear factor models.

Keywords: asset pricing, factor models, characteristics, covariances, mean-variance efficient portfolio, stochastic discount factor, covariance estimation
JEL classification: G11, G12, C38

1 Introduction

Since the capital asset pricing model (Sharpe, 1964) and the arbitrage pricing theory of Ross (1976), Chamberlain and Rothschild (1983), Chamberlain (1983), Admati and Pfleiderer (1985), and Reisman (1988), academia and industry alike have shown a strong desire to compress vast asset pricing panels into low-dimensional linear factor representations. Accordingly, an extensive econometric literature has simultaneously developed that exploits statistical arbitrage relations and asymptotic results in the

*EPFL and SFI, damir.filipovic@epfl.ch
†Università della Svizzera italiana and SFI, paul.schneider@usi.ch

time series and cross section to develop estimators for factor loadings and return covariances (Onatski, 2015; Gagliardini et al., 2016; Raponi et al., 2019; Zaffaroni, 2019; Fortin et al., 2023). Recent literature devises models for conditional means of panels (Freyberger et al., 2020; Gu et al., 2020; Kelly et al., 2019), or conditional covariances (Engle et al., 2019). To date, however, there is no comprehensive collection of properties that any conditional linear factor model should satisfy, how they relate to first and second conditional moments, and how these properties change when the factors are tradable.

In this short note, we fill the gap in this literature and discuss exhaustively and from first principles the fundamental properties obeyed by (un)conditional linear factor models. Our analysis focuses on conditional means and covariances to characterize the cross-sectional and inter-temporal properties of returns and factors as well as their interrelationships. The focus is on a concise, rigorous discussion with minimal assumptions and the identification of pitfalls. For example, we do not assume invertibility of the covariance matrices, which is a strong assumption given the large cross-sectional dimensions in modern unbalanced asset pricing panels.

In particular, we investigate under which conditions

(i) factors and residuals are conditionally uncorrelated,

(ii) tradable factors (i.e. factor portfolios) span the conditional mean-variance efficient portfolio,

(iii) a generative risk factor model can be represented by tradable factors that load on the same coefficient matrix and fulfill the first two properties,

and how these three questions are interrelated.[1] For example, we find that a linear factor model with tradable factors and conditionally uncorrelated residuals, as in (i), cannot possibly have an invertible residual covariance matrix. We also review the conditions outlined in Kozak and Nagel (2024) and show how the conditional mean-variance efficient portfolio of an unbalanced panel can be spanned by low-dimensional factor portfolios, relating to (ii). We find that property (iii) holds almost universally, which is remarkable and suggests a general structure for consistent estimators of conditional means and covariance matrices for unbalanced asset pricing panels.

[1] Question (ii) is also studied in Kozak and Nagel (2024), but under more restrictive assumptions. We further extend their paper by relating (ii) to aspects of (i) and (iii).

Our analysis is relevant in particular also for the broader context of models that are linear in nonlinear functions of factors, as they are commonly used in financial machine learning based upon neural nets, or kernel-based methods. It is complemented by extensive examples and provides a comprehensive foundation for the specification and estimation of conditional linear factor models.

The structure of this short note is as follows. In Section 2 we introduce the formal setup for conditional linear factor models and give an overview of our main results. In Section 3 we answer question (i) and derive related results based on the covariances. We also give an example that serves as counterexample in some of our proofs. In Section 4 we discuss risk premia and show that characteristics are covariances under certain conditions. In Section 5 we address question (ii). Using the counterexample, we also show some pitfalls with false implications. In Section 6 we address question (iii). In Section 7 we conclude. The appendix collects all the proofs.

2 Conditional linear factor models

We study conditional linear factor models of the form

$$\boldsymbol{x}_{t+1} = \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}, \tag{1}$$

where $\boldsymbol{x}_{t+1}$ denotes the vector of excess returns of n_t assets over period $[t, t+1]$, for a $n_t \times m$ matrix $\boldsymbol{\Phi}_t$ of asset characteristics that are measurable with respect to the information set (i.e. observable) at t, loading on a vector of m common risk factors $\boldsymbol{f}_{t+1}$, and residuals $\boldsymbol{\epsilon}_{t+1}$.[2]

The above panel specification inherently combines two notions, one within generative statistical models, and one within the study of optimal portfolios. The first interprets the right-hand side of (1) as a data-generating process, where abstract risk factors $\boldsymbol{f}_{t+1}$ and idiosyncratic risk $\boldsymbol{\epsilon}_{t+1}$ generate returns $\boldsymbol{x}_{t+1}$, given $\boldsymbol{\Phi}_t$. The second

[2]Formally, all random variables are modeled on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ along with a sequence of information sets (σ-algebras) $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \cdots \subseteq \mathcal{F}$. We write $\mathbb{E}_t[\cdot]$ and $\mathrm{Cov}_t[\cdot]$ for the conditional mean and covariance given $\mathcal{F}_t$. We assume a regular conditional probability $\mathbb{P}_t$ such that $\mathbb{E}_t[1_A] = \mathbb{P}_t[A]$ for any event $A \in \mathcal{F}$. Conditional expectations given $\mathcal{F}_t$ hence amount to expectations under $\mathbb{P}_t$, which we always assume to exist and be finite. Equality between random variables means $\mathbb{P}_t$-almost sure equality. Assuming a trivial information set, $\mathcal{F}_t = \{\emptyset, \Omega\}$, this setup includes also unconditional moments.

assumes that the factors are *tradable* portfolios of the form

$$\boldsymbol{f}_{t+1} = \boldsymbol{W}_t^\top \boldsymbol{x}_{t+1}, \tag{2}$$

for a $n_t \times m$ matrix $\boldsymbol{W}_t$ of weights that are measurable with respect to the information set at t. Such tradable factors should serve as an accurate low-rank representation of the full cross section $\boldsymbol{x}_{t+1}$, with implied residuals $\boldsymbol{\epsilon}_{t+1} = \boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}$, given $\boldsymbol{\Phi}_t$. In this portfolio context an eminent question concerns the conditional Sharpe ratios that can be attained from trading in the full cross section $\boldsymbol{x}_{t+1}$ or only in the factor portfolios $\boldsymbol{f}_{t+1}$, respectively.[3]

Our analysis of the introductory questions (i)–(iii) is based exclusively on the first two conditional moments. We therefore denote by $\boldsymbol{\mu}_t \coloneqq \mathbb{E}_t[\boldsymbol{x}_{t+1}]$ and $\boldsymbol{\Sigma}_t \coloneqq \mathrm{Cov}_t[\boldsymbol{x}_{t+1}]$ the conditional mean and covariance matrix of the excess returns, and analogously we write $\boldsymbol{\mu}_{\boldsymbol{f},t} \coloneqq \mathbb{E}_t[\boldsymbol{f}_{t+1}]$, $\boldsymbol{\Sigma}_{\boldsymbol{f},t} \coloneqq \mathrm{Cov}_t[\boldsymbol{f}_{t+1}]$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} \coloneqq \mathrm{Cov}_t[\boldsymbol{\epsilon}_{t+1}]$ for those of the factors and residuals, respectively. We do not assume that these matrices have full rank. Instead of the regular matrix inverses we use the pseudoinverse, denoted by $\boldsymbol{A}^+$ for any matrix $\boldsymbol{A}$.[4]

To simplify the exposition, we will omit the qualifier "conditional" from probabilities, expectations, covariances, and correlations in the following. We will also refer to excess returns simply as "returns". Figure 1 gives an overview of some of our main findings.

3 Factor and residual correlation

Our first result gives necessary and sufficient conditions that factors and residuals are uncorrelated, which answers the introductory question (i), and relates this property to the covariance matrix of the returns. We refer to $\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}$ as the *factor-spanned component* of the returns whose covariance is given by $\mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}] = \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{f},t} \boldsymbol{\Phi}_t^\top$.

[3] Our setup is the same as in Kozak and Nagel (2024) and in fact more general, as they assume that matrices have full rank. In view of Footnote 4, our analysis even applies for $m \geq n_t$.

[4] Also called Moore–Penrose generalized inverse. Let $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{W}^\top$ be the singular value decomposition of a matrix $\boldsymbol{A}$, with orthogonal matrices $\boldsymbol{V}$ and $\boldsymbol{W}$. Then the pseudoinverse of $\boldsymbol{A}$ is given by $\boldsymbol{A}^+ = \boldsymbol{W}\boldsymbol{D}^+\boldsymbol{V}^\top$, where $\boldsymbol{D}^+$ is the transpose of $\boldsymbol{D}$ in which the positive singular values are replaced by their reciprocals, see (Horn and Johnson, 1990, Problem 7.3.7) or (Schott, 2017, Chapter 5). Consequently, $\boldsymbol{A}^+\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^+$ are the orthogonal projections on the images of $\boldsymbol{A}^\top$ and $\boldsymbol{A}$, respectively. If $\boldsymbol{A}^\top\boldsymbol{A}$ is invertible then $\boldsymbol{A}^+ = (\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top$.

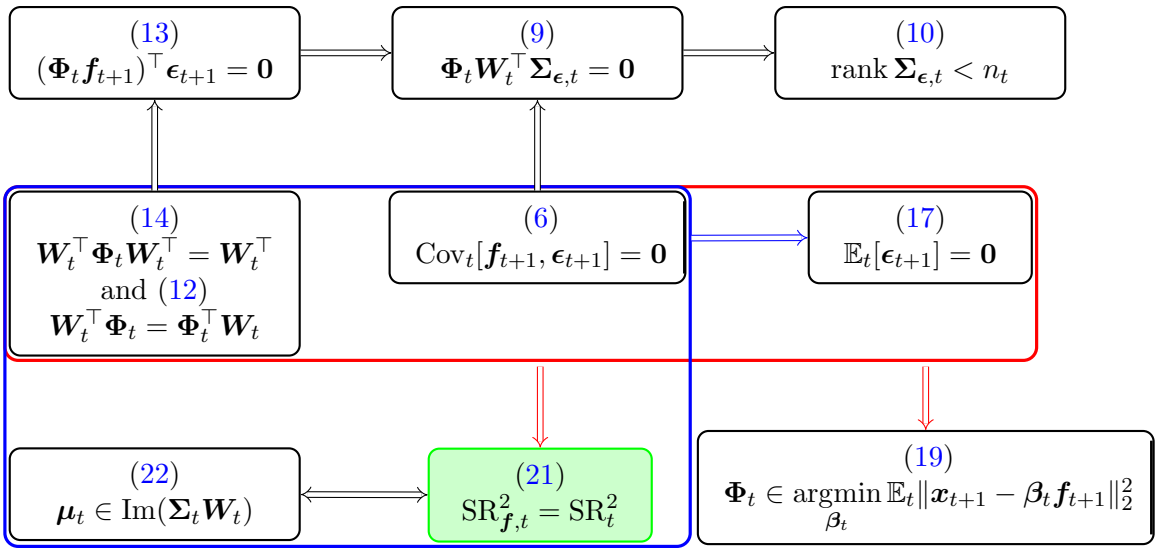


Figure 1: Main results at a glance. This figure illustrates the relationship between several properties of conditional linear factor models assuming that the factors are tradable (2). One-sided implications are strict, meaning the converse does not generally hold. We illustrate this with counterexamples. $\mathrm{SR}^2_{\boldsymbol{f},t}$ and SR^2_t are the maximum squared Sharpe ratios that are attainable by the factors and the full cross section.

Lemma 3.1. *The factor-spanned component and the residuals are uncorrelated,*

$$\operatorname{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \mathbf{0}, \tag{3}$$

if and only if the covariance matrix of the returns is given by the sum

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{f},t} \boldsymbol{\Phi}_t^\top + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t}. \tag{4}$$

Moreover, if

$$\operatorname{rank}(\boldsymbol{\Phi}_t) = m \tag{5}$$

then any of (3) *or* (4) *is equivalent to*

$$\operatorname{Cov}_t[\boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \mathbf{0}. \tag{6}$$

We obtain stronger results under the assumption that the factors are tradable.

Proposition 3.2. *Assume that* (2) *holds. Then any of* (3) *or* (4) *is equivalent to*

$$\boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t. \tag{7}$$

Moreover, if

$$\operatorname{Im} \boldsymbol{W}_t^\top \cap \ker \boldsymbol{\Phi}_t = \{\mathbf{0}\}, \tag{8}$$

then any of (3), (4) *or* (7) *is equivalent to* (6).

Factor-spanned components that are uncorrelated with residuals imply a zero matrix-product relation between covariance matrices, but not conversely, as the following proposition shows.

Proposition 3.3. *Assume that* (2) *holds. Then any of* (3), (4) *or* (7) *implies the zero matrix-product relation*

$$\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} = \mathbf{0}. \tag{9}$$

However, the converse is not true, as (9) *does not imply any of* (3), (4) *or* (7).

As an important corollary of Proposition 3.3, we furthermore obtain an impossibility result.

Corollary 3.4. *Assume that* (2) *holds and the factor model* (1) *is non-degenerate,* $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \neq \mathbf{0}$. *Then the residual covariance matrix cannot have full rank,*

$$\operatorname{rank} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} < n_t, \tag{10}$$

under any of the equivalent conditions (3), (4) *or* (7).

Corollary 3.4 implies in particular that there exists no linear factor model (1) with tradable factors (2), which are portfolios in $\boldsymbol{x}_{t+1}$, and uncorrelated idiosyncratic risk components $\boldsymbol{\epsilon}_{t+1}$ with nonsingular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t}$. Corollary 3.4 thus has ramifications for the literature on shrinkage of covariance matrices with factor structure (Ledoit and Wolf, 2020, Section 5), where oftentimes the spectrum of the residual covariance is lifted in order to obtain invertibility.

Tradable factors with weight matrix inducing a projection imply *cross-sectional orthogonality*, which is stronger than (9), in the following sense.

Lemma 3.5. *Assume that* (2) *holds for a weight matrix such that the matrix product*

$$\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = (\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top)^2 \tag{11}$$

is a (generally non-orthogonal) projection. Then the rows of $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top$ *and the residuals are cross-sectionally orthogonal,* $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\epsilon}_{t+1} = \mathbf{0}$*, and therefore* (9) *holds. If in addition the matrix product*

$$\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top \tag{12}$$

is self-adjoint, and thus an orthogonal projection, then the factor-spanned component and the residuals are cross-sectionally orthogonal,

$$(\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1})^\top \boldsymbol{\epsilon}_{t+1} = \mathbf{0}. \tag{13}$$

Even in this case, uncorrelatedness (3) *generally does not follow.*

We are able to further qualify the above results under more specific assumptions on the weight matrix, as the following lemma shows.

Lemma 3.6. *The weight matrix* $\boldsymbol{W}_t$ *satisfies*

$$\boldsymbol{W}_t^\top \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{W}_t^\top \tag{14}$$

if and only if (8) *and* (11) *hold. In this case, we also have* $\ker(\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) = \ker \boldsymbol{W}_t^\top$*. Property* (14) *holds in particular for weight matrices of the form*

$$\boldsymbol{W}_t^\top = \boldsymbol{R}_t (\boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t)^+ \boldsymbol{S}_t, \tag{15}$$

for some $m \times m$*-matrix* $\boldsymbol{R}_t$ *and* $n_t \times n_t$*-matrix* $\boldsymbol{S}_t$*. If in addition we have*

$$\operatorname{Im}(\boldsymbol{\Phi}_t \boldsymbol{R}_t) \cap \ker \boldsymbol{S}_t = \{\mathbf{0}\}, \tag{16}$$

then $\operatorname{Im}(\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) = \operatorname{Im}(\boldsymbol{\Phi}_t \boldsymbol{R}_t)$*.*

Weight matrices of the form (15) include the following two examples.

Example 1. *For* $\boldsymbol{R}_t = \boldsymbol{I}_m$ *and* $\boldsymbol{S}_t = \boldsymbol{I}_{n_t}$ *in* (15)*, we obtain the* OLS factors

$$\boldsymbol{f}_{t+1}^{\mathrm{OLS}} := \boldsymbol{\Phi}_t^+ \boldsymbol{x}_{t+1}.$$

Here, $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^+$ is the orthogonal projection in $\mathbb{R}^{n_t}$ onto the image of $\boldsymbol{\Phi}_t$. The OLS factors $\boldsymbol{f}_{t+1}^{\mathrm{OLS}}$ minimize the cross-sectional least-squares problem $\|\boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}\|_2$ over all factors $\boldsymbol{f}_{t+1}$ in $\mathbb{R}^m$.[5]

Example 2. *For $\boldsymbol{R}_t = \boldsymbol{I}_m$ and $\boldsymbol{S}_t$ in (15) such that $\boldsymbol{S}_t^\top \boldsymbol{S}_t = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t}^+$, we obtain the* GLS factors

$$\boldsymbol{f}_{t+1}^{\mathrm{GLS}} := (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{S}_t \boldsymbol{x}_{t+1}.$$

The GLS factors $\boldsymbol{f}_{t+1}^{\mathrm{GLS}}$ minimize the squared Mahalanobis length $(\boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1})^\top \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t}^+ (\boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1})$ over all factors $\boldsymbol{f}_{t+1}$ in $\mathbb{R}^m$.[6] *More generally, we see that (15) allows for rotated GLS factors by an appropriate choice of $\boldsymbol{R}_t$.*

The following example illustrates the dichotomy between cross-sectional orthogonality as defined in Lemma 3.5, and orthogonality in the time series, as it shows OLS factors that are correlated with residuals. It serves as counterexample in the proofs of Proposition 3.3 and Lemmas 3.5 and 5.3, and prepares and complements Proposition 5.2 below.

Example 3. *Let $n_t = 3$, $m = 2$, and let ξ_1, ξ_2, ξ_3 be uncorrelated random variables with variances $\mathrm{Cov}_t[\xi_i] = a_i > 0$ and means $b_i = \mathbb{E}_t[\xi_i] \in \mathbb{R}$, for $i = 1, 2, 3$. Assume returns and their characteristics are given by*

$$\begin{bmatrix} \boldsymbol{x}_{t+1,1} \\ \boldsymbol{x}_{t+1,2} \\ \boldsymbol{x}_{t+1,3} \end{bmatrix} := \begin{bmatrix} \xi_1 \\ \xi_2 \\ \frac{\rho}{a_1}\xi_1 + \frac{\rho}{a_2}\xi_2 + \xi_3 \end{bmatrix}, \quad \boldsymbol{\Phi}_t := \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

for some $\rho \in \mathbb{R}$. This gives the following mean and covariance matrix of $\boldsymbol{x}_{t+1}$,

$$\boldsymbol{\mu}_t = \begin{bmatrix} b_1 \\ b_2 \\ \frac{\rho}{a_1}b_1 + \frac{\rho}{a_2}b_2 + b_3 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t = \begin{bmatrix} a_1 & 0 & \rho \\ 0 & a_2 & \rho \\ \rho & \rho & \frac{\rho}{a_1} + \frac{\rho}{a_2} + a_3 \end{bmatrix}.$$

[5] If $\mathrm{rank}(\boldsymbol{\Phi}_t) = m$ then the OLS problem has a unique solution. In general, all factors of the form $\boldsymbol{f}_{t+1} = \boldsymbol{f}_{t+1}^{\mathrm{OLS}} + (\boldsymbol{I}_m - \boldsymbol{\Phi}_t^+ \boldsymbol{\Phi}_t)\boldsymbol{w}_{t+1}$, for any $\boldsymbol{w}_{t+1} \in \mathbb{R}^m$, are OLS solutions. Hence $\boldsymbol{f}_{t+1}^{\mathrm{OLS}}$ is the unique OLS solution $\boldsymbol{f}_{t+1}$ with minimal norm $\|\boldsymbol{f}_{t+1}\|_2$.

[6] Subject to similar aspects as discussed in Footnote 5.

Matrix $\boldsymbol{\Phi}_t$ *has full column rank* $m = 2$ *and its pseudoinverse equals*

$$\boldsymbol{\Phi}_t^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We set the portfolio weight matrix $\boldsymbol{W}_t^\top \coloneqq \boldsymbol{\Phi}_t^+$*, so that we obtain the OLS factors (cf. Example 1),*

$$\boldsymbol{f}_{t+1} = \boldsymbol{f}_{t+1}^{\mathrm{OLS}} = \boldsymbol{\Phi}_t^+ \boldsymbol{x}_{t+1} = \begin{bmatrix} \frac{1}{2}\boldsymbol{x}_{t+1,1} + \frac{1}{2}\boldsymbol{x}_{t+1,2} \\ \boldsymbol{x}_{t+1,3} \end{bmatrix},$$

and

$$\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{\Phi}_t \boldsymbol{\Phi}_t^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the orthogonal projection onto the image of $\boldsymbol{\Phi}_t$*. As shown in (the first part of) Lemma 3.5, the factor-spanned component and residuals, given by*

$$\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1} = \begin{bmatrix} \frac{1}{2}\boldsymbol{x}_{t+1,1} + \frac{1}{2}\boldsymbol{x}_{t+1,2} \\ \frac{1}{2}\boldsymbol{x}_{t+1,1} + \frac{1}{2}\boldsymbol{x}_{t+1,2} \\ \boldsymbol{x}_{t+1,3} \end{bmatrix} \quad \textit{and} \quad \boldsymbol{\epsilon}_{t+1} = \begin{bmatrix} \frac{1}{2}\boldsymbol{x}_{t+1,1} - \frac{1}{2}\boldsymbol{x}_{t+1,2} \\ \frac{1}{2}\boldsymbol{x}_{t+1,2} - \frac{1}{2}\boldsymbol{x}_{t+1,1} \\ 0 \end{bmatrix},$$

are cross-sectionally orthogonal, (13), and the zero matrix-product condition (9) holds. However, their covariance,

$$\mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \begin{bmatrix} \frac{1}{2}(a_1 - a_2) & \frac{1}{2}(a_2 - a_1) & 0 \\ \frac{1}{2}(a_1 - a_2) & \frac{1}{2}(a_2 - a_1) & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

is nonzero if $a_1 \neq a_2$*.*

4 Risk premia

The preceding properties of linear factor models have not yet directly used the asset risk premia $\boldsymbol{\mu}_t$ so far. In this section, we investigate the role of the mean in particular with respect to the time series properties of linear factor models. The following elementary lemma gives an algebraic condition in terms of $\boldsymbol{\mu}_t$ such that residual risk

is unpriced for tradable factors.

Lemma 4.1. *Assume that (2) holds. Then residual risk is unpriced,*

$$\mathbb{E}_t[\boldsymbol{\epsilon}_{t+1}] = \mathbf{0}, \tag{17}$$

if and only if

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\mu}_t. \tag{18}$$

The following proposition further extends our knowledge of the properties of linear factor models, by showing that characteristics are covariances under the assumption of uncorrelated factors and residuals and unpriced residual risk. This property in turn concerns optimality of the characteristics matrix $\boldsymbol{\Phi}_t$ in the time series. It extends the discussion of OLS factors in Fama and French (2020), and GLS factors in Kozak and Nagel (2024) with respect to their times series, cross-sectional, and asset pricing properties (see also the IPCA model of Kelly et al., 2019).

Proposition 4.2. *Assume that (6) and (17) hold. Then the characteristics $\boldsymbol{\Phi}_t$ are also covariances for the factors, in the sense that*

$$\boldsymbol{\Phi}_t \in \operatorname*{argmin}_{\boldsymbol{\beta}_t} \mathbb{E}_t \|\boldsymbol{x}_{t+1} - \boldsymbol{\beta}_t \boldsymbol{f}_{t+1}\|_2^2 \tag{19}$$

where the minimum is taken over all $n_t \times m$-matrices $\boldsymbol{\beta}_t$ that are measurable with respect to the information set at t. The minimizer in (19) is unique if and only if the Gram matrix of the factors, $\boldsymbol{\Sigma}_{\boldsymbol{f},t} + \boldsymbol{\mu}_{\boldsymbol{f},t}\boldsymbol{\mu}_{\boldsymbol{f},t}^\top$, has full rank. Sufficient conditions for (6) to hold for tradable factors are given in Proposition 3.2 and Lemma 3.6.

5 Spanning factors

We henceforth assume absence of arbitrage, in the weak sense that[7]

$$\boldsymbol{\mu}_t \in \operatorname{Im} \boldsymbol{\Sigma}_t. \tag{20}$$

[7] An arbitrage is a strategy $\boldsymbol{w}_t$ that yields a nonnegative return, $\boldsymbol{w}_t^\top \boldsymbol{x}_{t+1} \geq 0$, and such that $\mathbb{P}_t[\boldsymbol{w}_t^\top \boldsymbol{x}_{t+1} > 0] > 0$. Condition (20) is necessary but not sufficient for the absence of such arbitrage in the strict sense, which is outside the scope of this short note. Indeed, the orthogonal decomposition $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t,0} + \boldsymbol{\mu}_{t,1}$ according to $\ker \boldsymbol{\Sigma}_t \oplus \operatorname{Im} \boldsymbol{\Sigma}_t$ shows that $\boldsymbol{\mu}_{t,0}^\top \boldsymbol{x}_{t+1} = \|\boldsymbol{\mu}_{t,0}\|_2^2$ is risk-free and strictly positive, and thus $\boldsymbol{\mu}_{t,0}$ is an arbitrage, unless $\boldsymbol{\mu}_{t,0} = \mathbf{0}$.

This is equivalent to the existence of the *mean-variance efficient (MVE)* portfolio maximizing the objective $\boldsymbol{w}_t^\top \boldsymbol{\mu}_t - \frac{1}{2}\boldsymbol{w}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{w}_t$ over $\boldsymbol{w}_t \in \mathbb{R}^{n_t}$.[8] Its weights are given by $\boldsymbol{w}_t = \boldsymbol{\Sigma}_t^+ \boldsymbol{\mu}_t$, and it attains the maximum squared Sharpe ratio, $\mathrm{SR}_t^2 = \boldsymbol{\mu}_t^\top \boldsymbol{\Sigma}_t^+ \boldsymbol{\mu}_t$.

Absence of arbitrage (20) is also equivalent to the existence of the *minimum-variance stochastic discount factor (SDF)*, which solves the problem

$$\begin{aligned} &\underset{M_{t+1} \in L^2_{\mathbb{P}_t}}{\text{minimize}}\ \mathbb{E}_t[M_{t+1}^2] \\ &\text{subject to } \mathbb{E}_t[M_{t+1}\boldsymbol{x}_{t+1}] = \boldsymbol{0} \text{ and } \mathbb{E}_t[M_{t+1}] = 1, \end{aligned}$$

and is given in terms of the MVE portfolio by $M_{t+1} = 1 - \boldsymbol{\mu}_t^\top \boldsymbol{\Sigma}_t^+ (\boldsymbol{x}_{t+1} - \boldsymbol{\mu}_t)$.

We henceforth assume that factors are tradable, and consider m factor portfolios $\boldsymbol{f}_{t+1}$ with returns given by (2). We denote by $\boldsymbol{\mu}_{\boldsymbol{f},t} \coloneqq \mathbb{E}_t[\boldsymbol{f}_{t+1}] = \boldsymbol{W}_t^\top \boldsymbol{\mu}_t$ their mean and, as above, by $\boldsymbol{\Sigma}_{\boldsymbol{f},t} = \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t$ their covariance. Then the MVE portfolio spanned by $\boldsymbol{f}_{t+1}$ with weights given by $\boldsymbol{\Sigma}_{\boldsymbol{f},t}^+ \boldsymbol{\mu}_{\boldsymbol{f},t}$ is equally well defined and yields the squared Sharpe ratio $\mathrm{SR}_{\boldsymbol{f},t}^2 = \boldsymbol{\mu}_{\boldsymbol{f},t}^\top \boldsymbol{\Sigma}_{\boldsymbol{f},t}^+ \boldsymbol{\mu}_{\boldsymbol{f},t}$.[9] There exists also a minimum-variance SDF that prices only factors.

In general, we have $\mathrm{SR}_{\boldsymbol{f},t}^2 \leq \mathrm{SR}_t^2$. The following proposition gives sufficient and necessary conditions for equality and answers the introductory question (ii). It generalizes Kozak and Nagel (2024, Lemma 1) to the case where the matrices $\boldsymbol{\Sigma}_t$, $\boldsymbol{\Sigma}_{\boldsymbol{f},t}$ and $\boldsymbol{\Phi}_t$ do not have full rank.[10]

Proposition 5.1. *The following are equivalent:*

(i) *the factor MVE portfolio attains the maximum Sharpe ratio,*

$$\mathrm{SR}_{\boldsymbol{f},t}^2 = \mathrm{SR}_t^2; \tag{21}$$

(ii) *asset risk premia $\boldsymbol{\mu}_t$ are given as covariances with the factors,* $\mathrm{Cov}_t[\boldsymbol{x}_{t+1}, \boldsymbol{f}_{t+1}] = \boldsymbol{\Sigma}_t \boldsymbol{W}_t$, *as*

$$\boldsymbol{\mu}_t \in \mathrm{Im}(\boldsymbol{\Sigma}_t \boldsymbol{W}_t); \tag{22}$$

[8] The MVE portfolio is unique up to scaling by the risk aversion parameter, which cancels out in the Sharpe ratio and which we therefore set equal to one.

[9] In fact, we have $\boldsymbol{\mu}_{\boldsymbol{f},t} \in \mathrm{Im}\, \boldsymbol{\Sigma}_{\boldsymbol{f},t}$, which follows from (20) and using that $\mathrm{Im}\, \boldsymbol{A} = \mathrm{Im}\, \boldsymbol{A}\boldsymbol{A}^\top$ for any matrix $\boldsymbol{A}$.

[10] Kozak and Nagel (2024) derive the proof of their Lemma 1 from (Lu and Schmidt, 2012, Theorem 3 A, B), which assumes that matrices are invertible.

(iii) *factors* $\boldsymbol{f}_{t+1}$ *span the full MVE portfolio,*

$$\boldsymbol{\mu}_{\boldsymbol{f},t}^\top \boldsymbol{\Sigma}_{\boldsymbol{f},t}^+ \boldsymbol{f}_{t+1} = \boldsymbol{\mu}_t^\top \boldsymbol{\Sigma}_t^+ \boldsymbol{x}_{t+1}; \tag{23}$$

(iv) *factors* $\boldsymbol{f}_{t+1}$ *span the minimum-variance SDF,* $M_{t+1} = 1 - \boldsymbol{\mu}_{\boldsymbol{f},t}^\top \boldsymbol{\Sigma}_{\boldsymbol{f},t}^+ (\boldsymbol{f}_{t+1} - \boldsymbol{\mu}_{\boldsymbol{f},t})$.

For a large class of tradable factors, and under the assumption that factors and residuals are uncorrelated, the spanning condition (22) holds if and only if the residual risk is unpriced. This is the content of the following proposition.

Proposition 5.2. *Assume that any of* (3), (4) *or* (7) *holds. Then any of* (17) *or* (18) *implies the spanning condition* (22)*. Conversely, if the weight matrix* $\boldsymbol{W}_t$ *satisfies* (14)*, then the spanning condition* (22) *implies any of* (17) *or* (18)*.*

From Proposition 5.2 we deduce that any of (3), (4) or (7) and any of (17) or (18) together imply the spanning condition (22). However, the converse is not true, as the following continuation of Example 3 shows.[11]

Example 3 (continued)**.** *Assume additionally* $b_1 = b_2$ *and* $\rho \neq 0$*. Then* (18) *follows by inspection. Moreover, some algebra shows that*

$$\boldsymbol{\Sigma}_t \boldsymbol{W}_t \begin{bmatrix} 0 \\ \frac{b_1}{\rho} \end{bmatrix} = \boldsymbol{\Sigma}_t (\boldsymbol{\Phi}_t^+)^\top \begin{bmatrix} 0 \\ \frac{b_1}{\rho} \end{bmatrix} = \boldsymbol{\mu}_t,$$

if $b_3 = (1-\rho)\frac{b_1}{a_1} + (1-\rho)\frac{b_1}{a_2} + \frac{b_1}{\rho} a_3$*, which shows* (22)*. Hence* (17), (18) *and the spanning condition* (22) *hold, while* (3), (4) *and* (7) *do not.*

The following lemma puts Example 3 into a larger perspective and shows what goes wrong with the failed implication.[12]

[11] Proposition 2 in Kozak and Nagel (2024) is similar to our Proposition 5.2, but differs in some important points. First, their proposition assumes that the weight matrix is of the form $\boldsymbol{W}_t = \boldsymbol{\Phi}_t \boldsymbol{R}_t$, for some invertible $m \times m$ matrix $\boldsymbol{R}_t$. However, this excludes GLS factors (Example 2) unless they are OLS factors (Example 1) and $\ker \boldsymbol{\Phi}_t = \{\boldsymbol{0}\}$ holds, in which case $\boldsymbol{R}_t = (\boldsymbol{\Phi}_t^\top \boldsymbol{\Phi}_t)^{-1}$. Second, they replace the spanning condition (22) by the stronger assumption that $\operatorname{Im} \boldsymbol{\Phi}_t \subseteq \operatorname{Im}(\boldsymbol{\Sigma}_t \boldsymbol{W}_t)$ for the converse implication in their proposition. Third, they replace our decomposition (4) by the weaker property that $\boldsymbol{\Sigma}_t = \boldsymbol{\Phi}_t \boldsymbol{\Psi}_t \boldsymbol{\Phi}_t^\top + \boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t^\top$ for some conformable matrices $\boldsymbol{\Psi}_t$, $\boldsymbol{\Omega}_t$, and a $n_t \times (n_t - m)$ matrix $\boldsymbol{U}_t$ for which $\boldsymbol{U}_t^\top \boldsymbol{\Phi}_t = \boldsymbol{0}$. The latter holds automatically in our case, see (9). Hence their proposition makes no inference on the correlation between factors and residuals as we do.

[12] In fact, in Example 3 we have $\boldsymbol{b}_t = \boldsymbol{W}_t \boldsymbol{c}_t$ for $\boldsymbol{c}_t = \begin{bmatrix} 0 \\ \frac{b_1}{\rho} \end{bmatrix}$.

Lemma 5.3. *Assume that the weight matrix $\boldsymbol{W}_t$ satisfies either (14) or $\ker \boldsymbol{\Phi}_t \subseteq \ker \boldsymbol{W}_t$. Then any of (17) or (18) and the spanning condition (22) imply that there exists a vector $\boldsymbol{b}_t \in \mathbb{R}^{n_t}$ such that the vector equality $\boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top \boldsymbol{b}_t = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top \boldsymbol{b}_t$ holds. However, not the matrix equality (7), in general.*

6 Generative models have spanning factors

In the following we answer the introductory question (iii). Concretely, we establish sufficient conditions on a data-generating linear factor model (1), which we write here as

$$\boldsymbol{x}_{t+1} = \boldsymbol{\Phi}_t \boldsymbol{g}_{t+1} + \boldsymbol{z}_{t+1},$$

such that GLS-type factors are spanning. Here $\boldsymbol{g}_{t+1} \in \mathbb{R}^m$ denote some abstract (generally non-tradable) systematic risk factors and $\boldsymbol{z}_{t+1}$ the idiosyncratic risk components of the returns $\boldsymbol{x}_{t+1}$. We denote their mean and covariance matrices by $\boldsymbol{\mu}_{\boldsymbol{g},t} \coloneqq \mathbb{E}_t[\boldsymbol{g}_{t+1}]$, $\boldsymbol{\Sigma}_{\boldsymbol{g},t} \coloneqq \mathrm{Cov}_t[\boldsymbol{g}_{t+1}]$, and $\boldsymbol{\Sigma}_{\boldsymbol{z},t} \coloneqq \mathrm{Cov}_t[\boldsymbol{z}_{t+1}]$. We first prove two technical lemmas.

Lemma 6.1. *Assume that systematic risk factors $\boldsymbol{g}_{t+1}$ and idiosyncratic risk $\boldsymbol{z}_{t+1}$ are uncorrelated, $\mathrm{Cov}_t[\boldsymbol{g}_{t+1}, \boldsymbol{z}_{t+1}] = \boldsymbol{0}$, covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{g},t}$ is invertible, and idiosyncratic risk is unpriced, $\mathbb{E}_t[\boldsymbol{z}_{t+1}] = \boldsymbol{0}$. Then the implied return covariance matrix decomposes as*

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{g},t} \boldsymbol{\Phi}_t^\top + \boldsymbol{\Sigma}_{\boldsymbol{z},t}, \tag{24}$$

and absence of arbitrage (20) holds.

The following auxiliary lemma gives a matrix extension in the case where the idiosyncratic covariance matrix does not have full rank.

Lemma 6.2. *Let $\boldsymbol{U}_t$ be a $n_t \times n_t$-matrix such that $\boldsymbol{U}_t^\top \boldsymbol{U}_t = \boldsymbol{\Sigma}_{\boldsymbol{z},t}^+$. Then $\mathrm{Im}\, \boldsymbol{\Sigma}_{\boldsymbol{z},t} = \mathrm{Im}\, \boldsymbol{U}_t^\top$ and $\ker \boldsymbol{\Sigma}_{\boldsymbol{z},t} = \ker \boldsymbol{U}_t$. Moreover, there exists an invertible $n_t \times n_t$-matrix $\boldsymbol{S}_t$ extending $\boldsymbol{U}_t$ in the sense that $\boldsymbol{S}_t = \boldsymbol{U}_t$ on $\mathrm{Im}\, \boldsymbol{U}_t^\top$ and $\boldsymbol{S}_t^\top = \boldsymbol{U}_t^\top$ on $\mathrm{Im}\, \boldsymbol{U}_t$, and $\boldsymbol{S}_t : \ker \boldsymbol{U}_t \to \ker \boldsymbol{U}_t^\top$ is a bijection. In particular, $\boldsymbol{S}_t = \boldsymbol{U}_t$ if $\boldsymbol{\Sigma}_{\boldsymbol{z},t}$ is invertible, and a possible choice is $\boldsymbol{S}_t = \boldsymbol{I}_{n_t}$ in the degenerate case where $\boldsymbol{\Sigma}_{\boldsymbol{z},t} = \boldsymbol{0}$.*

We can now prove our announced result.

Proposition 6.3. *Under the assumptions of Lemmas 6.1 and 6.2, the GLS-type factors*

$$\boldsymbol{f}_{t+1} \coloneqq (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{S}_t \boldsymbol{x}_{t+1} \tag{25}$$

satisfy spanning condition (22). *The implied residuals* $\boldsymbol{\epsilon}_{t+1} \coloneqq \boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}$ *have zero mean* (17) *and are uncorrelated with the factors* (6). *The implied matrix*

$$\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{\Phi}_t (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{S}_t \text{ is a (generally non-orthogonal) projection onto } \operatorname{Im} \boldsymbol{\Phi}_t, \quad (26)$$

and the residuals are equal to the projection of idiosyncratic risk,

$$\boldsymbol{\epsilon}_{t+1} = (\boldsymbol{I}_{n_t} - \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) \boldsymbol{z}_{t+1}. \quad (27)$$

Mean and covariance matrices of $\boldsymbol{f}_{t+1}$ *and* $\boldsymbol{\epsilon}_{t+1}$ *are given by the expressions*

$$\boldsymbol{\mu}_{\boldsymbol{f},t} = (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{\mu}_{\boldsymbol{g},t}, \quad (28)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{f},t} = (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{g},t} \boldsymbol{\Phi}_t^\top \boldsymbol{S}_t^\top (\boldsymbol{\Phi}_t^\top \boldsymbol{S}_t^\top)^+ + \boldsymbol{Q}_t, \quad (29)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} = \boldsymbol{\Sigma}_{\boldsymbol{z},t} - \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{z},t} \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top = \boldsymbol{\Sigma}_{\boldsymbol{z},t} - \boldsymbol{\Phi}_t \boldsymbol{Q}_t \boldsymbol{\Phi}_t^\top, \quad (30)$$

where $\boldsymbol{Q}_t \coloneqq (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{S}_t \boldsymbol{\Sigma}_{\boldsymbol{z},t} \boldsymbol{S}_t^\top (\boldsymbol{\Phi}_t^\top \boldsymbol{S}_t^\top)^+ = (\boldsymbol{S}_t \boldsymbol{\Phi}_t)^+ \boldsymbol{U}_t \boldsymbol{U}_t^+ (\boldsymbol{\Phi}_t^\top \boldsymbol{S}_t^\top)^+$.

Note that $\boldsymbol{S}_t \boldsymbol{S}_t^\top \neq \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t}$ in general in the above statement. Hence, the factors $\boldsymbol{f}_{t+1}$ in Proposition 6.3 are only of "GLS-type", rather than the GLS factors as described in Example 2. Nevertheless, Proposition 6.3 remarkably promises that any return panel linearly generated by abstract risk factors, which are not necessarily tradable, also has a linear representation in terms of tradable factors $\boldsymbol{f}_{t+1}$ that load on the same characteristics $\boldsymbol{\Phi}_t$. However, the same statement also reveals that the properties of these tradable factors crucially rely on $\boldsymbol{\Phi}_t$, as well as on $\boldsymbol{\Sigma}_{g,t}$ and $\boldsymbol{\Sigma}_{z,t}$ in their population formulation. Any empirical application thus necessitates consistent estimators of $\boldsymbol{\Phi}_t$ and $\boldsymbol{\Sigma}_t$, or, alternatively $\boldsymbol{\Sigma}_{\boldsymbol{g},t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{z},t}$.

7 Conclusion

This short note provides a collection of elementary and fundamental properties of conditional linear factor models for unbalanced panels, in particular in the context of asset pricing. Our results are derived for a finite cross section, they are comprehensive, and based upon elementary linear algebra.

We exhaustively describe the relation between the covariance matrix of returns, the covariance matrix of the residuals, risk premia, and Sharpe ratios. Our results

range from the question when a covariance matrix of a panel generated by a linear factor structure can be decomposed into a factor, and a residual part, to the question when the maximum Sharpe ratio attained by the full return panel can also be attained a linear factor portfolio. If the factors are themselves portfolios of the returns, a number of additional powerful properties and qualifications arise that are useful in a portfolio context. We show that any return panel whose data-generating process is linear in abstract, non-tradable factors, has a linear factor representation in terms of tradable factors that load on the same characteristics. The same result also shows that consistent estimators for factor, residual, and return covariances are indispensable for econometric factor analysis.

Our analysis provides a simple framework together with simple guidelines for the estimation and specification of linear factor models, comprising also modern formulations based upon neural nets or kernels. Future work could tackle the connection of the results in this short note, to the asymptotic notions introduced in Chamberlain (1983) and Chamberlain and Rothschild (1983).

A Proofs

The appendix collects all proofs.

A.1 Proof of Lemma 3.1

Equivalence of (3) and (4) follows from the elementary decomposition

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{f},t} \boldsymbol{\Phi}_t^\top + \mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] + \mathrm{Cov}_t[\boldsymbol{\epsilon}_{t+1}, \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}] + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t}.$$

Assuming (5), equivalence of (3) and (6) follows because $\mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \boldsymbol{\Phi}_t \mathrm{Cov}_t[\boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}]$.

A.2 Proof of Proposition 3.2

Given (2), it follows that

$$\mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}] = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t - \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top.$$

Hence (3) is equivalent to (7). Assuming (8), equivalence of (3) and (6) follows because $\mathrm{Cov}_t[\boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \boldsymbol{\Phi}_t \mathrm{Cov}_t[\boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}]$ and $\mathrm{Cov}_t[\boldsymbol{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}] = \boldsymbol{W}_t^\top \mathrm{Cov}_t[\boldsymbol{x}_{t+1}, \boldsymbol{\epsilon}_{t+1}]$.

A.3 Proof of Proposition 3.3

Given (2), we have

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} = \mathrm{Cov}_t[\boldsymbol{x}_{t+1} - \boldsymbol{\Phi}_t \boldsymbol{f}_{t+1}] = \boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top - \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t + \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top.$$

Hence (7) implies $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} = \boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top$ and therefore $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t - \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top = \boldsymbol{0}$, which yields (9). That the converse implication is not true is proved by means of the counterexample given in Example 3.

A.4 Proof of Corollary 3.4

This follows from (9).

A.5 Proof of Lemma 3.5

Given (11), the residuals $\boldsymbol{\epsilon}_{t+1} = (\boldsymbol{I}_{n_t} - \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) \boldsymbol{x}_{t+1}$ satisfy $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\epsilon}_{t+1} = \boldsymbol{0}$, and therefore (9). The second statement follows by elementary linear algebra. The last statement follows by means of the counterexample given in Example 3.

A.6 Proof of Lemma 3.6

Property (14) clearly implies (11). Now let $\boldsymbol{v} \in \operatorname{Im} \boldsymbol{W}_t^\top \cap \ker \boldsymbol{\Phi}_t$. This means that $\boldsymbol{v} = \boldsymbol{W}_t^\top \boldsymbol{c}$ for some $\boldsymbol{c} \in \mathbb{R}^{n_t}$ such that $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{c} = \boldsymbol{0}$. Given (14), this implies that $\boldsymbol{v} = \boldsymbol{W}_t^\top \boldsymbol{c} = \boldsymbol{W}_t^\top \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{c} = \boldsymbol{0}$, which proves (8). This also implies that $\ker(\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) = \ker \boldsymbol{W}_t^\top$.[13] Conversely, (11) reads $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top$, and given (8) this again implies (14), which proves the first part of the lemma.

For the second part of the lemma, given (15), matrix algebra shows that $\boldsymbol{W}_t^\top \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top = \boldsymbol{R}_t (\boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t)^+ \boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t (\boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t)^+ \boldsymbol{S}_t = \boldsymbol{R}_t (\boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t)^+ \boldsymbol{S}_t = \boldsymbol{W}_t^\top$, where we used that $\boldsymbol{A}^+ \boldsymbol{A} \boldsymbol{A}^+ = \boldsymbol{A}^+$ for any matrix $\boldsymbol{A}$, which implies (14).

[13] By elementary linear algebra, this again implies that the adjoint $\boldsymbol{W}_t \boldsymbol{\Phi}_t^\top$ is a projection with $\operatorname{Im}(\boldsymbol{W}_t \boldsymbol{\Phi}_t^\top) = \operatorname{Im} \boldsymbol{W}_t^\top$. On the other hand, we only have $\operatorname{Im}(\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) \subseteq \operatorname{Im} \boldsymbol{\Phi}_t$, without equality in general.

For the last statement, we first note that $\mathrm{Im}(\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) \subseteq \mathrm{Im}(\boldsymbol{\Phi}_t \boldsymbol{R}_t)$ and consider the identity $\boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top (\boldsymbol{\Phi}_t \boldsymbol{R}_t) = \boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t (\boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t)^+ \boldsymbol{S}_t \boldsymbol{\Phi}_t \boldsymbol{R}_t = \boldsymbol{S}_t (\boldsymbol{\Phi}_t \boldsymbol{R}_t)$. Given (16), we infer that $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top (\boldsymbol{\Phi}_t \boldsymbol{R}_t) = \boldsymbol{\Phi}_t \boldsymbol{R}_t$, and therefore $\mathrm{Im}(\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top) = \mathrm{Im}(\boldsymbol{\Phi}_t \boldsymbol{R}_t)$.

A.7 Proof of Lemma 4.1

This follows from the identity $\boldsymbol{x}_{t+1} = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{x}_{t+1} + \boldsymbol{\epsilon}_{t+1}$.

A.8 Proof of Proposition 4.2

The first claim follows from the $L^2_{\mathbb{P}_t}$-orthogonality of the factors and residuals, $\mathbb{E}_t[\boldsymbol{f}_{t+1} \boldsymbol{\epsilon}_{t+1}^\top] = \boldsymbol{0}$, which holds due to (17) and (6). The second claim about the uniqueness of the minimizer follows elementary.

A.9 Proof of Proposition 5.1

(i)⇔(ii): Given (20), we can write $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\nu}_t$, and thus $\boldsymbol{\mu}_{\boldsymbol{f},t} = \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\nu}_t$, for some $\boldsymbol{\nu}_t \in \mathbb{R}^{n_t}$. We obtain $\mathrm{SR}_t^2 = \boldsymbol{\nu}_t^\top \boldsymbol{\nu}_t$ and, using the matrix identity $\boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^+ = \boldsymbol{A}^+$ for any matrix $\boldsymbol{A}$,

$$\begin{aligned}\mathrm{SR}_{\boldsymbol{f},t}^2 &= \boldsymbol{\nu}_t^\top (\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2})^\top \big((\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2}) (\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2})^\top \big)^+ (\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2}) \boldsymbol{\nu}_t \\ &= \boldsymbol{\nu}_t^\top (\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2})^+ (\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2}) \boldsymbol{\nu}_t = \boldsymbol{\nu}_t^\top \boldsymbol{P}_{\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{W}_t} \boldsymbol{\nu}_t,\end{aligned}$$

where $\boldsymbol{P}_{\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{W}_t}$ denotes the orthogonal projection onto the image of $\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{W}_t$. It follows that (21) holds if and only if $\boldsymbol{\nu}_t = \boldsymbol{P}_{\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{W}_t} \boldsymbol{\nu}_t$ lies in the image of $\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{W}_t$, which again is equivalent to (22).

(ii)⇒(iii): Given (22), and because $\ker(\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{W}_t) \subseteq \ker(\boldsymbol{\Sigma}_t \boldsymbol{W}_t)$, there exists a vector $\boldsymbol{b}_t \in \mathrm{Im}(\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t^{1/2})$ such that $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{b}_t$ and therefore $\boldsymbol{\Sigma}_{\boldsymbol{f},t}^+ \boldsymbol{\mu}_{\boldsymbol{f},t} = (\boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t)^+ \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{b}_t = \boldsymbol{b}_t$. On the other hand, we have $\boldsymbol{x}_{t+1} = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t^+ \boldsymbol{x}_{t+1}$ with probability one. We obtain $\boldsymbol{\mu}_{\boldsymbol{f},t}^\top \boldsymbol{\Sigma}_{\boldsymbol{f},t}^+ \boldsymbol{f}_{t+1} = \boldsymbol{b}_t^\top \boldsymbol{W}_t^\top \boldsymbol{x}_{t+1} = \boldsymbol{b}_t^\top \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t^+ \boldsymbol{x}_{t+1} = \boldsymbol{\mu}_t^\top \boldsymbol{\Sigma}_t^+ \boldsymbol{x}_{t+1}$, with probability one, which is (23).

(iii)⇒(i): This direction is trivial.

(iii)⇔(iv): This follows from (23) and the aforementioned expression of M_t in terms of the full MVE portfolio.

A.10 Proof of Proposition 5.2

Using (7) and (20), property (18) implies $\boldsymbol{\mu}_t = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\mu}_t = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{b}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top \boldsymbol{b}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{c}_t$, for $\boldsymbol{c}_t \coloneqq \boldsymbol{\Phi}_t^\top \boldsymbol{b}_t$, for some $\boldsymbol{b}_t \in \mathbb{R}^{n_t}$, which proves (18) $\Rightarrow$ (22).

Conversely, using (7) and (14), the spanning condition (22) implies $\boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\mu}_t = \boldsymbol{\Phi}_t \boldsymbol{W}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{c}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top \boldsymbol{W}_t \boldsymbol{c}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{c}_t = \boldsymbol{\mu}_t$, for some $\boldsymbol{c}_t \in \mathbb{R}^m$, which proves (22) $\Rightarrow$ (18).

A.11 Proof of Lemma 5.3

Under either assumption, we have that (22) implies that there exists a vector $\boldsymbol{b}_t \in \mathbb{R}^{n_t}$ such that $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{\Phi}_t^\top \boldsymbol{b}_t$. The claimed vector equality now follows from (18). The last statement follows by means of the counterexample given in Example 3.

A.12 Proof of Lemma 6.1

The decomposition (24) follows from Lemma 3.1. This clearly implies that $\operatorname{Im} \boldsymbol{\Sigma}_t = \operatorname{Im}(\boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{g},t} \boldsymbol{\Phi}_t^\top) + \operatorname{Im} \boldsymbol{\Sigma}_{\boldsymbol{z},t}$, and as $\boldsymbol{\mu}_t = \boldsymbol{\Phi}_t \boldsymbol{\mu}_{\boldsymbol{g},t} \in \operatorname{Im} \boldsymbol{\Phi}_t = \operatorname{Im}(\boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{\boldsymbol{g},t} \boldsymbol{\Phi}_t^\top)$ by assumption, absence of arbitrage (20) follows.

A.13 Proof of Lemma 6.2

We have $\operatorname{Im} \boldsymbol{\Sigma}_{\boldsymbol{z},t} = \operatorname{Im}(\boldsymbol{U}_t^+ (\boldsymbol{U}_t^+)^\top) = \operatorname{Im} \boldsymbol{U}_t^+ = \operatorname{Im} \boldsymbol{U}_t^\top$, and $\ker \boldsymbol{\Sigma}_{\boldsymbol{z},t} = \ker \boldsymbol{U}_t$, where we used that $(\boldsymbol{A}^\top \boldsymbol{A})^+ = \boldsymbol{A}^+ (\boldsymbol{A}^+)^\top$ for any matrix $\boldsymbol{A}$. This proves the first statement.

We now construct $\boldsymbol{S}_t$. By elementary linear algebra, we have $\mathbb{R}^{n_t} = \ker \boldsymbol{U}_t \oplus \operatorname{Im} \boldsymbol{U}_t^\top = \ker \boldsymbol{U}_t^\top \oplus \operatorname{Im} \boldsymbol{U}_t$, and $\boldsymbol{U}_t : \operatorname{Im} \boldsymbol{U}_t^\top \to \operatorname{Im} \boldsymbol{U}_t$ is a bijection, and $\dim \ker \boldsymbol{U}_t = \dim \ker \boldsymbol{U}_t^\top =: k \geq 0$. Let $\{\xi_1, \ldots, \xi_k\}$ and $\{\zeta_1, \ldots, \zeta_k\}$ be any orthonormal bases of $\ker \boldsymbol{U}_t$ and $\ker \boldsymbol{U}_t^\top$, respectively. Setting $\boldsymbol{S}_t \coloneqq \boldsymbol{U}_t$ on $\operatorname{Im} \boldsymbol{U}_t^\top$ and $\boldsymbol{S}_t \xi_i \coloneqq \zeta_i$ on $\ker \boldsymbol{U}_t$ gives the desired extension.

A.14 Proof of Proposition 6.3

Lemma 6.2 implies that $\boldsymbol{S}_t \boldsymbol{\Sigma}_{\boldsymbol{z},t} \boldsymbol{S}_t^\top = \boldsymbol{U}_t (\boldsymbol{U}_t^\top \boldsymbol{U}_t)^+ \boldsymbol{U}_t^\top = \boldsymbol{U}_t \boldsymbol{U}_t^+$ is the orthogonal projection on $\operatorname{Im} \boldsymbol{U}_t$, where we used that $(\boldsymbol{A}^\top \boldsymbol{A})^+ \boldsymbol{A}^\top = \boldsymbol{A}^+$ for any matrix $\boldsymbol{A}$. We obtain $(\boldsymbol{S}_t \boldsymbol{\Sigma}_{\boldsymbol{z},t} \boldsymbol{S}_t^\top) \boldsymbol{S}_t = \boldsymbol{U}_t \boldsymbol{U}_t^+ \boldsymbol{U}_t = \boldsymbol{U}_t$. Given that $\operatorname{Im}(\boldsymbol{U}_t \boldsymbol{\Phi}) \subseteq \operatorname{Im}(\boldsymbol{S}_t \boldsymbol{\Phi})$, by definition of

$\boldsymbol{S}_t$, we thus have proved that

$$\boldsymbol{S}_t\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{S}_t^\top = \boldsymbol{U}_t\boldsymbol{U}_t^+ \text{ maps } \mathrm{Im}(\boldsymbol{S}_t\boldsymbol{\Phi}_t) \text{ into itself.} \tag{31}$$

As $\boldsymbol{S}_t$ is invertible, condition (16) for $\boldsymbol{R}_t = \boldsymbol{I}_m$ holds, and Lemma 3.6 implies property (26). This again implies (27) and (18) and, by Lemma 4.1, thus (17).

By the same token and given (26), condition (7) is equivalent to $\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top = \boldsymbol{\Phi}_t\boldsymbol{W}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top$, which again is equivalent to

$$\boldsymbol{S}_t\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top = \boldsymbol{S}_t\boldsymbol{\Phi}_t\boldsymbol{W}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top. \tag{32}$$

Plugging in (26) for $\boldsymbol{\Phi}_t\boldsymbol{W}_t^\top$, the right hand side of (32) can be written as $\boldsymbol{S}_t\boldsymbol{\Phi}_t(\boldsymbol{S}_t\boldsymbol{\Phi}_t)^+\boldsymbol{S}_t\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{S}_t^\top(\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+\boldsymbol{\Phi}_t^\top$. Given (31), and as $\mathrm{Im}((\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+\boldsymbol{\Phi}_t^\top) \subseteq \mathrm{Im}((\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+) = \mathrm{Im}(\boldsymbol{S}_t\boldsymbol{\Phi}_t)$ and $\boldsymbol{S}_t\boldsymbol{\Phi}_t(\boldsymbol{S}_t\boldsymbol{\Phi}_t)^+$ is the orthogonal projection on $\mathrm{Im}(\boldsymbol{S}_t\boldsymbol{\Phi}_t)$, we infer that the right hand side of (32) is equal to $\boldsymbol{S}_t\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{S}_t^\top(\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+\boldsymbol{\Phi}_t^\top$. This again is equal to the left hand side of (32), which implies (7). Proposition 3.2 and Lemma 3.6 prove (6).

Proposition 5.2 now implies that the spanning condition (22) holds for $\boldsymbol{f}_{t+1}$.

Expression (28) and the first summand in expression (29) follow directly from (25). For the second summand in (29), we have the identity $(\boldsymbol{S}_t\boldsymbol{\Phi}_t)^+\boldsymbol{S}_t\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{S}_t^\top(\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+ = (\boldsymbol{S}_t\boldsymbol{\Phi}_t)^+\boldsymbol{U}_t\boldsymbol{U}_t^+(\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+$, using (31). For (30), as in the proof of Proposition 3.3, we first derive that (7) implies $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},t} = \boldsymbol{\Sigma}_t - \boldsymbol{\Phi}_t\boldsymbol{W}_t^\top\boldsymbol{\Sigma}_t\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top = \boldsymbol{\Sigma}_{\boldsymbol{z},t} - \boldsymbol{\Phi}_t\boldsymbol{W}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top$, where we used (26). By similar calculations as above, we derive $\boldsymbol{\Phi}_t\boldsymbol{W}_t^\top\boldsymbol{\Sigma}_{\boldsymbol{z},t}\boldsymbol{W}_t\boldsymbol{\Phi}_t^\top = \boldsymbol{\Phi}_t(\boldsymbol{S}_t\boldsymbol{\Phi}_t)^+\boldsymbol{U}_t\boldsymbol{U}_t^+(\boldsymbol{\Phi}_t^\top\boldsymbol{S}_t^\top)^+\boldsymbol{\Phi}_t^\top$, which completes the proof.

References

Admati, A. R. and Pfleiderer, P. (1985). Interpreting the factor risk premia in the arbitrage pricing theory. *Journal of Economic Theory*, 35(1):191–195.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5):1305–1323.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Fama, E. F. and French, K. R. (2020). Comparing cross-section and time-series factor models. *Review of Financial Studies*, 33(5):1891–1926.

Fortin, A.-P., Gagliardini, P., and Scaillet, O. (2023). Latent factor analysis in short panels. working paper, University of Geneva, and Universitá delle Svizzera italiana.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting Characteristics Nonparametrically. *Review of Financial Studies*, 33(5):2326–2377.

Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Horn, R. A. and Johnson, C. R. (1990). *Matrix analysis*. Cambridge University Press, Cambridge. Corrected reprint of the 1985 original.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Kozak, S. and Nagel, S. (2024). When do cross-sectional asset pricing factors span the stochastic discount factor? Working Paper 31275, National Bureau of Economic Research.

Ledoit, O. and Wolf, M. (2020). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. *Journal of Financial Econometrics*, 20(1):187–218.

Lu, C. and Schmidt, P. (2012). Conditions for the numerical equality of the OLS, GLS and Amemiya–Cragg estimators. *Economics Letters*, 116(3):538–540.

Onatski, A. (2015). Asymptotic analysis of the squared estimation error in misspecified factor models. *Journal of Econometrics*, 186(2):388–406. High Dimensional Problems in Econometrics.

Raponi, V., Robotti, C., and Zaffaroni, P. (2019). Testing Beta-Pricing Models Using Large Cross-Sections. *Review of Financial Studies*, 33(6):2796–2842.

Reisman, H. (1988). A general approach to the arbitrage pricing theory (apt). *Econometrica*, 56(2):473–476.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Schott, J. R. (2017). *Matrix analysis for statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, third edition.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk*. *Journal of Finance*, 19(3):425–442.

Zaffaroni, P. (2019). Factor models for asset pricing. working paper, Imperial College.