A Dynamic Factor Model of Price Impacts*

Yu An† Zeyu Zheng‡

This version: November 27, 2023
First version: November 9, 2022

Abstract

We develop a dynamic model showing how the predictability of noise trading flows impacts return predictability at both individual asset and factor levels. Our model first aggregates flow predictability from individual assets to factors. Unlike static models, different factors can have differential price sensitivity to current flows, because flow-absorbing arbitrageurs anticipate differential future flows. This factor-level price impact generates momentum or reversal dynamics in factor price, which in turn determines asset-level return predictability. Additionally, when flows manifest excessive momentum, asset price bubbles can emerge, because providing liquidity under this scenario would require running a Ponzi scheme.

Keywords: dynamic, factor, flow, predictability, price impact

JEL Codes: G12

*For helpful comments and suggestions, we thank Darrell Duffie, Vadim Elenev, Jiacui Li, Alessandro Rebucci, Xiaofei Shi, Semih Üslü, Xingtan Zhang, and participants at JHU Carey finance brownbag, Renmin University of China, and Fudan University. We especially thank Yuhang Wu for his insights and help in solving stochastic differential equations with jumps. All errors are our own.

†Johns Hopkins Carey Business School; yua@jhu.edu.
‡University of California, Berkeley; zyzheng@berkeley.edu.

1 Introduction

Numerous studies document that uninformed noise trading flows significantly impact asset prices. The theory suggests that arbitrageurs, who have limited risk-bearing capacity, provide liquidity to these flows. As these flows change arbitrageurs' holdings in the risky assets, price impacts arise as arbitrageurs' compensation for bearing extra risks.

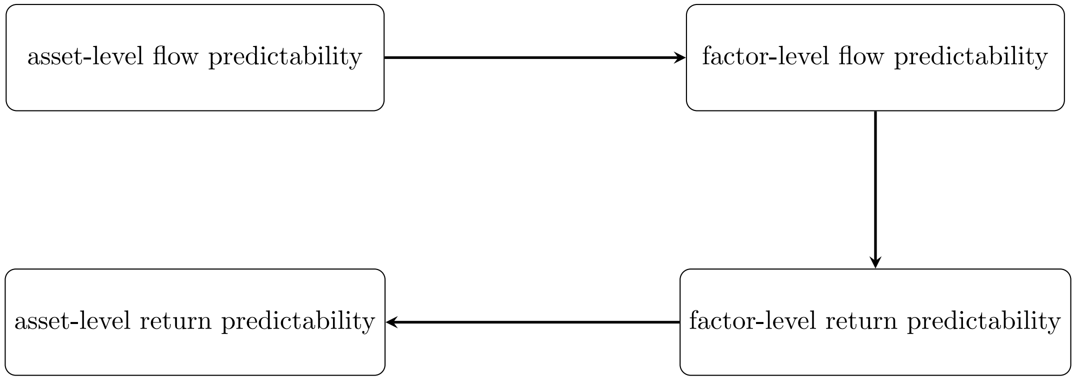
Empirically, a notable feature of noise trading flows is their predictability. For instance, if noise traders buy some asset today, they are likely to buy more of the asset tomorrow, a pattern exemplified during the GameStop frenzy.[1] This observation leads to the intuitive hypothesis that the predictability of noise trading flows, which impact asset prices, could partly explain the predictability in returns. This hypothesis receives empirical support from Coval and Stafford (2007) and Lou (2012). Additionally, the phenomenon of predictable flows and returns is not limited to individual assets but is also observed at the factor level, as documented by Kelly, Moskowitz, and Pruitt (2021), Ben-David, Li, Rossi, and Song (2022), Ehsani and Linnainmaa (2022), and Arnott, Kalesnik, and Linnainmaa (2023).

The empirical evidence shows a gap in the theoretical literature: what is the linkage between flow predictability and return predictability at both individual asset and factor levels? This paper answers this question by developing a dynamic factor model of price impacts, as illustrated in Figure 1. This model shows that flow predictability generates return predictability through factors, and provides a unifying framework to interpret various strands of empirical evidence. The model first aggregates flow predictability from individual assets to factors. Unlike static models, different factors can have differential price sensitivity to current flows, because flow-absorbing arbitrageurs anticipate differential future flows. This factor-level price impact generates momentum or reversal dynamics in factor price, which in turn determines asset-level return predictability. We now delve into the details of the model.

Our model examines noisy flows into N risky assets. A group of arbitrageurs accommo-

[1] A summary of the GameStop frenzy can be found at https://corpgov.law.harvard.edu/2022/04/08/gamestop-and-the-reemergence-of-the-retail-investor/.

Figure 1. A dynamic factor model of price impacts



Note: Flow predictability generates return predictability through factors.

dates these flows. Equilibrium asset prices are determined by market-clearing conditions and the arbitrageurs' optimal consumption and portfolio choices. Serving as a multi-asset generalization of Campbell and Kyle (1993) and Wang (1993), our model addresses a challenging issue in order to link individual assets to factors. Specifically, the N assets might experience correlated dividend shocks. Additionally, asset flows may be correlated, and flows into one asset might predict future flows into another. This results in three separate $N \times N$ matrices: for dividend covariance, flow covariance, and flow predictability. A dynamic portfolio problem becomes unsolvable with arbitrary matrices for all three. Prior studies oversimplify by assuming at least one matrix is diagonal, which implies no cross-asset relationships.[2] This is not only inconsistent with empirical evidence but also masks the key economic insight that flow predictability generates return predictability through factors.

We solve the modeling challenge by applying a mathematical result known as commuting matrices. In essence, two $N \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$ commute (i.e., $\mathbf{AB} = \mathbf{BA}$) if and only if they share the same set of eigenvectors, often referred to as "factors" in finance. Here, factors mean portfolios of the N assets. At this stage, we have not yet considered dimension reduction, which we shall examine later.

Applying commuting matrices, we develop a new restriction on the three matrices that

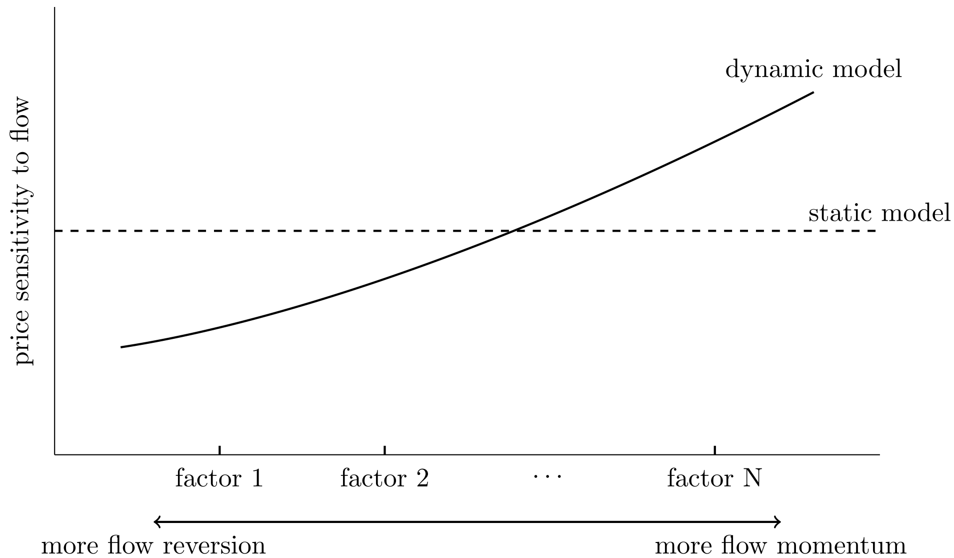
[2] See, for example, the survey article by Rostek and Yoon (2020) or the models in Bogousslavsky (2016) and Lu, Malliaris, and Qin (2023).

describe dividend covariance, flow covariance, and flow predictability. We refer to this restriction as *Factorizable Flow Predictability (FFP)*. The FFP is the necessary and sufficient condition such that one can use N factors to characterize the $N \times N$ price impacts between any pair of assets. These factors are specifically constructed such that they have uncorrelated dividends and uncorrelated flows, with future flows into any factor relying solely on its own past flows. Intuitively, while the N assets may exhibit complex interactions, the N factors are truly "independent" of each other, so arbitrageurs respond to each factor independently. Hence, studying the cross-section of price impacts via the N factors simplifies the theoretical analysis, while maintaining a realistic covariance structure for the N underlying assets.

The heart of our model is simple: flow predictability generates return predictability through factors. The FFP condition aggregates flow predictability from assets to factors. For each factor, as flow is predictable and impacts the price, factor-level return predictability emerges. In the cross-section, the law of one price implies that when factor prices shift, asset prices must adjust based on their risk exposure. Thus, cross-sectional arbitrage channels factor-level return predictability to individual assets.

The cross-sectional pricing implications of flow predictability in our model can be best comprehended through an analogy with the Merton (1973) ICAPM. Merton shows that future return predictors become new risk factors, thereby generalizing the static single-factor CAPM into a dynamic multi-factor model. In our model, flow predictability varies among factors. A factor with more flow momentum has a higher price sensitivity to current flow. The intuition is that if a flow has stronger momentum, arbitrageurs absorbing it today expect more tomorrow. This dynamic consideration implies that arbitrageurs demand greater price compensation for absorbing flow today. Thus, different factors can react differently to flows due to their differential flow predictability. This effect allows us to generalize the static mean-variance price impact model, where all factors must have the same price sensitivity, to a dynamic factor model, where different factors can exhibit differential price sensitivity. This improvement over the static model is illustrated in Figure 2. Empirically, using mutual

Figure 2. Improvement over static mean-variance price impact model



Note: In the static mean-variance price impact model, all factors must have the same price sensitivity to flow. In our dynamic model, a factor with more flow momentum has a higher price sensitivity.

fund flows, An, Su, and Wang (2023) indeed find significantly different price sensitivity for the Fama and French (1993) three factors.

From an empirical perspective, our theory views factors as having price momentum. The model shows that price momentum stems from factors rather than individual assets, aligning with recent empirical studies (Kelly, Moskowitz, and Pruitt, 2021; Ehsani and Linnainmaa, 2022; Arnott, Kalesnik, and Linnainmaa, 2023). It also supports the evidence that momentum in factor flow generates momentum in factor price (Ben-David, Li, Rossi, and Song, 2022). Therefore, our model differs from Merton's ICAPM, which views momentum as a separate risk factor (Jegadeesh and Titman, 1993; Carhart, 1997).

Moreover, we show that when flows manifest excessive momentum, asset price bubbles can emerge. This occurrence is linked to the momentum in flows, which results in momentum in price dislocation. Arbitrageurs absorbing these flows have to endure short-term mark-to-market losses. These losses are temporary, given that flows eventually revert in the long term. This reversion prompts the asset price to align with its fundamental value, allowing arbitrageurs to recoup all losses while also profiting from trading against the dislocated price. However, if the flow's momentum is too strong, providing liquidity to flow would require arbitrageurs to run a Ponzi scheme—the expected temporary loss becomes infinitely large. In this case, rational arbitrageurs refuse to provide liquidity and correct mispricing. Notably, there is a specific price sensitivity level where this scenario happens, meaning that the price cannot be overly responsive to flow in the dynamic equilibrium.

This Ponzi result explains Brunnermeier and Nagel (2004), who document that hedge funds ride the technology bubble rather than correcting mispricing. Abreu and Brunnermeier (2002, 2003) attribute this phenomenon to synchronized arbitrageur actions. Our model differs by showing that the flow momentum by itself produces a Ponzi region, where homogeneous competitive arbitrageurs refrain from absorbing flow and correcting mispricing.

So far, we have studied price impacts for the N assets using N factors. However, when N is large, selecting a smaller set of K factors is often necessary for an approximate model. Utilizing K instead of N factors raises a further question: assuming the FFP holds for the N underlying assets, does it necessarily apply to the K selected factors as well?

The answer is no. We prove that the FFP holds for the K factors if and only if these factors preserve the orthogonality structure of the underlying N-dimensional data. For instance, with $N = 3$ true factors generating the data and $K = 2$ selected factors, the FFP holds if the first selected factor aligns with the first true factor and the second is a combination of the remaining two. However, if the two selected factors overlap, the FFP generally fails. The empirical relevance of this result is that it is impractical to directly estimate the true model with N assets and then assess how good the K selected factors are. However, by testing whether the FFP holds for the K factors, econometricians can ascertain the effectiveness of the K selected factors.

The remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 presents the model setup. Section 4 presents the FFP condition. Section 5 solves the model. Section 6 reduces the number of factors. Section 7 concludes. The appendices provide proofs and additional theoretical results.

2 Related Literature

As discussed in the introduction, mounting empirical evidence has documented the predictability patterns of asset returns and trading flows at both individual asset and factor levels. However, the literature lacks a theory that links flow predictability and return predictability at both these levels. Our model fills this gap by showing that flow predictability generates return predictability through factors, offering a unified framework to interpret different empirical studies. The cross-sectional analysis sets our model apart from existing single-asset models on predictable trading flows (Campbell and Kyle, 1993; Wang, 1993; Bernhardt and Taub, 2008; Guo and Ou-Yang, 2015; Sadzik and Woolnough, 2021; Waters, Zhang, Zhou, and Santosh, 2022). Our new insight is that flow predictability causes different factors to have differential price sensitivity to flows. This insight enables us to generalize the static mean-variance price impact model to a dynamic factor model of price impacts.

Our FFP condition contributes to the literature on dynamic multi-asset price impact models, as seen in works like Bogousslavsky (2016) and Lu, Malliaris, and Qin (2023). Economically, the FFP reveals that price impacts across assets are channeled through factors. Moreover, the FFP introduces commuting matrices to the literature, facilitating a more theoretically general and empirically realistic characterization of dividend covariance, flow covariance, and flow predictability, and expanding the class of analytically tractable models.

The canonical foundation of asset-pricing factors concerns only returns. Employing a dynamic equilibrium and a static arbitrage setup, the ICAPM by Merton (1973) and the APT by Ross (1976), respectively, derive the canonical factor model for expected returns. But what about an economy with commonality in both returns and flows? We address this question by showing that in a dynamic equilibrium, flow predictability generates return predictability through factors. Unlike traditional theories, factors are uniquely determined in our model through the covariance structures of returns and flows. In a related study, An (2023) offers an alternative arbitrage-pricing foundation for the same factor model of price impacts presented in this paper, while An, Su, and Wang (2023) develop empirical methodologies to estimate this model.

3 Model Setup

In this section, we present the model setup.

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a continuous-time filtration $\{\mathcal{F}_t : t \geq 0\}$. The risk-free rate is a constant $r > 0$. In our notation, we employ bold fonts to represent matrices and vectors, and we use $\mathbf{A}^\top$ to indicate the transpose of a matrix $\mathbf{A}$.

The model encompasses N assets, with the price of these assets at time t represented as $\mathbf{P}_t = (P_{1,t}, P_{2,t}, \ldots, P_{N,t})^\top$. Asset n has S_n shares outstanding, expressed as $\mathbf{S} = (S_1, S_2, \ldots, S_N)^\top$. Each share of asset n yields a dividend rate of $D_{n,t}$ at time t, denoted as $\mathbf{D}_t = (D_{1,t}, D_{2,t}, \ldots, D_{N,t})^\top$. The process of the N assets' dividends satisfies

$$d\mathbf{D}_t = \boldsymbol{\phi} dt + d\mathbf{B}_t^{\{D\}}, \tag{1}$$

where $\mathbf{B}_t^{\{D\}} = \left(B_{1,t}^{\{D\}}, B_{2,t}^{\{D\}}, \ldots, B_{N,t}^{\{D\}}\right)^\top$ are correlated Brownian motions with a variance-covariance matrix $\boldsymbol{\Sigma}^{\{D\}}$, and the $N \times 1$ constant vector $\boldsymbol{\phi}$ dentoes the dividend growth rate. The fluctuation of dividends $\mathbf{D}_t$ generates the fundamental fluctuation of asset prices.

The total amount of flow (expressed in the number of shares) into asset n from time 0 to time t is denoted as $f_{n,t}$, with $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{N,t})^\top$. Consequently, the instantaneous flow into asset n at time t is $df_{n,t}$, which is assumed to satisfy

$$d\mathbf{f}_t = \boldsymbol{\kappa} \mathbf{f}_t dt + d\mathbf{B}_t^{\{f\}} - \mathbf{f}_{t-} dM_t, \tag{2}$$

where $\mathbf{B}_t^{\{f\}} = \left(B_{1,t}^{\{f\}}, B_{2,t}^{\{f\}}, \ldots, B_{N,t}^{\{f\}}\right)^\top$ are correlated Brownian motions with a variance-covariance matrix $\boldsymbol{\Sigma}^{\{f\}}$ and M_t is a Poisson process with a constant intensity $\eta \geq 0$. Without loss of generality, we assume that $\boldsymbol{\Sigma}^{\{f\}}$ and $\boldsymbol{\Sigma}^{\{D\}}$ are of full rank.[3] The flows are assumed

[3]Otherwise, we can choose lower-dimensional linearly independent portfolios and rotate the dividend $\mathbf{D}_t$

noisy in that the Brownian motions $\mathbf{B}_t^{\{D\}}$ and $\mathbf{B}_t^{\{f\}}$ are independent.

The $N \times N$ matrix $\boldsymbol{\kappa} = \{\kappa_{n,m}\}$ quantifies the short-run predictability of the flows into the N assets. If $\kappa_{n,m}$ is positive, a higher flow into asset m predicts an increased future flow into asset n. Technically speaking, the matrix $\boldsymbol{\kappa}$ characterizes the structure of the continuous-time Ornstein–Uhlenbeck process.

In the long term, flows always revert due to the last term $-\mathbf{f}_{t-}dM_t$. This term means that with an exponential clock of intensity η, the flows $\mathbf{f}_t$ jump back to zero, leading to a reversion of asset prices to their fundamental values. Ideally, one might prefer a model incorporating a more gradual long-term reversion. However, such models typically require the introduction of additional state variables. We use a continuous-time Ornstein–Uhlenbeck setup with jumps to retain analytical tractability. Moreover, in process (2), flows are inelastic to asset prices, and all assets' flows reset to zero at the same time. These assumptions could be relaxed, and we present the model extension in Section 5.4.

Our model features a continuum of arbitrageurs with a total mass of $\mu > 0$. All arbitrageurs are identical and have access to all available information up to time t, denoted as $\mathcal{F}_t = \{\mathbf{D}_s, \mathbf{f}_s | s \leq t\}$. Arbitrageurs have a CARA utility function $u(c) = -\exp(-\gamma c)$, with c denoting the consumption rate and γ the risk aversion parameter. The equilibrium satisfies two conditions:

- Arbitrageurs choose consumption c_t and investment in the N assets $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})^\top$ at time t to maximize their discounted utility

$$\mathbb{E}\left[\int_t^\infty e^{-\rho(s-t)} u(c_s)\, ds \,\middle|\, \mathcal{F}_t\right], \tag{3}$$

  where $\rho > 0$ represents the subjective discount rate.

and flow $\mathbf{f}_t$ to these portfolios.

- Markets clear for all assets at all time,

$$\mu \mathbf{y}_t = \mathbf{S} - \mathbf{f}_t. \tag{4}$$

4 Factorizable Flow Predictability

In this section, we present the FFP condition in order to solve the model.

In general, the model presented so far is intractable, because the dividend covariance $\mathbf{\Sigma}^{\{D\}}$, the flow covariance $\mathbf{\Sigma}^{\{f\}}$, and the flow predictability $\boldsymbol{\kappa}$ can take any arbitrary matrix form. While one can achieve tractability by assuming diagonal matrices, such assumptions are both empirically counterfactual and theoretically restrictive. However, the model becomes tractable under the following condition.

ASSUMPTION 1. ***Factorizable Flow Predictability (FFP)***

We assume that

$$\mathbf{\Sigma}^{\{f\}}\mathbf{\Sigma}^{\{D\}}\boldsymbol{\kappa} = \boldsymbol{\kappa}\mathbf{\Sigma}^{\{f\}}\mathbf{\Sigma}^{\{D\}}. \tag{5}$$

As discussed in the introduction, the FFP employs commuting matrices. Essentially, the FFP is the necessary and sufficient condition ensuring that the three matrices $\mathbf{\Sigma}^{\{D\}}$, $\mathbf{\Sigma}^{\{f\}}$, and $\boldsymbol{\kappa}$ can be rotated to become diagonal at the same time. The next proposition clarifies this, showing that N factors emerge uniquely and endogenously. These N factors have uncorrelated dividends and flows, and each factor's flow predictability depends only on its own flow. Appendix A.1 gives a proof.

PROPOSITION 1. *Let the Cholesky decomposition of* $\mathbf{\Sigma}^{\{D\}}$ *be* $\mathbf{U}^\top\mathbf{U}$*. Under the technical assumption that* $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$ *has distinct eigenvalues, the FFP holds if and only if there exist* N *unique factors* $\mathbf{o}_n = (o_{1,n}, o_{2,n}, \ldots, o_{N,n})^\top$ *(up to multiplication by* -1*), which we denote as an* $N \times N$ *matrix* $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_N)$*, satisfying* *i. factor decomposition of flow*

$$\mathbf{f}_t = \sum_{n=1}^{N} \mathbf{o}_n \tilde{f}_{n,t} = \mathbf{O}\tilde{\mathbf{f}}_t, \tag{6}$$

where $\tilde{f}_{n,t}$ represents the flow into factor n from time 0 to t, and $\tilde{\mathbf{f}}_t = (\tilde{f}_{1,t}, \tilde{f}_{2,t}, \ldots, \tilde{f}_{N,t})^\top$.

ii. uncorrelated dividend

$$\mathbf{O}^\top \mathbf{\Sigma}^{\{D\}} \mathbf{O} = r^2 \mathbf{I}_N. \tag{7}$$

iii. uncorrelated flow

$$\mathbf{O}^{-1} \mathbf{\Sigma}^{\{f\}} (\mathbf{O}^{-1})^\top = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2), \tag{8}$$

where $\sigma_1 > \sigma_2 > \cdots > \sigma_N > 0$.

iv. each factor's flow predictability depends only on its own flow

$$\mathbf{O}^{-1} \boldsymbol{\kappa} \mathbf{O} = \operatorname{diag}(\tilde{\kappa}_1, \tilde{\kappa}_2, \ldots, \tilde{\kappa}_N) := \tilde{\boldsymbol{\kappa}}. \tag{9}$$

Equation (6) merits further discussion. Given that one share of factor n holds $o_{m,n}$ shares of asset m, a one-share flow into factor n increases the flow of asset m by $o_{m,n}$ shares. This rationale explains why $f_{m,t} = \sum_{n=1}^{N} o_{m,n} \tilde{f}_{n,t}$, which in matrix form is equation (6). Conversely, to derive factor flows $\tilde{\mathbf{f}}_t$ from asset flows $\mathbf{f}_t$ and portfolio weights $\mathbf{O}$, one should apply the equation $\tilde{\mathbf{f}}_t = \mathbf{O}^{-1}\mathbf{f}_t$, which differs from the construction of portfolio returns (i.e., $\tilde{\mathbf{f}}_t \neq \mathbf{O}^\top \mathbf{f}_t$).

At its essence, Proposition 1 constructs N factors, namely $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_N$. These factors are characterized by uncorrelated dividends (as per equation (7)) and uncorrelated flows[4] (as per equation (8)). The FFP imposes an economic constraint on the N factors, specifically as illustrated in (9)—each factor n's flow predictability relies solely on its own flow, but not on the flow into other factors. Each factor's flow predictability is thus represented by a single parameter $\tilde{\kappa}_n$.

[4] We have the freedom to scale the N factors that have uncorrelated dividends and flows. Proposition 1 adopts a convenient normalization that sets the dividend risks to $r^2 \mathbf{I}_N$.

Proposition 1 imposes a technical requirement that $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$ has distinct eigenvalues. Subject to this constraint, the FFP corresponds to the existence of N unique factors. However, this equivalence does not hold when $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$ has duplicate eigenvalues. The FFP theory also covers this duplicate-eigenvalue scenario, which is presented in the Online Appendix. In this context, the N factors are no longer uniquely determined, thereby the theory acknowledges the data limitation and permits a factor's flow predictability to depend on the flows into all factors residing within the same eigenspace.

Given Proposition 1, we can reinterpret the model setup in Section 3 through the lens of these N factors, leading to significant simplification. Specifically, the flow process (2) becomes

$$d\tilde{\mathbf{f}}_t = \mathbf{O}^{-1} d\mathbf{f}_t = \tilde{\boldsymbol{\kappa}}\tilde{\mathbf{f}}_t dt + \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N) d\tilde{\mathbf{B}}_t^{\{f\}} - \tilde{\mathbf{f}}_{t-} dM_t, \tag{10}$$

where $\tilde{\mathbf{B}}_t^{\{f\}} = \left(\tilde{B}_{1,t}^{\{f\}}, \tilde{B}_{2,t}^{\{f\}}, \ldots, \tilde{B}_{N,t}^{\{f\}}\right)^\top$ are independent standard Brownian motions. Therefore, we can express equation (10) element-wise as

$$d\tilde{f}_{n,t} = \tilde{\kappa}_n \tilde{f}_{n,t} dt + \sigma_n d\tilde{B}_{n,t}^{\{f\}} - \tilde{f}_{n,t-} dM_t, \tag{11}$$

which implies that each factor's flow $\tilde{f}_{n,t}$ is independent of the flow of other factors. Similarly, the dividend process (1) becomes

$$d\tilde{\mathbf{D}}_t = \mathbf{O}^\top d\mathbf{D}_t = \tilde{\boldsymbol{\phi}} dt + r d\tilde{\mathbf{B}}_t^{\{D\}}, \tag{12}$$

where $\tilde{\mathbf{B}}_t^{\{D\}} = (\tilde{B}_{1,t}^{\{D\}}, \tilde{B}_{2,t}^{\{D\}}, \ldots, \tilde{B}_{N,t}^{\{D\}})^\top$ are independent standard Brownian motions, and $\tilde{\boldsymbol{\phi}} := \mathbf{O}^\top \boldsymbol{\phi}$ denotes the dividend growth rate of the N factors. We can express equation (12) element-wise as

$$d\tilde{D}_{n,t} = \tilde{\phi}_n dt + r d\tilde{B}_{n,t}^{\{D\}}, \tag{13}$$

which implies that the dividends $\tilde{D}_{n,t}$ of each factor are also independent of the dividends of other factors. The N factors have $\tilde{\mathbf{S}} = \mathbf{O}^{-1}\mathbf{S}$ outstanding shares. The market clears for the N assets if and only if the market clears for the N factors, so equation (4) can be equivalently expressed as

$$\mu\tilde{\mathbf{y}}_t = \tilde{\mathbf{S}} - \tilde{\mathbf{f}}_t, \tag{14}$$

where $\tilde{\mathbf{y}}_t = (\tilde{y}_{1,t}, \tilde{y}_{2,t}, \dots, \tilde{y}_{N,t})^\top$ represent the arbitrageurs' optimal investments in the N factors at time t.

In summary, this section demonstrates the application of the FFP condition to restructure the general problem concerning N assets into N factors. This process results in a significant simplification, as we shall show in the following section that the price equilibrium for each factor can be determined independently of the other factors.

5 Model Solution

In this section, we present the model solution. Section 5.1 solves for the factor price. Section 5.2 presents the factor model of price impacts for the N original assets. Section 5.3 considers an equivalent model formulation using actual returns. Section 5.4 considers model extension.

5.1 Factor Price

To solve for the factor price, we conjecture that the arbitrageurs' value function is separable for the N factors and takes the form

$$J(W, \tilde{\mathbf{f}}, \tilde{\mathbf{D}}) = -\frac{1}{r} e^{-r\gamma W - b_0 - \sum_{n=1}^N b_{1,n}\tilde{f}_n - \frac{1}{2}\sum_{n=1}^N b_{2,n}\tilde{f}_n^2 - \sum_{n=1}^N a_{1,n}\tilde{D}_n - \frac{1}{2}\sum_{n=1}^N a_{2,n}\tilde{D}_n^2}, \tag{15}$$

where W is the arbitrageurs' wealth. Furthermore, we conjecture that the price of each factor linearly depends on its own flows and dividends. Proposition 2 provides the solution for the factor price, and its proof can be found in Appendix A.2. Specifically, we assume the no-Ponzi condition, $\tilde{\kappa}_n < (r + \eta)/2$. We shall delve more deeply into the implications of this condition later. Additionally, we employ the standard method of linearizing the jump term in the Bellman equation.[5]

PROPOSITION 2. *The equilibrium price of factor n satisfies*

$$d\tilde{P}_{n,t} = \lambda_n d\tilde{f}_{n,t} + d\tilde{D}_{n,t}/r, \tag{16}$$

where λ_n is the unique solution to the following equation

$$\tilde{\kappa}_n(\tilde{\kappa}_n - r - \eta) = \frac{r\gamma}{\mu}\left(\frac{r\gamma}{\mu} - \lambda_n(r+\eta)\right)\left(\frac{1}{\lambda_n^2} + \sigma_n^2\right). \tag{17}$$

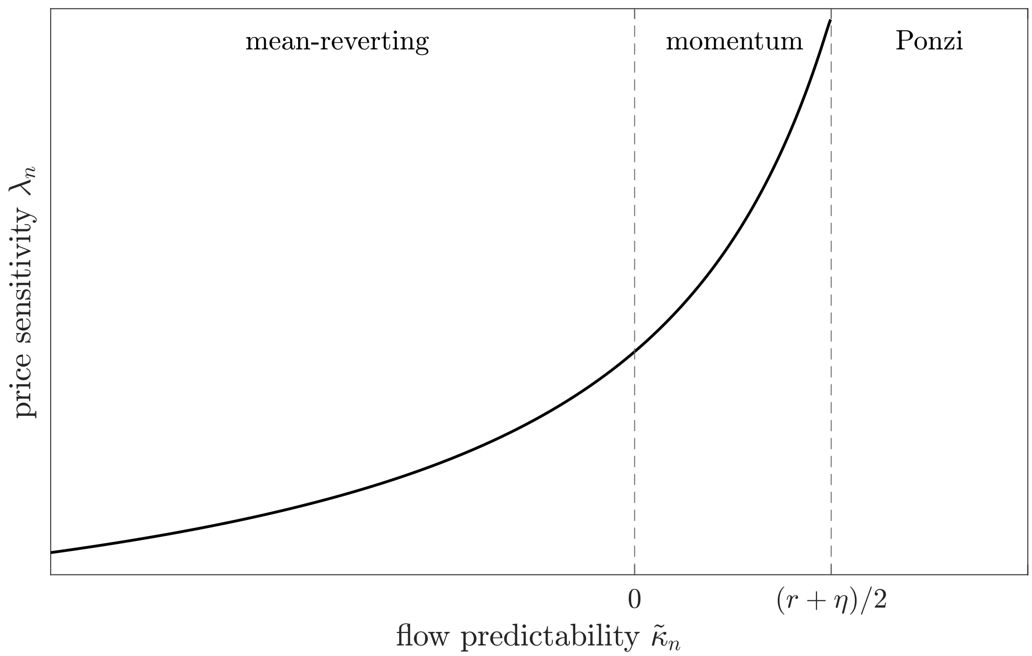
Proposition 2 reveals that the price fluctuations $d\tilde{P}_{n,t}$ for each factor consist of a flow-driven component $\lambda_n d\tilde{f}_{n,t}$ and a fundamental-driven component $d\tilde{D}_{n,t}/r$. The crucial parameter, λ_n, gauges the price sensitivity of factor n to its flow. Derived from the arbitrageurs' optimality conditions, equation (17) ties the price sensitivity λ_n with short-term flow predictability $\tilde{\kappa}_n$, long-term flow reversal η, arbitrageurs' risk aversion γ and mass μ, and flow volatility σ_n. Figure 3 illustrates the equilibrium's different regions.

First, for $\tilde{\kappa}_n < 0$, flows show short-term mean-reversion, as indicated by the negative drift term $\tilde{\kappa}_n \tilde{f}_{n,t} dt$ in (11). An inflow into factor n today predicts an outflow from it tomorrow. Previous studies often assume mean-reverting flows for analytical simplicity, as this ensures well-behaved equilibria (Campbell and Kyle, 1993; Wang, 1993).

Though theoretically simple, mean-reverting flows differ significantly from empirical findings. Specifically, mutual fund and retail investor flows often display short-term momentum at daily, weekly, and monthly intervals, as highlighted in studies like Lee, Liu, Roll, and Subrahmanyam (2004), Kelley and Tetlock (2013), Barber, Huang, Odean, and Schwarz (2021), and Boehmer, Jones, Zhang, and Zhang (2021). As a result, it is vital to move beyond the mean-reverting flow assumption and explore the equilibrium of momentum-driven flows.

[5] This approximation method has been previously utilized by Biais (1993), Duffie, Gârleanu, and Pedersen (2007), Vayanos and Weill (2008), Gârleanu (2009), Praz (2014), Uslu and Velioglu (2021), and among others.

Figure 3. Relationship between the price sensitivity λ_n and flow predictability $\tilde{\kappa}_n$



Note: This figure illustrates the relationship between a factor's price sensitivity λ_n and flow predictability $\tilde{\kappa}_n$. The given λ_n values are computed using equation (17) with the parameters set as follows: $\gamma = 3$, $\mu = 0.1$, $r = 0.1$, $\sigma = 0.01$, and $\eta = 10$.

Second, we establish the equilibrium existence for $\tilde{\kappa}_n$ in the interval $(0, (r+\eta)/2)$. Here, short-term momentum is evident as today's inflow into factor n forecasts a similar inflow tomorrow. Within this domain, the price sensitivity λ_n increases sharply with $\tilde{\kappa}_n$. Intuitively, with stronger flow momentum, arbitrageurs taking in one flow unit today anticipate more inflow tomorrow. Such dynamic consideration prompts arbitrageurs to demand greater compensation for absorbing today's flow, leading to an elevated λ_n. We formally prove this comparative static in Proposition 4.

Third, equilibrium no longer exists when the flow momentum $\tilde{\kappa}_n$ surpasses $(r+\eta)/2$. This threshold is approximately half of the flow's long-term reversion speed η. In such cases, asset price bubbles form. This occurrence is linked to the momentum in flows, which results in momentum in price dislocation. Arbitrageurs absorbing these flows have to endure short-term mark-to-market losses. However, these losses are temporary, given that flows revert to zero at an intensity of η. This reversion prompts the asset price to align with its fundamental value, allowing arbitrageurs to recoup all losses while also profiting from trading against the dislocated price. The following proposition proves that if $\tilde{\kappa}_n$ surpasses $(r+\eta)/2$, providing liquidity to flow would require arbitrageurs to run a Ponzi scheme—the expected temporary loss becomes infinitely large. Consequently, rational arbitrageurs refuse to provide liquidity and correct mispricing. Appendix A.3 provides a proof.

PROPOSITION 3. *If* $\tilde{\kappa}_n < (r+\eta)/2$, *then* $\lim_{t\to\infty} e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2) = 0$. *If* $\tilde{\kappa}_n \geq (r+\eta)/2$, *then* $\limsup_{t\to\infty} e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2) = \infty$.

Proposition 3 shows that the threshold $\tilde{\kappa}_n = (r+\eta)/2$ decides if $e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2)$ approaches zero or explodes to infinity. To comprehend why an explosive $e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2)$ is akin to a Ponzi scheme, consider the arbitrageurs trading against the flows $\tilde{f}_{n,t}$. Since the price dislocation $\lambda_n \tilde{f}_{n,t}$ is proportional to these flows, the discounted expected temporary mark-to-market loss is proportional to $e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2)$. Proposition 3 emphasizes that if $\tilde{\kappa}_n \geq (r+\eta)/2$, arbitrageurs effectively engage in a Ponzi scheme when absorbing flows, as they incur an infinite expected loss before realizing any profit.[6]

Moreover, the threshold for the flow momentum $\tilde{\kappa}_n = (r+\eta)/2$ corresponds to a finite value of price sensitivity λ_n. This implies that prices cannot be excessively responsive to flow. This insight also follows from the Ponzi scheme intuition. If prices respond too dramatically to flows, arbitrageurs would be disinclined to provide liquidity for the initial unit of flows in a dynamic environment.

Having delineated the different equilibrium regions, Proposition 4 states the key properties of the price sensitivity λ_n. Appendix A.4 provides a proof.

PROPOSITION 4. *The price sensitivity* λ_n *is strictly increasing in flow predictability* $\tilde{\kappa}_n$, *strictly increasing in risk aversion* γ, *strictly decreasing in the mass of arbitrageurs* μ, *and strictly increasing in flow volatility* σ_n *if* $\tilde{\kappa}_n < 0$ *and strictly decreasing in* σ_n *if* $\tilde{\kappa}_n > 0$.

[6] If $\tilde{\kappa}_n < r+\eta$, neither the discounted price dislocation $\lambda_n \tilde{f}_{n,t}$ nor flow $\tilde{f}_{n,t}$ explode to infinity. Hence, the key threshold $\tilde{\kappa}_n = (r+\eta)/2$ is derived from the Ponzi scheme rationale rather than any erratic behavior of the flow process itself.

The price sensitivity λ_n increases with flow predictability $\tilde{\kappa}_n$. This not only is a comparative static but also holds across different factors. Because factors can have unique $\tilde{\kappa}_n$ values, their λ_n can also differ. In Section 5.2, we demonstrate how varying λ_n enriches the dynamics of price impacts cross-sectionally.

Proposition 4 shows that λ_n rises with risk aversion γ and falls with the number of arbitrageurs μ. This is intuitive. The nuanced aspect of Proposition 4 is the tie between λ_n and flow volatility σ_n. For mean-reverting flows ($\tilde{\kappa}_n < 0$), arbitrageurs prefer this reversion. An increased σ_n means more uncertainty in this reversion, making arbitrageurs seek a higher λ_n. However, for momentum flows ($\tilde{\kappa}_n > 0$), arbitrageurs dislike this trend. A greater σ_n suggests more doubt in this momentum, leading arbitrageurs to demand a smaller λ_n.

5.2 Factor Model of Price Impacts

This section uses the factor price in Proposition 2 to derive the factor model of price impacts.

Consider any asset that holds w_n shares of factor n. The asset's dividend rate at time t is $D_t = \sum_{n=1}^{N} w_n \tilde{D}_{n,t}$. Because the factors' dividend rates are independent by (13), we have $\text{cov}(dD_t, d\tilde{D}_{n,t}) = w_n \text{var}(d\tilde{D}_{n,t}) = w_n r^2 dt$. Therefore, by the no-arbitrage condition, the asset's price satisfies

$$dP_t = \sum_{n=1}^{N} w_n d\tilde{P}_{n,t} = \sum_{n=1}^{N} \lambda_n \frac{\text{cov}(dD_t/r, d\tilde{D}_{n,t}/r)}{dt} d\tilde{f}_{n,t} + dD_t/r. \tag{18}$$

Defining $dq_{n,t} = \tilde{P}_{n,t} d\tilde{f}_{n,t}$ as the flow into factor n at time t measured in the unit of dollars, we obtain the following factor model.

PROPOSITION 5. *The factor model of price impacts is*

$$\underbrace{\frac{dP_t}{P_t}}_{\text{actual return}} = \sum_{n=1}^{N} \overbrace{\lambda_n}^{\text{price sensitivity}} \underbrace{\text{cov}\left(\frac{dD_t}{rP_t}, \frac{d\tilde{D}_{n,t}}{r\tilde{P}_{n,t}}\right)/dt}_{\text{quantity of fundamental risk}} \overbrace{dq_{n,t}}^{\text{factor flow}} + \underbrace{\frac{dD_t}{rP_t}}_{\text{fundamental return}}. \tag{19}$$

Under model (19), the instantaneous return of any asset dP_t/P_t consists of two components. The first component $dD_t/(rP_t)$ corresponds to the variation driven by fundamentals. The second component is price impact, which emerges due to the arbitrageurs' required compensation for absorbing the extra fundamental risk induced by the flow.

The price impact component is a product of three elements. The instantaneous dollar-amount flow into factor n at time t is $dq_{n,t}$. The quantity of fundamental risk per dollar of factor flow is determined by the covariance between the asset's fundamental return $dD_t/(rP_t)$ and the factor's fundamental return $d\tilde{D}_{n,t}/(r\tilde{P}_{n,t})$. The price sensitivity λ_n quantifies the extent to which the price is affected by each unit of flow-induced fundamental risk.

Our dynamic model allows λ_n to vary across factors. The static CARA-normal price impact model is a special case of the dynamic model (19), but with the additional constraint[7] $\lambda_1 = \lambda_2 = \cdots = \lambda_N$. This constraint implies that arbitrageurs demand the same amount of compensation for absorbing flow-induced risk, irrespective of the factor. This restriction, however, is empirically rejected by An, Su, and Wang (2023). In contrast, our dynamic model does not bind to this same price sensitivity across factors. According to Proposition 4, when factor flows show more momentum, its price sensitivity rises, reflecting increased risk aversion by arbitrageurs for that factor.

5.3 Equivalent Formulation Using Actual Return

The price impact component of our factor model (19) is based on the asset's fundamental return, denoted by $dD_t/(rP_t)$, and the factor's fundamental return, denoted by $d\tilde{D}_{n,t}/(r\tilde{P}_{n,t})$. An empirical challenge arises from the fact that these fundamental returns are not directly observable. This section demonstrates the feasibility of employing actual returns in lieu of fundamental returns within the same factor model framework, offering a practical solution to the aforementioned empirical issue.

The key insight from our analysis is the specific construction of factors in Proposition 1,

[7] The static model can be obtained by setting $\tilde{\kappa}_n = 0$ and $\eta = 0$ in our dynamic model.

which results in uncorrelated flows and dividends. This construction implies that the actual returns of these factors, comprising both flow-driven and fundamental-driven components, are also uncorrelated. Under the factor model (19), the covariance between the actual returns of any asset and factor n can be expressed as

$$\text{cov}\left(\frac{dP_t}{P_t}, \frac{d\tilde{P}_{n,t}}{\tilde{P}_{n,t}}\right) = (1 + \lambda_n^2 \text{var}(d\tilde{f}_{n,t})/dt)\,\text{cov}\left(\frac{dD_t}{rP_t}, \frac{d\tilde{D}_{n,t}}{r\tilde{P}_{n,t}}\right), \tag{20}$$

where we use the definition $dq_{n,t} = \tilde{P}_{n,t} d\tilde{f}_{n,t}$, and the fact that factors have uncorrelated flows from (11) and uncorrelated dividends from (13). Equation (20) shows that the covariance of actual returns between any asset and a factor includes both a fundamental-driven component and a flow-driven component, which is influenced by the factor's price sensitivity λ_n and flow variation $\text{var}(d\tilde{f}_{n,t})/dt$.

By equation (20), we can reformulate the factor model of price impacts using actual returns as follows,

$$\underbrace{\frac{dP_t}{P_t}}_{\text{actual return}} = \sum_{n=1}^{N} \overbrace{\frac{\lambda_n}{1 + \lambda_n^2 \text{var}(d\tilde{f}_{n,t})/dt}}^{\text{price sensitivity}} \underbrace{\text{cov}\left(\frac{dP_t}{P_t}, \frac{d\tilde{P}_{n,t}}{\tilde{P}_{n,t}}\right)/dt}_{\text{quantity of risk}} \overbrace{dq_{n,t}}^{\text{factor flow}} + \underbrace{\frac{dD_t}{rP_t}}_{\text{fundamental return}}. \tag{21}$$

The key difference between models (19) and (21) lies in how the quantity of risk is measured. In model (19), the risk is assessed using fundamental returns, resulting in a price sensitivity of λ_n. Conversely, model (21) assesses risk using actual returns, which yield a larger measure of risk compared to fundamental returns. Consequently, for the same price impact, the corresponding price sensitivity is now reduced to $\lambda_n/(1 + \lambda_n^2 \text{var}(d\tilde{f}_{n,t})/dt))$.

5.4 Model Extension

In this section, we extend the baseline model to accommodate price-elastic flows and differential long-term reversion speeds for different factors.

We explicitly specify the dynamics for the flows of the N factors. This specification is without loss of generality, as we can equivalently rotate the flow dynamics to the N original assets by applying Proposition 1. Instead of (11), we assume that the flow into the n-th factor satisfies

$$d\tilde{f}_{n,t} = (\tilde{\kappa}_n \tilde{f}_{n,t} + \chi_n(\tilde{P}_{n,t}(\tilde{\mathbf{f}}_t) - \tilde{P}_{n,t}(\mathbf{0})))dt + \sigma_n d\tilde{B}^{\{f\}}_{n,t} - \tilde{f}_{n,t-} dM_{n,t}, \tag{22}$$

where $\tilde{P}_{n,t}(\tilde{\mathbf{f}}_t) - \tilde{P}_{n,t}(\mathbf{0})$ denotes the flow-induced price dislocation of factor n at time t. Parameter χ_n gauges how noise trading flows react to the price dislocation of factor n. In our baseline model, where flows are fully inelastic, $\chi_n = 0$. If χ_n is positive, noise traders engage in positive feedback trading, as empirically documented by Ben-David, Li, Rossi, and Song (2022) for mutual fund flows. On the other hand, a negative χ_n implies that noise traders mechanically sell when prices are high and buy when they are low. This mechanical liquidity provision by noise traders differs from our modeled arbitrageurs, who engage in dynamic portfolio and consumption optimization.

Moreover, $M_{n,t}$ represents a Poisson process with a constant intensity $\eta_n \geq 0$. Different $M_{n,t}$ processes are independent of one another. Unlike the baseline model, where all flows revert to zero simultaneously, in this extended model, flows into different factors can have differential long-run reversion speeds, represented by η_n.

We generalize Proposition 2 to derive the factor price for the extended model. Appendix A.5 provides a proof.

PROPOSITION 6. *The equilibrium price of factor n satisfies*

$$d\tilde{P}_{n,t} = \lambda_n d\tilde{f}_{n,t} + d\tilde{D}_{n,t}/r, \tag{23}$$

where λ_n is the unique solution to the following equations

$$\hat{\kappa}_n = \tilde{\kappa}_n + \chi_n \lambda_n, \tag{24}$$

$$\hat{\kappa}_n(\hat{\kappa}_n - r - \eta_n) = \frac{r\gamma}{\mu}\left(\frac{r\gamma}{\mu} - \lambda_n(r+\eta_n)\right)\left(\frac{1}{\lambda_n^2} + \sigma_n^2\right). \tag{25}$$

For each factor n, the noise traders' response χ_n and reversion speed η_n influence its own price sensitivity λ_n to flow. As shown by (24), a higher χ_n increases the momentum of the factor flow from $\tilde{\kappa}_n$ to $\hat{\kappa}_n$, which in turn increases the price sensitivity λ_n. Equation (25) represents a refined version of the baseline equation (17), where each factor n has a unique short-run momentum $\hat{\kappa}_n$ and long-run reversion speed η_n. Apart from the adjusted λ_n, the factor model of price impacts (19) stays the same in the extended model.

6 Reducing the Number of Factors

So far, we have solved the factor model of price impacts (19) using N factors for N assets. This method requires the knowledge of the $N \times N$ matrices of flow covariance $\mathbf{\Sigma}^{\{f\}}$, the dividend covariance $\mathbf{\Sigma}^{\{D\}}$, and the flow predictability $\boldsymbol{\kappa}$. However, reliably estimating these quantities is not feasible when the number of assets N is large. Instead, we could select a smaller set of K factors. From these, we can estimate the corresponding $K \times K$ matrices $\check{\mathbf{\Sigma}}^{\{f\}}$, $\check{\mathbf{\Sigma}}^{\{D\}}$, and $\check{\boldsymbol{\kappa}}$, and construct an approximate K-factor model of price impacts.

In this section, we answer one question. Assuming the FFP $\mathbf{\Sigma}^{\{f\}}\mathbf{\Sigma}^{\{D\}}\boldsymbol{\kappa} = \boldsymbol{\kappa}\mathbf{\Sigma}^{\{f\}}\mathbf{\Sigma}^{\{D\}}$ holds for the N underlying assets, does it necessarily apply to the K selected factors as well? In other words, do we find that $\check{\mathbf{\Sigma}}^{\{f\}}\check{\mathbf{\Sigma}}^{\{D\}}\check{\boldsymbol{\kappa}} = \check{\boldsymbol{\kappa}}\check{\mathbf{\Sigma}}^{\{f\}}\check{\mathbf{\Sigma}}^{\{D\}}$?

The answer is no. We prove that the FFP holds for the K factors if and only if the K factors preserve the orthogonality structure inherent in the N-dimensional data. This implies that by testing whether the FFP holds for the K factors, econometricians can assess the effectiveness of the selected K factors empirically.

We now formally state the theoretical result. As per Proposition 1, when the FFP holds for the N assets, there exist N unique factors $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_N$. We identify the selected K factors as an $N \times K$ portfolio-weight matrix $\mathbf{X}$, where factor k holds $x_{n,k}$ share of the true factor[8] $\mathbf{o}_n$. The K factors are non-redundant such that $\mathbf{X}$ has full column rank. By equation (12), the dividend process of the K factors is

$$d\check{\mathbf{D}}_t = \mathbf{X}^\top d\tilde{\mathbf{D}}_t = \mathbf{X}^\top \tilde{\boldsymbol{\phi}} dt + r d\check{\mathbf{B}}_t^{\{D\}}, \tag{26}$$

where the variance-covariance matrix of Brownian motions $\check{\mathbf{B}}_t^{\{D\}}$ is $\check{\boldsymbol{\Sigma}}^{\{D\}} := \mathbf{X}^\top \mathbf{X}$.

We use the N true factors' flows $\tilde{\mathbf{f}}_t$ and portfolio weights $\mathbf{X}$ to construct the flows $\check{\mathbf{f}}_t$ of the K factors. As discussed in Proposition 1, our aim is to ensure that $\tilde{\mathbf{f}}_t \approx \mathbf{X}\check{\mathbf{f}}_t$. This is an approximation because $K < N$, implying that the flows $\check{\mathbf{f}}_t$ of the K factors cannot entirely span the flows $\tilde{\mathbf{f}}_t$ of the N true factors. Thus, we define the flows of the K factors as $\check{\mathbf{f}}_t = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\mathbf{f}}_t$, where $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ is the generalized inverse. By (10) and the approximation $\tilde{\mathbf{f}}_t \approx \mathbf{X}\check{\mathbf{f}}_t$, the flow process of the K factors is

$$\begin{aligned} d\check{\mathbf{f}}_t &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathrm{diag}(\tilde{\kappa}_1, \tilde{\kappa}_2, \ldots, \tilde{\kappa}_N) \tilde{\mathbf{f}}_t dt + d\check{\mathbf{B}}_t^{\{f\}} - \check{\mathbf{f}}_{t-} dM_t \\ &\approx \check{\boldsymbol{\kappa}} \check{\mathbf{f}}_t dt + d\check{\mathbf{B}}_t^{\{f\}} - \check{\mathbf{f}}_{t-} dM_t, \end{aligned} \tag{27}$$

where $\check{\boldsymbol{\kappa}} := (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathrm{diag}(\tilde{\kappa}_1, \tilde{\kappa}_2, \ldots, \tilde{\kappa}_N) \mathbf{X}$ is the predictability matrix of the flow of the K factors, and $\check{\boldsymbol{\Sigma}}^{\{f\}} := (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}$ is the variance-covariance matrix of Brownian motions $\check{\mathbf{B}}_t^{\{f\}}$.

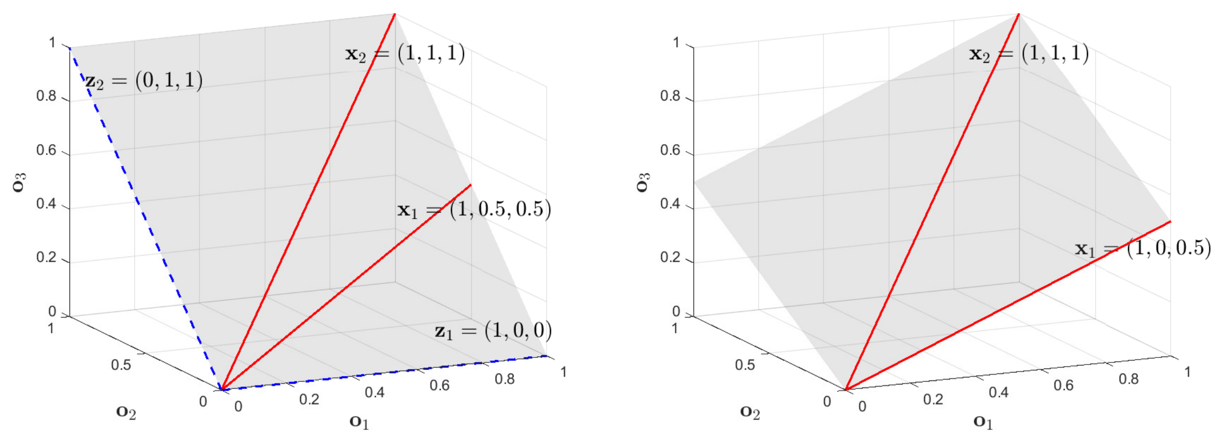
The following proposition offers a complete characterization of $\mathbf{X}$ for which the FFP holds for the K factors. Appendix A.6 provides a proof.

PROPOSITION 7. *The following holds true:*

$$\check{\boldsymbol{\Sigma}}^{\{f\}} \check{\boldsymbol{\Sigma}}^{\{D\}} \check{\boldsymbol{\kappa}} = \check{\boldsymbol{\kappa}} \check{\boldsymbol{\Sigma}}^{\{f\}} \check{\boldsymbol{\Sigma}}^{\{D\}} \tag{28}$$

[8] Note that $\mathbf{X}$ is the portfolio weights expressed in terms of the N factors, not the N underlying assets. The portfolio weights in terms of the N assets are $\mathbf{OX}$. Because $\mathbf{O}$ is an $N \times N$ invertible matrix, both expressions of portfolio weights are equivalent.

Figure 4. Illustration for when the FFP holds for K factors



Good case: FFP holds for K factors **Bad case: FFP fails for K factors**

Note: We illustrate when the FFP holds in a scenario with $N = 3$ assets and $K = 2$ factors. The left panel illustrates the good case when the FFP holds. Here, through rotation, the two factors $\mathbf{x}_1$ and $\mathbf{x}_2$ yield $\mathbf{z}_1 = \mathbf{o}_1$ and $\mathbf{z}_2 = \mathbf{o}_2 + \mathbf{o}_3$. The right panel illustrates the bad case when the FFP fails, making it impossible to rotate $\mathbf{x}_1$ and $\mathbf{x}_2$ into non-overlapping combinations of the true factors $\mathbf{o}_1$, $\mathbf{o}_2$, and $\mathbf{o}_3$.

for any parameters $\tilde{\kappa}_1, \tilde{\kappa}_2, \ldots, \tilde{\kappa}_N$ and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$ if and only if $\mathbf{X} = \mathbf{ZH}$, where $\mathbf{Z}$ is an $N \times K$ matrix of full column rank having at most one non-zero entry in each row, and $\mathbf{H}$ is a $K \times K$ invertible matrix.

Proposition 7 shows that the FFP holds for the K factors if and only if these factors can be decomposed as $\mathbf{X} = \mathbf{ZH}$. To understand this decomposition, we note that $\mathbf{Z}$ represents the portfolio weights of the K factors in terms of the N true factors. The unique structure of $\mathbf{Z}$ (having at most one non-zero entry in each row) indicates that each true factor $\mathbf{o}_n$ may feature in at most one of the K factors. Essentially, $\mathbf{Z}$ partitions the N true factors into K distinct, non-overlapping segments. These K factors are then rotated by the matrix $\mathbf{H}$ to generate all "good cases" of $\mathbf{X}$ for which the FFP holds.

The left panel of Figure 4 illustrates the good case for $N = 3$ assets and $K = 2$ factors. Here, through rotation, the two factors $\mathbf{x}_1$ and $\mathbf{x}_2$ yield $\mathbf{z}_1 = \mathbf{o}_1$ and $\mathbf{z}_2 = \mathbf{o}_2 + \mathbf{o}_3$, which are non-overlapping combinations of the true factors $\mathbf{o}_1$, $\mathbf{o}_2$, and $\mathbf{o}_3$. Essentially, the two factors $\mathbf{x}_1$ and $\mathbf{x}_2$ align with the proper plane that maintains the orthogonality structure of the underlying N-dimensional data.

The right panel illustrates the bad case when the FFP fails for the K factors, making it impossible to rotate $\mathbf{x}_1$ and $\mathbf{x}_2$ into non-overlapping combinations of the true factors $\mathbf{o}_1$, $\mathbf{o}_2$, and $\mathbf{o}_3$. The plane that goes through $\mathbf{x}_1$ and $\mathbf{x}_2$ does not preserve the orthogonality structure of the underlying N-dimensional data.

To sum up, whether the FFP holds for the K factors is a simple economic test that answers a difficult question. We cannot directly assess whether the K factors that we select preserve the orthogonality structure of the underlying N-dimensional data, because we cannot reliably estimate all true factors $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_N$. However, as Proposition 7 shows, by testing whether the FFP holds for the K factors, we can assess whether these K factors preserve the orthogonality structure of the N-dimensional data.

7 Conclusion

In conclusion, we develop a dynamic model showing how the predictability of noise trading flows impacts return predictability at both individual asset and factor levels. Our model first aggregates flow predictability from individual assets to factors. Unlike static models, different factors can have differential price sensitivity to current flows, because flow-absorbing arbitrageurs anticipate differential future flows. This factor-level price impact generates momentum or reversal dynamics in factor price, which in turn determines asset-level return predictability. Additionally, when flows manifest excessive momentum, asset price bubbles can emerge, because providing liquidity under this scenario would require running a Ponzi scheme.

References

Abreu, Dilip, and Markus K Brunnermeier, 2002, Synchronization risk and delayed arbitrage, *Journal of Financial Economics* 66, 341–360.

Abreu, Dilip, and Markus K Brunnermeier, 2003, Bubbles and crashes, *Econometrica* 71, 173–204.

An, Yu, 2023, Flow-based arbitrage pricing theory, Working paper, Johns Hopkins University.

An, Yu, Yinan Su, and Chen Wang, 2023, A factor framework for cross-sectional price impacts, Working paper, Johns Hopkins University.

Arnott, Robert D, Vitali Kalesnik, and Juhani T Linnainmaa, 2023, Factor momentum, *Review of Financial Studies* 36, 3034–3070.

Barber, Brad M, Xing Huang, Terrance Odean, and Christopher Schwarz, 2021, Attention induced trading and returns: Evidence from robinhood users, *Journal of Finance* .

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022, Discontinued positive feedback trading and the decline of return predictability, *Journal of Financial and Quantitative Analysis* Forthcoming.

Bernhardt, Dan, and Bart Taub, 2008, Front-running dynamics, *Journal of Economic Theory* 138, 288–296.

Biais, Bruno, 1993, Price formation and equilibrium liquidity in fragmented and centralized markets, *Journal of Finance* 48, 157–185.

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *Journal of Finance* 76, 2249–2305.

Bogousslavsky, Vincent, 2016, Infrequent rebalancing, return autocorrelation, and seasonality, *Journal of Finance* 71, 2967–3006.

Brunnermeier, Markus K., and Stefan Nagel, 2004, Hedge funds and the technology bubble, *Journal of Finance* 59, 2013–2040.

Campbell, John Y, and Albert S Kyle, 1993, Smart money, noise trading and stock price behaviour, *Review of Economic Studies* 60, 1–34.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, 2007, Valuation in over-the-counter markets, *Review of Financial Studies* 20, 1865–1900.

Ehsani, Sina, and Juhani T Linnainmaa, 2022, Factor momentum and the momentum factor, *The Journal of Finance* 77, 1877–1919.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Gârleanu, Nicolae, 2009, Portfolio choice and pricing in illiquid markets, *Journal of Economic Theory* 144, 532–564.

Guo, Ming, and Hui Ou-Yang, 2015, Feedback trading between fundamental and nonfundamental information, *Review of Financial Studies* 28, 247–296.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Kelley, Eric K, and Paul C Tetlock, 2013, How wise are crowds? insights from retail orders and stock returns, *Journal of Finance* 68, 1229–1265.

Kelly, Bryan T, Tobias J Moskowitz, and Seth Pruitt, 2021, Understanding momentum and reversal, *Journal of Financial Economics* 140, 726–743.

Lee, Yi-Tsung, Yu-Jane Liu, Richard Roll, and Avanidhar Subrahmanyam, 2004, Order imbalances and market efficiency: Evidence from the taiwan stock exchange, *Journal of Financial and Quantitative Analysis* 39, 327–341.

Lou, Dong, 2012, A flow-based explanation for return predictability, *Review of Financial Studies* 25, 3457–3489.

Lu, Zhongjin, Steven Malliaris, and Zhongling Qin, 2023, Heterogeneous liquidity providers and night-minus-day return predictability, *Journal of Financial Economics* 148, 175–200.

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica* 867–887.

Praz, Remy, 2014, *Essays in Asset Pricing with Search Frictions*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne.

Ross, Stephen A, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–60.

Rostek, Marzena J, and Ji Hee Yoon, 2020, Equilibrium theory of financial markets: Recent developments, Working paper, University of Wisconsin - Madison.

Sadzik, Tomasz, and Chris Woolnough, 2021, Rational destabilization in a frictionless market, *Journal of Economic Theory* 192, 105169.

Uslu, Semih, and Guner Velioglu, 2021, Liquidity in the cross section of OTC assets, Working paper, Johns Hopkins University.

Vayanos, Dimitri, and Pierre-Olivier Weill, 2008, A search-based theory of the on-the-run phenomenon, *Journal of Finance* 63, 1361–1398.

Wang, Jiang, 1993, A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies* 60, 249–282.

Waters, Brian, Xingtan Zhang, Wei Zhou, and Shrihari Santosh, 2022, Smart money and investor sentiment, Working paper, University of Colorado Boulder.

Appendices

The appendices provide proofs omitted in the main text.

A Proofs

In this appendix, we provide proofs omitted in the main text.

A.1 Proof of Proposition 1

We conduct the Cholesky decomposition of $\mathbf{\Sigma}^{\{D\}}/r^2 = \mathbf{U}^\top\mathbf{U}$. First, we show that the FFP condition (5) implies the orthogonalization. Condition (5) implies that

$$(\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top)(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}) = (\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})(\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top). \tag{A.1}$$

We denote the eigenvalue decomposition of matrix $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$ as

$$\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top\mathbf{G} = \mathbf{G}\mathbf{\Pi}, \tag{A.2}$$

where $\mathbf{G}$ is an orthonormal matrix and the eigenvalue matrix $\mathbf{\Pi} := \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$, where the eigenvalues $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$ are distinct by assumption.

We define $\mathbf{O} = \mathbf{U}^{-1}\mathbf{G}$ and prove that $\mathbf{O}$ satisfies the orthogonalization. First, by $\mathbf{G}^\top = \mathbf{G}^{-1}$ and $\mathbf{\Sigma}^{\{D\}}/r^2 = \mathbf{U}^\top\mathbf{U}$, we have

$$\mathbf{O}^\top\mathbf{\Sigma}^{\{D\}}\mathbf{O} = r^2\mathbf{G}^\top(\mathbf{U}^{-1})^\top\mathbf{U}^\top\mathbf{U}\mathbf{U}^{-1}\mathbf{G} = r^2\mathbf{I}_N, \tag{A.3}$$

which satisfies (7). Second, by $\mathbf{G}^\top = \mathbf{G}^{-1}$ and (A.2), we have

$$\mathbf{O}^{-1}\mathbf{\Sigma}^{\{f\}}(\mathbf{O}^{-1})^\top = \mathbf{G}^{-1}\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top\mathbf{G} = \mathbf{\Pi}, \tag{A.4}$$

which satisfies (8). Third, we define the matrix

$$\tilde{\boldsymbol{\kappa}} = \mathbf{O}^{-1}\boldsymbol{\kappa}\mathbf{O} = \mathbf{G}^\top\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}\mathbf{G}. \tag{A.5}$$

Using equations (A.1), (A.2), and (A.5), we have

$$\tilde{\boldsymbol{\kappa}}\boldsymbol{\Pi} = \mathbf{G}^\top(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})\mathbf{G}\mathbf{G}^\top(\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top)\mathbf{G} = \mathbf{G}^\top(\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top)\mathbf{G}\mathbf{G}^\top(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})\mathbf{G} = \boldsymbol{\Pi}\tilde{\boldsymbol{\kappa}}. \tag{A.6}$$

Note that $\boldsymbol{\Pi} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ is a diagonal matrix with distinct diagonal entries. Therefore, by $\tilde{\boldsymbol{\kappa}}\boldsymbol{\Pi} = \boldsymbol{\Pi}\tilde{\boldsymbol{\kappa}}$, we know that $\tilde{\boldsymbol{\kappa}}$ must be diagonal. This proves (9).

Conversely, we show that the orthogonalization implies the FFP condition (5). We define $\mathbf{G} = \mathbf{U}\mathbf{O}$. By equation (7), we have

$$\mathbf{G}^\top\mathbf{G} = \mathbf{O}^\top\mathbf{U}^\top\mathbf{U}\mathbf{O} = \mathbf{O}^\top\boldsymbol{\Sigma}^{\{D\}}\mathbf{O}/r^2 = \mathbf{I}_N, \tag{A.7}$$

so the $N \times N$ matrix $\mathbf{G}$ is orthonormal. By equation (8) and $\mathbf{G}$ being orthonormal,

$$\boldsymbol{\Pi} = \mathbf{O}^{-1}\boldsymbol{\Sigma}^{\{f\}}(\mathbf{O}^{-1})^\top = \mathbf{G}^\top\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top\mathbf{G}, \tag{A.8}$$

which is the eigenvalue decomposition of matrix $\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top$. By equation (9), we have

$$\tilde{\boldsymbol{\kappa}} = \mathbf{O}^{-1}\boldsymbol{\kappa}\mathbf{O} = \mathbf{G}^\top\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}\mathbf{G}. \tag{A.9}$$

Because $\tilde{\boldsymbol{\kappa}}$ is diagonal, we have $\tilde{\boldsymbol{\kappa}}\boldsymbol{\Pi} = \boldsymbol{\Pi}\tilde{\boldsymbol{\kappa}}$. Therefore, by (A.8) and (A.9), we have

$$(\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top)(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}) = (\mathbf{G}\boldsymbol{\Pi}\mathbf{G}^\top)(\mathbf{G}\tilde{\boldsymbol{\kappa}}\mathbf{G}^\top) = (\mathbf{G}\tilde{\boldsymbol{\kappa}}\mathbf{G}^\top)(\mathbf{G}\boldsymbol{\Pi}\mathbf{G}^\top) = (\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})(\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top), \tag{A.10}$$

which implies (5).

Additionally, we need to show that each column of $\mathbf{O}$ is unique up to multiplication by -1. Suppose that another $\tilde{\mathbf{O}}$ satisfies the orthogonalization, we can then use the same argument in equations (A.7) and (A.8) to show that $\mathbf{U}\tilde{\mathbf{O}}$ are the full set of eigenvectors of $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$. Because $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$ has distinct eigenvalues, each eigenvector is unique up to multiplication by -1. Therefore, each column of $\mathbf{O}$ is also unique up to multiplication by -1.

A.2 Proof of Proposition 2

We drop the tilde notation for factors in this proof for simplicity. We conjecture that the factor price is linear in state variables,

$$P_{n,t} = \lambda_n f_{n,t} + \zeta D_{n,t} + C_n. \tag{A.11}$$

By equations (11) and (13), the agent's wealth process is

$$\begin{aligned} dW_t &= \left(rW_t - c_t + \sum_{n=1}^{N} y_{n,t}(D_{n,t} - rP_{n,t}) \right) dt + \sum_{n=1}^{N} y_{n,t} dP_{n,t} \\ &= \left(rW_t - c_t + \sum_{n=1}^{N} y_{n,t}(D_{n,t} - rP_{n,t} + \lambda_n \kappa_n f_{n,t} + \zeta \phi_n) \right) dt \\ &\quad + \sum_{n=1}^{N} y_{n,t}(\lambda_n \sigma_n dB^{\{f\}}_{n,t} + r\zeta dB^{\{D\}}_{n,t}) - \sum_{n=1}^{N} y_{n,t} \lambda_n f_{n,t-} dM_t. \end{aligned} \tag{A.12}$$

The Bellman equation is

$$\rho J = \max_{c,\mathbf{y}} -e^{-\gamma c} - r\gamma J\left(rW - c + \sum_{n=1}^{N} y_n(D_n - rP_n + \lambda_n\kappa_n f_n + \zeta\phi_n)\right)$$
$$- \sum_{n=1}^{N} J(b_{1,n} + b_{2,n}f_n)\kappa_n f_n - \sum_{n=1}^{N} J(a_{1,n} + a_{2,n}D_n)\phi_n + \frac{1}{2}r^2\gamma^2 J\sum_{n=1}^{N} y_n^2(\lambda_n^2\sigma_n^2 + r^2\zeta^2)$$
$$+ \sum_{n=1}^{N} r\gamma J(b_{1,n} + b_{2,n}f_n)y_n\lambda_n\sigma_n^2 + \sum_{n=1}^{N} r^3\gamma J(a_{1,n} + a_{2,n}D_n)y_n\zeta$$
$$+ \frac{1}{2}\sum_{n=1}^{N} J\left((b_{1,n} + b_{2,n}f_n)^2 - b_{2,n}\right)\sigma_n^2 + \frac{1}{2}\sum_{n=1}^{N} r^2 J\left((a_{1,n} + a_{2,n}D_n)^2 - a_{2,n}\right)$$
$$+ \eta\left(J\left(W - \sum_{n=1}^{N} y_n\lambda_n f_n, \mathbf{0}, \mathbf{D}\right) - J\right). \tag{A.13}$$

Taking the first-order condition (FOC) with respect to consumption c, we have $\gamma e^{-\gamma c} + r\gamma J = 0$, which implies that the optimal consumption

$$c^* = rW + \left(b_0 + \sum_{n=1}^{N} b_{1,n}f_n + \frac{1}{2}\sum_{n=1}^{N} b_{2,n}f_n^2 + \sum_{n=1}^{N} a_{1,n}D_n + \frac{1}{2}\sum_{n=1}^{N} a_{2,n}D_n^2\right)/\gamma. \tag{A.14}$$

Substituting back into the Bellman equation, we obtain

$$\rho - r = \max_{\mathbf{y}} r\left(b_0 + \sum_{n=1}^{N} b_{1,n}f_n + \frac{1}{2}\sum_{n=1}^{N} b_{2,n}f_n^2 + \sum_{n=1}^{N} a_{1,n}D_n + \frac{1}{2}\sum_{n=1}^{N} a_{2,n}D_n^2\right)$$
$$- r\gamma\sum_{n=1}^{N} y_n(D_n - rP_n + \lambda_n\kappa_n f_n + \zeta\phi_n) + \frac{1}{2}r^2\gamma^2\sum_{n=1}^{N} y_n^2(\lambda_n^2\sigma_n^2 + r^2\zeta^2)$$
$$- \sum_{n=1}^{N}(b_{1,n} + b_{2,n}f_n)\kappa_n f_n - \sum_{n=1}^{N}(a_{1,n} + a_{2,n}D_n)\phi_n$$
$$+ \sum_{n=1}^{N} r\gamma(b_{1,n} + b_{2,n}f_n)y_n\lambda_n\sigma_n^2 + \sum_{n=1}^{N} r^3\gamma(a_{1,n} + a_{2,n}D_n)y_n\zeta$$
$$+ \frac{1}{2}\sum_{n=1}^{N}\left((b_{1,n} + b_{2,n}f_n)^2 - b_{2,n}\right)\sigma_n^2 + \frac{1}{2}\sum_{n=1}^{N} r^2\left((a_{1,n} + a_{2,n}D_n)^2 - a_{2,n}\right)$$
$$+ \eta\left(J\left(W - \sum_{n=1}^{N} y_n\lambda_n f_n, \mathbf{0}, \mathbf{D}\right)/J - 1\right). \tag{A.15}$$

Using the linearization $e^x \approx 1 + x$ for the jump term, we have

$$J\left(W - \sum_{n=1}^{N} y_n \lambda_n f_n, \mathbf{0}, \mathbf{D}\right) / J - 1 \approx r\gamma \sum_{n=1}^{N} y_n \lambda_n f_n + \sum_{n=1}^{N} b_{1,n} f_n + \frac{1}{2} \sum_{n=1}^{N} b_{2,n} f_n^2. \qquad \text{(A.16)}$$

Taking the FOC with respect to y_n, we have

$$\begin{aligned} 0 =& - r\gamma(D_n - rP_n + \lambda_n \kappa_n f_n + \zeta\phi_n) + r^2\gamma^2 y_n(\lambda_n^2 \sigma_n^2 + r^2\zeta^2) \\ & + r\gamma(b_{1,n} + b_{2,n} f_n)\lambda_n \sigma_n^2 + r^3\gamma(a_{1,n} + a_{2,n} D_n)\zeta + \eta r\gamma \lambda_n f_n, \end{aligned} \qquad \text{(A.17)}$$

which gives the optimal investment

$$y_n^* = \frac{(D_n - rP_n + \lambda_n \kappa_n f_n + \zeta\phi_n) - (b_{1,n} + b_{2,n} f_n)\lambda_n \sigma_n^2 - (a_{1,n} + a_{2,n} D_n)\zeta r^2 - \eta\lambda_n f_n}{r\gamma(\lambda_n^2 \sigma_n^2 + r^2\zeta^2)}. \qquad \text{(A.18)}$$

Substituting back into the Bellman equation, we obtain

$$\begin{aligned} \rho - r =& r\left(b_0 + \sum_{n=1}^{N} b_{1,n} f_n + \frac{1}{2}\sum_{n=1}^{N} b_{2,n} f_n^2 + \sum_{n=1}^{N} a_{1,n} D_n + \frac{1}{2}\sum_{n=1}^{N} a_{2,n} D_n^2\right) \\ & - \frac{1}{2} r^2\gamma^2 \sum_{n=1}^{N} (y_n^*)^2 (\lambda_n^2 \sigma_n^2 + r^2\zeta^2) - \sum_{n=1}^{N} (b_{1,n} + b_{2,n} f_n)\kappa_n f_n - \sum_{n=1}^{N} (a_{1,n} + a_{2,n} D_n)\phi_n \\ & + \frac{1}{2}\sum_{n=1}^{N} \left((b_{1,n} + b_{2,n} f_n)^2 - b_{2,n}\right)\sigma_n^2 + \frac{1}{2}\sum_{n=1}^{N} \left((a_{1,n} + a_{2,n} D_n)^2 - a_{2,n}\right) r^2 \\ & + \eta \sum_{n=1}^{N} b_{1,n} f_n + \frac{\eta}{2}\sum_{n=1}^{N} b_{2,n} f_n^2. \end{aligned} \qquad \text{(A.19)}$$

Note that

$$\begin{aligned} & (D_n - rP_n + \lambda_n \kappa_n f_n + \zeta\phi_n) - (b_{1,n} + b_{2,n} f_n)\lambda_n \sigma_n^2 - (a_{1,n} + a_{2,n} D_n)\zeta r^2 - \eta\lambda_n f_n \\ =& D_n(1 - r\zeta - a_{2,n}\zeta r^2) + f_n \lambda_n(\kappa_n - r - b_{2,n}\sigma_n^2 - \eta) - rC_n + \zeta\phi_n - b_{1,n}\lambda_n \sigma_n^2 - a_{1,n}\zeta r^2. \end{aligned} \qquad \text{(A.20)}$$

Matching coefficients on $D_n f_n$, D_n, and D_n^2 terms, we obtain $\zeta = 1/r$, $a_{1,n} = 0$, and $a_{2,n} = 0$.

That is, the value function does not depend on D_n. Market clearing implies

$$y_n = \frac{f_n \lambda_n(\kappa_n - r - b_{2,n}\sigma_n^2 - \eta) - rC_n + \phi_n/r - b_{1,n}\lambda_n\sigma_n^2}{r\gamma(\lambda_n^2\sigma_n^2 + 1)} = \frac{S_n - f_n}{\mu}, \tag{A.21}$$

which implies that

$$\mu\lambda_n(\kappa_n - r - b_{2,n}\sigma_n^2 - \eta) + r\gamma(\lambda_n^2\sigma_n^2 + 1) = 0, \tag{A.22}$$

$$\mu(rC_n - \phi_n/r + b_{1,n}\lambda_n\sigma_n^2) + S_n r\gamma(\lambda_n^2\sigma_n^2 + 1) = 0. \tag{A.23}$$

Therefore, the Bellman equation simplifies to

$$\begin{aligned}\rho - r = & r\left(b_0 + \sum_{n=1}^{N} b_{1,n} f_n + \frac{1}{2}\sum_{n=1}^{N} b_{2,n} f_n^2\right) - \frac{1}{2}\sum_{n=1}^{N}(\lambda_n^2\sigma_n^2 + 1)r^2\gamma^2(S_n - f_n)^2/\mu^2 \\ & - \sum_{n=1}^{N}(b_{1,n} + b_{2,n}f_n)\kappa_n f_n + \frac{1}{2}\sum_{n=1}^{N}\left((b_{1,n} + b_{2,n}f_n)^2 - b_{2,n}\right)\sigma_n^2 \\ & + \eta\sum_{n=1}^{N} b_{1,n} f_n + \frac{\eta}{2}\sum_{n=1}^{N} b_{2,n} f_n^2. \end{aligned} \tag{A.24}$$

Matching coefficients, we have

$$\frac{1}{2} r b_{2,n} - \frac{1}{2}(\lambda_n^2\sigma_n^2 + 1)r^2\gamma^2/\mu^2 - b_{2,n}\kappa_n + \frac{1}{2} b_{2,n}^2\sigma_n^2 + \frac{1}{2}\eta b_{2,n} = 0, \tag{A.25}$$

$$r b_{1,n} + (\lambda_n^2\sigma_n^2 + 1)r^2\gamma^2 S_n/\mu^2 - b_{1,n}\kappa_n + b_{1,n}b_{2,n}\sigma_n^2 + \eta b_{1,n} = 0, \tag{A.26}$$

$$\rho - r = r b_0 - \frac{1}{2}\sum_{n=1}^{N}(\lambda_n^2\sigma_n^2 + 1)r^2\gamma^2 S_n^2/\mu^2 + \frac{1}{2}\sum_{n=1}^{N}(b_{1,n}^2 - b_{2,n})\sigma_n^2. \tag{A.27}$$

Equations (A.22) and (A.25) jointly solve λ_n and $b_{2,n}$. Once λ_n and $b_{2,n}$ are known, equation (A.23) solves for C_n, equation (A.26) solves for $b_{1,n}$, and equation (A.27) solves for b_0.

To solve equations (A.22) and (A.25), we have

$$\mu\lambda_n(r + b_{2,n}\sigma_n^2 + \eta - \kappa_n) = r\gamma(\lambda_n^2\sigma_n^2 + 1), \tag{A.28}$$

$$\mu^2 b_{2,n}(r + b_{2,n}\sigma_n^2 + \eta - 2\kappa_n) = r^2\gamma^2(\lambda_n^2\sigma_n^2 + 1). \tag{A.29}$$

We define $x_n := b_{2,n}\mu/(r\gamma\lambda_n)$ and then have

$$(x_n - 1)r\gamma\lambda_n^2\sigma_n^2 + \mu\lambda_n(r + \eta - \kappa_n) = r\gamma, \tag{A.30}$$

$$(x_n^2 - 1)r\gamma\lambda_n^2\sigma_n^2 + \mu\lambda_n x_n(r + \eta - 2\kappa_n) = r\gamma. \tag{A.31}$$

Multiplying (A.30) by x_n and then subtracting (A.31), we have

$$(1 - x_n)r\gamma\lambda_n^2\sigma_n^2 + \mu\lambda_n x_n\kappa_n = (x_n - 1)r\gamma. \tag{A.32}$$

Using (A.30) to solve for $x_n - 1$ and then plugging into (A.32), we obtain

$$\mu\lambda_n\kappa_n\left(1 + \frac{r\gamma - \mu\lambda_n(r + \eta - \kappa_n)}{r\gamma\lambda_n^2\sigma_n^2}\right) = \frac{r\gamma - \mu\lambda_n(r + \eta - \kappa_n)}{r\gamma\lambda_n^2\sigma_n^2}\left(r\gamma + r\gamma\lambda_n^2\sigma_n^2\right), \tag{A.33}$$

which simplifies to the equilibrium condition (17),

$$\kappa_n(\kappa_n - r - \eta) = \frac{r\gamma}{\mu}\left(\frac{r\gamma}{\mu} - \lambda_n(r + \eta)\right)\left(\frac{1}{\lambda_n^2} + \sigma_n^2\right). \tag{A.34}$$

We now provide verification for optimality. We first prove the following transversality condition,

$$\lim_{t\to\infty} e^{-rt}\mathbb{E}(J(W_t, \mathbf{f}_t, \mathbf{D}_t)) = 0. \tag{A.35}$$

Proof. By equation (15), when $\tilde{\kappa}_n < (r + \eta)/2$, it suffices to show that $b_{2,n} > 0$ for $n = 1, 2, \cdots N$. Note that $b_{2,n}$ is solved by equations (A.28) and (A.29). We define $b(\kappa)$ as the solution $b_{2,n}$ for the following set of two equations

$$\mu\lambda_n(r + b_{2,n}\sigma_n^2 + \eta - \kappa) = r\gamma(\lambda_n^2\sigma_n^2 + 1/r^2), \tag{A.36}$$

$$\mu^2 b_{2,n}(r + b_{2,n}\sigma_n^2 + \eta - 2\kappa) = r^2\gamma^2(\lambda_n^2\sigma_n^2 + 1/r^2). \tag{A.37}$$

Note that equation (17) implies that $\lambda_n > 0$. When $\kappa < 0$, it must be that $b_{2,n} > 0$. Because otherwise, from (A.37), $b_{2,n} < 0$ and $r + b_{2,n}\sigma_n^2 + \eta - 2\kappa < 0$ must hold simultaneously. However, if $r+b_{2,n}\sigma_n^2+\eta-2\kappa < 0$, then $r+b_{2,n}\sigma_n^2+\eta-\kappa < r+b_{2,n}\sigma_n^2+\eta-2\kappa < 0$ because $\kappa < 0$, leading to a contradiction because the right-hand side of (A.36) is positive. When $\kappa = 0$, equation (A.36) implies that $r + b_{2,n}\sigma_n^2 + \eta > 0$ and therefore equation (A.37) implies that $b_{2,n} > 0$, i.e., $b(\kappa) > 0$ when $\kappa = 0$. Now we claim that $b(\kappa) > 0$ for any $0 < \kappa < (r + \eta)/2$. Suppose not, because of continuity, there must exist some $\kappa' \in (0, (r + \eta)/2)$ such that $b(\kappa') = 0$. In this case, the left-hand side of (A.37) is zero, but the right-hand side is strictly positive. This is a contradiction. □

We now use the transversality condition (A.35) to verify optimality. Specifically, we show that the solution $J(W, \mathbf{f}, \mathbf{D})$ from (15) satisfies

$$J(W_0, \mathbf{f}_0, \mathbf{D}_0) = \max_{c_t, \mathbf{y}_t} \mathbb{E}\left[\int_0^\infty -e^{-\rho t - \gamma c_t}\, dt\right]. \tag{A.38}$$

First, note that for any admissible controls $\{c_t, \mathbf{y}_t |\, t \geq 0\}$, we have by (A.13),

$$\rho J(W_t, \mathbf{f}_t, \mathbf{D}_t) \geq -e^{-\gamma c_t} + \mathcal{D}J(W_t, \mathbf{f}_t, \mathbf{D}_t), \tag{A.39}$$

where $\mathcal{D}J(W_t, \mathbf{f}_t, \mathbf{D}_t)$ is the infinitesimal generator (i.e., all remaining terms of (A.13)). By

Ito's Lemma, we have (note that $a_{1,n} = a_{2,n} = 0$)

$$
\begin{aligned}
e^{-\rho T} J(W_T, \mathbf{f}_T, \mathbf{D}_T) - J(W_0, \mathbf{f}_0, \mathbf{D}_0) = &\int_0^T e^{-\rho t}(-\rho J(W_t, \mathbf{f}_t, \mathbf{D}_t) + \mathcal{D}J(W_t, \mathbf{f}_t, \mathbf{D}_t))dt \\
&- \int_0^T r\gamma J(W_t, \mathbf{f}_t, \mathbf{D}_t) \sum_{n=1}^N y_{n,t}(\lambda_n \sigma_n dB_{n,t}^{\{f\}} + \zeta dB_{n,t}^{\{D\}}) \\
&- \int_0^T \sum_{n=1}^N J(W_t, \mathbf{f}_t, \mathbf{D}_t)(b_{1,n} + b_{2,n} f_n)\sigma_n dB_{n,t}^{\{f\}}.
\end{aligned}
\tag{A.40}
$$

Assuming local martingales are martingales under admissible controls and taking expectation on both sides, we have

$$
e^{-\rho T}\mathbb{E}(J(W_T, \mathbf{f}_T, \mathbf{D}_T)) - J(W_0, \mathbf{f}_0, \mathbf{D}_0) = \mathbb{E}\left(\int_0^T e^{-\rho t}(-\rho J(W_t, \tilde{\mathbf{f}}_t, \tilde{\mathbf{D}}_t) + \mathcal{D}J(W_t, \tilde{\mathbf{f}}_t, \tilde{\mathbf{D}}_t))dt\right). \tag{A.41}
$$

Taking T to ∞ and using (A.39), the transversality condition (A.35), and Fatou's Lemma, we have

$$
J(W_0, \mathbf{f}_0, \mathbf{D}_0) \geq \mathbb{E}\left[\int_0^\infty -e^{-\rho t - \gamma c_t}\, dt\right]. \tag{A.42}
$$

Second, we re-run all the arguments in the first step, but with the optimal control $\{c_t^*, \mathbf{y}_t^* | \, t \geq 0\}$, and obtain

$$
\rho J(W_t, \mathbf{f}_t, \mathbf{D}_t) = -e^{-\gamma c_t^*} + \mathcal{D}J(W_t, \mathbf{f}_t, \mathbf{D}_t). \tag{A.43}
$$

The same argument implies that the optimality is achieved under $\{c_t^*, \mathbf{y}_t^* | \, t \geq 0\}$,

$$
J(W_0, \mathbf{f}_0, \mathbf{D}_0) = \mathbb{E}\left[\int_0^\infty -e^{-\rho t - \gamma c_t^*}\, dt\right]. \tag{A.44}
$$

A.3 Proof of Proposition 3

First, we show that if $\tilde{\kappa}_n < (r+\eta)/2$, then $\lim_{t\to\infty} e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2) = 0$.

Because $\tilde{\kappa}_n < (r+\eta)/2$, there exists an $r_0 \in (0, r)$ such that $\tilde{\kappa}_n < (r_0+\eta)/2 < (r+\eta)/2$. We perform a change of variable by letting, for $t \geq 0$,

$$g_t = e^{-\frac{r_0 t}{2\eta}} \tilde{f}_{n,t/\eta}. \tag{A.45}$$

By equation (11), the dynamics of g_t is given by

$$dg_t = \frac{1}{\eta}\left(\tilde{\kappa}_n - \frac{r_0}{2}\right) g_t dt + e^{-\frac{r_0 t}{2\eta}} \frac{\sigma}{\sqrt{\eta}} d\tilde{B}_t - g_{t-} d\tilde{M}_t, \tag{A.46}$$

where $\{\tilde{B}_t | t \geq 0\}$ is a standard Brownian motion and $\{\tilde{M}_t | t \geq 0\}$ is a standard rate-one Poisson process. We next show that

$$\limsup_{t\to\infty} \mathbb{E}(g_t^2) < \infty. \tag{A.47}$$

We denote $0 < \tau_1 < \tau_2 < \ldots$ as the jump time epochs for $\{\tilde{M}_t | t \geq 0\}$. Define $\tau_0 = 0$. Let $\mu_0 = (\tilde{\kappa}_n - r_0/2)/\eta$ and $\sigma_0 = \sigma/\sqrt{\eta}$ in this proof. For any t such that $t < \tau_1$, we have when $\mu_0 \neq -r_0/(2\eta)$,

$$\mathbb{E}(g_t^2) = e^{2\mu_0 t} g_0^2 + e^{2\mu_0 t}\sigma_0^2 \int_0^t e^{-(2\mu_0 + \frac{r_0}{\eta})s} ds = \frac{\sigma_0^2}{2\mu_0 + \frac{r_0}{\eta}}(e^{2\mu_0 t} - e^{-\frac{r_0 t}{\eta}}), \tag{A.48}$$

where we use $g_0 = \tilde{f}_{n,0} = 0$. We have that when $\mu_0 = -r_0/(2\eta)$,

$$\mathbb{E}(g_t^2) = e^{2\mu_0 t}\sigma_0^2 t. \tag{A.49}$$

Note that $\mu_0 = -r_0/(2\eta) < 0$. Consider a function $h(t) = e^{2\mu_0 t}\sigma_0^2 t$ defined on $t \in [0, \infty)$. It is evident that $h'(t) = \sigma_0^2 e^{2\mu_0 t}(1 + 2\mu_0 t)$ takes a negative sign when $t > -1/(2\mu_0)$ and takes a positive sign when $t \in [0, -1/(2\mu_0))$. Therefore $h(t) \leq h(-1/(2\mu_0)) = -e^{-1}\sigma_0^2/(2\mu_0)$ for any $t \in [0, \infty)$. As a result, when $\mu_0 = -r_0/(2\eta)$, for any $t \in [0, \infty)$,

$$\mathbb{E}(g_t^2) = e^{2\mu_0 t}\sigma_0^2 t \le \frac{e^{-1}\sigma_0^2\eta}{r_0}. \tag{A.50}$$

When $\mu_0 < -r_0/(2\eta)$, we have $2\mu_0 + r_0/\eta < 0$ and therefore

$$\mathbb{E}(g_t^2) = e^{-\frac{r_0 t}{\eta}} \frac{\sigma_0^2}{2\mu_0 + \frac{r_0}{\eta}}(e^{(2\mu_0 + \frac{r_0}{\eta})t} - 1) \le e^{-\frac{r_0 t}{\eta}} \frac{\sigma_0^2}{-(2\mu_0 + \frac{r_0}{\eta})} \le \frac{\sigma_0^2}{-(2\mu_0 + \frac{r_0}{\eta})}. \tag{A.51}$$

When $\mu_0 > -r_0/(2\eta)$, we have

$$\mathbb{E}(g_t^2) = \frac{\sigma_0^2}{2\mu_0 + \frac{r_0}{\eta}}(e^{2\mu_0 t} - e^{-\frac{r_0 t}{\eta}}) \le \frac{\sigma_0^2}{2\mu_0 + \frac{r_0}{\eta}} e^{2\mu_0 t}. \tag{A.52}$$

Note that $\mu_0 = (\tilde{\kappa}_n - r_0/2)/\eta < 1/2$. We therefore have for any t such that $t < \tau_1$,

$$\mathbb{E}(g_t^2) \le \mathbb{E}(g_{\tau_1-}^2) \le \frac{\sigma_0^2}{2\mu_0 + \frac{r_0}{\eta}} \mathbb{E}(e^{2\mu_0\tau_1}) = \frac{\sigma_0^2}{(2\mu_0 + \frac{r_0}{\eta})(1 - 2\mu_0)}. \tag{A.53}$$

Above we have discussed t such that $t < \tau_1$. If $t > \tau_1$, for any t, there must exist some $i \ge 1$, such that $\tau_i \le t < \tau_{i+1}$. In this case, because $g_{\tau_i} = 0$, we can use similar arguments on $\mathbb{E}(g_t^2)$. As a result, we have shown that for any $\tilde{\kappa}_n < (r + \eta)/2$, we have $\limsup_{t\to\infty} \mathbb{E}(g_t^2) < \infty$ and therefore

$$\limsup_{t\to\infty} e^{-\frac{(r-r_0)t}{\eta}} \mathbb{E}(g_t^2) = 0. \tag{A.54}$$

Using definition (A.45), we have proved $\lim_{t\to\infty} e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2) = 0$.

Next, we show that if $\tilde{\kappa}_n \ge (r + \eta)/2$, then $\limsup_{t\to\infty} e^{-rt}\mathbb{E}(\tilde{f}_{n,t}^2) = \infty$.

We perform a change of variable by letting, for $t \ge 0$,

$$g_t = \tilde{f}_{n,t/\eta}. \tag{A.55}$$

The dynamics of g_t is then given by

$$dg_t = \frac{\tilde{\kappa}_n}{\eta} g_t dt + \frac{\sigma}{\sqrt{\eta}} d\tilde{B}_t - g_{t-} d\tilde{M}_t, \tag{A.56}$$

where $\{\tilde{B}_t | t \geq 0\}$ is a standard Brownian motion and $\{\tilde{M}_t | t \geq 0\}$ is a standard rate-one Poisson process. We denote $0 < \tau_1 < \tau_2 < \ldots$ as the jump time epochs for $\{\tilde{M}_t | t \geq 0\}$. Let $\mu_0 = \tilde{\kappa}_n/\eta$ and $\sigma_0 = \sigma/\sqrt{\eta}$ in this proof. For any fixed t such that $t < \tau_1$, we have that

$$\mathbb{E}(g_t^2) = e^{2\mu_0 t} g_0^2 + e^{2\mu_0 t} \sigma_0^2 \int_0^t e^{-2\mu_0 s} ds = \frac{\sigma_0^2}{2\mu_0}(e^{2\mu_0 t} - 1), \tag{A.57}$$

where we use $g_0 = \tilde{f}_{n,0} = 0$. Therefore, we have

$$\mathbb{E}(e^{-rt/\eta} g_t^2) = \frac{\sigma_0^2}{2\mu_0} e^{-rt/\eta}(e^{2\mu_0 t} - 1). \tag{A.58}$$

Sending t to τ_1-, we have

$$\liminf_{t \to \tau_1 -} \mathbb{E}(e^{-rt/\eta} g_t^2) \geq \frac{\sigma_0^2}{2\mu_0} \mathbb{E}(e^{(2\mu_0 - r/\eta)\tau_1} - 1). \tag{A.59}$$

Note that $2\mu_0 - r/\eta = 2\tilde{\kappa}_n/\eta - r/\eta \geq 1$. Because τ_1 is an exponential random variable with rate parameter one, we have $\mathbb{E}(e^{(2\mu_0 - r/\eta)\tau_1}) = \infty$. It is therefore the case that when $\tilde{\kappa}_n \geq (r+\eta)/2$, $\lim_{t \to \tau_1 -} \mathbb{E}(e^{-rt/\eta} g_t^2) = \infty$. Similar arguments apply to each interval $[\tau_i, \tau_{i+1})$ for $i = 1, 2, \ldots$ so that $\lim_{t \to \tau_i -} \mathbb{E}(e^{-rt/\eta} g_t^2) = \infty$. Because $\lim_{i \to \infty} \tau_i = \infty$ almost surely, we have that $\limsup_{t \to \infty} e^{-rt/\eta} \mathbb{E}(g_t^2) = \infty$. Therefore, by (A.55), we have

$$\limsup_{t \to \infty} e^{-rt} \mathbb{E}(\tilde{f}_{n,t}^2) = \limsup_{t \to \infty} \mathbb{E}(e^{-rt/\eta} g_t^2) = \infty. \tag{A.60}$$

A.4 Proof of Proposition 4

We first show that λ_n is strictly increasing in $\tilde{\kappa}_n$. We define $\lambda^* := \gamma/(\mu(1+\eta/r))$. From equation (17), we see that if $\tilde{\kappa}_n < 0$, then $0 < \lambda_n < \lambda^*$; if $\tilde{\kappa}_n = 0$, then $\lambda_n = \lambda^*$; if $\tilde{\kappa}_n \in (0, (r+\eta)/2)$, then $\lambda_n > \lambda^*$. When $\tilde{\kappa}_n < 0$, the LHS of (17) is strictly decreasing in $\tilde{\kappa}_n$ and the RHS of (17) is strictly decreasing in λ_n. Therefore, λ_n is strictly increasing in $\tilde{\kappa}_n$ when $\tilde{\kappa}_n < 0$.

For the case of $\tilde{\kappa}_n \in (0, (r+\eta)/2)$, we have by (17),

$$\frac{\tilde{\kappa}_n(r+\eta-\tilde{\kappa}_n)}{(r+\eta)^2} = \lambda^*(\lambda_n - \lambda^*)\left(1/\lambda_n^2 + \sigma_n^2\right). \tag{A.61}$$

The LHS is between 0 and 1/4 and is strictly increasing in $\tilde{\kappa}_n$. Note that if $\lambda_n = 2\lambda^*$, then

$$\lambda^*(\lambda_n - \lambda^*)\left(1/\lambda_n^2 + \sigma_n^2\right) > 1/4, \tag{A.62}$$

and the function $\lambda^*(\lambda_n - \lambda^*)\left(1/\lambda_n^2 + \sigma_n^2\right)$ is strictly increasing in $\lambda_n \in [\lambda^*, 2\lambda^*]$. Therefore, λ_n is strictly increasing in $\tilde{\kappa}_n \in (0, (r+\eta)/2)$.

Next, we consider the comparative statics of λ_n with respect to σ_n. If $\tilde{\kappa}_n = 0$, then $\lambda_n = \lambda^*$ does not depend on σ_n. If $\tilde{\kappa}_n \in (0, (r+\eta)/2)$, the RHS of (A.61) is strictly increasing in λ_n and σ_n. Therefore, λ_n is strictly decreasing in σ_n. Finally, if $\tilde{\kappa}_n < 0$, the RHS of (17) is strictly decreasing in λ_n and strictly increasing in σ_n, so λ_n is strictly increasing in σ_n.

Next, we show that λ_n is strictly increasing in γ. If $\tilde{\kappa}_n < 0$, then the RHS of (17) is strictly decreasing in λ_n and strictly increasing in γ, so λ_n is strictly increasing in γ. If $\tilde{\kappa}_n \in (0, (r+\eta)/2)$, then the RHS of (A.61) is strictly increasing in λ_n and strictly decreasing in λ^*. The reason for the latter is because $\lambda^* \in [\lambda_n/2, \lambda_n]$. Therefore, λ_n is strictly increasing in λ^* and thus in γ.

Finally, the fact that λ_n is strictly decreasing in μ follows effectively the same argument as the comparative statics with respect to γ, so is omitted.

A.5 Proof of Proposition 6

The proof of Proposition 6 is a straightforward extension of that of Proposition 2, so we just note the key differences.

First, using flow dynamics (22) and price conjecture (23), we obtain

$$d\tilde{f}_{n,t} = \hat{\kappa}_n \tilde{f}_{n,t} dt + \sigma_n d\tilde{B}_{n,t}^{\{f\}} - \tilde{f}_{n,t-} dM_{n,t}, \tag{A.63}$$

where the effective flow predictability is $\hat{\kappa}_n = \tilde{\kappa}_n + \chi_n \lambda_n$.

Second, the last term of the Bellman equation (A.13) needs to be modified to account for the differential reversion speed η_n,

$$\begin{aligned}\rho J = \max_{c,\mathbf{y}} -e^{-\gamma c} - r\gamma J \left(rW - c + \sum_{n=1}^N y_n (D_n - rP_n + \lambda_n \hat{\kappa}_n f_n + \zeta \phi_n) \right) \\ - \sum_{n=1}^N J(b_{1,n} + b_{2,n} f_n) \hat{\kappa}_n f_n - \sum_{n=1}^N J(a_{1,n} + a_{2,n} D_n)\phi_n + \frac{1}{2} r^2 \gamma^2 J \sum_{n=1}^N y_n^2 (\lambda_n^2 \sigma_n^2 + r^2 \zeta^2) \\ + \sum_{n=1}^N r\gamma J(b_{1,n} + b_{2,n} f_n) y_n \lambda_n \sigma_n^2 + \sum_{n=1}^N r^3 \gamma J (a_{1,n} + a_{2,n} D_n) y_n \zeta \\ + \frac{1}{2} \sum_{n=1}^N J \left((b_{1,n} + b_{2,n} f_n)^2 - b_{2,n} \right) \sigma_n^2 + \frac{1}{2} \sum_{n=1}^N r^2 J \left((a_{1,n} + a_{2,n} D_n)^2 - a_{2,n} \right) \\ + \sum_{n=1}^N \eta_n \left(J \left(W - y_n \lambda_n f_n, \hat{\mathbf{f}}_n, \mathbf{D} \right) - J \right),\end{aligned} \tag{A.64}$$

where $\hat{\mathbf{f}}_n$ is the same as the $\mathbf{f} = (f_1, f_2, \ldots, f_N)^\top$ except that the n-th entry is zero. Using the linearization $e^x \approx 1 + x$ for the jump term, we have

$$\sum_{n=1}^N \eta_n \left(J \left(W - y_n \lambda_n f_n, \hat{\mathbf{f}}_n, \mathbf{D} \right) / J - 1 \right) \approx \sum_{n=1}^N \eta_n (r\gamma y_n \lambda_n f_n + b_{1,n} f_n + \frac{1}{2} b_{2,n} f_n^2). \tag{A.65}$$

It is then straightforward to see that the problem is still separable for the N factors, where each factor has its own reversion speed η_n. We can therefore derive (25).

Finally, we note that equations (24) and (25) determine a unique λ_n. This is because $\tilde{\kappa}_n$ is strictly increasing in λ_n in equation (25) by Proposition 4 and $\tilde{\kappa}_n$ is decreasing in λ_n in equation (24).

A.6 Proof of Proposition 7

We first simplify equation (28). By definition, (28) is equivalent to

$$\begin{aligned}&\mathbf{X}^\top \mathrm{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \mathrm{diag}(\tilde{\kappa}_1, \tilde{\kappa}_2, \dots, \tilde{\kappa}_N)\mathbf{X}\\ =&\mathbf{X}^\top \mathrm{diag}(\tilde{\kappa}_1, \tilde{\kappa}_2, \dots, \tilde{\kappa}_N)\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \mathrm{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)\mathbf{X},\end{aligned} \tag{A.66}$$

because $\mathbf{X}^\top\mathbf{X}$ is an invertible $K \times K$ matrix. That (A.66) holds for any $\tilde{\kappa}_1, \tilde{\kappa}_2, \dots, \tilde{\kappa}_N$ and $\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2$ is equivalent to

$$\mathbf{X}^\top \boldsymbol{\iota}_n \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\iota}_m \mathbf{X} = \mathbf{X}^\top \boldsymbol{\iota}_m \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\iota}_n \mathbf{X}, \tag{A.67}$$

for any $n, m = 1, 2, \dots, N$, where $\boldsymbol{\iota}_n$ is an $N \times N$ matrix with only (n,n)-th entry being 1 and all other entries being 0. Note that (A.67) is equivalent to

$$\mathbf{x}_{n,\cdot}^\top \mathbf{x}_{n,\cdot}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_{m,\cdot}^\top \mathbf{x}_{m,\cdot} = \mathbf{x}_{m,\cdot}^\top \mathbf{x}_{m,\cdot}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_{n,\cdot}^\top \mathbf{x}_{n,\cdot}, \tag{A.68}$$

where $\mathbf{x}_{n,\cdot}$ is the n-th row of matrix $\mathbf{X}$.

With condition (28) simplified as (A.68), we first show that if $\mathbf{X} = \mathbf{ZH}$ for some $N \times K$ full-column-rank matrix $\mathbf{Z}$ that has at most one non-zero element each row and some $K \times K$ invertible matrix $\mathbf{H}$, then (A.68) holds. Because of the special form of the matrix $\mathbf{Z}$. we have that for any $n, m = 1, 2, \dots, N$,

$$\mathbf{z}_{n,\cdot}^\top \mathbf{z}_{n,\cdot}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{z}_{m,\cdot}^\top \mathbf{z}_{m,\cdot} = \mathbf{z}_{m,\cdot}^\top \mathbf{z}_{m,\cdot}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{z}_{n,\cdot}^\top \mathbf{z}_{n,\cdot}. \tag{A.69}$$

Because $\mathbf{H}$ is invertible, we have

$$\mathbf{H}^\top \mathbf{z}_{n,\cdot}^\top \mathbf{z}_{n,\cdot} \mathbf{H}(\mathbf{H}^\top \mathbf{Z}^\top \mathbf{Z}\mathbf{H})^{-1}\mathbf{H}^\top \mathbf{z}_{m,\cdot}^\top \mathbf{z}_{m,\cdot} \mathbf{H} = \mathbf{H}^\top \mathbf{z}_{m,\cdot}^\top \mathbf{z}_{m,\cdot} \mathbf{H}(\mathbf{H}^\top \mathbf{Z}^\top \mathbf{Z}\mathbf{H})^{-1}\mathbf{H}^\top \mathbf{z}_{n,\cdot}^\top \mathbf{z}_{n,\cdot} \mathbf{H}, \quad \text{(A.70)}$$

which implies (A.68) by definition $\mathbf{X} = \mathbf{Z}\mathbf{H}$.

Conversely, we assume that (A.68) holds and derive the form of $\mathbf{X}$. First, we do Cholesky decomposition and obtain $\mathbf{U}^\top \mathbf{U} = \mathbf{X}^\top \mathbf{X}$, where $\mathbf{U}$ is a $K \times K$ upper-triangular invertible matrix. We define the $N \times K$ matrix $\mathbf{Y} := \mathbf{X}\mathbf{U}^{-1}$. Because matrix $\mathbf{Y}$ has rank K, we can pick K linearly independent rows of $\mathbf{Y}$. Without loss of generality, we assume that these are the first K rows and we define the resulting $K \times K$ matrix as $\tilde{\mathbf{Y}} = (\mathbf{y}_{1,\cdot}; \mathbf{y}_{2,\cdot}; \ldots; \mathbf{y}_{K,\cdot})$. Second, we conduct Cholesky decomposition and obtain $\mathbf{L}\mathbf{L}^\top = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top$, where $\mathbf{L}$ is a $K \times K$ lower-triangular invertible matrix.

We now define the matrix $\mathbf{H} := \mathbf{L}^{-1}\tilde{\mathbf{Y}}\mathbf{U}$ and $\mathbf{Z} := \mathbf{X}\mathbf{H}^{-1}$. We then have

$$\mathbf{Z}^\top \mathbf{Z} = \mathbf{L}^\top (\tilde{\mathbf{Y}}^{-1})^\top (\mathbf{U}^{-1})^\top \mathbf{X}^\top \mathbf{X}\mathbf{U}^{-1}\tilde{\mathbf{Y}}^{-1}\mathbf{L} = (\mathbf{L}^{-1}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top (\mathbf{L}^\top)^{-1})^{-1} = \mathbf{I}_K. \quad \text{(A.71)}$$

Moreover, we have

$$\mathbf{Z} = \mathbf{X}\mathbf{U}^{-1}\tilde{\mathbf{Y}}^{-1}\mathbf{L} = \mathbf{Y}\tilde{\mathbf{Y}}^{-1}\mathbf{L}. \quad \text{(A.72)}$$

In particular, the first K rows of $\mathbf{Z}$ is $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{-1}\mathbf{L} = \mathbf{L}$, a lower-triangular matrix with non-zero diagonal entry.

Using equation (A.68) and $\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_K$, we have

$$\mathbf{z}_{n,\cdot}^\top \mathbf{z}_{n,\cdot} \mathbf{z}_{m,\cdot}^\top \mathbf{z}_{m,\cdot} = \mathbf{z}_{m,\cdot}^\top \mathbf{z}_{m,\cdot} \mathbf{z}_{n,\cdot}^\top \mathbf{z}_{n,\cdot}. \quad \text{(A.73)}$$

Because the first K rows of $\mathbf{Z}$ is lower-triangular, we know that $\mathbf{z}_{1,\cdot}$ has only the first entry being non-zero and $\mathbf{z}_{2,\cdot}$ can have both the first and second entries being non-zero. Because

$$\mathbf{z}_{1,\cdot}^\top \mathbf{z}_{1,\cdot} \mathbf{z}_{2,\cdot}^\top \mathbf{z}_{2,\cdot} = \mathbf{z}_{2,\cdot}^\top \mathbf{z}_{2,\cdot} \mathbf{z}_{1,\cdot}^\top \mathbf{z}_{1,\cdot}, \quad \text{(A.74)}$$

we see that $\mathbf{z}_{2,\cdot}^{\top}\mathbf{z}_{2,\cdot}$ must be a diagonal matrix, which implies that $\mathbf{z}_{2,\cdot}$ only has the second entry being non-zero. Using this logic inductively, we see that for $n = 1, 2, \ldots, K$, $\mathbf{z}_{n,\cdot}$ only has the n-th entry being non-zero. For the remaining rows $n = K+1, K+2, \ldots, N$, $\mathbf{z}_{n,\cdot}$ can have at most one non-zero entry. Therefore, $\mathbf{X} = \mathbf{Z}\mathbf{H}$, where $\mathbf{Z}$ is a full-column-rank matrix that has at most one non-zero element in each row and $\mathbf{H}$ is invertible.

Online Appendix of

"A Dynamic Factor Model of Price Impacts"

The online appendix provides additional theoretical results omitted in the paper.

A General Theory of Factorizable Flow Predictability

We present the general theory of FFP when $\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top$ can have duplicate eigenvalues. We denote the eigenvalues of matrix $\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top$ as $\sigma_1^2 > \sigma_2^2 > \cdots > \sigma_J^2 > 0$, where each eigenvalue π_j having r_j degrees of duplication for $j = 1, 2, \ldots, J$ and $\sum_{j=1}^J r_j = N$. In the general case, Proposition 1 of the main text needs to be modified.

PROPOSITION O.1. *The FFP holds if and only if there exist N factor portfolios $\mathbf{o}_n = (o_{1,n}, o_{2,n}, \ldots, o_{N,n})^\top$, which we denote as an $N \times N$ matrix $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_N)$, satisfying*

i. factor decomposition of flow

$$\mathbf{f}_t = \sum_{n=1}^{N} \mathbf{o}_n \tilde{f}_{n,t} = \mathbf{O}\tilde{\mathbf{f}}_t, \tag{OA.1}$$

where $\tilde{f}_{n,t}$ is the flow into factor n up to time t and $\tilde{\mathbf{f}}_t = (\tilde{f}_{1,t}, \tilde{f}_{2,t}, \ldots, \tilde{f}_{N,t})^\top$.

ii. uncorrelated dividends

$$\mathbf{O}^\top \boldsymbol{\Sigma}^{\{D\}} \mathbf{O} = r^2 \mathbf{I}_N. \tag{OA.2}$$

iii. uncorrelated flows

$$\mathbf{O}^{-1}\boldsymbol{\Sigma}^{\{f\}}(\mathbf{O}^{-1})^\top = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{r_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{r_2} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_J^2 \mathbf{I}_{r_J} \end{pmatrix} := \boldsymbol{\Pi}. \tag{OA.3}$$

iv. each factor's flow predictability depends only on flow into factors within the same eigenspace

$$\mathbf{O}^{-1}\boldsymbol{\kappa}\mathbf{O} = \begin{pmatrix} \tilde{\boldsymbol{\kappa}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\kappa}}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tilde{\boldsymbol{\kappa}}_J \end{pmatrix} := \tilde{\boldsymbol{\kappa}}, \tag{OA.4}$$

where each $\tilde{\boldsymbol{\kappa}}_j$ is an $r_j \times r_j$ matrix for $j = 1, 2, \ldots, J$.

Appendix B.1 gives a proof of Proposition O.1. The key change is the last condition (OA.4). With distinct eigenvalues in the main text, each factor's flow predictability depends only on its own flows. However, with duplicate eigenvalues, the N factors are no longer uniquely pinned down, so each factor's flow predictability can depend on flow into all factors with the same eigenspace. As a result, the factors' predictability matrix $\tilde{\boldsymbol{\kappa}}$ is block-diagonal, with each block corresponding to a unique eigenvalue of $\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top$.

Other than $\tilde{\boldsymbol{\kappa}}$ being block-diagonal, the transformation from the N assets to the N factors remains the same as in the main text. We next solve the price for the N factors. Appendix B.2 provides a proof.

PROPOSITION O.2. *We denote the price of N factors at time t as $\tilde{\mathbf{P}}_t = (\tilde{P}_{1,t}, \tilde{P}_{2,t}, \ldots, \tilde{P}_{N,t})^\top$. In a linear equilibrium, if it exists, we have*

$$d\tilde{\mathbf{P}}_t = \boldsymbol{\Lambda} d\tilde{\mathbf{f}}_t + d\tilde{\mathbf{D}}_t / r, \tag{OA.5}$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Lambda}_J \end{pmatrix}, \tag{OA.6}$$

with each $\mathbf{\Lambda}_j$ *being an* $r_j \times r_j$ *matrix for* $j = 1, 2, \ldots, J$.

The key difference between Proposition O.2 and Proposition 2 in the main text is the form of the price sensitivity $\mathbf{\Lambda}$ in equation (OA.6). With duplicate eigenvalues, the price of a factor can depend on the flow into other factors within the same eigenspace. Intuitively, dividend covariance $\mathbf{\Sigma}^{\{D\}}$ and flow covariance $\mathbf{\Sigma}^{\{f\}}$ fail to uniquely pin down N factors, so the theory acknowledges the data limitation and allows $\mathbf{\Lambda}$ to be less restrictive.

Given the factor price in Proposition O.2, we derive the N assets' price. Consider any asset that holds w_n shares of factor n, with $\mathbf{w} = (w_1, w_2, \ldots, w_N)^\top$. The asset's dividend rate at time t is $D_t = \sum_{n=1}^{N} w_n \tilde{D}_{n,t}$. Because the factors' dividend rates are independent by (OA.2), we have $\text{cov}(dD_t, d\tilde{D}_{n,t}) = w_n \text{var}(d\tilde{D}_{n,t}) = w_n r^2 dt$. Therefore, by the no-arbitrage condition, the asset's price satisfies

$$dP_t = \mathbf{w}^\top d\tilde{\mathbf{P}}_t = \frac{\text{cov}(dD_t/r, d\tilde{\mathbf{D}}_t/r)}{dt} \mathbf{\Lambda} d\tilde{\mathbf{f}}_t + dD_t/r, \tag{OA.7}$$

which generalizes the factor model of price impacts in the main text.

B Proof

In this appendix, we provide proofs.

B.1 Proof of Proposition O.1

We conduct the Cholesky decomposition of $\mathbf{\Sigma}^{\{D\}}/r^2 = \mathbf{U}^\top\mathbf{U}$. First, we show that FFP $\mathbf{\Sigma}^{\{f\}}\mathbf{\Sigma}^{\{D\}}\boldsymbol{\kappa} = \boldsymbol{\kappa}\mathbf{\Sigma}^{\{f\}}\mathbf{\Sigma}^{\{D\}}$ implies the orthogonalization. The FFP implies that

$$(\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top)(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}) = (\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})(\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top). \tag{OB.1}$$

We denote the eigenvalue decomposition of matrix $\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top$ as

$$\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top\mathbf{G} = \mathbf{G}\mathbf{\Pi}, \tag{OB.2}$$

where $\mathbf{G}$ is an orthonormal matrix and the eigenvalue matrix $\mathbf{\Pi}$ is defined in (OA.3). In general, $\mathbf{G}$ is not unique, so we just pick an arbitrary one.

We define $\mathbf{O} = \mathbf{U}^{-1}\mathbf{G}$ and prove that $\mathbf{O}$ satisfies the orthogonalization. First, by $\mathbf{G}^\top = \mathbf{G}^{-1}$ and $\mathbf{\Sigma}^{\{D\}}/r^2 = \mathbf{U}^\top\mathbf{U}$, we have

$$\mathbf{O}^\top\mathbf{\Sigma}^{\{D\}}\mathbf{O} = r^2\mathbf{G}^\top(\mathbf{U}^{-1})^\top\mathbf{U}^\top\mathbf{U}\mathbf{U}^{-1}\mathbf{G} = r^2\mathbf{I}_N, \tag{OB.3}$$

which satisfies (OA.2). Second, by $\mathbf{G}^\top = \mathbf{G}^{-1}$ and (OB.2), we have

$$\mathbf{O}^{-1}\mathbf{\Sigma}^{\{f\}}(\mathbf{O}^{-1})^\top = \mathbf{G}^{-1}\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top\mathbf{G} = \mathbf{\Pi}, \tag{OB.4}$$

which satisfies (OA.3). Third, we define the matrix

$$\tilde{\boldsymbol{\kappa}} = \mathbf{O}^{-1}\boldsymbol{\kappa}\mathbf{O} = \mathbf{G}^\top\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}\mathbf{G} \tag{OB.5}$$

Using equations (OB.1), (OB.2), and (OB.5), we have

$$\tilde{\boldsymbol{\kappa}}\mathbf{\Pi} = \mathbf{G}^\top(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})\mathbf{G}\mathbf{G}^\top(\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top)\mathbf{G} = \mathbf{G}^\top(\mathbf{U}\mathbf{\Sigma}^{\{f\}}\mathbf{U}^\top)\mathbf{G}\mathbf{G}^\top(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})\mathbf{G} = \mathbf{\Pi}\tilde{\boldsymbol{\kappa}}. \tag{OB.6}$$

For any vector $\mathbf{v}$ in the span of the j-th part of the partition of matrix $\mathbf{\Pi}$ in (OB.2), we have $\mathbf{\Pi}\mathbf{v} = \pi_j\mathbf{v}$. Therefore, we have

$$\mathbf{\Pi}(\tilde{\boldsymbol{\kappa}}\mathbf{v}) = (\mathbf{\Pi}\tilde{\boldsymbol{\kappa}})\mathbf{v} = (\tilde{\boldsymbol{\kappa}}\mathbf{\Pi})\mathbf{v} = \tilde{\boldsymbol{\kappa}}\pi_j\mathbf{v} = \pi_j(\tilde{\boldsymbol{\kappa}}\mathbf{v}). \tag{OB.7}$$

Therefore, $\tilde{\boldsymbol{\kappa}}\mathbf{v}$ is also a vector in the span of the j-th part of the partition. Because we can arbitrarily choose the vector $\mathbf{v}$ and the part $j = 1, 2, \ldots, J$, the matrix $\tilde{\boldsymbol{\kappa}}$ must be

$$\tilde{\boldsymbol{\kappa}} = \begin{pmatrix} \tilde{\boldsymbol{\kappa}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\kappa}}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tilde{\boldsymbol{\kappa}}_J \end{pmatrix}, \tag{OB.8}$$

where each $\tilde{\boldsymbol{\kappa}}_j$ is an $r_j \times r_j$ matrix for $j = 1, 2, \ldots, J$. This proves equations (OA.4).

Conversely, we show that the orthogonalization implies the FFP. We define $\mathbf{G} = \mathbf{U}\mathbf{O}$. By equation (OA.2), we have

$$\mathbf{G}^\top \mathbf{G} = \mathbf{O}^\top \mathbf{U}^\top \mathbf{U}\mathbf{O} = \mathbf{O}^\top \boldsymbol{\Sigma}^{\{D\}} \mathbf{O}/r^2 = \mathbf{I}_N, \tag{OB.9}$$

so the $N \times N$ matrix $\mathbf{G}$ is orthonormal. By equation (OA.3) and $\mathbf{G}$ being orthonormal,

$$\boldsymbol{\Pi} = \mathbf{O}^{-1} \boldsymbol{\Sigma}^{\{f\}} (\mathbf{O}^{-1})^\top = \mathbf{G}^\top \mathbf{U} \boldsymbol{\Sigma}^{\{f\}} \mathbf{U}^\top \mathbf{G}, \tag{OB.10}$$

which is the eigenvalue decomposition of matrix $\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top$. By equation (OA.4), we have

$$\tilde{\boldsymbol{\kappa}} = \mathbf{O}^{-1} \boldsymbol{\kappa} \mathbf{O} = \mathbf{G}^\top \mathbf{U} \boldsymbol{\kappa} \mathbf{U}^{-1} \mathbf{G}. \tag{OB.11}$$

Because each block of $\tilde{\boldsymbol{\kappa}}$ corresponds to a distinct eigenvalue of $\boldsymbol{\Pi}$, we have

$$\boldsymbol{\Pi}\tilde{\boldsymbol{\kappa}} = \begin{pmatrix} \sigma_1^2 \tilde{\boldsymbol{\kappa}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \tilde{\boldsymbol{\kappa}}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_J^2 \tilde{\boldsymbol{\kappa}}_J \end{pmatrix} = \tilde{\boldsymbol{\kappa}}\boldsymbol{\Pi}. \tag{OB.12}$$

Therefore, by (OB.10), (OB.11), and (OB.12), we have

$$(\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top)(\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1}) = (\mathbf{G}\boldsymbol{\Pi}\mathbf{G}^\top)(\mathbf{G}\tilde{\boldsymbol{\kappa}}\mathbf{G}^\top) = (\mathbf{G}\tilde{\boldsymbol{\kappa}}\mathbf{G}^\top)(\mathbf{G}\boldsymbol{\Pi}\mathbf{G}^\top) = (\mathbf{U}\boldsymbol{\kappa}\mathbf{U}^{-1})(\mathbf{U}\boldsymbol{\Sigma}^{\{f\}}\mathbf{U}^\top), \tag{OB.13}$$

which implies the FFP.

B.2 Proof of Proposition O.2

We drop the tilde notation for factors in this proof for simplicity.

We conjecture the value function as

$$J(W, \mathbf{f}, \mathbf{D}) = -\frac{1}{r} e^{-r\gamma W - b_0 - \mathbf{b}_1^\top \mathbf{f} - \frac{1}{2}\mathbf{f}^\top \mathbf{b}_2 \mathbf{f} - \mathbf{a}_1^\top \mathbf{D} - \frac{1}{2}\mathbf{D}^\top \mathbf{a}_2 \mathbf{D}}. \tag{OB.14}$$

We conjecture that the factor price is linear in state variables,

$$\mathbf{P}_t = \mathbf{C}_0 + \boldsymbol{\Lambda}\mathbf{f}_t + \zeta \mathbf{D}_t, \tag{OB.15}$$

In the general case, the factor flows $\mathbf{f}_t$ satisfy

$$d\mathbf{f}_t = \boldsymbol{\kappa}\mathbf{f}_t dt + \sqrt{\boldsymbol{\Pi}} d\mathbf{B}_t^{\{f\}} - \mathbf{f}_{t-} dM_t, \tag{OB.16}$$

where $\sqrt{\boldsymbol{\Pi}}$ takes the square root of the diagonal terms of $\boldsymbol{\Pi}$ in (OA.3). The agent's wealth process is

$$\begin{aligned} dW_t &= \left(rW_t - c_t + \mathbf{y}_t^\top(\mathbf{D}_t - r\mathbf{P}_t)\right) dt + \mathbf{y}_t^\top d\mathbf{P}_t \\ &= \left(rW_t - c_t + \mathbf{y}_t^\top(\mathbf{D}_t - r\mathbf{P}_t + \boldsymbol{\Lambda}\boldsymbol{\kappa}\mathbf{f}_t + \zeta\boldsymbol{\phi})\right) dt \\ &\quad + \mathbf{y}_t^\top(\boldsymbol{\Lambda}\sqrt{\boldsymbol{\Pi}} d\mathbf{B}_t^{\{f\}} + r\zeta d\mathbf{B}_t^{\{D\}}) - \mathbf{y}_t^\top \boldsymbol{\Lambda}\mathbf{f}_{t-} dM_t. \end{aligned} \tag{OB.17}$$

The Bellman equation is

$$
\begin{aligned}
\rho J = \max_{c,\mathbf{y}} & -e^{-\gamma c} - r\gamma J \left(rW - c + \mathbf{y}^\top(\mathbf{D} - r\mathbf{P} + \boldsymbol{\Lambda}\boldsymbol{\kappa}\mathbf{f} + \zeta\boldsymbol{\phi})\right) \\
& - J\left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\kappa}\mathbf{f} - J\left(\mathbf{a}_1^\top + \mathbf{D}^\top \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right)\boldsymbol{\phi} \\
& + r\gamma J\left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\Pi}\boldsymbol{\Lambda}^\top\mathbf{y} + r^3\gamma\zeta J\left(\mathbf{a}_1^\top + \mathbf{D}^\top \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right)\mathbf{y} \\
& + \frac{1}{2}J\mathrm{tr}\left[\left(\left(\mathbf{b}_1 + \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f}\right)\left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right) - \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\Pi}\right] \\
& + \frac{1}{2}r^2 J\mathrm{tr}\left[\left(\left(\mathbf{a}_1 + \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\mathbf{D}\right)\left(\mathbf{a}_1^\top + \mathbf{D}^\top \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right) - \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right)\right] \\
& + \frac{1}{2}r^2\gamma^2 J\mathbf{y}^\top(\boldsymbol{\Lambda}\boldsymbol{\Pi}\boldsymbol{\Lambda}^\top + r^2\zeta^2\mathbf{I}_N)\mathbf{y} + \eta\left(J\left(W - \mathbf{y}^\top\boldsymbol{\Lambda}\mathbf{f}, \mathbf{0}, \mathbf{D}\right) - J\right). \qquad \text{(OB.18)}
\end{aligned}
$$

Taking the FOC with respect to consumption c, we have $\gamma e^{-\gamma c} + r\gamma J = 0$, which implies that the optimal consumption

$$
c^* = rW + \left(b_0 + \mathbf{b}_1^\top\mathbf{f} + \frac{1}{2}\mathbf{f}^\top\mathbf{b}_2\mathbf{f} + \mathbf{a}_1^\top\mathbf{D} + \frac{1}{2}\mathbf{D}^\top\mathbf{a}_2\mathbf{D}\right)/\gamma. \qquad \text{(OB.19)}
$$

Substituting back into the Bellman equation, we obtain

$$
\begin{aligned}
\rho - r = \max_{\mathbf{y}} & \, r\left(b_0 + \mathbf{b}_1^\top\mathbf{f} + \frac{1}{2}\mathbf{f}^\top\mathbf{b}_2\mathbf{f} + \mathbf{a}_1^\top\mathbf{D} + \frac{1}{2}\mathbf{D}^\top\mathbf{a}_2\mathbf{D}\right) - r\gamma\mathbf{y}^\top(\mathbf{D} - r\mathbf{P} + \boldsymbol{\Lambda}\boldsymbol{\kappa}\mathbf{f} + \zeta\boldsymbol{\phi}) \\
& - \left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\kappa}\mathbf{f} - \left(\mathbf{a}_1^\top + \mathbf{D}^\top \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right)\boldsymbol{\phi} \\
& + r\gamma\left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\Pi}\boldsymbol{\Lambda}^\top\mathbf{y} + r^3\gamma\zeta\left(\mathbf{a}_1^\top + \mathbf{D}^\top \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right)\mathbf{y} \\
& + \frac{1}{2}\mathrm{tr}\left[\left(\left(\mathbf{b}_1 + \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f}\right)\left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right) - \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\Pi}\right] \\
& + \frac{1}{2}r^2\mathrm{tr}\left[\left(\left(\mathbf{a}_1 + \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\mathbf{D}\right)\left(\mathbf{a}_1^\top + \mathbf{D}^\top \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right) - \frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right)\right] \\
& + \frac{1}{2}r^2\gamma^2\mathbf{y}^\top(\boldsymbol{\Lambda}\boldsymbol{\Pi}\boldsymbol{\Lambda}^\top + r^2\zeta^2\mathbf{I}_N)\mathbf{y} + \eta\left(J\left(W - \mathbf{y}^\top\boldsymbol{\Lambda}\mathbf{f}, \mathbf{0}, \mathbf{D}\right)/J - 1\right). \qquad \text{(OB.20)}
\end{aligned}
$$

Using the linearization $e^x \approx 1 + x$ for the jump term, we have

$$J\left(W - \mathbf{y}^\top \mathbf{\Lambda f}, \mathbf{0}, \mathbf{D}\right)/J - 1 \approx r\gamma \mathbf{y}^\top \mathbf{\Lambda f} + \mathbf{b}_1^\top \mathbf{f} + \frac{1}{2}\mathbf{f}^\top \mathbf{b}_2 \mathbf{f}. \tag{OB.21}$$

Taking FOC with respect to $\mathbf{y}$, we have

$$\begin{aligned} 0 = & - r\gamma(\mathbf{D} - r\mathbf{P} + \mathbf{\Lambda}\boldsymbol{\kappa}\mathbf{f} + \zeta\boldsymbol{\phi})^\top + r^2\gamma^2\mathbf{y}^\top(\mathbf{\Lambda\Pi\Lambda}^\top + r^2\zeta^2\mathbf{I}_N) \\ & + r\gamma\left(\mathbf{b}_1^\top + \mathbf{f}^\top\frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\mathbf{\Pi\Lambda}^\top + r^3\gamma\zeta\left(\mathbf{a}_1^\top + \mathbf{D}^\top\frac{\mathbf{a}_2 + \mathbf{a}_2^\top}{2}\right) + \eta r\gamma\mathbf{f}^\top\mathbf{\Lambda}^\top. \end{aligned} \tag{OB.22}$$

Substituting back into the Bellman equation and matching coefficients on $D_n f_n$, D_n, and D_n^2 terms, we obtain $\zeta = 1/r$, $\mathbf{a}_1 = \mathbf{0}$, and $\mathbf{a}_2 = \mathbf{0}$. Market clearing implies

$$\begin{aligned} \frac{\mathbf{S} - \mathbf{f}}{\mu} = & (\mathbf{\Lambda\Pi\Lambda}^\top + \mathbf{I}_N)^{-1}/(r\gamma) \\ & \left[(-r\mathbf{C}_0 - r\mathbf{\Lambda f} + \mathbf{\Lambda}\boldsymbol{\kappa}\mathbf{f} + \boldsymbol{\phi}/r) - \mathbf{\Lambda\Pi}\left(\mathbf{b}_1 + \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f}\right) - \eta\mathbf{\Lambda f}\right]. \end{aligned} \tag{OB.23}$$

We now conjecture that

$$\mathbf{b}_2 = \begin{pmatrix} \mathbf{b}_{2,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_{2,2} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{b}_{2,J} \end{pmatrix} \qquad \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Lambda}_J \end{pmatrix}, \tag{OB.24}$$

where $\mathbf{b}_{2,j}$ and $\mathbf{\Lambda}_j$ are both $r_j \times r_j$ matrices for $j = 1, 2, \ldots, J$ corresponding to the same eigenvalue σ_j^2 of $\mathbf{\Pi}$ in (OA.3). We also break the $N \times 1$ vectors into J corresponding blocks, $\mathbf{C}_0 = (\mathbf{C}_{0,1}; \mathbf{C}_{0,2}; \ldots; \mathbf{C}_{0,J})$, $\mathbf{S} = (\mathbf{S}_1; \mathbf{S}_2; \ldots; \mathbf{S}_J)$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1; \boldsymbol{\phi}_2; \ldots; \boldsymbol{\phi}_J)$, $\mathbf{b}_1 = (\mathbf{b}_{1,1}; \mathbf{b}_{1,2}; \ldots; \mathbf{b}_{1,J})$, and $\mathbf{f} = (\mathbf{f}_1; \mathbf{f}_2; \ldots; \mathbf{f}_J)$. We therefore simplify (OB.23) as

$$\mu \boldsymbol{\Lambda}_j \left(\boldsymbol{\kappa}_j - r\mathbf{I}_{r_j} - \sigma_j^2 \frac{\mathbf{b}_{2,j} + \mathbf{b}_{2,j}^\top}{2} - \eta \mathbf{I}_{r_j} \right) + r\gamma(\sigma_j^2 \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^\top + \mathbf{I}_{r_j}) = \mathbf{0}, \tag{OB.25}$$

$$\mu(r\mathbf{C}_{0,j} - \boldsymbol{\phi}_j / r + \sigma_j^2 \boldsymbol{\Lambda}_j \mathbf{b}_{1,j}) + r\gamma(\sigma_j^2 \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^\top + \mathbf{I}_{r_j})\mathbf{S}_j = \mathbf{0}. \tag{OB.26}$$

With the optimal portfolio holding $\mathbf{y}^*$, the Bellman equation simplifies to

$$\begin{aligned} \rho - r = & r\left(b_0 + \mathbf{b}_1^\top \mathbf{f} + \frac{1}{2}\mathbf{f}^\top \mathbf{b}_2 \mathbf{f}\right) - \frac{1}{2} r^2 \gamma^2 (\mathbf{S} - \mathbf{f})^\top (\boldsymbol{\Lambda}\boldsymbol{\Pi}\boldsymbol{\Lambda}^\top + \mathbf{I}_N)(\mathbf{S} - \mathbf{f})/\mu^2 \\ & - \left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right) \boldsymbol{\kappa}\mathbf{f} + \eta \mathbf{b}_1^\top \mathbf{f} + \frac{\eta}{2}\mathbf{f}^\top \mathbf{b}_2 \mathbf{f} \\ & + \frac{1}{2}\mathrm{tr}\left[\left(\left(\mathbf{b}_1 + \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f}\right)\left(\mathbf{b}_1^\top + \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right) - \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\Pi}\right]. \end{aligned} \tag{OB.27}$$

Note that

$$\mathrm{tr}\left[\frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f}\mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\boldsymbol{\Pi}\right] = \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\boldsymbol{\Pi}\frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f} = \sum_{j=1}^J \sigma_j^2 \mathbf{f}_j^\top \frac{(\mathbf{b}_{2,j} + \mathbf{b}_{2,j}^\top)^2}{4}\mathbf{f}_j, \tag{OB.28}$$

$$\frac{1}{2}\mathrm{tr}\left[\frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f}\mathbf{b}_1^\top \boldsymbol{\Pi} + \mathbf{b}_1 \mathbf{f}^\top \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\boldsymbol{\Pi}\right] = \mathbf{b}_1^\top \boldsymbol{\Pi}\frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\mathbf{f} = \sum_{j=1}^J \sigma_j^2 \mathbf{b}_{1,j}^\top \frac{\mathbf{b}_{2,j} + \mathbf{b}_{2,j}^\top}{2}\mathbf{f}_j, \tag{OB.29}$$

$$\mathrm{tr}\left[\left(\mathbf{b}_1 \mathbf{b}_1^\top - \frac{\mathbf{b}_2 + \mathbf{b}_2^\top}{2}\right)\boldsymbol{\Pi}\right] = \sum_{j=1}^J \sigma_j^2 (\mathbf{b}_{1,j}^\top \mathbf{b}_{1,j} - \mathrm{tr}(\mathbf{b}_{2,j})). \tag{OB.30}$$

Matching coefficients in the Bellman equation, we have

$$\frac{1}{2} r \mathbf{b}_{2,j} - \frac{1}{2}(\sigma_j^2 \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^\top + \mathbf{I}_{r_j}) r^2 \gamma^2 / \mu^2 - \frac{\mathbf{b}_{2,j} + \mathbf{b}_{2,j}^\top}{2}\boldsymbol{\kappa}_j + \sigma_j^2 \frac{(\mathbf{b}_{2,j} + \mathbf{b}_{2,j}^\top)^2}{8} + \frac{\eta}{2}\mathbf{b}_{2,j} = \mathbf{0}, \tag{OB.31}$$

$$r\mathbf{b}_{1,j} + (\sigma_j^2 \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^\top + \mathbf{I}_{r_j})\mathbf{S}_j r^2 \gamma^2 / \mu^2 - \boldsymbol{\kappa}_j^\top \mathbf{b}_{1,j} + \sigma_j^2 \frac{\mathbf{b}_{2,j} + \mathbf{b}_{2,j}^\top}{2}\mathbf{b}_{1,j} + \eta \mathbf{b}_{1,j} = \mathbf{0}, \tag{OB.32}$$

$$\rho - r = r b_0 - \frac{1}{2}\mathbf{S}_j^\top (\sigma_j^2 \boldsymbol{\Lambda}_j \boldsymbol{\Lambda}_j^\top + \mathbf{I}_{r_j})\mathbf{S}_j r^2 \gamma^2 / \mu^2 + \frac{1}{2}\sum_{j=1}^J \sigma_j^2 (\mathbf{b}_{1,j}^\top \mathbf{b}_{1,j} - \mathrm{tr}(\mathbf{b}_{2,j})). \tag{OB.33}$$

Equations (OB.25) and (OB.31) jointly determine $\mathbf{\Lambda}_j$ and $\mathbf{b}_{2,j}$. Unlike the distinct-eigenvalue case in the main text, both (OB.25) and (OB.31) are now $r_j \times r_j$ matrix equations, so can only be solved numerically for $\mathbf{\Lambda}_j$ and $\mathbf{b}_{2,j}$. Once $\mathbf{\Lambda}_j$ and $\mathbf{b}_{2,j}$ are known, it is straightforward to see that equation (OB.26) solves for $\mathbf{C}_{0,j}$, equation (OB.32) solves for $\mathbf{b}_{1,j}$, and equation (OB.33) solves for b_0.