Insurance: Mathematics and Economics 111 (2023) 57–83

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Dynamic asset-liability management with frictions

Tingjin Yan[a], Jinhui Han[b,d], Guiyuan Ma[c], Chi Chung Siu[d,*]

[a] *Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, and School of Statistics and Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai, China*
[b] *Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, China*
[c] *School of Economics and Finance, Xi'an Jiaotong University, Xi'an, China*
[d] *Department of Mathematics, Statistics and Insurance, School of Decision Sciences, The Hang Seng University of Hong Kong, Shatin, Hong Kong, China*

ARTICLE INFO

Article history:
Received September 2022
Received in revised form January 2023
Accepted 6 March 2023
Available online 15 March 2023

JEL classification:
C61
G11

Keywords:
Asset-liability management
Temporary and persistent price impacts
Return predictability
Target-chasing strategy
Coupled matrix Riccati differential system

ABSTRACT

This paper studies a dynamic asset-liability management problem of a company with market frictions. Specifically, the asset prices are modeled by a multivariate geometric Brownian motion with their excess returns driven by some correlated stochastic signals; and the liability process is modeled by another geometric Brownian motion correlated to the asset price dynamics. The company trades dynamically to offset the risks from its liability and each trade induces both temporary and persistent price impacts. We characterize the optimal trading strategies in terms of the solutions to the coupled matrix Riccati differential systems. Due to the price impacts, the company should adopt a target-chasing strategy in which the dynamic target portfolio is expressed in terms of the return-predicting signals and realized liability. We also derive some sufficient conditions, based on the model parameters alone, to ensure the well-posedness of the coupled Riccati systems. Our numerical results indicate that the temporary and persistent price impacts have opposite implications on the company's trading behavior. While the temporary price impact slows down the company's trading speed toward the target portfolio, the persistent price impact may encourage the company to trade more aggressively to enhance the expected returns.

1. Introduction

The problem of jointly managing assets and liabilities is prevalent in various industries such as insurance, pension funds, and banking. Asset-liability management (ALM) typically seeks the best asset investment plan to accomplish the profit goals while meeting the current and future liabilities. Indeed, this concept is described in Society of Actuaries (2003) as follows:

"ALM is the ongoing process of formulating, implementing, monitoring and revising strategies related to assets and liabilities to achieve an organization's financial objectives, given the organization's risk tolerances and other constraints."

Considering its importance, ALM has always been a heated topic among practitioners and academics in related fields. For instance, Kusy and Ziemba (1986) developed a multi-period stochastic linear programming ALM model to find an optimal trade-off between risk, return, and liquidity in a bank; also see Choudhry (2011) for a comprehensive review for ALM in banking. Leippold et al. (2004) proposed a geometric method to solve the multi-period ALM problem with explicit investment strategies and an efficient frontier using the mean-variance criterion. Within a stochastic linear-quadratic framework, Chiu and Li (2006) and Xie et al. (2008) respectively addressed a continuous-time ALM problem, with the underlying liability dynamics being a geometric Brownian motion and an arithmetic Brownian motion. The Markovian regime-switching extensions were also implemented in Chen et al. (2008), Xie (2009), and Wei et al. (2013) to accommodate the randomly changing market state so that the ALM investment strategies are adaptive to the regime-switching environment. Chiu and Wong (2012, 2014) investigated the mean-variance ALM strategies under the cointegration effect and correlation risk, respectively. Yao et al. (2013) considered the case with endogenous liabilities and derived the optimal allocation among multiple risky assets and liabilities. Recently, more interesting and practical factors have been incorporated to advance the analysis on the ALM problems, which include, but not limited to, open-loop equilibrium strategy (Wei and Wang, 2017), derivative-based investment (Li et al., 2018), risk constraints

* Corresponding author.
E-mail addresses: tjyan@fem.ecnu.edu.cn (T. Yan), jinhuizjcu@gmail.com (J. Han), guiyuanma@xjtu.edu.cn (G. Ma), ccsiu@hsu.edu.hk (C.C. Siu).

https://doi.org/10.1016/j.insmatheco.2023.03.001
0167-6687/

on liquidity (Pan and Xiao, 2017) and debt ratio (Zhang et al., 2020), inflation risk (Zhu et al., 2020), non-Markovian regime-switching environments (Shen et al., 2020), and general ALM problems with delay (A et al., 2022).

In this paper, we solve a continuous-time ALM problem with market frictions and return predictability under a local mean-variance (MV) optimization criterion. Indeed, market frictions play an important role in the daily operations of the insurance companies. Berry-Stölzle (2008a) pointed out that property and casualty insurance companies often need to liquidate financial assets to cover claims and hence face liquidity risk. To the best of our knowledge, Berry-Stölzle (2008a) is the first paper to study asset allocation and liquidation with frictions in the context of property and casualty insurance. Berry-Stölzle (2008b) further explored the impact of illiquidity and interpreted it as transaction costs in the ALM problem. Berry-Stölzle (2008a) and Berry-Stölzle (2008b) resorted to numerical methods when computing optimal asset allocation and hedging strategies, but their numerical schemes prove to be difficult to be extended to the multi-asset settings. Due to the technical difficulty, there has been a scarcity of actuarial literature providing analytical results for the ALM problem with market frictions since then. In this respect, to the best of our knowledge, our work is the first one to fill this gap in a comprehensive setup with both temporary and persistent price impacts.

Our ALM problem has its roots dating back to the pioneering works of Samuelson (1969) and Merton (1969); also see Yan and Wong (2019, 2020), Han and Yam (2022), Han et al. (2022), and the references therein for some recent developments. Specifically, we build on the tractable frameworks of Gârleanu and Pedersen (2013, 2016) (hereinafter, the GP model). Considering the fact that the real financial market is not perfectly efficient and trading typically involves market frictions, Gârleanu and Pedersen (2013) introduced a quadratic transaction cost from a linear price impact assumption and derived a closed-form optimal investment strategy for an infinitely-lived mean-variance agent in a discrete-time framework. The continuous-time extension of Gârleanu and Pedersen (2013) was further investigated in Gârleanu and Pedersen (2016), and the highly tractable frameworks in these two seminar papers have sparked a widespread attention (Glasserman and Xu, 2013; Moallemi and Sağlam, 2017; Ma et al., 2019, 2020a,b, 2022,b; Bensoussan et al., 2022). Moreover, the GP model allows for return predictability, which has been a stylized fact validated by many empirical studies (Ang and Bekaert, 2007; Campbell and Thompson, 2008; Welch and Goyal, 2008). As a related work to ALM problems with return predictability, Ferstl and Weissensteiner (2011) proposed a stochastic linear programming framework and solved it numerically.

We investigate a local mean-variance ALM problem over a finite time horizon inspired by the GP model. Different from the pre-commitment solution approach to the mean-variance problems (Chiu and Li, 2006; Chiu and Wong, 2012, 2014; Zhang and Chen, 2016), the local mean-variance analysis provides an alternative convenient yet reasonable formulation to avoid the time-inconsistent issue and therefore it is widely adopted by the practitioners. More importantly, we do not follow the adoption of the arithmetic Brownian motions in the GP model. Instead, we assume that the asset prices follow a multi-dimensional geometric Brownian motion. Under some appropriate and reasonable model assumptions, the nice tractability is still preserved in our scenario. Meanwhile, we incorporate the mean-reverting return-predicting signals such that the expected returns of asset prices are linear in these predictors. The company is faced with an uncertain liability over time, with the randomness correlated with that of asset prices and predictors so as to account for the common risk in financial markets. Moreover, when the manager employs a dynamic investment strategy to maximize the objective function, each trade in the market not only incurs a quadratic transaction cost due to the linear temporary price impact, but also induces a persistent price impact that distorts the asset return persistently. These price impacts are motivated by the fact that institutional trades are typically of great volume, which may affect the trading decisions of other market participants and then hence the price movement (Bertsimas and Lo, 1998; Sannikov and Skrzypacz, 2016; Van Kervel and Menkveld, 2019).

Using the dynamic programming principle, we solve the corresponding Hamilton-Jacobi-Bellman (HJB) equation up to a coupled Riccati system of ordinary differential equations (ODEs). In line with the GP models, this coupled Riccati system admits no explicit solutions, although its well-posedness can be established when only a linear temporary price impact is present. However, the incorporation of the persistent price impact complicates the solution significantly. We follow the methodology of Bensoussan et al. (2022) and apply the comparison principle for matrix Riccati equation system in Abou-Kandil et al. (2003) to develop a refined sufficient condition for guaranteeing a unique local solution. We also derive some tight sufficient conditions for a unique global solution under some special important cases. These conditions can be verified easily in practice. Furthermore, we provide a comprehensive analysis on the discrete-time counterpart of our formulated problem to demonstrate the superiority of our concise sufficient conditions in continuous time.

Our solution features the company dynamically trading towards a target portfolio. Indeed, the analytical characterizations enable the ALM manager to dynamically audit the target portfolio and implement the current trading strategy accordingly, based on the integrated information on return predictors and realized liabilities. In general, a higher growth rate or variation in liability induces the manager to invest more in stocks. The trading speed will be lowered when the temporary price impact becomes larger. The persistent price impact, in particular, has a greater influence on the trading behavior than the temporary price impact, which is consistent with Berry-Stölzle (2008b). Meanwhile, the presence of persistent price impact enables the manager to manipulate the price via engaging in an aggressive trading.

We conclude this introduction by highlighting the key differences between this paper and the existing literature. From the modeling perspective, this paper is the first one to study an ALM problem of an insurance company with multiple assets and market frictions (temporary and persistent price impacts). The objective is to maximize the company's trading performance while netting trading costs and liabilities over a finite time horizon. Different from the existing literature (see, e.g. Gârleanu and Pedersen (2013, 2016); Glasserman and Xu (2013); Ma et al. (2020b); Bensoussan et al. (2022)), we do not use the arithmetic Brownian motions to model the asset price dynamics, and adopt a multi-dimensional geometric Brownian motion to ensure the price positiveness. Moreover, inspired by Gârleanu and Pedersen (2013, 2016), we characterize the persistent price impact with market resiliency, which introduces an additional state variable, known as the price distortion, into the asset price dynamics. On the other hand, Ma et al. (2020b) investigated a different optimal execution problem, where the investor should purchase/liquidate a predetermined amount of assets while minimizing the execution costs over a fixed time period. The price dynamics in Ma et al. (2020b) follow an arithmetic Brownian motion. Although Ma et al. (2020b) also considered temporary and permanent price impacts, they did not consider the market resiliency when formulating the persistent price impact.

From the mathematical perspective, the additional state variable from the newly introduced price distortion process leads to a more complicated coupled matrix Riccati differential equations (RDE) system compared with that in Ma et al. (2020b). Since the solution techniques in Ma et al. (2020b) can not be employed in our problem, we have henceforth developed some new results for the existence of solutions to the RDEs herein. In sum, this paper aims to offer a highly tractable framework to study the ALM problems with trading frictions to the actuarial literature, showing that the optimal trading strategies of the multiple assets can be characterized in terms of the solutions to the coupled matrix RDEs, which can then be efficiently computed using the existing numerical schemes for ODEs.

The remainder of this paper is structured as follows. Section 2 describes the model settings and formulates our target ALM problem. In Section 3, we first explore the situation with only temporary price impact. Then, we proceed to tackling the case with both temporary and persistent price impacts in Section 4. In both cases, analytical optimal ALM strategies are characterized up to the solutions of the coupled differential Riccati system. In Section 5, we further investigate the optimal initial capital structure for the company by calculating the optimal funding ratio. Numerical examples are provided in Section 6 to better illustrate the financial implications. Section 7 concludes the paper. The detailed proofs of related claims are relegated to the Appendix.

2. Problem formulation

2.1. Notations

Before describing the problem, we introduce some basic mathematical notations that will be used frequently in this paper for readers' convenience. We hereinafter use boldfaces to distinguish matrices (including vectors) from scalars, and all vectors are column vectors. Let $n > 0$ be a generic integer, $\mathbf{1} := (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$ be a vector of ones, and $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ denote the n-dimensional identity matrix. Let $\left(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ be a complete filtered probability space satisfying the usual conditions.

The spectral norm is denoted by $\|\cdot\|$. Specifically, for any $\mathbf{X} \in \mathbb{R}^{n\times n}$, $\|\mathbf{X}\| = \sqrt{\lambda_{\max}\left(\mathbf{X}^\top\mathbf{X}\right)}$, where $\lambda_{\max}(\mathbf{X})$ represents the largest eigenvalue of $\mathbf{X}$. For two symmetric matrices $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{n\times n}$, $\mathbf{X}_1 <$ (resp. $>$)$\mathbf{X}_2$ represents $\mathbf{X}_2 - \mathbf{X}_1$ is positive (resp. negative) definite, and $\mathbf{X}_1 \leq$ (resp. $\geq$)$\mathbf{X}_2$ represents $\mathbf{X}_2 - \mathbf{X}_1$ is positive (reps. negative) semidefinite.

2.2. State processes

We consider a financial market with one risk-free asset and n risky assets. The price dynamics for risky assets are jointly modeled by a multivariate geometric Brownian motion:

$$d\mathbf{P}_t = \mathrm{diag}\left(\mathbf{P}_t\right)\left[\left(r_f\mathbf{1} + \mathbf{B}\mathbf{f}_t\right)dt + \boldsymbol{\sigma}_P d\mathbf{W}_t^P\right], \tag{1}$$

where $\mathbf{P}_t = (P_{1t}, P_{2t}, \cdots, P_{nt})^\top$ and $\mathrm{diag}\left(\mathbf{P}_t\right)$ represents a diagonal matrix with the diagonal entries being the components of $\mathbf{P}_t$; $\mathbf{W}_t^P$ is an $\mathbb{F}$-adapted n-dimensional Brownian motion; r_f is the risk-free interest rate, and $\boldsymbol{\mu}_{P_t} := r_f\mathbf{1} + \mathbf{B}\mathbf{f}_t \in \mathbb{R}^n$ and $\boldsymbol{\sigma}_P \in \mathbb{R}^{n\times n}$ denote the return and volatility coefficients for n stocks, respectively; and $\mathbf{f}_t = (f_{1t}, f_{2t}, \cdots, f_{mt})^\top \in \mathbb{R}^m$ is a vector of return-predicting factors that could be of a different dimension of stocks, associated with the factor loading matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$. Specifically, motivated by the empirical evidences in Gârleanu and Pedersen (2013) and Lehalle and Neuman (2019) where the mean-reverting nature of common predictors is observed, we assume $\mathbf{f}_t$ to be a stationary Ornstein-Uhlenbeck process

$$d\mathbf{f}_t = -\boldsymbol{\Phi}\mathbf{f}_t dt + \boldsymbol{\sigma}_f d\mathbf{W}_t^f,$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{m\times m}$ represents the mean-reversion speed, $\mathbf{W}_t^f$ is an $\mathbb{F}$-adapted m-dimensional Brownian motion which can be correlated with $\mathbf{W}_t^P$, and $\boldsymbol{\sigma}_f \in \mathbb{R}^{m\times m}$ represents the volatility coefficient for m factors.

The asset-liability manager takes a dynamic investment in these stocks over a finite horizon $[0, T]$. The corresponding portfolio strategy is denoted by a vector $\mathbf{x}_t = (x_{1t}, x_{2t}, \cdots, x_{nt})^\top$ specifying the *dollar amount* invested into each stock at time t. Following the setup in Gârleanu and Pedersen (2016) and Collin-Dufresne et al. (2020), we assume that $\mathbf{x}_t$ is absolutely continuous with an instantaneous trading intensity $\mathbf{u}_t = (u_{1t}, u_{2t}, \cdots, u_{nt})^\top$ such that[1]

$$d\mathbf{x}_t = \mathbf{u}_t dt. \tag{2}$$

In particular, a trading strategy $\mathbf{u}$ is said to be admissible if

$$\mathbb{E}\left[\int_0^T |\mathbf{u}_s|^2 ds\right] < \infty,$$

and the set of all admissible trading strategies is defined as $\mathcal{A}$.

If we define the asset value of the company as a_t at time t, it then satisfies the following dynamics in a frictionless market

$$da_t = r_f\left(a_t - \mathbf{x}_t^\top\mathbf{1}\right)dt + \mathbf{x}_t^\top\,\mathrm{diag}(\mathbf{P}_t)^{-1}d\mathbf{P}_t.$$

However, the practical financial market is not entirely efficient and trading frictions always exist for each transaction. We first illustrate the setup for the temporary price impact. Bertsimas and Lo (1998) note that it is more plausible to consider a *linear-percentage* price impact than a linear price impact, because the latter has a percentage price impact that decreases in the price level, which is counterfactual (see Bertsimas and Lo (1998) and the references therein). Specifically, for an instantaneous trading intensity $\mathbf{u}_t$, we assume the execution

[1] In the continuous-time framework, Gârleanu and Pedersen (2016) shows that any portfolio $\mathbf{x}_t$ that is not absolutely continuous, with non-smooth changes such as discrete jumps and quadratic variations, would incur an infinite trading costs, which is then excluded from our analysis.

price is raised on average by a percentage of $\frac{1}{2}\mathbf{\Lambda}\mathbf{u}_t$ where $\mathbf{\Lambda}\in\mathbb{R}^{n\times n}$ is a positive definite matrix. Let $\tilde{\mathbf{P}}_t$ represent the average execution price under $\mathbf{u}_t$, then

$$\tilde{\mathbf{P}}_t=\mathbf{P}_t+\frac{1}{2}\operatorname{diag}(\mathbf{P}_t)\,\mathbf{\Lambda}\mathbf{u}_t=\operatorname{diag}(\mathbf{P}_t)\left(\mathbf{1}+\frac{1}{2}\mathbf{\Lambda}\mathbf{u}_t\right).$$

Therefore, the temporary transaction cost $\mathrm{TC}(\mathbf{u}_t)dt$ incurred by $\mathbf{u}_t dt$ can be calculated by

$$\mathrm{TC}(\mathbf{u}_t)dt=\left(\operatorname{diag}(\mathbf{P}_t)^{-1}\mathbf{u}_t dt\right)^{\top}\tilde{\mathbf{P}}_t-\mathbf{u}_t{}^{\top}\mathbf{1}dt=\frac{1}{2}\mathbf{u}_t^{\top}\mathbf{\Lambda}\mathbf{u}_t dt.$$

The assumption that $\mathbf{\Lambda}$ is symmetric is without loss of generality (also see Footnote 5 in Gârleanu and Pedersen (2013)). One can assume a non-symmetric temporary price impact level $\bar{\mathbf{\Lambda}}$ and then the corresponding trading cost becomes $\frac{1}{2}\mathbf{u}_t^{\top}\bar{\mathbf{\Lambda}}\mathbf{u}_t$, which is equivalent to the trading cost under a symmetric $\mathbf{\Lambda}=\frac{1}{2}\left(\bar{\mathbf{\Lambda}}+\bar{\mathbf{\Lambda}}^{\top}\right)$. The positive definite assumption on $\mathbf{\Lambda}$ is reasonable since a non-zero trading speed $\mathbf{u}_t$ should induce positive trading cost. The quadratic transaction cost has been commonly adopted in the literature (Obizhaeva and Wang, 2013; Gârleanu and Pedersen, 2016; Collin-Dufresne et al., 2020; Isaenko, 2022). Indeed, the temporary price impact naturally arises from the instantaneous imbalance between the supply and demand of the stocks caused by the manager's coming trading, which eventually renders the manager make a price concession in order to execute the trade.

In addition, we further consider the case that the manager's trade shall cause a persistent impact on the market expectations as other market participants believe that institutional trades are usually accompanied with certain information leakage on the prices (Sannikov and Skrzypacz, 2016; Van Kervel and Menkveld, 2019). To study this situation, we again assume a linear-percentage price impact which distorts the assets' return persistently. Specifically, we introduce a return distortion process $\mathbf{D}_t$ satisfying

$$d\mathbf{D}_t=-\mathbf{R}\mathbf{D}_t dt+\mathbf{C}\mathbf{u}_t dt,$$

where $\mathbf{C}\in\mathbb{R}^{n\times n}$ is a positive definite matrix measuring the persistent price impact level and $\mathbf{R}\in\mathbb{R}^{n\times n}$ is a positive definite matrix representing the resiliency of the persistent price impact. In other words, a trading speed of $\mathbf{u}_t$ will raise or reduce the return by $\mathbf{C}\mathbf{u}_t$, and the corresponding expected return in the interval $[t,t+dt)$ becomes $\tilde{\boldsymbol{\mu}}_{Pt}dt=\boldsymbol{\mu}_{Pt}dt+d\mathbf{D}_t$. We naturally assume $\mathbf{D}_0=\mathbf{0}_{n\times 1}$ to indicate that there is no price distortion before the manager's trade.

Remark 2.1. In addition, Gârleanu and Pedersen (2016) employs the arithmetic Brownian motion to model the price dynamic, which implies a positive probability of negative price. Nonetheless, our introduced models can get rid of this problematic setting in that the price dynamics are always strictly positive. Meanwhile, to the best of our knowledge, our work derives the first analytical result in the literature that tackles the portfolio selection following the framework of Gârleanu and Pedersen (2016), with the underlying prices assumed to follow geometric Brownian motions.

Employing the dynamic portfolio strategy $\mathbf{u}_t$, the firm's asset value dynamic, with the presence of both persistent and temporary price impacts, evolves according to

$$\begin{aligned}da_t&=r_f(a_t-\mathbf{x}_t^{\top}\mathbf{1})dt+\mathbf{x}_t^{\top}\left[(\operatorname{diag}\mathbf{P}_t)^{-1}d\mathbf{P}_t+d\mathbf{D}_t\right]-\mathrm{TC}(\mathbf{u}_t)dt\\&=\left\{r_f a_t+\mathbf{x}_t^{\top}\left[\mathbf{B}\mathbf{f}_t dt-\mathbf{R}\mathbf{D}_t+\mathbf{C}\mathbf{u}_t\right]-\frac{1}{2}\mathbf{u}_t^{\top}\mathbf{\Lambda}\mathbf{u}_t\right\}dt+\mathbf{x}_t^{\top}\boldsymbol{\sigma}_P d\mathbf{W}_t^P.\end{aligned}$$

Following Xie (2009); Zhang and Chen (2016); Zhu et al. (2020), the company's uncontrollable liability process l_t is modeled by a geometric Brownian motion:

$$dl_t=l_t\left[\mu_l dt+\sigma_l dW_t^l\right],\tag{3}$$

where W_t^l is an $\mathbb{F}$-adapted standard Brownian motion, which could be correlated with $\mathbf{W}_t^P$ and $\mathbf{W}_t^f$. The respective vectors of correlation coefficients are denoted by $\boldsymbol{\rho}_{Pl}\in\mathbb{R}^n$ and $\boldsymbol{\rho}_{fl}\in\mathbb{R}^n$. For simplicity, we define $\mathbf{\Sigma}_P:=\boldsymbol{\sigma}_P\boldsymbol{\sigma}_P^{\top}\in\mathbb{R}^{n\times n}$, $\Sigma_l:=\sigma_l^2\in\mathbb{R}$, $\mathbf{\Sigma}_{Pl}:=\boldsymbol{\sigma}_P\boldsymbol{\rho}_{Pl}\sigma_l\in\mathbb{R}^n$, and $\mathbf{\Sigma}_{fl}:=\boldsymbol{\sigma}_f\boldsymbol{\rho}_{fl}\sigma_l\in\mathbb{R}^m$.

Remark 2.2. One can also choose a jump process, such as compound Poisson process, to model liability, especially when the companies face random payments for insurance claims or liability (Chiu and Wong, 2014). In this paper, we are interested in how asset allocation hedges liability risk while achieving the overall profit goal. Indeed, the hedging demand arises from the correlation between investment opportunity and liability. Therefore, to maintain this correlation in a jump model, one may also need to formulate a "co-jump" component in asset price dynamics. For convenience, we here adopt the geometric Brownian motion setting so that the correlation between asset and liability is more intuitive and accessible. Nonetheless, adding jumps to (3) will not lead to further mathematical complications and the main results of the current work remain valid.

Then the net asset value $S_t:=a_t-l_t$ satisfies

$$dS_t=\left\{r_f S_t+\left(r_f-\mu_l\right)l_t+\mathbf{x}_t^{\top}\left[\mathbf{B}\mathbf{f}_t-\mathbf{R}\mathbf{D}_t+\mathbf{C}\mathbf{u}_t\right]-\frac{1}{2}\mathbf{u}_t^{\top}\mathbf{\Lambda}\mathbf{u}_t\right\}dt+\mathbf{x}_t^{\top}\boldsymbol{\sigma}_P d\mathbf{W}_t^P-\sigma_l l_t dW_t^l.$$

2.3. Objective function

At time $t \in [0, T]$, the manager aims to maximize the discounted sum of her *local* mean-variance criterion of the net asset return over each infinitesimal time interval $[s, s+ds]$ for $s \in (t, T]$. Specifically, we consider the following objective function:

$$\begin{aligned} J(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l;\ \mathbf{u}) &:= \mathbb{E}_t\left[\int_t^T e^{-\rho(s-t)}\left(\mathbb{E}_s\left[dS_s - r_f S_s ds\right] - \frac{\gamma}{2}\mathrm{Var}_s\left[dS_s - r_f S_s ds\right]\right)\right] \\ &= \mathbb{E}_t\left[\int_t^T e^{-\rho(s-t)}\left\{\left(r_f - \mu_l\right) l_s + \mathbf{x}_s^\top\left[\mathbf{B}\mathbf{f}_s - \mathbf{R}\mathbf{D}_s + \mathbf{C}\mathbf{u}_s\right] - \frac{1}{2}\mathbf{u}_s^\top \boldsymbol{\Lambda}\mathbf{u}_s - \frac{\gamma}{2}\left[\mathbf{x}_s^\top \boldsymbol{\Sigma}_P \mathbf{x}_s + \Sigma_l l_s^2 - 2\mathbf{x}_s^\top \boldsymbol{\Sigma}_{Pl} l_s\right]\right\} ds\right], \end{aligned}$$

where $\rho > 0$ is the discount rate, $\gamma > 0$ represents the manager's risk aversion, and $\mathbb{E}_t[\cdot]$ and $\mathrm{Var}_t[\cdot]$ respectively denote the conditional expectation and variance operators given $\mathcal{F}_t$. We refer to such an objective as the *local mean-variance criterion*. This criterion characterizes the company's cumulative sum of its mean-variance preference on the changes in its net asset values in each infinitesimal time period. Indeed, it is widely adopted by practitioners. Moreover, it provides a simple alternative to circumvent the notion of time inconsistency in the dynamic mean-variance criterion (see the notable work by Basak and Chabakauri (2010) and also the recent work by Bensoussan et al. (2022)). Due to its high tractability, this criterion is popular in the extant literature on the portfolio choice problems with trading frictions (see, e.g., Gârleanu and Pedersen (2013, 2016); Glasserman and Xu (2013); Collin-Dufresne et al. (2020); Demiguel et al. (2016)). In this respect, this paper introduces the local mean-variance criterion as a tractable way to study the ALM problems with trading frictions to the actuarial literature.

The manager aims to find an admissible trading strategy $\mathbf{u}^* \in \mathcal{A}$ which maximizes the objective function, i.e.,

$$J(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l; \mathbf{u}^*) = \max_{\mathbf{u}\in\mathcal{A}} J(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l; \mathbf{u}). \tag{4}$$

According to the standard dynamic programming theory, the Hamilton–Jacobi–Bellman (HJB) equation associated with the value function V is

$$\begin{aligned} \rho V = \max_{\mathbf{u}\in\mathbb{R}^n}\Bigg\{ &\frac{\partial V}{\partial t} + \frac{\partial V}{\partial \mathbf{x}}^\top \mathbf{u} - \frac{\partial V}{\partial \mathbf{f}}^\top \Phi\mathbf{f} + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 V}{\partial \mathbf{f}\partial \mathbf{f}^\top}\boldsymbol{\Sigma}_f\right) + \frac{\partial V}{\partial l}\mu_l l + \frac{1}{2}\frac{\partial^2 V}{\partial l^2}\Sigma_l l^2 + \frac{\partial^2 V}{\partial \mathbf{f}\partial l}^\top \boldsymbol{\Sigma}_{fl} l \\ &+ \frac{\partial V}{\partial \mathbf{D}}^\top(-\mathbf{R}\mathbf{D} + \mathbf{C}\mathbf{u}) + \left(r_f - \mu_l\right) l + \mathbf{x}^\top\left[\mathbf{B}\mathbf{f} - \mathbf{R}\mathbf{D} + \mathbf{C}\mathbf{u}\right] - \frac{1}{2}\mathbf{u}^\top\boldsymbol{\Lambda}\mathbf{u} - \frac{\gamma}{2}\left(\mathbf{x}^\top\boldsymbol{\Sigma}_P\mathbf{x} + \Sigma_l l^2 - 2\mathbf{x}^\top\boldsymbol{\Sigma}_{Pl} l\right)\Bigg\} \end{aligned} \tag{5}$$

with the boundary condition $V(T, \mathbf{x}, \mathbf{D}, \mathbf{f}, l) = 0$, where we have suppressed the arguments $(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l)$ for simplicity.

The choice of the control variable (portfolio position vs. trading intensity) can be explained in terms of trading frictions in continuous time. In absence of trading frictions, the existing literature on the ALM problems takes the portfolio position as the control variable to reflect that the utility-maximizing manager can adjust his portfolio freely to capture the instantaneous changes of the return-predicting factors and the liability. However, when trading frictions are present, the manager can no longer adjust his portfolio freely and must take the trading costs in consideration (Gârleanu and Pedersen (2013, 2016)). Since trading costs are incurred only when he rebalances his portfolio, we therefore adopt the manager's trading intensity in continuous time as the control variable to account for the cumulative trading costs during the finite investment horizon.

Remark 2.3. In Section 4.2, the optimal solution to the frictionless version of our ALM problem (43) is $\mathbf{x}^{\mathrm{nf}}(\mathbf{f}_t, l_t) = \frac{1}{\gamma}\boldsymbol{\Sigma}_P^{-1}\mathbf{B}\mathbf{f}_t + \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{Pl} l_t$, which is of a Brownian-motion type. Economically, the optimal frictionless portfolio $\mathbf{x}^{\mathrm{nf}}$ implies that the manager should rebalance his portfolio whenever the return-predicting factors $\mathbf{f}$ and liability l change. However, as first stated in Gârleanu and Pedersen (2016), adopting $\mathbf{x}^{\mathrm{nf}}$ in the frictional market would incur infinite trading costs in continuous time, contributed by the non-differentiability of the Brownian paths. To quantify trading costs while maintaining tractability, we have therefore adopted the instantaneous rate of change of the portfolio position, i.e., trading intensity, as the control variable so that the portfolio position is absolutely continuous, with bounded trading costs.

Remark 2.4. For completeness, we have also included a discrete-time formulation of our ALM problem in Appendix B. For each time t, the control variable can be either $\mathbf{x}_t$ or $\Delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$, and the state variable is $\mathbf{x}_{t-1}$. In other words, one can also adopt the commonly-used portfolio position or the change of the portfolio position as the control variable when solving the ALM problem in discrete time. Due to the feedback nature of our optimal solution, the optimal portfolio position $\mathbf{x}_t$ is expressed in terms of the portfolio position in the previous period, $\mathbf{x}_{t-1}$, the observed return-predicting factors $\mathbf{f}_t$ and liability l_t, with the coefficients expressed in terms of the coupled system of Riccati difference equations.

3. Asset-liability management with only temporary price impact

In this section, we investigate this case and solve for the optimal trading strategy when the persistent distortion is absent, i.e., $\mathbf{D} \equiv \mathbf{0}_{n\times 1}$ ($\mathbf{R} = \mathbf{C} = \mathbf{0}_{n\times n}$). In other words, the market friction only comes from the linear temporary price impact.

3.1. Optimal solution

The HJB equation (5) in this case degenerates to

$$\rho V=\max_{\mathbf{u}\in\mathbb{R}^n}\left\{\frac{\partial V}{\partial t}+\frac{\partial V}{\partial \mathbf{x}}^{\top}\mathbf{u}-\frac{\partial V}{\partial \mathbf{f}}^{\top}\Phi\mathbf{f}+\frac{1}{2}\operatorname{tr}\left(\frac{\partial^2 V}{\partial \mathbf{f}\partial \mathbf{f}^{\top}}\mathbf{\Sigma}_f\right)+\frac{\partial V}{\partial l}\mu_l l+\frac{1}{2}\frac{\partial^2 V}{\partial l^2}\Sigma_l l^2+\frac{\partial^2 V}{\partial \mathbf{f}\partial l}^{\top}\mathbf{\Sigma}_{fl}l\right.$$
$$\left.+\left(r_f-\mu_l\right)l+\mathbf{x}^{\top}\mathbf{B}\mathbf{f}-\frac{1}{2}\mathbf{u}^{\top}\mathbf{\Lambda}\mathbf{u}-\frac{\gamma}{2}\left(\mathbf{x}^{\top}\mathbf{\Sigma}_P\mathbf{x}+\Sigma_l l^2-2\mathbf{x}^{\top}\mathbf{\Sigma}_{Pl}l\right)\right\}, \quad (6)$$

where $V=V(t,\mathbf{x},\mathbf{f},l)$ and $V(T,\mathbf{x},\mathbf{f},l)=0$.

Consider the following *ansatz* for the value function:

$$V(t,\mathbf{x},\mathbf{f},l)=-\frac{1}{2}\mathbf{x}^{\top}\mathbf{A}_{xx}(t)\mathbf{x}+\frac{1}{2}\mathbf{f}^{\top}\mathbf{A}_{ff}(t)\mathbf{f}+\frac{1}{2}A_{ll}(t)l^2+\mathbf{x}^{\top}\mathbf{A}_{xf}(t)\mathbf{f}+\mathbf{x}^{\top}\mathbf{A}_{xl}(t)l+\mathbf{f}^{\top}\mathbf{A}_{fl}(t)l+A_l(t)l+A_c(t), \quad (7)$$

where $\mathbf{A}_{xx}\in\mathbb{R}^{n\times n}$, $\mathbf{A}_{ff}\in\mathbb{R}^{m\times m}$, $\mathbf{A}_{xf}\in\mathbb{R}^{n\times m}$, $\mathbf{A}_{xl}\in\mathbb{R}^{n}$, $\mathbf{A}_{fl}\in\mathbb{R}^{m}$, and $A_{ll},A_l,A_c\in\mathbb{R}$. Applying the first-order condition then yields $\mathbf{u}^*=\mathbf{\Lambda}^{-1}\frac{\partial V}{\partial \mathbf{x}}$. Substituting $\mathbf{u}^*$ and (7) back into (6), we obtain

$$\rho V=\frac{\partial V}{\partial t}+\frac{1}{2}\left(-\mathbf{A}_{xx}\mathbf{x}+\mathbf{A}_{xf}\mathbf{f}+\mathbf{A}_{xl}l\right)^{\top}\mathbf{\Lambda}^{-1}\left(-\mathbf{A}_{xx}\mathbf{x}+\mathbf{A}_{xf}\mathbf{f}+\mathbf{A}_{xl}l\right)+\left(r_f-\mu_l\right)l+\frac{1}{2}\Sigma_l l^2$$
$$+\left(A_{ll}l+\mathbf{A}_{xl}^{\top}\mathbf{x}+\mathbf{A}_{fl}^{\top}\mathbf{f}+A_l\right)\mu_l l-\left(\mathbf{A}_{ff}\mathbf{f}+\mathbf{A}_{xf}^{\top}\mathbf{x}+\mathbf{A}_{fl}l\right)^{\top}\mathbf{\Phi}\mathbf{f}+\frac{1}{2}\operatorname{tr}\left(\mathbf{A}_{ff}\mathbf{\Sigma}_f\right)$$
$$+\mathbf{A}_{fl}^{\top}\mathbf{\Sigma}_{fl}l+\mathbf{x}^{\top}\mathbf{B}\mathbf{f}-\frac{\gamma}{2}\left(\mathbf{x}^{\top}\mathbf{\Sigma}_P\mathbf{x}-2\mathbf{x}^{\top}\mathbf{\Sigma}_{Pl}l+\Sigma_l l^2\right).$$

Comparing the coefficients of $\mathbf{x},\mathbf{f},l$ and the constant terms on both sides, we obtain

$$\dot{\mathbf{A}}_{xx}(t)=\rho\mathbf{A}_{xx}(t)+\mathbf{A}_{xx}(t)^{\top}\mathbf{\Lambda}^{-1}\mathbf{A}_{xx}(t)-\gamma\mathbf{\Sigma}_P, \quad (8)$$
$$\dot{\mathbf{A}}_{ff}(t)=\rho\mathbf{A}_{ff}(t)-\mathbf{A}_{xf}^{\top}(t)\mathbf{\Lambda}^{-1}\mathbf{A}_{xf}(t)+\mathbf{A}_{ff}(t)^{\top}\mathbf{\Phi}+\mathbf{\Phi}^{\top}\mathbf{A}_{ff}(t), \quad (9)$$
$$\dot{A}_{ll}(t)=\rho A_{ll}(t)-\mathbf{A}_{xl}(t)^{\top}\mathbf{\Lambda}^{-1}\mathbf{A}_{xl}(t)-2\mu_l A_{ll}(t)-A_{ll}(t)\Sigma_l+\gamma\Sigma_l, \quad (10)$$
$$\dot{\mathbf{A}}_{xf}(t)=\rho\mathbf{A}_{xf}(t)+\mathbf{A}_{xx}(t)^{\top}\mathbf{\Lambda}^{-1}\mathbf{A}_{xf}(t)+\mathbf{A}_{xf}(t)\mathbf{\Phi}-\mathbf{B}, \quad (11)$$
$$\dot{\mathbf{A}}_{xl}(t)=\rho\mathbf{A}_{xl}(t)+\mathbf{A}_{xx}(t)^{\top}\mathbf{\Lambda}^{-1}\mathbf{A}_{xl}(t)-\mathbf{A}_{xl}(t)\mu_l-\gamma\mathbf{\Sigma}_{Pl}, \quad (12)$$
$$\dot{\mathbf{A}}_{fl}(t)=\rho\mathbf{A}_{fl}(t)-\mathbf{A}_{xf}(t)^{\top}\mathbf{\Lambda}^{-1}\mathbf{A}_{xl}(t)-\mathbf{A}_{fl}(t)\mu_l+\mathbf{\Phi}^{\top}\mathbf{A}_{fl}(t), \quad (13)$$
$$\dot{A}_l(t)=\rho A_l(t)-\mu_l A_l(t)-\mathbf{A}_{fl}(t)^{\top}\mathbf{\Sigma}_{fl}+\mu_l-r_f, \quad (14)$$
$$\dot{A}_c(t)=\rho A_c(t)-\frac{1}{2}\operatorname{tr}\left(\mathbf{A}_{ff}(t)\mathbf{\Sigma}_f\right), \quad (15)$$

with the following boundary condition

$$\mathbf{A}_{xx}(T)=\mathbf{0}_{n\times n},\ \ \mathbf{A}_{ff}(T)=\mathbf{0}_{m\times m},\ \ \mathbf{A}_{xf}(T)=\mathbf{0}_{n\times m},$$
$$\mathbf{A}_{xl}(T)=\mathbf{0}_{n\times 1},\ \ \mathbf{A}_{fl}(T)=\mathbf{0}_{m\times 1},\ \ A_{ll}(T)=A_l(T)=A_c(T)=0. \quad (16)$$

Still, we need to ensure that the matrix Riccati equation system admits a unique classical solution so that the corresponding *ansatz* of V in (7) as well as the optimal strategy is well-defined.

Lemma 3.1. *The matrix RDE system* (8)-(16) *admits a unique classical solution on* $[0,T]$. *Moreover,* $\mathbf{A}_{xx}(t)$ *is positive definite on* $[0,T)$ *and* $\mathbf{A}_{xx}(t)=\mathbf{K}_2(t)\mathbf{K}_1^{-1}(t)$ *where*

$$\frac{d}{dt}\begin{pmatrix}\mathbf{K}_1(t)\\ \mathbf{K}_2(t)\end{pmatrix}=\begin{pmatrix}-\frac{1}{\rho}\mathbf{I}_n & -\mathbf{\Lambda}^{-1}\\ -\gamma\mathbf{\Sigma}_P & \frac{1}{\rho}\mathbf{I}_n\end{pmatrix}\begin{pmatrix}\mathbf{K}_1(t)\\ \mathbf{K}_2(t)\end{pmatrix}. \quad (17)$$

Proof. Note that once $\mathbf{A}_{xx}$ is solved, $\mathbf{A}_{xf},\mathbf{A}_{ff},\mathbf{A}_{xl},\mathbf{A}_{ll},\mathbf{A}_{fl},\mathbf{A}_l$, and $\mathbf{A}_c$ can be solved sequentially as their associated differential equations in (8)-(16) become linear. Therefore, it suffices to prove the existence and uniqueness of $\mathbf{A}_{xx}$ to (8). This can be done by following the proof of Proposition 1 in Ma et al. (2019), where a comparison principle for matrix Riccati equation is utilized to find the upper bound and lower bound of the solution. We thus omit the details of the proof. □

While Lemma 3.1 establishes the well-posedness of the RDE system (8)-(16), the following assumption on a special price impact structure enables us to obtain simpler expressions for the optimal trading strategy and portfolio.

Assumption 1. $\mathbf{\Lambda}=\lambda\mathbf{\Sigma}_P$ with a constant $\lambda>0$.

As noted by Gârleanu and Pedersen (2013, 2016), Assumption 1 is natural and has its foundation in market microstructure. Particularly, a dealer, who stands in the opposite side of the manager's trade $\triangle \mathbf{x}_t$, holds the position for a period of time dt and unwinds it at the end, is exposed to the risk $\triangle \mathbf{x}_t^\top \boldsymbol{\Sigma}_P \triangle \mathbf{x}_t dt$. The trading cost is a compensation for risk. The positive constant λ can be understood as the risk aversion of the dealer. At time t, we define the single-period Markowitz mean-variance portfolio as $\mathbf{x}_t^{\mathrm{mv}} = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{B}\mathbf{f}_t$. The following theorem characterizes the optimal trading strategy.

Theorem 3.2. *With only the temporary price impact, i.e.,* $\mathbf{D} \equiv \mathbf{0}_{n\times 1}$ *(*$\mathbf{R} = \mathbf{C} = \mathbf{0}_{n\times n}$*),*

(i) *the optimal trading intensity is*

$$\mathbf{u}_t^* = \boldsymbol{\Lambda}^{-1}\left[\mathbf{A}_{xf}(t)\mathbf{f}_t + \mathbf{A}_{xl}(t)l_t - \mathbf{A}_{xx}(t)\mathbf{x}_t\right], \quad \text{for } t \in [0, T]. \tag{18}$$

The function (7) identifies the optimal value function in (4);

(ii) *for* $t \in [0, T)$, $\mathbf{u}_t^*$ *has the following representation*

$$\mathbf{u}_t^* = \mathbf{M}_t^{\mathrm{rate}}\left[\mathbf{M}_t^{\mathrm{aim}} - \mathbf{x}_t\right], \tag{19}$$

where

$$\mathbf{M}_t^{\mathrm{rate}} = \boldsymbol{\Lambda}^{-1}\mathbf{A}_{xx}(t),$$

$$\mathbf{M}_t^{\mathrm{aim}} = \gamma \mathbf{A}_{xx}(t)^{-1}\mathcal{E}_1(t)^{-1}\int_t^T e^{-\rho(s-t)}\mathcal{E}_1(s)\boldsymbol{\Sigma}_P\mathbb{E}_t\left[\left(\mathbf{x}_s^{\mathrm{mv}} + \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{Pl}l_s\right)\right]ds,$$

and $\mathcal{E}_1(t)$ *is the solution to the following linear ODE*

$$\dot{\mathcal{E}}_1(t) = -\mathbf{A}_{xx}(t)\boldsymbol{\Lambda}^{-1}\mathcal{E}_1(t), \quad \mathcal{E}_1(T) = \mathbf{I}_n; \tag{20}$$

(iii) *under Assumption 1, for* $t \in [0, T)$,

$$\mathbf{M}_t^{\mathrm{rate}} = \frac{a_{xx}(t)}{\lambda},$$

$$\mathbf{M}_t^{\mathrm{aim}} = \int_t^T \omega_u(s,t)\mathbb{E}_t\left[\mathbf{x}_s^{\mathrm{mv}} + \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{Pl}l_s\right]ds,$$

where $\omega_u(s,t) = \frac{\gamma}{a_{xx}(t)}e^{-\int_t^s\left(\rho + \frac{a_{xx}(u)}{\lambda}\right)du}$ *for* $0 \le t \le s \le T$, $a_{xx}(t)$ *is a nonnegative function given by*

$$a_{xx}(t) = -\frac{\gamma\lambda\left(1 - e^{(T-t)(y_+ - y_-)}\right)}{y_+ - y_- e^{(T-t)(y_+ - y_-)}} \ge 0,$$

and $y_\pm = \frac{\lambda}{2}\left(-\rho \pm \sqrt{\rho^2 + 4\frac{\gamma}{\lambda}}\right)$.

Proof. See Appendix A.1 □

Remark 3.3. As shown in Gârleanu and Pedersen (2013) and Glasserman and Xu (2013), the empirical estimates of $\boldsymbol{\Lambda}$ are typically small. Therefore, it is important to assess the stability of the solutions of the matrix Riccati equation system when $\boldsymbol{\Lambda}$ is small. To investigate this issue more closely, let us consider a special case when Assumption 1 holds. In this case, the $\mathbf{A}_{xx}$ in Equation (8) becomes

$$\dot{\mathbf{A}}_{xx}(t) = \rho\mathbf{A}_{xx}(t) + \frac{1}{\lambda}\mathbf{A}_{xx}(t)^\top\boldsymbol{\Sigma}_P^{-1}\mathbf{A}_{xx}(t) - \gamma\boldsymbol{\Sigma}_P.$$

Similar to Proposition 5.9 in Bensoussan et al. (2022), for a small $\lambda > 0$, one can show that

$$\|\mathbf{A}_{xx}(t)\| = O(\sqrt{\lambda}), \quad t \ge 0. \tag{21}$$

This implies that the right-hand side of the differential equation for $\mathbf{A}_{xx}$ is of order $O(1)$. In other words, the derivative of $\mathbf{A}_{xx}$ is bounded for small λ, which leads to a stable solution to the Riccati system. Similarly, one can show that terms $\mathbf{A}_{xf}, \mathbf{A}_{xl}$ are of order $O(\sqrt{\lambda})$ and terms $\mathbf{A}_{ff}, A_{ll}, \mathbf{A}_{fl}, A_l, A_c$ are of order $O(1)$. In view of these small-λ analysis, the Riccati system will not explode as λ goes to zero.

The trading intensity in (19) is of a target-chasing form such that the portfolio is rebalanced towards a benchmark $\mathbf{M}^{\mathrm{aim}}$ at the tracking rate $\mathbf{M}^{\mathrm{rate}}$. Roughly speaking, the target process $\mathbf{M}_t^{\mathrm{aim}}$ can be understood as the expectation of a weighted sum of all the future Markowitz portfolios within the investment horizon adjusted by the liability dynamic. We distinguish such a trading behavior from many existing results in classical dynamic ALM problems such as Leippold et al. (2004); Chiu and Li (2006); Yao et al. (2013); Wei et al. (2013) where the optimal strategy is to realize a target portfolio immediately. This target-chasing behavior appears because of the market frictions.

Corollary 3.4. *With only the temporary price impact,*

(i) *the optimal portfolio is*

$$\mathbf{x}_t^* = \mathcal{E}_2(t)^{-1}\mathbf{x}_0 + \gamma \int_0^t \int_s^T \mathcal{E}_2(s)\mathbf{\Lambda}^{-1}\mathcal{E}_1(s)^{-1}e^{-\rho(u-s)}\mathcal{E}_1(u)\mathbf{\Sigma}_P\mathbb{E}_s\left[\mathbf{x}_u^{\mathrm{mv}} + \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl}l_u\right]duds,$$

where $\mathcal{E}_1(t)$ *is given by* (20) *and* $\mathcal{E}_2(t)$ *solves*

$$\dot{\mathcal{E}}_2(t) = \mathbf{\Lambda}^{-1}\mathbf{A}_{xx}(t)\mathcal{E}_2(t), \quad \mathcal{E}_2(0) = \mathbf{I}_n, \tag{22}$$

(ii) *suppose Assumption 1 is enforced, the optimal portfolio is*

$$\mathbf{x}_t^* = e^{\int_0^t \frac{a_{xx}(l)}{\lambda}dl}\mathbf{x}_0 + \int_0^t \int_s^T \omega_x(s,u)\mathbb{E}_s\left[\mathbf{x}_u^{\mathrm{mv}} + \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl}l_u\right]duds,$$

where $\omega_x(s,u) = \frac{\gamma}{\lambda}e^{\int_0^s \frac{a_{xx}(l)}{\lambda}dl - \int_s^u\left(\rho + \frac{a_{xx}(l)}{\lambda}\right)dl}$ *for* $0 \le s \le u \le T$ *and* a_{xx} *is given by* (A.4)*.*

Proof. The results in (22) can be straightforwardly obtained by utilizing (A.1) and (A.3). Then one can verify (ii) using the expression $\mathbf{A}_{xx}(t) = a_{xx}(t)\mathbf{\Sigma}_P$. $\square$

4. Temporary and persistent price impact

In this section, we investigate the optimal trading strategy in our framework when the market frictions consists of two components: the temporary price impact and persistent price impact.

4.1. Optimal trading intensity and portfolio

For the ease of illustration, we first define a new vector $\mathbf{y} := (\mathbf{x}^\top, \mathbf{D}^\top)^\top \in \mathbb{R}^{2n}$ and some new matrices $\mathbf{Q}, \mathbf{N}_1 \in \mathbb{R}^{2n\times 2n}, \mathbf{N}_2, \mathbf{M}_1 \in \mathbb{R}^{2n\times n}, \mathbf{M}_2 \in \mathbb{R}^{2n\times m}, \mathbf{M}_3 \in \mathbb{R}^{2n}$ as follows:

$$\mathbf{Q} = \begin{pmatrix} \gamma\mathbf{\Sigma}_P & \mathbf{R} \\ \mathbf{R}^\top & \mathbf{0}_{n\times n}\end{pmatrix}, \quad \mathbf{N}_1 = \begin{pmatrix} \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{R}\end{pmatrix}, \quad \mathbf{N}_2 = \begin{pmatrix}\mathbf{I}_n \\ \mathbf{C}\end{pmatrix}, \tag{23}$$

$$\mathbf{M}_1 = \begin{pmatrix}\mathbf{C} \\ \mathbf{0}_{n\times n}\end{pmatrix}, \quad \mathbf{M}_2 = \begin{pmatrix}\mathbf{B} \\ \mathbf{0}_{n\times n}\end{pmatrix}, \quad \mathbf{M}_3 = \begin{pmatrix}\gamma\mathbf{\Sigma}_{Pl} \\ \mathbf{0}_{n\times n}\end{pmatrix}. \tag{24}$$

The original HJB equation (5) can then be rewritten as

$$\begin{aligned}\rho V =& \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \mathbf{y}}^\top(-\mathbf{N}_1\mathbf{y} + \mathbf{N}_2\mathbf{u}) - \frac{\partial V}{\partial \mathbf{f}}^\top\mathbf{\Phi}\mathbf{f} + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 V}{\partial\mathbf{f}\partial\mathbf{f}^\top}\mathbf{\Sigma}_f\right) + \frac{\partial V}{\partial l}\mu_l l + \frac{1}{2}\frac{\partial^2 V}{\partial l^2}\Sigma_l l^2 \\ &+ \frac{\partial^2 V}{\partial\mathbf{f}\partial l}^\top\mathbf{\Sigma}_{fl}l + (r_f - \mu_l)l - \frac{1}{2}\mathbf{y}^\top\mathbf{Q}\mathbf{y} + \mathbf{y}^\top\mathbf{M}_1\mathbf{u} - \frac{1}{2}\mathbf{u}^\top\mathbf{\Lambda}\mathbf{u} + \mathbf{y}^\top\mathbf{M}_2\mathbf{f} + \mathbf{y}^\top\mathbf{M}_3 l - \frac{\gamma}{2}\Sigma_l l^2.\end{aligned} \tag{25}$$

The first-order condition yields

$$\mathbf{u}^* = \mathbf{\Lambda}^{-1}\left(\mathbf{M}_1^\top\mathbf{y} + \mathbf{N}_2^\top\frac{\partial V}{\partial\mathbf{y}}\right). \tag{26}$$

Define $\mathbf{E}_{yy} := \begin{pmatrix}\mathbf{E}_{xx} & \mathbf{E}_{xD} \\ \mathbf{E}_{xD}^\top & \mathbf{E}_{DD}\end{pmatrix} \in \mathbb{R}^{2n\times 2n}$, $\mathbf{E}_{yf} := \begin{pmatrix}\mathbf{E}_{xf} \\ \mathbf{E}_{Df}\end{pmatrix} \in \mathbb{R}^{2n\times n}$, and $\mathbf{E}_{yl} := \begin{pmatrix}\mathbf{E}_{xl} \\ \mathbf{E}_{Dl}\end{pmatrix} \in \mathbb{R}^{2n}$. Consider the following ansatz for V:

$$V(t,\mathbf{y},\mathbf{f},l) = -\frac{1}{2}\mathbf{y}^\top\mathbf{E}_{yy}(t)\mathbf{y} + \frac{1}{2}\mathbf{f}^\top\mathbf{E}_{ff}(t)\mathbf{f} + \frac{1}{2}E_{ll}(t)l^2 + \mathbf{y}^\top\mathbf{E}_{yf}(t)\mathbf{f} + \mathbf{y}^\top\mathbf{E}_{yl}(t)l + \mathbf{f}^\top\mathbf{E}_{fl}(t)l + E_l(t)l + E_c(t), \tag{27}$$

where $\mathbf{E}_{yy} \in \mathbb{R}^{2n\times 2n}$, $\mathbf{E}_{ff} \in \mathbb{R}^{m\times m}$, $\mathbf{E}_{yf} \in \mathbb{R}^{2n\times m}$, $\mathbf{E}_{yl} \in \mathbb{R}^{2n}$, $\mathbf{E}_{fl} \in \mathbb{R}^m$, and $E_{ll}, E_l, E_c \in \mathbb{R}$. Substituting (26) and (27) into (25), and comparing the coefficients, we can similarly derive

$$\dot{\mathbf{E}}_{yy} = \left(\mathbf{N}_2^\top\mathbf{E}_{yy} - \mathbf{M}_1^\top\right)^\top\mathbf{\Lambda}^{-1}\left(\mathbf{N}_2^\top\mathbf{E}_{yy} - \mathbf{M}_1^\top\right) + \mathbf{E}_{yy}\left(\frac{1}{2}\rho\mathbf{I} + \mathbf{N}_1\right) + \left(\frac{1}{2}\rho\mathbf{I} + \mathbf{N}_1\right)^\top\mathbf{E}_{yy} - \mathbf{Q}, \tag{28}$$

$$\dot{\mathbf{E}}_{ff} = -\mathbf{E}_{yf}^\top\mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top\mathbf{E}_{yf} + \rho\mathbf{E}_{ff} + \mathbf{E}_{ff}^\top\mathbf{\Phi} + \mathbf{\Phi}^\top\mathbf{E}_{ff}, \tag{29}$$

$$\dot{E}_{ll} = -\mathbf{E}_{yl}^\top\mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top\mathbf{E}_{yl} + (\rho - 2\mu_l - \Sigma_l)E_{ll} + \gamma\Sigma_l, \tag{30}$$

$$\dot{\mathbf{E}}_{yf} = \left(\mathbf{N}_2^\top\mathbf{E}_{yy} - \mathbf{M}_1^\top\right)^\top\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top\mathbf{E}_{yf} + \rho\mathbf{E}_{yf} + \mathbf{N}_1^\top\mathbf{E}_{yf} + \mathbf{E}_{yf}\mathbf{\Phi} - \mathbf{M}_2, \tag{31}$$

$$\dot{\mathbf{E}}_{yl} = \left(\mathbf{N}_2^\top\mathbf{E}_{yy} - \mathbf{M}_1^\top\right)^\top\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top\mathbf{E}_{yl} + (\rho - \mu_l)\mathbf{E}_{yl} + \mathbf{N}_1^\top\mathbf{E}_{yl} - \mathbf{M}_3, \tag{32}$$

$$\dot{\mathbf{E}}_{fl} = -\mathbf{E}_{yf}^{\top}\mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{N}_2^{\top}\mathbf{E}_{yl} + (\rho-\mu_l)\,\mathbf{E}_{fl} + \mathbf{\Phi}^{\top}\mathbf{E}_{fl}, \tag{33}$$

$$\dot{E}_l = (\rho-\mu_l)\,E_l + \mathbf{E}_{fl}^{\top}\mathbf{\Sigma}_{fl} + \mu_l - r_f, \tag{34}$$

$$\dot{E}_c = \rho E_c - \frac{1}{2}\operatorname{tr}\left(\mathbf{E}_{ff}\mathbf{\Sigma}_f\right), \tag{35}$$

with the boundary conditions

$$\begin{aligned} &\mathbf{E}_{yy}(T) = \mathbf{0}_{2n\times 2n},\ \ \mathbf{E}_{ff}(T) = \mathbf{0}_{m\times m},\ \ \mathbf{E}_{yf}(T) = \mathbf{0}_{2n\times m},\\ &\mathbf{E}_{yl}(T) = \mathbf{0}_{2n\times 1},\ \ \mathbf{E}_{fl}(T) = \mathbf{0}_{m\times 1},\ \ E_{ll}(T) = E_l(T) = E_c(T) = 0. \end{aligned} \tag{36}$$

Obviously, we again obtain a matrix RDE system in (28)–(36). If the solution of (28) exists, then (32) – (35) become linear, and they can be solved sequentially using variation of constants formula. As for the matrix RDE (28), we can apply the well-known Radon's lemma, which connects the non-linear equations of Riccati type and the linear equations, and then provide the solution with an explicit representation as follows.

Lemma 4.1. *Consider the solution $(\mathbf{H}_1(t), \mathbf{H}_2(t))^{\top}$ of the following linear equation system:*

$$\frac{d}{dt}\begin{pmatrix}\mathbf{H}_1(t)\\ \mathbf{H}_2(t)\end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\rho\mathbf{I} - \mathbf{N}_1 + \mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{M}_1^{\top} & -\mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{N}_2^{\top}\\ \mathbf{M}_1\mathbf{\Lambda}^{-1}\mathbf{M}_1^{\top} - \mathbf{Q} & \frac{1}{2}\rho\mathbf{I} + \mathbf{N}_1 - \mathbf{M}_1\mathbf{\Lambda}^{-1}\mathbf{N}_2^{\top}\end{pmatrix}\begin{pmatrix}\mathbf{H}_1(t)\\ \mathbf{H}_2(t)\end{pmatrix}, \tag{37}$$

with $\mathbf{H}_1(T) = \mathbf{I}_{2n}$ and $\mathbf{H}_2(T) = \mathbf{0}_{2n\times 2n}$. If $\mathbf{H}_1(t)$ is non-singular for $t \in [0, T]$, then $\mathbf{E}_{yy}(t) = \mathbf{H}_2(t)\mathbf{H}_1^{-1}(t)$ is a solution of (28).

Proof. We rewrite (28) as follows

$$\dot{\mathbf{E}}_{yy} = \mathbf{E}_{yy}\mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{N}_2^{\top}\mathbf{E}_{yy} + \left(\frac{1}{2}\rho\mathbf{I} + \mathbf{N}_1 - \mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{M}_1^{\top}\right)^{\top}\mathbf{E}_{yy} + \mathbf{E}_{yy}\left(\frac{1}{2}\rho\mathbf{I} + \mathbf{N}_1 - \mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{M}_1^{\top}\right) + \mathbf{M}_1\mathbf{\Lambda}^{-1}\mathbf{M}_1^{\top} - \mathbf{Q}. \tag{38}$$

Then the result can be established by applying Theorem 3.1.1 in Abou-Kandil et al. (2003). □

If we define the finite escape time of the matrix $\mathbf{E}_{yy}$ as $t_{\infty} := \inf\left\{t < T; \det|\mathbf{H}_1(t)| = 0\right\}$, Lemma 4.2 provides a sufficient condition such that $t_{\infty} \notin [0, T]$, i.e., $\mathbf{E}_{yy}(t)$ will never explode on $[0, T]$.

Lemma 4.2. *Suppose that $\gamma\mathbf{\Sigma}_P - \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}^{\top}$ is positive definite. The matrix RDE system (28)-(36) admits a unique bounded solution on $[0, T]$ if $T < \frac{1}{2(\psi_1+\sqrt{\psi_0\psi_2})}$, where*

$$\begin{aligned} \psi_0 &= \left\|\mathbf{R}^{\top}\left(\gamma\mathbf{\Sigma}_P - \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}^{\top}\right)^{-1}\mathbf{R}\right\|,\\ \psi_1 &= \left\|\tfrac{1}{2}\rho\mathbf{I} + \mathbf{R}\right\|, \quad \psi_2 = \left\|\mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}^{\top}\right\|. \end{aligned}$$

Moreover, $\mathbf{E}_{xx}(t)$ is positive semidefinite for $t \in [0, T]$.

Proof. See Appendix A.2. □

The aforementioned lemma provides a sufficient condition to ensure the local existence for the target matrix RDE system. However, the length of time period is highly sensible to the model parameters and could be quite short in some extreme circumstances. As a result, we give another sufficient condition below under which a unique global solution always exists, regardless of the length of the planning horizon.

Lemma 4.3. *Suppose that $\mathbf{\Sigma}_P$, $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{\Lambda}$ are all diagonal matrices with the i-th diagonal element being $\sigma_{P,i}^2, r_i, c_i$, and λ_i for $i = 1, 2, \cdots, n$, respectively. If $\gamma\sigma_{P,i}^2 > \frac{c_i^2}{\lambda_i}$ and $\frac{c_i}{\lambda_i} < r_i < \rho$, then the matrix RDE system (28)-(36) admits a unique bounded solution for any finite $T > 0$.*

Proof. See Appendix A.3. □

Roughly speaking, the conditions in Lemma 4.2 and 4.3 commonly indicate that the (squared) persistent price impact $\mathbf{C}$ discounted by the temporary price impact $\mathbf{\Lambda}$ cannot be too large compared to the trader's perceived risk cost $\gamma\mathbf{\Sigma}_P$. Otherwise, it is cheap and risk-affordable for the trader to manipulate the stock prices to achieve an abnormally high profit in the finite horizon $[0, T]$. This is partially consistent with the well-posedness condition in Lemma 1 of Gârleanu and Pedersen (2016) that requires $\mathbf{C}$ to be moderate. We will also conduct a numerical study in Section 6 to investigate the general feature of the solutions.

Define $\mathbf{L}_x(t) := \mathbf{E}_{xx}(t) + \mathbf{C}^{\top}\mathbf{E}_{xD}(t)^{\top} - \mathbf{C}^{\top}$. The following theorem provides the optimal trading strategy when both temporary and persistent price impacts are present.

Theorem 4.4. *With both temporary and persistent price impacts, and suppose that the matrix RDE system (28)-(36) admits a bounded solution, then*

(i) *the optimal trading intensity is*

$$\mathbf{u}_t^* = \mathbf{\Lambda}^{-1}\Big[-\Big(\mathbf{E}_{xD}(t)+\mathbf{C}^\top\mathbf{E}_{DD}(t)\Big)\mathbf{D}_t + \mathbf{N}_2^\top\mathbf{E}_{yf}(t)\mathbf{f}_t + \mathbf{N}_2^\top\mathbf{E}_{yl}(t)l_t - \mathbf{L}_x(t)\mathbf{x}_t\Big]. \tag{39}$$

The function (27) *identifies the optimal value function in* (4)*;*

(ii) *for $t \in [0,T)$, if $\mathbf{L}_x(t)$ is non-singular, $\mathbf{u}_t^*$ has the following representation*

$$\mathbf{u}_t^* = \tilde{\mathbf{M}}_t^{\text{rate}}\Big[\tilde{\mathbf{M}}_{flt}^{\text{aim}} + \tilde{\mathbf{M}}_{Dt}^{\text{aim}} - \mathbf{x}_t\Big], \tag{40}$$

where

$$\tilde{\mathbf{M}}_t^{\text{rate}} = \mathbf{\Lambda}^{-1}\mathbf{L}_x(t),$$

$$\tilde{\mathbf{M}}_{flt}^{\text{aim}} = \mathbf{L}_x(t)^{-1}\mathbf{N}_2^\top\int_t^T e^{-\rho(s-t)}\tilde{\mathcal{E}}_1(t)^{-1}\tilde{\mathcal{E}}_1(s)\begin{pmatrix}\gamma\mathbf{\Sigma}_P\\ \mathbf{0}_{n\times n}\end{pmatrix}\mathbb{E}_t\Big[\mathbf{x}_s^{\text{mv}} + \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl}l_s\Big]ds,$$

$$\tilde{\mathbf{M}}_{Dt}^{\text{aim}} = -\mathbf{L}_x(t)^{-1}\Big(\mathbf{E}_{xD}(t)+\mathbf{C}^\top\mathbf{E}_{DD}(t)\Big)\mathbf{D}_t,$$

and $\tilde{\mathcal{E}}_1(t) \in \mathbb{R}^{2n\times 2n}$ is the solution to the following linear ODE

$$\dot{\tilde{\mathcal{E}}}_1(t) = -\Big[\Big(\mathbf{N}_2^\top\mathbf{E}_{yy}(t) - \mathbf{M}_1^\top\Big)^\top\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top + \mathbf{N}_1\Big]\tilde{\mathcal{E}}_1(t), \quad \tilde{\mathcal{E}}_1(T) = \mathbf{I}_{2n}.$$

Proof. See Appendix A.4. □

The following corollary, as a direct result of Theorem 4.4, offers the corresponding optimal portfolio.

Corollary 4.5. *Under the assumptions of Theorem 4.4 and assume that $\mathbf{L}_x(t)$ is non-singular, then the optimal portfolio is*

$$\begin{aligned}\mathbf{x}_t^* &= \tilde{\mathcal{E}}_2(t)^{-1}\mathbf{x}_0 - \int_0^t \tilde{\mathcal{E}}_2(t)^{-1}\tilde{\mathcal{E}}_2(s)\mathbf{\Lambda}^{-1}\Big(\mathbf{E}_{xD}(s)+\mathbf{C}^\top\mathbf{E}_{DD}(s)\Big)\mathbf{D}_s ds\\ &\quad+\int_0^t\int_s^T e^{-\rho(u-s)}\tilde{\mathcal{E}}_2(t)^{-1}\tilde{\mathcal{E}}_2(s)\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top\tilde{\mathcal{E}}_1(s)^{-1}\tilde{\mathcal{E}}_1(u)\begin{pmatrix}\gamma\mathbf{\Sigma}_P\\ \mathbf{0}_{n\times n}\end{pmatrix}\mathbb{E}_s\Big[\mathbf{x}_u^{\text{mv}} + \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl}l_u\Big]duds,\end{aligned}$$

where $\tilde{\mathcal{E}}_2(t) \in \mathbb{R}^{n\times n}$ is the solution to the following linear ODE.

$$\dot{\tilde{\mathcal{E}}}_2(t) = \tilde{\mathbf{M}}_t^{\text{rate}}\mathcal{E}_2(t), \quad \tilde{\mathcal{E}}_2(0) = \mathbf{I}_n. \tag{41}$$

The previous discussions are built on a continuous-time economy. As noted by Gârleanu and Pedersen (2016), such a formulation addresses the dynamic trading problems under market frictions that arise when an investor trades at a relatively high trading frequency. Meanwhile, we offer a discrete-time solution in Appendix B for an associated low-frequency trading problem, similar to what has been done in Gârleanu and Pedersen (2013) and Collin-Dufresne et al. (2020), so that practitioners can choose suitable ALM strategies in different situations. Moreover, we find that the optimal strategies are consistent in these two formulations, which also echoes the results of Gârleanu and Pedersen (2013) and Collin-Dufresne et al. (2020).

4.2. Small-price-impact asymptotics

Theorem 4.4 characterizes the optimal trading strategy under the fixed parameters of market frictions, i.e., $\mathbf{\Lambda}, \mathbf{R}$ and $\mathbf{C}$. To study how these parameters affect the value function and the trading strategy, we now conduct a small-price-impact asymptotic analysis. Following Ekren and Muhle-Karbe (2019), we consider the following assumption,

$$\mathbf{\Lambda} = \epsilon^2\mathbf{\Lambda}_1, \quad \mathbf{C} = \epsilon\mathbf{C}_1, \quad \mathbf{R} = \epsilon^{-1}\mathbf{R}_1, \tag{42}$$

where $\epsilon > 0$ is a small value. This assumption indicates a highly liquid market where both the temporary and persistent price impacts are small and the market resiliency is large. As noted by Ekren and Muhle-Karbe (2019), this scaling is chosen so that neither of the frictions dominates the other in the limit.

Before proceeding to the asymptotic analysis, we first study the benchmark case when there is no frictions. Suppose $\mathbf{\Lambda} = \mathbf{C} = \mathbf{R} = \mathbf{0}_{n\times n}$ and $\mathbf{D} \equiv \mathbf{0}_{n\times 1}$, one can directly solve the following ALM problem,

$$V^0(t,\mathbf{f},l) = \max_{\mathbf{x}_t}\mathbb{E}_t\left[\int_t^T e^{-\rho(s-t)}\Big\{\big(r_f-\mu_l\big)l_s + \mathbf{x}_s^\top\mathbf{B}\mathbf{f}_s - \frac{\gamma}{2}\Big[\mathbf{x}_s^\top\mathbf{\Sigma}_P\mathbf{x}_s + \Sigma_l l_s^2 - 2\mathbf{x}_s^\top\mathbf{\Sigma}_{Pl}l_s\Big]\Big\}ds\right]. \tag{43}$$

According to the classic dynamic programming equation, the HJB equation for Problem (43) is

$$\frac{\partial V^0}{\partial t} + \max_{\mathbf{x}\in\mathbb{R}^n} \mathcal{L}^{\mathbf{x}} V^0 = \rho V^0, \tag{44}$$

where the operator $\mathcal{L}^{\mathbf{x}}$ is defined as follows:

$$\begin{aligned}\mathcal{L}^{\mathbf{x}}\hat{V} := &-\frac{\partial \hat{V}}{\partial \mathbf{f}}^{\top} \mathbf{\Phi f} + \frac{1}{2}\operatorname{tr}\left(\frac{\partial^2 \hat{V}}{\partial \mathbf{f}\partial \mathbf{f}^{\top}}\mathbf{\Sigma}_f\right) + \frac{\partial \hat{V}}{\partial l}\mu_l l + \frac{1}{2}\frac{\partial^2 \hat{V}}{\partial l^2}\Sigma_l l^2 + \frac{\partial^2 \hat{V}}{\partial \mathbf{f}\partial l}^{\top}\mathbf{\Sigma}_{fl} l \\ &+ (r_f-\mu_l)l + \mathbf{x}^{\top}\mathbf{B f} - \frac{\gamma}{2}\left[\mathbf{x}^{\top}\mathbf{\Sigma}_P\mathbf{x} + \Sigma_l l^2 - 2\mathbf{x}^{\top}\mathbf{\Sigma}_{Pl} l\right].\end{aligned} \tag{45}$$

The first-order condition from (44) yields the optimal portfolio choice in absence of market frictions as $\mathbf{x}^{\mathrm{nf}}(\mathbf{f}, l) := \frac{1}{\gamma}\mathbf{\Sigma}_P^{-1}\mathbf{B f} + \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl} l$. Since $\mathbf{f}$ and l are the drifted Brownian motions, the optimal portfolio without frictions $\mathbf{x}^{\mathrm{nf}}$ is of a Brownian-motion type, which is in stark contrast to the optimal portfolio with frictions $\mathbf{x}^*$ in Corollary 4.5, which is absolutely continuous.

For simplicity, we write $\hat{\mathbf{C}}_1^{\top} := \begin{pmatrix}\mathbf{I}_n & \mathbf{C}_1^{\top}\end{pmatrix}$. The following theorem shows the relationship between the optimal value function V^{ϵ} for problem (4) and its frictionless counterpart V^0 in (43).

Theorem 4.6. *Under the regime (42), suppose $\mathbf{C}_1^{-1}\mathbf{R}_1 + \mathbf{R}_1\mathbf{C}_1^{-1}$ is positive definite, then the value function V^{ϵ} of Problem (4) has the following first-order expansion around the small liquidity parameter $\epsilon > 0$,*

$$\begin{aligned}&V^{\epsilon}(t, \mathbf{x}, \epsilon\mathbf{D}, f, l) \\ &= \underbrace{V^0(t, \mathbf{f}, l)}_{\textit{benchmark}} - \underbrace{\epsilon\left[V^1(t,l) + \frac{1}{2}\left(\mathbf{x}^{\top} - \mathbf{x}^{\mathrm{nf}}(\mathbf{f}, l)^{\top}, \mathbf{D}^{\top}\right)\mathbf{K}\left(\mathbf{x}^{\top} - \mathbf{x}^{\mathrm{nf}}(\mathbf{f}, l)^{\top}, \mathbf{D}^{\top}\right)^{\top} + \mathbf{x}^{\top}\mathbf{D} - \frac{1}{2}\mathbf{D}^{\top}\mathbf{C}_1^{-1}\mathbf{D}\right]}_{\textit{first-order approximation of the loss from frictions}} + o(\epsilon),\end{aligned} \tag{46}$$

where $\mathbf{K} \in \mathbb{R}^{2n\times 2n}$ is positive definite and is the maximal solution of the following algebraic Riccati equation:

$$\mathbf{K}^{\top}\hat{\mathbf{C}}_1\mathbf{\Lambda}_1^{-1}\hat{\mathbf{C}}_1^{\top}\mathbf{K} + \begin{pmatrix}\mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{R}_1\end{pmatrix}\mathbf{K}^{\top} + \mathbf{K}\begin{pmatrix}\mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{R}_1\end{pmatrix} - \begin{pmatrix}\gamma\mathbf{\Sigma}_P & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \rho\left(\mathbf{C}_1^{-1}\mathbf{R}_1 + \mathbf{R}_1\mathbf{C}_1^{-1}\right)\end{pmatrix} = \mathbf{0}_{2n\times 2n}, \tag{47}$$

and V^1 admits the following representation

$$\begin{aligned}V^1(t,l) &= \mathbb{E}_t\left[\int_t^T e^{-\rho(s-t)} a(l_s) ds\right], \\ a(l) &:= \frac{1}{2}\mathbf{\Sigma}_{Pl}^{\top}{\mathbf{\Sigma}_P^{-1}}^{\top}\mathbf{K}_{11}\mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl}\Sigma_l l^2 + \frac{1}{\gamma}\mathbf{\Sigma}_{Pl}^{\top}\mathbf{K}_{11}{\mathbf{\Sigma}_P^{-1}}^{\top}\mathbf{B}\mathbf{\Sigma}_{fl} l + \frac{1}{2\gamma^2}\operatorname{tr}\left(\mathbf{B}^{\top}{\mathbf{\Sigma}_P^{-1}}^{\top}\mathbf{K}_{11}\mathbf{\Sigma}_P^{-1}\mathbf{B}\mathbf{\Sigma}_f\right),\end{aligned} \tag{48}$$

where $\mathbf{K}_{11} \in \mathbb{R}^{n\times n}$ is the submatrix of $\mathbf{K}$ formed by intersecting the first n rows with the first n columns.

Moreover, the first-order expansion of V^{ϵ} in (46) is achieved when the manager adopts the almost-optimal trading strategy $\mathbf{u}^{\epsilon}$:

$$\mathbf{u}^{\epsilon}(\mathbf{x}_t, \mathbf{D}_t, \mathbf{f}_t, l_t) = \epsilon^{-1}\mathbf{\Lambda}_1^{-1}\hat{\mathbf{C}}_1^{\top}\mathbf{K}\begin{pmatrix}\mathbf{x}^{\mathrm{nf}}(\mathbf{f}_t, l_t) - \mathbf{x}_t\\ -\epsilon\mathbf{D}_t\end{pmatrix}. \tag{49}$$

Proof. See Appendix A.5. $\square$

In (46), we scale the distortion process and consider $\epsilon\mathbf{D}$ among the arguments of V^{ϵ}, which is consistent with Ekren and Muhle-Karbe (2019). This setting is reasonable as market frictions are small in our regime (42) and we expect the distortion to converge to zero. At time zero, recall that $\mathbf{D}_0 = \mathbf{0}_{n\times 1}$, one may simply let $\mathbf{D} = \mathbf{0}_{n\times 1}$ in (46) and obtain the first-order approximation of V^{ϵ}.

According to (49), it is clear that when market frictions are small, or equivalently, ϵ and $\mathbf{D}_t$ go to zero, the almost-optimal trading strategy $\mathbf{u}^{\epsilon}$ adjusts toward the frictionless portfolio $\mathbf{x}^{\mathrm{nf}}$ with a speed magnified by ϵ^{-1}. In addition, the value function V^{ϵ} in Equation (46) converges to the frictionless counterpart V^0 as $\epsilon \downarrow 0$. In other words, despite the fact that the asymptotic strategy gradually follows its frictionless counterpart, the presence of market frictions forbids it to converge to its frictionless counterpart as $\epsilon \downarrow 0$, and the value function with small market frictions can be well approximated by its frictionless counterpart.

5. Optimal funding ratio

The previous analysis answers the question that, given an initial asset value, stock holding and liability value, how can the company manage dynamically the risk from future asset allocation and future liability with frictions in the financial market. In this section, we provide one way for the manager about how to determine an initial capital structure at time zero.

Choosing a proper initial capital structure is important in ALM practice; indeed, it can be decided using the funding ratio, i.e. the ratio of the asset value to the liability see, e.g., Leippold et al. (2004) and Chiu and Li (2006). Following Chiu and Li (2006), we assume that the asset value is a linear function of the liability at $t = 0$. This dependency is realistic as argued by Chiu and Li (2006) because the company can increase the capital of investment by issuing loans. Mathematically, we write $a_0 = \alpha + \beta l_0$, where $\alpha > 0$ represents the wealth base for investment and $\beta \in (0, 1]$ denotes the proportion of liability that is raised to increase the asset value.

We consider the case that the initial portfolio is determined by $\mathbf{x}_0 = a_0(\hat{\mathbf{x}}^\top \mathbf{1})^{-1}\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}^\top$ is a target portfolio predetermined by the company. A noteworthy example is the static Markowitz portfolio $\hat{\mathbf{x}} = \mathbf{x}^{\mathrm{mv}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_P^{-1}\mathbf{B}\mathbf{f}_0$. Recall that under the temporary and persistent price impact, the optimal value function at time 0 is $V(0, \mathbf{x}_0, \mathbf{D}_0, \mathbf{f}_0, l_0)$ which is in a quadratic form. We define the optimal funding ratio as $\kappa^* := \frac{a_0^*}{l_0^*}$, where (a_0^*, l_0^*) solves the following quadratic programming problem:

$$\begin{aligned} &\max_{a_0, l_0} V(0, a_0(\hat{\mathbf{x}}^\top \mathbf{1})^{-1}\hat{\mathbf{x}}, \mathbf{D}_0, \mathbf{f}_0, l_0) \\ &\text{s.t.} \quad a_0 = \alpha + \beta l_0, \qquad l_0 > 0. \end{aligned} \tag{50}$$

To obtain κ^* more explicitly, we here assume temporary price impact only. We further impose Assumption 1 and assume that $\boldsymbol{\Phi} = \boldsymbol{\Sigma}_f = \mathbf{0}_{m\times m}$ and $\boldsymbol{\Sigma}_{fl} = \mathbf{0}_{m\times 1}$ so that the dynamic predictors are not concerned and $\mathbf{f}_t \equiv \mathbf{f}_0$. Therefore, the vector of excess returns is time-invariant and is denoted by $\boldsymbol{\mu}_P := \mathbf{B}\mathbf{f}_0$. Recall that a_{xx} is given by (A.4). We define the following non-negative scalar functions

$$a_{xl}(t) := \int_t^T e^{-\int_t^s \left(\rho - \mu_l + \frac{a_{xx}(u)}{\lambda}\right)du} ds, \qquad a_{xf}(t) := \int_t^T e^{-\int_t^s \left(\rho + \frac{a_{xx}(u)}{\lambda}\right)du} ds,$$
$$a_{ff}(t) := \int_t^T e^{-\rho(s-t)} \frac{a_{xf}^2(s)}{\lambda} ds, \qquad a_{fl}(t) := \int_t^T e^{-(\rho-\mu_l)(s-t)} \frac{a_{xf}(s)a_{xl}(s)}{\lambda} ds.$$

In this simplified case, the solution of the RDE system (8)–(16) is explicitly given by

$$\begin{aligned} &\mathbf{A}_{xx}(t) = a_{xx}(t)\boldsymbol{\Sigma}_P, \qquad \mathbf{A}_{ff}(t) = a_{ff}(t)\mathbf{B}^\top \boldsymbol{\Sigma}_P^{-1}\mathbf{B}, \\ &A_{ll}(t) = \int_t^T e^{-(\rho - 2\mu_l - \Sigma_l)(s-t)} \left(a_{xl}^2(s) \frac{\gamma^2}{\lambda^2} \boldsymbol{\Sigma}_{Pl}^\top \boldsymbol{\Sigma}_P^{-1} \boldsymbol{\Sigma}_{Pl} - \gamma \Sigma_l \right) ds, \\ &\mathbf{A}_{xf}(t) = a_{xf}(t)\mathbf{B}, \qquad \mathbf{A}_{xl}(t) = \gamma a_{xl}(t)\boldsymbol{\Sigma}_{Pl}, \qquad \mathbf{A}_{fl}(t) = \gamma a_{fl}(t)\mathbf{B}^\top \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{Pl}, \\ &A_l(t) = \int_t^T e^{-(\rho-\mu_l)(s-t)}(r_f - \mu_l) ds, \qquad A_c(t) \equiv 0. \end{aligned} \tag{51}$$

By Theorem 3.2, the value function at time $t = 0$ with only the temporary price impact is exactly $V(0, \mathbf{x}_0, \mathbf{f}_0, l_0)$, where V satisfies (7) with the parameters solved above. Let $\mathbf{z} := (a_0, l_0)^\top \in \mathbb{R}^2$. The optimal funding ratio between the initial asset and initial liability is then calculated through $\kappa^* = \frac{a_0^*}{l_0^*}$, where (a_0^*, l_0^*) solves:

$$\begin{aligned} &\max_{\mathbf{z}} \frac{1}{2}\mathbf{z}^\top \mathbf{Q}_{al}\mathbf{z} + \mathbf{c}_{al}^\top \mathbf{z} + c_0 \\ &\text{s.t.} \quad a_0 = \alpha + \beta l_0, \qquad l_0 > 0. \end{aligned} \tag{52}$$

where

$$\mathbf{Q}_{al} = \begin{pmatrix} -a_{xx}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-2}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_P \hat{\mathbf{x}} & \gamma a_{xl}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_{Pl} \\ \gamma a_{xl}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_{Pl} & A_{ll}(0) \end{pmatrix}. \tag{53}$$

The quadratic programming problem (52) admits a unique optimal solution pair (a_0^*, l_0^*) when $\mathbf{Q}_{al} < 0$, which clearly holds when $\boldsymbol{\Sigma}_{Pl} = 0$. In the following, we provide an explicit representation for κ^* under a special case. We first ignore the constraint $l_0 > 0$ and solve the optimal z^* through the Karush–Kuhn–Tucker condition:

$$\mathbf{Q}_{al}\mathbf{z}^* + \mathbf{c}_{al} + \nu(1, -\beta)^\top = \mathbf{0}_{2\times 1},$$

where $\nu \in \mathbb{R}$ is the Lagrange multiplier. Combining the above equation and the constraint $(1, -\beta)^\top \mathbf{z}^* = \alpha$, we then derive

$$\kappa^* = \frac{\beta\gamma a_{xl}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_{Pl} + A_{ll}(0) - \frac{\beta}{\alpha}\phi(\beta)}{\beta a_{xx}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-2}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_P \hat{\mathbf{x}} + \gamma a_{xl}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_{Pl} - \frac{1}{\alpha}\phi(\beta)}, \tag{54}$$

where $\phi(\beta) = \beta a_{xf}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1}\hat{\mathbf{x}}^\top \boldsymbol{\mu}_P + \gamma a_{fl}(0)\boldsymbol{\mu}_P^\top \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{Pl} + A_l(0)$. The above κ^* becomes a feasible and hence optimal funding ratio if the corresponding $l_0^* > 0$. The following proposition provides the conditions for that.

Proposition 5.1. *Suppose Assumption 1 holds and the return predicting-factor is not concerned, i.e.,* $\boldsymbol{\Phi} = \boldsymbol{\Sigma}_f = \mathbf{0}_{m\times m}$ *and* $\boldsymbol{\Sigma}_{fl} = \mathbf{0}_{m\times 1}$.

(i) *If the following cubic polynomial inequality in terms of* β *is satisfied*

$$\left[a_{xx}(0)(\hat{\mathbf{x}}^\top \mathbf{1})^{-2}\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}_P \hat{\mathbf{x}}\beta^2 - A_{ll}(0)\right][\zeta_1\beta + \zeta_0] < 0, \tag{55}$$

where ζ_1, ζ_0 are constants independent of β given by

$$\zeta_1 = a_{xx}(0)(\hat{\mathbf{x}}^\top \mathbf{1})^{-2}\hat{\mathbf{x}}^\top \mathbf{\Sigma}_P \hat{\mathbf{x}} - \frac{1}{\alpha} a_{xf}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1} \hat{\mathbf{x}}^\top \boldsymbol{\mu}_P,$$

$$\zeta_0 = \gamma a_{xl}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-1} \hat{\mathbf{x}}^\top \mathbf{\Sigma}_{Pl} - \frac{\gamma}{\alpha} a_{fl}(0)\boldsymbol{\mu}_P^\top \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl} - \frac{1}{\alpha}A_l(0),$$

then κ^* in (54) is a feasible funding ratio and attains the optimal value in Problem (50).

(ii) *Moreover, if the asset prices and the liability are independent, i.e., $\mathbf{\Sigma}_{Pl} = 0$, then the condition (55) reduces to*

$$\zeta_1 \beta < \frac{1}{\alpha} A_l(0).$$

Proof. We only need to show that under the conditions in Proposition 5.1, κ^* is a feasible funding ratio. It remains to address the constraint $l_0^* > 0$. Since $\kappa^* = \frac{a_0^*}{l_0^*} = \frac{\alpha}{l_0^*} + \beta$ with $\alpha > 0$, then $l_0^* > 0$ is equivalent to $\kappa^* > \beta$. Indeed, by utilizing the representation (54), we obtain

$$\kappa^* - \beta = \frac{-a_{xx}(0)\left(\hat{\mathbf{x}}^\top \mathbf{1}\right)^{-2}\hat{\mathbf{x}}^\top \mathbf{\Sigma}_P \hat{\mathbf{x}}\beta^2 + A_{ll}(0)}{\zeta_1\beta + \zeta_0}.$$

Therefore, from the above relationship, it is clear that $\kappa^* > \beta$ is further equivalent to the cubic polynomial inequality (55) of β. In other words, under the condition (55), κ^* given by (54) is a feasible funding ratio that optimizes Problem (50).

Suppose the asset prices and the liability are independent, i.e. $\mathbf{\Sigma}_{Pl} = 0$. According to the expressions in (51), $A_{ll}(0) = -\int_0^T e^{-(\rho - 2\mu_l - \Sigma_l)s}\gamma \Sigma_l ds < 0$. From Theorem 3.2 (iii), we have $a_{xx}(0) \geq 0$. Then the first term in the condition (55) is positive. Moreover, $\zeta_0 = -\frac{1}{\alpha}A_l(0)$. The condition (55) can be further simplified as $\zeta_1\beta < \frac{1}{\alpha}A_l(0)$. □

The inequalities in Proposition 5.1 can be efficiently verified in practice. In Section 6, the model parameters are chosen so that the condition in (55) is satisfied when we analyze the optimal initial funding ratio κ^*.

6. Numerical study

One of the main contributions of this paper is the consideration of the transaction costs in the ALM problem. In this section, we investigate numerically the effects of liability and the market frictions on the manager's trading strategies. First, a comparative statics analysis of the trading strategy with respect to the liability process and the temporary price impact level is given in Section 6.1. Second, based on the theory developed in Section 5, we study the optimal funding ratio and provide its sensitivity result in Section 6.2. Finally, in Section 6.3, we comprehensively analyze the scenario where both temporary and persistent price impacts are considered, and illustrate their different implications on the manager's trading behaviors.

In the subsequent analysis, for the sake of clarity, we consider the one-dimensional setting ($n = m = 1$). Furthermore, to focus on the analysis on the effects of liability and market frictions, we temporarily put the return predictability aside and assume $\Phi = \Sigma_f = \Sigma_{fl} = 0$ so that $f_t \equiv f_0$. Suppose that the planning horizon is 5 years, i.e., $T = 5$, and the asset-liability manager monthly updates the stock holdings, i.e., the time-discretization step has a length of $\Delta t = 1/12$. The initial values of stock holdings and liability are assumed to be $x_0 = 10^7$ and $l_0 = 5 \times 10^6$, respectively. Let the manager's risk aversion parameter be $\gamma = 2 \times 10^{-7}$, the discount rate be $\rho = 0$, and the annual risk-free interest rate be $r_f = 0.01$.

6.1. Trading strategy with only temporary price impact

When the trading frequency is substantially lower than the price resiliency, it is reasonable to shift away from the persistent price impact to assume $R = C = 0$ so that the return distortion process is negligible, i.e., $D \equiv 0$. To produce the financial insights that closely link with the practice, we follow the setting in Leippold et al. (2004) to use a realistic parameter set for modeling the stock price and liability processes, while, at the same time, we inherit the parameter setting for the temporary price impact from Berry-Stölzle (2008b). Specifically, we have

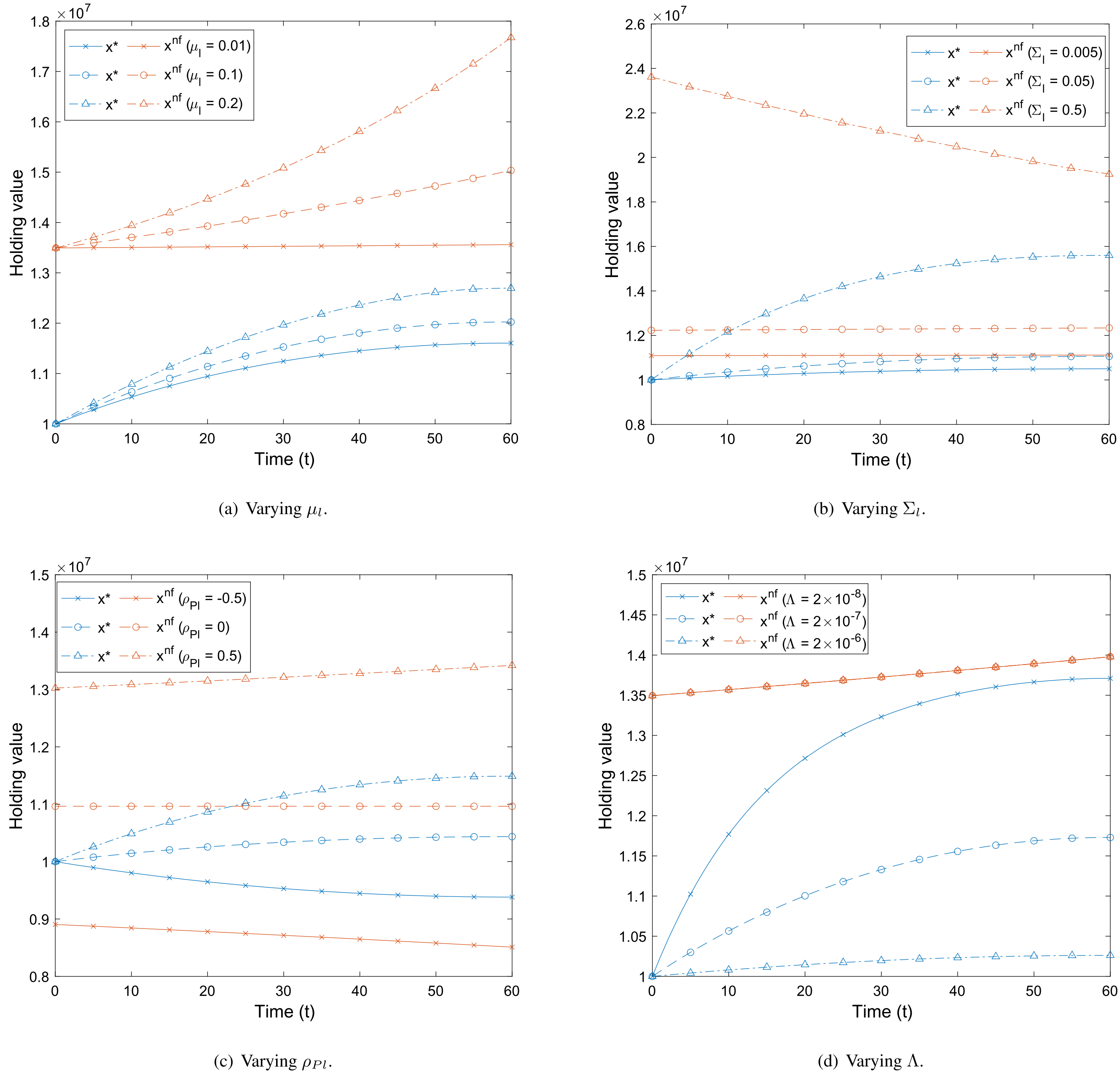
$$B = 1, \quad f_0 = 0.1200, \quad \Sigma_P = 0.0589, \quad \mu_l = 0.0400,$$
$$\Sigma_l = 0.0100, \quad \Sigma_{Pl} = 0.0149, \quad \tfrac{1}{2}\Lambda = 10^{-7}.$$

The above parameter setting implies a correlation between the stock price and the liability $\rho_{Pl} \simeq 0.3376$. Also, the last specification $\frac{1}{2}\Lambda = 10^{-7}$ on the temporary price impact indicates that the purchase of 100,000 shares incurs 1% average trading cost.

We are interested in the effects of the liability parameters, including μ_l, Σ_l, ρ_{Pl}, as well as the temporary price impact Λ on the optimal stock holding value x^*. When we vary one parameter, the remaining parameters are fixed so that every single effect can be clearly explored. For each case, we generate 100,000 random sample paths for l_t and calculate the average holding value in the stock x_t^*. Moreover, we introduce the following trading strategy:

$$x_t^{\text{nf}} = \frac{1}{\gamma}\Sigma_P^{-1}Bf_t + \Sigma_P^{-1}\Sigma_{Pl}l_t, \tag{56}$$

which represents the optimal stock holding value in a frictionless market and can be derived directly from the objective (4) under a similar procedure as illustrated in Sections 3 and 4. Fig. 1 presents the sensitivity result and the comparison between x^* and x^{nf}.



(a) Varying μ_l.

(b) Varying Σ_l.

(c) Varying ρ_{Pl}.

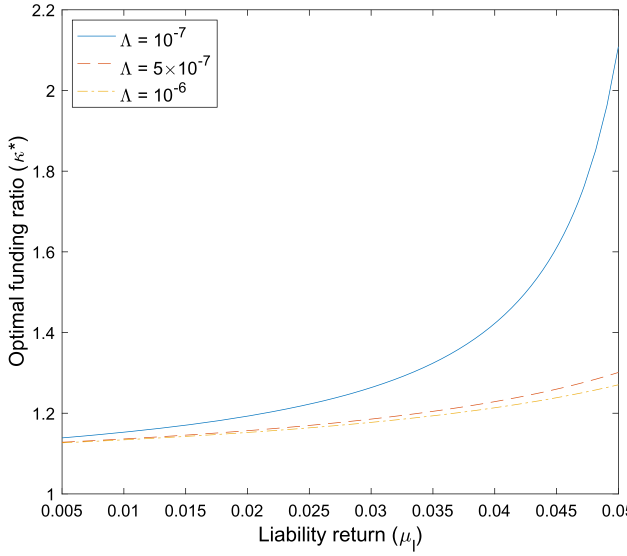
(d) Varying Λ.

Fig. 1. Sensitivity analysis of the stock holding values with varying model parameters. The holding values are calculated using the average of 100, 000 sample paths. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

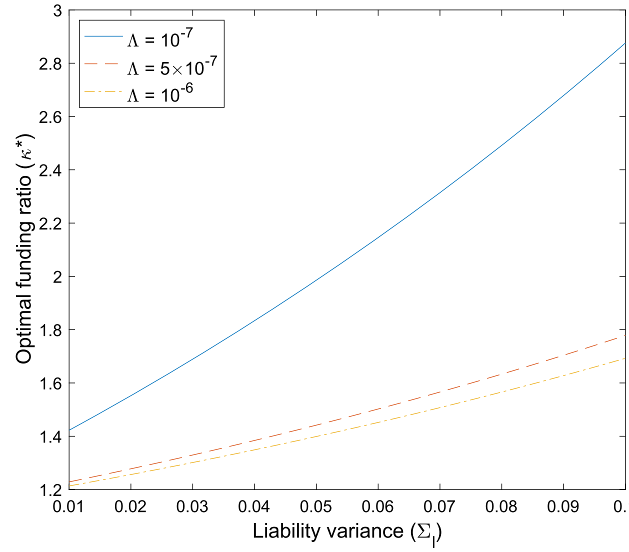
As shown in Fig. 1(a) and 1(b), increasing the growth rate of liability μ_l and the liability variance Σ_l improves the stock holding. The former is quite intuitive because the manager seeks more potential profit from the financial market by taking more risky investment, so as to cover a higher growth rate of liability. Meanwhile, a more volatile liability process also motivates the manager to hold more stocks to hedge the risk, as the stock price and liability are positively correlated. Fig. 1(c) demonstrates that the manager holds more (resp. less) stocks when the stock-liability correlation is positive (resp. negative), which again illustrates that the financial portfolio in ALM should not only be itself profit-seeking, but should also meet the need of hedging liability risk. Fig. 1(d) reveals how the manager's trading behavior changes as the price impact vanishes. Note that three red lines are overlapped since x^{nf} represents the optimal stock holding value in a frictionless market and is independent of Λ. We can see that x^* becomes more concave and moves up toward x^{nf} as Λ decreases. The concavity of x^* is due to the horizon effect (Gennotte and Jung, 1994). The manager is willing to adjust stock holdings at earlier times to enjoy the benefits from long-term investment. A higher price impact leads to a more conservative trading decision, which is as expected because the manager suffers a greater transaction cost with a higher price impact.

6.2. Optimal funding ratio

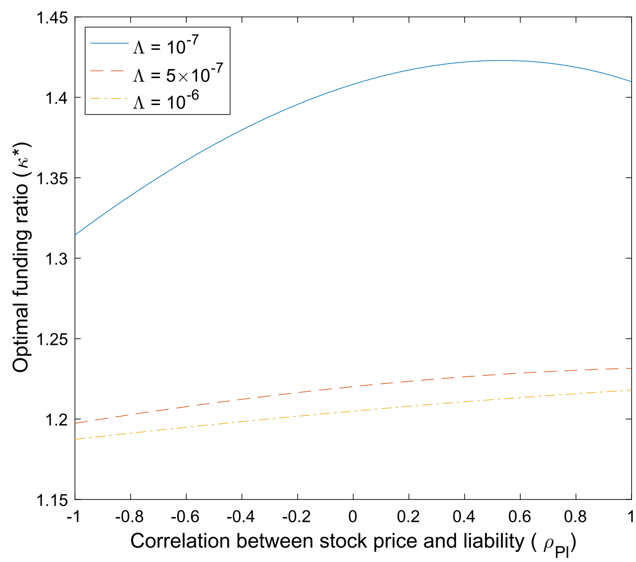
We keep using the same parameter set as in the previous subsection. Suppose the company needs to determine the optimal initial asset value and liability so as to maximize the objective function, and it can raise capital by issuing loans. Following the discussion in Section 5, the initial asset value is set up according to the following guideline:



(a) Varying μ_l.



(b) Varying Σ_l.



(c) Varying ρ_{Pl}.

Fig. 2. Sensitivity analysis of the optimal funding ratio with varying model parameters under different price impact levels.

$$a_0 = \alpha + \beta l_0.$$

In what follows, we set the basic value of asset as $\alpha = 10^6$ and the proportion of raised liability as $\beta = 1$. For the temporary price impact, we consider three different scenarios: 1) a high level, where $\Lambda = 10^{-6}$; 2) a normal level, where $\Lambda = 5 \times 10^{-7}$; and 3) a low level, where $\Lambda = 10^{-7}$. The relationships between the optimal funding ratio $\kappa^* = \frac{\alpha_0^*}{l_0^*}$ and the liability parameters under these three price impact levels are demonstrated in Fig. 2. Our model parameters are chosen such that the cubic inequality (55) in Proposition 5.1 is satisfied, indicating that κ^* is a feasible funding ratio. Fig. 2 also provides a visual confirmation that $\kappa^* > \beta = 1$.

From the definition of the function ratio that $\kappa = \frac{\alpha_0}{l_0} = \frac{\alpha}{l_0} + \beta$, we know that a low funding ratio implies a high liability level. Generally speaking, Fig. 2 predicts a lower optimal funding ratio when the price impact is higher. This is expected because the manager needs to raise more initial capital by issuing loans to support the scheduled trading strategy when faced with a larger transaction cost. Fig. 2(a) and 2(b) indicate that the optimal funding ratio is increasing in the growth rate and the variance of the liability. When μ_l and Σ_l are large, the risk from liability is high and hence the company issues less loans. Fig. 2(c) demonstrates delicate results. When the price impact is relatively high, κ^* is strictly increasing in ρ_{Pl}. As the stock-liability correlation increases, the manager tends to raise less liability to avoid more systemic risk commonly embedded in stock prices and future liabilities. However, when the price impact is comparatively low, as ρ_{Pl} increases, κ^* first moves up but later drops down. The rationale behind is that in this case, the financial market becomes more attractive with less frictions and the risk from the net asset value reduces when the correlation ρ_{Pl} gets larger. Specifically, when ρ_{Pl} is

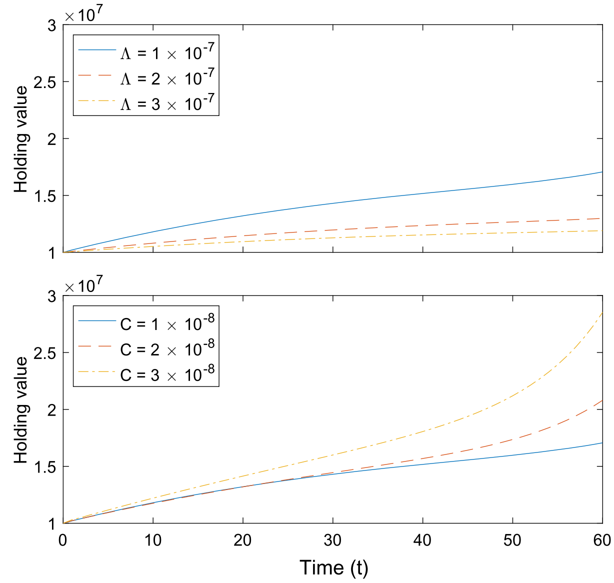


Fig. 3. Sensitivity analysis of the stock holding values with different joint effects of temporary and persistent price impacts. The processes are calculated using the average of 100, 000 sample paths.

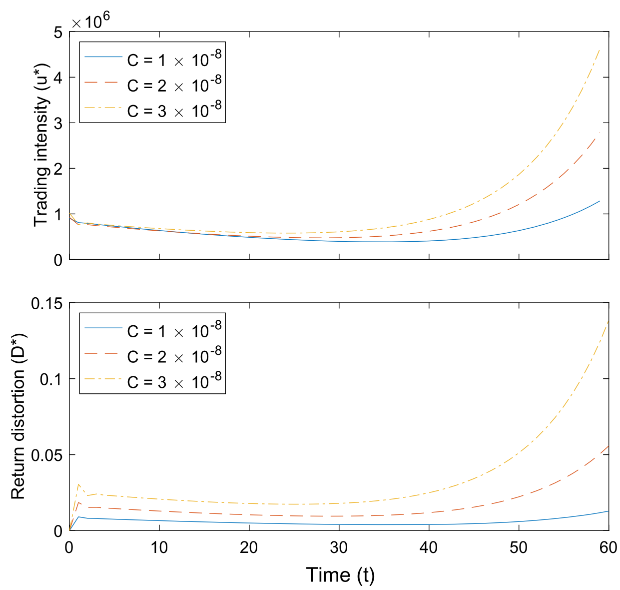


Fig. 4. The optimal trading intensity u^* and the return distortion process D^* with varying persistent price impact levels. The processes are calculated using the average of 100, 000 sample paths.

over certain threshold and becomes very close to 1, it is optimal for the manager to invest more in the market by raising liability, leading to the decrease in the funding ratio.

6.3. Temporary price impact and persistent price impact

To study the ALM problem under a comprehensive market friction setting, we here further consider the persistent price impact. The basic model parameters in Section 6.1 are again used. To compare the temporary effect with the persistent effect, we let $\Lambda = 10^{-7}$ and $C = 10^{-8}$ and vary one of these two parameters with the other fixed. For the return distortion process D_t, we set $D_0 = 0$ and the resiliency as $R = 1$. Fig. 3 shows the holding values under different market friction setting. Note that these processes are average values over 100, 000 sample paths. Different from that under varying temporary effect, as the persistent price impact increases, the stock holding value increases and the improvement becomes more significant as time approaches T. By comparing the magnitudes of Λ and C, we conclude that the persistent effect affects the manager's trading behavior more significantly compared to the temporary price impact, which is consistent with the numerical finding in Berry-Stölzle (2008b).

Fig. 4 presents the optimal trading intensity u^* and the corresponding return distortion process D^* with different persistent price impact levels. It is expected that the distortion D^* increases with C since a higher C implies a stronger ability to affect the prices. The result about u^* again demonstrates that the persistent price impact not only increases the trading intensity at the initial time, but also renders the manager to improve the trading speed as time approaches the terminal time T, leading to a reversed S-shaped curve for x^*.

We point out that such a pattern is indeed induced by the horizon effect. When the persistent price impact is present, purchasing stocks at a high speed will not only significantly increases the stock holding values but also improve the trading cost in the present and future. Since the ALM planning is $[0, T]$ and the company's objective function (4) does not involve what occurs after T, the high execution price after T will have no influence on it. As a result, when t is near the end time T, the manager decides to manipulate the price through the persistent price impact by trading aggressively to increase the stock holdings value.

7. Conclusion

In this paper, we study an ALM problem with return predictability and market frictions from both temporary and persistent price impacts. We adopt the local mean-variance framework in Gârleanu and Pedersen (2013, 2016) in order to avoid the time-inconsistency issue commonly encountered under a mean-variance optimization criterion. Different from Gârleanu and Pedersen (2013, 2016), we assume the asset dynamics to follow a multivariate geometric Brownian motion instead of a multivariate Arithmetic Brownian motion. It turns out that the optimal portfolio choice features a target-chasing manner so that the current position is rebalanced towards a target portfolio, which is dynamically audited by the manager based on the new information on predicting signals and liability.

The persistent price impact considered in this paper complicates the solvability of the associated matrix Riccati equation system. We provide sufficient conditions to address the well-posedness of the corresponding coupled Riccati differential system. To this end, we are able to examine the effects of liability and market frictions on the optimal trading behavior. On the one hand, the need to hedge the liability risk is incorporated into the dynamic target portfolio, which then influences the current trading decision. On the other hand, the temporary price impact leads to a quadratic transaction cost that hinders aggressive trading behavior. In addition, if the persistent price impact is large, it is possible for the ALM manager to manipulate the price through an adoption of an aggressive trading strategy.

Declaration of competing interest

There is no competing interest.

Data availability

No data was used for the research described in the article.

Acknowledgements

We sincerely appreciate the insightful comments from the Editor and two anonymous reviewers that have improved our paper significantly. We are also grateful for the comments from the seminar participants at The 25th International Congress on Insurance: Mathematics and Economics. Tingjin Yan acknowledges the financial support from Shanghai Pujiang Program (22PJC038), and the National Natural Science Foundation of China (71931004). Guiyuan Ma acknowledges the financial support from the National Natural Science Foundation of China (72101199). Chi Chung Siu acknowledges the financial support under the grant "Generalized Sethi Advertising Model and Extensions" (Project No. UGC/FDS14/P02/20) from the Research Grants Council of Hong Kong.

Appendix A. Proofs in this paper

A.1. Proof of Theorem 3.2

Proof. (i) According to Lemma 3.1, $\mathbf{u}^*$ and V, respectively given by (18) and (7), are well defined and solve the HJB equation (6). Applying standard result (see, for instance, Yong and Zhou (1999)), it suffices to prove that the candidate control $\mathbf{u}^*$ is admissible. We write the portfolio dynamic generated by $\mathbf{u}^*$ as $\mathbf{x}^*$. Plugging (18) into (2), we derive

$$\mathbf{x}_t^* = \mathcal{E}_2(t)^{-1}\mathbf{x}_0 + \int_0^t \mathcal{E}_2(s)\boldsymbol{\Lambda}^{-1}\left[\mathbf{A}_{xf}(s)\mathbf{f}_s + \mathbf{A}_{xl}(s)l_s\right]ds, \tag{A.1}$$

where $\mathcal{E}_2(t)$ is the bounded solution to the linear ODE (22). Since $\mathbf{f}_t$ and l_t are solutions to linear SDEs, it is clear from the standard theory (Yong and Zhou, 1999) that

$$\mathbb{E}\left[\sup_{t\in[0,T]}|\mathbf{f}_t|^2\right] < \infty, \qquad \mathbb{E}\left[\sup_{t\in[0,T]}|l_t|^2\right] < \infty. \tag{A.2}$$

Therefore, there exists a constant $c_1 > 0$ such that

$$\mathbb{E}\left[\sup_{t\in[0,T]}|\mathbf{x}_t^*|^2\right] \le c_1\left(1 + \mathbb{E}\left[\sup_{t\in[0,T]}|\mathbf{f}_t|^2\right] + \mathbb{E}\left[\sup_{t\in[0,T]}|l_t|^2\right]\right) < \infty.$$

Then we derive

$$\mathbb{E}\left[\int_0^T |\mathbf{u}_t^*|^2\,dt\right] = \mathbb{E}\left[\int_0^T \left|\boldsymbol{\Lambda}^{-1}\left[\mathbf{A}_{xf}(t)\mathbf{f}_t + \mathbf{A}_{xl}(t)l_t - \mathbf{A}_{xx}(t)\mathbf{x}_t\right]\right|^2 dt\right]$$

$$\leq c_2 \left(1 + \mathbb{E}\left[\sup_{t\in[0,T]} |\mathbf{f}_t|^2\right] + \mathbb{E}\left[\sup_{t\in[0,T]} |l_t|^2\right] + \mathbb{E}\left[\sup_{t\in[0,T]} \left|\mathbf{x}_t^*\right|^2\right]\right)$$
$$< \infty,$$

where $c_2 > 0$ is a proper constant. Therefore, we verify u^* given by (18) is the optimal trading intensity.

(ii) Based on the ODE for $\mathbf{A}_{xf}$ and $\mathbf{A}_{xl}$, i.e. (11) and (12), we apply the Feynman-Kac formula and derive

$$\mathbf{A}_{xf}(t)\mathbf{f}_t + \mathbf{A}_{xl}(t)l_t = \gamma \mathbb{E}_t\left[\mathcal{E}_1(t)^{-1}\int_t^T e^{-\rho(s-t)}\mathcal{E}_1(s)\mathbf{\Sigma}_P\mathbb{E}_t\left[\mathbf{x}_s^{\mathrm{mv}} + \mathbf{\Sigma}_P^{-1}\mathbf{\Sigma}_{Pl}l_s\right]ds\right]. \tag{A.3}$$

According to Lemma 3.1, $\mathbf{A}_{xx}(t)$ is positive definite for $t \in [0, T)$ and thus non-singular. Therefore, $\mathbf{M}_t^{\mathrm{aim}}$ is well defined. Plugging the above equation into (18) yields the desired result.

(iii) Under the condition $\mathbf{\Lambda} = \lambda\mathbf{\Sigma}_P$, we assume the ansatz $\mathbf{A}_{xx}(t) = a_{xx}(t)\mathbf{\Sigma}_P$. From (8), we derive

$$\dot{a}_{xx}(t) = \frac{1}{\lambda}a_{xx}^2(t) + \rho a_{xx}(t) - \gamma, \quad a_{xx}(T) = 0.$$

The above scalar Riccati equation admits the following explicit solution

$$a_{xx}(t) = -\frac{\gamma\lambda\left(1 - e^{(T-t)(y_+ - y_-)}\right)}{y_+ - y_- e^{(T-t)(y_+ - y_-)}} \geq 0, \tag{A.4}$$

where $y_\pm = \frac{\lambda}{2}\left(-\rho \pm \sqrt{\rho^2 + 4\frac{\gamma}{\lambda}}\right)$. Then $\mathcal{E}_1(t) = e^{\int_t^T \frac{a_{xx}(s)}{\lambda}ds}$. The representations of $\mathbf{M}^{\mathrm{rate}}$ and $\mathbf{M}^{\mathrm{aim}}$ follow by plugging the expressions of $\mathbf{A}_{xx}$ and $\mathcal{E}_1$ into the results in (20). □

A.2. Proof of Lemma 4.2

Proof. We first supplement some notations. Let $\mathbf{X} \in \mathbb{R}^{n\times n}$ be partitioned as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{pmatrix}. \tag{A.5}$$

If $\mathbf{X}_{11}$ is a square matrix and invertible, we denote the matrix $\mathbf{X}_{22} - \mathbf{X}_{21}\mathbf{X}_{11}^{-1}\mathbf{X}_{12}$ by $\mathbf{X}/\mathbf{X}_{11}$, which is referred to as the *Schur complement* of $\mathbf{X}_{11}$ in $\mathbf{X}$ (see Zhang (2006) for detailed illustrations). Similarly, if $\mathbf{X}_{22}^{-1}$ exists, we denote the matrix $\mathbf{X}_{11} - \mathbf{X}_{12}\mathbf{X}_{22}^{-1}\mathbf{X}_{21}$ by $\mathbf{X}/\mathbf{X}_{11}$ and call it the Schur complement of $\mathbf{X}_{22}$ in $\mathbf{X}$.

Then we introduce the following comparison result for matrix RDEs (see Theorem 4.1.19 in Abou-Kandil et al. (2003)).

Lemma A.1. *For $i = 1, 2$, let $\mathbf{E}_i(t) \in \mathbb{R}^{n\times n}$ be the solution of the following matrix RDE*

$$\dot{\mathbf{E}}_i(t) = \mathbf{E}_i(t)\mathbf{S}_i\mathbf{E}_i(t) + \mathbf{E}_i(t)\mathbf{H}_i + \mathbf{H}_i^\top\mathbf{E}_i(t) + \mathbf{Q}_i, \tag{A.6}$$

where $\mathbf{S}_i, \mathbf{Q}_i$ are symmetric. If $\begin{pmatrix} \mathbf{Q}_1 & \mathbf{H}_1 \\ \mathbf{H}_1^\top & \mathbf{S}_1 \end{pmatrix} \leq \begin{pmatrix} \mathbf{Q}_2 & \mathbf{H}_2 \\ \mathbf{H}_2^\top & \mathbf{S}_2 \end{pmatrix}$, and $\mathbf{E}_2(T) \leq \mathbf{E}_1(T)$, then $\mathbf{E}_2(t) \leq \mathbf{E}_1(t)$ for $t \in [0, T]$.

We proceed to the proof of Lemma 4.2. Note that if (28) admits a unique bounded solution, the existence and uniqueness result for the remaining equations follows immediately because they can be considered as linear ODEs and solved in sequence. It suffices to prove that (28) has a unique bounded solution on $[0, T]$. The main idea is to introduce two functions $\bar{\mathbf{E}}$, $\underline{\mathbf{E}}^\epsilon$ that are bounded solutions of two matrix RDEs respectively and show $\underline{\mathbf{E}}^\epsilon(t) \leq \mathbf{E}_{yy}(t) \leq \bar{\mathbf{E}}(t)$ for $t \in [0, T]$ using Lemma A.1. Recall that the equation (28) has a standard expression (38). Define $\mathbf{Q}_0 := \mathbf{M}_1\mathbf{\Lambda}^{-1}\mathbf{M}_1^\top - \mathbf{Q}$, $\mathbf{H}_0 := \frac{1}{2}\rho\mathbf{I} + \mathbf{N}_1 - \mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{M}_1^\top$, $\mathbf{S}_0 := \mathbf{N}_2\mathbf{\Lambda}^{-1}\mathbf{N}_2^\top$. Then (38) becomes

$$\dot{\mathbf{E}}_{yy}(t) = \mathbf{E}_{yy}(t)\mathbf{S}_0\mathbf{E}_{yy}(t) + \mathbf{E}_{yy}(t)\mathbf{H}_0 + \mathbf{H}_0^\top\mathbf{E}_{yy}(t) + \mathbf{Q}_0.$$

Step 1. Let $\bar{\mathbf{E}}(\cdot)$ solve the following linear equation:

$$\dot{\bar{\mathbf{E}}}(t) = \bar{\mathbf{E}}(t)\mathbf{H}_0 + \mathbf{H}_0^\top\bar{\mathbf{E}}(t) + \mathbf{Q}_0, \tag{A.7}$$

with $\bar{\mathbf{E}}(T) = \mathbf{0}_{n\times n}$. It is clear that this equation admits a bounded solution on $[0, T]$. Note that

$$\begin{pmatrix} \mathbf{Q}_0 & \mathbf{H}_0 \\ \mathbf{H}_0^\top & \mathbf{0}_{n\times n} \end{pmatrix} \leq \begin{pmatrix} \mathbf{Q}_0 & \mathbf{H}_0 \\ \mathbf{H}_0^\top & \mathbf{S}_0 \end{pmatrix}. \tag{A.8}$$

Therefore, according to Lemma A.1, we have $\bar{\mathbf{E}}(t) \geq \mathbf{E}_{yy}(t)$ for $t \in [0, T]$.

Step 2. Let $\epsilon > 0$ be a constant. For each ϵ, consider the matrix RDE:

$$\dot{\underline{\mathbf{E}}}^\epsilon(t) = \underline{\mathbf{E}}^\epsilon(t)\underline{\mathbf{S}}(\epsilon)\underline{\mathbf{E}}^\epsilon(t) + \underline{\mathbf{E}}^\epsilon(t)\underline{\mathbf{H}} + \underline{\mathbf{H}}^\top\underline{\mathbf{E}}^\epsilon(t) + \underline{\mathbf{Q}}(\epsilon), \qquad \underline{\mathbf{E}}^\epsilon(T) = \mathbf{0}_{n\times n}, \tag{A.9}$$

where

$$\underline{\mathbf{S}}(\epsilon)=\mathbf{S}_0+\begin{pmatrix}\boldsymbol{\Gamma}_1(\epsilon) & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\end{pmatrix},\quad \underline{\mathbf{H}}=\begin{pmatrix}\frac{1}{2}\rho\mathbf{I}_n-\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}\end{pmatrix},\quad \underline{\mathbf{Q}}(\epsilon)=\begin{pmatrix}\mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \boldsymbol{\Gamma}_2(\epsilon)\end{pmatrix},$$

and

$$\begin{aligned}\boldsymbol{\Gamma}_1(\epsilon)&=\frac{1}{\epsilon}\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\left(\mathbf{R}+r_f\mathbf{I}\right)^{-1}\left(\gamma\boldsymbol{\Sigma}_P-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right)\left(\mathbf{R}^\top+r_f\mathbf{I}\right)^{-1}\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top,\\ \boldsymbol{\Gamma}_2(\epsilon)&=(1+\epsilon)\mathbf{R}^\top\left(\gamma\boldsymbol{\Sigma}_P-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}.\end{aligned}\tag{A.10}$$

We will show in the following that $\underline{\mathbf{E}}^\epsilon(t)\le\mathbf{E}_{yy}(t)$ for each ϵ. Since $\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top<\gamma\boldsymbol{\Sigma}_P$, we have $\boldsymbol{\Gamma}_i\ge 0$ for $i=1,2$ and $\mathbf{S}_0\le\underline{\mathbf{S}}$. Moreover, since $\underline{\mathbf{Q}}(\epsilon)-\mathbf{Q}_0=\begin{pmatrix}\gamma\boldsymbol{\Sigma}_P-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top & \mathbf{R}\\ \mathbf{R}^\top & \boldsymbol{\Gamma}_2(\epsilon)\end{pmatrix}$. Then

$$\left(\underline{\mathbf{Q}}(\epsilon)-\mathbf{Q}_0\right)/\left(\gamma\boldsymbol{\Sigma}_P-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right)=\epsilon\mathbf{R}^\top\left(\gamma\boldsymbol{\Sigma}_P-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}>\mathbf{0}_{n\times n}.\tag{A.11}$$

According to Theorem 1.12 in Zhang (2006), we obtain $\underline{\mathbf{Q}}(\epsilon)-\mathbf{Q}_0>\mathbf{0}_{n\times n}$. Let $\mathbf{G}\in\mathbb{R}^n$ be the last n rows of the last n columns of $\left(\underline{\mathbf{Q}}(\epsilon)-\mathbf{Q}_0\right)^{-1}$. Theorem 1.2 in Zhang (2006) yields

$$\mathbf{G}=\frac{1}{\epsilon}\mathbf{R}^{-1}\left(\gamma\boldsymbol{\Sigma}_P-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right)\mathbf{R}^{\top-1}>\mathbf{0}_{n\times n}.$$

Then we consider the Schur complement of $\underline{\mathbf{Q}}(\epsilon)-\mathbf{Q}_0$ in the matrix $\begin{pmatrix}\underline{\mathbf{Q}}(\epsilon) & \underline{\mathbf{H}}\\ \underline{\mathbf{H}}^\top & \underline{\mathbf{S}}\end{pmatrix}-\begin{pmatrix}\mathbf{Q}_0 & \mathbf{H}_0\\ \mathbf{H}_0^\top & \mathbf{S}_0\end{pmatrix}$:

$$\begin{aligned}&\underline{\mathbf{S}}-\mathbf{S}_0-\begin{pmatrix}\mathbf{0}_{n\times n} & \mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\end{pmatrix}\left(\underline{\mathbf{Q}}(\epsilon)-\mathbf{Q}_0\right)^{-1}\begin{pmatrix}\mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\\ \mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top & \mathbf{0}_{n\times n}\end{pmatrix}\\ &=\begin{pmatrix}\boldsymbol{\Gamma}_1(\epsilon)-\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\mathbf{G}\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top & \mathbf{0}_{n\times n}\\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n}\end{pmatrix}=\mathbf{0}_{2n\times 2n}.\end{aligned}\tag{A.12}$$

Applying Theorem 1.12 in Zhang (2006) again, we have $\begin{pmatrix}\underline{\mathbf{Q}}(\epsilon) & \underline{\mathbf{H}}\\ \underline{\mathbf{H}}^\top & \underline{\mathbf{S}}\end{pmatrix}\ge\begin{pmatrix}\mathbf{Q}_0 & \mathbf{H}_0\\ \mathbf{H}_0^\top & \mathbf{S}_0\end{pmatrix}$. If $\underline{\mathbf{E}}^\epsilon(t)$ exist, Lemma A.1 yields $\mathbf{E}_{yy}(t)\ge\underline{\mathbf{E}}^\epsilon(t)$ for $t\in[0,T]$. In the following, we show that a bounded solution $\underline{\mathbf{E}}^\epsilon(t)$ exists under the given condition.

We write $\underline{\mathbf{E}}^\epsilon=\begin{pmatrix}\underline{\mathbf{E}}^\epsilon_{xx} & \underline{\mathbf{E}}^\epsilon_{xD}\\ \underline{\mathbf{E}}^{\epsilon\top}_{xD} & \underline{\mathbf{E}}^\epsilon_{DD}\end{pmatrix}$. Then it is clear from (A.9) that $\underline{\mathbf{E}}^\epsilon_{xx}=\underline{\mathbf{E}}^\epsilon_{xD}\equiv\mathbf{0}_{n\times n}$ and $\underline{\mathbf{E}}^\epsilon_{DD}$ satisfies

$$\dot{\underline{\mathbf{E}}}^\epsilon_{DD}(t)=\underline{\mathbf{E}}^\epsilon_{DD}(t)\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\underline{\mathbf{E}}^\epsilon_{DD}(t)+\underline{\mathbf{E}}^\epsilon_{DD}(t)^\top\left(\frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}\right)+\left(\frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}^\top\right)\underline{\mathbf{E}}^\epsilon_{DD}(t)+\boldsymbol{\Gamma}_2(\epsilon).\tag{A.13}$$

Next, we show the existence and uniqueness result for $\underline{\mathbf{E}}^\epsilon_{DD}(t)$. The proof is in the same spirit of that in Bensoussan et al. (2022); Chu et al. (2022). Let $\mathbb{C}_T$ be the Banach space of the continuous matrix functions $\mathbf{E}(t):[0,T]\to\mathbb{R}^{n\times n}$ with the maximum norm $\|\mathbf{E}\|_\infty:=\max_{t\in[0,T]}\|\mathbf{E}(t)\|$. Let $\mathbb{B}_T$ be a ball with radius H in $\mathbb{C}_T$. For each $\epsilon>0$, we define an operator $\mathcal{J}^\epsilon$ on $\mathbb{B}_T$ as follows

$$(\mathcal{J}^\epsilon\mathbf{E})(t):=\int_t^T\left[\mathbf{E}(s)\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\mathbf{E}(s)+\mathbf{E}(s)^\top\left(\frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}\right)+\left(\frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}^\top\right)\mathbf{E}(s)+\boldsymbol{\Gamma}_2(\epsilon)\right]ds,$$

for $t\in[0,T]$. If

$$T(\left\|\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right\|H^2+2\left\|\frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}\right\|H+\|\boldsymbol{\Gamma}_2(\epsilon)\|)<H,\tag{A.14}$$

we have for any function $\hat{\mathbf{E}}\in\mathbb{B}_T$, $\left\|\mathcal{J}^\epsilon(\hat{\mathbf{E}})\right\|_\infty<H$. Therefore, $\mathcal{J}^\epsilon(\mathbb{B}_T)\subset\mathbb{B}_T$ and $\mathcal{J}^\epsilon$ is a self-map. Note that (A.14) is satisfied if

$$T<\frac{1}{2\left(\left\|\frac{1}{2}\rho\mathbf{I}_n+\mathbf{R}\right\|+\sqrt{\left\|\mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}^\top\right\|\|\boldsymbol{\Gamma}_2(\epsilon)\|}\right)}.\tag{A.15}$$

Next we prove that $\mathcal{J}^\epsilon(\mathbb{B}_T)$ is relatively compact. Consider any sequence in $\mathcal{J}^\epsilon(\mathbb{B}_T)$ denoted by $\left\{\mathcal{J}^\epsilon(\mathbf{E}_n)\right\}_{n=1,2,\ldots}$. The sequence is uniformly bounded, and from the definition of $\mathcal{J}^\epsilon$, the derivatives of the functions in the sequence are also uniformly bounded. The latter implies $\left\{\mathcal{J}^\epsilon(\mathbf{E}_n)\right\}_{n=1,2,\ldots}$ is equicontinuous. By using Arzela-Ascoli theorem, $\left\{\mathcal{J}^\epsilon(\mathbf{E}_n)\right\}_{n=1,2,\ldots}$ has a uniformly convergent subsequence. And hence $\mathcal{J}^\epsilon(\mathbb{B}_T)$ is relatively compact.

According to the Schauder's fixed point theorem, the self-map $\mathcal{J}^\epsilon$ admits a fixed point, which is the solution to (A.13). Moreover, the uniqueness of the $\underline{\mathbf{E}}^\epsilon$ follows from the fact that the right-hand side of the equation (A.13) is locally Lipschitz continuous. Note that $\lim_{\epsilon\downarrow 0}\boldsymbol{\Gamma}_2(\epsilon)=\psi_2$. By letting $\epsilon\downarrow 0$, we obtain that if

$$T<\frac{1}{2\left(\psi_1+\sqrt{\psi_0\psi_2}\right)},$$

there exists an $\epsilon_0 > 0$ and a bounded function $\underline{\mathbf{E}}^{\epsilon_0}$ on $[0, T]$.

Step 3. Since $\underline{\mathbf{E}}^{\epsilon_0}(t) \le \mathbf{E}_{yy}(t) \le \bar{\mathbf{E}}(t)$, the existence result for (28) is proved. The uniqueness of the solution to (28) follows easily by applying Lemma A.1. $\mathbf{E}_{xx}$ is positive semidefinite since $\mathbf{E}_{xx} \ge \underline{\mathbf{E}}^{\epsilon_0}_{xx} = \mathbf{0}_{n\times n}$. □

A.3. Proof of Lemma 4.3

Proof. It suffices to show the result for equation (28). The following proof is in the spirit of that of Theorem 5 in Chu et al. (2022). Under the assumption of Lemma 4.3, one can verify that $\mathbf{E}_{xx}(t), \mathbf{E}_{xD}(t), \mathbf{E}_{DD}(t)$ are diagonal. We denote their diagonal elements by $E^i_{xx}(T-t), E^i_{DD}(T-t)$ for $i = 1, 2, \ldots, n$, respectively. Note that the timeline is reversed. Then, the existence and unique result of the matrix RDE (28) is equivalent to that of the following coupled RDEs:

$$\dot{E}^i_{xx}(t) = -\lambda_i^{-1}(E^i_{xx}(t) + c_i E^i_{xD}(t) - c_i)^2 - \rho E^i_{xx}(t) + \gamma\sigma^2_{P,i}, \tag{A.16}$$

$$\dot{E}^i_{xD}(t) = -\lambda_i^{-1}(E^i_{xx}(t) + c_i E^i_{xD}(t) - c_i)(E^i_{xD}(t) + c_i E^i_{DD}(t)) - (\rho + r_i)E_{xD}(t) + r_i, \tag{A.17}$$

$$\dot{E}^i_{DD}(t) = -\lambda_i^{-1}(E^i_{xD}(t) + c_i E^i_{DD}(t))^2 - E_{DD}(t)(\rho + 2r_i). \tag{A.18}$$

For simplicity, we suppress the superscript i in the following statement. Let $\mathbb{C} := \mathbb{C}([0, T]; \mathbb{R})$ denote the family of continuous functions $f : [0, T] \to \mathbb{R}$ with the norm $\|f\|_\infty = \max_{t\in[0,T]} |f(t)|$. Define the cube $\mathbb{H} := \{f \in \mathbb{C};\ f(t) \in [0, 1] \text{ for } t \in [0, T]\}$. Let $\hat{g}_{xD}$ be an arbitrary function in $\mathbb{H}$. We introduce a map $\mathcal{T}$: $\mathbb{H} \to \mathbb{C}$, where $\mathcal{T}(\hat{g}_{xD}) := g_{xD}$ is a part of the solutions of the following RDE system.

$$\dot{g}_{DD}(t) = -\lambda^{-1}c^2 g^2_{DD}(t) - (\rho + 2r + 2\lambda^{-1}c\hat{g}_{xD}(t))g_{DD}(t) - \lambda^{-1}\hat{g}^2_{xD}(t), \tag{A.19}$$

$$\dot{g}_{xx}(t) = -\lambda^{-1}g^2_{xx}(t) - (2\lambda^{-1}c(\hat{g}_{xD}(t) - 1) + \rho)g_{xx}(t) + \gamma\sigma^2_P - \lambda^{-1}c^2(\hat{g}_{xD}(t) - 1)^2, \tag{A.20}$$

$$\dot{g}_{xD}(t) = -\lambda^{-1}c g^2_{xD}(t) - \left[\lambda^{-1}(g_{xx}(t) - c) + \lambda^{-1}c^2 g_{DD}(t) + \rho + r\right] g_{xD}(t) - \lambda^{-1}(g_{xx}(t) - c)c g_{DD}(t) + r. \tag{A.21}$$

We introduce a lemma before showing that the map $\mathcal{T}$ is well defined.

Lemma A.2. *For a constant $a < 0$ and two functions $g_1, g_2 \in \mathbb{C}([0, \infty), \mathbb{R})$. Suppose there exist two constants $m_1, m_2 \in \mathbb{R}$ such that $g_1(t) < m_1 < g_2(t) < m_2$, for $t > 0$. Consider a non-autonomous system:*

$$\dot{y}(t) = a(y(t) - g_1(t))(y(t) - g_2(t)), \qquad y(0) = \xi \in \mathbb{R}. \tag{A.22}$$

Denote the solution to the above equation as $y(t; 0, \xi)$. Then every system trajectory starting from an initial point $\xi \in [m_1, m_2]$ stays in an interval $[m_1, m_2]$, i.e. for any $t \ge 0$,

$$y(t; 0, \xi) \in [m_1, m_2].$$

Proof. To prove this lemma, we first introduce a Lyapunov function $v(y) = (y - \frac{m_1+m_2}{2})^2$. Then it is clear that $v(y) \le \left(\frac{m_2-m_1}{2}\right)^2$ if and only if $y \in [m_1, m_2]$. And $v(y) = \left(\frac{m_2-m_1}{2}\right)^2$ if and only if $y \in \{m_1, m_2\}$. The evolution of the Lyapunov function along the curve $y(t; 0, \xi)$ in the system (A.22) is

$$\frac{dv(y(t; 0, \xi))}{dt} = 2a\left(y - \frac{m_1 + m_2}{2}\right)(y - g_1(t))(y - g_2(t))\Big|_{y=y(t;0,\xi)}. \tag{A.23}$$

By using the assumption, we have $\frac{dv(y(t;0,\xi))}{dt}\Big|_{y(t;0,\xi)=m_1} \le 0$ and $\frac{dv(y(t;0,\xi))}{dt}\Big|_{y(t;0,\xi)=m_2} \le 0$, which implies that $v(y(t; 0, \xi)) \le \left(\frac{m_2-m_1}{2}\right)^2$. Therefore, $y(t; 0, \xi) \in [m_1, m_2]$ for any $t \ge 0$. □

Then we proceed to show that $\mathcal{T}$ is a self-map, i.e. $\mathcal{T}(\mathbb{H}) \subset \mathbb{H}$:

First, we write the right-hand-side of (A.19) as $h_{DD}(t, g_{DD}(t))$. Then it is clear that, for each fixed t, $h_{DD}(t, x) = 0$ admits two roots $x_{DD,-}(t) < x_{DD,+}(t)$. Note that $h_{DD}(0) < 0$ and $h_{DD}(-c^{-1}\hat{g}_{xD}(t)) = c^{-1}\hat{g}_{xD}(t)(\rho + 2r) > 0$. Since $-\lambda^{-1}c^2 < 0$ and $-(\rho + 2r + 2\lambda^{-1}c\hat{g}_{xD}(t)) < 0$, we have $x_{DD,+}(t) \in (-c^{-1}\hat{g}_{xD}(t), 0)$. By applying Lemma A.2, we derive $g_{DD}(t) \in [-c^{-1}\hat{g}_{xD}(t), 0]$.

Second, we similarly define the right-hand-side of (A.20) as $h_{xx}(t, g_{xx}(t))$ and $x_{xx,-}(t) < x_{xx,+}(t)$ as two roots of $h_{xx}(t, x) = 0$ for each t. Recall that we assume $\gamma\sigma^2_P > \lambda^{-1}c^2$. Therefore,

$$h_{xx}(t, 0) = -\lambda^{-1}c^2(\hat{g}_{xD}(t) - 1)^2 + \gamma\sigma^2_P \ge -\lambda^{-1}c^2 + \gamma\sigma^2_P > 0.$$

Since $x_{xx,+}(t)$ is a continuous function for $t \in [0, T]$, there exists a constant M such that $x_{xx,+}(t) < M$. We conclude $x_{xx,+}(t) \in (0, M)$. Lemma A.2 yields $g_{xx}(t) \in [0, M]$.

Third, we write the right-hand-side of (A.21) as $h_{xD}(t, g_{xD}(t))$ and $x_{xD,-}(t) < x_{xD,+}(t)$ as two roots of $h_{xD}(t, x) = 0$ for each t. Note that

$$h_{xD}(t, 0) = -\lambda^{-1}c(g_{xx}(t) - c)g_{DD}(t) + r \ge -\lambda^{-1}c^2 + r > 0,$$

and $h_{xD}(t, 1) = -\lambda^{-1}g_{xx}(t)(1 + c g_{DD}(t)) - \rho < -\rho < 0$. Therefore $x_{xD,+}(t) \in [0, 1]$. Then, utilizing Lemma A.2, we obtain $g_{xD} \in \mathbb{H}$. Thus, $\mathcal{T}$ is a self-map.

By following the arguments in *Step 2* of the proof of Lemma 4.2, it can be proved that $\mathcal{T}(\mathbb{H})$ is a relative compact subset of $\mathbb{H}$ and $\mathcal{T}$ admits a unique fixed point, denoted by $g_{xD,\infty}$. With $\hat{g}_{xD}$ replaced by $g_{xD,\infty}$, we define the solutions to the equations (A.19), (A.20) as $g_{DD,\infty}, g_{xx,\infty}$. It is clear that the solutions of (A.16)-(A.18) coincides with $g_{xx,\infty}, g_{xD,\infty}, g_{DD,\infty}$. This proves the existence and uniqueness result for (28). □

A.4. Proof of Theorem 4.4

Proof. (i) Similar to the proof of Theorem 3.2 (i), it suffices to show $\mathbf{u}^*$ given by (39) is admissible. We write $\mathbf{y}^*(t) := \left(\mathbf{x}^*(t)^\top, \mathbf{D}^*(t)^\top\right)^\top$ as the state process generated by $\mathbf{u}^*$ and $\mathbf{y}_0 := \left(\mathbf{x}_0^\top, \mathbf{D}_0^\top\right)^\top$. Then $\mathbf{u}^*$ has the expression

$$\mathbf{u}^*(t) = \boldsymbol{\Lambda}^{-1}\left[\left(\mathbf{M}_1 - \mathbf{E}_{yy}(t)\mathbf{N}_2\right)^\top \mathbf{y}^*(t) + \mathbf{N}_2^\top \mathbf{E}_{yf}(t)\mathbf{f}(t) + \mathbf{N}_2^\top \mathbf{E}_{yl}(t) l(t)\right].$$

From $d\mathbf{y}^*(t) = (-\mathbf{N}_1\mathbf{y}^*(t) + \mathbf{N}_2\mathbf{u}^*(t))\,dt$, we derive

$$\mathbf{y}^*(t) = \hat{\mathcal{E}}(t)^{-1}\mathbf{y}_0 + \int_0^t \hat{\mathcal{E}}(t)^{-1}\hat{\mathcal{E}}(s)\left(\mathbf{N}_2\boldsymbol{\Lambda}^{-1}\mathbf{N}_2^\top \mathbf{E}_{yf}(s)\mathbf{f}_s + \mathbf{N}_2\boldsymbol{\Lambda}^{-1}\mathbf{N}_2^\top \mathbf{E}_{yl}(s) l_s\right) ds,$$

where $\hat{\mathcal{E}}$ solves

$$\dot{\hat{\mathcal{E}}}(t) = \left[\mathbf{N}_2\boldsymbol{\Lambda}^{-1}\left(\mathbf{N}_2^\top \mathbf{E}_{yy}(t) - \mathbf{M}_1^\top\right) + \mathbf{N}_1\right]\hat{\mathcal{E}}(t), \quad \hat{\mathcal{E}}(0) = \mathbf{I}_{2n}.$$

Following the proof of Theorem 3.2 (i), one can show $\mathbb{E}\left[\sup_{t\in[0,T]}|\mathbf{y}^*(t)|^2\right] < \infty$ and $\mathbb{E}\left[\sup_{t\in[0,T]}|\mathbf{u}^*(t)|^2\right] < \infty$.

(ii) Using the ODE for $\mathbf{E}_{yf}$ and $\mathbf{E}_{yl}$, i.e. (31) and (32), we apply the Feynman-Kac formula and derive

$$\mathbf{E}_{yf}(t)\mathbf{f}_t + \mathbf{E}_{yl}(t) l_t = \mathbb{E}_t\left[\int_t^T e^{-\rho(s-t)}\tilde{\mathcal{E}}_1(s)^{-1}\tilde{\mathcal{E}}_1(s)\left[\mathbf{M}_2\mathbf{f}_s + \mathbf{M}_3 l_s\right] ds\right].$$

Plugging the above expression into (39) yields the desired result. □

A.5. Proof of Theorem 4.6

Proof. The first-order condition in the HJB equation (5) is $\mathbf{u}^* = \boldsymbol{\Lambda}^{-1}\left(\frac{\partial V}{\partial \mathbf{x}} + \mathbf{C}^\top \frac{\partial V}{\partial \mathbf{D}} + \mathbf{C}^\top \mathbf{x}\right)$. Plugging $\mathbf{u}^*$ to (5) and replacing V with V^ϵ lead to

$$-\rho V^\epsilon + \frac{\partial V^\epsilon}{\partial t} + \mathcal{L}^{\mathbf{x}} V^\epsilon - \frac{1}{\epsilon}\mathbf{x}^\top \mathbf{R}_1\mathbf{D} - \frac{1}{\epsilon}\frac{\partial V^\epsilon}{\partial \mathbf{D}}^\top \mathbf{R}_1\mathbf{D} + \frac{1}{2\epsilon}\left(\frac{\partial V^\epsilon}{\partial \mathbf{x}} + \epsilon\mathbf{C}_1^\top\frac{\partial V^\epsilon}{\partial \mathbf{D}} + \epsilon\mathbf{C}_1^\top\mathbf{x}\right)^\top \boldsymbol{\Lambda}_1^{-1}\left(\frac{\partial V^\epsilon}{\partial \mathbf{x}} + \epsilon\mathbf{C}_1^\top\frac{\partial V^\epsilon}{\partial \mathbf{D}} + \epsilon\mathbf{C}_1^\top\mathbf{x}\right) = 0, \tag{A.24}$$

where the operator $\mathcal{L}^{\mathbf{x}}$ is given by (45).

Following Ekren and Muhle-Karbe (2019), we consider the asymptotic expansion of the following function

$$\tilde{V}^\epsilon(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l) = V^\epsilon(t, \mathbf{x}, \epsilon\mathbf{D}, \mathbf{f}, l) + \epsilon\left(\mathbf{x}^\top\mathbf{D} - \frac{1}{2}\mathbf{D}^\top\mathbf{C}_1^{-1}\mathbf{D}\right), \tag{A.25}$$

where a small distortion $\epsilon\mathbf{D}$ is considered in order to characterize the limiting trading behavior under large liquidity. Then $\tilde{V}^\epsilon$ satisfies

$$-\rho\tilde{V}^\epsilon + \frac{\partial\tilde{V}^\epsilon}{\partial t} + \mathcal{L}^{\mathbf{x}}\tilde{V}^\epsilon + \mathcal{H}\tilde{V}^\epsilon + \rho\epsilon\left(\mathbf{x}^\top\mathbf{D} - \frac{1}{2}\mathbf{D}^\top\mathbf{C}_1^{-1}\mathbf{D}\right) - \mathbf{D}^\top\mathbf{C}_1^{-1}\mathbf{R}_1\mathbf{D} = 0, \tag{A.26}$$

where

$$\mathcal{H}\tilde{V}^\epsilon = -\frac{1}{\epsilon}\frac{\partial\tilde{V}^\epsilon}{\partial\mathbf{D}}^\top\mathbf{R}_1\mathbf{D} + \frac{1}{2\epsilon^2}\frac{\partial\tilde{V}^\epsilon}{\partial\mathbf{y}}^\top\hat{\mathbf{C}}_1\boldsymbol{\Lambda}_1^{-1}\hat{\mathbf{C}}_1^\top\frac{\partial\tilde{V}^\epsilon}{\partial\mathbf{y}}.$$

Next, we consider the following ansatz for $\tilde{V}^\epsilon$:

$$\tilde{V}^\epsilon(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l) = V^0(t, \mathbf{f}, l) - \epsilon V^1(t, l) - \epsilon^2\left(\omega\circ\boldsymbol{\xi}^\epsilon\right)(\mathbf{x}, \mathbf{D}, \mathbf{f}, l) + o(\epsilon), \tag{A.27}$$

where $\left(\omega\circ\boldsymbol{\xi}^\epsilon\right)(\mathbf{x}, \mathbf{D}, \mathbf{f}, l) = \omega\left(\boldsymbol{\xi}^\epsilon\right)$ and

$$\boldsymbol{\xi}^\epsilon := \left(\boldsymbol{\xi}_1^\epsilon, \boldsymbol{\xi}_2^\epsilon\right) = \left(\epsilon^{-\frac{1}{2}}\left(\mathbf{x} - \mathbf{x}^{\mathrm{nf}}(\mathbf{f}, l)\right), \epsilon^{-\frac{1}{2}}\mathbf{D}\right). \tag{A.28}$$

We note that $V^1 = V^1(t, l), \omega = \omega(\boldsymbol{\xi})$ are functions to be determined later. Then (A.26) becomes

$$\begin{aligned}&\underbrace{\left(-\rho V^0 + \frac{\partial V^0}{\partial t} + \mathcal{L}^{\mathbf{x}} V^0\right)}_{I_0} + \epsilon\underbrace{\left(\rho V^1 - \frac{\partial V^1}{\partial t} - \mathcal{L}^{\mathbf{x}} V^1 - \rho{\boldsymbol{\xi}_2^\epsilon}^\top\mathbf{C}_1^{-1}\mathbf{R}_1\boldsymbol{\xi}_2^\epsilon\right)}_{I_1} \\ &+ \epsilon^2\underbrace{\left(\rho\left(\omega\circ\boldsymbol{\xi}^\epsilon\right) - \mathcal{L}^{\mathbf{x}}\left(\omega\circ\boldsymbol{\xi}^\epsilon\right) + \frac{1}{\epsilon}\frac{\partial\left(\omega\circ\boldsymbol{\xi}^\epsilon\right)}{\partial\mathbf{D}}^\top\mathbf{R}_1\mathbf{D} + \frac{1}{2\epsilon^2}\frac{\partial\left(\omega\circ\boldsymbol{\xi}^\epsilon\right)}{\partial\mathbf{y}}^\top\hat{\mathbf{C}}_1\boldsymbol{\Lambda}_1^{-1}\hat{\mathbf{C}}_1^\top\frac{\partial\left(\omega\circ\boldsymbol{\xi}^\epsilon\right)}{\partial\mathbf{y}}\right)}_{I_2} \\ &+ \rho\epsilon\left(\epsilon{\boldsymbol{\xi}_1^\epsilon}^\top\boldsymbol{\xi}_2^\epsilon + \sqrt{\epsilon}{\mathbf{x}^{\mathrm{nf}}}^\top\boldsymbol{\xi}_2^\epsilon - \frac{1}{2}\epsilon{\boldsymbol{\xi}_2^\epsilon}^\top\mathbf{C}_1^{-1}\boldsymbol{\xi}_2^\epsilon\right) - \epsilon{\boldsymbol{\xi}_2^\epsilon}^\top\mathbf{C}_1^{-1}\mathbf{R}_1\boldsymbol{\xi}_2^\epsilon + o(\epsilon) = 0.\end{aligned} \tag{A.29}$$

Using (45), (44) and (A.28), we derive

$$I_0 = -\rho V^0 + \frac{\partial V^0}{\partial t} + \mathcal{L}^{\mathbf{x}^{\mathrm{nf}}} V^0 - \frac{\gamma}{2}\epsilon {\boldsymbol{\xi}_1^\epsilon}^\top \boldsymbol{\Sigma}_P \boldsymbol{\xi}_1^\epsilon = -\frac{\gamma}{2}\epsilon {\boldsymbol{\xi}_1^\epsilon}^\top \boldsymbol{\Sigma}_P \boldsymbol{\xi}_1^\epsilon.$$

Then

$$I_1 = \epsilon\left(\rho V^1 - \frac{\partial V^1}{\partial t} - \mathcal{L}^{\mathbf{x}^{\mathrm{nf}}} V^1 - \rho {\boldsymbol{\xi}_2^\epsilon}^\top \mathbf{C}_1^{-1}\mathbf{R}_1 \boldsymbol{\xi}_2^\epsilon\right) + \frac{\gamma}{2}\epsilon^2 {\boldsymbol{\xi}_1^\epsilon}^\top \boldsymbol{\Sigma}_P \boldsymbol{\xi}_1^\epsilon.$$

Next, we simplify I_2. For $\boldsymbol{\zeta} \in \{\mathbf{x}, \mathbf{D}, \mathbf{f}, l\}$, by applying the chain rule to the composite function $\omega \circ \boldsymbol{\xi}$, we obtain

$$\frac{\partial\left(\omega \circ \boldsymbol{\xi}^\epsilon\right)}{\partial \boldsymbol{\zeta}} = \epsilon^{-\frac{1}{2}}\left[\frac{\partial\left(\mathbf{x} - \mathbf{x}^{\mathrm{nf}}\right)}{\partial \boldsymbol{\zeta}}^\top \frac{\partial \omega}{\partial \boldsymbol{\xi}_1} + \frac{\partial \mathbf{D}}{\partial \boldsymbol{\zeta}}^\top \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}\right].$$

Moreover, since $\mathbf{x}^{\mathrm{nf}}$ is linear in $\mathbf{f}$ and l, for $\boldsymbol{\zeta}, \tilde{\boldsymbol{\zeta}} \in \{\mathbf{f}, l\}$, we derive

$$\frac{\partial^2\left(\omega \circ \boldsymbol{\xi}^\epsilon\right)}{\partial \boldsymbol{\zeta}\partial \tilde{\boldsymbol{\zeta}}^\top} = \epsilon^{-1} \frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial \boldsymbol{\zeta}}^\top \frac{\partial^2 \omega}{\partial \boldsymbol{\xi}_1 \partial \boldsymbol{\xi}_1^\top} \frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial \tilde{\boldsymbol{\zeta}}}.$$

Then the terms in I_2 can be reformulated as

$$I_2 = \epsilon\Big[-\frac{1}{2}\mathcal{Q}\mathcal{V}^l(\omega) + \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}^\top \mathbf{R}_1 \boldsymbol{\xi}_2^\epsilon - \frac{1}{2}\left(\frac{\partial \omega}{\partial \boldsymbol{\xi}_1} + \mathbf{C}_1^\top \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}\right)^\top \boldsymbol{\Lambda}_1^{-1}\left(\frac{\partial \omega}{\partial \boldsymbol{\xi}_1} + \mathbf{C}_1^\top \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}\right)\Big] + o(\epsilon),$$

where

$$\mathcal{Q}\mathcal{V}^l(\omega) = \mathrm{tr}\left(\frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial \mathbf{f}}^\top \frac{\partial^2 \omega}{\partial \boldsymbol{\xi}_1 \partial \boldsymbol{\xi}_1^\top} \frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial \mathbf{f}} \boldsymbol{\Sigma}_f\right) + \frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial l}^\top \frac{\partial^2 \omega}{\partial \boldsymbol{\xi}_1 \boldsymbol{\xi}_1^\top} \frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial l} \Sigma_l l^2 + 2\frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial l}^\top \frac{\partial^2 \omega}{\partial \boldsymbol{\xi}_1 \boldsymbol{\xi}_1^\top} \frac{\partial \mathbf{x}^{\mathrm{nf}}}{\partial \mathbf{f}} \boldsymbol{\Sigma}_{fl} l.$$

We collect the terms of order ϵ in (A.29) and let their summation equal to zero. Following Ekren and Muhle-Karbe (2019), we introduce the following two differential equations for function $(\omega, a) = (\omega(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2), a(l))$: The first corrector equation

$$-\frac{1}{2}\mathcal{Q}\mathcal{V}^l(\omega) - \frac{\gamma}{2}\boldsymbol{\xi}_1^\top \boldsymbol{\Sigma}_P \boldsymbol{\xi}_1 - \rho \boldsymbol{\xi}_2^\top \mathbf{C}_1^{-1}\mathbf{R}_1 \boldsymbol{\xi}_2 + \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}^\top \mathbf{R}_1 \boldsymbol{\xi}_2 + \frac{1}{2}\left(\frac{\partial \omega}{\partial \boldsymbol{\xi}_1} + \mathbf{C}_1^\top \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}\right)^\top \boldsymbol{\Lambda}_1^{-1}\left(\frac{\partial \omega}{\partial \boldsymbol{\xi}_1} + \mathbf{C}_1^\top \frac{\partial \omega}{\partial \boldsymbol{\xi}_2}\right) + a(l) = 0, \tag{A.30}$$

and the second corrector equation

$$\rho V^1 - \frac{\partial V^1}{\partial t} - \mathcal{L}^{\mathbf{x}^{\mathrm{nf}}} V^1 = a(l), \quad V^1(T, l) = 0. \tag{A.31}$$

To proceed, we introduce the following lemma.

Lemma A.3. *Suppose $\mathbf{C}_1^{-1}\mathbf{R}_1 + \mathbf{R}_1\mathbf{C}_1^{-1}$ is positive definite. The algebraic Riccati equation (47) has a maximal solution $\mathbf{K} \in \mathbb{R}^{2n\times 2n}$. Moreover, $\mathbf{K}$ is positive definite.*

Proof. This lemma can be proved similarly following the proof of Lemma 3.2 in Ekren and Muhle-Karbe (2019). □

Then, by a simple validation and the Feynman-Kac formula, the following two lemmas are valid.

Lemma A.4. *The pair (ω, a) that solves the first corrector equation (A.30) admits the following form*

$$\omega(\boldsymbol{\xi}) = \frac{1}{2}\boldsymbol{\xi}^\top \mathbf{K}\boldsymbol{\xi},$$
$$a(l) = \frac{1}{2}\boldsymbol{\Sigma}_{Pl}^\top {\boldsymbol{\Sigma}_P^{-1}}^\top \mathbf{K}_{11}\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{Pl}\Sigma_l l^2 + \frac{1}{\gamma}\boldsymbol{\Sigma}_{Pl}^\top \mathbf{K}_{11}{\boldsymbol{\Sigma}_P^{-1}}^\top \mathbf{B}\boldsymbol{\Sigma}_{fl} l + \frac{1}{2\gamma^2}\mathrm{tr}\left(\mathbf{B}^\top {\boldsymbol{\Sigma}_P^{-1}}^\top \mathbf{K}_{11}\boldsymbol{\Sigma}_P^{-1}\mathbf{B}\boldsymbol{\Sigma}_f\right),$$

where $\mathbf{K} \in \mathbb{R}^{2n\times 2n}$ is the maximal solution in Lemma A.3 and $\mathbf{K}_{11} \in \mathbb{R}^{n\times n}$ is the submatrix of $\mathbf{K}$ formed by intersecting the first n rows with the first n columns.

Lemma A.5. *The solution to (A.31), i.e. V^1, admits a classical solution and it has the probability representation (48).*

We define

$$\tilde{\Psi}^\epsilon(t, \mathbf{x}, \mathbf{D}, \mathbf{f}, l) := V^0(t, \mathbf{f}, l) - \epsilon V^1(t, l) - \epsilon^2\left(\omega \circ \boldsymbol{\xi}^\epsilon\right)(\mathbf{x}, \mathbf{D}, \mathbf{f}, l).$$

By the previous discussion, we have

$$-\rho\tilde{\Psi}^\epsilon + \frac{\partial \tilde{\Psi}^\epsilon}{\partial t} + \mathcal{L}^{\mathbf{x}}\tilde{\Psi}^\epsilon + \mathcal{H}\tilde{\Psi}^\epsilon + \rho\epsilon\left(\mathbf{x}^\top \mathbf{D} - \frac{1}{2}\mathbf{D}^\top \mathbf{C}_1^{-1}\mathbf{D}\right) - \mathbf{D}^\top \mathbf{C}_1^{-1}\mathbf{R}_1\mathbf{D} = o(\epsilon).$$

Let

$$\Psi^{\epsilon}(t,\mathbf{x},\mathbf{D},\mathbf{f},l) := \tilde{\Psi}^{\epsilon}(t,\mathbf{x},\mathbf{D},\mathbf{f},l) - \epsilon(\mathbf{x}^{\top}\mathbf{D} - \frac{1}{2}\mathbf{D}^{\top}\mathbf{C}_1^{-1}\mathbf{D}).$$

Then we have

$$-\rho\Psi^{\epsilon} + \frac{\partial\Psi^{\epsilon}}{\partial t} + \mathcal{L}^{\mathbf{x}}\Psi^{\epsilon} - \frac{1}{\epsilon}\mathbf{x}^{\top}\mathbf{R}_1\mathbf{D} - \frac{1}{\epsilon}\frac{\partial\Psi^{\epsilon}}{\partial\mathbf{D}}^{\top}\mathbf{R}_1\mathbf{D} + \frac{1}{2\epsilon}\left(\frac{\partial\Psi^{\epsilon}}{\partial\mathbf{x}} + \epsilon\mathbf{C}_1^{\top}\frac{\partial\Psi^{\epsilon}}{\partial\mathbf{D}} + \epsilon\mathbf{C}_1^{\top}\mathbf{x}\right)^{\top}\mathbf{\Lambda}_1^{-1}\left(\frac{\partial\Psi^{\epsilon}}{\partial\mathbf{x}} + \epsilon\mathbf{C}_1^{\top}\frac{\partial\Psi^{\epsilon}}{\partial\mathbf{D}} + \epsilon\mathbf{C}_1^{\top}\mathbf{x}\right) = o(\epsilon),$$

which implies that the function Ψ^{ϵ} satisfies the equation (A.24) at the leading order. By utilizing the first-order condition, the optimal leading-order performance is attained by

$$\mathbf{u}^{\epsilon}(\mathbf{x},\mathbf{D},\mathbf{f},l) = \epsilon^{-1}\mathbf{\Lambda}_1^{-1}\hat{\mathbf{C}}_1^{\top}\mathbf{K}\begin{pmatrix}\mathbf{x}^{\mathrm{nf}}(\mathbf{f},l) - \mathbf{x} \\ -\epsilon\mathbf{D}\end{pmatrix}.$$

We complete the proof. □

Appendix B. A discrete-time formulation

In this section, we provide a discrete-time formulation of the asset-liability management under market frictions. Let the time variable $t = 1, 2, \ldots, T$. In the financial market, we consider an economy with n stocks: $\mathbf{P}_t = (P_{1t}, P_{2t}, \ldots, P_{nt})^{\top}$. The percentage excess returns of these assets are denoted by a random vector $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{nt})^{\top}$. In other words, $r_{it} = P_{it+1}/P_{it} - 1 - r_f$, where r_f denotes the risk-free interest rate and $i = 1, 2, \ldots, n$. As our main focus is the impact of market liquidity on the asset-liability management, for simplicity, we assume $\mathbf{r}_t$'s are independent and identically distributed with mean vector $\boldsymbol{\mu}_r$ and variance-covariance matrix $\mathbf{\Sigma}_r$.

The manager rebalances the strategy at the end of each time point. The vector of the dollar holdings, i.e. the portfolio, is denoted by $\mathbf{x}_t$ and the dollar trade vector is denoted by $\Delta\mathbf{x}_t := \mathbf{x}_t - \mathbf{x}_{t-1}$. We assume a linear temporary price impact in which trading $\mathbf{u}_t$ dollar amount incurs a trading cost of $\frac{1}{2}\Delta\mathbf{x}_t^{\top}\mathbf{\Lambda}\Delta\mathbf{x}_t$. We assume $\mathbf{\Lambda}$ is positive definite. Let the current asset value of the insurance company be a_t. By the means of self-financing principle, we obtain the dynamics of a_t:

$$a_{t+1} = (1 + r_f)a_t + \mathbf{x}_t^{\top}\mathbf{r}(t+1) - \frac{1}{2}\Delta\mathbf{x}_t^{\top}\mathbf{\Lambda}\Delta\mathbf{x}_t. \tag{B.1}$$

Following Leippold et al. (2004), the liability process is defined as

$$l_{t+1} = l_t q_{t+1}, \tag{B.2}$$

where q_t's are independent and identically distributed non-negative random variables representing the liability return with mean μ_l and variance σ_l^2. We assume a constant covariance between $\mathbf{r}$ and q, i.e. $\mathrm{Cov}(\boldsymbol{\varepsilon}_{rt}, q_t) = \mathbf{\Sigma}_{rl} \in \mathbb{R}^n$.

The persistent price impact is addressed using the return distortion $\mathbf{D}_t$ satisfies the following dynamics:

$$\mathbf{D}_{t+1} = (\mathbf{I} - \mathbf{R})(\mathbf{D}_t + \mathbf{C}\Delta\mathbf{x}_t), \tag{B.3}$$

where we assume a trading speed of $\Delta\mathbf{x}_t$ will raise or reduce the return by $\mathbf{C}\Delta\mathbf{x}_t$, where $\mathbf{C} \in \mathbb{R}^{n\times n}$ is a positive definite matrix measuring the persistent price impact level. $\mathbf{R} \in \mathbb{R}^{n\times n}$ is a positive definite matrix representing the mean-reverting speed (resiliency of the persistent price impact). The process $\mathbf{D}$ extends the preceding model by having a return vector $\mathbf{r}_t + \mathbf{D}_{t+1} - \mathbf{D}_t$.

Similar to the analysis in Gârleanu and Pedersen (2013), for each period from t to $t+1$, the return rate of the net asset value due to the posttrade return distortion is

$$\mathbf{D}_{t+1} - (\mathbf{D}_t + \mathbf{C}\Delta\mathbf{x}_t) = -\mathbf{R}\left(\mathbf{D}_t + \mathbf{C}\Delta\mathbf{x}_t\right),$$

and the persistent price impact also raises the stock value by

$$\mathbf{x}_{t-1}\mathbf{C}\Delta\mathbf{x}_t + \frac{1}{2}\Delta\mathbf{x}_t^{\top}\mathbf{C}\Delta\mathbf{x}_t.$$

The first term represents the mark-to-market gain from the old position $\mathbf{x}_{t-1}$ from the price impact of the new trade $\mathbf{C}\mathbf{u}_t$. The second term reflects that the traded assets $\Delta\mathbf{x}_t$ are assumed to be executed at the average distortion $\mathbf{D}_t + \frac{1}{2}\mathbf{C}\Delta\mathbf{x}_t$ and hence, $\mathbf{u}_t$ earns a mark-to-market gain of $\frac{1}{2}\Delta\mathbf{x}_t^{\top}\mathbf{C}\Delta\mathbf{x}_t$ as the price moves up an additional $\frac{1}{2}\mathbf{C}\Delta\mathbf{x}_t$. Moreover, at time t, the conditional variance of the dollar excess return of the net asset value in the next period is

$$\mathbf{x}_t^{\top}\mathbf{\Sigma}_r\mathbf{x}_t - 2\mathbf{x}_t^{\top}\mathbf{\Sigma}_{rl}l_t + \sigma_l^2 l_t^2.$$

The manager aims to optimize the local mean-variance criterion which is a cumulative sum of the conditional expectation and variance of the dollar excess return in each period. Specifically, the objective function is

$$\begin{aligned}\max_{\mathbf{x}_1,\ldots,\mathbf{x}_T}\mathbb{E}\Big[\sum_{t=0}^{T}\rho^{t+1}\big\{\left(1 + r_f - q_{t+1}\right)l_t + \mathbf{x}_t^{\top}\left[\boldsymbol{\mu}_r - \mathbf{R}(\mathbf{D}_t + \mathbf{C}\Delta\mathbf{x}_t)\right] - \frac{1}{2}\gamma\left(\mathbf{x}_t^{\top}\mathbf{\Sigma}_r\mathbf{x}_t - 2\mathbf{x}_t^{\top}\mathbf{\Sigma}_{rl}l_t + \sigma_l^2 l_t^2\right)\big\} \\ + \sum_{t=0}^{T}\rho^{t}\left(-\frac{1}{2}\Delta\mathbf{x}_t^{\top}\mathbf{\Lambda}\Delta\mathbf{x}_t + \mathbf{x}_{t-1}^{\top}\mathbf{C}\Delta\mathbf{x}_t + \frac{1}{2}\Delta\mathbf{x}_t^{\top}\mathbf{C}\Delta\mathbf{x}_t\right)\Big],\end{aligned} \tag{B.4}$$

where $\gamma > 0$ represents the risk aversion parameter, $\rho \in (0,1)$ is the discount factor. The third line is discounted by ρ^t because these cash flows are incurred at time t, not time $t+1$.

For the state vector $\left(\mathbf{x}_{t-1}^\top, \mathbf{D}_t^\top, l_t\right)^\top = (\mathbf{x}^\top, \mathbf{D}^\top, l)^\top \in \mathbb{R}^{2n+1}$, the Bellman equations for the optimization problem (B.4) is

$$V_{T+1}(\mathbf{x}, \mathbf{D}, l) = 0, \tag{B.5}$$

$$\begin{aligned} V_t(\mathbf{x}, \mathbf{D}, l) = \max_{\mathbf{x}_t} \Big\{ & \rho^{t+1} \big\{ \left(1 + r_f - \mu_l\right) l + \mathbf{x}_t^\top \left[\boldsymbol{\mu}_r - \mathbf{R}\left(\mathbf{D} + \mathbf{C}\left(\mathbf{x}_t - \mathbf{x}\right)\right)\right] \\ & \qquad - \frac{1}{2}\gamma \left(\mathbf{x}_t^\top \boldsymbol{\Sigma}_r \mathbf{x}_t - 2\mathbf{x}_t^\top \boldsymbol{\Sigma}_{rl} l + \sigma_l^2 l^2\right) \big\} \\ & + \rho^t \big\{ - \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}\right)^\top \boldsymbol{\Lambda} \left(\mathbf{x}_t - \mathbf{x}\right) + \mathbf{x}^\top \mathbf{C}\left(\mathbf{x}_t - \mathbf{x}\right) \\ & \qquad + \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}\right)^\top \mathbf{C}\left(\mathbf{x}_t - \mathbf{x}\right) \big\} \\ & + \mathbb{E}_t \left[V_{t+1}\left(\mathbf{x}_t, \left(\mathbf{I} - \mathbf{R}\right)\left(\mathbf{D} + \mathbf{C}\mathbf{x}_t - \mathbf{C}\mathbf{x}\right), l^2 q_{t+1}^2\right) \right] \Big\}, \end{aligned} \tag{B.6}$$

where $\mathbb{E}_t[\cdot] = \mathbb{E}\left[\cdot \middle| \mathbf{x}_{t-1} = \mathbf{x}, \mathbf{D}_t = \mathbf{D}, l_t = l\right]$.

For $t = 0, 1, \ldots, T+1$, consider the following ansatz for V_t in the Bellman equation (B.6):

$$\begin{aligned} V_t(\mathbf{x}, \mathbf{D}, l) = \rho^t \Big[& \frac{1}{2}\mathbf{x}^\top \mathbf{E}_{xx}(t)\mathbf{x} + \frac{1}{2}\mathbf{D}^\top \mathbf{E}_{DD}(t)\mathbf{D} + \frac{1}{2}E_{ll}(t)l^2 + \mathbf{x}^\top \mathbf{E}_{xD}(t)\mathbf{D} + \mathbf{x}^\top \mathbf{E}_{xl}(t)l \\ & + \mathbf{D}^\top \mathbf{E}_{Dl}(t)l + \mathbf{x}^\top \mathbf{E}_x(t) + \mathbf{D}^\top \mathbf{E}_D(t) + E_l(t)l + E_c(t) \Big], \end{aligned} \tag{B.7}$$

with $\mathbf{E}_{xx}(T) = \mathbf{E}_{DD}(T) = \mathbf{E}_{xD}(T) = \mathbf{0}_{n\times n}$, $\mathbf{E}_{xl}(T) = \mathbf{E}_{Dl}(T) = \mathbf{E}_x(T) = \mathbf{E}_D(T) = \mathbf{0}_{n\times 1}$, $E_{ll}(T) = E_l(T) = E_c(T) = 0$.

For the ease of illustration, we define $\mathbf{y} := (\mathbf{x}^\top, \mathbf{D}^\top)^\top$, $\mathbf{E}_{yy} := \begin{pmatrix} \mathbf{E}_{xx} & \mathbf{E}_{xD} \\ \mathbf{E}_{xD}^\top & \mathbf{E}_{DD} \end{pmatrix}$, $\mathbf{E}_{yl} := \begin{pmatrix} \mathbf{E}_{xl} \\ \mathbf{E}_{Dl} \end{pmatrix}$, $\mathbf{E}_y := \begin{pmatrix} \mathbf{E}_x \\ \mathbf{E}_D \end{pmatrix}$ and the matrices $\mathbf{M}_1, \mathbf{N}_1 \in \mathbb{R}^{2n\times 2n}$, $\mathbf{M}_2 \in \mathbb{R}^{n\times n}$, $\mathbf{N}_2 \in \mathbb{R}^{2n\times n}$, $\mathbf{M}_3 \in \mathbb{R}^{n\times 2n}$:

$$\begin{aligned} \mathbf{M}_1 &= \begin{pmatrix} \boldsymbol{\Lambda} + \mathbf{C} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \end{pmatrix}, \\ \mathbf{M}_2 &= \rho \mathbf{R}\mathbf{C} + \rho \mathbf{C}^\top \mathbf{R}^\top + \rho\gamma \boldsymbol{\Sigma}_r + \boldsymbol{\Lambda} - \mathbf{C}, \\ \mathbf{M}_3 &= \begin{pmatrix} \boldsymbol{\Lambda} + \rho \mathbf{R}\mathbf{C} & -\rho \mathbf{R} \end{pmatrix}, \\ \mathbf{N}_1 &= \begin{pmatrix} \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ -\left(\mathbf{I} - \mathbf{R}\right)\mathbf{C} & \mathbf{I} - \mathbf{R} \end{pmatrix}, \quad \mathbf{N}_2 = \begin{pmatrix} \mathbf{I} \\ \left(\mathbf{I} - \mathbf{R}\right)\mathbf{C} \end{pmatrix}. \end{aligned}$$

Note that at stage t, the state variable is $\mathbf{y}_t = \left(\mathbf{x}_{t-1}^\top, \mathbf{D}_t^\top\right)$ and the control variable is $\mathbf{x}_t$. For $\mathbf{y}_t = \mathbf{y}$, according to the special form of V_{t+1} in (B.7), we obtain

$$\begin{aligned} & \mathbb{E}_t \left[V_{t+1}\left(\mathbf{x}_t, \left(\mathbf{I} - \mathbf{R}\right)\left(\mathbf{D} + \mathbf{C}\mathbf{x}_t - \mathbf{C}\mathbf{x}\right), l^2 q_{t+1}^2\right) \right] \\ = {} & \rho^{t+1} \Big[\frac{1}{2}\left(\mathbf{N}_1\mathbf{y} + \mathbf{N}_2\mathbf{x}_t\right)^\top \mathbf{E}_{yy}(t+1)\left(\mathbf{N}_1\mathbf{y} + \mathbf{N}_2\mathbf{x}_t\right) + \frac{1}{2}E_{ll}(t+1)\left(\mu_l^2 + \sigma_l^2\right) l^2 \\ & + \left(\mathbf{N}_1\mathbf{y} + \mathbf{N}_2\mathbf{x}_t\right)^\top \mathbf{E}_{yl}(t+1)\mu_l l + \left(\mathbf{N}_1\mathbf{y} + \mathbf{N}_2\mathbf{x}_t\right)^\top \mathbf{E}_y(t+1) + E_l(t+1)\mu_l l + E_c(t+1) \Big]. \end{aligned}$$

Plugging the above into (B.6), we obtain

$$\begin{aligned} V_t(\mathbf{y}, l) = \max_{\mathbf{x}_t} \rho^t \Big\{ & - \frac{1}{2}\mathbf{x}_t^\top \mathbf{H}(t)\mathbf{x}_t + \mathbf{x}_t^\top \left[\mathbf{L}_y(t)\mathbf{y} + \mathbf{L}_l(t)l + \mathbf{L}_c(t)\right] - \frac{1}{2}\mathbf{y}^\top \mathbf{M}_1 \mathbf{y} \\ & + \frac{\rho}{2}\mathbf{y}^\top \mathbf{N}_1^\top \mathbf{E}_{yy}(t+1)\mathbf{N}_1\mathbf{y} + \rho \mathbf{y}^\top \mathbf{N}_1 \mathbf{E}_{yl}(t+1)\mu_l l + \rho \mathbf{y}^\top \mathbf{N}_1 \mathbf{E}_y(t+1) \\ & + \frac{\rho}{2} l^2 \left[\mathbf{E}_{ll}(t+1)\left(\mu_l^2 + \sigma_l^2\right) - \gamma\sigma_l^2\right] + \rho\left(1 + r_f - \mu_l + E_l(t+1)\mu_l\right) l + \rho E_c(t+1). \Big\} \end{aligned} \tag{B.8}$$

Let $\mathbf{H}(t) \in \mathbb{R}^{n\times n}$, $\mathbf{L}_y(t) \in \mathbb{R}^{n\times 2n}$, $\mathbf{L}_l(t), \mathbf{L}_c(t) \in \mathbb{R}^{n\times 1}$ be given by

$$\begin{aligned} \mathbf{H}(t) &= \mathbf{M}_2 - \rho \mathbf{N}_2^\top \mathbf{E}_{yy}(t+1)\mathbf{N}_2, \\ \mathbf{L}_y(t) &= \mathbf{M}_3 + \rho \mathbf{N}_2^\top \mathbf{E}_{yy}(t+1)\mathbf{N}_1, \\ \mathbf{L}_l(t) &= \rho\gamma \boldsymbol{\Sigma}_{rl} + \rho \mathbf{N}_2^\top \mathbf{E}_{yl}(t+1)\mu_l, \\ \mathbf{L}_c(t) &= \rho \boldsymbol{\mu}_r + \rho \mathbf{N}_2^\top \mathbf{E}_y(t+1). \end{aligned} \tag{B.9}$$

Before proceeding to solve (B.8), we introduce the following assumption to ensure the global maximizer exists.

Assumption 2. For $t = T, T-1, \ldots, 0$, $\mathbf{H}(t) > 0$.

We note that the above assumption is easy to verify in practice since $\mathbf{H}(t)$ can be derived by solving the recursive equation of $\mathbf{E}_{yy}(t)$, i.e. equation (B.11). See Glasserman and Xu (2013) for a similar assumption. Immediately, a sufficient condition for Assumption 2 is $\mathbf{M}_2 = \rho\mathbf{RC} + \rho\mathbf{C}^\top\mathbf{R}^\top + \rho\gamma\mathbf{\Sigma}_r + \mathbf{\Lambda} - \mathbf{C} > 0$ and $\mathbf{E}_{yy}(t) \le 0$ for $t = T, T-1, \ldots, 0$.

Theorem B.1. *Suppose Assumption 2 is enforced. Let $\mathbf{H}(t), \mathbf{L}_x(t), \mathbf{L}_D(t), \mathbf{L}_l(t), \mathbf{L}_c(t)$ be matrix valued functions given by (B.9) and (B.18). For $t = 0, 1, \ldots, T$,*

1. *the optimal portfolio is*

$$\mathbf{x}_t^* = \mathbf{H}(t)^{-1}\left[\mathbf{L}_x(t)\mathbf{x}_{t-1}^* + \mathbf{L}_D(t)\mathbf{D}_t + \mathbf{L}_l(t)l_t + \mathbf{L}_c(t)\right].$$

2. *if $\mathbf{H}(t) - \mathbf{L}_x(t)$ is non-singular,*

$$\mathbf{x}_t^* = \mathbf{x}_{t-1}^* + \mathbf{M}_t^{\text{rate}}\left[\mathbf{M}_t^{\text{aim}} - \mathbf{x}_{t-1}^*\right],$$

where

$$\begin{aligned}
\mathbf{M}_t^{\text{rate}} &= \mathbf{I} - \mathbf{H}^{-1}(t)\mathbf{L}_x(t),\\
\mathbf{M}_t^{\text{aim}} &= (\mathbf{H}(t) - \mathbf{L}_x(t))^{-1}\left[\mathbf{L}_D(t)\mathbf{D}_t + \mathbf{L}_l(t)l_t + \mathbf{L}_c(t)\right].
\end{aligned}$$

Proof. Since (ii) is a direct result of (i), we only prove (i). The first-order derivative condition for (B.8) yields the maximizer

$$\mathbf{x}_t^* = \mathbf{H}(t)^{-1}\left[\mathbf{L}_y(t)\mathbf{y} + \mathbf{L}_l(t)l + \mathbf{L}_c(t)\right]. \tag{B.10}$$

Substituting (B.10) into the Bellman equation (B.6) and comparing the coefficients, we derive the following discrete Riccati equation:

$$\mathbf{E}_{yy}(t) = \mathbf{L}_y(t)^\top\mathbf{H}(t)^{-1}\mathbf{L}_y(t) + \rho\mathbf{N}_1^\top\mathbf{E}_{yy}(t+1)\mathbf{N}_1 - \mathbf{M}_1, \tag{B.11}$$

$$E_{ll}(t) = \mathbf{L}_l(t)^\top\mathbf{H}(t)^{-1}\mathbf{L}_l(t) + \rho E_{ll}(t+1)\left(\mu_l^2 + \sigma_l^2\right) - \rho\gamma\sigma_l^2, \tag{B.12}$$

$$\mathbf{E}_{yl}(t) = \mathbf{L}_y(t)^\top\mathbf{H}(t)^{-1}\mathbf{L}_l(t) + \rho\mu_l\mathbf{N}_1\mathbf{E}_{yl}(t+1), \tag{B.13}$$

$$\mathbf{E}_y(t) = \mathbf{L}_y(t)^\top\mathbf{H}(t)^{-1}\mathbf{L}_c(t) + \rho\mathbf{N}_1\mathbf{E}_y(t+1), \tag{B.14}$$

$$E_l(t) = \mathbf{L}_l(t)^\top\mathbf{H}(t)^{-1}\mathbf{L}_c(t) + \rho\mu_l E_l(t+1) + \rho(1 + r_f - \mu_l), \tag{B.15}$$

$$E_c(t) = \frac{1}{2}\mathbf{L}_c(t)^\top\mathbf{H}(t)^{-1}\mathbf{L}_c(t) + \rho E_c(t+1), \tag{B.16}$$

for $t = 1, \ldots, T$ and

$$\begin{aligned}
\mathbf{E}_{yy}(T+1) = \mathbf{0}_{2n\times 2n},\ \ \mathbf{E}_{yl}(T+1) = \mathbf{E}_y(T+1) = \mathbf{0}_{2n\times 1},&\\
E_{ll}(T+1) = E_l(T+1) = E_c(T+1) = 0.&
\end{aligned} \tag{B.17}$$

The solution to the system (B.11)–(B.17) is well defined under Assumption 2. Let $\mathbf{L}_x, \mathbf{L}_D$ be the first n columns and the last n columns respectively, then

$$\begin{aligned}
\mathbf{L}_x(t) &= -\rho\left[\mathbf{E}_{xD}(t+1) + \mathbf{C}^\top(\mathbf{I} - \mathbf{R})\mathbf{E}_{DD}(t+1)\right](\mathbf{I} - \mathbf{R})\mathbf{C} + \mathbf{\Lambda} + \rho\mathbf{RC},\\
\mathbf{L}_D(t) &= -\rho\mathbf{R} + \rho\left[\mathbf{E}_{xD}(t+1) + \mathbf{C}^\top(\mathbf{I} - \mathbf{R})\mathbf{E}_{DD}(t+1)\right](\mathbf{I} - \mathbf{R}).
\end{aligned} \tag{B.18}$$

We rewrite the optimal solution (B.10) as

$$\begin{aligned}
\mathbf{x}_t^* &= \mathbf{H}(t)^{-1}\left[\mathbf{L}_x(t)\mathbf{x}_{t-1}^* + \mathbf{L}_D(t)\mathbf{D}_t + \mathbf{L}_l(t)l_t + \mathbf{L}_c(t)\right],\\
&= \mathbf{x}_{t-1}^* + \mathbf{M}_t^{\text{rate}}\left[\mathbf{M}_t^{\text{aim}} - \mathbf{x}_{t-1}^*\right]. \quad \square
\end{aligned}$$

Define the single-period Markowitz mean-variance portfolio as $\mathbf{x}_t^{\text{mv}} = \frac{1}{\gamma}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_r$. The following corollary states the optimal portfolio with only temporary price impact.

Proposition B.2. *Suppose there is only temporary price impact. Then for $t = 0, 1, \ldots, T$,*

(i) *The optimal trading strategy is*

$$\mathbf{x}_t^* = \mathbf{x}_{t-1}^* + \mathbf{M}_t^{\text{rate}}\left[\mathbf{M}_t^{\text{aim}} - \mathbf{x}_{t-1}^*\right],$$

where

$$\mathbf{M}_t^{\text{rate}} = \mathbf{I} - \left(\rho\gamma\boldsymbol{\Sigma}_r + \boldsymbol{\Lambda} - \rho\mathbf{E}_{xx}(t+1)\right)^{-1}\boldsymbol{\Lambda},$$
$$\mathbf{M}_t^{\text{aim}} = \mathbf{Z}(t)\left(\mathbf{x}_t^{\text{mv}} + \boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rl}l_t\right) + (\mathbf{I} - \mathbf{Z}(t))\,\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right],$$

$\mathbf{Z}(t) = \gamma\left(\gamma\boldsymbol{\Sigma}_r - \mathbf{E}_{xx}(t+1)\right)^{-1}\boldsymbol{\Sigma}_r$ and $\mathbf{E}_{xx}(t)$ satisfies the recursive equation (B.19).

(ii) *if the temporary price impact level $\boldsymbol{\Lambda} = \lambda\boldsymbol{\Sigma}_r$ with a constant $\lambda > 0$, then*

$$\mathbf{M}_t^{\text{rate}} = 1 - \frac{\lambda}{\rho\left(\gamma + a_{xx}(t+1)\right) + \lambda},$$
$$\mathbf{M}_t^{\text{aim}} = z(t)\left(\mathbf{x}_t^{\text{mv}} + \boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rl}l_t\right) + (1 - z(t))\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right],$$

where $z(t) = \frac{\gamma}{\gamma + a_{xx}(t+1)}$, and $\{a_{xx}(t)\}_{t=0,1,\dots,T+1}$ is a non-negative sequence solving the recursive equations (B.32). Moreover, $a_{xx}(t)$ is increasing in λ and $\frac{a_{xx}(t)}{\gamma}$ is decreasing in γ.

Proof. In this case, the equation system (B.11)–(B.17) can be simplified. We have $\mathbf{E}_{DD} = \mathbf{0}_{n\times n}$, $\mathbf{E}_{Dl} = \mathbf{E}_D = \mathbf{0}_{n\times 1}$ and

$$\mathbf{E}_{xx}(t) = \boldsymbol{\Lambda}\mathbf{H}(t)^{-1}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}, \tag{B.19}$$
$$E_{ll}(t) = \rho^2\left(\gamma\boldsymbol{\Sigma}_{rl} + \mu_l\mathbf{E}_{xl}(t+1)\right)^\top\mathbf{H}(t)^{-1}\left(\gamma\boldsymbol{\Sigma}_{rl} + \mu_l\mathbf{E}_{xl}(t+1)\right) + \rho E_{ll}(t+1)\left(\mu_l^2 + \sigma_l^2\right) - \gamma\rho\sigma_l^2, \tag{B.20}$$
$$\mathbf{E}_{xl}(t) = \rho\boldsymbol{\Lambda}\mathbf{H}(t)^{-1}\left(\gamma\boldsymbol{\Sigma}_{rl} + \mu_l\mathbf{E}_{xl}(t+1)\right), \tag{B.21}$$
$$\mathbf{E}_x(t) = \rho\boldsymbol{\Lambda}\mathbf{H}(t)^{-1}\left(\boldsymbol{\mu}_r + \mathbf{E}_x(t+1)\right), \tag{B.22}$$
$$E_l(t) = \rho^2\left(\gamma\boldsymbol{\Sigma}_{rl} + \mu_l\mathbf{E}_{xl}(t+1)\right)\mathbf{H}(t)^{-1}\left(\mathbf{E}_x(t+1) + \mu_l\right), \tag{B.23}$$
$$E_c(t) = \rho^2\left(\mathbf{E}_x(t+1) + \boldsymbol{\mu}_r\right)^\top\mathbf{H}(t)^{-1}\left(\mathbf{E}_x(t+1) + \boldsymbol{\mu}_r\right) + \rho E_c(t+1), \tag{B.24}$$

with $\mathbf{E}_{xx}(T+1) = \mathbf{0}_{n\times n}$, $\mathbf{E}_{xl}(T+1) = \mathbf{E}_x(T+1) = \mathbf{0}_{n\times 1}$, $E_{ll}(T+1) = E_l(T+1) = E_c(T+1) = 0$ and $\mathbf{H}(t) = \rho\gamma\boldsymbol{\Sigma}_r + \boldsymbol{\Lambda} - \rho\mathbf{E}_{xx}(t+1)$.

The optimal trading strategy is given by

$$\mathbf{x}_t^* = \mathbf{H}(t)^{-1}\left[\boldsymbol{\Lambda}\mathbf{x}_{t-1}^* + \rho\left(\gamma\boldsymbol{\Sigma}_{rl} + \mu_l\mathbf{E}_{xl}(t+1)\right)l_t + \rho\mathbf{E}_x(t+1) + \rho\mu_r\right],$$
$$= \mathbf{x}_{t-1}^* + \mathbf{M}_t^{\text{rate}}\left[\mathbf{M}_t^{\text{aim}} - \mathbf{x}_{t-1}^*\right],$$

where

$$\mathbf{M}_t^{\text{rate}} = \mathbf{I} - \mathbf{H}(t)^{-1}\boldsymbol{\Lambda}, \tag{B.25}$$
$$\mathbf{M}_t^{\text{aim}} = \rho\left(\mathbf{H}(t) - \boldsymbol{\Lambda}\right)^{-1}\left[\left(\gamma\boldsymbol{\Sigma}_{rl} + \mu_l\mathbf{E}_{xl}(t+1)\right)l_t + \mathbf{E}_x(t+1) + \mu_r\right]. \tag{B.26}$$

Then we express (B.26) using the single-period Markowitz mean-variance portfolio

$$\mathbf{M}_t^{\text{aim}} = \rho\left(\mathbf{H}(t) - \boldsymbol{\Lambda}\right)^{-1}\left[\gamma\boldsymbol{\Sigma}_r\mathbf{x}_t^{\text{mv}} + \gamma\boldsymbol{\Sigma}_{rl}l_t + \mu_l\mathbf{E}_{xl}(t+1)l_t + \mathbf{E}_x(t+1)\right]. \tag{B.27}$$

On the other hand, according to (B.26), (B.21) and (B.23), we have

$$\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right] = \left(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\mathbf{H}(t+1)^{-1}\boldsymbol{\Lambda}\right)^{-1}\left[\mu_l\mathbf{E}_{xl}(t+1)l_t + \mathbf{E}_x(t+1)\right]. \tag{B.28}$$

Combining (B.27) and (B.28) and utilizing (B.19), we obtain

$$\mathbf{M}_t^{\text{aim}} = \left(\mathbf{H}(t) - \boldsymbol{\Lambda}\right)^{-1}\left[\rho\gamma\boldsymbol{\Sigma}_r\mathbf{x}_t^{\text{mv}} + \rho\gamma\boldsymbol{\Sigma}_{rl}l_t + \rho\left(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\mathbf{H}(t+1)^{-1}\boldsymbol{\Lambda}\right)\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right]\right] \tag{B.29}$$
$$= \left(\mathbf{H}(t) - \boldsymbol{\Lambda}\right)^{-1}\left[\rho\gamma\boldsymbol{\Sigma}_r\mathbf{x}_t^{\text{mv}} + \rho\gamma\boldsymbol{\Sigma}_{rl}l_t + \left(\mathbf{H}(t) - \boldsymbol{\Lambda} - \rho\gamma\boldsymbol{\Sigma}_r\right)\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right]\right] \tag{B.30}$$
$$= \mathbf{Z}(t)\left(\mathbf{x}_t^{\text{mv}} + \boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rl}l_t\right) + (\mathbf{I} - \mathbf{Z}(t))\,\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right], \tag{B.31}$$

where $\mathbf{Z}(t) = \rho\gamma\left(\mathbf{H}(t) - \boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Sigma}_r$.

Suppose $\boldsymbol{\Lambda} = \lambda\boldsymbol{\Sigma}_r$. We consider the ansatz for $\mathbf{E}_{xx}(t) = -a_{xx}(t)\boldsymbol{\Sigma}_r$ with $a_{xx}(t) \in \mathbb{R}$. From (B.19), we derive, for $t = T, T-1, \dots, 0$,

$$a_{xx}(t) = \lambda - \frac{\lambda^2}{\rho\gamma + \lambda + \rho a_{xx}(t+1)}, \qquad a_{xx}(T+1) = 0. \tag{B.32}$$

By employing mathematical induction, we have $a_{xx}(t) > 0$ for $t < T+1$. Therefore, the above recursive equations is well defined for all T. Then (B.29) can be simplified as

$$\mathbf{M}_t^{\text{aim}} = \frac{\gamma}{\gamma + a_{xx}(t+1)}\left(\mathbf{x}_t^{\text{mv}} + \boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rl}l_t\right) + \frac{a_{xx}(t+1)}{\gamma + a_{xx}(t+1)}\mathbb{E}_t\left[\mathbf{M}_{t+1}^{\text{aim}}\right]. \tag{B.33}$$

We write $b_{xx}(t) := \frac{a_{xx}(t)}{\gamma}$. Note that

$$\frac{\partial a_{xx}(t)}{\partial \lambda} = \frac{\rho^2(\gamma + a_{xx}(t+1))(\gamma + a_{xx}(t+1)) + \lambda^2 \rho \frac{\partial a_{xx}(t+1)}{\partial \lambda}}{(\rho\gamma + \lambda + \rho a_{xx}(t+1))^2},$$
$$\frac{\partial b_{xx}(t)}{\partial \gamma} = \frac{\rho\lambda^2 \frac{\partial b_{xx}(t+1)}{\partial \gamma} - \rho^2\lambda(1 + b_{xx}(t+1))^2}{(\rho\gamma + \lambda + \rho a_{xx}(t+1))^2}.$$

Again, by using mathematical induction, $a_{xx}(t)$ is increasing in λ and $\frac{a_{xx}(t)}{\gamma}$ is decreasing in γ. $\square$

References

A, C., Shen, Y., Zeng, Y., 2022. Dynamic asset-liability management problem in a continuous-time model with delay. International Journal of Control 95, 1315–1336.
Abou-Kandil, H., Freiling, G., Ionescu, V., Jank, G., 2003. Matrix Riccati Equations in Control and Systems Theory. Birkhäuser.
Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? The Review of Financial Studies 20, 651–707.
Basak, S., Chabakauri, G., 2010. Dynamic mean-variance asset allocation. The Review of Financial Studies 23, 2970–3016.
Bensoussan, A., Ma, G., Siu, C.C., Yam, S.C.P., 2022. Dynamic mean–variance problem with frictions. Finance and Stochastics 26, 267–300.
Berry-Stölzle, T.R., 2008a. Evaluating liquidation strategies for insurance companies. The Journal of Risk and Insurance 75, 207–230.
Berry-Stölzle, T.R., 2008b. The impact of illiquidity on the asset management of insurance companies. Insurance. Mathematics & Economics 43, 1–14.
Bertsimas, D., Lo, A.W., 1998. Optimal control of execution costs. Journal of Financial Markets 1, 1–50.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? The Review of Financial Studies 21, 1509–1531.
Chen, P., Yang, H., Yin, G., 2008. Markowitz's mean-variance asset-liability management with regime switching: a continuous-time model. Insurance. Mathematics & Economics 43, 456–465.
Chiu, M.C., Li, D., 2006. Asset and liability management under a continuous-time mean–variance optimization framework. Insurance. Mathematics & Economics 39, 330–355.
Chiu, M.C., Wong, H.Y., 2012. Mean–variance asset–liability management: cointegrated assets and insurance liability. European Journal of Operational Research 223, 785–793.
Chiu, M.C., Wong, H.Y., 2014. Mean–variance asset–liability management with asset correlation risk and insurance liabilities. Insurance. Mathematics & Economics 59, 300–310.
Choudhry, M., 2011. Bank Asset and Liability Management: Strategy, Trading, Analysis. John Wiley & Sons.
Chu, D., Ma, G., Siu, C.C., Yam, S.C.P., 2022. Robust portfolio selection with price impact. SSRN Electronic Journal. http://ssrn.com/abstract=4159192.
Collin-Dufresne, P., Daniel, K., Sağlam, M., 2020. Liquidity regimes and optimal dynamic asset allocation. Journal of Financial Economics 136, 379–406.
Demiguel, V., Martín-Utrera, A., Nogales, F.J., 2016. Parameter uncertainty in multiperiod portfolio optimization with transaction costs. Journal of Financial and Quantitative Analysis 50, 1443–1471.
Ekren, I., Muhle-Karbe, J., 2019. Portfolio choice with small temporary and transient price impact. Mathematical Finance 29.
Ferstl, R., Weissensteiner, A., 2011. Asset-liability management under time-varying investment opportunities. Journal of Banking & Finance 35, 182–192.
Gârleanu, N., Pedersen, L.H., 2013. Dynamic trading with predictable returns and transaction costs. The Journal of Finance 68, 2309–2340.
Gârleanu, N., Pedersen, L.H., 2016. Dynamic portfolio choice with frictions. Journal of Economic Theory 165, 487–516.
Gennotte, G., Jung, A., 1994. Investment strategies under transaction costs: the finite horizon case. Management Science 40, 385–404.
Glasserman, P., Xu, X., 2013. Robust portfolio control with stochastic factor dynamics. Operations Research 61, 874–893.
Han, J., Ma, G., Yam, S.C.P., 2022. Relative performance evaluation for dynamic contracts in a large competitive market. European Journal of Operational Research 302, 768–780.
Han, J., Yam, S.C.P., 2022. A probabilistic method for a class of non-Lipschitz bsdes with application to fund management. SIAM Journal on Control and Optimization 60, 1193–1222.
Isaenko, S., 2022. Transaction costs, frequent trading, and stock prices. Journal of Financial Markets 100775.
Kusy, M.I., Ziemba, W.T., 1986. A bank asset and liability management model. Operations Research 34, 356–376.
Lehalle, C.A., Neuman, E., 2019. Incorporating signals into optimal trading. Finance and Stochastics 23, 275–311.
Leippold, M., Trojani, F., Vanini, P., 2004. A geometric approach to multiperiod mean variance optimization of assets and liabilities. Journal of Economic Dynamics and Control 28, 1079–1113.
Li, D., Shen, Y., Zeng, Y., 2018. Dynamic derivative-based investment strategy for mean–variance asset–liability management with stochastic volatility. Insurance. Mathematics & Economics 78, 72–86.
Ma, G., Siu, C.C., Yam, S., Zhou, Z., 2022. Dynamic trading with Markov liquidity switching. Automatica. https://ssrn.com/abstract=4202586.
Ma, G., Siu, C.C., Zhu, S.P., 2019. Dynamic portfolio choice with return predictability and transaction costs. European Journal of Operational Research 278, 976–988.
Ma, G., Siu, C.C., Zhu, S.P., 2020a. Optimal investment and consumption with return predictability and execution costs. Economic Modelling 88, 408–419.
Ma, G., Siu, C.C., Zhu, S.P., 2022b. Portfolio choice with return predictability and small trading frictions. Economic Modelling 111, 105823.
Ma, G., Siu, C.C., Zhu, S.P., Elliott, R.J., 2020b. Optimal portfolio execution problem with stochastic price impact. Automatica 112, 108739.
Merton, R.C., 1969. Lifetime portfolio selection under uncertainty: the continuous-time case. Review of Economics and Statistics, 247–257.
Moallemi, C.C., Sağlam, M., 2017. Dynamic portfolio choice with linear rebalancing rules. Journal of Financial and Quantitative Analysis 52, 1247–1278.
Obizhaeva, A.A., Wang, J., 2013. Optimal trading strategy and supply/demand dynamics. Journal of Financial Markets 16, 1–32.
Pan, J., Xiao, Q., 2017. Optimal asset–liability management with liquidity constraints and stochastic interest rates in the expected utility framework. Journal of Computational and Applied Mathematics 317, 371–387.
Samuelson, P.A., 1969. Lifetime portfolio selection by dynamic stochastic programming. Review of Economics and Statistics 51, 239–246.
Sannikov, Y., Skrzypacz, A., 2016. Dynamic trading: price inertia and front-running. Preprint. https://web.stanford.edu/~skrz/Dynamic_Trading.pdf.
Shen, Y., Wei, J., Zhao, Q., 2020. Mean–variance asset–liability management problem under non-Markovian regime-switching models. Applied Mathematics & Optimization 81, 859–897.
Society of Actuaries, 2003. Professional actuarial specialty guide: asset-liability management. https://nexusrisk.com/docs/SOA%20ALM%20Specialty%20Guide.pdf.
Van Kervel, V., Menkveld, A.J., 2019. High-frequency trading around large institutional orders. The Journal of Finance 74, 1091–1137.
Wei, J., Wang, T., 2017. Time-consistent mean–variance asset–liability management with random coefficients. Insurance. Mathematics & Economics 77, 84–96.
Wei, J., Wong, K., Yam, S., Yung, S., 2013. Markowitz's mean–variance asset–liability management with regime switching: a time-consistent approach. Insurance. Mathematics & Economics 53, 281–291.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.
Xie, S., 2009. Continuous-time mean–variance portfolio selection with liability and regime switching. Insurance. Mathematics & Economics 45, 148–155.
Xie, S., Li, Z., Wang, S., 2008. Continuous-time portfolio selection with liability: mean–variance model and stochastic lq approach. Insurance. Mathematics & Economics 42, 943–953.
Yan, T., Wong, H.Y., 2019. Open-loop equilibrium strategy for mean–variance portfolio problem under stochastic volatility. Automatica 107, 211–223.
Yan, T., Wong, H.Y., 2020. Open-loop equilibrium reinsurance-investment strategy under mean–variance criterion with stochastic volatility. Insurance. Mathematics & Economics 90, 105–119.
Yao, H., Lai, Y., Li, Y., 2013. Continuous-time mean–variance asset–liability management with endogenous liabilities. Insurance. Mathematics & Economics 52, 6–17.
Yong, J., Zhou, X.Y., 1999. Stochastic Controls: Hamiltonian Systems and HJB Equations, vol. 43. Springer Science & Business Media.
Zhang, F., 2006. The Schur Complement and Its Applications, vol. 4. Springer Science & Business Media.
Zhang, J., Chen, P., Jin, Z., Li, S., 2020. Open-loop equilibrium strategy for mean–variance asset–liability management portfolio selection problem with debt ratio. Journal of Computational and Applied Mathematics 380, 112951.
Zhang, M., Chen, P., 2016. Mean–variance asset–liability management under constant elasticity of variance process. Insurance. Mathematics & Economics 70, 11–18.
Zhu, H.N., Zhang, C.K., Jin, Z., 2020. Continuous-time mean-variance asset-liability management with stochastic interest rates and inflation risks. Journal of Industrial and Management Optimization 16, 813–834.