A Dynamic Equilibrium Model of ETFs *

Semyon Malamud†

September 17, 2015

Abstract

I develop a dynamic general equilibrium model of exchange traded funds (ETFs) that accounts for the two-tier ETF market structure with both a centralized exchange (secondary market) and a creation/redemption mechanism (primary market) operating through market-making firms known as Authorized Participants (APs). The model is tractable and allows for any number of ETFs and basket securities. I show that the creation/redemption mechanism serves as a shock propagation channel through which temporary demand shocks may have long-lasting impacts on future prices. In particular, they may lead to a momentum in asset returns and a persistent ETF pricing gap. Improving liquidity in the primary market stimulates creation/redemption and therefore strengthens the shock propagation channel. As a result, it may amplify the volatility of both the underlying assets and the ETF pricing gap. At the same time, introducing new ETFs may reduce both the volatility and co-movement in the returns and may improve the liquidity of the underlying securities.

JEL Classification:
Keywords: exchange-traded funds, liquidity, volatility, co-movement, mis-pricing

1 Introduction

Exchange traded funds (ETFs) and other exchange traded products (ETPs) have grown tremendously over the past several years and have attracted considerable attention from investors, regulators and academics.[1] Numerous research papers have been written investigating *empirically* both the pricing of ETFs and their impact on market risk and liquidity. In particular, regulatory concerns about the hidden risks to which ETF investors are exposed and the threat that ETFs pose to market stability have been a topic of continuous debate. For example, Ramaswamy (2011) suggests that ETFs may increase systemic risks, and the U.S. Securities and Exchange Commission (SEC)

*I thank ETF Research Academy for their financial support and Guillaume de Martel and Marlene Hassine for important feedback concerning the functioning of ETF markets. I also thank Bruno Biais, Anna Cieslak, Christian Julliard, Grigory Vilkov, Hongjun Yan, Ban Zheng, and conference participants at the ESSFM 2015 in Gerzensee for helpful comments. Finally, I thank Swiss Finance Institute and the Swiss National Science Foundation for their financial support.

†Swiss Finance Institute, EPF Lausanne, and CEPR; E-mail: `semyon.malamud@epfl.ch`.

[1] As of June 2014, there were 5,217 ETFs with $2.63 trillion in total net assets (Source: BlackRock). The average annual growth rate of ETF assets since the early 1990s is 26 percent, which is two times more than that of actively managed assets.

has begun investigating the role of ETFs in increasing market volatility. Despite all these concerns, however, there is not yet a single theoretical model that could be used to analyze the impact of ETFs on the functioning of financial markets. Without such an equilibrium model, it is difficult to assess the role and interactions of all the complex mechanisms that shape the underlying market microstructure. Thus, the goal of this paper is to develop such a model.

An ETF is an investment fund traded on stock exchanges. An ETF originates with an "ETF sponsor" who chooses the fund's investment objective. A wide variety of ETFs are actively traded in modern markets, with different ETFs holding stocks, commodities, or bonds. Most ETFs are passive index tracking funds, and these have become popular investment devices because of their low costs and tax efficiency. Two major features distinguish ETFs from open-end mutual funds. First, retail investors can buy and sell ETF shares on a stock exchange through a broker-dealer, in contrast to mutual fund shares, which are not listed on stock exchanges. Second, ETF shares can only be created by large financial institutions called "Authorized Participants" (APs) that serve as broker-dealers and ETF market makers on the stock exchange and have a special agreement with the ETF sponsor. Specifically, at the end of each trading day, authorized participants can create/redeem ETF shares[2] using one or a combination of two different mechanisms: "in-kind" and "cash" transactions. In an in-kind transaction, an AP delivers the constituents of the index that the ETF holds in the exact same weights as the index and receives in exchange the corresponding number of ETF shares. In a cash transaction, ETF shares can be created/redeemed at the ETF Net Asset Value (NAV), which is the market value of the portfolio that the ETF holds.[3] The NAV is typically calculated at the close of the same day (or the open of the next day if there is a time zone difference between the ETF trading venue and that of the index constituents) after the creation/redemption orders have been placed by the APs. Both creation/redemption mechanisms serve two purposes. First, ETF sponsors effectively offer an additional primary market to APs that serves as a source of complementary liquidity and therefore improves liquidity in the secondary ETF market. Second, the creation/redemption mechanism ensures that ETF shares trade in line with the underlying NAV; if the ETF price deviates from the NAV (that is, if ETF trades at a premium or at a discount), APs can exploit this arbitrage opportunity by taking opposite positions in the ETF and the underlying basket and then offloading this inventory to the ETF sponsor at the end of the trading day. This built-in arbitrage mechanism is one of the critical ways in which ETFs differ from closed-end funds; with the latter, there is no mechanism available to control supply and demand pressures, and therefore closed-end funds often trade at massive premiums or discounts to their NAV.[4]

The above institutional details are slightly different for synthetic ETFs. In contrast to the

[2]Creation/redemption is only allowed in large blocks ("creation units"), with a typical size of about 25,000 to 200,000 shares.

[3]The in-kind transactions allow for some flexibility: each business day, ETFs are required to make available a portfolio composition file that describes the makeup of their "creation and redemption baskets" ? a specific list of names and quantities of securities or other assets designed to track the performance of the portfolio as a whole. In the case of an index-based ETF, the creation and redemption baskets are either a duplicate or a sample of the ETF portfolio (which samples the index).

[4]See, e.g., Cherkes, Sagi, and Stanton (2009). Another important difference is that closed-end funds are usually actively managed.

physical ETFs discussed above, a synthetic ETF does not hold the underlying basket but rather enters a total return swap agreement with an investment bank. In turn, the bank usually replicates the underlying basket to hedge the total return swap position. Thus, from the supply/demand point of view, the pair "ETF sponsor-investment bank" has a joint market impact that is similar to that of the sponsor himself for the case of physical ETFs. Note also that synthetic ETFs can only be created/redeemed in cash.

The presence of a primary market, the nature of the creation/redemption mechanism, and the specific role played by APs together imply that standard market microstructure models cannot be directly applied to studying ETFs. In this paper, I develop a new model that accounts for these important institutional details of the ETF markets. In my model, there are multiple "basic" securities traded in the market as well as multiple index-tracking ETFs. The market is populated by four classes of participants: ETF sponsors that provide liquidity to APs in the primary market; basic securities dealers who supply liquidity in the basic securities market and are subject to income shocks;[5] ETF clients that are subject to income shocks and demand liquidity in the ETF market; and APs that supply liquidity in the ETF market and demand liquidity in the basic securities market. In particular, in my model (as in the real world), APs perform a dual role of being both market makers and arbitrageurs. I assume that markets are segmented and only APs can trade both ETFs and the basic securities. Anecdotal evidence suggests that most large (institutional) arbitrageurs that exploit price discrepancies between ETFs and their NAV also have the status of APs.[6] I assume that information is symmetric and that trading is done purely for risk-sharing reasons.[7] When income shocks are realized, ETF clients buy ETFs, and APs are forced to absorb these shocks in the secondary market. After trading is done, APs use the creation/redemption mechanism to adjust their positions and load some of their inventory onto the ETF sponsor. This creates a "lock-in effect," whereby the shares of basic securities get locked inside the ETF, while the number of floating ETF shares increases. These endogenous fluctuations in the total number of shares available for trading generates a CAPM risk premium dynamic because the limits to arbitrage imply that aggregate risk changes over time. As a result, asset prices exhibit non-trivial dynamics, despite the fact that dividends and demand shocks are i.i.d. These dynamics are driven purely by the creation/redemption mechanism that effectively serves as a *shock propagation channel* through which temporary (i.i.d.) demand shocks influence future prices. In particular, the creation/redemption mechanism, combined with the arbitrage activity of APs, may make the markets unstable and lead to a momentum in asset returns and a persistent ETF pricing gap. The latter effect is particularly surprising given that the creation/redemption mechanism has been

[5]This assumption is made for simplicity. Standard aggregation results imply that having two classes of CARA investors with access to the same market, one class subject to shocks (say, hedgers), and one not (say, dealers), would be equivalent to a single class.

[6]For example, Ben-David, Franzoni, and Mussawi (2014) argue that "ETFs answered need of speculators (traders with relatively high turnover) for a tool that allows taking market positions at low transaction costs; ... ETF sponsors indeed testify that their clientele is composed of such institutional traders. ... ETF sponsors facilitate arbitrageurs' activity by disseminating NAV values at the fifteen-second frequency throughout the trading day."

[7]Asymmetric information can be easily incorporated into my model. This is an important direction for future research.

designed precisely to reduce fluctuations in the pricing gap.

These results imply that improving liquidity in the primary market may have an adverse effect on equilibrium dynamics, as reducing primary market trading costs stimulates creation/redemption and therefore strengthens the shock propagation channel. As a result, higher primary market liquidity may amplify the volatility of both the underlying assets and the ETF pricing gap. At the same time, introducing new ETFs may offset the negative effects of existing ETFs through a *demand substitution effect:* If the newly introduced ETFs are properly designed,[8] they may substitute for some of the demand for existing ETFs and, as result, lead to a reduction in both the volatility and co-movement in the returns. As a consequence, newly introduced ETFs may also improve liquidity of the underlying securities.

In my model, ETFs are non-redundant securities because arbitrageurs (APs) face limits to arbitrage that I model through execution risk. Specifically, I assume that APs are limited in their ability to serve as liquidity suppliers in the basic securities market and are not able and/or willing to absorb demand shocks in the basic securities market. As a result, APs face execution risk and (due to risk aversion) cannot infinitely lever their positions.[9] As a result, arbitrage is risky. Consequently, the price gap (the difference between the ETF price and its NAV) is non-zero and moves over time in response to demand shocks. In the limit, as the volatility of demand shocks goes to zero, execution risk vanishes and ETF become redundant. Yet, surprisingly, even when redundant, ETFs still have a significant impact on the prices of basic securities. The reason for this is the arbitrage activity of APs: As the execution risk vanishes, arbitrage between ETFs and their underlying basket becomes almost riskless and arbitrageurs start taking very large positions exploiting the price gap. At the same time, the price gap itself converges to zero, and the two effects offset each other and lead to a finite "correction" in basic securities prices. It turns out that, in this case, all equilibrium quantities can be computed in closed form, making the model amenable to a detailed comparative statics analysis.

In recent years, several academic and industry-oriented papers have been written regarding various "myths" and common beliefs about ETFs. I use the explicit expressions for equilibrium prices to address these issues in a rigorous theoretical framework. Here, I list several questions and claims taken from a recent paper by BlackRock (2013). [10]

- *Do ETFs increase volatility and co-movement?* This question has received particularly substantial attention in the news,[11] and many arguments have been proposed in support of the notion that ETFs increase overall systemic risk. However, I show that this is generally not

[8]The ETF design in the model is characterized by the weights of the index that the ETF replicates, as well as the primary market liquidity chosen by the sponsor.

[9]Anecdotal evidence suggests that execution risk is an important driver of the ETF price gap. For example, BlackRock (2013) states: "When this exchange is difficult (e.g., due to difficulty in locating the basket securities), ETFs tend to trade at a premium to the fair value of their underlying holdings."

[10]BlackRock is the world's largest asset manager and is the owner of iShares, the largest ETF sponsor in the world with more than 500 funds, including over 275 ETFs in the U.S., and more than $645 billion of assets under management.

[11]See, e.g., "SEC Reviewing Effects of ETFs on Volatility" by Andrew Ackerman, Wall Street Journal, 19 October 2011, and "Volatility, Thy Name is E.T.F.? by Andrew Ross Sorkin, New York Times, October 10, 2011.

true. In fact, I show that introducing new ETFs may lead to a reduction in volatility and co-movement for some assets.

- *There is concern that ETF trading substitutes for and takes away from volume and liquidity in the underlying securities.* I show that this is generally not true. In fact, introducing new ETFs may lead to an increase in liquidity in the underlying securities. This is particularly surprising given that, in my model, APs consume liquidity in the underlying securities for their hedging/arbitrage activity.

- *Costs of trading ETFs can be much lower than the costs of trading the underlying securities.* I show that this is indeed always the case when the demand shock volatility is small. Indeed, in this case, arbitrage trade between an ETF and the underlying basket is almost riskless, and, therefore, APs take a very high leverage on this trade. As a result, they are willing to absorb even very large demand shocks at a small premium, implying that equilibrium price sensitivity to shocks is small.

- *The existence of both a primary and a secondary market increases overall pricing efficiency and enhances liquidity.* I show that, indeed, the level of (endogenous) secondary market liquidity always increases in the level of (exogenous) primary market liquidity of ETFs. However, the behavior of the pricing efficiency (defined as the pricing gap) is more complicated, as discussed in the next item.

- *The relative ease with which ETF shares can be exchanged for the underlying securities in the basket determines how closely the ETF share price in the secondary market will track the fair value of its underlying holdings.* I show that this is generally not true. In fact, an increase in the primary market ETF liquidity may lead to a larger and more volatile price gap.

Overall, my results show how careful one should be in the empirical analysis of ETF returns, given the serious endogeneity problems due to general equilibrium effects. Furthermore, given the existence of thousands of ETFs and the constant growth of the ETF universe, it is important, from both the academic and the policy perspective, to study the behavior of the ETF universe as a whole. My results imply that the joint trading of multiple ETFs may dampen the shock propagation mechanism that ETFs give rise to. In particular, my results imply that regulators may potentially achieve volatility and co-movement reduction by regulating the design of newly created ETFs.

2 Literature Review

There is a growing body of empirical studies on ETFs and their effects on financial markets.

Da and Shive (2013) and Broman (2014) provide empirical evidence that ETF arbitrage activity leads to excess co-movement of returns. As a result, they argue that ETFs may actually reduce diversification. My model shows that the presence of ETFs per se may indeed increase co-movement; however, introducing new ETFs may have an offsetting effect and may actually reduce co-movement.

Staer (2014) finds a strong positive relation between daily ETF flows and the underlying stocks' returns. This finding agrees with the price pressure effect of demand shocks that is present in my model. Furthermore, Staer (2014) shows that the effect of fund flows may last for several days, in agreement with the shocks propagation channel discussed above.

Madhavan and Sobczyk (2014) and Ben-David, Franzoni, and Moussawi (2014) find evidence that ETFs are significantly more liquid than the underlying basket. In particular, the Amihud (2002) illiquidity measure is much lower for ETFs than for the underlying basket. In my model, I show that this is indeed always the case when the demand shock volatility is small. The reason for this finding is that, for their arbitrage activity, APs are willing to absorb demand shocks at a small premium, implying that equilibrium price sensitivity to demand shocks is also small.

Ben-David, Franzoni, and Moussawi (2014) show that demand shocks in the ETF market propagate through arbitrage activity to the prices of underlying securities, leading to an increase in the non-fundamental volatility.[12] My results agree with these empirical findings, showing that, in the presence of ETFs, new non-fundamental shocks make underlying securities volatility higher than it would have been in the absence of ETFs. However, I show that the effects of introducing additional ETFs on the underlying volatility and co-movement can be ambiguous: when properly designed, the growth in the ETF variety may actually reduce volatility and co-movement in the underlying securities. In a way, new ETFs partially undo the effects of "old" ETFs on market volatility.

Hamm (2014) finds that ETFs exacerbate illiquidity in stocks with high degrees of adverse selection through a demand substitution effect. In my model, information is symmetric, and hence I cannot address the issues related to adverse selection. Nonetheless, I show that ETF trading may indeed reduce liquidity in the underlying securities simply due to the increase in risk and volatility. However, introducing new ETFs may offset these effects and improve liquidity in the constituents of the underlying basket.

Petajisto (2013), Madhavan and Sobczyk (2014), and Broman (2014) find evidence for systematic ETF mis-pricing. Madhavan and Sobczyk (2014) and Broman (2014) also show that ETF mis-pricing co-moves excessively across ETFs and is strongly related to cross-sectional measures of liquidity. In my model, ETF mispricing co-moves due to the APs arbitrage activity in both the primary and secondary markets. In particular, consistent with the findings of Petajisto (2013), mis-pricing can be positively auto-correlated and hence *persistent.* Surprisingly, contrary to the conventional wisdom, price gap persistence arises despite the fact that, in my model, demand shocks are i.i.d. over time. The mechanism responsible for such a "mis-pricing momentum" is driven by the shock propagation channel of the ETF creation/redemption mechanism.

In my model, APs fulfil a dual role as liquidity providers in the ETF market and arbitrageurs who exploit deviations from the law of one price. This structure links my model to the seminal contribution by Gromb and Vayanos (2002) (see also Gromb and Vayanos (2010) for a recent survey of the literature on limits-to-arbitrage). The key differences between my model and that of Gromb and Vayanos (2002) are the nature of limits-to-arbitrage (execution risk in the former

[12]See Marshall, Nguyen, and Visaltanachoti (2010) and Richie, Daigler, and Gleason (2008) for evidence on ETF arbitrage activity.

versus collateral constraints in the latter) and the underlying market microstructure (the existence of a primary market for ETFs). In addition, I allow for any number of assets and ETFs. To solve the model, I use the techniques from Vayanos (1999, 2001) that are based on small shock volatility approximations. I model liquidity supply and demand through demand schedules (as in Kyle, 1989, and Vayanos, 1999) and market orders (as in Kyle, 1985 and Vayanos, 2001) respectively.

Similar to the present paper, Cespa and Foucault (2014) also study endogenous liquidity with multiple investor classes and market fragmentation. However, in their model liquidity is, to a large extent, driven by asymmetric information. They show how asymmetric information implies that illiquidity in one asset can spill over into the illiquidity of another asset through a cross-asset learning effect and may lead to multiple equilibria. Thus, a large drop in the liquidity of both assets may occur if the economy jumps from one equilibrium to another. In my model, information is symmetric, but markets are segmented, as in Cespa and Foucault (2014). Thus, APs' arbitrage activity serves as the only source of liquidity spillover. In particular, ETF (exogenous) liquidity in the primary market spills over to the liquidity in the underlying securities. The effect can be non-monotonic: An increase in primary ETF liquidity can lead to a drop in the liquidity of underlying securities. To the best of my knowledge, this liquidity spillover effect is novel to my paper.

In my model, the limits-to-arbitrage are directly linked to the idiosyncratic volatility. This links my model to the paper by Stambaugh, Yu, and Yuan (2014) who show that idiosyncratic volatility represents risk that deters arbitrage and may lead to mis-pricing and the emergence of the idiosyncratic volatility puzzle (i.e., the empirically observed negative link between idiosyncratic volatility and returns).

In my paper, ETF sponsors and APs are institutional investors whose behavior influences asset prices. This relates my paper to a growing body of literature on the effect of institutional trading on asset prices. Several papers find evidence that institutional trading may amplify volatility and correlations (see, e.g., Greenwood and Thesmar, 2011; Anton and Polk, 2014). Ramaswamy (2010) and Wurgler (2010) suggest that index-linked investing may distort stock prices and risk-return tradeoffs and that these effects may have strong consequences for the real side of the economy. In my model, I show that ETF trading and arbitrage activity by APs indeed distorts underlying asset prices and that these distortions move over time in a systematic way in response to the creation/redemption activities by APs. However, these effects may be welfare-improving and lead to a more efficient equilibrium risk sharing. The overall effects of ETF trading on the real economy may be non-trivial; thus, understanding these effects is an important direction for future research.

There is also a growing theoretical literature on the effect of indexing on asset prices. Basak and Pavlova (2013a, 2013b) and Buffa, Vayanos and Woolley (2014) show theoretically that the presence of agents whose performance is benchmarked against an index leads to excess co-movement and volatility. In their models, the existence of tradable index funds would have no impact on the model behavior. By contrast, in my model, the existence of ETFs has a significant impact on equilibrium dynamics *even when ETFs are redundant.* The mechanism responsible for this effect is based on the APs' arbitrage activity. When the risk of arbitrage vanishes, APs start making very large bets to arbitrage the deviations from the law of one price; when multiplied by the small price gap, the two effects offset each other, and the remaining impact is finite and non-zero. Chabakauri and Rytchkov (2014) study how index investing affects capital market equilibrium in an economy with two Lucas trees and one index (a portfolio of the two trees). They find that indexing can either increase or decrease the correlation between stock returns and that it also decreases market volatility. Such non-monotonicities are also present in my model.

Vayanos and Woolley (2013) show how investor flows between investment funds triggered by changes in fund managers' efficiency may lead to momentum, reversal, and co-movement. My results imply that ETF flows may have a similar effect on equilibrium prices, even when ETFs are redundant and are managed efficiently.

Finally, several papers discuss the market structure of leveraged and inverse ETFs (see, e.g., Madhavan and Cheng (2009) and Jiang and Yan (2015)). They show that the structure of these funds impacts market volatility and liquidity and can lead to value destruction for a buy-and-hold investor. Investigating leveraged ETFs in a theoretical setting is thus an important direction for future research.

3 Model Setup

Time is discrete, $t = 0, 1, \cdots$. There are N "basic" securities, indexed by $i = 1, \cdots, N$ that pay the vector of dividends $d_t \sim N(\bar{d}, \Sigma_d)$ with $\bar{d} \in \mathbb{R}^N$ and a positive definite matrix $\Sigma_d \in \mathbb{R}^{N \times N}$. Dividends are i.i.d. over time. There is also a riskless asset with interest rate r in a perfectly elastic supply. I denote by $p_t = (p_{t,i})_{i=1}^N$ the time-t vector of basic securities prices. In addition to basic securities, there are L ETFs. ETF number m is supposed to replicate a basket (an index) of the underlying securities with the dividend vector $f_m^T d_t$ where $f_m = (f_{m,i})_{i=1}^N$ is the vector of ETF portfolio weights. I denote by $\mathbf{F} = (f_{m,i})_{m,i=1}^{L,N} \in \mathbb{R}^{L \times N}$ the matrix of ETF basket weights. An agent holding one unit of the m-th ETF receives dividends $f_m^T d_t$ at time t.[13] The vector of ETF dividends can be written as

$$D_t = \mathbf{F}\, d_t\,. \tag{1}$$

I will use Var_t to denote the conditional variance-covariance matrix of a random vector. In this notation, $\mathrm{Var}_t[d_{t+1}] = \Sigma_d$ whereas $\mathrm{Var}_t[D_{t+1}] = \mathbf{F}\Sigma_d\mathbf{F}^T$. I also define the (vector of) ETF NAV as

$$\mathrm{NAV}_t = \mathbf{F} p_t\,.$$

The vector of time-t ETF prices is denoted by P_t. The difference between the ETF prices and the NAV, given by $P_t - \mathbf{F}p_t$, is called the ETF mis-pricing or the *pricing gap.*

The market is populated by four classes of agents who differ in their roles: ETF sponsors, authorized participants (APs), basic securities dealers, and ETF clients. ETF sponsors offer liquidity in the *primary ETF market.* Namely, at the end of each trading day, at a time denoted by $t+ \in (t, t+1)$, each authorized participant has the right to contact an ETF sponsor to cre-

[13] ETFs are known for their tax efficiency. In particular, dividends are typically reinvested directly and are not paid out. I leave these important considerations for future research.

ate/redeem ETF shares. The creation/redemption mechanism operates in two forms: "in-kind" transactions and "cash" transactions.

In an in-kind transaction, an authorized participant has the right to exchange any number of ETF shares for the replicating basket of basic securities:[14] that is, an AP has the right to exchange a portfolio $Z_m(f_{m,i})_{i=1}^N$ of basic securities for Z_m shares of the m-th ETF. In the vector form, we can see that a portfolio $\mathbf{F}^T Z \in \mathbb{R}^N$ of basic securities can be exchanged for the portfolio $Z \in \mathbb{R}^L$ of ETFs. Here, $\mathbf{F}^T$ denotes the transpose of the matrix $\mathbf{F}$.

In a cash transaction, an AP can acquire Z_m shares of m-th ETF at its NAV, *calculated in the next trading round*; that is, at time $t+1$. This is an important friction in the primary ETF market: at the time of a cash transaction, an AP does not know the exact price at which the transaction will be executed.[15]

The creation/redemption mechanism of the ETF primary market is a major characteristic of ETFs that makes them different from other financial instruments. The liquidity provided in the primary market serves as a natural arbitrage mechanism that ensures that ETF prices stay sufficiently close to their NAV. The degree of primary market liquidity is determined by the fees (bid-ask spreads) that the ETF sponsors charge to APs. I assume that these fees are quadratic in the number of ETF shares. Specifically, a creation/redemption of Z_m shares of ETF m costs $0.5\lambda_{I,m}Z_m^2$ in an In-kind transaction, and it costs $0.5\lambda_{C,m}Z_m^2$ in a Cash transaction.[16] With this notation, transacting a vector of Z ETF shares in the primary market costs $0.5Z_I^T\Lambda_I Z_I$ for an in-kind transaction of a vector of Z_I shares, while the cost is $0.5Z^T\Lambda_C Z$ for a cash transaction of Z_C shares. Here, $\Lambda_I = \text{diag}(\lambda_{I,m}),\ \Lambda_C = \text{diag}(\Lambda_{C,m})$ denotes the diagonal matrix of transaction costs.

I assume that all ETFs are physical (securities-based). This means that an ETF sponsor physically holds the basket of securities that replicate the ETF. To this end, after receiving the total (both in-kind and cash) creation/redemptions at time $t+$, the corresponding ETF sponsor goes to the primary market and acquires the necessary basket.[17] Note that the in-kind transactions do

[14] In reality, the "in-kind" transaction rules are slightly more flexible and allow some freedom in the choice of the replicating basket. Instead of exactly replicating the ETF, the basket may contain fewer securities or securities with characteristics that are very similar to those of the securities in the underlying basket. The exact analysis of these covenants is beyond the scope of these paper.

[15] Creations/redemptions are often made shortly before the 4:00 p.m. close of a given trading day and are often executed at the closing market prices for securities traded in the same time zone as the ETF. However, for international securities, due to differences in time zones, transactions are usually executed on the morning of the next trading day. Naturally, the so-called slippage (i.e., unanticipated price change) is larger in such transactions. In my model, a "trading day" can be viewed as a combination of the evening and morning trading rounds. This way, the model partially accounts for the possible time zone differences. It would be interesting to introduce multiple (at least two) intra-day trading rounds, accounting for pre-creation and post-creation periods. I leave this direction for future research.

[16] The assumption of quadratic transaction costs is made for tractability. See, for example, Garleanu and Pedersen (2013). However, many real-world ETFs have transaction costs that increase with the number of shares. For example, ProShares 30 Year TIPS/TSY Spread ETF has the following comment on its fee structure: "For orders of $15 million or more, the Advisor may charge, in its sole discretion, a variable Create/Redeem fee between 0.00% and 0.50%." Similar covenants apply to most of their ETFs. See http://www.proshares.com/resources/creation_and_redemption_fees.html. Fees are typically higher for cash transactions because they need to cover replication costs.

[17] The largest ETF provider, iShares, uses the physical replication method. However, many providers also use the

not require any action from the sponsor, as the underlying securities have already been delivered. However, a cash transaction of Z_C shares forces the ETF sponsor to acquire $\mathbf{F}^T Z_C$ shares of underlying securities in the corresponding market. However, in both cases, the total replicating basket (i.e., $\mathbf{F}^T(Z_I + Z_C)$ shares of basic securities) gets *"locked" inside the ETF sponsor holdings.* This "lock-in" mechanism implies that ETF creations/redemptions change the total supply of available securities. This is *the ETF demand shock propagation channel.*

I now come to describing other market participants. I assume that the basic securities market is populated by dealers who provide (supply) liquidity in this market. Dealers maximize constant absolute risk aversion (CARA) preferences given by

$$E\left[\sum_{t=0}^{\infty} -e^{-\beta t - \alpha_D c_t}\right]. \tag{2}$$

Here, β is their discount rate, c_t is their time-t consumption, and α_D is their absolute risk aversion. While I assume that this is a single class of traders in this market, standard results imply that we can view this class as an aggregate of several classes of agents. I assume that dealers do not trade ETFs and only use basket securities to hedge their positions. Some of the basket securities can be, for example, futures contracts that dealers typically use for hedging purposes. I assume that dealers are subject to income shocks $\varepsilon_t^T d_{t+1}$ where $\varepsilon_t \sim N(\delta\bar{\varepsilon}, \delta\Sigma_\varepsilon)$ are i.i.d. across time. Here, δ is a parameter that measures the overall magnitude of income shock volatility. Then, given their time-t portfolio x_t^D of basic securities, dealers' money market account M_t^D evolves according to

$$M_{t+1}^D \;=\; (M_t^D - c_t - (x_t^D)^T p_t)\, e^r \;+\; \varepsilon_t^T d_{t+1} \;+\; (x_t^D)^T (p_{t+1} + d_{t+1})\,. \tag{3}$$

Income shocks $\varepsilon_t d_{t+1}$ play an important role in my model, as they lead to endogenous supply/demand shocks in the underlying securities without the need to introduce noise traders. This makes my model amenable to welfare analysis.

In contrast to dealers, authorized participants can trade in both markets (ETFs and basic securities) and have CARA preferences given by

$$E\left[\sum_{t=0}^{\infty} -e^{-\beta t - \alpha c_t}\right] \tag{4}$$

where α is their risk aversion. The timeline describing their actions is more involved than that for other classes of agents due to their participation in the primary ETF market. Namely, at time t,

so-called synthetic replication method, whereby an ETF sponsor enters into a total return swap agreement with an investment bank. In this total return swap transaction, the investment bank delivers the return on the underlying basket in exchange for a fee. In such a synthetic ETF, the sponsor does not need to replicate the underlying basket. However, the investment bank would naturally perform the same function and go to the underlying market to hedge the total return swap position by physically replicating (a part of) the basket. Of course, in contrast to physical ETFs, this replication does not have to be perfect, and the degree of hedging depends on the risk aversion of the investment bank. However, the same endogenous supply mechanism is present even for synthetic ETFs. We could view the physical ETF case as a limit of the synthetic ETF case when the risk aversion of the investment bank becomes infinitely large.

APs trade in the ETF and basic securities markets and end up with portfolio holdings $x_t^A \in \mathbb{R}^N$ in the basic securities and portfolio holdings $y_t^A \in \mathbb{R}^M$ in ETFs. Then, at time $t+$, they contact ETF sponsors and create/redeem $Z_{I,t+}$ and $Z_{C,t+}$ ETF shares through in-kind and cash transactions, respectively, and pay the corresponding fee

$$0.5(Z_{I,t+}^T \Lambda_I Z_{I,t+} + Z_{C,t+}^T \Lambda_C Z_{C,t+})\,.$$

After these transactions, they end up with

$$x_{t+}^A \;=\; x_t^A \;-\; \mathbf{F}^T Z_{I,t+} \tag{5}$$

shares of basic securities shares and with

$$y_{t+}^A \;=\; y_t^A \;+\; Z_{I,t+} \;+\; Z_{C,t+} \tag{6}$$

ETF shares. Here, we can see the difference between in-kind and cash transactions: Since in-kind transactions require delivery of underlying securities, they change APs' security holdings. At the same time, the effect of primary market transactions on ETF holdings is independent of the choice of the creation/redemption mechanism. At time $t+1$, APs receive dividends of $(y_t^A)^T \mathbf{F} d_{t+1}$ and pay the realized NAV, $Z_{C,t+}^T \mathbf{F} p_{t+1}$ of the ETF shares they acquired at time $t+$ through the cash transaction. Summarizing, we can see that the money market account M_t^A of APs evolves according to

$$\begin{aligned} M_{t+1}^A \;=\;& \Big(M_t^A - c_t - (x_t^A)^T p_t - (y_t^A)^T P_t - 0.5(Z_{I,t+}^T \Lambda_I Z_{I,t+} + Z_{C,t+}^T \Lambda_C Z_{C,t+})\Big)\, e^r \\ &+\; (x_{t+}^A)^T (p_{t+1} + d_{t+1}) \;+\; (y_{t+}^A)^T (P_{t+1} + \mathbf{F} d_{t+1}) - Z_{C,t+}^T \mathbf{F} p_{t+1} \\ =\;& \Big(M_t^A - c_t - (x_t^A)^T p_t - (y_t^A)^T P_t - 0.5(Z_{I,t+}^T \Lambda_I Z_{I,t+} + Z_{C,t+}^T \Lambda_C Z_{C,t+})\Big)\, e^r \\ &+ (x_t^A + \mathbf{F}^T (y_t^A + Z_{C,t+}))^T d_{t+1} \;+\; (y_t^A)^T P_{t+1} \;+\; (x_t^A)^T p_{t+1} \;+\; (Z_{I,t+} + Z_{C,t+})^T (P_{t+1} - \mathbf{F} p_{t+1})\,. \end{aligned} \tag{7}$$

Formula (7) shows that the two creation/redemption mechanisms differ in their impact on the nature of risk to which the APs are exposed. Namely, an in-kind transaction has no impact on the exposure to dividend risk, because a loss of $\mathbf{F}^T Z_{I,t+}$ basic securities dividends is compensated by an exactly offsetting gain in ETF dividends. By contrast, both creation/redemption mechanisms contribute equally to exposure to the price gap $P_{t+1} - \mathbf{F} p_{t+1}$. In particular, when ETFs are redundant (so that the price gap is identically zero), in-kind transactions are also redundant.

The last class of market participants is that of ETF clients. I assume that they can only trade ETFs, do not participate in the market for basic securities, and do not have access to the primary market. They maximize

$$E\left[\sum_{t=0}^{\infty} -e^{-\beta t - \alpha_E c_t}\right]\,, \tag{8}$$

where α_E is their risk aversion. I assume that ETF clients are subject to income shocks $\xi_t^T d_{t+1}$

where $\xi_t \sim N(\bar{\xi}, \delta\Sigma_\xi)$ are i.i.d. across time and are independent of dealer income shocks ε_t. Then, given their time-t portfolio y_t^E of ETFs, their money market account M_t^E evolves according to

$$M_{t+1}^E \;=\; (M_t^E - c_t - (y_t^E)^T P_t)\, e^r \;+\; \xi_t^T d_{t+1} \;+\; (y_t^E)^T (P_{t+1} + \mathbf{F} d_{t+1})\,. \tag{9}$$

Income shocks $\xi_t d_{t+1}$ play the role of clientele in my model: depending on their structure, there may be different ETFs that better suit the clients' hedging needs. For example, an ETF with portfolio weights equal to $\bar{\xi}$ would be the "best average hedge" that would allow ETF clients to eliminate average risk of their income shocks. The assumption that ETFs' clients use ETFs only for hedging purposes is stylized. One can view the ξ_t shocks as a reduced form of demand fluctuations that may have origins related to beliefs, capital flows, or tax efficiency considerations. The assumption of endogenous demand shocks allows me to study general equilibrium effects of ETFs on social welfare.

4 Equilibrium

Up to this point, except for the presence of the primary market for ETFs, the structure of my model resembles that of Gromb and Vayanos (2002). As in their model, there are redundant securities that are traded in segmented markets, and there are arbitrageurs (APs in my model) that can trade in both markets and arbitrage deviations from the law of one price. In Gromb and Vayanos (2002), perfect arbitrage is not possible because arbitrageurs are financially constrained and have to separately collateralize their positions in each of the securities. In my model, I follow a different route: I assume that APs face execution risk in the underlying securities.[18] To model this execution risk (or, equivalently, arbitrage risk), I assume that APs cannot supply liquidity both in the ETF market and in the basic securities markets. Thus, as in the real world, they serve as liquidity providers (market makers) in the ETF market, while they demand (consume) liquidity in the basic securities market to hedge their ETF exposures and perform their arbitrage trades. Formally, I assume that APs trade through limit orders in the ETF market, while they can only submit market orders in the basic securities market.

This model of risky arbitrage could also be viewed as a reduced form of financial constraint Indeed, in the real world most APs cannot serve as market makers both in the ETF and in the basic securities markets. Being a market maker often means holdings large inventories of the underlying assets and being willing to absorb the corresponding supply shocks. This can be quite costly in terms of margin and collateral requirements. As is common in the market microstructure literature, I model liquidity supply by assuming that APs submit price-contingent demand schedules, $\mathcal{S}^A(P_t)$

[18] For example, execution may be subject to significant delay, shares may not be available for short selling, or the costs of short selling (stock lending fees) may be high. Execution risk is also closely related to the liquidity of the underlying market: exercising a large trade immediately is typically impossible, and execution has to be spread over a significant time period to reduce price impact. Such a delayed execution is obviously associated with a significant execution risk. These liquidity considerations are important even for very liquid securities such as foreign exchange instruments. See Mancini, Ranaldo, and Wrimpelmeyer (2013).

in the ETF market, whereas they submit market orders in the basic securities market.[19] The latter assumption implies that the number of shares of basic securities that APs acquire is independent of the realization of ε_t shocks. In agreement with the above discussion, this assumption means that APs only consume liquidity in basic securities and are not willing to act as absorbers of these (security-specific) shocks, leaving this role to security dealers who submit price-contingent demand schedules in the basic securities market. Thus, while APs make markets in ETFs and absorb demand shocks ξ_t of ETF clients, they face arbitrage risk when hedging their ETF positions in the basic securities markets. In agreement with the findings of Stambaugh, Yu, and Yuan (2014), the riskiness of arbitrage is directly linked to idiosyncratic volatility.

In equilibrium, prices are determined by market clearing. Absent sunspots, standard intertemporal CAPM logic implies that only aggregate supply matters for pricing; hence, the only relevant state variables are the total supply of the basic securities and the total supply of ETFs. Given that the portfolio $\mathbf{F}^T Z_{C,t+}$ of basic securities is absorbed by ETF sponsors at time $t+1$, the aggregate supply available for sharing between the other three classes of market participants is given by

$$\begin{aligned} \bar{x}_{t+} &= \bar{x}_{t-1+} - \mathbf{F}^T(Z_{I,t+} + Z_{C,t+}) \\ \bar{y}_{t+} &= \bar{y}_{t-1+} + Z_{I,t+} + Z_{C,t+}\,. \end{aligned} \tag{10}$$

Equation (10) implies that the aggregate supply always satisfies the following *invariance law:*

$$\bar{x}_{t+} + \mathbf{F}^T \bar{y}_{t+} \;=\; \bar{x}_{t-1+} + \mathbf{F}^T \bar{y}_{t-1+}\,. \tag{11}$$

Formula (11) means that the "effective supply" of basic securities, $\bar{x}_{t+} + \mathbf{F}^T \bar{y}_{t+}$, remains constant over time because the basket of underlying securities simply moves from being freely traded to being traded inside an ETF.

I will denote the vector of these state variables by $\bar{\mathcal{X}}_{t+} = \binom{\bar{y}_{t+}}{\bar{x}_{t+}}$. Then, the market operates as follows. At time t, APs supply liquidity in the ETF market by submitting a demand schedule $y_t^A(P_t) = \mathcal{S}^A(P_t, \bar{\mathcal{X}}_{t-1+})$; at the same time, APs demand (consume) liquidity in the basic securities market by submitting market orders $x_t^A = \mathcal{D}^A(P_t, \bar{\mathcal{X}}_{t-1+})$. Similarly, basic securities dealers observe the ε_t shock and then submit demand schedules $x_t^D(p_t) = \mathcal{S}^D(p_t, \bar{\mathcal{X}}_{t-1+}, \varepsilon_t)$, while ETF clients submit market orders $y_t^E = \mathcal{D}^E(\bar{\mathcal{X}}_{t-1+}, \xi_t)$ in the ETF markets. Prices (p_t, P_t) are then determined by market clearing:

$$\begin{aligned} \bar{x}_{t-1+} &= x_t^A + x_t^D(p_t) \\ \bar{y}_{t-1+} &= y_t^A(P_t) + y_t^E\,. \end{aligned} \tag{12}$$

After the time-t trading round, the optimal choice of creation/redemption policies $Z_{I,t+}, Z_{C,t+}$ is made at time $t+$ between the trading rounds taking place at times t and $t+1$. Naturally, these policies depend not only on the aggregate supply but also on the holdings x_t^A, y_t^A that the APs have accumulated during the day. Due to their market-making role, APs have to supply liquidity and absorb a part of the demand shocks ξ_t in the ETF market; they use the creation/redemption mechanism to "correct" their inventories in the desired direction. Thus, we expect these policies

[19] This assumption is often used in the literature. See, e.g., Kyle (1985) and Vayanos (2001).

$Z_{I,t+}, Z_{C,t+}$ to be a function of all four state variables

$$\mathcal{X}_t \;\equiv\; \begin{pmatrix} y_t^A \\ x_t^A \\ y_t^E \\ x_t^D \end{pmatrix}. \tag{13}$$

I can now state a formal definition of a Markov perfect equilibrium.

Definition 4.1 *A Markov perfect linear stationary equilibrium is characterized by prices and strategies such that*

- *prices of ETFs and the underlying securities are given by*

$$\begin{aligned} P_t &= \bar{P} \;+\; \Pi_X^E \bar{\mathcal{X}}_{t-1+} \;+\; \Pi_\xi^E \xi_t \\ p_t &= \bar{p} \;+\; \Pi_X^D \bar{\mathcal{X}}_{t-1+} \;+\; \Pi_\xi^D \xi_t \;+\; \Pi_\varepsilon^D \varepsilon_t \end{aligned} \tag{14}$$

  with some matrices $\Pi_X^E \in \mathbb{R}^{L\times(L+N)}$, $\Pi_\xi^E \in \mathbb{R}^{L\times N}$, $\Pi_X^D \in \mathbb{R}^{N\times(L+N)}$, $\Pi_\xi^D \in \mathbb{R}^{N\times N}$, *and vectors* $\bar{P} \in \mathbb{R}^L$, $\bar{p} \in \mathbb{R}^N$;

- *the vector of total supply* $\bar{\mathcal{X}}_{t+}$ *follows a linear evolution rule*

$$\bar{\mathcal{X}}_{t+} \;=\; \bar{X}^* \;+\; \bar{A}_{\mathcal{X}}\, \bar{\mathcal{X}}_{t-1+} \;+\; \bar{A}_\xi\, \xi_t \tag{15}$$

  with some matrices $\bar{A}_{\mathcal{X}} \in \mathbb{R}^{(L+N)\times(L+N)}$, $\bar{A}_\xi \in \mathbb{R}^{(L+N)\times N}$;

- *the optimal strategies for the market participants are such that*

  - *liquidity supply by APs in the ETF market is given by*

$$\mathcal{S}^A(P_t, \bar{\mathcal{X}}_{t-1+}) \;=\; \widetilde{C}_A^E + \widetilde{\Theta}_{\mathcal{X}}^A \bar{\mathcal{X}}_{t-1+} \;-\; \widetilde{\mathbf{B}}_A P_t \tag{16}$$

    with some matrices $\widetilde{\mathbf{B}}_A \in \mathbb{R}^{L\times L}$, $\widetilde{\Theta}_{\mathcal{X}}^A \in \mathbb{R}^{L\times(L+N)}$ *and a vector* $\widetilde{C}_A^E \in \mathbb{R}^L$;

  - *liquidity demand by APs in the market for underlying securities is given by*

$$\mathcal{D}^A(P_t, \bar{\mathcal{X}}_{t-1+}) \;=\; \widetilde{C}_A^D + \widetilde{\theta}_{\mathcal{X}}^A \bar{\mathcal{X}}_{t-1+} + \widetilde{\theta}_\xi^A P_t \tag{17}$$

    with some matrices $\widetilde{\theta}_{\mathcal{X}}^A \in \mathbb{R}^{N\times(L+N)}$, $\widetilde{\theta}_\xi^A \in \mathbb{R}^{N\times N}$, *and a vector* $\widetilde{C}_A^D \in \mathbb{R}^N$;

  - *the creation/redemption rule for the APs is given by*

$$\begin{pmatrix} Z_{I,t+} \\ Z_{C,t+} \end{pmatrix} \;=\; \begin{pmatrix} \bar{Z}_I^0 \\ \bar{Z}_C^0 \end{pmatrix} \;+\; \begin{pmatrix} \mathcal{Z}_I \\ \mathcal{Z}_C \end{pmatrix} \mathcal{X}_t \tag{18}$$

    with some matrices $\mathcal{Z}_I, \mathcal{Z}_C \in \mathbb{R}^{L\times(L+N)}$ *and vectors* $\bar{Z}_I^0, \bar{Z}_C^0 \in \mathbb{R}^L$;

– *liquidity supply by dealers in the market for basic securities is given by*

$$\mathcal{S}^D(p_t, \bar{\mathcal{X}}_{t-1+}, \varepsilon_t) \;=\; \widetilde{Q}^*_D \;+\; \widetilde{\Theta}^D_{\mathcal{X}} \bar{\mathcal{X}}_{t-1+} \;+\; \widetilde{\Theta}^D_{\varepsilon} \varepsilon_t \;-\; \widetilde{\mathbf{B}}_D p_t \tag{19}$$

with some matrices $\widetilde{\Theta}^D_{\mathcal{X}} \in \mathbb{R}^{N\times(L+N)}$, $\widetilde{\Theta}_{\varepsilon} \in \mathbb{R}^{N\times N}$ *and a vector* $\widetilde{Q}^*_D \in \mathbb{R}^N$;

– *liquidity demand by customers in the ETF markets is given by*

$$\mathcal{D}^E(\bar{\mathcal{X}}_{t-1+}, \xi_t) \;=\; Q^*_E \;+\; \Theta^E_{\mathcal{X}} \bar{\mathcal{X}}_{t-1+} \;+\; \Theta^E_{\xi} \xi_t \tag{20}$$

with some matrices $\Theta^E_{\mathcal{X}} \in \mathbb{R}^{L\times(L+N)}$, $\Theta^E_{\xi} \in \mathbb{R}^{L\times N}$;

- *both the ETF and the underlying securities markets clear:*

$$\begin{aligned} \mathcal{S}^D(p_t, \bar{\mathcal{X}}_{t-1+}, \varepsilon_t) \;+\; \mathcal{D}^A(P_t, \bar{\mathcal{X}}_{t-1+}) \;&=\; \bar{x}_{t-1+} \\ \mathcal{S}^A(P_t, \bar{\mathcal{X}}_{t-1+}) \;+\; \mathcal{D}^E(\bar{\mathcal{X}}_{t-1+}, P_t) \;&=\; \bar{y}_{t-1+}\,. \end{aligned} \tag{21}$$

I will always refer to a Markov perfect linear stationary equilibrium as simply "an equilibrium." In order to solve for an equilibrium, we first need to understand the role of demand shocks ξ_t and ε_t. I start by noting that, in the above equilibrium definition, ε_t shocks do not enter the evolution of state variables. This is a direct consequence of the assumption that APs do not provide liquidity in basic securities and hence do not absorb idiosyncratic shocks. As a result, ε_t realization only influences the prices at which APs can hedge their ETF positions in the basic securities market; it has no impact on the APs' asset holdings. Thus, neither $\mathcal{X}$ nor $\bar{\mathcal{X}}$ depends on the ε_t realization.

By contrast, ETF demand shocks ξ_t play a major role in equilibrium dynamics; since APs are forced to absorb these shocks during the day, they offload their ETF inventory to the ETF sponsor through the primary market mechanism. This way, creation/redemption policies $Z_{I,t+}$, $Z_{C,t+}$ depend on ξ_t realization and hence (by (10)) so does $\bar{\mathcal{X}}_{t+} = \binom{\bar{y}_{t+}}{\bar{x}_{t+}}$. This is precisely the shock propagation channel of ETFs through which transitory (i.i.d.) demand shocks ξ_t have an impact on future prices.

The next important observation is that, in order to make optimal time-t equilibrium trading decisions, it suffices for the agents to know the realization of ξ_t.[20] If they know ξ_t, the agents do not need to condition their demand on prices, because ξ_t contains all the relevant information about time t and time $t+$ equilibrium asset holdings, given that the agents know $\bar{\mathcal{X}}_{t-1+}$. In particular, the assumption that dealers in the basic securities market do not condition their demand on ETF prices is without loss of generality if the dealers can recover the relevant part of ξ_t from basic securities prices p_t. Making use of this observation, I proceed as follows. First, I assume that the shock ξ_t is publicly observable at time t before the trade starts; then, I solve for the optimal equilibrium policies as a function of ξ_t. Second, I verify that the relevant part of ξ_t can be recovered from prices by all market participants. This implies that an equilibrium with observable ξ_t is equivalent to an equilibrium with price-contingent demand and supply, as described in Definition 4.1.

[20] In fact, as I show below, only the "part" of the shock given by $\mathbf{F}\Sigma_d\xi_t$ appears in the demand of ETF clients, and hence it is only this part of ξ_t that matters for equilibrium prices.

The next important question concerns the definition of equilibrium illiquidity. In linear rational expectation models with noise traders, illiquidity is commonly defined as the sensitivity of prices to noisy demand and supply shocks.[21] The price contingent demand schedule (19) incorporates both the hedging demand and the effect of recovering the shock ξ_t from prices. To distinguish the pure supply shock effect, I consider the (ex-post equivalent) situation where the dealers observe ξ_t directly. Then, the demand schedule can be rewritten as

$$\mathcal{S}^D(p_t, \bar{\mathcal{X}}_{t-1+}, \xi_t, \varepsilon_t) \;=\; Q_D^* \;+\; \Theta_{\mathcal{X}}^D \bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\xi^D \xi_t + \Theta_\varepsilon^D \varepsilon_t \;-\; \mathbf{B}_D p_t \tag{22}$$

with some coefficients $Q_D^*, \Theta_{\mathcal{X}}^D, \Theta_\xi^D, \Theta_\varepsilon^D, \mathbf{B}_D$. Then, $\hat{\varepsilon}_t = \Theta_\varepsilon^D \varepsilon_t$ is the "pure" demand shock, and the basic securities illiquidity is given by $\mathbf{B}_D^{-1}$. A similar result holds for the ETF market: the slope $\mathbf{B}_A$ of the APs' demand schedule already accounts for the fact that APs recover the shocks ξ_t from prices. If they could observe supply shocks directly,[22] they could condition their demand schedules on ξ_t and submit

$$\mathcal{S}^A(P_t, \bar{\mathcal{X}}_{t-1+}, \xi_t) \;=\; C_A^E + \Theta_{\mathcal{X}}^A \bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\xi^A \xi_t - \; \mathbf{B}_A P_t\,, \tag{23}$$

with some coefficients $C_A^E, \Theta_{\mathcal{X}}^A, \Theta_\xi^A, \mathbf{B}_A$, where $\mathbf{B}_A^{-1}$ is the ETF illiquidity that is "free from learning the shocks from prices." In the sequel, I will always use $\mathbf{B}_D$ and $\mathbf{B}_A$ as measures of equilibrium liquidity.

Note that equilibrium dynamics (15) makes prices potentially prone to bubbles. Denote by $\mathrm{eig}(\mathcal{A})$ the set of eigenvalues of a matrix $\mathcal{A}$. If $\max(|\,\mathrm{eig}(\mathcal{A}_{\mathcal{X}})|) > 1$, equilibrium supply process $\mathcal{X}_{t+}$ is non-stationary. As a result, price level and volatility have the potential to grow over time. In order to ensure the absence of bubbles, we need to check that the transversality condition $e^{-rt}E[P_t] \;\to\; 0$ holds at infinity. This is equivalent to requiring that $\max(|\,\mathrm{eig}(\mathcal{A}_{\mathcal{X}})|) < e^r$. Absent this condition, agents might be able to attain infinite utility by building Ponzi-like schemes and earning infinite profits in the long run. As I will show below, ETF trading indeed often makes the dynamics non-stationary (i.e., we often have $\max(|\,\mathrm{eig}(\mathcal{A}_{\mathcal{X}})|) > 1$). As I show below, a stronger "no-bubble" condition $\max(|\,\mathrm{eig}(\mathcal{A}_{\mathcal{X}})|) < e^{r/2}$ is sufficient to prevent Ponzi schemes and ensure equilibrium existence.

5 Individual Optimization and Equilibrium Dynamics

To solve for an equilibrium, I follow the standard approach in the literature on linear rational expectations equilibria. Conditional on their beliefs about equilibrium dynamics, agents choose utility-maximizing policies. This determines their optimal demand which (by market clearing) in turn determines the dynamics of equilibrium prices and state variables. As I explained above, I will work directly with the case in which the price-relevant part of the shock ξ_t is publicly observable. As I show below, only the part of the shock given by $\mathbf{F}\Sigma_d \xi_t$ appears in the demand of ETF clients,

[21] See, e.g., Cespa and Foucault (2014).

[22] This is often the case for market makers who can directly observe clients' order flow.

and hence it is only this part of ξ_t that matters for equilibrium prices.

We start with the optimization problem of dealers. Denote by $V^D(M_t^D, \bar{\mathcal{X}}_{t-1+}, \varepsilon_t, \xi_t)$ their time-t value function and let M_t^D be their money-market account (see (3)). Following the standard approach (see Vayanos, 1999, 2001), I conjecture that the value function is of the form

$$V^D(M_t^D, \bar{\mathcal{X}}_{t-1+}, \varepsilon_t, \xi_t) \;=\; e^{-\alpha_D((1-e^{-r})M_t^D + \bar{v}^D + \mathcal{V}^D\cdot(\varepsilon_t,\xi_t,\bar{\mathcal{X}}_{t-1+}) - 0.5(\xi_t,\varepsilon_t,\bar{\mathcal{X}}_{t-1+})^T\mathcal{W}^D(\varepsilon_t,\xi_t,\bar{\mathcal{X}}_{t-1+}))} \tag{24}$$

for some matrix $\mathcal{W}^D \in \mathbb{R}^{(3N+L)\times(3N+L)}$, a vector $\mathcal{V}^D \in \mathbb{R}^{3N+L}$, and a constant $\bar{v}^D \in \mathbb{R}$. Substituting this Ansatz into the dynamic programming equation and performing some (quite tedious) calculations, I arrive at the following result.[23]

Proposition 5.1 *Suppose that* $\max(|\operatorname{eig}(\bar{A}_{\mathcal{X}})|) < e^{r/2}$. *Then, the optimal demand schedule of securities dealers is given by*

$$\mathcal{S}^D(p_t) \;=\; Q_D^* \;+\; \Theta_{\mathcal{X}}^D \bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\xi^D \xi_t \;+\; \Theta_\varepsilon^D \varepsilon_t \;-\; \mathbf{B}_D p_t\,,$$

where $\zeta_t = \binom{\varepsilon_t}{\xi_t}$ *is the vector of demand shocks and*

$$\begin{aligned}
\mathbf{B}_D^{-1} &= e^{-r}\alpha_D(1-e^{-r})\Big(\Sigma_d + \delta\Pi_\zeta^D(\Sigma_\zeta^{-1} - \delta\alpha_D\mathcal{W}_\zeta^D)^{-1}(\Pi_\zeta^D)^T\Big)\\
Q_D^* &= e^{-r}\mathbf{B}_D\left(\bar{d} \;+\; \bar{p} \;+\; \Pi_X^D\bar{\mathcal{X}}^* + \Pi_\zeta^D(\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}\left(\Sigma_\zeta^{-1}\bar{\zeta} - \alpha_D\left(\mathcal{V}_\zeta^D + \mathcal{W}_{\zeta,\mathcal{X}}^D\bar{\mathcal{X}}^*\right)\right)\right)\\
\Theta_{\mathcal{X}}^D &= e^{-r}\mathbf{B}_D\Big(\Pi_X^D - \delta\Pi_\zeta^D(\Sigma_\zeta^{-1} - \delta\alpha_D\mathcal{W}_\zeta^D)^{-1}\alpha_D\mathcal{W}_{\zeta,\mathcal{X}}^D\Big)\bar{A}_{\mathcal{X}}\\
\Theta_\xi^D &= e^{-r}\mathbf{B}_D\Big(\Pi_X^D - \delta\Pi_\zeta^D(\Sigma_\zeta^{-1} - \delta\alpha_D\mathcal{W}_\zeta^D)^{-1}\alpha_D\mathcal{W}_{\zeta,\mathcal{X}}^D\Big)\bar{A}_\xi\\
\Theta_\varepsilon^D &= -\alpha_D(1-e^{-r})e^{-r}\mathbf{B}_D\Sigma_d\,,
\end{aligned} \tag{25}$$

with $\Pi_\zeta^D = (\Pi_\varepsilon^D\ \Pi_\xi^D)$. *The coefficients* $\mathcal{W}^D$ *and* $\mathcal{V}^D$ *satisfy the dynamic programming equations*

$$\begin{aligned}
e^r\mathcal{W}^D &= -(\Theta^D - \mathbf{B}_D\Pi^D)^T(e^{-r}\mathbf{B}_D)^{-1}(\Theta^D - \mathbf{B}_D\Pi^D) + \begin{pmatrix}0 & 0\\ 0 & (\bar{\mathcal{A}})^T\mathcal{W}_{\mathcal{X}}^D\bar{\mathcal{A}}\end{pmatrix} + \alpha_D(1-e^{-r})^2\begin{pmatrix}\Sigma_d & 0\\ 0 & 0\end{pmatrix}\\
e^r\mathcal{V}^D &= \begin{pmatrix}(1-e^{-r})\bar{d}\\ 0\end{pmatrix} \;+\; \bar{\mathcal{A}}^T\left(\mathcal{V}_{\mathcal{X}}^D - \mathcal{W}_{\mathcal{X}}^D\bar{\mathcal{X}}^* + (\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}(\Sigma_\zeta^{-1}\bar{\zeta} - \alpha_D(\mathcal{V}_\zeta^D + \mathcal{W}_{\zeta,\mathcal{X}}^D\bar{X}^*))\right)\\
&\quad + (\Theta^D - e^r\mathbf{B}_D\Pi^D)^T(e^{-r}\mathbf{B}_D)^{-1}(Q_D^* - \mathbf{B}_D\bar{p}).
\end{aligned} \tag{26}$$

Proposition 5.1 characterizes dealers' optimal policy as a function of price dynamics and the dynamics of equilibrium state variables $\bar{\mathcal{X}}_{t-1+}$. The slope $\mathbf{B}_D$ is inversely proportional to the risk aversion α_D times the effective riskiness of a dealer's exposure to securities returns. Absent future utility contributions captured by $\mathcal{W}_\zeta^D$, this is simply the conditional variance-covariance matrix of

[23] As usual, I assume that the discount rate β is sufficiently large to ensure the validity of the transversality condition.

asset payoffs, given by

$$\mathrm{Var}_t[p_{t+1} + d_{t+1}] \;=\; \mathrm{Cov}_t[d_{t+1} + \Pi^D_\zeta \zeta_{t+1}] \;=\; \Sigma_d + \delta \Pi^D_\zeta \Sigma_\zeta (\Pi^D_\zeta)^T \,.$$

However, since the agents are non-myopic, they take into account the co-movement of asset returns with future state variables (and hence also with future returns). These anticipated changes in the investment opportunity set lead to a hedging demand that effectively modifies the risk perception of dealers from $\mathrm{Var}_t[p_{t+1} + d_{t+1}]$ to $\Sigma_d + \delta\Pi^D_\zeta(\Sigma_\zeta^{-1} - \delta\alpha_D \mathcal{W}^D_\zeta)^{-1}(\Pi^D_\zeta)^T$.[24] The coefficients $\Theta^D_{\mathcal{X}}$ and Θ^D_ζ can be interpreted similarly. Here, I only discuss $\Theta^D_\xi, \Theta^D_\varepsilon$. Absent creation/redemption, asset supply stays constant and does not respond to ETF demand shocks, and hence $\bar{\mathcal{A}}_\xi = 0$. In this case, $\Theta^D_\xi = 0$ and dealers' demand only depends on ε_t and is proportional to dividend riskiness Σ_d. However, creation/redemption activity influences the amount of aggregate risk and causes total asset supply to respond to demand shocks. In particular, $\bar{\mathcal{A}}_\xi$ is non-zero, and a time-t demand shock ξ_t changes future basic securities' prices by $\Pi^D_X \bar{\mathcal{A}}_\xi \xi_t$. Dealers choose the optimal exposure to this expected price change, and this exposure is adjusted to take into account covariances of returns with future state variables; this adjustment is captured by $\mathcal{W}^D$.

The optimization problem of ETF clients can be solved similarly. I conjecture that their value function is given by

$$V^E(M^E_t, \bar{\mathcal{X}}_{t-1+}, \xi_t) \;=\; e^{-\alpha_E((1-e^{-r})M^E_t + \bar{v}^E + \mathcal{V}^E \cdot (\xi_t, \bar{\mathcal{X}}_{t-1+}) - 0.5(\xi_t, \bar{\mathcal{X}}_{t-1+})^T \mathcal{W}^E (\xi_t, \bar{\mathcal{X}}_{t-1+}))} \tag{27}$$

for some matrix $\mathcal{W}^E \in \mathbb{R}^{(2N+L)\times(2N+L)}$, a vector $\mathcal{V}^E \in \mathbb{R}^{2N+L}$, and a constant $\bar{v}^E \in \mathbb{R}$. Here, M^E_t is their money market account, the dynamics of which is given by (9). Substituting this Ansatz into the dynamic programming equation, we arrive at the following result.[25]

Proposition 5.2 *Suppose that* $\max(|\operatorname{eig}(\bar{\mathcal{A}}_{\mathcal{X}})|) < e^{r/2}$. *Then, the optimal demand of ETF clients is given by*

$$\mathcal{D}^E(\bar{\mathcal{X}}_{t-1+}, \xi_t) \;=\; Q^*_E \;+\; \Theta^E_{\mathcal{X}} \bar{\mathcal{X}}_{t-1+} \;+\; \Theta^E_\xi \xi_t \,, \tag{28}$$

where

$$\begin{aligned}
Q^*_E \;&=\; \mathcal{S}_E\Bigg(\mathbf{F}\bar{d} - e^r \bar{P} \;+\; \Pi^E_X \bar{\mathcal{X}}^* + \Pi^E_\xi(\Sigma_\xi^{-1} - \alpha_E \mathcal{W}^E_\xi)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \alpha_E\Bigg(\mathcal{V}^E_\xi + \mathcal{W}^E_{\xi,\mathcal{X}}\bar{\mathcal{X}}^*\Bigg)\Bigg)\Bigg) \\
\Theta^E_{\mathcal{X}} \;&=\; \mathcal{S}_E\Big(\Pi^E_X - \Pi^E_\xi(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}^E_\xi)^{-1}\alpha_E\mathcal{W}^E_{\xi,\mathcal{X}}\Big)\bar{\mathcal{A}}_{\mathcal{X}} - \mathcal{S}_E e^r \Pi^E_X \\
\Theta^E_\xi \;&=\; \mathcal{S}_E\Big(\Pi^E_X - \Pi^E_\xi(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}^E_\xi)^{-1}\alpha_E\mathcal{W}^E_{\xi,\mathcal{X}}\Big)\bar{\mathcal{A}}_\xi \; - \mathcal{S}_E(e^r\Pi^E_\xi + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\,,
\end{aligned} \tag{29}$$

with

$$\mathcal{S}_E^{-1} = (1-e^{-r})\alpha_E\Big(\mathbf{F}\Sigma_d\mathbf{F}^T + \Pi^E_\xi(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}^E_\xi)^{-1}(\Pi^E_\xi)^T\Big)\,.$$

[24] Note that the difference between the two is proportional to δ and is therefore small when δ is small.

[25] As usual, I assume that the discount rate β is sufficiently large to ensure the validity of the transversality condition.

The coefficients $\mathcal{W}^E$ and $\mathcal{V}^E$ satisfy the dynamic programming equations

$$
\begin{aligned}
e^r \mathcal{W}^E &= -(\Theta^E)^T (\mathcal{S}^E)^{-1} \Theta^E + (\bar{\mathcal{A}})^T \mathcal{W}^E_{\mathcal{X}} \bar{\mathcal{A}} + \alpha_E (1-e^{-r})^2 \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_d \end{pmatrix} \\
e^r \mathcal{V}^E &= \begin{pmatrix} 0 \\ (1-e^{-r}) \mathbf{F} \bar{d} \end{pmatrix} + \bar{\mathcal{A}}^T \left(\mathcal{V}^E_{\mathcal{X}} - \mathcal{W}^E_{\mathcal{X}} \bar{\mathcal{X}}^* + (\Sigma_\xi^{-1} - \alpha_E \mathcal{W}^E_\xi)^{-1} (\Sigma_\xi^{-1} \bar{\xi} - \alpha_E (\mathcal{V}^E_\xi + \mathcal{W}^E_{\xi,\mathcal{X}} \bar{X}^*)) \right) \\
&+ (\Theta^E)^T \mathcal{S}_E^{-1} Q_E^* .
\end{aligned} \tag{30}
$$

The expressions for the optimal policy of an ETF client are similar to those of a basic securities dealer. The main differences are: (i) the variance-covariance matrix Σ_d is replaced by the ETF dividends' variance-covariance matrix $\mathbf{F}\Sigma_d\mathbf{F}^T$, and (ii) the demand schedule is not price-contingent. Note, however, that the latter assumption in fact has no impact on equilibrium behavior; since ETF clients know the realization of ξ_t, they know the exact value of ETF prices and hence are indifferent between submitting a market order or a limit order. Nevertheless, this assumption is important for calculating ETF market liquidity. If ETF clients submitted limit orders, they would be supplying liquidity to the ETF market.[26] This would be counterfactual: Anecdotal evidence suggests that most ETF liquidity suppliers are APs.

It remains to characterize the optimal policies of APs. I conjecture that their value function is of the form

$$
V(M_t, \bar{\mathcal{X}}_{t-1+}, \xi_t) = e^{-\alpha_E((1-e^{-r})M_t + \bar{v} + \mathcal{V}\cdot(\xi_t, \bar{\mathcal{X}}_{t-1+}) - 0.5(\xi_t, \bar{\mathcal{X}}_{t-1+})^T \mathcal{W}(\xi_t, \bar{\mathcal{X}}_{t-1+}))} \tag{31}
$$

for some matrix $\mathcal{W} \in \mathbb{R}^{(2N+L)\times(2N+L)}$, a vector $\mathcal{V} \in \mathbb{R}^{2N+L}$, and a constant $\bar{v} \in \mathbb{R}$. However, the dynamic programming equation for APs is more complicated than that for other classes of agents because of an intermediate optimization problem that they need to solve at time $t+$ when they trade in the primary ETF market through the creation/redemption mechanism. At that intermediate trading round, APs need to have rational expectations about the dynamics of equilibrium state variables. Namely, APs rationally anticipate that $\bar{\mathcal{X}}_{t+}$ is linearly related to the vector

$$
\mathcal{X}_t \equiv \begin{pmatrix} y_t^A \\ x_t^A \\ y_t^E \\ x_t^D \end{pmatrix} \tag{32}
$$

of post-trade time-t asset holdings: In a linear equilibrium, there exists a vector $X_0^+ \in \mathbb{R}^{L+N}$ and a matrix $\mathcal{B} \in \mathbb{R}^{(L+N)\times(2N+2L)}$ such that

$$
\bar{\mathcal{X}}_{t+} = X_0^+ + \mathcal{B}\mathcal{X}_t .
$$

[26] In this case, total liquidity would have to be defined as the sum of the slopes of clients' and APs' demand schedules.

In fact, substituting (18) and (12) into (10), we can see that $X_0^+ = \binom{\bar{Z}_C^0+\bar{Z}_I^0}{-\mathbf{F}^T(\bar{Z}_C^0+\bar{Z}_I^0)}$, whereas[27]

$$\mathcal{B} \;=\; \begin{pmatrix} \mathrm{Id} & 0 & \mathrm{Id} & 0 \\ 0 & \mathrm{Id} & 0 & \mathrm{Id} \end{pmatrix} \;+\; \begin{pmatrix} \mathcal{Z}_I + \mathcal{Z}_C \\ -\mathbf{F}^T(\mathcal{Z}_I + \mathcal{Z}_C) \end{pmatrix}. \tag{33}$$

Here, $\mathcal{Z}_I, \mathcal{Z}_C, \bar{Z}_C^0, \bar{Z}_I^0$ are the coefficients of the equilibrium creation/redemption policy. Note that, conditional on the post-trade allocation $\mathcal{X}_t$, the realization of past shock ξ_t is irrelevant for APs, as its effects have already been incorporated into the allocation.

Now, recalling the dynamics of the money market account,

$$\begin{aligned} M_{t+1}^A \;=\; & \Big(M_t^A - c_t - (x_t^A)^T p_t - (y_t^A)^T P_t - 0.5(Z_{I,t+}^T \Lambda_I Z_{I,t+} + Z_{C,t+}^T \Lambda_C Z_{C,t+})\Big)\, e^r \\ & + (x_t^A + \mathbf{F}^T(y_t^A + Z_{C,t+}))^T d_{t+1} \;+\; (y_t^A)^T P_{t+1} \;+\; (x_t^A)^T p_{t+1} \;+\; (Z_{I,t+} + Z_{C,t+})^T (P_{t+1} - \mathbf{F} p_{t+1})\,, \end{aligned} \tag{34}$$

we see that creation/redemption exposes an AP to two types of risk: (1) the *mis-pricing risk,* captured by the exposure $\Pi_\xi^E - \mathbf{F}\Pi_\xi^D$ of the pricing gap $P_{t+1} - \mathbf{F}p_{t+1}$ to ETF demand shocks ξ_{t+1} and the exposure $-\mathbf{F}\Pi_\varepsilon^D$ to basic securities demand shocks ε_{t+1} and (2) the dividend risk that depends only on the cash transaction $Z_{C,t+}$. As a result, the total effective cost of creation/redemption is given by the matrix

$$\mathcal{R}_Z \;=\; \begin{pmatrix} (e^r-1)\Lambda_I & 0 \\ 0 & (e^r-1)\Lambda_C + \alpha(1-e^{-r})^2 \mathbf{F}\Sigma_d \mathbf{F}^T \end{pmatrix} \;+\; \alpha(1-e^{-r})^2 \begin{pmatrix} \widehat{\Pi} & \widehat{\Pi} \\ \widehat{\Pi} & \widehat{\Pi} \end{pmatrix} \tag{35}$$

with

$$\widehat{\Pi} \;=\; \mathrm{Var}_t[P_{t+1} - \mathbf{F}p_{t+1}] \;=\; \mathbf{F}\Pi_\varepsilon^D \Sigma_\varepsilon (\mathbf{F}\Pi_\varepsilon^D)^T + (\Pi_\xi^E - \mathbf{F}\Pi_\xi^D)(\Sigma_\xi^{-1} - \alpha \mathcal{W}_\xi)^{-1} (\Pi_\xi^E - \mathbf{F}\Pi_\xi^D)^T\,.$$

The first term in (35) is simply the quadratic creation/redemption cost plus the additional dividend risk acquired through a cash transaction and given by $\alpha(1-e^{-r})^2 \mathrm{Var}_t[\mathbf{F}d_{t+1}] \;=\; \alpha(1-e^{-r})^2 \mathbf{F}\Sigma_d \mathbf{F}^T$. The second term is the cost of the additional exposure to mis-pricing risk that comprises the exposures to ETF demand shocks ξ_{t+1} and basic securities demand shocks ε_{t+1}.

Denote by $\mathcal{G}(\mathcal{X}_t, y, x) \;=\; \binom{\mathcal{G}_I}{\mathcal{G}_C}(\mathcal{X}_t, y, x)$ the vector of expected utility gains from creating one unit of an ETF conditional on APs' time t holdings of ETFs and basic securities being given by y and x respectively; here, $\mathcal{G}_I$ and $\mathcal{G}_C$ are the expected gains from in-kind and cash transactions,

[27] Note that here $\mathcal{Z}_I, \mathcal{Z}_C$ have the same dimensions as the matrix (Id 0 Id 0).

respectively.[28] It is possible to show[29] that

$$\mathcal{G}(\mathcal{X}_t, y, x) = (1 - e^{-r})\begin{pmatrix} E_t[P_{t+1} - \mathbf{F}p_{t+1}] + g(\mathcal{X}_t, y, x) \\ E_t[P_{t+1} - \mathbf{F}p_{t+1}] + g(\mathcal{X}_t, y, x) \end{pmatrix} + \ (1 - e^{-r})\begin{pmatrix} 0 \\ \mathbf{F}\bar{d} - \alpha(1 - e^{-r})\mathbf{F}\Sigma_d(x + \mathbf{F}^T y) \end{pmatrix}. \tag{36}$$

Here, the expected mis-pricing is given by

$$E_t[P_{t+1} - \mathbf{F}p_{t+1}] \ = \ (\bar{P} - \mathbf{F}\bar{p}) + (\Pi_X^E - \mathbf{F}\Pi_X^D)(X_0^+ + \mathcal{B}\mathcal{X}_t) - \delta\mathbf{F}\Pi_\varepsilon^D\bar{\varepsilon} \ + \ (\Pi_\xi^E - \mathbf{F}\Pi_\xi^D)\bar{\xi}, \tag{37}$$

whereas $\mathbf{F}\bar{d} - \alpha(1 - e^{-r})\mathbf{F}\Sigma_d(x + \mathbf{F}^T y)$ represents expected dividends gained from the cash transaction net of the marginal utility loss from the existing exposure $x + \mathbf{F}^T y$ to dividend risk. Finally, $g(\mathcal{X}_t, y, x)$ is a hedging term that reflects marginal utility loss from the existing exposure to mispricing risk. It is given by

$$g(\mathcal{X}_t, y, x) \ = \ (\Pi_\xi^E - \mathbf{F}\Pi_\xi^D)\Bigg(- \bar{\xi} + (\Sigma_\xi^{-1} - \delta\alpha\mathcal{W}_\xi)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \delta\alpha\Bigg(\mathcal{V}_\xi + (1 - e^{-r})\Big[(\Pi_\xi^D)^T x + (\Pi_\xi^E)^T y\Big] - \mathcal{W}_{\xi,\mathcal{X}}\Big(X_0^+ \ + \ \mathcal{B}\mathcal{X}_t\Big)\Bigg)\Bigg)\Bigg) + \delta\alpha(1 - e^{-r})^2\mathbf{F}\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T x\,. \tag{38}$$

Here, the term $\alpha(1-e^{-r})^2\mathbf{F}\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T x$ is the marginal utility loss from the exposure x to shocks ε_{t+1}, while the first term is the marginal utility loss from the exposure to ETF demand shock ξ_{t+1}, accounting for covariances of asset returns with future state variables that are captured by $\mathcal{W}$ and $\mathcal{V}$. Given the above expressions, APs choose a creation/redemption policy to achieve the optimal risk-return tradeoff given the marginal risk $\mathcal{R}_Z$ and the expected payoff $\mathcal{G}$. This gives

$$\begin{pmatrix} Z_{I,t+} \\ Z_{C,t+} \end{pmatrix} \ = \ \mathcal{R}_Z^{-1}\,\mathcal{G}(\mathcal{X}_t, y, x)\,. \tag{39}$$

Furthermore, the total expected utility gain is given by $0.5(\mathcal{G}(\mathcal{X}_t, y, x))^T\mathcal{R}_Z^{-1}\mathcal{G}(\mathcal{X}_t, y, x)$.

The just-computed optimal creation/redemption policy at time $t+$ depends on the asset holdings x, y achieved by trading at time t. We can now proceed by backward induction to find the optimal holdings $y_t^A,\ x_t^A$, of the APs in the ETF and basic securities markets, respectively. To this end, we have to identify the risk-return tradeoff that the agents face at time t. The total risk that the APs face consists of several components. Absent the primary market, the risk is given by the variance of (additive) asset returns plus a component responsible for fluctuations in the investment

[28]Note that $\mathcal{X}_t$ contains asset holdings y_t^A, x_t^A of APs. However, it order to solve the optimization problem of APs at time t, we need to compute their utility for any holdings (y, x). Since all APs behave identically, in equilibrium we will have $x = x_t^A,\ y = y_t^A$.

[29]See the Appendix for details of this calculation.

opportunity set:

$$
\begin{aligned}
\mathcal{R}_{[yx]} \;&=\; \alpha(1-e^{-r})^2 \mathrm{Var}_{t-1+,\xi_t}\left[\begin{pmatrix} P_{t+1} + \mathbf{F} d_{t+1} - e^r P_t \\ p_{t+1} + d_{t+1} - e^r p_t \end{pmatrix}\right] \\
&+ \delta\alpha(1-e^{-r})^2 \begin{pmatrix} \Pi^E_\xi \\ \Pi^D_\xi \end{pmatrix} ((\Sigma_\xi^{-1} - \delta\alpha \mathcal{W}_\xi)^{-1} - \Sigma_\xi) \begin{pmatrix} \Pi^E_\xi \\ \Pi^D_\xi \end{pmatrix}^T
\end{aligned}
\tag{40}
$$

Here,

$$
\begin{aligned}
&\mathrm{Var}_{t-1+,\xi_t}\left[\begin{pmatrix} P_{t+1} + \mathbf{F} d_{t+1} - e^r P_t \\ p_{t+1} + d_{t+1} - e^r p_t \end{pmatrix}\right] \\
&= \begin{pmatrix} \mathbf{F}\Sigma_d \mathbf{F}^T & \mathbf{F}\Sigma_d \\ \Sigma_d \mathbf{F}^T & \Sigma_d + (1+e^{2r})\Pi^D_\varepsilon \Sigma_\varepsilon (\Pi^D_\varepsilon)^T \end{pmatrix} + \begin{pmatrix} \Pi^E_\xi \\ \Pi^D_\xi \end{pmatrix} \Sigma_\xi \begin{pmatrix} \Pi^E_\xi \\ \Pi^D_\xi \end{pmatrix}^T .
\end{aligned}
\tag{41}
$$

is the variance of asset returns given the information that APs have at time t before the trade occurs. Here, I have used the fact that

$$
\mathrm{Var}_{t-1+,\xi_t}\left[\begin{pmatrix} P_{t+1} \\ p_{t+1} \end{pmatrix}\right] \;=\; \begin{pmatrix} \Pi^E_\xi \\ \Pi^D_\xi \end{pmatrix} \Sigma_\xi \begin{pmatrix} \Pi^E_\xi \\ \Pi^D_\xi \end{pmatrix}^T + \begin{pmatrix} 0 & 0 \\ 0 & \Pi^D_\varepsilon \Sigma_\varepsilon (\Pi^D_\varepsilon)^T \end{pmatrix} .
$$

The coefficient $(1+e^{2r})$ in front of $\Pi^D_\varepsilon \Sigma_\varepsilon (\Pi^D_\varepsilon)^T$ reflects the fact that, due to the execution risk at time t, APs are exposed twice to the idiosyncratic demand shocks of basic securities, through the shocks ε_t and ε_{t+1}. The total idiosyncratic risk is thus given by

$$
\mathrm{Var}_{t-1+,\xi_t}[\Pi^D_\varepsilon \varepsilon_{t+1} - e^r \Pi^D_\varepsilon \varepsilon_t] \;=\; (1+e^{2r})\Pi^D_\varepsilon \Sigma_\varepsilon (\Pi^D_\varepsilon)^T .
$$

The last term in (40) is a hedging component that arises due to co-movement of returns with asset supply; the latter is captured by the matrix $\mathcal{W}_\xi$. This hedging component is driven by the difference $((\Sigma_\xi^{-1} - \delta\alpha\mathcal{W}_\xi)^{-1} - \Sigma_\xi)$ and by the sensitivities Π^E_ξ, Π^D_ξ of prices to demand shocks, and is identically zero if $\mathcal{W}_\xi = 0$.

In the presence of the primary market, some of this risk acquired by APs at the time-t trading round can be offloaded to the ETF sponsor who effectively serves as a liquidity provider for APs. Due to this additional trading opportunity, the effective risk that APs face is smaller than that described in (40) and is given by

$$
\mathcal{R}_{\mathcal{G}} \;=\; \mathcal{R}_{[yx]} - \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix} \mathcal{R}_Z^{-1} (\mathcal{G}_y \; \mathcal{G}_x) , \tag{42}
$$

where $\mathcal{G}_y$ and $\mathcal{G}_x$ are the sensitivities of the primary market marginal utility gains $\mathcal{G}(\mathcal{X}_t, y, x)$ to y and x respectively:

$$
\mathcal{G}_y = \frac{\partial}{\partial y}\mathcal{G}(\mathcal{X}_t, y, x),\;\; \mathcal{G}_x = \frac{\partial}{\partial x}\mathcal{G}(\mathcal{X}_t, y, x) .
$$

Indeed, this follows from the fact that the quadratic part of the utility from holding (y, x) is given by

$$-\begin{pmatrix} y \\ x \end{pmatrix}^T \mathcal{R}_{[yx]} \begin{pmatrix} y \\ x \end{pmatrix} + (\mathcal{G}(\mathcal{X}_t, y, x))^T \mathcal{R}_Z^{-1} \mathcal{G}(\mathcal{X}_t, y, x).$$

In order to compute the optimal position $(y, x) = (y_t^A, x_t^A)$, we need to know the corresponding risk/return tradeoff. To this end, I write down the ETF creation gains $\mathcal{G}$ (see (36)) in terms of their sensitivities $\mathcal{G}_{\mathcal{X}}, \mathcal{G}_x, \mathcal{G}_y$ to $\mathcal{X}_t$ (equilibrium portfolio holdings of all three classes of agents); the given AP's ETF holdings y; and the given AP's basic securities holdings x :

$$\mathcal{G}(\mathcal{X}_t, y, x) \;=\; \mathcal{G}_0 + \mathcal{G}_{\mathcal{X}} \mathcal{X}_t + \mathcal{G}_y y + \mathcal{G}_x x\,. \tag{43}$$

Then, the expected gains from primary market trading per unit of (y, x) are given by

$$\begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix} \mathcal{R}_Z^{-1} (\mathcal{G}_0 + \mathcal{G}_{\mathcal{X}} \mathcal{X}_t)\,,$$

and hence the total expected gains are given by

$$\begin{aligned} &(1-e^{-r}) \begin{pmatrix} E_{t-1+,\xi_t}[P_{t+1} + \mathbf{F} d_{t+1} - e^r P_t] \\ E_{t-1+,\xi_t}[p_{t+1} + d_{t+1} - e^r p_t] \end{pmatrix} \\ &+ \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix} \mathcal{R}_Z^{-1} (\mathcal{G}_0 + \mathcal{G}_{\mathcal{X}} \mathcal{X}_t) \\ &+ (1-e^{-r}) \begin{pmatrix} \Pi_\xi^E \\ \Pi_\xi^D \end{pmatrix} \left(-\bar{\xi} + (\Sigma_\xi^{-1} - \delta\alpha \mathcal{W}_\xi)^{-1} \left(\Sigma_\xi^{-1} \bar{\xi} - \delta\alpha \left(\mathcal{V}_\xi - \mathcal{W}_{\xi,\mathcal{X}} \bar{\mathcal{X}}_{t+} \right) \right) \right), \end{aligned} \tag{44}$$

where the last term is the hedging component due to anticipated changes in the investment opportunity set. Here,

$$\begin{aligned} &\begin{pmatrix} E_{t-1+,\xi_t}[P_{t+1} + \mathbf{F} d_{t+1} - e^r P_t] \\ E_{t-1+,\xi_t}[p_{t+1} + d_{t+1} - e^r p_t] \end{pmatrix} \\ &= \begin{pmatrix} \bar{P} \;+\; \Pi_X^E(\bar{X}^* + \bar{\mathcal{A}}_{\mathcal{X}} \bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi \xi_t) + \Pi_\xi^E \bar{\xi} + \mathbf{F}\bar{d} - e^r P_t \\ \bar{p}(1-e^r) + \Pi_{\mathcal{X}}^D(\bar{X}^* + (\bar{\mathcal{A}}_{\mathcal{X}} - e^r \mathrm{Id}) \bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi \xi_t) + \Pi_\xi^E(\bar{\xi} - e^r \xi_t) + \Pi_\varepsilon^D \bar{\varepsilon}(1-e^r) + \bar{d} \end{pmatrix}. \end{aligned} \tag{45}$$

To complete the calculation, we need to express optimal policies (y, x) in terms of the state variables $\bar{\mathcal{X}}_{t-1+}, \xi_t$ known to the APs prior to the time-t trading round. To this end, we need to rewrite the vector of post-trade holdings, $\mathcal{X}_t$, in terms of these variables. In a linear equilibrium, we have

$$\mathcal{X}_t \;=\; X_0^- + \bar{\mathcal{A}}_{\mathcal{X}}^- \bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi \xi_t \tag{46}$$

for some vector $X_0^- \in \mathbb{R}^{(2N+2L)\times(2N+2L)}$ and matrices $\bar{\mathcal{A}}_{\mathcal{X}}^- \in \mathbb{R}^{(2N+2L)\times(N+L)}$, $\bar{\mathcal{A}}_\xi^- \in \mathbb{R}^{(2N+2L)\times N}$.[30] Summarizing, we can see that there exists a vector $\mathcal{G}_0^* \in \mathbb{R}^{L+N}$ and a matrix $\mathcal{G}^+ \in \mathbb{R}^{(L+N)\times(L+2N)}$

[30]The coefficients X_0^-, $\bar{\mathcal{A}}^-$ can be computed by substituting the agents' optimal policies and using market clearing. See the Appendix for details.

such that[31]

$$
\begin{aligned}
\mathcal{G}_0^* + \mathcal{G}^+\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix} - \begin{pmatrix}(e^r-1)P_t\\ 0\end{pmatrix} \;&=\; (1-e^{-r})\begin{pmatrix}E_{t-1+,\xi_t}[P_{t+1}+\mathbf{F}d_{t+1}-e^rP_t]\\ E_{t-1+,\xi_t}[p_{t+1}+d_{t+1}-e^rp_t]\end{pmatrix}\\
+\;\begin{pmatrix}\mathcal{G}_y^T\\ \mathcal{G}_x^T\end{pmatrix}\mathcal{R}_Z^{-1}(\mathcal{G}_0+\mathcal{G}_{\mathcal{X}}\mathcal{X}_t)\\
+\,(1-e^{-r})\begin{pmatrix}\Pi_\xi^E\\ \Pi_\xi^D\end{pmatrix}\Bigg(-\bar{\xi}+(\Sigma_\xi^{-1}-\delta\alpha\mathcal{W}_\xi)^{-1}\Big(\Sigma_\xi^{-1}\bar{\xi}-\delta\alpha\Big(\mathcal{V}_\xi-\mathcal{W}_{\xi,\mathcal{X}}\bar{\mathcal{X}}_{t+}\Big)\Big)\Bigg).
\end{aligned}
\tag{47}
$$

Thus, we arrive at the following result.[32]

Proposition 5.3 *Suppose that* $\max(|\operatorname{eig}(\bar{\mathcal{A}}_{\mathcal{X}})|) < e^{r/2}$. *Then, the APs' optimization problem has a unique solution given by*

$$
\begin{pmatrix}y_t^A\\ x_t^A\end{pmatrix} \;=\; \mathcal{R}_{\mathcal{G}}^{-1}\left(\mathcal{G}_0^*+\mathcal{G}^+\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}-\begin{pmatrix}(e^r-1)P_t\\ 0\end{pmatrix}\right), \tag{48}
$$

while the optimal creation/redemption policy is given by (39). *In particular, optimal equilibrium policies can be written as*

$$
\begin{aligned}
y_t^A \;&=\; C_A^E + \Theta_{\mathcal{X}}^A\bar{\mathcal{X}}_{t-1+} + \Theta_\xi^A\xi_t \;-\; \mathbf{B}_AP_t\\
x_t^A \;&=\; C_A^D + \theta_\xi^A\xi_t + \theta_{\mathcal{X}}^A\bar{\mathcal{X}}_{t-1+}\,,
\end{aligned}
\tag{49}
$$

with

$$
\begin{pmatrix}C_A^E\\ C_A^D\end{pmatrix} = \mathcal{R}_{\mathcal{G}}^{-1}\mathcal{G}_0^*,\; \begin{pmatrix}\Theta_{\mathcal{X}}^A & \Theta_\xi^A\\ \theta_{\mathcal{X}}^A & \theta_\xi^A\end{pmatrix} \;=\; \mathcal{R}_{\mathcal{G}}^{-1}\mathcal{G}^+\;,\; \mathbf{B}_A \;=\; (e^r-1)(\mathcal{R}_{\mathcal{G}}^{-1})_y\,. \tag{50}
$$

The coefficients $\mathcal{V}$, $\mathcal{W}$ *satisfy dynamic programming equations*

$$
\begin{aligned}
\mathcal{W} \;&=\; e^{-r}\left(\bar{\mathcal{A}}^T(\mathcal{W}_{\mathcal{X}}+\alpha(\mathcal{W}_{\xi,\mathcal{X}}\mathcal{B})^T(\Sigma_\xi^{-1}-\alpha\mathcal{W}_\xi)^{-1}\mathcal{W}_{\xi,\mathcal{X}})\bar{\mathcal{A}}-(\mathcal{G}_{\mathcal{X}}\bar{\mathcal{A}}^-)^T\mathcal{R}_Z^{-1}\mathcal{G}_{\mathcal{X}}\bar{\mathcal{A}}^- - (\mathcal{G}^+)^T\mathcal{R}_{\mathcal{G}}^{-1}\mathcal{G}^+\right)\\
\mathcal{V} \;&=\; e^{-r}(\mathcal{Y}^+ + (\mathcal{G}^+)^T\mathcal{R}_{\mathcal{G}}^{-1}\mathcal{G}_0^*)\,.
\end{aligned}
\tag{51}
$$

Having computed the optimal policies, I need to impose market clearing and solve for the coefficients of price dynamics; that is, I need to make sure that the market clearing conditions for ETFs

$$
\begin{aligned}
&\underbrace{C_A^E+\Theta_{\mathcal{X}}^A\bar{\mathcal{X}}_{t-1+}+\Theta_\xi^A\xi_t \;-\; \mathbf{B}_A(\bar{P}+\Pi_X^E\bar{\mathcal{X}}_{t-1+}+\Pi_\xi^E\xi_t)}_{APs'\ demand\ for\ ETFs} \;+\; \underbrace{Q_E^* \;+\; \Theta_{\mathcal{X}}^E\bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\xi^E\xi_t}_{clients'\ demand\ for\ ETFs}\\
&= \qquad \underbrace{\bar{y}_{t-1+}}_{endogenous\ supply\ of\ ETFs}
\end{aligned}
\tag{52}
$$

[31] Exact expressions for these coefficients are provided in the Appendix.

[32] As usual, I assume that the discount rate β is sufficiently large to ensure the validity of the transversality condition.

and for basic securities

$$\underbrace{C_A^D + \theta_\xi^A \xi_t + \theta_{\mathcal{X}}^A \bar{\mathcal{X}}_{t-1+}}_{APs'\ demand\ for\ basic\ securities} \;+\; \underbrace{Q_D^* \;+\; \Theta_{\mathcal{X}}^D \bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\zeta^D \zeta_t - \mathbf{B}_D(\bar{p} + \Pi_X^D \bar{\mathcal{X}}_{t-1+} + \Pi_\zeta^E \zeta_t)}_{dealers'\ demand\ for\ basic\ securities}$$
$$= \underbrace{\bar{x}_{t-1+}}_{endogenous\ supply\ of\ basic\ securities} \tag{53}$$

hold for any realization of supply $\bar{\mathcal{X}}_{t-1+} = \binom{y_{t-1+}}{x_{t-1+}}$ and any realization of demand shocks $\zeta_t = \binom{\varepsilon_t}{\xi_t}$. Equating the coefficients, we arrive at the following system for equilibrium coefficients:

$$\begin{aligned}
&C_A^E + Q_E^* - \mathbf{B}_A \bar{P} = 0\\
&\Theta_y^A - \mathbf{B}_A \Pi_y^E + \Theta_y^E = \mathrm{Id}\\
&\Theta_x^A - \mathbf{B}_A \Pi_x^E + \Theta_x^E = 0\\
&\Theta_\xi^A - \mathbf{B}_A \Pi_\xi^E + \Theta_\xi^E = 0\\
&C_A^D + Q_D^* - \mathbf{B}_D \bar{p} = 0\\
&\theta_x^A + \Theta_x^D - \mathbf{B}_D \Pi_x^D = \mathrm{Id}\\
&\theta_y^A + \Theta_y^D - \mathbf{B}_D \Pi_y^D = 0\\
&\theta_\xi^A + \Theta_\xi^D - \mathbf{B}_D \Pi_\xi^D = 0\\
&\Theta_\varepsilon^D - \mathbf{B}_D \Pi_\varepsilon^D = 0\,,
\end{aligned} \tag{54}$$

where I use the notation $\Theta_{\mathcal{X}} = (\theta_y\ \theta_x)$. Finding an explicit solution to such a system is generally not possible, as the coefficients of the optimal policies depend in a non-linear way on the coefficients of price and state variables dynamics. However, as I show in the next section, the system can be explicitly solved when the volatility of demand shocks is small.

6 Equilibrium Dynamics for Small Demand Shock Volatility

In this section, I follow Vayanos (1999, 2001) and assume that the demand shock volatility (as measured by the parameter δ) is sufficiently small.

I start with the important observation that, in the limit of vanishing demand shock volatility, ETFs are redundant. Indeed, if $\delta = 0$, then there is no difference between market and limit orders for APs; after learning the value of $\mathbf{F}\Sigma_d \xi_t$, they know exactly what p_t is going to be and hence are able to perform a riskless arbitrage trade between an ETF and its underlying basket. I formalize this result in the following proposition:

Proposition 6.1 *In the limit, as $\delta \to 0$, ETFs are always priced by arbitrage: $P_t = \mathbf{F} p_t$. In particular, in the limit, as $\delta \to 0$, we have $\bar{P} = \mathbf{F}\bar{p}$, $\Pi_X^E = \mathbf{F}\Pi_X^D$, $\Pi_\xi^E = \mathbf{F}\Pi_\xi^D$.*

Proposition 6.1 implies that, in the limit as $\delta \to 0$, the problem becomes degenerate; since ETFs are redundant in that limit, APs are indifferent between holding a portfolio y of ETFs and holding a portfolio $\mathbf{F}^T y$ of basic securities. However, the effect of ETF trading on equilibrium dynamics is non-trivial and does not vanish when supply shock volatility is small. In particular, many equilibrium effects can be already seen in this limit, without investigating higher order terms.[33] The reason for this is that, when arbitrage riskiness is small, APs take bets of the order δ^{-1} against ETF mis-pricing, while the mis-pricing itself is of the order δ. When δ is small, the two effects cancel (because $\delta^{-1} \cdot \delta \sim 1$), and the arbitrage activity of APs has a finite impact on asset prices.

In order to state the next result, I need to define several auxiliary objects. Define the rescaled risk aversions

$$\alpha_* = \alpha(1-e^{-r}),\ \alpha_{*E} = \alpha_E(1-e^{-r}),\ \alpha_{*D} = \alpha_D(1-e^{-r}),$$

and let

$$\bar{\Lambda}_C \ \equiv\ e^r \alpha_*^{-1} \Lambda_C + \mathbf{F}\Sigma_d \mathbf{F}^T \tag{55}$$

be the effective cost of creation/redemption in cash (normalized by risk aversion). Here, I have This cost accounts for both the actual physical trading cost and the implicit cost due to an increased exposure to dividend risk that APs incur after a cash transaction. By contrast, the effective cost of an in-kind transaction,

$$\bar{\Lambda}_I \ \equiv\ e^r \alpha_*^{-1} \Lambda_I\,, \tag{56}$$

does not contain a dividend risk term because an in-kind transaction does not change the exposure to dividend risk.

Since APs do not provide liquidity in the basic securities market, they do not absorb idiosyncratic demand shocks, and therefore the sensitivity of prices to these shocks is determined directly by the hedging component of dealers' demand. Namely, Proposition 5.1 implies that we always have

$$\Pi_\varepsilon^D \ = \ -\alpha_{*D} e^{-r} \Sigma_d\,.$$

Thus, prices p_t depend on idiosyncratic demand shocks ε_t through $\widetilde{\varepsilon}_t = \Pi_\varepsilon^D \varepsilon_t$, and I denote by

$$\widetilde{\Sigma}_\varepsilon \ \equiv\ \mathrm{Var}[\widetilde{\varepsilon}_t] \ = \ (e^{-r}\alpha_{*D})^2 \Sigma_d \Sigma_\varepsilon \Sigma_d$$

the variance-covariance matrix of the "true" idiosyncratic shocks. Let also

$$\Sigma_E \ \equiv\ \mathrm{Var}_t[D_{t+1}] = \mathbf{F}\Sigma_d \mathbf{F}^T$$

[33]When $\delta = 0$, optimal policies are not well defined because APs are indifferent between holding an ETF and holding the underlying basket. By contrast, when δ is small, the model is well defined, and the unique equilibrium converges to a finite limit when $\delta \to 0$.

be the covariance matrix of ETF dividends, and let

$$\begin{aligned}
\mathbf{A} &= -\Sigma_d \mathbf{F}^T \bar{\Lambda}_C^{-1} \mathbf{F} \Sigma_d + \Sigma_d \\
\mathbf{C} &\equiv ((1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon \mathbf{F}^T)^{-1}((1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon + \mathbf{F}\widetilde{\Sigma}_\varepsilon \mathbf{F}^T \bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d)\mathbf{A}^{-1} \\
\Omega &= (\mathrm{Id} + \alpha_{*D}\Sigma_d \alpha_*^{-1}(\mathbf{A}^{-1} - \mathbf{F}^T \mathbf{C}))^{-1} \\
\Gamma &= \alpha_*^{-1} \begin{pmatrix} \bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d \mathbf{A}^{-1} \\ -\mathbf{F}^T \bar{\Lambda}_C^{-1} \mathbf{F} \Sigma_d \mathbf{A}^{-1} \end{pmatrix}.
\end{aligned} \tag{57}$$

Let also

$$\begin{aligned}
\Psi_{**} &= \alpha_* \Big(\mathrm{Id} + \alpha_{*D}\Big(\alpha_*^{-1}\mathbf{C} + \alpha_{*E}^{-1}\Sigma_E^{-1}\mathbf{F}\Big)\Omega\Sigma_d \mathbf{F}^T\Big)^{-1} \\
&\times \Big[-\mathbf{1}_y + \alpha_{*D}\Big(\alpha_*^{-1}\mathbf{C} + \alpha_{*E}^{-1}\Sigma_E^{-1}\mathbf{F}\Big)\Omega\Sigma_d \mathbf{1}_x\Big]; \\
\Psi_* &= \alpha_* \Big(\mathrm{Id} + \alpha_{*D}\Big(\alpha_*^{-1}\mathbf{C} + \alpha_{*E}^{-1}\Sigma_E^{-1}\mathbf{F}\Big)\Omega\Sigma_d \mathbf{F}^T\Big)^{-1} \Sigma_E^{-1}\mathbf{F}\Sigma_d,
\end{aligned} \tag{58}$$

and then define

$$\Pi_{**} = (1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon \mathbf{F}^T \Psi_{**}\Big((e^r - 1)\mathrm{Id} \; - \Gamma\Omega\alpha_{*D}\Sigma_d\Big[\alpha_*^{-1}\mathbf{F}^T\Psi_{**} - \mathbf{1}_x\Big]\Big)^{-1}$$

and

$$\Pi_* = [(1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon \mathbf{F}^T + \Pi_{**}\Gamma\,\Omega\alpha_{*D}\Sigma_d\alpha_*^{-1}\mathbf{F}^T]\Psi_*\,.$$

Here, $\mathbf{1}_x$ and $\mathbf{1}_y$ denote projections: $\mathbf{1}_x\binom{y}{x} = x,\ \mathbf{1}_y\binom{y}{x} = y.$

The following is true.

Theorem 6.1 *Suppose that δ is sufficiently small. Then, for generic $\alpha \in \mathbb{R}_+$, there exists a unique equilibrium if the matrix $\bar{\mathcal{A}}^*_{\mathcal{X}} \equiv \mathrm{Id} + \Gamma\Omega\alpha_{*D}\Sigma_d\Big[\alpha_*^{-1}\mathbf{F}^T\Psi_{**} - \mathbf{1}_x\Big]$ satisfies $\max(|\,\mathrm{eig}(\bar{\mathcal{A}}^*_{\mathcal{X}})|) < e^{r/2}$. In this case, equilibrium quantities are given approximately by the following expressions:*[34]

- *aggregate supply dynamics*

$$\begin{aligned}
\bar{\mathcal{A}}_{\mathcal{X}} &\approx \mathrm{Id} + \Gamma\Omega\alpha_{*D}\Sigma_d\Big[\alpha_*^{-1}\mathbf{F}^T\Psi_{**} - \mathbf{1}_x\Big] \\
\bar{\mathcal{A}}_\xi &\approx \Gamma\,\Omega\alpha_{*D}\Sigma_d\alpha_*^{-1}\mathbf{F}^T\Psi_*
\end{aligned} \tag{59}$$

- *basic securities' prices' dynamics and illiquidity*

$$\begin{aligned}
\Pi_X^D &\approx \Omega\alpha_{*D}\Sigma_d\Big[\alpha_*^{-1}\mathbf{F}^T\Psi_{**} - \mathbf{1}_x\Big](e^r\mathrm{Id} - \bar{\mathcal{A}}_{\mathcal{X}})^{-1} \\
\Pi_\xi^D &\approx e^{-r}(\Pi_X^D\Gamma + \mathrm{Id})\,\Omega\alpha_{*D}\Sigma_d\alpha_*^{-1}\mathbf{F}^T\Psi_* \\
\mathbf{B}_D^{-1} &\approx e^{-r}\alpha_{*D}\Big(\Sigma_d + \delta(\Pi_\xi^D\Sigma_\xi(\Pi_\xi^D)^T + \widetilde{\Sigma}_\varepsilon)\Big)
\end{aligned} \tag{60}$$

[34] Up to terms of the order $O(\delta)$.

- *ETF prices' dynamics and illiquidity*

$$\begin{aligned}
&\Pi_X^E \approx \mathbf{F}\Pi_X^D \;+\; \delta\Pi_{**} \\
&\Pi_\xi^E \approx \;\mathbf{F}\Pi_\xi^D \;+\; \delta\Pi_{*} \\
&\mathbf{B}_A^{-1} \approx \; e^{-r}\alpha_*\delta\Big((1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T \\
&-\delta\mathbf{F}\widetilde{\Sigma}_\varepsilon((1+e^{2r})\mathrm{Id}+\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F})(\Sigma_d^{-1}-\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F})^{-1}((1+e^{2r})\mathrm{Id}+\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F})\widetilde{\Sigma}_\varepsilon\mathbf{F}^T\Big)\,.
\end{aligned} \tag{61}$$

I now discuss several interesting features of the equilibrium characterized in Theorem 6.1. First of all, intuitively, one would expect only the effective supply (11) to matter when δ is small and ETFs become redundant. Indeed, if arbitrageurs can costlessly exchange an ETF for the underlying basket, it should not matter for the aggregate risk in the economy whether the agents hold Z units of ETFs or $\mathbf{F}^T Z$ shares of basic securities. As a result, by the invariance law (11), prices should stay constant, because creation/redemption does not change the effective supply. Formally, this can only be the case if the sensitivities of prices to the aggregate supply satisfy $\Pi_y^D = \Pi_x^D\mathbf{F}^T$, so that prices only depend on the effective supply. Indeed, in this case, $\Pi_y^D\bar{y}_{t-1+} + \Pi_x^D\bar{x}_{t-1+} = \Pi_x^D(\bar{x}_{t-1+}+\mathbf{F}^T\bar{y}_{t-1+})$. Surprisingly, Theorem 6.1 implies that this intuition is incorrect. Even when δ is very small but non-zero and the "degree of ETF redundancy" (the price gap) is very small (of the order δ), the arbitrage activity (which is of the order δ^{-1}) amplifies the effects of ETF flows and makes ETF supply a non-redundant state variable. The following is true.

Corollary 6.1 *In the limit, when demand shock volatility becomes small and ETFs become redundant, ETF creation/redemption is non-zero*[35] *and has a non-negligible impact on equilibrium prices.*

Theorem 6.1 also provides an explicit expression for the impact of ETFs on basic securities' illiquidity (as measured by $\mathbf{B}_D^{-1}$) and volatility. Namely, the willingness of basic securities dealers to provide liquidity is determined by the riskiness of the underlying returns:

$$\begin{aligned}
&\mathbf{B}_D^{-1} \;\approx\; e^{-r}\alpha_D(1-e^{-r})\mathrm{Var}_t[d_{t+1}+p_{t+1}]\,, \\
&\mathrm{Var}_t[d_{t+1}+p_{t+1}] \;=\; \mathrm{Var}_t[d_{t+1}+\Pi_\xi^D\xi_{t+1}+\Pi_\varepsilon^D\varepsilon_{t+1}] \;=\; \Sigma_d+\delta(\Pi_\xi^D\Sigma_\xi(\Pi_\xi^D)^T+\widetilde{\Sigma}_\varepsilon)\,.
\end{aligned} \tag{62}$$

In agreement with the empirical evidence in Ben-David, Franzoni, and Moussawi (2014), ETF trading propagates ETF demand shocks into the basic securities prices and introduces an additional term into return volatility and illiquidity. Thus, one might argue that ETF trading increases volatility. However, the natural question here is: what is the counter-factual? If there were no ETFs, ETF clients would still require some assets to hedge their income shocks and would either directly acquire basic securities or contact some intermediaries (analogs of APs in my model) who would have to directly acquire the underlying basket and then deliver it to the ETF clients. Independent of the nature of this intermediation, ETF demand shocks would still propagate into

[35] While in-kind transactions are useless in this case ($Z_I \to 0$ as $\delta \to 0$), cash transactions remain non-trivial even when δ is very small.

the prices of basic securities, and it is not obvious whether such alternative market designs would lead to lower volatility.

ETF demand shocks influence price volatility through two channels: the direct channel, represented by the term $\Pi_\xi^D \xi_t$, and the indirect channel, represented by the term $\Pi_X^D \bar{\mathcal{X}}_{t-1+}$ and by the fact that $\bar{\mathcal{X}}_{t-1+}$ depends on ξ_t through the term $\bar{\mathcal{A}}_\xi \xi_t$. Using these observations, we can decompose the conditional variances of price changes as follows:

$$\begin{aligned}
\mathrm{Var}_{t-1+}[p_{t+1}-p_t] &= \underbrace{(\Pi_X^D \bar{\mathcal{A}}_\xi - \Pi_\xi^D)\Sigma_\xi(\Pi_X^D \bar{\mathcal{A}}_\xi - \Pi_\xi^D)^T}_{creation/redemption\ effect} + \underbrace{\Pi_\xi^D \Sigma_\xi (\Pi_\xi^D)^T}_{ETF\ demand\ shocks} + 2\underbrace{\Pi_\varepsilon^D \Sigma_\varepsilon (\Pi_\varepsilon^D)^T}_{idiosyncratic\ shocks} \\
\mathrm{Var}_t[p_{t+1}-p_t] &= \Pi_\xi^D \Sigma_\xi (\Pi_\xi^D)^T + \Pi_\varepsilon^D \Sigma_\varepsilon (\Pi_\varepsilon^D)^T .
\end{aligned} \tag{63}$$

The first expression could be viewed as a measure of longer-term volatility that accounts for the dynamics of aggregate supply due to ETF creation/redemption, while the second expression can be interpreted as a measure of short-term (intra-day) volatility that only accounts for demand shocks. As I show below, the link between both volatilities and the structure of the ETF markets (primary market liquidity and basket weights) may exhibit unexpected behavior.

Another important consequence of Theorem 6.1 is the fact that ETF liquidity is always higher than the liquidity of the underlying basket. In fact, Theorem 6.1 implies that ETF illiquidity $\mathbf{B}_A^{-1}$ is always small (of the order δ) when δ is small. Since $\mathbf{B}_D^{-1}$ does not vanish when $\delta \to 0$, I arrive at the following result.

Corollary 6.2 *Suppose that the demand shock volatility is sufficiently small. In that case, ETF liquidity is always higher than the liquidity of the underlying securities.*

The key difference between ETF illiquidity $\mathbf{B}_A^{-1}$ and the basic securities illiquidity is that $\mathbf{B}_A^{-1}$ does not contain the term accounting for the dividend risk. The reason for this is that dividend riskiness does not directly impact ETF illiquidity[36] when arbitrage riskiness is sufficiently small: Indeed, dividend risk can be hedged by taking an opposite position in the underlying basket; similarly, since ETFs are redundant in the limit, we can see that the exposure to future shocks ξ_{t+1} can be also perfectly hedged with the same underlying basket. Thus, only the arbitrage riskiness determines how much liquidity arbitrageurs are willing to provide. In turn, arbitrage riskiness is determined by the idiosyncratic variance of the basket:

$$\begin{aligned}
&\mathrm{Var}_t[(P_{t+1}+D_{t+1}-e^r P_t) - \mathbf{F}(p_{t+1}+d_{t+1}-e^r p_t)] \approx \mathrm{Var}_t[\mathbf{F}(\Pi_\varepsilon^D \varepsilon_{t+1} - e^r \Pi_\varepsilon^D \varepsilon_t)] \\
&= \delta(1+e^{2r})\mathbf{F}\Pi_\varepsilon^D \Sigma_\varepsilon (\Pi_\varepsilon^D)^T \mathbf{F}^T .
\end{aligned} \tag{64}$$

Thus, ETFs play the role of "liquidity transformation," as they bundle less liquid securities into more liquid ones. Theorem 6.1 also shows that, naturally, ETF illiquidity depends on the liquidity Λ in the primary market, but only in the higher order term (in δ.)

Throughout the sequel, I assume that demand shock volatility is sufficiently small, such that Theorem 6.1 is directly applicable.

[36] Σ_d appears indirectly through $\Pi_\varepsilon^D \Sigma_\varepsilon (\Pi_\varepsilon^D)^T \approx (e^{-r}\alpha_{*D})^2 \Sigma_d \Sigma_\varepsilon \Sigma_d$.

7 Single ETF

In this section, I consider the case in which there is only one ETF available for trading. While this is a stylized example, it allows me to illustrate several important mechanisms through which ETFs influence equilibrium dynamics. Let

$$\begin{aligned} \bar{\xi}_t &\equiv \operatorname{Cov}_t(\xi_t^T d_{t+1}, D_{t+1}), \\ \bar{\zeta}^1_{t-1+} &= \operatorname{Cov}_t(\bar{x}_{t-1+} \cdot d_{t+1}, D_{t+1}), \\ \bar{\zeta}^2_{t-1+} &= \operatorname{Cov}_t(\bar{x}_{t-1+} \cdot \widetilde{\varepsilon}_{t+1}, f \cdot \widetilde{\varepsilon}_{t+1}) \end{aligned} \tag{65}$$

By definition, $\bar{\xi}_t$ is the covariance of ETF clients' income shocks with the ETF dividend; $\bar{\zeta}^1_{t-1+}$ is the covariance of the ETF dividend with the dividend of the "market portfolio" with weights given by the total supply of basic securities; and $\bar{\zeta}^2_{t-1+}$ is the covariance of the idiosyncratic shock $f \cdot \widetilde{\varepsilon}_{t+1}$ to the ETF NAV with the idiosyncratic shock $\bar{x}_{t-1+} \cdot \widetilde{\varepsilon}_{t+1}$ to the price of the market portfolio.

Proposition 7.1 *There exist $\psi, \psi_1, \psi_2, \psi_3 \in \mathbb{R}$ such that basic securities prices are given by*

$$\begin{aligned} p_{it} &= \bar{p}_i - \alpha_{*D} e^{-r}(1 + \alpha_{*D}\alpha_*^{-1}) \operatorname{Cov}_t(d_{i,t+1}, \bar{x}_{t-1+} \cdot d_{t+1}) - \alpha_{*D} e^{-r} \operatorname{Cov}_t(d_{i,t+1}, \varepsilon_t^T d_{t+1}) \\ &+ \operatorname{Cov}(d_{i,t+1}, D_{t+1})(\psi \bar{\xi}_t + \psi_1 \bar{\zeta}^1_{t-1+} + \psi_2 \bar{\zeta}^2_{t-1+} + \psi_3 \bar{y}_{t-1+}) \end{aligned} \tag{66}$$

for $i = 1, \cdots, N$.

Absent ETF trading, equilibrium prices are given by a simple CAPM-like expression: Since dividends are i.i.d., prices are given by the (risk-adjusted) present value $\bar{p}_i$ of dividends net of a risk premium given by the covariance of security's fundamentals d_i with two risk factors: the market portfolio and the dealers' income shock $\varepsilon_t^T d_{t+1}$. Proposition 7.1 shows that ETF trading introduces an additional risk factor into equilibrium dynamics: the ETF dividend. This is intuitive: the price pressure on the security i arises due to the hedging activity of APs who use basic securities to hedge their ETF positions, and the quality of the hedge is determined by the covariance with ETF dividends; furthermore, it does not matter whether a security belongs to the ETF basket; what matters is how useful it is as a hedge. Interestingly, the risk premium for the D_{t+1} risk factor is given by $\psi \bar{\xi}_t + \psi_1 \bar{\zeta}^1_{t-1+} + \psi_2 \bar{\zeta}^2_{t-1+} + \psi_3 \bar{y}_{t-1+}$ and moves over time together with the ETF demand shocks. These demand shocks lead to excess volatility because of the changes in the risk premium and in the market portfolio. Note, however, that the market portfolio always changes by multiples of the vector of ETF weights, f, and hence the covariance $\operatorname{Cov}_t(\bar{x}_{t-1+} \cdot d_{t+1}, d_{i,t+1})$ always changes by multiples of $\operatorname{Cov}(d_{i,t+1}, D_{t+1})$. In particular, Proposition 7.1 implies that two securities i and j respond to ETF demand shocks in the same direction if and only if $\operatorname{Cov}(d_{i,t+1}, D_{t+1}) \operatorname{Cov}(d_{j,t+1}, D_{t+1}) > 0$. Otherwise, when covariances are of opposite signs, APs create buying pressure in one security and selling pressure in another, pushing them in opposite directions. Thus, we arrive at the following result.

Corollary 7.1 *ETF trading increases co-movement of securities i and j if and only if*

$$\mathrm{Cov}(d_i, D)\,\mathrm{Cov}(d_j, D) > 0.$$

Another interesting observation concerns the coefficient $\alpha_{*D}e^{-r}(1+\alpha_{*D}\alpha_*^{-1})$ for the covariance with the dividend of the market portfolio; while it is increasing in dealers' risk aversion α_D (just as in the standard CAPM), it is decreasing in APs risk aversion α. The reason is that APs serve as liquidity demanders in the basic securities, and the dealers are forced to absorb their orders; when APs' risk aversion is low, they take on a lot of arbitrage risk and demand a lot of liquidity from the dealers, who require a higher premium for this liquidity provision.

I now discuss the nature of endogenous supply dynamics and its impact on the equilibrium behavior of prices. Absent ETFs and ETF demand shocks, prices are driven only by the idiosyncratic demand shocks ε_t, and price changes are naturally mean-reverting because shocks are i.i.d. over time and prices always revert back to their mean:[37]

$$\mathrm{Cov}_t[p_{t+1}-p_t, p_{t+2}-p_{t+1}] \;=\; \mathrm{Cov}_t[\Pi_\varepsilon^D(\varepsilon_{t+1}-\varepsilon_t), \Pi_\varepsilon^D(\varepsilon_{t+2}-\varepsilon_{t+1})] \;=\; -\delta\widetilde{\Sigma}_\varepsilon < 0\,. \tag{67}$$

As I will show now, ETF trading may give rise to a positive auto-covariance of returns (the "momentum" effect), because the endogenously time-varying supply makes the effect of ETF demand shocks persistent. In order to understand how this mechanism works, I decompose the auto-covariance of returns as follows:

$$\begin{aligned}
\mathrm{Cov}_t[p_{t+1}-p_t, p_{t+2}-p_{t+1}] \;&=\; \mathrm{Cov}_t[\Pi_X^D(\bar{\mathcal{X}}_{t+}-\bar{\mathcal{X}}_{t-1+}) \;+\; \Pi_\xi^D(\xi_{t+1}-\xi_t) \;+\; \Pi_\varepsilon^D(\varepsilon_{t+1}-\varepsilon_t),\\
&\Pi_X^D(\bar{\mathcal{X}}_{t+1+}-\bar{\mathcal{X}}_{t+}) \;+\; \Pi_\xi^D(\xi_{t+2}-\xi_{t+1}) \;+\; \Pi_\varepsilon^D(\varepsilon_{t+2}-\varepsilon_{t+1})]\\
&=\; \mathrm{Cov}_t[\Pi_\xi^D\xi_{t+1} \;+\; \Pi_\varepsilon^D\varepsilon_{t+1}\,,\,(\Pi_X^D\bar{\mathcal{A}}_\xi-\Pi_\xi^D)\xi_{t+1} \;-\Pi_\varepsilon^D\varepsilon_{t+1}]\\
&=\; \delta(\Pi_\xi^D\Sigma_\xi(\Pi_X^D\bar{\mathcal{A}}_\xi-\Pi_\xi^D)^T-\widetilde{\Sigma}_\varepsilon)
\end{aligned} \tag{68}$$

In the abive calculation, I have used the fact that $\bar{\mathcal{X}}_{t+1+}$ contains a term $\bar{\mathcal{A}}_\xi\xi_t$ and therefore covaries with past demand shocks. Thus, the time $t+1$ demand shock ξ_{t+1} enters the price change $p_{t+2}-p_{t+1}$ two times: in p_{t+1} and in p_{t+2}. If ETF demand shocks do not influence aggregate supply, $\bar{\mathcal{A}}_\xi = 0$ and we get the same result as in (67): prices recover from idiosyncratic shocks and revert to their means. However, if $\bar{\mathcal{A}}_\xi$ is sufficiently large, the effect may be reversed, as is shown by the following proposition.

Proposition 7.2 (Prices Autocovariance) *For generic α, there exists an open set of Σ_ξ for which ETF NAV changes exhibit momentum.*

Proposition 7.2 relates my paper to Vayanos and Woolley (2013), who show that mutual fund flows generate momentum in returns when investors learn about fund managers' skill. Proposition 7.2 shows that ETF fund flows are also able to generate momentum through the shock propagation

[37] See Greenwood (2005).

channel that arises from the creation/redemption mechanism. Namely, when a demand shock arrives, it pushes the ETF price up and forces APs to absorb this demand shock and sell ETF shares to clients. APs respond by redeeming their ETF shares in the primary market and reduce the aggregate supply of ETFs, which in turn reduces the risk premium and pushes the ETF price even further up.

I now turn to the behavior of the ETF pricing gap $P_t - \mathrm{NAV}_t$. One of the major motivations for the two-tier structure of the ETF market is the ability of the creation/redemption mechanism to efficiently correct ETF mispricing. Every time the ETF price deviates from the NAV, APs are able create/redeem shares and correct the mis-pricing. Formally, this intuition suggests that ETF mis-pricing should be mean-reverting. However, empirical evidence (see Petajisto, 2013) suggests that price gaps are often highly persistent, and this phenomenon is commonly attributed to the persistence of ETF demand shocks combined with limits to arbitrage that prevent arbitrageurs from correcting this pricing inefficiency. As I will now show, my model is able to generate price gap persistence even though demand shocks are i.i.d.

The only source of potential auto-covariance in the mis-pricing is the time-varying aggregate supply that depends on the realization of ETF demand shocks. Specifically, we have

$$\begin{aligned} &\mathrm{Cov}_{t-1+}[P_{t+1} - \mathrm{NAV}_{t+1}, P_t - \mathrm{NAV}_t] \;\approx\; \delta^2\,\mathrm{Cov}_{t-1+}[\Pi_{**}\bar{\mathcal{X}}_{t+} + \Pi_*\xi_{t+1}, \Pi_{**}\bar{\mathcal{X}}_{t-1+} + \Pi_*\xi_t] \\ &=\; \delta^2\,\mathrm{Cov}_{t-1+}[\Pi_{**}\bar{\mathcal{A}}_\xi\xi_t, \Pi_*\xi_t] \;=\; \delta^3\Pi_{**}\bar{\mathcal{A}}_\xi\Sigma_\xi\Pi_*^T\,, \end{aligned} \tag{69}$$

and there is no obvious reason why this quantity has to be negative. Indeed, the following is true.

Proposition 7.3 (Price Gap Persistence) *For generic α, there exists an open set of Σ_ξ for which the price gap is positively autocorrelated.*

In my model, mis-pricing persistence arises because demand shocks influence future supply of ETF shares and the corresponding risk premia. When a demand shock hits ETF prices and pushes them up w.r.t. to NAV, the willingness of arbitrageurs to correct this mis-pricing depends on their expectations of tomorrow's prices. If they anticipate that prices will increase even further tomorrow because creation/redemption changes the total supply, they will be long ETF shares even though the latter are over-priced. Thus, mis-pricing becomes a self-fulfilling equilibrium effect: anticipating an even higher mis-pricing tomorrow prevents APs from betting against the mis-pricing today, which further amplifies mis-pricing.

I now discuss the impact of primary market liquidity on equilibrium dynamics. If cash transactions in the primary market were almost costless, deviations of ETF price from the NAV would be hard to sustain because APs would exploit these deviations and create/redeem shares to correct them. However, since creations/redemptions only happen with a delay, significant intraday deviations may still occur before the APs have the chance to correct them in the primary market. In particular, it is natural to expect that the price gap will mean-revert very quickly to zero when Λ_C is sufficiently small. This may have a non-trivial effect on the volatility of the price gap and actually push this volatility up when Λ_C decreases. The following proposition confirms this intuition.

Proposition 7.4 (Primary market liquidity and the response of price gap to supply shocks) *The sensitivity* Π_* *of the price gap to supply shocks and the price gap volatility can be both increasing and decreasing in* Λ_C.

Interestingly, the same intuition also applies to the dynamics of prices of both the ETF and the basic securities. If we increase Λ_C, prices become more sensitive to demand shocks today because ETF liquidity drops; however, at the same time, tomorrow's prices may become less sensitive to today's demand shocks because there is less creation/redemption. Dealers, anticipating less creation/redemption and hence less arbitrage activity by APs are willing to provide more liquidity in the basic securities markets. As a result, longer-term and shorter-term volatilities may be non-monotonic in Λ_C. Namely, intuition is confirmed by the following proposition.

Proposition 7.5 (ETF liquidity and volatility) *The volatility of basic securities' prices may be non-monotonic in the ETF primary market illiquidity* Λ_C.

I now present several numerical examples of the behavior of equilibrium volatility. I separately consider the cases of high and low risk aversion of APs. When APs' risk aversion is low, they will be providing a lot of liquidity in the ETF markets and will be less constrained by their hedging needs. As a result, both ETF prices and the prices of underlying securities will be much less sensitive to ETF demand shocks, and hence both Π_ξ^D and the shorter-term volatility (see (63)) will be very small, as is illustrated by Figure 2. When risk aversion is higher, the magnitude of Π_ξ^D is larger and one can more clearly see the U-shaped dependence of the shorter-term volatility $\Pi_\xi^D \Sigma_\xi (\Pi_\xi^D)'$ on Λ_C (Figure 1).[38] Surprisingly, Figure 2 shows that the longer-term volatility is monotone increasing in Λ_C, which is again counter-intuitive, given that longer-term volatility (see (63)) is driven by the creation/redemption. The reason is that, when Λ_C increases, ETF liquidity drops, and this makes prices more sensitive to the amount of aggregate risk. That is, Π_X^D increases in Λ_C. Thus, while $\bar{\mathcal{A}}_\xi$ drops in response to an increase in Λ_C, this effect is offset by the increase in Π_X^D, the the product $\Pi_X^D \bar{\mathcal{A}}_\xi$ increases, making the longer-term volatility go up.[39] It is also important to note that the ranges on which the Λ_C varies on Figures 1 and 2 are different. The reason is that, for Λ_C below that range the no-bubble condition $\max(|\operatorname{eig}(\bar{\mathcal{A}}_\mathcal{X})|) < e^{r/2}$ fails to hold, and hence equilibrium existence cannot be guaranteed. One can clearly see this instability on Figure 1: As Λ_C decreases to the lower threshold of the region, creation/redemption intensifies, pricing become more non-stationary, and volatility spikes. This again emphasizes the important role that the primary market illiquidity plays in stabilizing the market: when Λ_C is too low, equilibrium may simply fail to exist and markets may collapse.

[38] The same U-shaped pattern is also present in Figure 2, but is difficult to observe due to an insufficient magnitude of the effect.

[39] The creation-redemption effect also makes longer-term volatility increase relative to the shorter-term volatility on Figure 1, but the effect is small in magnitude.

8 Multiple ETFs

In this section, I investigate the effects of the presence of multiple ETFs. I start with the question "Do ETFs increase volatility and co-movement?" which was recently investigated empirically by Ben-David, Franzoni, and Moussawi (2014). I address this question by considering the effect of introducing new ETFs on the volatility and co-movement of basic securities. After the introduction of a new ETF, demand shocks continue to influence the dynamics of basic securities' prices, but the nature of ETF trading changes, as does the way demand shocks are distributed across different markets. The following is true.

Proposition 8.1 (Does the introduction of new ETFs increase volatility and co-movement?) *Fix a pair of securities i, j. For a generic new ETF, there exists an open set of matrices Σ_ξ such that the introduction of this ETF reduces the volatility of these two securities and their co-movement.*

The intuition behind the result of Proposition 8.1 can be gained from the single-ETF case. Suppose for simplicity that there are only two ETFs with portfolio weights f_1, f_2. Then, the shocks ξ_t affect equilibrium prices through the quantities $\bar{\xi}_{kt} = \mathrm{Cov}_t(\xi_t \cdot d_{t+1}, D_{k,t+1})$, where $k = 1, 2$ is the corresponding ETF. Then, the exposure of the price p_{it} to the shocks ξ_t can be re-expressed as

$$p_{i,t} \sim \sum_{k_1,k_2=1}^{2} \psi_{k_1,k_2} \, \mathrm{Cov}(d_{i,t+1}, D_{k_1,t+1}) \bar{\xi}_{k_2,t} \; = \; A_{i,1}\bar{\xi}_{1,t} + A_{i,2}\bar{\xi}_{2,t}$$

for some coefficients ψ_{k_1,k_2}, where

$$A_{i,k} \; = \; \sum_{k_1=1}^{2} \psi_{k_1,k} \, \mathrm{Cov}(d_{i,t+1}, D_{k_1,t+1}) \, .$$

Suppose for simplicity that the coefficients ψ_{k_1,k_2} are fixed.[40] Then, the introduction of a new ETF number 2 effectively creates exposure to an additional shock $\bar{\xi}_2$. By properly choosing the new ETF weights, we can find that the two demand shocks $\bar{\xi}_{1,t}$, $\bar{\xi}_{2,t}$ are negatively correlated while at the same time the coefficients $A_{1,i}$ and $A_{2,i}$ are of the same sign, such that the new shock offsets the old shocks and reduces the overall volatility of security i. Note that, interestingly, this "volatility substitution effect" works even if security dividends are not correlated with the new ETF dividends. The reason for this is that trading in the new ETF attracts some of the trading from the old ETF, which exposes the security price to the $\bar{\xi}_{2,t}$ shocks even if security's fundamentals are not correlated with those of the new ETF.

Figure 3 illustrates these effects. Initially, introducing new ETFs reduces volatility through the demand substitution effect described above. However, eventually, the ETF market becomes "over-crowded", and there are "too many" ETFs that make volatility go up.

The next important question concerns the role of ETF trading for the liquidity of basic securities. According to Theorem 6.1, illiquidity is (approximately) proportional to the basic securities

[40] Of course, in equilibrium, these coefficients are endogenous and depend on the ETF weights.

variance, and hence the overall effect of ETF demand shocks on the liquidity is negative; when risk averse dealers anticipate more demand shocks and larger volatility, they reduce their liquidity provision. However, Proposition 8.1 above immediately implies that introducing new ETFs may improve liquidity in some of the securities. I formalize this observation in the following proposition.

Proposition 8.2 (Do ETFs improve liquidity in the underlying securities?) *Fix a pair of securities i, j. For a generic new ETF, there exists an open set of matrices Σ_ξ such that the introduction of this ETF reduces the volatilities of these two securities and their co-movement.*

The presence of a primary market with an exogenously given illiquidity Λ (chosen by the ETF sponsors) naturally raises the question of the link between the primary market liquidity and the secondary market liquidity. Given that the primary market serves as an additional source of liquidity for APs, it is natural to expect that improvements in the primary market liquidity (that is, a reduction in the spreads and costs charged by the ETF sponsors to APs) should positively influence liquidity in the secondary market. This intuition is partially confirmed by the following proposition.

Proposition 8.3 (Primary market liquidity and secondary market liquidity) *Suppose that Λ_C is sufficiently small. Then, secondary market liquidity of any ETF is increasing in the primary market liquidity of any ETF. At the same time, the introduction of a new ETF may lead to a drop in liquidity for some of the existing ETFs.*

The fact that liquidity spills over from one ETF to all other ETFs is a positive result that shows that there are complementarities in liquidity provision by ETF sponsors. At the same time, Proposition 8.3 shows that financial innovation (introducing new ETFs) may have unintended consequences for existing markets. The intuition behind this result is based on the demand substitution effect: when there are more ETFs, APs may decide to provide more liquidity in the new market at the cost of reducing liquidity provision in some of the existing ETFs. This observation may have important implications for the design of new ETFs by sponsors. Specifically, when introducing a new ETF, sponsors should ensure that it does not hurt liquidity in other existing markets.

9 Conclusions

The unprecedented growth of the ETF industry over the past two decades has attracted a lot of attention from academics, practitioners, and regulators. The dynamics of risks and liquidity for the increasingly complex ETF landscape have become a major concern of many market participants. In this paper, I develop a rigorous theoretical framework for studying the pricing, risk premia, and liquidity of ETFs. My model allows for any number of ETFs and underlying securities, making it possible to study the ETF universe jointly and to investigate the mutual impact of trading different ETFs. My results indicate that ETFs may be both a blessing and a curse. That is, introducing new ETFs may lead to a significant amplification of speculative behavior of arbitrageurs, destabilize the market, and lead to a spike in volatility; however, at the same time, a "good" ETF may actually stabilize the economy, lead to a significant reduction in volatility, and improve the liquidity of the underlying securities. Regulators should be aware of this phenomenon and investigate the consequences of different ETF designs on the market.

In my paper, all ETFs are physical and hence do not involve any counter-party risk. At the same time, many of the modern ETFs are synthetic; they effectively serve as securitization devices for pooling less-liquid instruments and expose ETF investors to counter-party risk. These effects in general equilibrium is an important topic of investigation for future research.

Finally, in my paper, information is symmetric, and ETFs play a pure risk-sharing role and are very liquid because fundamental risk can be hedged. In reality, ETFs may also be more liquid than the underlying securities because ETF investors are less subject to adverse selection. Investigating the role of ETFs in price discovery and their interactions with asymmetric information is another promising direction for future research.

A Figures

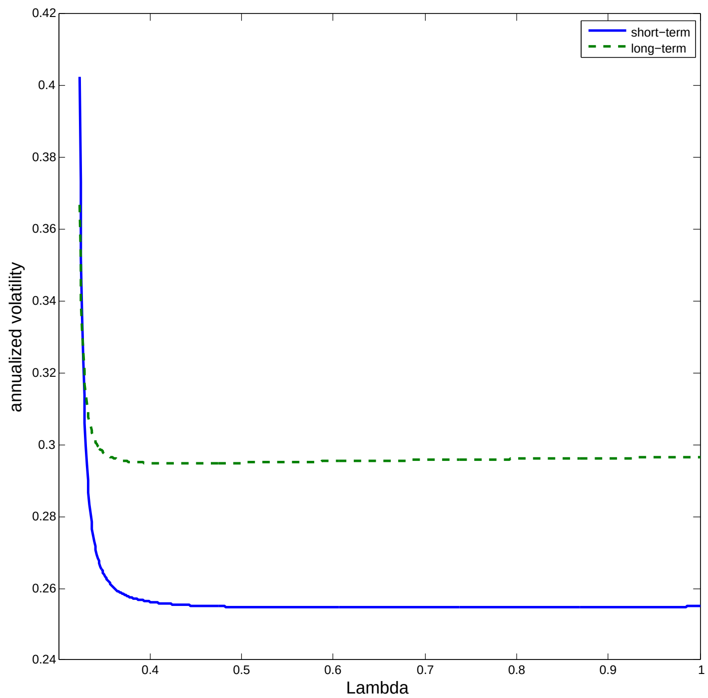

Figure 1: Volatility of basic securities and primary market liquidity $,\Lambda_C$. Short- and long-term volatilities are defined as square roots of expressions in (63), but annualized: the short-term volatility is multiplied by $252^{1/2}$, the long-term one is multiplied by $(252/2)^{1/2}$. Parameter values: $N = 10$; $L = 1$; $\Sigma_d = \mathrm{Id} + \mathbf{1}_{N\times N}$; $\Sigma_\varepsilon = \Sigma_\xi = 0.05\,\mathrm{Id}$, $\alpha_* = 2$, $\alpha_{*E} = 0.2$; $\alpha_{*D} = 0.02$; $r = 0.05/252$; $\mathbf{F} = \frac{1}{N}\mathbf{1}$.

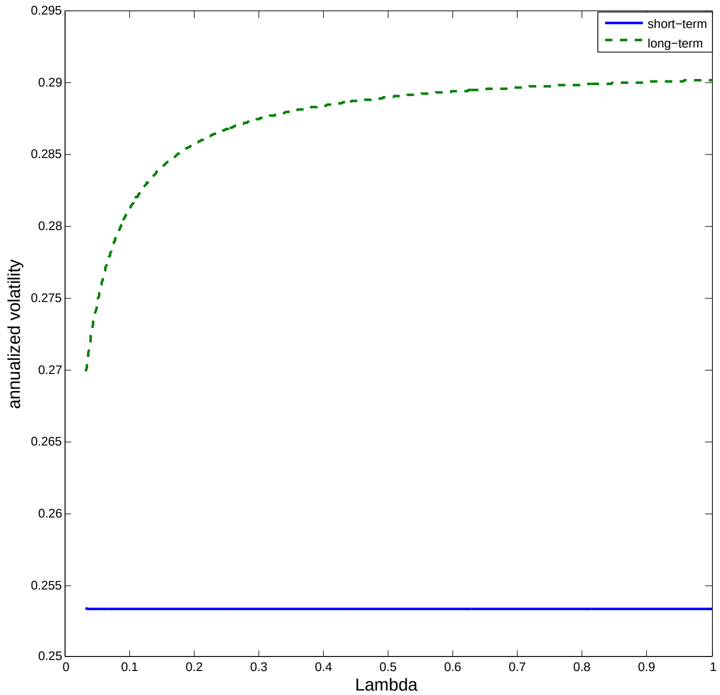

Figure 2: Volatility of basic securities and primary market liquidity $,\Lambda_C$. Short- and long-term volatilities are defined as square roots of expressions in (63), but annualized: the short-term volatility is multiplied by $252^{1/2}$, the long-term one is multiplied by $(252/2)^{1/2}$. Parameter values: $N = 10;\ L = 1;\ \Sigma_d = \mathrm{Id} + \mathbf{1}_{N\times N};\ \Sigma_\varepsilon = \Sigma_\xi = 0.05\,\mathrm{Id},\ \alpha_* = 0.2,\ \alpha_{*E} = 0.2;\ \alpha_{*D} = 0.02;\ r = 0.05/252;\ \mathbf{F} = \frac{1}{N}\mathbf{1}$.

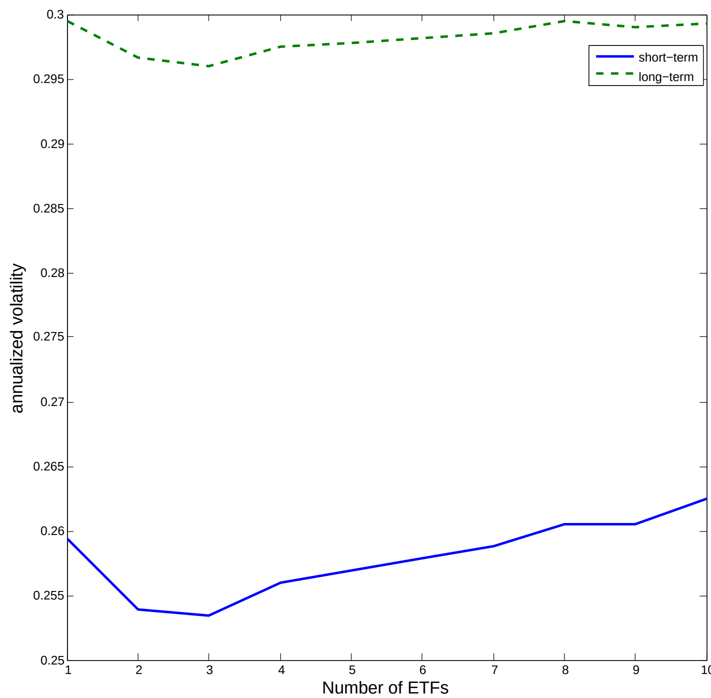

Figure 3: Volatility of basic securities and new ETFs. Short- and long-term volatilities are defined as square roots of expressions in (63) but annualized: the short-term volatility is multiplied by $252^{1/2}$, and the long-term one is multiplied by $(252/2)^{1/2}$. Parameter values: $N = 10;\ L = 1;\ \Sigma_d = \mathrm{Id} + \mathbf{1}_{N\times N};\ \Sigma_\varepsilon = \Sigma_\xi = 0.05\,\mathrm{Id},\ \alpha_* = 2,\ \alpha_{*E} = 0.2;\ \alpha_{*D} = 0.02;\ r = 0.05/252$. The ETF universe starts with $\mathbf{F} = \frac{1}{N}\mathbf{1};\ \Lambda_C = 3$. Then, every new ETF is introduced using a Gaussian random number generator, and then weights are exponentiated and normalized to add up to one. The shape of the graph is robust and looks similar for most random ETFs that have been generated.

B Securities Dealers: Proof of Proposition 5.1

To solve the HJB equation, we first compute

$$
\begin{aligned}
&E_t\Bigg[\exp\Bigg\{-\alpha_D((1-e^{-r})(e^r(M_t-c_t)\ +\ (x_t+\varepsilon_t)\cdot d_{t+1}\\
&+\ x_t\cdot(\bar p\ +\ \Pi_X^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)+\ \Pi_\zeta^D\zeta_{t+1}-e^rp_t))\\
&+\bar v^D+\mathcal{V}^D\cdot(\zeta_{t+1},\bar{\mathcal{X}}_{t+})-0.5(\zeta_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big))^T\mathcal{W}^D(\zeta_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)))\Bigg\}\Bigg]\\
&=\ E_t\Bigg[\exp\Bigg\{-\alpha_E((1-e^{-r})(e^r(M_t-c_t)\ +\ (x_t+\varepsilon_t)\cdot\bar d-0.5\alpha_D(x_t+\varepsilon_t)^T\Sigma_d(x_t+\varepsilon_t)\\
&+\ x_t\cdot(\bar p\ +\ \Pi_X^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)+\ \Pi_\zeta^D\zeta_{t+1}-e^rp_t))\\
&+\bar v^D+\mathcal{V}^D\cdot(\zeta_{t+1},\bar{\mathcal{X}}_{t+})-0.5(\zeta_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big))^T\mathcal{W}^D(\zeta_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)))\Bigg\}\Bigg]\\
&=\exp\Bigg\{-\alpha_D\Bigg((1-e^{-r})(e^r(M_t-c_t)\ +\ (x_t+\varepsilon_t)\cdot\bar d-0.5\alpha_D(1-e^{-r})(x_t+\varepsilon_t)^T\Sigma_d(x_t+\varepsilon_t)\\
&+\ x_t\cdot(\bar p\ +\ \Pi_X^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)-e^rp_t))\\
&+\bar v^D+\mathcal{V}_{\mathcal{X}}^D\cdot\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)-0.5\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)^T\mathcal{W}_{\mathcal{X}}^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg\}\\
&\times\frac{\det((\Sigma_\zeta^{-1}-\alpha_E\mathcal{W}_\zeta^D)^{-1})e^{0.5(\bar Y^T(\Sigma_\zeta^{-1}-\alpha\mathcal{W}_\zeta^D)\bar Y-\bar\zeta^T\Sigma_\zeta^{-1}\bar\zeta)}}{(\det\Sigma_\zeta)^{1/2}}
\end{aligned}
\tag{70}
$$

where

$$
\bar Y\ =\ (\Sigma_\zeta^{-1}-\alpha_D\mathcal{W}_\zeta^D)^{-1}\Bigg(\Sigma_\zeta^{-1}\bar\zeta-\alpha_D\Bigg((1-e^{-r})(\Pi_\zeta^E)^Tx_t+\mathcal{V}_\zeta^E+\mathcal{W}_{\zeta,\mathcal{X}}^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg)
$$

Thus, the optimization problem over x_t becomes to maximize

$$
\begin{aligned}
x_t(1-e^{-r})\cdot\Bigg(&\bar d-(1-e^{-r})\alpha_D\Sigma_d\varepsilon_t+(\bar p\ +\ \Pi_X^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)-e^rp_t)\\
&+\Pi_\zeta^D(\Sigma_\zeta^{-1}-\alpha_D\mathcal{W}_\zeta^D)^{-1}\Bigg(\Sigma_\zeta^{-1}\bar\zeta-\alpha_D\Bigg(\mathcal{V}_\zeta^D+\mathcal{W}_{\zeta,\mathcal{X}}^D\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg)\Bigg)\\
&-0.5\alpha_D(1-e^{-r})^2x_t^T\Big(\Sigma_d+\Pi_\zeta^D(\Sigma_\zeta^{-1}-\alpha_D\mathcal{W}_\zeta^D)^{-1}(\Pi_\zeta^D)^T\Big)x_t\,,
\end{aligned}
\tag{71}
$$

which gives the optimal trade

$$
\begin{aligned}
q^D + x_{t-1} \;&=\; e^{-r}\mathbf{B}_D\Bigg(\bar d - (1-e^{-r})\alpha_D\Sigma_d\varepsilon_t + (\bar p \;+\; \Pi_X^D\Big(\bar{\mathcal{X}}^* + \bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi\xi_t\Big) - e^r p_t)\\
&+ \Pi_\zeta^D(\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}\Bigg(\Sigma_\zeta^{-1}\bar\zeta - \alpha_D\Bigg(\mathcal{V}_\zeta^D + \mathcal{W}_{\zeta,\mathcal{X}}^D\Big(\bar{\mathcal{X}}^* + \bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg)\Bigg) \qquad (72)\\
&=\; Q_D^* \;+\; \Theta_{\mathcal{X}}^D\bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\zeta^D\zeta_t - \mathbf{B}_D p_t\\
&=\; (Q_D^* - \mathbf{B}_D\bar p) \;+\; (\Theta_{\mathcal{X}}^D - \mathbf{B}_D\Pi_X^D)\mathcal{X}_{t-1+} \;+\; (\Theta_\xi^D - \mathbf{B}_D\Pi_\xi^D)\xi_t
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{B}_D^{-1} &= e^{-r}\alpha_D(1-e^{-r})\Big(\Sigma_d + \Pi_\zeta^D(\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}(\Pi_\zeta^D)^T\Big)\\
Q_D^* \;&=\; e^{-r}\mathbf{B}_D\Bigg(\bar d \;+\; \bar p \;+\; \Pi_X^D\bar{\mathcal{X}}^* + \Pi_\zeta^D(\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}\Bigg(\Sigma_\zeta^{-1}\bar\zeta - \alpha_D\Bigg(\mathcal{V}_\zeta^D + \mathcal{W}_{\zeta,\mathcal{X}}^D\bar{\mathcal{X}}^*\Bigg)\Bigg)\Bigg)\\
\Theta_{\mathcal{X}}^D \;&=\; e^{-r}\mathbf{B}_D\Big(\Pi_X^D - \Pi_\zeta^D(\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}\alpha_D\mathcal{W}_{\zeta,\mathcal{X}}^D\Big)\bar{\mathcal{A}}_{\mathcal{X}}\\
\Theta_\zeta^D \;&=\; e^{-r}\mathbf{B}_D\Big(\Pi_X^D - \Pi_\zeta^D(\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}\alpha_D\mathcal{W}_{\zeta,\mathcal{X}}^D\Big)\bar{\mathcal{A}}_\xi\mathbf{1}_\xi \;-\alpha_D(1-e^{-r})e^{-r}\mathbf{B}_D\Sigma_d\mathbf{1}_\varepsilon
\end{aligned} \qquad (73)
$$

This defines a fixed point system for $\mathcal{W}^D$ and $\mathcal{V}^D$:

$$
\begin{aligned}
e^r\mathcal{W}^D \;&=\; -(\Theta^D - \mathbf{B}_D\Pi^D)^T(e^{-r}\mathbf{B}_D)^{-1}(\Theta^D - \mathbf{B}_D\Pi^D) + (\bar{\mathcal{A}})^T\mathcal{W}_{\mathcal{X}}^D\bar{\mathcal{A}} + \alpha_D(1-e^{-r})^2\begin{pmatrix}\Sigma_d & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\\
e^r\mathcal{V}^D \;&=\; \begin{pmatrix}(1-e^{-r})\bar d\\ 0\\ 0\end{pmatrix} \;+\; \bar{\mathcal{A}}^T\Bigg(\mathcal{V}_{\mathcal{X}}^D - \mathcal{W}_{\mathcal{X}}^D\bar{\mathcal{X}}^* + (\Sigma_\zeta^{-1} - \alpha_D\mathcal{W}_\zeta^D)^{-1}(\Sigma_\zeta^{-1}\bar\zeta - \alpha_D(\mathcal{V}_\zeta^D + \mathcal{W}_{\zeta,\mathcal{X}}^D\bar X^*))\Bigg)\\
&+ (\Theta^D - e^r\mathbf{B}_D\Pi^D)^T(e^{-r}\mathbf{B}_D)^{-1}(Q_D^* - \mathbf{B}_D\bar p).
\end{aligned} \qquad (74)
$$

This defines the optimal policy matrices as a function of just three matrices $(\bar{\mathcal{A}},\ \Pi_X^D,\ \Pi_\zeta^D)$.

Now, the optimization problem over consumption can be rewritten as

$$
\max_c(-e^{-\alpha c} \;-\; e^{-\beta}e^{-\alpha(1-e^{-r})(M-c)+W^*})
$$

where W^* is the quadratic expression computed above. This gives $\alpha c = \alpha(1-e^{-r})M - e^{-r}W^* + c^*$ for some constant c^*, and the dynamic programming equation follows by direct calculation.

Finally, the transversality condition takes the form

$$
\lim_{T\to\infty} E_t[e^{-\beta(T-t)}V_T] \;=\; 0\,.
$$

It is straightforward to show that, when $\mathcal{W}_{\mathcal{X}}^D$ is negative definite, any exponential growth in the $\bar{\mathcal{X}}_{t-1+}$ process is always offset by the quadratic term $\bar{\mathcal{X}}_{t-1+}^T\mathcal{W}_{\mathcal{X}}^D\bar{\mathcal{X}}_{t-1+}$, and the claim follows if β is sufficiently large. The fact that $\mathcal{W}_{\mathcal{X}}^D$ is negative definite follows from the following lemma.

Lemma B.1 *Suppose that a matrix* $\mathcal{W}$ *satisfies*

$$e^r\mathcal{W} \;=\; \mathcal{A}^T\mathcal{W}\mathcal{A} \;-\; \mathcal{B}$$

where $\mathcal{B}$ *is positive definite and* $\mathcal{A}$ *satisfies* $\max(|\operatorname{eig}(\mathcal{A})|) < e^{r/2}$. *Then* $\mathcal{W}$ *is negative definite.*

Proof of Lemma ??. Consider the operator A^* on $\mathbb{R}^{N\times N}$ acting as $A^*(\mathcal{W}) = \mathcal{A}^T\mathcal{W}\mathcal{A}$. Then, it is straightforward to show that $\operatorname{eig}(A^*) = \operatorname{eig}(\mathcal{A}) \times \operatorname{eig}(\mathcal{A}) = \{\lambda_1\lambda_2 : \; \lambda_1, \lambda_2 \in \operatorname{eig}(\mathcal{A})\}$. Rewriting our equation as

$$(\mathrm{Id} - e^{-r}A^*)\mathcal{W} \;=\; -\mathcal{B}$$

we get

$$\mathcal{W} \;=\; -e^{-r}\sum_{n=0}^{\infty}(e^{-r}A^*)^n(\mathcal{B}) \;=\; -\sum_{n=0}^{\infty}(e^{-r/2}\mathcal{A}^T)^n e^{-r}\mathcal{B}(e^{-r/2}\mathcal{A})^n\,,$$

and the claim follows. ■

Suppose now that supply shock volatility is small. Then, we have

$$\begin{aligned}
\mathbf{B}_D^{-1} \;&\approx\; e^{-r}\alpha_D(1-e^{-r})(\Sigma_d + \delta\Pi_\zeta^D\Sigma_\zeta(\Pi_\zeta^D)^T)\\
\mathbf{B}_D \;&\approx\; e^{r}(\alpha_D(1-e^{-r}))^{-1}(\Sigma_d^{-1} - \delta\Sigma_d^{-1}\Pi_\zeta^D\Sigma_\zeta(\bar\Pi_\zeta^D)^T\Sigma_d^{-1})\\
\Theta_{\mathcal{X}}^D \;&\approx\; (\alpha_D(1-e^{-r}))^{-1}(\Sigma_d^{-1} - \delta\Sigma_d^{-1}\Pi_\zeta^D\Sigma_\zeta(\bar\Pi_\zeta^D)^T\Sigma_d^{-1})\Big(\Pi_X^D - \delta\Pi_\zeta^D\Sigma_\zeta\alpha_D\mathcal{W}_{\zeta,\mathcal{X}}^D\Big)\bar{\mathcal{A}}_{\mathcal{X}}\\
\Theta_\zeta^D \;&\approx\; (\alpha_D(1-e^{-r}))^{-1}(\Sigma_d^{-1} - \delta\Sigma_d^{-1}\Pi_\zeta^D\Sigma_\zeta(\bar\Pi_\zeta^D)^T\Sigma_d^{-1})\Big(\Big(\Pi_X^D - \delta\Pi_\zeta^D\Sigma_\zeta\alpha_D\mathcal{W}_{\zeta,\mathcal{X}}^D\Big)\bar{\mathcal{A}}_\xi\mathbf{1}_\xi \;-\alpha_D(1-e^{-r})\Sigma_d\mathbf{1}_\varepsilon\Big)
\end{aligned} \tag{75}$$

C ETF Clients: Proof of Proposition 5.2

The dynamics of the clients' money market account is given by

$$M_{t+1} \;=\; e^r(M_t - c_t - P_t\cdot y_t) \;+\; (\mathbf{F}^T y_t + \xi_t)\cdot d_{t+1} \;+\; y_t\cdot P_{t+1}\,.$$

To solve the dynamic programming equation, we first compute

$$
\begin{aligned}
&E_t\Bigg[\exp\Bigg\{-\alpha_E((1-e^{-r})(e^r(M_t-c_t)\ +\ (\mathbf{F}^T y_t+\xi_t)\cdot d_{t+1} \\
&+\ y_t\cdot(\bar{P}\ +\ \Pi_X^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)+\ \Pi_\xi^E\xi_{t+1}-e^rP_t)) \\
&+\bar{v}^E+\mathcal{V}^E\cdot(\xi_{t+1},\bar{\mathcal{X}}_{t+})-0.5(\xi_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big))^T\mathcal{W}^E(\xi_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)))\Bigg\}\Bigg] \\
&=\ E_t\Bigg[\exp\Bigg\{-\alpha_E((1-e^{-r})(e^r(M_t-c_t)\ +\ (\mathbf{F}^T y_t+\xi_t)\cdot\bar{d}-0.5\alpha_E(\mathbf{F}^T y_t+\xi_t)^T\Sigma_d(\mathbf{F}^T y_t+\xi_t) \\
&+\ y_t\cdot(\bar{P}\ +\ \Pi_X^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)+\ \Pi_\xi^E\xi_{t+1}-e^rP_t)) \\
&+\bar{v}^E+\mathcal{V}^E\cdot(\xi_{t+1},\bar{\mathcal{X}}_{t+})-0.5(\xi_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big))^T\mathcal{W}^E(\xi_{t+1},\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)))\Bigg\}\Bigg] \\
&=\exp\Bigg\{-\alpha_E\Bigg((1-e^{-r})(e^r(M_t-c_t)\ +\ (\mathbf{F}^T y_t+\xi_t)\cdot\bar{d}-0.5\alpha_E(1-e^{-r})(\mathbf{F}^T y_t+\xi_t)^T\Sigma_d(\mathbf{F}^T y_t+\xi_t) \\
&+\ y_t\cdot(\bar{P}\ +\ \Pi_X^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)-e^rP_t)) \\
&+\bar{v}^E+\mathcal{V}_{\mathcal{X}}^E\cdot\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)-0.5\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)^T\mathcal{W}_{\mathcal{X}}^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg\} \\
&\times\frac{\det((\Sigma_\xi^{-1}-\alpha_E\mathcal{W}_\xi^E)^{-1})e^{0.5(\bar{Y}^T(\Sigma_\xi^{-1}-\alpha\mathcal{W}_\xi^E)\bar{Y}-\bar{\xi}^T\Sigma_\xi^{-1}\bar{\xi})}}{(\det\Sigma_\xi)^{1/2}}
\end{aligned}
\tag{76}
$$

where

$$
\bar{Y}\ =\ (\Sigma_\xi^{-1}-\alpha_E\mathcal{W}_\xi^E)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi}-\alpha_E\Bigg((1-e^{-r})(\Pi_\xi^E)^Ty_t+\mathcal{V}_\xi^E+\mathcal{W}_{\xi,\mathcal{X}}^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg).
$$

Thus, the optimization problem over y_t becomes to maximize

$$
\begin{aligned}
y_t\cdot\Bigg(&(1-e^{-r})\mathbf{F}\bar{d}-\alpha_E(1-e^{-r})^2\mathbf{F}\Sigma_d\xi_t+(1-e^{-r})(\bar{P}\ +\ \Pi_X^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)-e^rP_t) \\
&+(1-e^{-r})\Pi_\xi^E(\Sigma_\xi^{-1}-\alpha_E\mathcal{W}_\xi^E)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi}-\alpha_E\Bigg(\mathcal{V}_\xi^E+\mathcal{W}_{\xi,\mathcal{X}}^E\Big(\bar{\mathcal{X}}^*+\bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+}+\bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg)\Bigg) \\
&-0.5\alpha_E y_t^T(1-e^{-r})^2\Big(\mathbf{F}\Sigma_d\mathbf{F}^T+\Pi_\xi^E(\Sigma_\xi^{-1}-\alpha_E\mathcal{W}_\xi^E)^{-1}(\Pi_\xi^E)^T\Big)y_t\,,
\end{aligned}
\tag{77}
$$

which gives the optimal trade

$$\begin{aligned} y_t \;&=\; \mathcal{S}_E\Bigg(\mathbf{F}\bar{d} - \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d\xi_t + (\bar{P} \;+\; \Pi_X^E\Big(\bar{\mathcal{X}}^* + \bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi\xi_t\Big) - e^r P_t) \\ &+ \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \alpha_E\Bigg(\mathcal{V}_\xi^E + \mathcal{W}_{\xi,\mathcal{X}}^E\Big(\bar{\mathcal{X}}^* + \bar{\mathcal{A}}_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \bar{\mathcal{A}}_\xi\xi_t\Big)\Bigg)\Bigg)\Bigg) \\ &=\; \hat{Q}_E^* \;+\; \hat{\Theta}_{\mathcal{X}}^E\bar{\mathcal{X}}_{t-1+} \;+\; \hat{\Theta}_\xi^E\xi_t - \mathcal{S}_E e^r P_t \\ &=\; \hat{Q}_E^* \;+\; \hat{\Theta}_{\mathcal{X}}^E\bar{\mathcal{X}}_{t-1+} \;+\; \hat{\Theta}_\xi^E\xi_t - \mathcal{S}_E e^r(\bar{P} \;+\; \Pi_X^E\bar{\mathcal{X}}_{t-1+} \;+\; \Pi_\xi^E\xi_t) \\ &=\; Q_E^* \;+\; \Theta_{\mathcal{X}}^E\bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\xi^E\xi_t \end{aligned} \tag{78}$$

where

$$\begin{aligned} \mathcal{S}_E^{-1} &= (1-e^{-r})\alpha_E\Big(\mathbf{F}\Sigma_d\mathbf{F}^T + \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}(\Pi_\xi^E)^T\Big) \\ Q_E^* \;&=\; \mathcal{S}_E\Bigg(\mathbf{F}\bar{d} - e^r\bar{P} \;+\; \Pi_X^E\bar{\mathcal{X}}^* + \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \alpha_E\Bigg(\mathcal{V}_\xi^E + \mathcal{W}_{\xi,\mathcal{X}}^E\bar{\mathcal{X}}^*\Bigg)\Bigg)\Bigg) \\ \Theta_{\mathcal{X}}^E \;&=\; \mathcal{S}_E\Big(\Pi_X^E - \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\alpha_E\mathcal{W}_{\xi,\mathcal{X}}^E\Big)\bar{\mathcal{A}}_{\mathcal{X}} - \mathcal{S}_E e^r\Pi_X^E \\ \Theta_\xi^E \;&=\; \mathcal{S}_E\Big(\Pi_X^E - \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\alpha_E\mathcal{W}_{\xi,\mathcal{X}}^E\Big)\bar{\mathcal{A}}_\xi \;-\mathcal{S}_E(e^r\Pi_\xi^E + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d). \end{aligned} \tag{79}$$

This defines a fixed point system for $\mathcal{W}^E$ and $\mathcal{V}^E$:

$$\begin{aligned} e^r\begin{pmatrix}\mathcal{W}_{\mathcal{X}}^E & \mathcal{W}_{\mathcal{X},\xi}^E \\ \mathcal{W}_{\xi,\mathcal{X}}^E & \mathcal{W}_\xi^E\end{pmatrix} \;&=\; -(\Theta^E)^T(\mathcal{S}^E)^{-1}\Theta^E + (\bar{\mathcal{A}})^T\mathcal{W}_{\mathcal{X}}^E\bar{\mathcal{A}} + \alpha_E(1-e^{-r})^2\begin{pmatrix}0 & 0 \\ 0 & \Sigma_d\end{pmatrix} \\ e^r\mathcal{V}^E \;&=\; \begin{pmatrix}0 \\ (1-e^{-r})\mathbf{F}\bar{d}\end{pmatrix} + \bar{\mathcal{A}}^T\Bigg(\mathcal{V}_{\mathcal{X}}^E - \mathcal{W}_{\mathcal{X}}^E\bar{\mathcal{X}}^* + (\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}(\Sigma_\xi^{-1}\bar{\xi} - \alpha_E(\mathcal{V}_\xi^E + \mathcal{W}_{\xi,\mathcal{X}}^E\bar{X}^*))\Bigg) \\ &+ (\Theta^E)^T\mathcal{S}_E^{-1}Q_E^*. \end{aligned} \tag{80}$$

This defines the optimal policy matrices as a function of just three matrices $(\bar{\mathcal{A}}, \Pi_X^E,\ \Pi_\xi^E)$.

Finally, the transversality condition takes the form

$$\lim_{T\to\infty} E_t[e^{-\beta(T-t)}V_T] \;=\; 0\,.$$

It is straightforward to show that, when $\mathcal{W}_{\mathcal{X}}^E$ is negative definite, any exponential growth in the $\bar{\mathcal{X}}_{t-1+}$ process is always offset by the quadratic term $\bar{\mathcal{X}}_{t-1+}^T\mathcal{W}_{\mathcal{X}}^E\bar{\mathcal{X}}_{t-1+}$, and the claim follows if β is sufficiently large.

When δ is small, we get

$$
\begin{aligned}
&\mathcal{S}_E^{-1} \approx (1-e^{-r})\alpha_E\Big(\Sigma_E + \Pi_\xi^E \Sigma_\xi (\Pi_\xi^E)^T\Big) \\
&\mathcal{S}_E \approx ((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi_\xi^E\Sigma_\xi(\Pi_\xi^E)^T\Sigma_E^{-1}) \\
&Q_E^* \;=\; \mathcal{S}_E\left(\mathbf{F}\bar{d} \;+\; \Pi_X^E\bar{\mathcal{X}}^* + \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\left(\Sigma_\xi^{-1}\bar{\xi} - \alpha_E\left(\mathcal{V}_\xi^E + \mathcal{W}_{\xi,\mathcal{X}}^E\bar{\mathcal{X}}^*\right)\right)\right) \\
&\Theta_{\mathcal{X}}^E \;\approx\; ((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi_\xi^E\Sigma_\xi(\Pi_\xi^E)^T\Sigma_E^{-1})\Big(\Big(\Pi_X^E - \delta\Pi_\xi^E\Sigma_\xi\alpha_E\mathcal{W}_{\xi,\mathcal{X}}^E\Big)\bar{\mathcal{A}}_{\mathcal{X}} - e^r\Pi_X^E\Big) \\
&\Theta_\xi^E \;\approx\; ((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi_\xi^E\Sigma_\xi(\Pi_\xi^E)^T\Sigma_E^{-1})\Big(\Big(\Pi_X^E - \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\alpha_E\mathcal{W}_{\xi,\mathcal{X}}^E\Big)\bar{\mathcal{A}}_\xi \\
&-(e^r\Pi_\xi^E + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big)
\end{aligned}
\tag{81}
$$

D APs: Proof of Proposition 5.3

I will use x, y to denote individual AP portfolio holdings at time t. We first compute everything conditional on the results of time t trade:

$$
\begin{aligned}
&\left(1-e^{-r}\right)\hat{M}_{l,t+1} \;+\; \mathcal{V}_{\mathcal{X}}^T\mathcal{X}_{l,t+} \;-\; 0.5\bar{\mathcal{X}}_{t+}^T\mathcal{W}\bar{\mathcal{X}}_{t+} \\
&+\; (\mathcal{V}_{\xi}^T\xi_{t+1} - 0.5\xi_{t+1}^T\mathcal{W}_{\xi}\xi_{t+1} - \xi_{t+1}^T\mathcal{W}_{\xi,\mathcal{X}}\bar{\mathcal{X}}_{t+}) \\
&=\; \left(1-e^{-r}\right)\Bigg(e^r(\hat{M}_{l,t} - c_{l,t} - 0.5Z_C^T\Lambda_C Z_C - 0.5Z_I^T\Lambda_I Z_I - P_t\cdot y \;-\; p_t\cdot x) \\
&-\; \mathbf{F}^T Z_C\cdot p_{t+1} \;+\; y_{t+}\cdot(\mathbf{F}\mathbf{d}_{t+1} + P_{t+1}) \;+\; x_{t+}\cdot(\mathbf{d}_{t+1} + p_{t+1})\Bigg) \\
&+\; \mathcal{V}^T\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right) \;-\; 0.5\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)^T\mathcal{W}_{\mathcal{X}}\left(X_0^+ \;+\mathcal{B}\mathcal{X}_t\right) \\
&+\; (\mathcal{V}_{\xi}^T\xi_{t+1} - 0.5\xi_{t+1}^T\mathcal{W}_{\xi}\xi_{t+1} - \xi_{t+1}^T\mathcal{W}_{\xi,\mathcal{X}}\bar{\mathcal{X}}_{t+}) \\
&=\; \left(1-e^{-r}\right)\Bigg(e^r\left(\hat{M}_{l,t} - c_{l,t} - 0.5Z_C^T\Lambda_C Z_C - 0.5Z_I^T\Lambda_I Z_I - P_t\cdot y \;-\; p_t\cdot x\right) \\
&+\left(\bar{p} \;+\; \Pi_X^D\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right) \;+\; \Pi_{\varepsilon}^D(\Theta_{\varepsilon}^D\varepsilon_{t+1})\right)\cdot\left(x - \mathbf{F}^T(Z_I + Z_C)\right) \\
&+\; \left(x + \mathbf{F}^T(y + Z_C)\right)\cdot\mathbf{d}_{t+1} \;+\; (y + Z_I + Z_C)\cdot\left(\bar{P} \;+\; \Pi_X^E\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)\right)\Bigg) \\
&+\xi_{t+1}^T\left(\mathcal{V}_{\xi} + (1-e^{-r})\left[(\Pi_{\xi}^D)^T\left(x - \mathbf{F}^T(Z_I + Z_C)\right) + (\Pi_{\xi}^E)^T\left(y + Z_I + Z_C\right)\right] - \mathcal{W}_{\xi,\mathcal{X}}\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)\right) \\
&-0.5\xi_{t+1}^T\mathcal{W}_{\xi}\xi_{t+1} \\
&+\; \mathcal{V}_{\mathcal{X}}^T\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right) \;-\; 0.5\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)^T\mathcal{W}_{\mathcal{X}}\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)
\end{aligned}
\tag{82}
$$

Now, we can integrate out the $\varepsilon_{t+1}, \xi_{t+1}$ shocks as well as the d_{t+1} shock:

$$
\begin{aligned}
&- E_{t,\xi}\left[\exp\left(-\alpha\left((1-e^{-r})M_{l,t+1} \;+\; \mathcal{V}^T\bar{\mathcal{X}}_{t+} \;-\; 0.5\bar{\mathcal{X}}_{t+}^T\mathcal{W}\bar{\mathcal{X}}_{t+}\right)\right)\right] \\
&= \; -e^{-\alpha((e^r-1)(M_{l,t}-c_{l,t}-0.5Z_C^T\Lambda_C Z_C-0.5Z_I^T\Lambda_I Z_I-P_t\cdot y \;-\; p_t\cdot x))} \\
&\times \exp\left(-\alpha\left(\mathcal{V}_{\mathcal{X}}^T\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right) \;-\; 0.5\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)^T\mathcal{W}_{\mathcal{X}}\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)\right)\right) \\
&\times \exp\Bigg(-\alpha(1-e^{-r})\left(\bar{p}+\Pi_X^D\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)+\Pi_\varepsilon^D\bar{\varepsilon}\right)^T(x-\mathbf{F}^T(Z_I+Z_C)) \\
&-\alpha(1-e^{-r})\cdot\left(\bar{P} \;+\; \Pi_X^E\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)\right)^T(y+Z_I+Z_C) \\
&+0.5\alpha^2(1-e^{-r})^2(x-\mathbf{F}^T(Z_I+Z_C))^T\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T(x-\mathbf{F}^T(Z_I+Z_C))\Bigg) \\
&\times \exp\left(-\alpha(1-e^{-r})\bar{d}^T(x+\mathbf{F}^T(y+Z_C))+0.5\alpha^2(1-e^{-r})^2(x+\mathbf{F}^T(y+Z_C))^T\Sigma_d(x+\mathbf{F}^T(y+Z_C))\right) \\
&\times \frac{\det((\Sigma_\xi^{-1}-\alpha\mathcal{W}_\xi)^{-1})\,e^{0.5(\bar{\mathcal{V}}_\xi^T(\Sigma_\xi^{-1}-\alpha\mathcal{W}_\xi)^{-1}\bar{\mathcal{V}}_\xi-\bar{\xi}^T\Sigma_\xi^{-1}\bar{\xi})}}{(\det\Sigma_\xi)^{1/2}} \\
&= \; -e^{-\alpha((e^r-1)(M_{l,t}-c_{l,t}-P_t\cdot y \;-\; p_t\cdot x))} \\
&\times \exp\left(v^+-\alpha\left((\mathcal{P}_0+\mathcal{P}_{\mathcal{X}}\mathcal{X}_t)\cdot\begin{pmatrix}y\\x\end{pmatrix}+\mathcal{Y}_0\cdot\mathcal{X}_t-0.5\mathcal{X}_t^T\mathcal{R}_{\mathcal{X}}\mathcal{X}_t-0.5\begin{pmatrix}y\\x\end{pmatrix}^T\mathcal{R}_{[yx]}\begin{pmatrix}y\\x\end{pmatrix}\right)\right) \\
&\times \exp\left(-\alpha\left((\widehat{\mathcal{G}}+\mathcal{G}_{\mathcal{X}}\mathcal{X}_t+\mathcal{G}_x x+\mathcal{G}_y y)\cdot\begin{pmatrix}Z_I\\Z_C\end{pmatrix}-0.5\begin{pmatrix}Z_I\\Z_C\end{pmatrix}^T\mathcal{R}_Z\begin{pmatrix}Z_I\\Z_C\end{pmatrix}\right)\right)
\end{aligned}
\tag{83}
$$

where I have defined

$$
\begin{aligned}
\bar{\mathcal{V}}_\xi = &\left(\Sigma_\xi^{-1}\bar{\xi}-\alpha\left(\mathcal{V}_\xi+(1-e^{-r})\left[(\Pi_\xi^D)^T\left(x-\mathbf{F}^T(Z_I+Z_C)\right)+(\Pi_\xi^E)^T\left(y+Z_I+Z_C\right)\right]\right.\right. \\
&\left.\left.-\mathcal{W}_{\xi,\mathcal{X}}\left(X_0^+ \;+\; \mathcal{B}\mathcal{X}_t\right)\right)\right)
\end{aligned}
\tag{84}
$$

Here,

$$
\begin{aligned}
\mathcal{R}_Z \;=\; &(e^r-1)\begin{pmatrix}\Lambda_I & 0\\ 0 & \Lambda_C\end{pmatrix} \;+\; (1-e^{-r})^2\alpha\mathbf{F}\Pi_\varepsilon^D\Sigma_\varepsilon(\mathbf{F}\Pi_\varepsilon^D)^T\otimes\begin{pmatrix}\mathrm{Id} & \mathrm{Id}\\ \mathrm{Id} & \mathrm{Id}\end{pmatrix} \\
&+(1-e^{-r})^2\alpha\begin{pmatrix}0 & 0\\ 0 & \mathbf{F}\Sigma_d\mathbf{F}^T\end{pmatrix} \;+\; (1-e^{-r})^2\alpha\Pi_*(\Sigma_\xi^{-1}-\alpha\mathcal{W}_\xi)^{-1}\Pi_*^T\otimes\begin{pmatrix}\mathrm{Id} & \mathrm{Id}\\ \mathrm{Id} & \mathrm{Id}\end{pmatrix}
\end{aligned}
\tag{85}
$$

where

$$
\Pi_* \;\equiv\; \Pi_\xi^E-\mathbf{F}\Pi_\xi^D\,.
$$

Furthermore,

$$
\begin{aligned}
&(\widehat{\mathcal{G}} + \mathcal{G}_{\mathcal{X}}\mathcal{X}_t + \mathcal{G}_x x + \mathcal{G}_y y) \\
&= (1-e^{-r})\Big((\bar{P} - \mathbf{F}\bar{p}) + (\Pi_X^E - \mathbf{F}\Pi_X^D)(X_0^+ + \mathcal{B}\mathcal{X}_t) - \mathbf{F}\Pi_\varepsilon^D\bar{\varepsilon}\Big) \otimes \begin{pmatrix}1\\1\end{pmatrix} \\
&+\ \alpha(1-e^{-r})^2\mathbf{F}\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T x \otimes \begin{pmatrix}1\\1\end{pmatrix}\ + (1-e^{-r})\begin{pmatrix}0\\ \mathbf{F}\bar{d}\end{pmatrix} \\
&- \alpha(1-e^{-r})^2\begin{pmatrix}0\\ \mathbf{F}\Sigma_d(x+\mathbf{F}^T y)\end{pmatrix} \\
&+ (1-e^{-r})\Pi_*(\Sigma_\xi^{-1} - \alpha\mathcal{W}_\xi)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \alpha\Bigg(\mathcal{V}_\xi + (1-e^{-r})\Big[(\Pi_\xi^D)^T x + (\Pi_\xi^E)^T y\Big] \\
&- \mathcal{W}_{\xi,\mathcal{X}}\Big(X_0^+\ +\ \mathcal{B}\mathcal{X}_t\Big)\Bigg)\Bigg) \otimes \begin{pmatrix}1\\1\end{pmatrix}
\end{aligned} \tag{86}
$$

whereas

$$
\begin{aligned}
\mathcal{P}_0 + \mathcal{P}_{\mathcal{X}}\mathcal{X}_t \ &=\ (1-e^{-r})\begin{pmatrix}\bar{P}\ +\ \Pi_X^E\Big(X_0^+\ +\ \mathcal{B}\mathcal{X}_t\Big) + \mathbf{F}\bar{d} \\ \bar{p} + \Pi_X^D\Big(X_0^+\ +\ \mathcal{B}\mathcal{X}_t\Big) + \Pi_\varepsilon^D\bar{\varepsilon} + \bar{d}\end{pmatrix} \\
&+ (1-e^{-r})\begin{pmatrix}\Pi_\xi^E\\ \Pi_\xi^D\end{pmatrix}(\Sigma_\xi^{-1} - \alpha\mathcal{W}_\xi)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \alpha\Bigg(\mathcal{V}_\xi - \mathcal{W}_{\xi,\mathcal{X}}\Big(X_0^+\ +\ \mathcal{B}\mathcal{X}_t\Big)\Bigg)\Bigg)
\end{aligned} \tag{87}
$$

and

$$
\begin{aligned}
\mathcal{R}_{[yx]}\ &=\ \alpha(1-e^{-r})^2\begin{pmatrix}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \\ \Sigma_d\mathbf{F}^T & \Sigma_d + \Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T\end{pmatrix} \\
&+ \alpha(1-e^{-r})^2\begin{pmatrix}\Pi_\xi^E\\ \Pi_\xi^D\end{pmatrix}(\Sigma_\xi^{-1} - \alpha\mathcal{W}_\xi)^{-1}\begin{pmatrix}\Pi_\xi^E\\ \Pi_\xi^D\end{pmatrix}^T
\end{aligned} \tag{88}
$$

and

$$
\mathcal{Y}_0\ =\ \mathcal{B}^T\mathcal{V}_{\mathcal{X}} - \mathcal{B}^T\mathcal{W}_{\mathcal{X}}X_0^+ - (\mathcal{W}_{\xi,\mathcal{X}}\mathcal{B})^T(\Sigma_\xi^{-1} - \alpha\mathcal{W}_\xi)^{-1}\Bigg(\Sigma_\xi^{-1}\bar{\xi} - \alpha\Bigg(\mathcal{V}_\xi - \mathcal{W}_{\xi,\mathcal{X}}X_0^+\Bigg)\Bigg).
$$

and

$$
\mathcal{R}_{\mathcal{X}}\ =\ \mathcal{B}^T\mathcal{W}_{\mathcal{X}}\mathcal{B} + \alpha(\mathcal{W}_{\xi,\mathcal{X}}\mathcal{B})^T(\Sigma_\xi^{-1} - \alpha\mathcal{W}_\xi)^{-1}\mathcal{W}_{\xi,\mathcal{X}}\mathcal{B} \tag{89}
$$

This gives

$$
\begin{pmatrix}Z_I\\ Z_C\end{pmatrix}\ =\ \mathcal{R}_Z^{-1}(\widehat{\mathcal{G}} + \mathcal{G}_{\mathcal{X}}\mathcal{X}_t + \mathcal{G}_x x + \mathcal{G}_y y) \tag{90}
$$

The optimization problem (after integrating out the shock ε_t) takes the form

$$
\begin{aligned}
&(e^r-1)\Bigg(-(\bar{p} \;+\; \Pi_X^D \mathcal{X}_{t-1+} \;+\; \Pi_\varepsilon^D(\Theta_\varepsilon^D \bar{\varepsilon}+\theta_\xi^A \xi_t))\cdot x - 0.5\alpha(e^r-1)x^T \Pi_\varepsilon^D \Sigma_\varepsilon (\Pi_\varepsilon^D)^T x \\
&-(\bar{P}+\Pi_X^E \bar{\mathcal{X}}_{t-1+}+\Theta_\xi^E \xi_t)\cdot y\Bigg) \\
&+\left(\mathcal{P}_0+\mathcal{P}_{\mathcal{X}}\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)\right)\cdot\begin{pmatrix}y\\ x\end{pmatrix}+\mathcal{Y}_0\cdot\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right) \\
&-0.5\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)^T \mathcal{R}_{\mathcal{X}}\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)-0.5\begin{pmatrix}y\\ x\end{pmatrix}^T\mathcal{R}_{[yx]}\begin{pmatrix}y\\ x\end{pmatrix} \\
&+0.5\left(\widehat{\mathcal{G}}+\mathcal{G}_{\mathcal{X}}\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)+\mathcal{G}_x x+\mathcal{G}_y y\right)^T\mathcal{R}_Z^{-1} \\
&\times\left(\widehat{\mathcal{G}}+\mathcal{G}_{\mathcal{X}}\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)+\mathcal{G}_x x+\mathcal{G}_y y\right) \\
&=\;\mathcal{Y}^+\cdot\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\;-\;0.5\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}^T\mathcal{R}^+\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix} \\
&+\;\left(\mathcal{G}_0^*+\mathcal{G}^+\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)\cdot\begin{pmatrix}y\\ x\end{pmatrix}-0.5\begin{pmatrix}y\\ x\end{pmatrix}^T\mathcal{R}_{\mathcal{G}}\begin{pmatrix}y\\ x\end{pmatrix}
\end{aligned}
\tag{91}
$$

Here,

$$
\begin{aligned}
\mathcal{R}_{\mathcal{G}} \;&=\; \mathcal{R}_{[yx]}-\begin{pmatrix}\mathcal{G}_y^T\\ \mathcal{G}_x^T\end{pmatrix}\mathcal{R}_Z^{-1}(\mathcal{G}_y\;\mathcal{G}_x)+\alpha(e^r-1)^2\begin{pmatrix}0 & 0\\ 0 & \Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T\end{pmatrix} \\
\mathcal{G}_0^*+\mathcal{G}^+\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix} \;&=\; -(e^r-1)\begin{pmatrix}\bar{P}+\Pi_X^E\bar{\mathcal{X}}_{t-1+}+\Pi_\xi^E\xi_t\\ \bar{p}\;+\;\Pi_X^D\mathcal{X}_{t-1+}\;+\;\Pi_\varepsilon^D\bar{\varepsilon}+\Pi_\xi^D\xi_t\end{pmatrix} \\
&+\mathcal{P}_0+\mathcal{P}_{\mathcal{X}}\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right) \\
&+\begin{pmatrix}\mathcal{G}_y^T\\ \mathcal{G}_x^T\end{pmatrix}\mathcal{R}_Z^{-1}\left(\widehat{\mathcal{G}}+\mathcal{G}_{\mathcal{X}}\left(X_0^- + \bar{\mathcal{A}}^-\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)\right) \\
\mathcal{Y}^+ \;&=\; (\bar{\mathcal{A}}^-)^T\mathcal{Y}_0-(\bar{\mathcal{A}}^-)^T\mathcal{R}_{\mathcal{X}}X_0^-+(\mathcal{G}_{\mathcal{X}}\bar{\mathcal{A}}^-)^T\mathcal{R}_Z^{-1}\left(\widehat{\mathcal{G}}+\mathcal{G}_{\mathcal{X}}X_0^-\right) \\
\mathcal{R}^+ \;&=\; (\bar{\mathcal{A}}^-)^T\mathcal{R}_{\mathcal{X}}\bar{\mathcal{A}}^- - (\mathcal{G}_{\mathcal{X}}\bar{\mathcal{A}}^-)^T\mathcal{R}_Z^{-1}\mathcal{G}_{\mathcal{X}}\bar{\mathcal{A}}^-\,.
\end{aligned}
\tag{92}
$$

The optimal position is then given by

$$
\begin{pmatrix}y\\ x\end{pmatrix}\;=\;\mathcal{R}_{\mathcal{G}}^{-1}\left(\mathcal{G}_0^*+\mathcal{G}^+\begin{pmatrix}\bar{\mathcal{X}}_{t-1+}\\ \xi_t\end{pmatrix}\right)
\tag{93}
$$

and the outcome of this maximization is

$$
\begin{aligned}
&\mathcal{Y}^+ \cdot \begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix} \;-\; 0.5 \begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix}^T \mathcal{R}^+ \begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix} \\
&+\; 0.5 \left(\mathcal{G}_0^* + \mathcal{G}^+ \begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix} \right)^T \mathcal{R}_{\mathcal{G}}^{-1} \left(\mathcal{G}_0^* + \mathcal{G}^+ \begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix} \right) \\
&=\; 0.5(\mathcal{G}_0^*)^T \mathcal{R}_{\mathcal{G}}^{-1} \mathcal{G}_0^* \;+\; (\mathcal{Y}^+ + (\mathcal{G}^+)^T \mathcal{R}_{\mathcal{G}}^{-1} \mathcal{G}_0^*) \cdot \begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix} - 0.5 \left(\mathcal{R}^+ - (\mathcal{G}^+)^T \mathcal{R}_{\mathcal{G}}^{-1} \mathcal{G}^+ \right) .
\end{aligned}
\tag{94}
$$

This gives the HJB fixed point system

$$
\begin{aligned}
\mathcal{W} \;&=\; e^{-r} \left((\bar{\mathcal{A}}^-)^T \mathcal{R}_{\mathcal{X}} \bar{\mathcal{A}}^- - (\mathcal{G}_{\mathcal{X}} \bar{\mathcal{A}}^-)^T \mathcal{R}_Z^{-1} \mathcal{G}_{\mathcal{X}} \bar{\mathcal{A}}^- - (\mathcal{G}^+)^T \mathcal{R}_{\mathcal{G}}^{-1} \mathcal{G}^+ \right) \\
\mathcal{V} \;&=\; e^{-r} (\mathcal{Y}^+ + (\mathcal{G}^+)^T \mathcal{R}_{\mathcal{G}}^{-1} \mathcal{G}_0^*)
\end{aligned}
\tag{95}
$$

Finally, the transversality condition takes the form

$$
\lim_{T \to \infty} E_t[e^{-\beta(T-t)} V_T] \;=\; 0\,.
$$

It is straightforward to show that, when $\mathcal{W}_{\mathcal{X}}$ is negative definite, any exponential growth in the $\bar{\mathcal{X}}_{t-1+}$ process is always offset by the quadratic term $\bar{\mathcal{X}}_{t-1+}^T \mathcal{W}_{\mathcal{X}} \bar{\mathcal{X}}_{t-1+}$, and the claim follows if β is sufficiently large.

E Small Demand Shock Volatility: Proof of Theorem 6.1

We have, keeping only first order terms in δ :

$$
\begin{aligned}
\mathcal{R}_{\mathcal{X}} \;&\approx\; (\mathcal{B}(0) + \delta \mathcal{B}(1))^T (\mathcal{W}_{\mathcal{X}}(0) + \delta \mathcal{W}_{\mathcal{X}}(1))(\mathcal{B}(0) + \delta \mathcal{B}(1)) \\
&+ \alpha(\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(0) + \delta(\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(1) + \mathcal{W}_{\xi,\mathcal{X}}(1)\mathcal{B}(0)))^T \delta(\Sigma_\xi^{-1} - \delta\alpha \mathcal{W}_\xi)^{-1} \\
&\times (\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(0) + \delta(\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(1) + \mathcal{W}_{\xi,\mathcal{X}}(1)\mathcal{B}(0))) \\
&\approx \mathcal{B}(0)^T \mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(0) + \delta \Big(\mathcal{B}(1)\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(0) + \mathcal{B}(0)\mathcal{W}_{\mathcal{X}}(1)\mathcal{B}(0) + \mathcal{B}(0)\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(1) \\
&+ \alpha(\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(0))^T \Sigma_\xi \mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(0) \Big)
\end{aligned}
\tag{96}
$$

To compute the matrix $\bar{\mathcal{A}}^-$ describing the change in the total holdings, we note that, by market clearing, we have

$$
\begin{aligned}
y_t^A + y_t^E \;&=\; y_{t-1+}^A + y_{t-1}^E \\
x_t^A + x_t^D \;&=\; x_{t-1+}^A + x_{t-1}^D \;-\; \mathbf{F} Z_{C,t-1+}
\end{aligned}
\tag{97}
$$

Now, we have

$$\mathcal{R}_Z \approx \begin{pmatrix} (e^r-1)\Lambda_I & 0 \\ 0 & (e^r-1)\Lambda_C + (1-e^{-r})^2\alpha\mathbf{F}\Sigma_d\mathbf{F}^T \end{pmatrix} + \delta(1-e^{-r})^2\alpha\Bigg(\mathbf{F}\Pi_\varepsilon^D(0)\Sigma_\varepsilon(\mathbf{F}\Pi_\varepsilon^D(0))^T \otimes \begin{pmatrix} \mathrm{Id} & \mathrm{Id} \\ \mathrm{Id} & \mathrm{Id} \end{pmatrix} + \Pi_*^T(0)\Sigma_\xi\Pi_*(0) \otimes \begin{pmatrix} \mathrm{Id} & \mathrm{Id} \\ \mathrm{Id} & \mathrm{Id} \end{pmatrix}\Bigg) \tag{98}$$

However, since in the limit as $\delta \to$ ETFs become redundant, we have $\Pi_*(0) = 0$ and hence the second term vanishes. Therefore,

$$\mathcal{R}_Z^{-1} \approx \mathcal{R}_Z^{-1}(0) - \delta\mathcal{R}_Z^{-1}(0)(1-e^{-r})^2\alpha\mathbf{F}\Pi_\varepsilon^D(0)\Sigma_\varepsilon(\mathbf{F}\Pi_\varepsilon^D(0))^T \otimes \begin{pmatrix} \mathrm{Id} & \mathrm{Id} \\ \mathrm{Id} & \mathrm{Id} \end{pmatrix} \mathcal{R}_Z^{-1}(0)$$

The matrix $\mathcal{R}_Z^{-1}(0)$ is the only place where ETF transaction costs enter equilibrium equations. Now,

$$\mathcal{G}_{\mathcal{X}} = \delta(1-e^{-r})(\Pi_X^E(1) - \mathbf{F}\Pi_X^D(1))\mathcal{B} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{99}$$

where I have used that $\Pi_X^E(0) - \mathbf{F}\Pi_X^D(0) = 0$ and $\Pi_*(0) = 0$ because ETFs are redundant in the limit. Similarly,

$$(\mathcal{G}_y\ \mathcal{G}_x) \approx -\alpha(1-e^{-r})^2 \begin{pmatrix} 0 & 0 \\ \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \end{pmatrix} + \delta\alpha(1-e^{-r})^2 \begin{pmatrix} 0 & \mathbf{F}\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T \\ 0 & \mathbf{F}\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T \end{pmatrix}. \tag{100}$$

Similarly, we have

$$\begin{aligned}\mathcal{R}_{\mathcal{G}} \approx\ & \alpha(1-e^{-r})^2 \begin{pmatrix} \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \\ \Sigma_d\mathbf{F}^T & \Sigma_d \end{pmatrix} \\ & + \delta\alpha((1-e^{-r})^2 + (e^r-1)^2) \begin{pmatrix} 0 & 0 \\ 0 & \Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T \end{pmatrix} \\ & + \delta\alpha(1-e^{-r})^2 \begin{pmatrix} \mathbf{F}\Pi_\xi^D \\ \Pi_\xi^D \end{pmatrix}(0)\Sigma_\xi \begin{pmatrix} \mathbf{F}\Pi_\xi^D \\ \Pi_\xi^D \end{pmatrix}(0)^T - \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix}(0)\mathcal{R}_Z(0)^{-1}(\mathcal{G}_y\ \mathcal{G}_x)(0) \\ & - \delta\Bigg(- \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix}(0)\mathcal{R}_Z^{-1}(0)(1-e^{-r})^2\alpha\mathbf{F}\Pi_\varepsilon^D(0)\Sigma_\varepsilon(\mathbf{F}\Pi_\varepsilon^D(0))^T \otimes \begin{pmatrix} \mathrm{Id} & \mathrm{Id} \\ \mathrm{Id} & \mathrm{Id} \end{pmatrix} \mathcal{R}_Z^{-1}(0)(\mathcal{G}_y\ \mathcal{G}_x)(0) \\ & + \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix}(1)\mathcal{R}_Z(0)^{-1}(\mathcal{G}_y\ \mathcal{G}_x)(0) + \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix}(0)\mathcal{R}_Z(0)^{-1}(\mathcal{G}_y\ \mathcal{G}_x)(1)\Bigg) \end{aligned} \tag{101}$$

where we have used that in the limit as $\delta \to 0$ we have $\Pi_\xi^E = \mathbf{F}\Pi_\xi^D$ because ETFs are redundant. Furthermore,

$$
\begin{aligned}
&\begin{pmatrix}\mathcal{G}_y^T\\ \mathcal{G}_x^T\end{pmatrix}(0)\mathcal{R}_Z(0)^{-1}(\mathcal{G}_y\ \mathcal{G}_x)(0)\\
&= \alpha(1-e^{-r})^2\begin{pmatrix}0 & \mathbf{F}\Sigma_d\mathbf{F}^T\\ 0 & \Sigma_d\mathbf{F}^T\end{pmatrix}\begin{pmatrix}(e^r-1)^{-1}\Lambda_I^{-1} & 0\\ 0 & \bar{\Lambda}_C^{-1}\end{pmatrix}\begin{pmatrix}0 & 0\\ \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d\end{pmatrix} \qquad (102)\\
&= \alpha(1-e^{-r})^2\begin{pmatrix}\mathbf{F}\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\\ \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\end{pmatrix}
\end{aligned}
$$

where we have define

$$
\bar{\Lambda}_C \equiv ((e^r-1)\Lambda_C + (1-e^{-r})^2\alpha\mathbf{F}\Sigma_d\mathbf{F}^T)/(\alpha(1-e^{-r})^2)\,.
$$

To compute the inverse we will need the following auxiliary result.

Lemma E.1 *We have*

$$
\begin{aligned}
&\begin{pmatrix}\delta\rho_{11} & \delta\rho_{12}\\ \delta\rho_{12}^T & A+\delta\rho_{22}\end{pmatrix}^{-1}\\
&= \begin{pmatrix}\delta^{-1}(\rho_{11}-\delta\rho_{12}(A+\delta\rho_{22})^{-1}\rho_{12}^T))^{-1} & -\rho_{11}^{-1}\rho_{12}(A+\delta\rho_{22}-\delta\rho_{12}^T\rho_{11}^{-1}\rho_{12})^{-1}\\ -(A+\delta\rho_{22}-\delta\rho_{12}^T\rho_{11}^{-1}\rho_{12})^{-1}\rho_{12}^T\rho_{11}^{-1} & (A+\delta\rho_{22}-\delta\rho_{12}^T\rho_{11}^{-1}\rho_{12})^{-1}\end{pmatrix}\\
&\approx \begin{pmatrix}\delta^{-1}\rho_{11}^{-1}+\rho_{11}^{-1}\rho_{12}A^{-1}\rho_{12}^T\rho_{11}^{-1} & -\rho_{11}^{-1}\rho_{12}A^{-1}\\ -A^{-1}\rho_{12}^T\rho_{11}^{-1} & A^{-1}\end{pmatrix} \qquad (103)\\
&+\delta\begin{pmatrix}0 & \rho_{11}^{-1}\rho_{12}A^{-1}(\rho_{22}-\rho_{12}^T\rho_{11}^{-1}\rho_{12})A^{-1}\\ A^{-1}(\rho_{22}-\rho_{12}^T\rho_{11}^{-1}\rho_{12})A^{-1}\rho_{12}^T\rho_{11}^{-1} & -A^{-1}(\rho_{22}-\rho_{12}^T\rho_{11}^{-1}\rho_{12})A^{-1}\end{pmatrix}\\
&+\begin{pmatrix}O(\delta) & O(\delta^2)\\ O(\delta^2) & O(\delta^2)\end{pmatrix}
\end{aligned}
$$

Now, to proceed further, we will assume that ETF are simply a subset of the existing assets. [41] Namely, we assume that

$$
\mathbf{F} = (\mathrm{Id}\ 0)
$$

where we have partitioned the covariance matrix of the basket securities as

$$
\Sigma_d = \begin{pmatrix}\Sigma_E & \Sigma_{ED}\\ \Sigma_{DE} & \Sigma_D\end{pmatrix}
$$

[41] It is straightforward to show that this assumption is without loss of generality. Namely, one can write $\mathbf{F} = U(\mathrm{Id}\ 0)V$ where U, V are orthogonal matrices, and then show that U and V cancel out in the final expressions.

Then, the total covariance matrix can be rewritten as

$$\begin{pmatrix} \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \\ \Sigma_d\mathbf{F}^T & \Sigma_d \end{pmatrix} = \begin{pmatrix} \Sigma_E & \Sigma_E & \Sigma_{ED} \\ \Sigma_E & \Sigma_E & \Sigma_{ED} \\ \Sigma_{DE} & \Sigma_{DE} & \Sigma_D \end{pmatrix} = \begin{pmatrix} \mathrm{Id} & \mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_d \end{pmatrix} \begin{pmatrix} \mathrm{Id} & 0 \\ \mathbf{F}^T & \mathrm{Id} \end{pmatrix}$$

and, similarly,

$$\begin{aligned} & \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix}(0)\mathcal{R}_Z(0)^{-1}(\mathcal{G}_y\ \mathcal{G}_x)(0) \\ &= \alpha(1-e^{-r})^2 \begin{pmatrix} \mathbf{F}\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d \\ \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d \end{pmatrix} \\ &= \alpha(1-e^{-r})^2 \begin{pmatrix} \mathrm{Id} & \mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d \end{pmatrix} \begin{pmatrix} \mathrm{Id} & 0 \\ \mathbf{F}^T & \mathrm{Id} \end{pmatrix} \end{aligned} \tag{104}$$

Thus,

$$\begin{aligned} \mathcal{R}_{\mathcal{G}}^{-1} &= \left(\alpha(1-e^{-r})^2 \begin{pmatrix} \mathrm{Id} & \mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d + \Sigma_d \end{pmatrix} \begin{pmatrix} \mathrm{Id} & 0 \\ \mathbf{F}^T & \mathrm{Id} \end{pmatrix} + \delta\mathcal{R}_{\mathcal{G}}(1) \right)^{-1} \\ &= \begin{pmatrix} \mathrm{Id} & 0 \\ -\mathbf{F}^T & \mathrm{Id} \end{pmatrix} \left(\alpha(1-e^{-r})^2 \begin{pmatrix} 0 & 0 \\ 0 & -\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d + \Sigma_d \end{pmatrix} + \delta \begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \mathcal{R}_{\mathcal{G}}(1) \begin{pmatrix} \mathrm{Id} & 0 \\ -\mathbf{F}^T & \mathrm{Id} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \end{aligned} \tag{105}$$

I will denote

$$\begin{aligned} \widetilde{\Sigma}_\varepsilon &\equiv (\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T)(0) \\ \Delta_\xi &\equiv (\Pi_\xi^D\Sigma_\xi(\Pi_\xi^D)^T)(0)\,. \end{aligned} \tag{106}$$

Then, defining

$$\Delta_\Lambda \equiv \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\,,$$

we get

$$
\begin{aligned}
\mathcal{R}_{\mathcal{G}}(1) \; &= \alpha((1-e^{-r})^2 + (e^r-1)^2)\begin{pmatrix} 0 & 0 \\ 0 & \widetilde{\Sigma}_\varepsilon \end{pmatrix} \\
&+ \alpha(1-e^{-r})^2 \begin{pmatrix} \mathbf{F}\Delta_\xi\mathbf{F}^T & \mathbf{F}\Delta_\xi \\ \Delta_\xi\mathbf{F}^T & \Delta_\xi \end{pmatrix} \\
&- \alpha(1-e^{-r})^2\Bigg(- \begin{pmatrix} 0 & \mathbf{F}\Sigma_d\mathbf{F}^T \\ 0 & \Sigma_d\mathbf{F}^T \end{pmatrix} \begin{pmatrix} \Lambda_I^{-1} & 0 \\ 0 & \bar{\Lambda}_C^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T \\ \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T \end{pmatrix} \\
&\times \begin{pmatrix} \Lambda_I^{-1} & 0 \\ 0 & \bar{\Lambda}_C^{-1} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \end{pmatrix} \\
&- \begin{pmatrix} 0 & 0 \\ \widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \widetilde{\Sigma}_\varepsilon\mathbf{F}^T \end{pmatrix} \begin{pmatrix} \Lambda_I^{-1} & 0 \\ 0 & \bar{\Lambda}_C^{-1} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \end{pmatrix} \\
&- \begin{pmatrix} 0 & \mathbf{F}\Sigma_d\mathbf{F}^T \\ 0 & \Sigma_d\mathbf{F}^T \end{pmatrix} \begin{pmatrix} \Lambda_I^{-1} & 0 \\ 0 & \bar{\Lambda}_C^{-1} \end{pmatrix} \begin{pmatrix} 0 & \mathbf{F}\widetilde{\Sigma}_\varepsilon \\ 0 & \mathbf{F}\widetilde{\Sigma}_\varepsilon \end{pmatrix} \Bigg) \\
&= \alpha((1-e^{-r})^2 + (e^r-1)^2)\begin{pmatrix} 0 & 0 \\ 0 & \widetilde{\Sigma}_\varepsilon \end{pmatrix} \\
&+ \alpha(1-e^{-r})^2 \begin{pmatrix} \mathbf{F}\Delta_\xi\mathbf{F}^T & \mathbf{F}\Delta_\xi \\ \Delta_\xi\mathbf{F}^T & \Delta_\xi \end{pmatrix} \\
&+ \alpha(1-e^{-r})^2\Bigg(\begin{pmatrix} \mathbf{F}\Delta_\Lambda\mathbf{F}^T & \mathbf{F}\Delta_\Lambda \\ \Delta_\Lambda\mathbf{F}^T & \Delta_\Lambda \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{F}\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\widetilde{\Sigma}_\varepsilon \\ \widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d + \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\widetilde{\Sigma}_\varepsilon \end{pmatrix} \Bigg)
\end{aligned} \tag{107}
$$

Hence,

$$
\begin{aligned}
&\begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \mathcal{R}_{\mathcal{G}}(1) \begin{pmatrix} \mathrm{Id} & 0 \\ -\mathbf{F}^T & \mathrm{Id} \end{pmatrix} \\
&= \alpha((1-e^{-r})^2 + (e^r-1)^2)\begin{pmatrix} \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & -\mathbf{F}\widetilde{\Sigma}_\varepsilon \\ -\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \widetilde{\Sigma}_\varepsilon \end{pmatrix} \\
&+ \alpha(1-e^{-r})^2 \begin{pmatrix} 0 & 0 \\ 0 & \Delta_\xi + \Delta_\Lambda \end{pmatrix} \\
&+ \alpha(1-e^{-r})^2 \begin{pmatrix} 0 & -\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d \\ -\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d + \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\widetilde{\Sigma}_\varepsilon \end{pmatrix} \\
&= \alpha(1-e^{-r})^2 \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{12}^T & \rho_{22} \end{pmatrix}
\end{aligned} \tag{108}
$$

with

$$
\begin{aligned}
\rho_{11} &= \; (1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T \\
\rho_{12} &= -(1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon - \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d \\
\rho_{22} &= (1+e^{2r})\widetilde{\Sigma}_\varepsilon + \Delta_\xi + \Delta_\Lambda + (\widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d + \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\widetilde{\Sigma}_\varepsilon)\,.
\end{aligned} \tag{109}
$$

Let also

$$\mathbf{A} = -\Sigma_d \mathbf{F}^T \bar{\Lambda}_C^{-1} \mathbf{F} \Sigma_d + \Sigma_d\,.$$

Then,

$$\begin{aligned} &\left(\alpha(1-e^{-r})^2 \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{A} \end{pmatrix} \;+\; \delta \begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \mathcal{R}_{\mathcal{G}}(1) \begin{pmatrix} \mathrm{Id} & 0 \\ -\mathbf{F}^T & \mathrm{Id} \end{pmatrix} \right)^{-1} \\ &\approx\; (\alpha(1-e^{-r})^2)^{-1} \begin{pmatrix} \delta^{-1}(B_{11}(0) + \delta B_{11}(1)) & B_{12}(0) + \delta B_{12}(1) \\ B_{21}(0) + \delta B_{21}(1) & B_{22}(0) + \delta B_{22}(1) \end{pmatrix} \end{aligned} \tag{110}$$

where

$$\begin{aligned} B_{11}(0) &= \rho_{11}^{-1} \\ B_{11}(1) &= \rho_{11}^{-1} \rho_{12} \mathbf{A}^{-1} \rho_{12}^T \rho_{11}^{-1} \\ B_{12}(0) &= -\rho_{11}^{-1} \rho_{12} \mathbf{A}^{-1} \\ B_{22}(0) &= \mathbf{A}^{-1} \\ B_{12}(1) &= \rho_{11}^{-1} \rho_{12} \mathbf{A}^{-1} (\rho_{22} - \rho_{12}^T \rho_{11}^{-1} \rho_{12}) \mathbf{A}^{-1} \\ B_{22}(1) &= -\mathbf{A}^{-1} (\rho_{22} - \rho_{12}^T \rho_{11}^{-1} \rho_{12}) \mathbf{A}^{-1} \end{aligned} \tag{111}$$

Furthermore,

$$\begin{aligned} \mathcal{P}_{\mathcal{X}} \;&=\; (1-e^{-r}) \begin{pmatrix} \Pi_X^E \mathcal{B} \\ \Pi_X^D \mathcal{B} \end{pmatrix} + (1-e^{-r}) \begin{pmatrix} \Pi_\xi^E \\ \Pi_\xi^D \end{pmatrix} (\Sigma_\xi^{-1} - \alpha \mathcal{W}_\xi)^{-1} \alpha \mathcal{W}_{\xi,\mathcal{X}} \mathcal{B} \\ &\approx\; (1-e^{-r}) \begin{pmatrix} (\mathbf{F}\Pi_X^D(0) + \delta \Pi_X^E(1))\mathcal{B} \\ (\Pi_X^D(0) + \delta \Pi_X^D(1))\mathcal{B} \end{pmatrix} + (1-e^{-r}) \begin{pmatrix} \mathbf{F}\Pi_\xi^D(0) \\ \Pi_\xi^D(0) \end{pmatrix} \delta \Sigma_\xi \alpha \mathcal{W}_{\xi,\mathcal{X}} \mathcal{B} \end{aligned} \tag{112}$$

and therefore

$$\begin{aligned} \mathcal{G}^+ \;&=\; -(e^r - 1) \begin{pmatrix} \Pi_X^E & \Pi_\xi^E \\ \Pi_X^D & \Pi_\xi^D \end{pmatrix} + \mathcal{P}_{\mathcal{X}} \bar{\mathcal{A}}^- + \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix} \mathcal{R}_Z^{-1} \mathcal{G}_{\mathcal{X}} \bar{\mathcal{A}}^- \\ &\approx\; -(e^r - 1) \begin{pmatrix} \Pi_X^E & \Pi_\xi^E \\ \Pi_X^D & \Pi_\xi^D \end{pmatrix} + \left((1-e^{-r}) \begin{pmatrix} (\mathbf{F}\Pi_X^D(0) + \delta \Pi_X^E(1))\mathcal{B} \\ (\Pi_X^D(0) + \delta \Pi_X^D(1))\mathcal{B} \end{pmatrix} + (1-e^{-r}) \begin{pmatrix} \mathbf{F}\Pi_\xi^D(0) \\ \Pi_\xi^D(0) \end{pmatrix} \delta \Sigma_\xi \alpha \mathcal{W}_{\xi,\mathcal{X}} \mathcal{B} \right) \bar{\mathcal{A}}^- \\ &- \alpha(1-e^{-r})^2 \begin{pmatrix} 0 & \mathbf{F}\Sigma_d \mathbf{F}^T \\ 0 & \Sigma_d \mathbf{F}^T \end{pmatrix} \mathcal{R}_Z^{-1}(0) \delta (1-e^{-r}) \begin{pmatrix} (\Pi_X^E(1) - \mathbf{F}\Pi_X^D(1))\mathcal{B}\bar{\mathcal{A}}^- \\ (\Pi_X^E(1) - \mathbf{F}\Pi_X^D(1))\mathcal{B}\bar{\mathcal{A}}^- \end{pmatrix}. \end{aligned} \tag{113}$$

Recall that

$$\Pi_* = \Pi_\xi^E - \mathbf{F}\Pi_\xi^D \;\approx\; \delta \Pi_*(1)$$

and let also

$$\Pi_{**} \;\equiv\; \Pi_X^E - \mathbf{F}\Pi_X^D \;\approx\; \delta \Pi_{**}(1)\,.$$

Then,

$$
\begin{aligned}
&\begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \mathcal{G}^+ \;\approx \\
&-(e^r-1)\begin{pmatrix} \delta\Pi_{**}(1) & \delta\Pi_*(1) \\ \Pi_X^D(0)+\delta\Pi_X^D(1) & (\Pi_\xi^D)(0)+\delta(\Pi_\xi^D)(1) \end{pmatrix} \\
&+\left((1-e^{-r})\begin{pmatrix} \delta\Pi_{**}(1) \\ (\Pi_X^D(0)+\delta\Pi_X^D(1)) \end{pmatrix} + (1-e^{-r})\begin{pmatrix} 0 \\ \Pi_\xi^D(0) \end{pmatrix}\delta\Sigma_\xi\alpha\mathcal{W}_{\xi,\mathcal{X}}\right)\mathcal{B}(\bar{\mathcal{A}}_{\mathcal{X}}^- \; \bar{\mathcal{A}}_\xi^-) \\
&-\delta(1-e^{-r})\begin{pmatrix} 0 & 0 \\ 0 & \Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1} \end{pmatrix}\begin{pmatrix} \Pi_{**}(1) \\ \Pi_{**}(1) \end{pmatrix}\mathcal{B}(\bar{\mathcal{A}}_{\mathcal{X}}^- \; \bar{\mathcal{A}}_\xi^-) \\
&= \;(1-e^{-r})\begin{pmatrix} \delta\mathcal{G}_{11}(1) & \delta\mathcal{G}_{12}(1) \\ \mathcal{G}_{21}(0)+\delta\mathcal{G}_{21}(1) & \mathcal{G}_{22}(0)+\delta\mathcal{G}_{22}(1) \end{pmatrix}
\end{aligned}
\tag{114}
$$

where we have defined

$$
\begin{aligned}
\mathcal{G}_{11}(1) &= -\Pi_{**}(1)(e^r-(\bar{\mathcal{A}}_{\mathcal{X}})(0)) \\
\mathcal{G}_{12}(1) &= -e^r\Pi_*(1)+\Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(0) \\
\mathcal{G}_{21}(0) &= -\Pi_X^D(0)(e^r-(\bar{\mathcal{A}}_{\mathcal{X}})(0)) \\
\mathcal{G}_{21}(1) &= \Pi_X^D(0)(\bar{\mathcal{A}}_{\mathcal{X}})(1)-\Pi_X^D(1)(e^r-\bar{\mathcal{A}}_{\mathcal{X}}(0))+\Pi_\xi^D(0)\Sigma_\xi\alpha\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}\mathcal{A}_{\mathcal{X}}^-(0)-\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\Pi_{**}(1)\bar{\mathcal{A}}_{\mathcal{X}}(0) \\
\mathcal{G}_{22}(0) &= -e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0) \\
\mathcal{G}_{22}(1) &= \Pi_X^D(0)\bar{\mathcal{A}}_\xi(1)-e^r(\Pi_\xi^D)(1)+\Pi_X^D(1)\bar{\mathcal{A}}_\xi(0)+\Pi_\xi^D(0)\Sigma_\xi\alpha\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}\mathcal{A}_\xi^-(0)-\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\Pi_{**}(1)\bar{\mathcal{A}}_\xi(0)\,.
\end{aligned}
\tag{115}
$$

Now, we have

$$
(\mathcal{G}^+)^T\mathcal{R}_{\mathcal{G}}^{-1}\mathcal{G}^+ \;\approx\; (\alpha(1-e^{-r}))^{-1}\begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} \\ \mathcal{C}_{12}^T & \mathcal{C}_{22} \end{pmatrix}
\tag{116}
$$

where

$$
\begin{aligned}
\mathcal{C}_{11} &= \mathcal{G}_{21}(0)^TB_{22}(0)\mathcal{G}_{21} \;+\; \delta\Big[\mathcal{G}_{11}(1)^TB_{11}(0)\mathcal{G}_{11}(1)+\mathcal{G}_{21}(0)^TB_{11}(1)\mathcal{G}_{21}(0) \\
&+(\mathcal{G}_{11}(1)^TB_{12}(0)\mathcal{G}_{21}(0)+\mathcal{G}_{21}(0)^TB_{12}(0)^TK_{11}(1))+(\mathcal{G}_{21}(0)^TB_{22}(0)\mathcal{G}_{21}(1)+\mathcal{G}_{21}(1)^TB_{22}(0)\mathcal{G}_{21}(0))\Big] \\
\mathcal{C}_{12} &= \mathcal{G}_{21}(0)^TB_{22}(0)\mathcal{G}_{22}(0) \;+\; \delta\Big[\mathcal{G}_{11}(1)^TB_{11}(0)\mathcal{G}_{12}(1)+\mathcal{G}_{21}(0)^TB_{22}(1)\mathcal{G}_{22}(0) \\
&+(\mathcal{G}_{11}(1)^TB_{12}(0)\mathcal{G}_{22}(0)+\mathcal{G}_{21}(0)^TB_{12}(0)^T\mathcal{G}_{12}(1))+(\mathcal{G}_{21}(1)^TB_{22}(0)\mathcal{G}_{22}(0)+\mathcal{G}_{21}(0)^TB_{22}(0)\mathcal{G}_{22}(1))\Big] \\
\mathcal{C}_{22} &= \mathcal{G}_{22}(0)^TB_{22}(0)\mathcal{G}_{22}(0) \;+\; \delta\Big[\mathcal{G}_{12}(1)^TB_{11}(0)\mathcal{G}_{12}(1)+\mathcal{G}_{22}^T(0)B_{22}(1)\mathcal{G}_{22}(0) \\
&+(\mathcal{G}_{12}(1)^TB_{12}(0)\mathcal{G}_{22}(0)+\mathcal{G}_{22}(0)^TB_{12}(0)^T\mathcal{G}_{12}(1))+(\mathcal{G}_{22}(0)^TB_{22}(0)\mathcal{G}_{22}(1)+\mathcal{G}_{22}(1)B_{22}(0)\mathcal{G}_{22}(0))\Big]
\end{aligned}
\tag{117}
$$

The fixed point equation for $\mathcal{W}$ takes the form

$$\begin{aligned}
\mathcal{W} \;&=\; e^{-r}\big((\bar{\mathcal{A}}^-)^T\mathcal{R}_{\mathcal{X}}\bar{\mathcal{A}}^- - (\mathcal{G}^+)^T\mathcal{R}_{\mathcal{G}}^{-1}\mathcal{G}^+\big)\\
&\approx\; e^{-r}(\bar{\mathcal{A}}^-)^T(0)\Bigg(\mathcal{B}(0)^T\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(0) + \delta\Big(\mathcal{B}(1)\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(0) + \mathcal{B}(0)\mathcal{W}_{\mathcal{X}}(1)\mathcal{B}(0) + \mathcal{B}(0)\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(1)\\
&+ \alpha(\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(0))^T\Sigma_\xi\mathcal{W}_{\xi,\mathcal{X}}(0)\mathcal{B}(0)\Big)\\
&- (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\mathcal{C}_{11} & \mathcal{C}_{12}\\ \mathcal{C}_{12}^T & \mathcal{C}_{22}\end{pmatrix}\Bigg)\bar{\mathcal{A}}^-(0)\\
&+ e^{-r}\delta(\bar{\mathcal{A}}^-)^T(1)\Bigg(\mathcal{B}(0)^T\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(0) - (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\mathcal{C}_{11}(0) & \mathcal{C}_{12}(0)\\ \mathcal{C}_{12}(0)^T & \mathcal{C}_{22}(0)\end{pmatrix}\Bigg)\bar{\mathcal{A}}^-(0)\\
&+\; e^{-r}\delta(\bar{\mathcal{A}}^-)^T(0)\Bigg(\mathcal{B}(0)^T\mathcal{W}_{\mathcal{X}}(0)\mathcal{B}(0) - (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\mathcal{C}_{11}(0) & \mathcal{C}_{12}(0)\\ \mathcal{C}_{12}(0)^T & \mathcal{C}_{22}(0)\end{pmatrix}\Bigg)\bar{\mathcal{A}}^-(1)
\end{aligned}\tag{118}$$

In particular, zeros order terms imply the fixed point for $\mathcal{W}(0)$:

$$\begin{aligned}
\mathcal{W}(0) &= e^{-r}(\bar{\mathcal{A}}^-)^T(0)\Bigg(\mathcal{B}(0)^T\mathcal{W}(0)\mathcal{B}(0) - (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\mathcal{C}_{11}(0) & \mathcal{C}_{12}(0)\\ \mathcal{C}_{12}(0)^T & \mathcal{C}_{22}(0)\end{pmatrix}\Bigg)\bar{\mathcal{A}}^-(0)\\
&=\; e^{-r}(\bar{\mathcal{A}}^-)^T(0)\Bigg(\mathcal{B}(0)^T\mathcal{W}(0)\mathcal{B}(0) - (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\mathcal{G}_{21}(0)^T\\ \mathcal{G}_{22}(0)^T\end{pmatrix}B_{22}(0)\begin{pmatrix}\mathcal{G}_{21}(0)^T\\ \mathcal{G}_{22}(0)^T\end{pmatrix}^T\Bigg)\bar{\mathcal{A}}^-(0)\\
&= e^{-r}(\bar{\mathcal{A}}^-)^T(0)\Bigg(\mathcal{B}(0)^T\mathcal{W}(0)\mathcal{B}(0)\\
&- (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}(-\Pi_X^D(0)(e^r-(\bar{\mathcal{A}}_{\mathcal{X}})(0)))^T\\ (-e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0))^T\end{pmatrix}\mathbf{A}^{-1}\begin{pmatrix}(-\Pi_X^D(0)(e^r-(\bar{\mathcal{A}}_{\mathcal{X}})(0)))^T\\ (-e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0))^T\end{pmatrix}^T\Bigg)\bar{\mathcal{A}}^-(0)
\end{aligned}\tag{119}$$

Now, recall that

$$\bar{\mathcal{X}}_{t-1+} \;=\; \begin{pmatrix} y_{t-1+}^A\\ x_{t-1+}^A\\ y_{t-1}^E\\ x_{t-1}^D\\ Z_{C,t-1+}\end{pmatrix}$$

however we know that $\mathcal{W}$ only depends on the total endowment $\bar{x}_{t-1+} = x_{t-1+}^A + x_{t-1}^D - \mathbf{F}Z_{C,t-1+}$ and $y_{t-1+}^+ = y_{t-1+}^A + y_{t-1}^E$ because market clearing will imply that all agents positions and also asset prices only functions of these aggregate risks. This will be reflected in Π_X^D and in $\mathcal{B},\mathcal{A}$ that will only depend on these aggregate risks. However, the matrix $\mathcal{B}$ involves everything in a non-trivial way, and hence we do need the whole matrix $\mathcal{A}$. We have

$$
\begin{aligned}
\mathcal{X}_t \;&=\; \begin{pmatrix} y_t^A \\ x_t^A \\ y_t^E \\ x_t^D \end{pmatrix} \;=\; \begin{pmatrix} \bar{y}_{t-1+} - (Q_E^* + \Theta_x^E \bar{x}_{t-1+} \;+\; \Theta_y^E \bar{y}_{t-1+} + \Theta_\xi^E \xi_t) \\ C_A^D + \;\theta_\xi^A \xi_t \;+\; \theta_x^A \bar{x}_{t-1+} + \;\theta_y^A \bar{y}_{t-1+} \\ Q_E^* + \Theta_x^E \bar{x}_{t-1+} \;+\; \Theta_y^E \bar{y}_{t-1+} + \Theta_\xi^E \xi_t \\ \bar{x}_{t-1+} - (C_A^D + \;\theta_\xi^A \xi_t \;+\; \theta_x^A \bar{x}_{t-1+} + \;\theta_y^A \bar{y}_{t-1+}) \end{pmatrix} \\
&=\; X_0^- + \begin{pmatrix} \mathrm{Id} - \Theta_y^E & -\Theta_x^E \\ \theta_y^A & \theta_x^A \\ \Theta_y^E & \Theta_x^E \\ -\theta_y^A & \mathrm{Id} - \theta_x^A \end{pmatrix} \begin{pmatrix} \bar{y}_{t-1+} \\ \bar{x}_{t-1+} \end{pmatrix} \;+\; \begin{pmatrix} -\Theta_\xi^E \\ \theta_\xi^A \\ \Theta_\xi^E \\ -\theta_\xi^A \end{pmatrix} \xi_t
\end{aligned}
\tag{120}
$$

where we have used that

$$
y_t^E \;=\; Q_E^* + \Theta_{\mathcal{X}}^E \bar{\mathcal{X}}_{t-1+} \;+\; \Theta_\xi^E \xi_t \;=\; Q_E^* + \Theta_x^E \bar{x}_{t-1+} \;+\; \Theta_y^E \bar{y}_{t-1+} + \Theta_\xi^E \xi_t
$$

and

$$
x_t^A = C_A^D + \;\theta_\xi^A \xi_t \;+\; \theta_x^A \bar{x}_{t-1+} + \;\theta_y^A \bar{y}_{t-1+}\,,
$$

and

$$
X_0^- \;=\; \begin{pmatrix} -Q_E^* \\ C_A^D \\ Q_E^* \\ -C_A^D \end{pmatrix}
$$

$$
\begin{aligned}
\mathcal{G}^+ \;&=\; -(e^r - 1) \begin{pmatrix} \Pi_X^E & \Pi_\xi^E \\ \Pi_X^D & \Pi_\xi^D \end{pmatrix} + \mathcal{P}_{\mathcal{X}} \bar{\mathcal{A}}^- + \begin{pmatrix} \mathcal{G}_y^T \\ \mathcal{G}_x^T \end{pmatrix} \mathcal{R}_Z^{-1} \mathcal{G}_{\mathcal{X}} \bar{\mathcal{A}}^- \\
&\approx\; -(e^r - 1) \begin{pmatrix} \Pi_X^E & \Pi_\xi^E \\ \Pi_X^D & \Pi_\xi^D \end{pmatrix} + \left((1 - e^{-r}) \begin{pmatrix} \Pi_X^E \\ \Pi_X^D \end{pmatrix} + (1 - e^{-r}) \begin{pmatrix} \Pi_\xi^E \\ \Pi_\xi^D \end{pmatrix} \delta(\Sigma_\xi + \delta \Sigma_\xi \alpha \mathcal{W}_\xi \Sigma_\xi) \alpha \mathcal{W}_{\xi,\mathcal{X}} \right) \bar{\mathcal{A}} \\
&- \alpha (1 - e^{-r})^2 \left(\begin{pmatrix} 0 & \mathbf{F} \Sigma_d \mathbf{F}^T \\ 0 & \Sigma_d \mathbf{F}^T \end{pmatrix} - \delta \begin{pmatrix} 0 & 0 \\ \widetilde{\Sigma}_\varepsilon \mathbf{F}^T & \widetilde{\Sigma}_\varepsilon \mathbf{F}^T \end{pmatrix} \right) \\
&\left(\mathcal{R}_Z^{-1}(0) - \delta \mathcal{R}_Z^{-1}(0) (1 - e^{-r})^2 \alpha \begin{pmatrix} \mathbf{F} \widetilde{\Sigma}_\varepsilon \mathbf{F}^T & \mathbf{F} \widetilde{\Sigma}_\varepsilon \mathbf{F}^T \\ \mathbf{F} \widetilde{\Sigma}_\varepsilon \mathbf{F}^T & \mathbf{F} \widetilde{\Sigma}_\varepsilon \mathbf{F}^T \end{pmatrix} \mathcal{R}_Z^{-1}(0) \right) \delta (1 - e^{-r}) \begin{pmatrix} \Pi_{**} + \delta \Pi_*^T \Sigma_\xi \alpha \mathcal{W}_{\xi,\mathcal{X}} \\ \Pi_{**} + \delta \Pi_*^T \Sigma_\xi \alpha \mathcal{W}_{\xi,\mathcal{X}} \end{pmatrix} \mathcal{B} \bar{\mathcal{A}}^- \,.
\end{aligned}
\tag{121}
$$

and hence

$$
\begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix} \mathcal{G}^+ = (1 - e^{-r}) \begin{pmatrix} \delta(\mathcal{G}_{11}(1) + \delta \mathcal{G}_{11}(2)) & \delta(\mathcal{G}_{12}(1) + \delta \mathcal{G}_{12}(2)) \\ \mathcal{G}_{21}(0) + \delta \mathcal{G}_{21}(1) & \mathcal{G}_{22}(0) + \delta \mathcal{G}_{22}(1) \end{pmatrix}
$$

with the new (second order) terms given by

$$\begin{aligned} \mathcal{G}_{11}(2) &= -e^r\Pi_{**}(2) + \Pi_{**}(2)(\bar{\mathcal{A}}_{\mathcal{X}})(0) + \Pi_{**}(1)(\bar{\mathcal{A}}_{\mathcal{X}})(1) + \Pi_*(1)\Sigma_\xi(\bar{\mathcal{A}}_{\mathcal{X}})(0) \\ \mathcal{G}_{12}(2) &= -e^r\Pi_*(2) + \Pi_{**}(2)(\bar{\mathcal{A}}_\xi)(0) + \Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(1) + \Pi_*(1)\Sigma_\xi(\bar{\mathcal{A}}_\xi)(0) \end{aligned} \tag{122}$$

Therefore,

$$\begin{aligned} \begin{pmatrix} y_t^A \\ x_t^A \end{pmatrix} &= \mathcal{R}_{\mathcal{G}}^{-1}\left(\mathcal{G}_0^* + \mathcal{G}^+\begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix}\right) \\ &\approx (\alpha(1-e^{-r})^2)^{-1}\begin{pmatrix} \mathrm{Id} & 0 \\ -\mathbf{F}^T & \mathrm{Id} \end{pmatrix}\begin{pmatrix} \delta^{-1}(B_{11}(0)+\delta B_{11}(1)) & B_{12}(0)+\delta B_{12}(1) \\ B_{21}(0)+\delta B_{21}(1) & B_{22}(0)+\delta B_{22}(1) \end{pmatrix} \\ &\times\left(\begin{pmatrix} \mathrm{Id} & -\mathbf{F} \\ 0 & \mathrm{Id} \end{pmatrix}\mathcal{G}_0^* + (1-e^{-r})\begin{pmatrix} \delta(\mathcal{G}_{11}(1)+\delta\mathcal{G}_{11}(2)) & \delta(\mathcal{G}_{12}(1)+\delta\mathcal{G}_{12}(2)) \\ \mathcal{G}_{21}(0)+\delta\mathcal{G}_{21}(1) & \mathcal{G}_{22}(0)+\delta\mathcal{G}_{22}(1) \end{pmatrix}\begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix}\right) \\ &= \widetilde{\mathcal{G}}_0^* + (\alpha(1-e^{-r}))^{-1}\begin{pmatrix} \mathcal{D}_{11}(0)+\delta\mathcal{D}_{11}(1) & \mathcal{D}_{12}(0)+\delta\mathcal{D}_{12}(1) \\ \mathcal{D}_{21}(0)+\delta\mathcal{D}_{21}(1) & \mathcal{D}_{22}(0)+\delta\mathcal{D}_{22}(1) \end{pmatrix}\begin{pmatrix} \bar{\mathcal{X}}_{t-1+} \\ \xi_t \end{pmatrix}. \end{aligned} \tag{123}$$

Here,

$$\begin{aligned} \mathcal{D}_{11}(0) &= B_{11}(0)\mathcal{G}_{11}(1) + B_{12}(0)\mathcal{G}_{21}(0) \\ &= -B_{11}(0)\Pi_{**}(1)(e^r - (\bar{\mathcal{A}}_{\mathcal{X}})(0)) - B_{12}(0)\Pi_X^D(0)(e^r - (\bar{\mathcal{A}}_{\mathcal{X}})(0)) \\ \mathcal{D}_{12}(0) &= B_{11}(0)\mathcal{G}_{12}(1) + B_{12}(0)\mathcal{G}_{22}(0) \\ &= B_{11}(0)(-e^r\Pi_*(1) + \Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(0)) + B_{12}(0)(-e^r(\Pi_\xi^D)(0) + \Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0)) \\ \mathcal{D}_{21}(0) &= -\mathbf{F}^T(B_{11}(0)\mathcal{G}_{11}(1) + B_{12}(0)\mathcal{G}_{21}(0)) + B_{22}(0)\mathcal{G}_{21}(0) \\ &= \mathbf{F}^T B_{11}(0)\Pi_{**}(1)(e^r - \bar{\mathcal{A}}_{\mathcal{X}}) + (\mathbf{F}^T B_{12}(0) - B_{22}(0))\Pi_X^D(0)(e^r - \bar{\mathcal{A}}_{\mathcal{X}}) \\ \mathcal{D}_{22}(0) &= -\mathbf{F}^T(B_{11}(0)\mathcal{G}_{12}(1) + B_{12}(0)\mathcal{G}_{22}(0)) + B_{22}(0)\mathcal{G}_{22}(0) \\ &= -\mathbf{F}^T B_{11}(0)(-e^r\Pi_*(1) + \Pi_{**}(1)\bar{\mathcal{A}}_\xi) + (B_{22}(0) - \mathbf{F}^T B_{12}(0))(-e^r(\Pi_\xi^D)(0) + \Pi_X^D(0)\bar{\mathcal{A}}_\xi) \end{aligned} \tag{124}$$

In order to compute the last stage of the dynamics, we need the optimal creation/redemption policies. We have

$$\begin{aligned} Z_I + Z_C = (\mathrm{Id}\ \mathrm{Id})\begin{pmatrix} Z_I \\ Z_C \end{pmatrix} &= (\mathrm{Id}\ \mathrm{Id})\,\mathcal{R}_Z^{-1}(\widehat{\mathcal{G}} + \mathcal{G}_{\mathcal{X}}\mathcal{X}_t + \mathcal{G}_x x + \mathcal{G}_y y) \\ &\approx (\mathrm{Id}\ \mathrm{Id})\,\mathcal{R}_Z^{-1}\widehat{\mathcal{G}} + (\mathrm{Id}\ \mathrm{Id})\,\mathcal{R}_Z(0)^{-1}\delta(1-e^{-r})\begin{pmatrix} \Pi_{**}(1) \\ \Pi_{**}(1) \end{pmatrix}\mathcal{B}\begin{pmatrix} y_t^A \\ x_t^A \\ y_t^E \\ x_t^D \end{pmatrix} \\ &- (\mathrm{Id}\ \mathrm{Id})\left(\mathcal{R}_Z^{-1}(0) - \delta\mathcal{R}_Z^{-1}(0)(1-e^{-r})^2\alpha\begin{pmatrix} \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T \\ \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T & \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T \end{pmatrix}\mathcal{R}_Z^{-1}(0)\right) \\ &\times\alpha(1-e^{-r})^2\left(\begin{pmatrix} 0 & 0 \\ \mathbf{F}\Sigma_d\mathbf{F}^T & \mathbf{F}\Sigma_d \end{pmatrix} - \delta\begin{pmatrix} 0 & \mathbf{F}\widetilde{\Sigma}_\varepsilon \\ 0 & \mathbf{F}\widetilde{\Sigma}_\varepsilon \end{pmatrix}\right)\begin{pmatrix} y_t^A \\ x_t^A \end{pmatrix} \end{aligned} \tag{125}$$

Furthermore,

$$\bar{\mathcal{X}}_{t+} = \begin{pmatrix} y^A_{t+} + y^E_t \\ x^A_{t+} + x^D_t - \mathbf{F}^T Z_{C,t+} \end{pmatrix} = \begin{pmatrix} y^A_t + y^E_t + Z_{I,t+} + Z_{C,t+} \\ x^A_t + x^D_t - \mathbf{F}^T (Z_{C,t+} + Z_{I,t+}) \end{pmatrix} = X^+_0 + (\mathcal{B}(0) + \delta\mathcal{B}(1)) \begin{pmatrix} y^A_t \\ x^A_t \\ y^E_t \\ x^D_t \end{pmatrix}$$

This gives

$$\mathcal{B}(0) = \begin{pmatrix} \mathrm{Id} - \bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & -\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & \mathrm{Id} & 0 \\ \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathrm{Id} + \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & 0 & \mathrm{Id} \end{pmatrix}$$

and hence

$$\bar{\mathcal{A}}_{\mathcal{X}} = \mathcal{B}(0)\bar{\mathcal{A}}^-_{\mathcal{X}} = \begin{pmatrix} \mathrm{Id} - \bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & -\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & \mathrm{Id} & 0 \\ \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathrm{Id} + \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} \mathrm{Id} - \Theta^E_y & -\Theta^E_x \\ \theta^A_y & \theta^A_x \\ \Theta^E_y & \Theta^E_x \\ -\theta^A_y & \mathrm{Id} - \theta^A_x \end{pmatrix}.$$

while

$$\bar{\mathcal{A}}_{\xi} = \mathcal{B}(0)\bar{\mathcal{A}}^-_{\xi} = \begin{pmatrix} \mathrm{Id} - \bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & -\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & \mathrm{Id} & 0 \\ \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathrm{Id} + \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} -\Theta^E_\xi \\ \theta^A_\xi \\ \Theta^E_\xi \\ -\theta^A_\xi \end{pmatrix}$$

Finally, market clearing

$$\begin{aligned} &Q^*_E + \Theta^E_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \Theta^E_\xi \xi_t + C^E_A + \Theta^A_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} - \widetilde{\mathbf{B}}_A P_t = \bar{y}_{t-1+} \\ &C^D_A + \theta^A_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \theta^A_\xi \xi_t + Q^*_D + \Theta^D_{\mathcal{X}}\bar{\mathcal{X}}_{t-1+} + \Theta^D_\zeta \zeta_t - \mathbf{B}_D p_t = \bar{x}_{t-1+} \end{aligned} \tag{126}$$

implies that

$$\begin{aligned} \Pi^E_X &= (\widetilde{\mathbf{B}}_A)^{-1}\Big(((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi^E_\xi\Sigma_\xi(\Pi^E_\xi)^T\Sigma_E^{-1})\Big(\Big(\Pi^E_X - \delta\Pi^E_\xi\Sigma_\xi\alpha_E\mathcal{W}^E_{\xi,\mathcal{X}}\Big)\bar{\mathcal{A}}_{\mathcal{X}} - e^r\Pi^E_X\Big) \\ &+ \Theta^A_{\mathcal{X}} - \mathbf{1}_y\Big) \\ \Pi^E_\xi &= (\widetilde{\mathbf{B}}_A)^{-1}((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi^E_\xi\Sigma_\xi(\Pi^E_\xi)^T\Sigma_E^{-1})\Big(\Big(\Pi^E_X - \Pi^E_\xi(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}^E_\xi)^{-1}\alpha_E\mathcal{W}^E_{\xi,\mathcal{X}}\Big)\bar{\mathcal{A}}_\xi \\ &- (e^r\Pi^E_\xi + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big) \\ e^r\Pi^D_X &= \Big(\Pi^D_X - \delta\Pi^D_\zeta\Sigma_\zeta\alpha_D\mathcal{W}^D_{\zeta,\mathcal{X}}\Big)\bar{\mathcal{A}}_{\mathcal{X}} \\ &+ \alpha_D(1-e^{-r})(\Sigma_d + \delta\Pi^D_\zeta\Sigma_\zeta(\Pi^D_\zeta)^T)[(\alpha(1-e^{-r}))^{-1}(\mathcal{D}_{21}(0) + \delta\mathcal{D}_{21}(1)) - \mathbf{1}_x] \\ e^r\Pi^D_\xi &= \Big(\Pi^D_X - \delta\Pi^D_\zeta\Sigma_\zeta\alpha_D\mathcal{W}^D_{\zeta,\mathcal{X}}\Big)\bar{\mathcal{A}}_\xi + \alpha_D(1-e^{-r})(\Sigma_d + \delta\Pi^D_\zeta\Sigma_\zeta(\Pi^D_\zeta)^T)(\alpha(1-e^{-r}))^{-1}(\mathcal{D}_{22}(0) + \delta\mathcal{D}_{22}(1)) \\ e^r\Pi^D_\varepsilon &= -\alpha_D(1-e^{-r})\Sigma_d \end{aligned} \tag{127}$$

while the identity

$$Q_E^* + \Theta_{\mathcal{X}}^E \bar{\mathcal{X}}_{t-1+} + \Theta_\xi^E \xi_t + (\widetilde{K}_0^*)_y + (\alpha(1-e^{-r}))^{-1}((\mathcal{D}_{11}(0)+\delta\mathcal{D}_{11}(1))\bar{\mathcal{X}}_{t-1+} + (\mathcal{D}_{12}(0)+\delta\mathcal{D}_{12}(1))\xi_t) = \bar{y}_{t-1+}$$

implies the following equations for $(\widetilde{\mathbf{B}}_A, \Theta_{\mathcal{X}}^A)$:

$$\begin{aligned}
&((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi_\xi^E\Sigma_\xi(\Pi_\xi^E)^T\Sigma_E^{-1})\Big(\Big(\Pi_X^E - \delta\Pi_\xi^E\Sigma_\xi\alpha_E\mathcal{W}_{\xi,\mathcal{X}}^E\Big)\bar{\mathcal{A}}_{\mathcal{X}} - e^r\Pi_X^E\Big)\\
&+\ (\alpha(1-e^{-r}))^{-1}(\mathcal{D}_{11}(0) + \delta\mathcal{D}_{11}(1)) = \mathbf{1}_y\\
&((1-e^{-r})\alpha_E)^{-1}(\Sigma_E^{-1} - \delta\Sigma_E^{-1}\Pi_\xi^E\Sigma_\xi(\Pi_\xi^E)^T\Sigma_E^{-1})\Big(\Big(\Pi_X^E - \Pi_\xi^E(\Sigma_\xi^{-1} - \alpha_E\mathcal{W}_\xi^E)^{-1}\alpha_E\mathcal{W}_{\xi,\mathcal{X}}^E\Big)\bar{\mathcal{A}}_\xi\\
&-(e^r\Pi_\xi^E + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big)\ +\ (\alpha(1-e^{-r}))^{-1}(\mathcal{D}_{12}(0) + \delta\mathcal{D}_{12}(1)) = 0
\end{aligned} \tag{128}$$

In particular, in the limit as $\delta \to 0$ we get that

$$\begin{aligned}
\Pi_X^E &= (\widetilde{\mathbf{B}}_A)^{-1}\Big(((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\Pi_X^E\bar{\mathcal{A}}_{\mathcal{X}} - e^r\Pi_X^E\Big) + \Theta_{\mathcal{X}}^A - \mathbf{1}_y\Big)\\
\Pi_\xi^E &= (\widetilde{\mathbf{B}}_A)^{-1}((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\Pi_X^E\bar{\mathcal{A}}_\xi - (e^r\Pi_\xi^E + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big)\\
e^r\Pi_X^D &= \Pi_X^D\bar{\mathcal{A}}_{\mathcal{X}} + \alpha_D\Sigma_d[\alpha^{-1}\mathcal{D}_{21}(0) - (1-e^{-r})\mathbf{1}_x]\\
&= \Pi_X^D\bar{\mathcal{A}}_{\mathcal{X}} + \alpha_D\Sigma_d\Big[\alpha^{-1}\Big(\mathbf{F}^TB_{11}(0)\Pi_{**}(1)(e^r - \bar{\mathcal{A}}_{\mathcal{X}}) + (\mathbf{F}^TB_{12}(0) - B_{22}(0))\Pi_X^D(0)(e^r - \bar{\mathcal{A}}_{\mathcal{X}})\Big) - (1-e^{-r})\mathbf{1}_x\Big]\\
e^r\Pi_\xi^D &= \Pi_X^D\bar{\mathcal{A}}_\xi + \alpha_D\Sigma_d\alpha^{-1}\mathcal{D}_{22}(0)\\
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\Pi_X^E\bar{\mathcal{A}}_{\mathcal{X}} - e^r\Pi_X^E\Big) +\ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0) = \mathbf{1}_y\\
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\Pi_X^E\bar{\mathcal{A}}_\xi - (e^r\Pi_\xi^E + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big)\ +\ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{12}(0) = 0
\end{aligned} \tag{129}$$

The first two equations are simply used to pin down $\widetilde{\mathbf{B}}_A$ and $\Psi_{\mathcal{X}}^A$ that do not appear anywhere else. Furthermore, we know that $\Pi_X^E(0) = \mathbf{F}\Pi_X^D,\ \Pi_\xi^E = \mathbf{F}\Pi_X^D$, and substituting this into the last two equations gives use equations for $\Pi_{**}(1),\ \Pi_*(1)$. We have

Substituting the equation for $\mathcal{D}_{22}(0)$ into the equation for Π_ξ^D, we get

$$\begin{aligned}
&e^r\Pi_\xi^D - \Pi_X^D\bar{\mathcal{A}}_\xi\\
&= \alpha_D\Sigma_d\alpha^{-1}\Big(-\mathbf{F}^TB_{11}(0)(-e^r\Pi_*(1) + \Pi_{**}(1)\bar{\mathcal{A}}_\xi) + (B_{22}(0) - \mathbf{F}^TB_{12}(0))(-e^r(\Pi_\xi^D)(0) + \Pi_X^D(0)\bar{\mathcal{A}}_\xi)\Big)
\end{aligned} \tag{130}$$

which implies

$$e^r\Pi_\xi^D - \Pi_X^D\bar{\mathcal{A}}_\xi \ = \ \Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^TB_{11}(0)(e^r\Pi_*(1) - \Pi_{**}(1)\bar{\mathcal{A}}_\xi) \tag{131}$$

with

$$\Omega \ = \ (\mathrm{Id} + \alpha_D\Sigma_d\alpha^{-1}(B_{22}(0) - \mathbf{F}^TB_{12}(0)))^{-1}\,.$$

Similarly,

$$\Pi_X^D(0)(e^r - \bar{\mathcal{A}}_{\mathcal{X}}(0)) \ = \ \Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^TB_{11}(0)\Pi_{**}(1)(e^r - \bar{\mathcal{A}}_{\mathcal{X}}) - (1-e^{-r})\mathbf{1}_x\Big]$$

Finally, the last two equations imply

$$
\begin{aligned}
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\Pi_X^E\bar{\mathcal{A}}_{\mathcal{X}}-e^r\Pi_X^E\Big)+\ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0)=\mathbf{1}_y\\
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\Pi_X^E\bar{\mathcal{A}}_{\xi}-(e^r\Pi_\xi^E+\alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big)\ +\ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{12}(0)=0
\end{aligned}
\tag{132}
$$

Here,

$$
\begin{aligned}
\bar{\mathcal{A}}_{\mathcal{X}}(0)=\mathcal{B}(0)\bar{\mathcal{A}}_{\mathcal{X}}^-(0)\ &=\ \begin{pmatrix}\mathrm{Id}-\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & -\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & \mathrm{Id} & 0\\ \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathrm{Id}+\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & 0 & \mathrm{Id}\end{pmatrix}\\
&\times\begin{pmatrix}(\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0)_y & (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0)_x\\ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{21}(0)_y & (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{21}(0)_x\\ \mathrm{Id}-(\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0)_y & -(\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0)_x\\ -(\alpha(1-e^{-r}))^{-1}\mathcal{D}_{21}(0)_y & \mathrm{Id}-(\alpha(1-e^{-r}))^{-1}\mathcal{D}_{21}(0)_x\end{pmatrix}\\
&=\ \mathrm{Id}\ +\ (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_dB_{22}(0)\\ -\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_dB_{22}(0)\end{pmatrix}\Pi_X^D(0)(e^r-\bar{\mathcal{A}}_{\mathcal{X}}(0))\\
&=\ \mathrm{Id}\ +\ (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_dB_{22}(0)\\ -\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_dB_{22}(0)\end{pmatrix}\Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^TB_{11}(0)\Pi_{**}(1)(e^r-\bar{\mathcal{A}}_{\mathcal{X}})-(1-e^{-r})\mathbf{1}_x\Big]
\end{aligned}
\tag{133}
$$

while

$$
\begin{aligned}
\bar{\mathcal{A}}_{\xi}(0)=\mathcal{B}(0)\bar{\mathcal{A}}_{\xi}^-(0)\ &=\ \begin{pmatrix}\mathrm{Id}-\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & -\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & \mathrm{Id} & 0\\ \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathrm{Id}+\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & 0 & \mathrm{Id}\end{pmatrix}(\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\mathcal{D}_{12}(0)\\ \mathcal{D}_{22}(0)\\ -\mathcal{D}_{12}(0)\\ -\mathcal{D}_{22}(0)\end{pmatrix}\\
&=\ \begin{pmatrix}\mathrm{Id}-\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & -\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & \mathrm{Id} & 0\\ \mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{F}^T & \mathrm{Id}+\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d & 0 & \mathrm{Id}\end{pmatrix}(\alpha(1-e^{-r}))^{-1}\\
&\times\begin{pmatrix}B_{11}(0)(-e^r\Pi_*(1)+\Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(0))+B_{12}(0)(-e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0))\\ \Big(-\mathbf{F}^TB_{11}(0)(-e^r\Pi_*(1)+\Pi_{**}(1)\bar{\mathcal{A}}_\xi)+(B_{22}(0)-\mathbf{F}^TB_{12}(0))(-e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)\bar{\mathcal{A}}_\xi)\Big)\\ -\Big(B_{11}(0)(-e^r\Pi_*(1)+\Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(0))+B_{12}(0)(-e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0))\Big)\\ -\Big(-\mathbf{F}^TB_{11}(0)(-e^r\Pi_*(1)+\Pi_{**}(1)\bar{\mathcal{A}}_\xi)+(B_{22}(0)-\mathbf{F}^TB_{12}(0))(-e^r(\Pi_\xi^D)(0)+\Pi_X^D(0)\bar{\mathcal{A}}_\xi)\Big)\end{pmatrix}\\
&=\ \Gamma\,\Omega\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^TB_{11}(0)(e^r\Pi_*(1)-\Pi_{**}(1)\bar{\mathcal{A}}_\xi)
\end{aligned}
\tag{134}
$$

where

$$
\Gamma\ =\ (\alpha(1-e^{-r}))^{-1}\begin{pmatrix}\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_dB_{22}(0)\\ -\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_dB_{22}(0)\end{pmatrix}
\tag{135}
$$

Now, in order to solve the system, it remains to solve for $\Pi_*(1), \Pi_{**}(1)$. To this end, we first substitute all the expressions to get

$$
\begin{aligned}
\mathcal{D}_{11}(0) &= -B_{11}(0)\Pi_{**}(1)(e^r - (\bar{\mathcal{A}}_{\mathcal{X}})(0)) - B_{12}(0)\Pi_X^D(0)(e^r - (\bar{\mathcal{A}}_{\mathcal{X}})(0)) \\
&= -B_{11}(0)\Pi_{**}(1)(e^r - (\bar{\mathcal{A}}_{\mathcal{X}})(0)) - B_{12}(0)\Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^T B_{11}(0)\Pi_{**}(1)(e^r - \bar{\mathcal{A}}_{\mathcal{X}}) - (1-e^{-r})\mathbf{1}_x\Big]; \\
\mathcal{D}_{12}(0) &= B_{11}(0)(-e^r\Pi_*(1) + \Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(0)) + B_{12}(0)(-e^r(\Pi_\xi^D)(0) + \Pi_X^D(0)(\bar{\mathcal{A}}_\xi)(0)) \\
&= B_{11}(0)(-e^r\Pi_*(1) + \Pi_{**}(1)(\bar{\mathcal{A}}_\xi)(0)) - B_{12}(0)\Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T B_{11}(0)(e^r\Pi_*(1) - \Pi_{**}(1)\bar{\mathcal{A}}_\xi)
\end{aligned}
\tag{136}
$$

Therefore, we can the rewrite the equilibrium equations

$$
\begin{aligned}
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(\mathbf{F}\Pi_X^D(\bar{\mathcal{A}}_{\mathcal{X}} - e^r)\Big) + \ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{11}(0) = \mathbf{1}_y \\
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(-\mathbf{F}\Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T B_{11}(0)(e^r\Pi_*(1) - \Pi_{**}(1)\bar{\mathcal{A}}_\xi) + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d\Big) \\
&\ + \ (\alpha(1-e^{-r}))^{-1}\mathcal{D}_{12}(0) = 0
\end{aligned}
\tag{137}
$$

as

$$
\begin{aligned}
&-((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\mathbf{F}\Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^T\Psi_{**} - (1-e^{-r})\mathbf{1}_x\Big] \\
&+ \ (\alpha(1-e^{-r}))^{-1}\mathrm{Id}\Big[-\Psi_{**} \ - B_{12}(0)\Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^T\Psi_{**} - (1-e^{-r})\mathbf{1}_x\Big]\Big] = \mathbf{1}_y\,, \\
&((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(-\mathbf{F}\Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\Psi_* + \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d)\Big) \\
&+ \ (\alpha(1-e^{-r}))^{-1}\Big[-\Psi_* - B_{12}(0)\Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\Psi_*\Big] = 0
\end{aligned}
\tag{138}
$$

where we have defined

$$
\Psi_{**} \ \equiv \ B_{11}(0)\Pi_{**}(1)(e^r - (\bar{\mathcal{A}}_{\mathcal{X}})(0)), \ \ \Psi_* \ \equiv \ B_{11}(0)(e^r\Pi_*(1) - \Pi_{**}(1)\bar{\mathcal{A}}_\xi(0))\,.
$$

Solving these equations, we get

$$
\begin{aligned}
\Psi_{**} &= \alpha(1-e^{-r})\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1} \\
&\times \Big[-\mathbf{1}_y + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{1}_x\Big]; \\
\Psi_* &= \alpha(1-e^{-r})\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1}\Sigma_E^{-1}\mathbf{F}\Sigma_d
\end{aligned}
\tag{139}
$$

Then,

$$
\begin{aligned}
\bar{\mathcal{A}}_{\mathcal{X}}(0) &= \mathrm{Id} \ + \ \Gamma\Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^T\Psi_{**} - (1-e^{-r})\mathbf{1}_x\Big] \\
\bar{\mathcal{A}}_\xi(0) &= \Gamma\,\Omega\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\Psi_*
\end{aligned}
\tag{140}
$$

and therefore we get

$$
\Pi_{**}(1) = B_{11}(0)^{-1}\Psi_{**}\Big(e^r - \Big(\mathrm{Id} \ + \ \Gamma\Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^T\Psi_{**} - (1-e^{-r})\mathbf{1}_x\Big]\Big)\Big)^{-1}
$$

whereas

$$e^r\Pi_*(1) = B_{11}(0)^{-1}\Psi_* + \Pi_{**}(1)\Gamma\,\Omega\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\Psi_*$$

whereas

$$\Pi_X^D(0) \;=\; \Omega\alpha_D\Sigma_d\Big[\alpha^{-1}\mathbf{F}^T\Psi_{**} - (1-e^{-r})\mathbf{1}_x\Big](e^r - \bar{\mathcal{A}}_{\mathcal{X}}(0))^{-1}\,. \tag{141}$$

with

$$B_{12}(0) = -((1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T)^{-1}(-(1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon - \mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d)\mathbf{A}^{-1} \tag{142}$$

and

$$\Omega \;=\; (\mathrm{Id} + \alpha_D\Sigma_d\alpha^{-1}(\mathbf{A}^{-1} - \mathbf{F}^T B_{12}(0)))^{-1}$$

whereas

$$\widetilde{\Sigma}_\varepsilon \;=\; \Pi_\varepsilon^D(0)\Sigma_\varepsilon(\Pi_\varepsilon^D(0))^T \;=\; (e^{-r}\alpha_D(1-e^{-r}))^2\Sigma_d\Sigma_\varepsilon\Sigma_d$$

and

$$B_{11}(0) = \;((1+e^{2r})\mathbf{F}\widetilde{\Sigma}_\varepsilon\mathbf{F}^T)^{-1}$$

and

$$\mathbf{A} = -\Sigma_d\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d + \Sigma_d\,.$$

Continuing like this, we get

$$e^r\Pi_\xi^D(0) \;=\; (\Pi_X^D(0)\Gamma + \mathrm{Id})\,\Omega\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\Psi_*$$

Now, in order to determine the slope $\widetilde{\mathbf{B}}_A$, we rewrite the fixed point equation as

$$\mathbf{F}\Pi_\xi^D(0) \;=\; \widetilde{\mathbf{B}}_A^{-1}((1-e^{-r})\alpha_E)^{-1}\Sigma_E^{-1}\Big(-\mathbf{F}\Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\Psi_* - \alpha_E(1-e^{-r})\mathbf{F}\Sigma_d\Big)\,, \tag{143}$$

that is

$$\begin{aligned}
&\widetilde{\mathbf{B}}_A(0)\mathbf{F}e^{-r}(\Pi_X^D(0)\Gamma + \mathrm{Id})\,\Omega\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\alpha(1-e^{-r})\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1}\Sigma_E^{-1}\mathbf{F}\Sigma_d \\
&= \;((1-e^{-r})\alpha_E)^{-1}\Big(-\Sigma_E^{-1}\mathbf{F}\Omega\,\alpha_D\Sigma_d\alpha^{-1}\mathbf{F}^T\alpha(1-e^{-r}) \\
&\times\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1}\Sigma_E^{-1}\mathbf{F}\Sigma_d - \alpha_E(1-e^{-r})\Sigma_E^{-1}\mathbf{F}\Sigma_d\Big)
\end{aligned} \tag{144}$$

which gives

$$\begin{aligned}
\widetilde{\mathbf{B}}_A(0) = &\Big(-\Sigma_E^{-1}\mathbf{F}\Omega\,\alpha_D\alpha_E^{-1}\Sigma_d\mathbf{F}^T\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1} \\
&- \mathrm{Id}\Big)\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)\Big(e^{-r}\mathbf{F}(\Pi_X^D(0)\Gamma + \mathrm{Id})\,\Omega\alpha_D\Sigma_d\mathbf{F}^T(1-e^{-r})\Big)^{-1}
\end{aligned} \tag{145}$$

and hence

$$\widetilde{\mathbf{B}}_A(0)^{-1} = \Big(e^{-r}\mathbf{F}(\Pi_X^D(0)\Gamma + \mathrm{Id})\,\Omega\alpha_D\Sigma_d\mathbf{F}^T(1-e^{-r})\Big)\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1}$$
$$\times\Big(-\Sigma_E^{-1}\mathbf{F}\Omega\,\alpha_D\alpha_E^{-1}\Sigma_d\mathbf{F}^T\Big(\mathrm{Id} + \alpha_D(\alpha^{-1}B_{12}(0) + \alpha_E^{-1}\Sigma_E^{-1}\mathbf{F})\Omega\Sigma_d\mathbf{F}^T\Big)^{-1} - \mathrm{Id}\Big)^{-1} \tag{146}$$

F Single ETF

Let f be the K vector of security weights in the portfolio. Then, we have

$$\begin{aligned}\mathbf{C} &\equiv ((1+e^{2r})f^T\Sigma_d\Sigma_\varepsilon\Sigma_d f)^{-1}((1+e^{2r})f^T\Sigma_d\Sigma_\varepsilon\Sigma_d + f^T\Sigma_d\Sigma_\varepsilon\Sigma_d f\bar{\Lambda}_C^{-1}f^T\Sigma_d)\left(\Sigma_d^{-1} + \frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}ff^T\right)\\ &= \frac{1}{c_*(1+e^{2r})}\left((1+e^{2r})(\Sigma_\varepsilon\Sigma_d f)^T + \frac{(1+e^{2r})c_*\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}f^T + c_*\bar{\Lambda}_C^{-1}f^T\right.\\ &\left.+\frac{c_*f^T\Sigma_d f}{\bar{\Lambda}_C^2(1-\bar{\Lambda}_C^{-1}\Sigma_E)}f^T\right) = \frac{1}{c_*}(\Sigma_\varepsilon\Sigma_d f)^T + b_*f^T\end{aligned} \tag{147}$$

where $c_* = f^T\Sigma_d\Sigma_\varepsilon\Sigma_d f$ and $\Sigma_E = f^T\Sigma_d f$ is the ETF variance, and

$$b_* = \frac{2+e^{2r}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}\,. \tag{148}$$

Then,

$$\begin{aligned}\mathbf{A} &= -\bar{\Lambda}_C^{-1}\Sigma_d f(\Sigma_d f)^T + \Sigma_d\\ \mathbf{A}^{-1} &= \Sigma_d^{-1} + \frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}ff^T\\ \Omega &= \left(\mathrm{Id} + \alpha_D\Sigma_d\alpha^{-1}\left(\Sigma_d^{-1} + \frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}ff^T - f\Big(\frac{1}{c_*}(\Sigma_\varepsilon\Sigma_d f)^T + b_*f^T\Big)\right)\right)^{-1}\\ &= \big((1+\alpha_D\alpha^{-1})\mathrm{Id} - \Sigma_d f(d_*f + e_*\Sigma_\varepsilon\Sigma_d f)^T)\big)^{-1}\\ &= (1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} + \frac{(1+\alpha_D\alpha^{-1})^{-1}}{(1+\alpha_D\alpha^{-1}) - (d_*f + e_*\Sigma_\varepsilon\Sigma_d f)^T\Sigma_d f}\Sigma_d f(d_*f + e_*\Sigma_\varepsilon\Sigma_d f)^T\\ &= (1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} + g_*\Sigma_d ff^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T\end{aligned} \tag{149}$$

with

$$\begin{aligned}g_* &= \frac{(1+\alpha_D\alpha^{-1})^{-1}}{1 - \frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}\\ h_* &= \frac{(1+\alpha_D\alpha^{-1})^{-1}}{1 - \frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\frac{\alpha_D\alpha^{-1}}{c_*}\,.\end{aligned} \tag{150}$$

$$\Gamma = \ (\alpha(1-e^{-r}))^{-1} \begin{pmatrix} \bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{A}^{-1} \\ -\mathbf{F}^T\bar{\Lambda}_C^{-1}\mathbf{F}\Sigma_d\mathbf{A}^{-1} \end{pmatrix} \ = \ (\alpha(1-e^{-r}))^{-1}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E} \begin{pmatrix} f^T \\ -ff^T \end{pmatrix} \ = \ \gamma_* \begin{pmatrix} f^T \\ -ff^T \end{pmatrix} \tag{151}$$

Then,

$$\begin{aligned}
\Psi_{**} \ &= \ \alpha(1-e^{-r})\Big(1+\alpha_D\Big(\alpha^{-1}\frac{1}{c_*}c_* + (\alpha^{-1}b_* + \alpha_E^{-1}\Sigma_E^{-1})\Sigma_E\Big) \\
&\times ((1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_*c_*)\Big)^{-1} \\
&\times \Big[-\mathbf{1}_y + \alpha_D\Big(\alpha^{-1}\frac{1}{c_*}(\Sigma_\varepsilon\Sigma_d f)^T + (\alpha^{-1}b_* + \alpha_E^{-1}\Sigma_E^{-1})f^T\Big)((1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} \\
&+ g_*\Sigma_d ff^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T)\Sigma_d\mathbf{1}_x\Big] \\
&= \ A_*\Big[-\mathbf{1}_y + \alpha_D\Big(\Big[\alpha^{-1}\frac{1}{c_*}[(1+\alpha_D\alpha^{-1})^{-1} + h_*c_*] + (\alpha^{-1}b_* + \alpha_E^{-1}\Sigma_E^{-1})h_*\Sigma_E\Big](\Sigma_\varepsilon\Sigma_d f)^T \\
&+ [\alpha^{-1}\frac{1}{c_*}g_*c_* + (\alpha^{-1}b_* + \alpha_E^{-1}\Sigma_E^{-1})((1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E)]f^T\Big)\,\Sigma_d\mathbf{1}_x\Big] \\
&= \ -A_*\mathbf{1}_y \ + \ [j_*(\Sigma_\varepsilon\Sigma_d f)^T + k_*f^T]\,\Sigma_d\mathbf{1}_x
\end{aligned} \tag{152}$$

where

$$\begin{aligned}
A_* &= \alpha(1-e^{-r})\Big(1+\alpha_D\Big(\alpha^{-1} + \Big(\alpha^{-1}\frac{2+e^{2r}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E} + \alpha_E^{-1}\Sigma_E^{-1}\Big)\Sigma_E\Big)B_*\Big)^{-1} \\
B_* &= \Big((1+\alpha_D\alpha^{-1})^{-1} + \frac{(1+\alpha_D\alpha^{-1})^{-1}}{1-\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}\Sigma_E + \frac{(1+\alpha_D\alpha^{-1})^{-1}}{1-\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\alpha_D\alpha^{-1}\Big) \\
j_* \ &= \ A_*\alpha_D\alpha^{-1}\frac{1}{c_*}\Big[(1+\alpha_D\alpha^{-1})^{-1} + \frac{(1+\alpha_D\alpha^{-1})^{-1}}{1-\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\alpha_D\alpha^{-1}\Big] \\
&+ A_*\alpha_D\Big[(\alpha^{-1}\frac{2+e^{2r}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E} + \alpha_E^{-1}\Sigma_E^{-1})\frac{(1+\alpha_D\alpha^{-1})^{-1}}{1-\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\frac{\alpha_D\alpha^{-1}}{c_*}\Sigma_E\Big] \\
k_* &= A_*\alpha_D\Big[\alpha^{-1}\frac{(1+\alpha_D\alpha^{-1})^{-1}}{1-\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E} \\
&+ (\alpha^{-1}\frac{2+e^{2r}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E} + \alpha_E^{-1}\Sigma_E^{-1})((1+\alpha_D\alpha^{-1})^{-1} \\
&+ \frac{(1+\alpha_D\alpha^{-1})^{-1}}{1-\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}\Sigma_E}{1-\bar{\Lambda}_C^{-1}\Sigma_E}}\frac{\alpha_D\alpha^{-1}}{1+e^{2r}}\frac{\bar{\Lambda}_C^{-1}}{1-\bar{\Lambda}_C^{-1}\Sigma_E}\Sigma_E)\Big]
\end{aligned} \tag{153}$$

and

$$\begin{aligned}\Psi_* \;&=\; \alpha(1-e^{-r})\Big(\mathrm{Id}+\alpha_D\Big(\alpha^{-1}(\frac{1}{c_*}(\Sigma_\varepsilon\Sigma_d f)^T + b_* f^T) + \alpha_E^{-1}\Sigma_E^{-1} f^T\Big)\\ &\times ((1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} + g_*\Sigma_d f f^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T)\Sigma_d f\Big)^{-1}\Sigma_E^{-1} f^T\Sigma_d \\ &= A^*\Sigma_E^{-1} f^T \Sigma_d\end{aligned} \tag{154}$$

Then, we have

$$\begin{aligned}\bar{\mathcal{A}}_{\mathcal{X}} \;&\approx\; \mathrm{Id} \;+\; \gamma_*\begin{pmatrix} f^T \\ -ff^T\end{pmatrix}\Big((1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} + g_*\Sigma_d f f^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T\Big)\\ &\times \alpha_D\Sigma_d\Big[\alpha^{-1} f\big[-i_*(1\ 0) \;+\; [j_*(\Sigma_\varepsilon\Sigma_d f)^T + k_* f^T]\ (0\ \Sigma_d)\big] - (1-e^{-r})(0\ \mathrm{Id})\Big]\\ &=\; \mathrm{Id} \;+\; \gamma_*\begin{pmatrix} [(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E] f^T + h_*\Sigma_E(\Sigma_\varepsilon\Sigma_d f)^T \\ -[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E] ff^T + h_*\Sigma_E f(\Sigma_\varepsilon\Sigma_d f)^T\end{pmatrix}\\ &\times \alpha_D\Sigma_d\Big[\alpha^{-1} f\big[-i_*(1\ 0) \;+\; [j_*(\Sigma_\varepsilon\Sigma_d f)^T + k_* f^T]\ (0\ \Sigma_d)\big] - (1-e^{-r})(0\ \mathrm{Id})\Big]\\ &=\; \mathrm{Id} \;+\; \gamma_*\begin{pmatrix} [(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E] f^T + h_*\Sigma_E(\Sigma_\varepsilon\Sigma_d f)^T \\ -[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E] ff^T + h_*\Sigma_E f(\Sigma_\varepsilon\Sigma_d f)^T\end{pmatrix}\\ &\times \alpha_D\Big(-\alpha^{-1} i_*\Sigma_d f \quad \alpha^{-1}[j_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T + k_*\Sigma_d f f^T - (1-e^{-r})\alpha\mathrm{Id}]\Sigma_d\Big)\\ &=\; \mathrm{Id} \;+\; \begin{pmatrix}\beta_{11} & \bar\beta^T \\ -f\beta_{11} & -f\bar\beta^T\end{pmatrix}\end{aligned} \tag{155}$$

where

$$\bar\beta^T = \beta_{12}(\Sigma_d f)^T + \beta_{13}(\Sigma_d\Sigma_\varepsilon\Sigma_d f)^T$$

and

$$\begin{aligned}\beta_{11} &= -\gamma_*\alpha_D\alpha^{-1} i_*([(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E]\Sigma_E + h_*\Sigma_E c_*)\\ \beta_{12} &= \gamma_*\alpha_D\alpha^{-1}[[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E] k_*\Sigma_E + h_*\Sigma_E k_* c_* - (1-e^{-r})\alpha[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E]]\\ \beta_{13} &= \gamma_*\alpha_D\alpha^{-1}[[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E] j_*\Sigma_E + h_*\Sigma_E j_* c_* - (1-e^{-r})\alpha h_*\Sigma_E]\,.\end{aligned} \tag{156}$$

Therefore,

$$(e^r\mathrm{Id} - \bar{\mathcal{A}}_{\mathcal{X}})^{-1} \;=\; \begin{pmatrix} U_{11} & (e^r-1-\beta_{11})^{-1}\bar\beta^T U_{22} \\ -U_{22} f\beta_{11}(e^r-1-\beta_{11})^{-1} & U_{22}\end{pmatrix} \tag{157}$$

where

$$\begin{aligned}U_{11} &= (e^r - 1 - \beta_{11} + \bar\beta^T((e^r-1)\mathrm{Id} + f\bar\beta^T)^{-1} f\beta_{11})^{-1}\\ &=\; \Big(e^r - 1 - \beta_{11} + \beta_{11}(\beta_{12}\Sigma_E + \beta_{13}c_*)(e^r-1)^{-1} - \beta_{11}\frac{(e^r-1)^{-1}(\beta_{12}\Sigma_E + \beta_{13}c_*)^2}{(e^r-1) + \beta_{12}\Sigma_E + \beta_{13}c_*}\Big)^{-1}\\ &=\; (e^r-1)^{-1}\frac{e^r - 1 + \beta_*}{e^r - 1 - \beta_{11} + \beta_*}\end{aligned} \tag{158}$$

where we have used that

$$\bar{\beta}^T f \;=\; (\beta_{12}(\Sigma_d f)^T + \beta_{13}(\Sigma_d \Sigma_\varepsilon \Sigma_d f)^T) f \;=\; \beta_{12}\Sigma_E + \beta_{13} c_* \;=\; \beta_*$$

Similarly,

$$\begin{aligned} U_{22} \;&=\; (e^r - 1)^{-1}\Big(\mathrm{Id} + \frac{1}{e^r - 1 - \beta_{11}} f \bar{\beta}^T\Big)^{-1} \\ &=\; (e^r - 1)^{-1}\Big(\mathrm{Id} - \frac{1}{e^r - 1 - \beta_{11} + \beta_*} f \bar{\beta}^T\Big) \end{aligned} \tag{159}$$

Furthermore,

$$\bar{\mathcal{A}}_\xi \;\approx\; \bar{\mathcal{A}}_\xi^* \begin{pmatrix} f^T \Sigma_d \\ -f f^T \Sigma_d \end{pmatrix}, \; \bar{\mathcal{A}}_\xi^* \;=\; \gamma_* \alpha^{-1} A^* \, \alpha_D \Big((1 + \alpha_D \alpha^{-1})^{-1} + g_* \Sigma_E + h_* c_*\Big) \tag{160}$$

and

$$\begin{aligned} \Pi_{**} &= (1 + e^{2r}) c_* (e^{-r} \alpha_D (1 - e^{-r}))^2 \Big(-i_* \quad [j_*(\Sigma_\varepsilon \Sigma_d f)^T + k_* f^T]\,\Sigma_d \Big) \\ &\times \begin{pmatrix} U_{11} & (e^r - 1 - \beta_{11})^{-1} \bar{\beta}^T U_{22} \\ -U_{22} f \beta_{11} (e^r - 1 - \beta_{11})^{-1} & U_{22} \end{pmatrix} \\ &= \; (\Pi_{**}^y \;\; \pi_{11} (\Sigma_d f)^T + \pi_{12} (\Sigma_d \Sigma_\varepsilon \Sigma_d f)^T) \end{aligned} \tag{161}$$

where

$$\begin{aligned} \Pi_{**}^y \;&=\; (1 + e^{2r}) c_* (e^{-r} \alpha_D (1 - e^{-r}))^2 \Big(-i_* U_{11} - [j_*(\Sigma_\varepsilon \Sigma_d f)^T + k_* f^T]\,\Sigma_d U_{22} f \beta_{11} (e^r - 1 - \beta_{11})^{-1} \Big) \\ \pi_{11}&(\Sigma_d f)^T + \pi_{12} (\Sigma_d \Sigma_\varepsilon \Sigma_d f)^T \\ &=\; (1 + e^{2r}) c_* (e^{-r} \alpha_D (1 - e^{-r}))^2 \Big(-i_* (e^r - 1 - \beta_{11})^{-1} \bar{\beta}^T U_{22} + [j_*(\Sigma_\varepsilon \Sigma_d f)^T + k_* f^T]\,\Sigma_d U_{22} \Big) \end{aligned} \tag{162}$$

Thus,

$$\begin{aligned} e^r \Pi_* &= (1 + e^{2r}) c_* (e^{-r} \alpha_D (1 - e^{-r}))^2 A^* \Sigma_E^{-1} f^T \Sigma_d + (\Pi_{**}^y \;\; \pi_{11} (\Sigma_d f)^T + \pi_{12} (\Sigma_d \Sigma_\varepsilon \Sigma_d f)^T) \\ &\times \gamma_* \alpha^{-1} A^* \, \alpha_D \Big((1 + \alpha_D \alpha^{-1})^{-1} + g_* \Sigma_E + h_* c_*\Big) \begin{pmatrix} f^T \Sigma_d \\ -f f^T \Sigma_d \end{pmatrix} \\ &= \; \bar{\pi}_* f^T \Sigma_d \end{aligned} \tag{163}$$

where

$$\begin{aligned} \bar{\pi}_* \;&=\; (1 + e^{2r}) c_* (e^{-r} \alpha_D (1 - e^{-r}))^2 A^* \Sigma_E^{-1} \\ &+ (\Pi_{**}^y - \pi_{11} \Sigma_E - \pi_{12} c_*) \gamma_* \alpha^{-1} A^* \, \alpha_D \Big((1 + \alpha_D \alpha^{-1})^{-1} + g_* \Sigma_E + h_* c_*\Big) \end{aligned} \tag{164}$$

Now we can compute underlying securities prices dynamics and liquidity. We have

$$\begin{aligned}
\Pi_X^D &\approx \Big((1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} + g_*\Sigma_d f f^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T\Big)\\
&\times \alpha_D\alpha^{-1}\Big(-i_*\Sigma_d f \quad \Sigma_d f\left[j_*(\Sigma_\varepsilon\Sigma_d f)^T + k_* f^T\right]\Sigma_d - (1-e^{-r})\alpha\Sigma_d\Big))\\
&\times \begin{pmatrix} U_{11} & (e^r-1-\beta_{11})^{-1}\bar\beta^T U_{22}\\ -U_{22} f\beta_{11}(e^r-1-\beta_{11})^{-1} & U_{22}\end{pmatrix}\\
&= \alpha_D\alpha^{-1}\Bigg(-i^*[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*]\Sigma_d f\\
&[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*]\Sigma_d f\left[j_*(\Sigma_\varepsilon\Sigma_d f)^T + k_* f^T\right]\Sigma_d\\
&-(1-e^{-r})\alpha\Big((1+\alpha_D\alpha^{-1})^{-1}\mathrm{Id} + g_*\Sigma_d f f^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T\Big)\Sigma_d\Bigg)\\
&\begin{pmatrix} U_{11} & (e^r-1-\beta_{11})^{-1}\bar\beta^T U_{22}\\ -U_{22} f\beta_{11}(e^r-1-\beta_{11})^{-1} & U_{22}\end{pmatrix}\\
&= \Big(\bar\pi_y^D\Sigma_d f \quad -\alpha_D e^{-r}(1+\alpha_D\alpha^{-1})\Sigma_d + \Sigma_d f(\bar\pi_{X1}^D\Sigma_d f + \bar\pi_{X2}^D\Sigma_d\Sigma_\varepsilon\Sigma_d f)^T\Big)
\end{aligned} \tag{165}$$

where

$$\begin{aligned}
\bar\pi_y^D &= -\alpha_D\alpha^{-1} i^*[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*]U_{11}\\
&-\alpha_D\alpha^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*][j_* c_* + k_*\Sigma_E]\\
&-(1-e^{-r})\alpha\Big((1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*\Big)\Bigg)\frac{(e^r-1)^{-1}\beta_{11}}{e^r-1-\beta_{11}+\beta_*}
\end{aligned} \tag{166}$$

while

$$\begin{aligned}
&\Sigma_d f(\bar\pi_{X1}^D\Sigma_d f + \bar\pi_{X2}^D\Sigma_d\Sigma_\varepsilon\Sigma_d f)^T\\
&= -i^*\alpha_D\alpha^{-1}[(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*]\Sigma_d f\bar\beta^T\frac{(e^r-1)^{-1}}{e^r-1-\beta_{11}+\beta_*}\\
&+\alpha_D\alpha^{-1}(e^r-1)^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*]\Sigma_d f\left[j_*(\Sigma_\varepsilon\Sigma_d f)^T + k_* f^T\right]\Sigma_d\\
&-(1-e^{-r})\alpha\Big(g_*\Sigma_d f f^T + h_*\Sigma_d f(\Sigma_\varepsilon\Sigma_d f)^T\Big)\Sigma_d\Bigg)\\
&-\alpha_D\alpha^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1} + g_*\Sigma_E + h_* c_*][j_* c_* + k_*\Sigma_E]\\
&-(1-e^{-r})\alpha\Big(g_*\Sigma_E + h_* c_*\Big)\Bigg)(e^r-1)^{-1}\frac{1}{e^r-1-\beta_{11}+\beta_*}\Sigma_d f\bar\beta^T
\end{aligned} \tag{167}$$

implies that

$$\begin{aligned}\bar{\pi}_{X1}^{D} \;=\; & \Bigg[-i^{*}\alpha_D\alpha^{-1}[(1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*]\frac{(e^{r}-1)^{-1}}{e^{r}-1-\beta_{11}+\beta_*} \\ & -\alpha_D\alpha^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*][j_*c_*+k_*\Sigma_E] \\ & -(1-e^{-r})\alpha\Big(g_*\Sigma_E+h_*c_*\Big)\Bigg)(e^{r}-1)^{-1}\frac{1}{e^{r}-1-\beta_{11}+\beta_*}\Bigg]\beta_{12} \\ & +\alpha_D\alpha^{-1}(e^{r}-1)^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*]k_*-(1-e^{-r})\alpha g_*\Bigg)\end{aligned} \tag{168}$$

and

$$\begin{aligned}\bar{\pi}_{X2}^{D} \;=\; & \Bigg[-i^{*}\alpha_D\alpha^{-1}[(1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*]\frac{(e^{r}-1)^{-1}}{e^{r}-1-\beta_{11}+\beta_*} \\ & -\alpha_D\alpha^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*][j_*c_*+k_*\Sigma_E] \\ & -(1-e^{-r})\alpha\Big(g_*\Sigma_E+h_*c_*\Big)\Bigg)(e^{r}-1)^{-1}\frac{1}{e^{r}-1-\beta_{11}+\beta_*}\Bigg]\beta_{13} \\ & +\alpha_D\alpha^{-1}(e^{r}-1)^{-1}\Bigg([(1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*]j_*-(1-e^{-r})\alpha h_*\Bigg)\end{aligned} \tag{169}$$

Therefore,

$$\begin{aligned}\Pi_X^D\gamma_*\begin{pmatrix} f^T \\ -ff^T\end{pmatrix} \;=\; & \gamma_*\Big(\bar{\pi}_y^D+\alpha_De^{-r}(1+\alpha_D\alpha^{-1})-(\bar{\pi}_{X1}^D\Sigma_E+\bar{\pi}_{X2}^Dc_*)\Big)\Sigma_dff^T \\ =\; & \hat{\gamma}\Sigma_dff^T\,.\end{aligned} \tag{170}$$

Hence,

$$\begin{aligned}\Pi_\xi^D \;\approx\; & e^{-r}\Bigg(\hat{\gamma}\Sigma_dff^T+\mathrm{Id}\Bigg)\,\Big((1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*\Big)\alpha_D\alpha^{-1}A^*\Sigma_E^{-1}\Sigma_dff^T\Sigma_d \;=\; \Pi_\xi^{D,*}\Sigma_dff^T\Sigma_d, \\ \Pi_\xi^{D,*} =\; & \Big((1+\alpha_D\alpha^{-1})^{-1}+g_*\Sigma_E+h_*c_*\Big)\alpha_D\alpha^{-1}A^*\Sigma_E^{-1}(1+\hat{\gamma}\Sigma_E) \\ \mathbf{B}_D^{-1} \;\approx\; & e^{-r}\alpha_D(1-e^{-r})\Big(\Sigma_d+\delta(\Pi_\xi^D\Sigma_\xi(\Pi_\xi^D)^T+\Pi_\varepsilon^D\Sigma_\varepsilon(\Pi_\varepsilon^D)^T)\Big)\end{aligned} \tag{171}$$

Proof of Proposition 7.1. We have

$$
\begin{aligned}
p_{it} \;&=\; \bar p_i \;+\; \Big(\bar\pi_y^D \Sigma_d f \;-\alpha_D e^{-r}(1+\alpha_D\alpha^{-1})\Sigma_d + \Sigma_d f(\bar\pi_{X1}^D\Sigma_d f + \bar\pi_{X2}^D\Sigma_d\Sigma_\varepsilon\Sigma_d f)^T\Big)_i \begin{pmatrix}\bar y_{t-1+}\\ \bar x_{t-1+}\end{pmatrix}\\
&+\; \Pi_\xi^{D,*}(\Sigma_d f)_i f^T\Sigma_d\xi_t \;+\; (\Pi_\varepsilon^D\varepsilon_t)_i\\
&=\; \bar p_i \;+\; \mathrm{Cov}_t(d_{i,t+1},D_{t+1})(\bar\pi_y^D\bar y_{t-1+} \;+\; \Pi_\xi^{D,*}\bar\xi_t + \bar\pi_{X1}^D\,\mathrm{Cov}_t(\bar x_{t-1+}\cdot d_{t+1},D_{t+1})\\
&+\bar\pi_{X2}^D\,\mathrm{Cov}_t(\bar x_{t-1+}\cdot d_{t+1},\Sigma_\varepsilon\Sigma_d f\;\cdot d_{t+1})) - \alpha_D e^{-r}(1+\alpha_D\alpha^{-1})\,\mathrm{Cov}_t(\bar x_{t-1+}\cdot d_{t+1},d_{i,t+1})\\
&-\;\alpha_D e^{-r}(1-e^{-r})\,\mathrm{Cov}_t(d_{i,t+1},\varepsilon_t^T d_{t+1})
\end{aligned}
\tag{172}
$$

and the claim follows. ■

G Many ETFs

Proof of Proposition 8.3. We will perform expansion in large $\alpha\Lambda_C^{-1}$ When α is large, we have

$$
\bar\Lambda_C^{-1} = (e^r(\alpha(1-e^{-r}))^{-1}\Lambda_C + \mathbf{F}\Sigma_d\mathbf{F}^T)^{-1} \;\approx\; \Sigma_E^{-1} - e^r(\alpha(1-e^{-r}))^{-1}\Sigma_E^{-1}\Lambda_C\Sigma_E^{-1} \tag{173}
$$

and hence

$$
\begin{aligned}
&\mathrm{Id} - \Sigma_d\mathbf{F}^T\bar\Lambda_C^{-1}\mathbf{F} \;\approx\; \mathrm{Id} - \begin{pmatrix}\Sigma_E & \Sigma_{ED}\\ \Sigma_{ED}^T & \Sigma_D\end{pmatrix}\begin{pmatrix}\Sigma_E^{-1} - e^r(\alpha(1-e^{-r}))^{-1}\Sigma_E^{-1}\Lambda_C\Sigma_E^{-1} & 0\\ 0 & 0\end{pmatrix}\\
&= \begin{pmatrix} e^r(\alpha(1-e^{-r}))^{-1}\Lambda_C\Sigma_E^{-1} & 0\\ -\Sigma_{ED}^T\Sigma_E^{-1}(\mathrm{Id} - e^r(\alpha(1-e^{-r}))^{-1}\Lambda_C\Sigma_E^{-1}) & \mathrm{Id}\end{pmatrix}
\end{aligned}
\tag{174}
$$

and hence

$$
(\mathrm{Id} - \Sigma_d\mathbf{F}^T\bar\Lambda_C^{-1}\mathbf{F})^{-1} \;\approx\; \begin{pmatrix} e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1} & 0\\ \Sigma_{ED}^T\Sigma_E^{-1}(\mathrm{Id} - e^r(\alpha(1-e^{-r}))^{-1}\Lambda_C\Sigma_E^{-1})e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1} & \mathrm{Id}\end{pmatrix}
\tag{175}
$$

and therefore

$$
\begin{aligned}
&(\Sigma_d^{-1} - \mathbf{F}^T\bar\Lambda_C^{-1}\mathbf{F})^{-1} \;\approx\; \begin{pmatrix} e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1} & 0\\ (e^{-r}\alpha(1-e^{-r})\Sigma_{ED}^T\Lambda_C^{-1} - \Sigma_{ED}^T\Sigma_E^{-1}) & \mathrm{Id}\end{pmatrix}\begin{pmatrix}\Sigma_E & \Sigma_{ED}\\ \Sigma_{ED}^T & \Sigma_D\end{pmatrix}\\
&= \begin{pmatrix} e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1}\Sigma_E & e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1}\Sigma_{ED}\\ e^{-r}\alpha(1-e^{-r})\Sigma_{ED}^T\Lambda_C^{-1}\Sigma_E) & \Sigma_{ED}^T(e^{-r}\alpha(1-e^{-r})\Lambda_C^{-1} - \Sigma_E^{-1})\Sigma_{ED} + \Sigma_D\end{pmatrix}
\end{aligned}
\tag{176}
$$

Thus, we have that the part of ETF liquidity that depends on Λ_C is approximately given by

$$\begin{aligned} &\mathbf{F}\widetilde{\Sigma}_\varepsilon \begin{pmatrix} (1+e^{2r})\mathrm{Id} + (\Sigma_E^{-1} - e^r(\alpha(1-e^{-r}))^{-1}\Sigma_E^{-1}\Lambda_C\Sigma_E^{-1}) & 0 \\ 0 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1}\Sigma_E & e^{-r}\alpha(1-e^{-r})\Sigma_E\Lambda_C^{-1}\Sigma_{ED} \\ e^{-r}\alpha(1-e^{-r})\Sigma_{ED}^T\Lambda_C^{-1}\Sigma_E & \Sigma_{ED}^T(e^{-r}\alpha(1-e^{-r})\Lambda_C^{-1} - \Sigma_E^{-1})\Sigma_{ED} + \Sigma_D \end{pmatrix} \\ &\times \begin{pmatrix} (1+e^{2r})\mathrm{Id} + (\Sigma_E^{-1} - e^r(\alpha(1-e^{-r}))^{-1}\Sigma_E^{-1}\Lambda_C\Sigma_E^{-1}) & 0 \\ 0 & 0 \end{pmatrix} \widetilde{\Sigma}_\varepsilon\mathbf{F}^T \end{aligned} \tag{177}$$

and the leading asymptotic term is given by

$$e^{-r}\alpha(1-e^{-r})\mathbf{F}\widetilde{\Sigma}_\varepsilon \begin{pmatrix} ((1+e^{2r})\mathrm{Id} + \Sigma_E^{-1})\Sigma_E\Lambda_C^{-1}\Sigma_E((1+e^{2r})\mathrm{Id} + \Sigma_E^{-1}) & 0 \\ 0 & 0 \end{pmatrix} \widetilde{\Sigma}_\varepsilon\mathbf{F}^T \tag{178}$$

and the claim follows. ■

H Bibliography

Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." Journal of Financial Markets 5, 31-56.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. "Comovement." Journal of Financial Economics 75(2), 283-317.

Basak, Suleyman, and Anna Pavlova. 2013a. "Asset Prices and Institutional Investors." American Economic Review 103, 1728-1758.

Basak, Suleyman, and Anna Pavlova. 2013b. "A Model of Financialization of Commodities." Working paper, London Business School.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi. 2014. "Do ETFs Increase Stock Volatility?" Dice Center WP 2011-20, Ohio State University

Broman, M. 2014. "Liquidity Clienteles, Correlated Demand and Excess Comovement of Exchange-Traded Fund Returns." York University.

Buffa, Andrea, Dimitri Vayanos, and Paul Woolley. 2014. "Asset Management Contracts and Equilibrium Prices." Working paper.

Cespa, Giovanni, and Thierry Foucault. 2014. "Illiquidity Contagion and Liquidity Crashes." forthcoming in the Review of Financial Studies

Chabakauri, Georgi, and Oleg Rytchkov. 2014. "Asset Pricing with Index Investors." Working Paper.

Cheng, Minder, and Ananth Madhavan. 2009. “The Dynamics of Leveraged and Inverse Exchange Traded Funds.” Journal of Investment Management 7 (4). Cherkes, Martin, Jacob Sagi, and

Richard Stanton. 2009. “A Liquidity-Based Theory of Closed-End Funds.” Review of Financial Studies 22(1), 257-297.

Da, Zhi and Shive, Sophie. 2013. “When the Bellwether Dances to Noise: Evidence from Exchange-Traded Funds.” Working paper, University of Notre Dame.

Gabaix, Xavier, Arvind Krishnamurthy, and Olivier Vigneron. 2007. “Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market.” Journal of Finance 62(2), 557-595,

Garleanu, Nicolae, and Lasse Heje Pedersen. 2013. “Dynamic Trading with Predictable Returns and Transaction Costs.” Journal of Finance 68(6), 2309-2340.

Greenwood, Robin. 2005. “Short- and Long-Term Demand Curves for Stocks: Theory and Evidence On the Dynamics of Arbitrage.” Journal of Financial Economics 75(3), 607-649.

Greenwood, Robin, and David Thesmar. 2011. Stock Price Fragility. Journal of Financial Economics 102(3), 471-490.

Greenwood, Robin, and Dimitri Vayanos. 2014. “Bond Supply and Excess Bond Returns.” Review of Financial Studies 27, 663-713.

Gromb, Denis, and Dimitri Vayanos. 2002. “Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs.” Journal of Financial Economics 2002, 66, 361-407.

Gromb, Denis, and Dimitri Vayanos. 2010. “Limits of Arbitrage: the State Of the Theory,” Annual Review of Financial Economics, 2, 251?275.

Hamm, Sophia J. W. 2014. “The Effect of ETFs on Stock Liquidity,” Working paper.

Jiang, Wenxi, and Hongjun Yan. 2015.“ Levered ETFs.” Working paper.

Kyle, A. S. 1989. “Informed Speculation and Imperfect Competition,” Review of Economic Studies 56, 517-556.

Madhavan, Ananth, and Minder Cheng. 2009. “The Dynamics of Leveraged and Inverse-Exchange Traded Funds.” Journal of Investment Management.

Madhavan, Ananth, and Alexander Sobczyk. 2014. “Price Dynamics and Liquidity of Exchange-Traded Funds.” BlackRock Investment Institute.

Mancini, Loriano, Angelo Ranaldo, and Jan Wrampelmeyer. 2013. “Liquidity in the Foreign Exchange Market: Measurement, Commonality, and Risk Premiums,” Journal of Finance 68, 1805-1841

Marshall, Ben R., Nhut H. Nguyen, and Nuttawat Visaltanachoti. 2010. "ETF Arbitrage." Working paper, Massey University.

Pastor, Lubos, and Robert F. Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." Journal of Political Economy 111, 642–685.

Petajisto, Antti. 2013. "Inefficiencies in the Pricing of Exchange-Traded Funds." Working paper, New York University

Ramaswamy, Srichander. 2010. "Market Structures and Systemic Risks of Exchange-Traded Funds." Bank of International Settlements, BIS Working paper No. 343.

Richie, Nivine, Robert Daigler, and Kimberly C. Gleason. 2008. "The Limits to Stock Index Arbitrage: Examining S&P 500 Futures and SPDRs," Journal of Futures Markets 28(12), 1182-1205.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam. 2007. "Liquidity and the Law of One Price: The Case of the Futures-Cash Basis." Journal of Finance 62 (5), 2201-2234.

Shleifer, Andrei. 1986. "Do Demand Curves for Stocks Slope Down?" Journal of Finance 41, 579–590.

Stambaugh, Robert, Jiafeng Yu, and Yu Yuan. 2014. "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle." Forthcoming in the Journal of Finance.

Staer, Arsenio. 2014. "Fund Flows and Underlying Returns: The Case of ETFs." Working paper.

Vayanos, Dimitri. 1999. "Strategic Trading and Welfare in a Dynamic Market." Review of Economic Studies 66, 219-254.

Vayanos, Dimitri. 2001. "Strategic Trading in a Dynamic Noisy Market." Journal of Finance 56, 131–171.

Vayanos, Dimitri, and Paul Woolley. 2013. "An Institutional Theory of Momentum and Reversal." Review of Financial Studies 26, 1087-1145.

Wurgler, Jeffrey. 2010. "On the Economic Consequences of Index-Linked Investing." NBER Working Paper 16376, National Bureau of Economic Research, Cambridge MA.