Asset-pricing puzzles and price-impact*

Xiao Chen

Rutgers University

Jin Hyuk Choi

Ulsan National Institute of Science and Technology (UNIST)

Kasper Larsen

Rutgers University

Duane J. Seppi

Carnegie Mellon University

November 13, 2020

ABSTRACT: We solve in closed-form a continuous-time Nash equilibrium model in which a finite number of exponential investors continuously consume and trade strategically with price-impact. Compared to the analogous Pareto-efficient equilibrium model, price-impact has an amplification effect on risk-sharing distortions that helps resolve the interest rate puzzle and the stock-price volatility puzzle. However, price impact has little effect on the equity premium puzzle.

KEYWORDS: Asset pricing, price-impact, Nash equilibrium, Radner equilibrium, risk-free rate puzzle, equity premium puzzle, volatility puzzle.

*The authors have benefited from helpful comments from Yashar Barardehi, Suleyman Basak, René Carmona, George Constantinides, Joel Hasbrouck, Burton Hollifield, Ioannis Karatzas, Lars Kuehn, Bryan Routledge, Mete Soner, Chris Telmer, and seminar participants at Tepper (Carnegie Mellon), ORFE (Princeton), and Intech. The third author has been supported by the National Science Foundation under Grant No. DMS 1812679 (2018 - 2021). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF). The corresponding author is Duane J. Seppi. Xiao Chen has email: xc206@math.rutgers.edu, Jin Hyuk Choi has email: jchoi@unist.ac.kr, Kasper Larsen has email: KL756@math.rutgers.edu, and Duane J. Seppi has email: ds64@andrew.cmu.edu.

1 Introduction

Understanding the effects of market frictions on pricing and trading is a long-standing topic of interest in financial economics. The market microstructure literature focuses on informational and liquidity-provision frictions (e.g., Kyle 1985, Stoll 1978, Grossman and Miller 1998). In contrast, the consumption-based asset pricing literature studies how various frictions affect risk-sharing across investors and, thus, affect interest rates, stock-price volatility, and the market price-of-risk.[1] This paper merges these two approaches and investigates the asset-pricing effects of trading by strategic investors with price-impact frictions on continuous-time stock-price dynamics and interest rates.

Much of our modeling approach is standard. A finite number of risk-averse investors with time-separable utility receive individual income over time and trade a stock that pays exogenous continuous dividends and a money market account. Consumption and trading decisions occur in continuous time over a finite time horizon. Investors trade due to initial stock-holding endowment imbalances. Following Vayanos (1999), investors are strategic with respect to the perceived price-impact of their asset holdings and trades. Our main theorem provides the continuous-time Nash equilibrium stock-price process and equilibrium interest rate with price-impact via solutions to a system of ODEs.

Our main application shows that price-impact in our Nash equilibrium model has material effects on the equilibrium interest rate and stock-price process relative to both the analogous competitive price-taking Radner equilibrium (with unspanned income shocks and no

[1] Previous research shows that model incompleteness and consequent non-efficient risk-sharing equilibria can arise from several channels including: (i) Unspanned labor income as in the continuous-time Radner models in Christensen, Larsen, and Munk (2012), Žitković (2012), Christensen and Larsen (2014), Choi and Larsen (2015), Kardaras, Xing, and Žitković (2015), Larsen and Sae-Sue (2016), and Weston and Žitković (2020). (ii) Limited stock-market participation and trading constraints as in the continuous-time Radner models in Basak and Cuoco (1998), Chien, Cole, and Lustig (2011), and Hugonnier (2012). (iii) Transaction costs and quadratic penalties as in the Radner models in Heaton and Lucas (1992, 1996), Vayanos and Vila (1999), Gârleanu and Pedersen (2016), Bouchard, Fukasawa, Herdegen, and Muhle-Karbe (2018), and Weston (2018). (iv) Trading targets as in the continuous-time Nash models in Brunnermeier and Petersen (2005), Sannikov and Skrzypacz (2016), and Choi, Larsen, and Seppi (2020). (v) Price-impact as in the discrete-time Nash model in Vayanos (1999) and the continuous-time Nash models in Basak (1996, 1997).

price-impact) and the analogous Pareto-efficient equilibrium (with spanned income shocks and no price-impact). More specifically, taking the Pareto-efficient equilibrium model as a baseline, price-impact in our Nash equilibrium model magnifies risk-sharing distortions and, as a result, simultaneously lowers the interest rate and increases stock-price volatility. Therefore, price-impact simultaneously helps resolve the risk-free rate puzzle of Weil (1989) and the volatility puzzle of LeRoy and Porter (1981) and Shiller (1981). However, price-impact has no impact on the instantaneous Sharpe ratio and has quantitatively little effect on the finite-horizon Sharpe ratios for reasonable model parameters. Thus, while price-impact does have asset pricing effects on the interest rate and stock volatility, it does not help resolve the equity premium puzzle of Mehra and Prescott (1985). To the best of our knowledge, our results about price-impact and asset-pricing puzzles are new.

A variety of other approaches have been proposed to resolve asset-pricing puzzles: (i) Constantinides and Duffie (1996) and related work in Storesletten, Telmer, and Yaron (2007, 2008) and Krueger and Lustig (2010) use permanent idiosyncratic income shocks to resolve the three asset-pricing puzzles. However, Cochrane (2005, p.478-9) argues that high levels of risk aversion are still needed to explain the equity premium puzzle in Constantinides and Duffie (1996). Cochrane (2008, p.310) argues that jumps help the continuous-time limiting model of Constantinides and Duffie (1996) to explain the puzzles.[2] Constantinides (2008) and Constantinides and Gosh (2017) make related points about skewedness. In contrast to approach (i), our price-impact equilibrium model has modest levels of risk aversion and no jumps or skewedness. In particular, we use correlated arithmetic Brownian motions to generate exogenous stock dividends and strategic investor idiosyncratic income shocks.[3] (ii) In a representative agent framework, Constantinides (1990) uses an internal habit process and Campbell and Cochrane (1999) use an external habit process to explain

[2]Larsen and Sae-Sue (2016) give a closed-form competitive Radner equilibrium model with exponential utility investors and dividend and income processes governed by continuous-time Lévy jump processes that can simultaneously explain the three puzzles.

[3]Additionally, Judd (1985), Feldman and Gilles (1985), and Uhlig (1996) present both mathematical and interpretation issues related to models with a continuum of investors — such as Constantinides and Duffie (1996) — because these models rely on average clearing conditions. In contrast, our equilibrium model's idiosyncratic income shocks do not average out at the aggregate level.

the puzzles. (iii) Bansal and Yaron (2004) combine long-run consumption risk and an Epstein-Zin representative agent to explain the puzzles. (iv) Barro (2006) and the extension to an Epstein-Zin representative agent in Wachter (2013) use rare disasters based on jump processes to resolve the puzzles. In contrast to approaches (ii)-(iv), our investors' utilities are time-additive separable exponential utility functions over continuous-time consumption rate processes. Furthermore, the models in approaches (ii)-(iv) are based on representative-agent frameworks in which the underlying model is effectively complete. However, our model incorporates unspanned income shocks and trading with price-impact. Of the models in (ii)-(iv), our model is closest to the external habit model in Campbell and Cochrane (1999). Indeed, by switching off our model's idiosyncratic income shocks, the resulting common income shocks can be interpreted as an external habit.

A non-standard feature is that our analysis is non-stationary in that the asset pricing effects of price-impact dissipate over time. In our model, investors start with endowed initial stock positions that are Pareto inefficient. However, due to price-impact, investors do not trade immediately to efficient risk-sharing; rather they trade gradually to optimize with respect to a trade-off between the benefits of improved risk-sharing and price-impact costs of faster trading. Over time, their gradual trading has a cumulative effect that improves risk-sharing. Thus, our analysis shows that price-impact can have a quantitatively material short-term amplification effect on asset pricing by prolonging risk-sharing distortions. In our model, risk-sharing distortions arise as a one-time occurrence via inequalities in initial endowed stock positions. In richer economic settings, however, risk-sharing distortions could arise on a reoccurring basis from shocks to stochastic habits, sentiment, income, heterogeneous beliefs, and asymmetric information. In such a reoccurring-shock environment, the asset pricing amplification effect due to price-impact could be part of asset pricing in a stationary equilibrium. Moreover, from a calibration perspective, fundamental risk-bearing shocks can be quantitatively smaller (i.e., more realistic) and still have material asset pricing effects because they would be magnified by the price-impact amplification effect.

Basak (1996, 1997), Vayanos (1999, 2001), and Pritsker (2009) are earlier equilibrium models of strategic trading with price-impact. The main differences between our model and Basak (1996, 1997) are: First, unspanned income shocks make our model incomplete. Second, we allow for multiple traders with price-impact. Third, our price-impact equilibrium model is time-consistent. Our analysis extends or differs from seminal work by Vayanos (1999, 2001) and from Pritsker (2009) in three ways: First, we focus directly on puzzles in asset pricing moments, whereas Vayanos (1999) studies welfare but not asset pricing moments. Second, we solve for an endogenous deterministic interest rate. This allows us to investigate the interest rate puzzle in Weil (1989). Importantly, it also allows us to derive the effect of price-impact on the equity premium taking into account the effect of price-impact on endogenous interest rates. In particular, we find that price-impact has a quantitatively larger effect on endogenous interest rates than on the equity Sharpe ratio. Third, our investors start with non-Pareto efficient initial stock endowments but then subsequently receive stochastic personal income shocks. This seems more realistic than dynamic stock-holding shocks as in Vayanos (1999).[4] Fourth, and more technically, our model is in continuous time, which makes the analysis mathematically tractable. Our main theoretical result ensures existence of a solution to a coupled forward-backward system of ODEs. This solution is subsequently used to give existence of a Nash equilibrium, which we solve in closed-form.

Our analysis is related to a long-standing question in financial economics about whether liquidity is priced (see, e.g., surveys in Easley and O'Hara (2003) and Amihud, Mendelson, and Pedersen (2006)). One literature holds that liquidity is priced because investors require compensation for holding securities that expose them to higher transaction costs. For example, Amihud and Mendelson (1986) provide a theoretical analysis of this cost-compensation effect. Acharya and Pedersen (2005) also show that systematic uncertainty in stochastic trading costs (seen as a type of random negative dividends) can be a priced risk factor. However, another literature argues that the quantitative impact of liquidity on asset pricing is small by showing in various economic settings that investors can, given trading

[4]Vayanos (2001) allows for exogenous noise traders, but all our investors are utility maximizers.

costs, reduce their trading with only small utility costs. This counter-argument was first presented in Constantinides (1986). In contrast, our model is not about bid-ask spreads and transactional forms of illiquidity, but rather about the price-impact of investor asset-demand imbalances on market-clearing prices. In particular, we show, in an analytically tractable version of a standard general equilibrium asset pricing framework, how gradually decaying distortions in risk-sharing due to how investors curtail their trading in response to price-impact has asset-pricing effects. We find that price-impact affects prices through interest rate and return volatility channels but not via the market price-of-risk. Thus, our analysis supports the equity-premium argument in Constantinides (1986).

2 Setup

We consider a real economy model with a single perishable consumption good, which we take as the model's numéraire. Trading and consumption take place continuously for $t \in [0, T]$ for a finite time-horizon $T \in (0, \infty)$. The model has two traded securities: A money market account and a stock. The money market account is in zero net supply, and the stock supply is a constant $L \in \mathbb{N}$. The stock pays exogenous random dividends given by a rate process $D = (D_t)_{t\in[0,T]}$ per share. We model the asset-holding and consumption decisions of a group of $i \in \{1, ..., I\}$, $I \in \mathbb{N}$, strategic traders. Investors receive individual income given by exogenous random rate processes $Y_i = (Y_{i,t})_{t\in[0,T]}$ for $i \in \{1, ..., I\}$. In Theorem 3.3 below, we determine endogenously the interest rate $r = \big(r(t)\big)_{t\in[0,T]}$ (a deterministic time-varying function) and the stock-price process $\hat{S} = (\hat{S}_t)_{t\in[0,T]}$ in a Nash equilibrium with price-impact.

2.1 Exogenous model inputs

Let $(B_t, W_{1,t}, ..., W_{I,t})_{t\in[0,T]}$ be a set of independent one-dimensional Brownian motions starting at zero with zero drifts and unit volatilities. The augmented standard Brownian filtration is denoted by

$$\mathcal{F}_t := \sigma(B_s, W_{1,s}, ..., W_{I,s})_{s\in[0,t]}, \quad t \in [0,T]. \tag{2.1}$$

An exogenous stock dividend rate process D_t has dynamics

$$dD_t := \mu_D dt + \sigma_D dB_t, \quad D_0 \in \mathbb{R}, \tag{2.2}$$

driven by the Brownian motion B_t, with a given initial value D_0, a constant drift μ_D, and a constant volatility coefficient $\sigma_D \geq 0$.[5] The dividend rate process plays two roles: First, it generates a running flow of instantaneous dividends where the associated cumulative dividend over $[0,t]$ is $\int_0^t D_s ds$ for $t \in [0,T]$. Second, the stock pays a terminal discrete liquidating dividend D_T at the terminal date T that pins down the terminal stock price:

$$\lim_{t\uparrow T} S_t = D_T, \quad \mathbb{P}\text{-a.s.} \tag{2.3}$$

The terminal condition (2.3) requires the stock-price process $S = (S_t)_{t\in[0,T]}$ to be left-continuous at time $t = T$. We refer to Ohasi (1991, 1992) for a discussion of (2.3). A boundary condition like (2.3) is needed since our model, for mathematical tractability, has a finite time horizon. However, by making T large, the terminal liquidating dividend D_T is small relative to total dividends $\int_0^T D_s ds + D_T$. In the next section, we require that (2.3) holds for each investor i's perceived stock-price process $S = S_i$ (to be defined in Section 2.3) and for the equilibrium stock-price process $S = \hat{S}$ (to be derived in Section 3). Models with terminal Gaussian dividends include Grossman and Stiglitz (1980) and Kyle (1985), and models with running Gaussian dividends include Campbell, Grossman, and Wang (1993), Vayanos (1999), and Christensen, Larsen, and Munk (2012).

We model the income rate process $Y_{i,t}$ for trader $i \in \{1, ..., I\}$ as in Christensen, Larsen,

[5] The assumption of Brownian motion dividends is for mathematical tractability. However, the initial dividend D_0 and the dividend drift and volatility can be set such that the probability of negative dividends is arbitrary small.

and Munk (2012) and define

$$dY_{i,t} := \mu_Y dt + \sigma_Y \big(\rho dB_t + \sqrt{1-\rho^2} dW_{i,t}\big), \quad Y_{i,0} \in \mathbb{R}, \tag{2.4}$$

where the Brownian motion $W_{i,t}$ generates idiosyncratic income shocks. Investor i's income consists of a flow of income over $[0,T]$ resulting in cumulative income given by $\int_0^T Y_{i,s} ds$ and then a lump-sum income payment $Y_{i,T}$ at the end. The terminal payment $Y_{i,T}$ is a reduced-form for the value of a flow of income after the terminal date T. Similar to the boundary condition for dividends, the terminal lump-sum income Y_T can be made small relative to total income $\int_0^T Y_s ds + Y_T$ by making T large. In the income rate dynamics (2.4), the given initial value is $Y_{i,0}$, the constant drift is μ_Y, the constant volatility coefficient is $\sigma_Y \geq 0$, and $\rho \in [-1,1]$ is a correlation parameter controlling the relative magnitudes of investor-specific (i.e., idiosyncratic) income shocks and systematic income shocks correlated with the dividend process in (2.2). For example, $\rho := 0$ makes all individual income shocks independent of dividend shocks. When $\rho^2 < 1$ in (2.4), no single stock-price process can span all risk because any model with multiple Brownian motions and only one stock is necessarily incomplete by the Second Fundamental Theorem of Asset Pricing. However, when $\rho^2 = 1$, all randomness in the model is due to the Brownian motion B_t, and model completeness is possible. The assumption of homogenous income coefficients across investors is common in many Nash equilibrium models.

The strategic traders' endowed money market balances are normalized to zero. Traders begin with exogenous initial individual stock endowments equal to constants $\theta_{i,0} \in \mathbb{R}$ for $i \in \{1, ..., I\}$. Their stock-holding processes over time are

$$\theta_{i,t} := \theta_{i,0} + \int_0^t \theta'_{i,u} du, \quad t \in [0,T]. \tag{2.5}$$

This restriction forces traders to use only holding processes given by continuous order-rate processes $\theta'_{i,t}$. This rate-process restriction has been used in various equilibrium models including Back, Cao, and Willard (2000), Brunnermeier and Pedersen (2005), Gârleanu and

Pedersen (2016), and Bouchard, Fukasawa, Herdegen, and Muhle-Karbe (2018). Section 5 below shows how to incorporate discrete orders (i.e., block trades) into the model.

At each time $t \in [0,T]$, trader i chooses an order-rate $\theta'_{i,t}$ and a consumption rate $c_{i,t}$. In aggregate, trading and consumption clear the stock and real-good consumption markets in the sense that

$$L = \sum_{i=1}^{I} \theta_{i,t}, \quad LD_t + \sum_{i=1}^{I} Y_{i,t} = \sum_{i=1}^{I} c_{i,t}, \quad t \in [0,T], \tag{2.6}$$

where L is the constant stock supply. Walras' law ensures that clearing in the stock and real-good consumption markets lead to clearing in the zero-supply money market. The terminal stock price (2.3) ensures clearing in the real good consumption market at the terminal time T.

Our model is constructed to investigate how price-impact affects risk-sharing and, thus, asset pricing. Two specific risk-sharing distortions are present in the model: The first is potential deviations of investors' initial endowments $\theta_{i,0}$ from equal holdings $\frac{L}{I}$. A second potential distortion is unspanned stochastic investor income. Section 3 investigates both distortions.

2.2 Individual utility-maximization problems

With price-impact in our model, traders perceive that their holdings $\theta_{i,t}$ and order rates $\theta'_{i,t}$ affect the prices at which they trade and their resulting wealth dynamics. In particular, price-impact here is due to a persistent impact of investor holdings on the market-clearing aggregate risk-bearing capacity of the market and due to a transitory microstructure impact of investor trading. Trader i's perceived wealth process is defined by

$$X_{i,t} := \theta_{i,t} S_{i,t} + M_{i,t}, \quad t \in [0,T], \quad i \in \{1, ..., I\}, \tag{2.7}$$

where $\theta_{i,t}$ denotes her stock holdings, $S_{i,t}$ is her perceived stock-price process, and $M_{i,t}$ is her money-market balance (all these processes are to be determined in equilibrium endogenously). In a Nash equilibrium model, the perceived stock-price processes $S_{i,t}$ in (2.7) can differ off-equilibrium across different traders i given their different hypothetical holdings $\theta_{i,t}$ and trades $\theta'_{i,t}$ but the equilibrium stock-price process $\hat{S}_t$ is identical for all traders. On the other hand, we assume all traders perceive the same deterministically time-varying interest rate $r(t)$, $t \in [0,T]$ (to be determined endogenously).

Recall that each strategic trader's initial money market account balance is normalized to zero, whereas the initial endowed stock holdings are exogenously given by $\theta_{i,0} \in \mathbb{R}$. The self-financing condition produces trader i's perceived wealth dynamics

$$dX_{i,t} = r(t)M_{i,t}dt + \theta_{i,t}(dS_{i,t} + D_t dt) + (Y_{i,t} - c_{i,t})dt, \quad X_{i,0} = \theta_{i,0}S_{i,0}. \tag{2.8}$$

As usual in continuous-time stochastic control problems, the traders' controls must satisfy various regularity conditions.

Definition 2.1 (Admissibility)**.** An order-rate process $\theta'_i = (\theta'_{i,t})_{t\in[0,T]}$ and a consumption-rate process $c_i = (c_{i,t})_{t\in[0,T]}$ are admissible, and we write $(\theta'_i, c_i) \in \mathcal{A}$ if:

(i) The processes (θ'_i, c_i) have continuous paths and are progressively measurable with respect to the filtration $\mathcal{F}_t$ in (2.1).

(ii) The stock-holding process $\theta_{i,t}$ defined by (2.5) is uniformly bounded.

(iii) The wealth process dynamics (2.8) as well as the corresponding money market account balance process $M_{i,t} := X_{i,t} - S_{i,t}\theta_{i,t}$ are well-defined and

$$\sup_{t\in[0,T]} \mathbb{E}[e^{\zeta M_{i,t}}] < \infty \quad \text{for all constants } \zeta \in \mathbb{R}. \tag{2.9}$$

(iv) The perceived stock-price process $S_{i,t}$ satisfies the terminal condition (2.3).

$\diamondsuit$

Each trader i seeks to solve[6]

$$\inf_{(\theta'_i, c_i)\in\mathcal{A}} \mathbb{E}\left[\int_0^T e^{-ac_{i,t}-\delta t}dt + e^{-a(X_{i,T}+Y_{i,T})-\delta T}\right], \quad i = 1, ..., I, \tag{2.10}$$

given the perceived stock-price process $S_{i,t}$ in her wealth dynamics (2.8). In (2.10), the term $\int_0^T e^{-ac_{i,t}-\delta t}dt$ denotes utility from the consumption flow rates, and the term $e^{-a(X_{i,T}+Y_{i,T})-\delta T}$ is a bequest value function for terminal wealth. Like the terminal dividend D_T and the lump-sum terminal income Y_T, the bequest utility function proxies for continuation utility past the terminal time in our model. For tractability, the common absolute risk-aversion coefficient $a > 0$ is the same for both the consumption flow utility and the bequest value function. The common time preference parameter is $\delta \geq 0$. The assumption of identical exponential utilities across investors is common in many Nash equilibrium models, see, e.g., Vayanos (1999).

The next subsection derives stock-price dynamics perceived by trader i when solving (2.10) as part of our Nash equilibrium with price-impact. These perceived price dynamics differ from those in the competitive Radner and Pareto-efficient equilibria where all traders perceive the same stock price and act as price-takers. We describe the analogous competitive Radner equilibrium in Appendix C and the analogous competitive Pareto-efficient equilibrium in Appendix D. As is shown below, neither our Nash model with price-impact nor the analogous competitive Radner model is Pareto efficient. In addition, when the idiosyncratic income shocks are turned off, the Radner equilibrium reduces to the Pareto-efficient equilibrium, whereas our Nash model remains Pareto inefficient due to price-impact.

2.3 Price-impact for the stock market

The perceived stock-price process $S_{i,t}$ for trader i depends on market-clearing given how the other traders $j \in \{1, ..., I\} \setminus \{i\}$ respond to trader i's hypothetical choices of $\theta'_{i,t}$. Thus, for a Nash equilibrium, we must model how traders j, $j \neq i$, respond to an arbitrary control

[6]The negative sign in the exponential utility is removed for simplicity, which leads to the minimization problem in (2.10).

$\theta'_{i,t}$ used by trader i.

Several different price-impact models are available in the literature: Kyle (1985) and Back (1992) use continuous-time price-impact functions in which price changes $dS_{i,t}$ are affine in orders $d\theta_{i,t}$. Cvitanić and Cuoco (1998) take the drift process in $dS_{i,t}$ to be a function of $\theta_{i,t}$. The affine price-impact function (2.14) we derive below can be found in the single-trader optimal order-execution models in Almgren (2003) and Schied and Schöneborn (2009). Our Nash equilibrium model with price-impact can be seen as a continuous-time version of the discrete-time Nash equilibrium model in Vayanos (1999) in which S_{i,t_n} is affine in discrete orders $\Delta\theta_{i,t_n}$.

We conjecture that each generic trader $i \in \{1, ..., I\}$ perceives the responses used by the other traders j, $j \neq i$, to hypothetical holdings $\theta_{i,t}$ and trades $\theta'_{i,t}$ by investor i as given by

$$\theta'_{j,t} := A_0(t)\big(F(t)D_t - S_{i,t}\big) + A_1(t)\theta_{j,t} + A_2(t)\theta_{i,t} + A_3(t)\theta'_{i,t}, \quad j \neq i, \tag{2.11}$$

for deterministic functions of time $A_0(t), ..., A_3(t)$. The intuition behind (2.11) is that investors $j \neq i$ are perceived by investor i to have base levels for their orders $\theta'_{j,t}$ that they then adjust given the controlled price level $S_{i,t}$ (which is affected by trader i's holdings $\theta_{i,t}$ and orders $\theta'_{i,t}$) relative to an adjusted dividend level $F(t)D_t$ where $F(t)$ is the annuity[7]

$$F(t) := \int_t^T e^{-\int_t^s r(u)du} ds + e^{-\int_t^T r(u)du}, \quad t \in [0, T]. \tag{2.12}$$

The response specification in (2.11) also allows the perceived responses of investors $j \neq i$ to depend directly on investor i's hypothetical holdings $\theta_{i,t}$ and orders $\theta'_{i,t}$. Thus, $S_{i,t}$ is not assumed to be a sufficient statistic for the effects of $\theta_{i,t}$ and $\theta'_{i,t}$ on $\theta'_{j,t}$. At the end of this subsection, we show that (2.11) can be rewritten as trader j deviating from j's equilibrium behavior in response to trader i's off-equilibrium behavior. The model is symmetric in that each trader $i \in \{1, ..., I\}$ has perceptions of the form in (2.11).

The perceived investor-response functions $A_0(t), ..., A_3(t)$ in (2.11) are not simply as-

[7]For future reference, note that (2.12) is equivalent to $F(T) = 1$ and $F'(t) = r(t)F(t) - 1$.

sumed. Rather, these functions are endogenously determined in equilibrium in Theorem 3.3 below given market-clearing, certain belief-consistency conditions (described in Definition 3.1 below), and given a microstructure parameter that implicitly determines the temporary (transitory) price-impact of trading $\theta'_{i,t}$ (see Theorem 3.3 below).

Investor i does not care directly about other investors' trades and holdings per se, but she does care about the impact other investor's responses to i's trades and holdings have on the market-clearing prices at which trader i trades. In particular, the stock-price process $S_{i,t}$ trader i perceives in her optimization problem (2.10) is found using the stock-market clearing conditions (2.6) given the perceived responses in (2.11):

$$\begin{aligned} 0 &= \theta'_{i,t} + \sum_{j\neq i} \theta'_{j,t} \\ &= \theta'_{i,t} + (I-1)A_0(t)\big(F(t)D_t - S_{i,t}\big) \\ &\quad + A_1(t)(L-\theta_{i,t}) + (I-1)\big(A_2(t)\theta_{i,t} + A_3(t)\theta'_{i,t}\big). \end{aligned} \tag{2.13}$$

Provided that $A_0(t) \neq 0$ for all $t \in [0,T]$, we can solve (2.13) for trader i's perceived market-clearing stock-price process:

$$S_{i,t} = D_tF(t) + \tfrac{A_1(t)L}{A_0(t)(I-1)} + \tfrac{A_2(t)(I-1)-A_1(t)}{A_0(t)(I-1)}\theta_{i,t} + \tfrac{A_3(t)(I-1)+1}{A_0(t)(I-1)}\theta'_{i,t}. \tag{2.14}$$

Trader i's stock holdings $\theta_{i,t}$ and orders $\theta'_{i,t}$ affect the perceived stock-price process (2.14) as follows. Similar to Almgren (2003), the sum $F(t)D_t + \frac{A_1(t)L}{A_0(t)(I-1)}$ in (2.14) is called the *fundamental stock-price process*. The coefficient $\frac{A_2(t)(I-1)-A_1(t)}{A_0(t)(I-1)}$ on holdings $\theta_{i,t}$ in (2.14) is the *permanent* price-impact (positive in equilibrium) because the price-impact effect of investor i's past trading persists even after trading stops (when $\theta'_{i,t} = 0$ and $\theta_{i,t} \neq 0$). The coefficient $\frac{A_3(t)(I-1)+1}{A_0(t)(I-1)}$ on the order rate $\theta'_{i,t}$ in (2.14) is the *temporary* price-impact (positive in equilibrium) because this component of the price-impact effect disappears when investor i stops trading (i.e., when $\theta'_{i,t} = 0$). Theorem 3.3 below provides $A_0(t), ..., A_3(t)$ via solutions to a system of coupled ODEs.

To see that (2.10) is a quadratic minimization problem given the perceived stock-price process (2.14), we use the perceived money-market account balance process $M_{i,t}$ from (2.7) defined by

$$M_{i,t} := X_{i,t} - \theta_{i,t} S_{i,t}, \quad i \in \{1, ..., I\}, \tag{2.15}$$

as a state-process. The wealth dynamics (2.8) produce the following perceived dynamics of the money-market account balance process

$$\begin{aligned} dM_{i,t} &= dX_{i,t} - d(\theta_{i,t} S_{i,t}) \\ &= r(t)M_{i,t}dt + \theta_{i,t}(dS_{i,t} + D_t dt) + (Y_{i,t} - c_{i,t})dt - \theta'_{i,t} S_{i,t}dt - \theta_{i,t} dS_{i,t} \\ &= \Big(r(t)M_{i,t} + \theta_{i,t} D_t - S_{i,t}\theta'_{i,t} + Y_{i,t} - c_{i,t} \Big) dt. \end{aligned} \tag{2.16}$$

The second equality in (2.16) uses the quadratic variation property $\langle \theta_i, S_i \rangle_t = 0$, which holds because $\theta_{i,t}$ satisfies the order-rate condition (2.5). As shown in the proof in Appendix B, the affinity in the price-impact function (2.14) and the last line in (2.16) make the individual optimization problems (2.10) tractable.

Trader i's control $\theta'_{i,t}$ appears in trader j's response (2.11) implicitly through the stock-price process $S_i = (S_{i,t})_{t \in [0,T]}$ and directly via $\theta_{i,t}$ and $\theta'_{i,t}$. Substituting (2.14) for $S_{i,t}$ into (2.11), the resulting response functions for $j \neq i$ give trader j's response directly in terms of trader i's orders $\theta'_{i,t}$ and associated holdings $\theta_{i,t}$, where trader j's response is affine in those quantities:

$$\theta'_{j,t} = A_1(t)\theta_{j,t} + \frac{A_1(t)}{1-I}(L - \theta_{i,t}) + \frac{1}{1-I}\theta'_{i,t}. \tag{2.17}$$

Furthermore, the equilibrium holdings $(\hat{\theta}_{i,t}, \hat{\theta}_{j,t})$ and order-rate processes $(\hat{\theta}'_{i,t}, \hat{\theta}'_{j,t})$ in Theorem 3.3 below are consistent with (2.11) in the sense

$$\hat{\theta}'_{j,t} = A_0(t)\big(F(t)D_t - \hat{S}_t\big) + A_1(t)\hat{\theta}_{j,t} + A_2(t)\hat{\theta}_{i,t} + A_3(t)\hat{\theta}'_{i,t}, \quad j \neq i, \tag{2.18}$$

given the equilibrium stock-price process $\hat{S}_t$. This allows us re-write (2.17) as

$$\theta'_{j,t} = \hat{\theta}'_{j,t} - A_1(t)(\hat{\theta}_{j,t} - \theta_{j,t}) + \frac{1}{I-1}(\hat{\theta}'_{i,t} - \theta'_{i,t}) - \frac{A_1(t)}{I-1}(\hat{\theta}_{i,t} - \theta_{i,t}). \tag{2.19}$$

Thus, the responses in (2.19) describe deviations of $\theta'_{j,t}$ from equilibrium behavior $\hat{\theta}'_{j,t}$ for trader j, $j \neq i$, in response to trader i's off-equilibrium deviations of $\theta'_{i,t}$ from $\hat{\theta}'_{i,t}$. Note here that the equilibrium holdings $(\hat{\theta}_{i,t}, \hat{\theta}_{j,t})$ and order-rate processes $(\hat{\theta}'_{i,t}, \hat{\theta}'_{j,t})$, $j \neq i$, in (2.19) do not depend on trader i's arbitrary orders $\theta'_{i,t}$ and holdings $\theta_{i,t}$.

2.4 Modeling approach

This section briefly describes modeling differences between our analysis and other asset pricing models and explains the motivation and reasons for these differences. Our analysis is at the intersection of research on continuous-time asset pricing and dynamic trading by heterogenous investors. Thus, for tractability, our model combines modeling elements from both types of models. Our main technical contribution gives the existence of a tractable continuous-time incomplete-market price-impact equilibrium. There are two key ingredients in its construction: Price-impact perceptions as in Almgren (2003) and exponential utilities. First, perceived price-impact, necessitates, for tractability, that we restrict investors to use trading-rate processes, which, although less common than other continuous-time processes, have been used in other equilibrium trading models including Back, Cao, and Willard (2000), Brunnermeier and Pedersen (2005), and Gârleanu and Pedersen (2016). Second, while not as common as power and Epstein-Zin utilities, exponential utilities are used in asset pricing equilibrium models in Calvet (2001), Wang (2003), Christensen, Larsen, and Munk (2012), and Christensen and Larsen (2014). On the other hand, exponential utilities are standard in the rational expectations trading literature as in, e.g., Grossman and Stiglitz (1980), Campbell, Grossman, and Wang (1993), and Vayanos (1999). Since our model requires trading and market-clearing by heterogenous investors (due to their heterogenous stock holdings), exponential utilities make market-clearing tractable.

Endowment effects: With pricing-taking exponential investors, the initial endowed stock-holding distribution across investors is irrelevant for asset-pricing models. This is well-known and is re-derived in Appendices C and D. However, our exponential investors are strategic in that they perceive their holdings and trades to have price-impact, which explains why our equilibrium model exhibits stock-endowment effects. However, these endowment effects are due to a risk-bearing mechanism rather than a wealth effect. When investors are endowed with non-Pareto efficient initial stock endowments in terms of risk-sharing, it is suboptimal for investors to trade immediately to their Radner allocations due to their perceived costs given their perceived price-impact.[8] The deviation of risk-sharing in the model relative to the Radner equilibrium, in turn, affects investor stock demands, which has price effects. It is this risk-sharing based endowment mechanism in our model that affects risk-free interest rates and stock-price dynamics as mentioned in the introduction and detailed in Section 4 below. The intuition behind the risk-sharing based endowment mechanism is simple: It is costly to rebalance to efficient positions given price-impact.

Timing: Our model's time horizon is finite but can be arbitrary long. Because of slow trading due to price-impact, our investors' heterogenous stock holdings converge gradually over time to the Radner allocations over the time horizon. Consequently, our model is non-stationary, and, in particular, the asset pricing effects of price-impact are short-term in nature. However, to the extent that investors are repeatedly shocked away from efficient risk sharing and need to trade, the model and its asset pricing effects could be made stationary. These could be periodic shocks to preferred holdings — due to fund flows causing fluctuations in institutional assets under management (e.g., Coval and Stafford (2017)) and changes in sentiment (e.g., Baker and Wurgler (2006)) — as well as to shocks to actual holdings as in Vayanos (1999).

Illiquidity: Liquidity in general equilibrium asset pricing can be modeled as price-impact or as transaction costs. Our price-impact approach follows Vayanos (1999) in modeling liquidity as price pressure. Alternatively, transaction costs can be explicitly modeled search

[8]The Radner equilibrium is not Pareto efficient when $\rho^2 < 1$ in (2.4). However, when $\rho^2 = 1$, the idiosyncratic income shocks disappear and the Radner equilibrium becomes Pareto efficient.

costs (as in, e.g., Duffie, Gârleanu, and Pedersen (2005)) or implicitly represented via a reduced-form quadratic penalty (e.g., as in Gârleanu and Pedersen (2016)), or can be fees paid to intermediaries (i.e., bid-ask spreads and commissions charged by market makers, brokers, and exchanges). Price-impact and transaction costs often produce similar implications for optimal investor consumption and portfolio choice, but there are other differences: One difference is market clearing. Price pressure in prices is paid and received by all buyers and sellers such that, by definition, markets clear. In contrast, transaction costs must either be modeled as dissipative deadweight costs (i.e., costs for some investors but not income for other market participants) or else market-making intermediaries must be modeled as additional agents in the general equilibrium. A second difference concerns trading dynamics. Bid-ask spread type transaction costs and commissions drive a wedge between the net prices received by sellers and paid by buyers. Thus, as in Noh and Weston (2020), such transaction costs can lead to endogenous possible no-trade. In contrast, there is no such wedge with price-impact, and so there is no endogenous no-trade in our model. Third, price pressure effects can be large if fundamental asset-holding imbalances are large. In contrast, bid-ask spreads are compensation for short-term inventory holding. Thus, the ultimate asset buyers and sellers can keep these fees from being too large, absent adverse selection and learning, by using dynamic trading to limit how much inventory intermediaries need to hold at any point in time. Fourth, the quadratic penalty literature takes both permanent and transitory transaction costs as exogenous. In contrast, Section 3 shows that persistent price impact costs in our model, as in Vayanos (1999), are endogenously determined once the transitory price impact is determined. Fifth, household trading constraints and exogenous market participation restrictions (which can be viewed as infinite costs for certain types of transactions) also affect asset pricing by distorting asset holdings (see footnote 1 for references). In contrast, asset-holding distortions in our model are endogenous given that price-impact slows investor adjustments responding to suboptimal risk-sharing.

3 Price-impact equilibrium

This section first defines a Nash equilibrium with price impact and then solves for the equilibrium in closed-form.

3.1 Equilibrium definition

The following definition is a continuous-time version of the Nash equilibrium in Vayanos (1999):

Definition 3.1 (Nash equilibrium)**.** Perceived investor response coefficients given by continuous functions of time $A_0, ..., A_3 : [0,T] \to \mathbb{R}$ constitute a *financial-market Nash equilibrium* if:

(i) The solution $(\hat{c}_{i,t}, \hat{\theta}'_{i,t})$ to trader i's individual optimization problem (2.10) with the price-impact function (2.14) exists for all $i \in \{1, ..., I\}$.

(ii) The stock-price processes resulting from inserting trader i's optimizer $\hat{\theta}'_{i,t}$ into the price-impact function $S_{i,t}$ in (2.14) are identical for all traders $i \in \{1, ..., I\}$. This common stock-price process, denoted by $\hat{S}_t$, satisfies the terminal dividend restriction (2.3).

(iii) The individual orders $(\hat{\theta}'_{i,t})_{i=1}^I$ and corresponding holding processes $(\hat{\theta}_{i,t})_{i=1}^I$ satisfy the consistency requirement (2.18).

(iv) The real-good consumption market clearing and the stock-market clearing conditions (2.6) hold at all times $t \in [0,T]$.

$\diamondsuit$

A financial-market Nash equilibrium combines both the idea of strategic behavior in a Nash equilibrium and the idea of market-clearing. In models of strategic trading, such as Kyle (1985), this is often achieved by a liquidity provider who supplies residual liquidity to clear the market. In contrast, following Vayanos (1999), our model constructs perceived response functions in (2.11) for other strategic investors $j \neq i$ such that markets clear given any potential off-equilibrium trade $\theta'_{i,t}$ by any generic investor i. In particular, this construction leads to equation (2.19), which shows how other investors $j \neq i$ deviate from their equilibrium strategies $\hat{\theta}'_{j,t}$ in order for markets to clear in response to off-equilibrium trading deviations $\theta'_{i,t}$ of a generic investor i from her equilibrium strategy $\hat{\theta}'_{i,t}$. Equilibrium strategies are, therefore, fixed-points in that they are optimal responses to the perceived collective off-equilibrium response functions of the other traders given the equilibrium strategies.[9] We also note here that our notion of equilibrium is defined in terms of the perceived price-impact coefficients $A_0, ..., A_3$. These coefficients are the underlying primitives in our model. Given the equilibrium price impact coefficients, we then solve for optimal strategies $(\hat{c}_{i,t}, \hat{\theta}_{i,t})$, $i \in \{1, ..., I\}$, for investors given their equilibrium price perceptions as in a standard Nash equilibrium.

3.2 Results

Our main existence equilibrium existence result is based on the following technical lemma (the proof is in Appendix B below). It guarantees the existence of a solution to an autonomous forward-backward system of coupled ODEs with forward component ψ and backward components (F, Q_2, Q_{22}, Q). Similar forward-backward systems have appeared in equilibrium theory. For example, in Kyle (1985), the forward component is the filter and the backward components are the value-function coefficients.

Lemma 3.2. *For all $\alpha > 0$, there exists a constant $w \geq \frac{L^2}{I}$ such that the unique solutions*

[9]There are two further technical details to note here: First, for a Nash equilibrium we only need to consider unilateral deviations from equilibrium strategies by individual investors rather than multilateral deviations. Second, there is no infinite regress problem in our model (see Townsend (1983)). In a Nash equilibrium, strategies only need to be optimal with respect to perceptions of other agents' behavior, but Nash equilibrium does not require off-equilibrium beliefs to support that perceived behavior.

of the coupled ODE system

$$\psi'(t) = 2\frac{F(t)Q_{22}(t)}{\alpha}\left(\psi(t) - \frac{L^2}{I}\right), \quad \psi(T) = w, \tag{3.1}$$

$$F'(t) = F(t)\Big(\delta - \tfrac{a^2\sigma_D^2}{2I}\psi(t) - \tfrac{a\big(aI\sigma_Y^2 + 2aL\rho\sigma_D\sigma_Y - 2I\mu_Y - 2L\mu_D\big)}{2I}\Big) - 1, \quad F(T) = 1, \tag{3.2}$$

$$Q_2'(t) = a\rho\sigma_D\sigma_Y + \frac{2LF(t)Q_{22}(t)^2}{\alpha I} + \frac{Q_2(t)}{F(t)} - \mu_D, \quad Q_2(T) = 0, \tag{3.3}$$

$$Q_{22}'(t) = a\sigma_D^2 - \frac{2F(t)Q_{22}(t)^2}{\alpha} + \frac{Q_{22}(t)}{F(t)}, \quad Q_{22}(T) = 0, \tag{3.4}$$

$$Q'(t) = -\frac{\delta}{a} + \frac{aQ(t) - \log\left(\frac{1}{F(t)}\right) + 1}{aF(t)} + \frac{a\sigma_Y^2}{2} - \frac{L^2F(t)Q_{22}(t)^2}{\alpha I^2} - \mu_Y, \quad Q(T) = 0, \tag{3.5}$$

satisfy $\psi(0) = \sum_{i=1}^I \theta_{i,0}^2$.

Next, we give our main theoretical result. In this theorem, the parameter $\alpha > 0$ is a free input parameter that controls the temporary price-impact effect (see (3.10) below). In Appendix E, we use calibrate α to match observed data.

Theorem 3.3. *Let* $(\psi, F, Q, Q_2, Q_{22})$ *be as in Lemma 3.2 for initial stock endowments* $\sum_{i=1}^I \theta_{i,0} = L$. *Then, a financial-market Nash equilibrium exists in which:*

(i) The perceived investor response coefficients in (2.11) *are*

$$A_0(t) := \frac{ILQ_{22}(t)}{\alpha(I-1)\big(IQ_2(t) + 2LQ_{22}(t)\big)}, \tag{3.6}$$

$$A_1(t) := \frac{A_0(t)(I-1)F(t)\big(IQ_2(t) + 2LQ_{22}(t)\big)}{IL}, \tag{3.7}$$

$$A_2(t) := \frac{A_0(t)F(t)\big(IQ_2(t) - (I-2)LQ_{22}(t)\big)}{IL}, \tag{3.8}$$

$$A_3(t) := A_0(t)\alpha + \frac{1}{1-I}, \tag{3.9}$$

which simplifies the perceived price-impact model (2.14) *to*

$$S_{i,t} = F(t)D_t + F(t)\left(\frac{2LQ_{22}(t)}{I} + Q_2(t)\right) - F(t)Q_{22}(t)\theta_{i,t} + \alpha\theta_{i,t}'. \tag{3.10}$$

(ii) The equilibrium interest rate $r(t)$ is given by

$$r(t) = \delta - \frac{a^2\sigma_D^2}{2I}\psi(t) - \frac{a\left(aI\sigma_Y^2 + 2aL\rho\sigma_D\sigma_Y - 2I\mu_Y - 2L\mu_D\right)}{2I}. \tag{3.11}$$

(iii) The equilibrium stock-price process is

$$\hat{S}_t = F(t)D_t + F(t)\left(\frac{LQ_{22}(t)}{I} + Q_2(t)\right), \tag{3.12}$$

where $F(t)$ is the annuity in (3.2) *with explicit solution* (2.12).

(iv) For $i \in \{1, ..., I\}$, trader i's optimal order and consumption rates are:

$$\hat{\theta}'_{i,t} = \gamma(t)\big(\hat{\theta}_{i,t} - \frac{L}{I}\big), \quad \gamma(t) := \frac{F(t)Q_{22}(t)}{\alpha}, \tag{3.13}$$

$$\hat{c}_{i,t} = \frac{\log\big(F(t)\big)}{a} + D_t\hat{\theta}_{i,t} + \frac{\hat{M}_{i,t}}{F(t)} + Q(t) + \hat{\theta}_{i,t}Q_2(t) + \frac{1}{2}\hat{\theta}_{i,t}^2 Q_{22}(t) + Y_{i,t}. \tag{3.14}$$

Remark 3.1.

1. The equilibrium stock-price process (3.12) is Gaussian. Such Bachelier stock-price models are common equilibrium prices in many settings including Kyle (1985), Grossman and Stiglitz (1980), and Hellwig (1980). By setting the initial dividend D_0 to be sufficiently high, the probability of negative stock prices can be made small.

   The equilibrium price dynamics given the prices in (3.12) are

$$\begin{aligned} d\hat{S}_t &= F(t)\Big(dD_t + \frac{LQ'_{22}(t)}{I}dt + Q'_2(t)dt\Big) + \frac{F'(t)}{F(t)}\hat{S}_t dt \\ &= \big(r(t)\hat{S}_t - D_t + F(t)\sigma_D\lambda\big)dt + F(t)\sigma_D dB_t, \end{aligned} \tag{3.15}$$

   where the *market price-of-risk* coefficient in our Nash model is

$$\lambda := \frac{a}{I}(L\sigma_D + I\sigma_Y\rho). \tag{3.16}$$

The dynamics (3.15) follow from Itô's Lemma and then rearranging using (3.1)-(3.4) and (3.11). Interestingly, the market price of risk λ is an intertemporal constant that only depends on exogenous parameters for dividends, income, preferences, and number of investors.

2. Price-impact is partially endogenous in our model. In particular, the perceived persistent price impact $F(t)Q_{22}(t)$ of an investor's stock holdings $\theta_{i,t}$ in (3.10) is endogenously determined in equilibrium given the perceived transitory price impact α of an investor's trading rate $\theta'_{i,t}$.

3. Our Nash equilibrium model with price-impact has stock-endowment effects because the equilibrium stock holdings $\hat{\theta}_{i,t}$ for trader i in (3.13) depend on the initial endowed holdings $\theta_{i,0}$:

$$\hat{\theta}_{i,t} = \frac{L}{I} + \Big(\theta_{i,0} - \frac{L}{I}\Big)e^{\int_0^t \gamma(s)ds}, \quad t \in [0,T], \tag{3.17}$$

where $\gamma(t)$ is given in (3.13). In contrast, in the competitive Radner equilibrium (with no price-impact), trader i's time $t \in (0,1]$ equilibrium holdings are L/I regardless of trader i's endowed holdings $\theta_{i,0}$. In particular, the absence of wealth effects due to stochastic dividends and incomes with exponential utility makes the model tractable by making equilibrium trades in (3.13) and the corresponding holdings (3.17) deterministic functions of time $t \in [0,1]$. Section 4 below shows that the stock-endowment dependency ultimately allows our Nash equilibrium model to simultaneously resolve some, but not all, asset pricing puzzles.

4. Heterogeneity in initial stock holdings leads to distortions in risk-sharing over time that affect asset pricing. Appendix B shows that the solution of (3.1) satisfies $\psi(t) = \sum_{i=1}^I \hat{\theta}_{i,t}^2$, which is our metric for stock-holding heterogeneity. If the initial stock endowments are equal with $\theta_{i,0} = \frac{L}{I}$, then $\psi'(t) = 0$ from (3.1); and hence, $\psi(t) = \frac{L^2}{I}$ for all $t \in [0,T]$. In this case, the equilibrium interest rate in (3.11) becomes the

analogous competitive Radner equilibrium interest rate given by (see Appendix C below):

$$r^{\text{Radner}} := \delta - \frac{a^2\sigma_D^2 L^2}{2I^2} - \frac{a\left(aI\sigma_Y^2 + 2aL\rho\sigma_D\sigma_Y - 2I\mu_Y - 2L\mu_D\right)}{2I}. \tag{3.18}$$

For non-equal endowments (i.e., non-Pareto efficient), Cauchy-Schwarz's inequality produces $\sum_{i=1}^{I}\theta_{i,0}^2 > \frac{L^2}{I}$, which leads to $\psi(t) = \sum_{i=1}^{I}\hat{\theta}_{i,t}^2 > \frac{L^2}{I}$ for all $t \in [0,T]$. In that case, the Nash equilibrium interest rate (3.11) is strictly smaller than the competitive Radner equilibrium interest rate in (3.18). Thus, inequality in investor stock endowments as measured by $\sum_{i=1}^{I}\hat{\theta}_{i,t}^2 - \frac{L^2}{I}$ is a key quantity in our model's ability to resolve the interest rate puzzle and, as shown in Section 4 below, also affects the stock volatility puzzle. However, over time, the equilibrium holdings in (3.17) converge to equal holdings (Pareto efficient), and so these asset pricing effects are temporary.

5. Even if the analogous competitive Radner equilibrium is Pareto-efficient (i.e., if investor income is spanned), our Nash equilibrium can be non-Pareto efficient. To see this, set $\rho^2 = 1$ in the income dynamics (2.4), which makes the analogous competitive Radner model complete. In this case, the interest rate (3.18) in the competitive Radner equilibrium agrees with the Pareto efficient interest rate given in (D.4) in Appendix D below. However, as long as $\theta_{i,0} \neq \frac{L}{I}$ for some trader i, we have $\sum_{i=1}^{I}\theta_{i,0}^2 > \frac{L^2}{I}$ by Cauchy-Schwarz's inequality and, consequently, $\psi(t) = \sum_{i=1}^{I}\hat{\theta}_{i,t}^2 > \frac{L^2}{I}$ by (3.1). Thus, even if the competitive Radner equilibrium is Pareto-efficient because $\rho^2 = 1$ in (2.4), the Nash equilibrium interest rate (3.11) is strictly smaller than the Pareto-efficient equilibrium interest rate (D.4) whenever $\theta_{i,0} \neq \frac{L}{I}$ for some trader i.

6. Unspanned investor-income randomness also affects risk-sharing and asset pricing. Individual investor income $Y_{i,t}$ is optimally consumed, as seen in (3.14), and, thus, income shocks do not directly affect optimal investor holdings. As a result, investor trading in (3.13) is deterministic, which simplifies the modeling of the stock endow-

ment effects. However, the parameters of the investor income process affect asset pricing in (3.11) and (3.12) and the optimal trading rate $\hat{\theta}'_{i,t}$ in (3.13). Thus, imperfect risk-sharing due to both distortions in initial stock endowments and unspanned (idiosyncratic) shocks to investor income has asset-pricing effects with price-impact.

7. The proof of Theorem 3.3 in Appendix B is based on the standard dynamical programming principle and HJB equations. Thus, by definition, the individual optimization problems in our Nash equilibrium model are time-consistent. However, it might appear that our Nash equilibrium model is time-inconsistent given that the optimal holdings $\hat{\theta}_{i,t}$ in (3.17) depend on the endowed holdings $\theta_{i,0}$ (see, e.g., the discussion in Remark 3 on p.455 in Basak, 1997). The explanation for why our Nash equilibrium is time-consistent while the equilibrium holdings $\hat{\theta}_{i,t}$ depend on the endowed holdings $\theta_{i,0}$ lies in the state-processes and controls used in the proof of Theorem 3.3, summarized in Table 1:

Table 1: State-processes and controls used.

	State processes	Controls
Nash	$M_{i,t}, D_t, \theta_{i,t}$	$c_{i,t}, \theta'_{i,t}$
Radner and Pareto	$X_{i,t}$	$c_{i,t}, \theta_{i,t}$

For time-consistent optimization problems, the initial control values cannot appear in the optimal controls. However, because the trading rate $\theta'_{i,t}$ is the control — not the stock holdings $\theta_{i,t}$ — in the Nash equilibrium model, the endowment $\theta_{i,0}$ can (and do) appear in the time-consistent individual optimal holdings $\hat{\theta}_{i,t}$ in (3.17). Likewise, the Radner and Pareto equilibrium models are time-consistent, and so the endowment $\theta_{i,0}$ cannot (and do not) appear in the individual optimal holdings $\frac{L}{I}$.

4 Asset-pricing puzzles

This section shows that our continuous-time price-impact equilibrium model produces material differences relative to the analogous Radner and Pareto-efficient equilibria. In particular, based on the C-CAPM from Breeden (1979), Appendix D derives the analogous Pareto-efficient equilibrium where all investors act as price-takers and markets are complete. We show how price-impact affects the three main asset-pricing puzzles (risk-free rate, equity premium, and volatility). We do this both analytically and by illustrating the equilibrium differences in a numerical example. The differences between our model and the Radner and Pareto-efficient equilibria are due to perceived price-impact, heterogenous stock holdings, and market incompleteness (due to idiosyncratic income risk when $\rho^2 < 1$).

Our conclusion is that, by using the Pareto-efficient equilibrium model as a benchmark, our price-impact Nash equilibrium model can simultaneously help resolve the risk-free rate puzzle of Weil (1989), and the volatility puzzle of LeRoy and Porter (1981) and Shiller (1981). However, price-impact has little quantitative effect on the Sharpe ratio and the equity premium puzzle of Mehra and Prescott (1985). These empirical works on asset-pricing puzzles compare a competitive representative agent model with historical data. Such representative agent models are (effectively) complete and therefore also Pareto efficient by the First Welfare theorem. Therefore, we use the Pareto efficient equilibrium interest rate and stock-price process as benchmarks. The comparison with the Radner model shows the incremental effect of price-impact alone.

4.1 Analytical Comparison

We start by recalling the definition of the equity premium over a time interval $[0, t]$:

$$\mathrm{EP}(t) := \mathbb{E}\left[\frac{\hat{S}_t - \hat{S}_0 + \int_0^t D_u e^{\int_u^t r(s)ds} du}{\hat{S}_0}\right] - \left(e^{\int_0^t r(u)du} - 1\right), \quad t \in [0, T]. \tag{4.1}$$

In (4.1), the interest rate $r(t)$ is given in (3.11) with the corresponding (deterministic) money market account price process is $e^{\int_0^t r(u)du}$ and the equilibrium stock-price process $\hat{S}_t$ is given in (3.12). Based on (4.1), we define the Sharpe ratio measured over a time interval

$[0, t]$ as

$$\mathrm{SR}(t) := \frac{\mathrm{EP}(t)}{\mathbb{V}\Big[\frac{\hat{S}_t - \hat{S}_0 + \int_0^t D_u \frac{S_t^{(0)}}{S_u^{(0)}} du}{\hat{S}_0} - \big(e^{\int_0^t r(u)du} - 1\big)\Big]^{\frac{1}{2}}}, \quad t \in (0, T], \tag{4.2}$$

where $\mathbb{V}[\cdot]$ in the denominator in (4.2) is the variance operator. Because models based on randomness generated by Brownian motions produce expected returns and variances growing linear in t for $t > 0$ small, we consider the annualized Sharpe ratio defined by $\frac{\mathrm{SR}(t)}{\sqrt{t}}$ for a horizon $t \in (0, T]$. The *instantaneous* Sharpe ratio is defined as the limit $\lim_{t \downarrow 0} \frac{\mathrm{SR}(t)}{\sqrt{t}}$. The Sharpe ratios $\mathrm{SR}^{\mathrm{Radner}}(t)$ and $\mathrm{SR}^{\mathrm{Pareto}}(t)$ are defined analogously for the competitive Radner and Pareto-efficient equilibrium stock-price processes S^{Radner} from (C.2) and S^{Pareto} from (D.5) and interest rates r^{Radner} from (3.18) and r^{Pareto} from (D.4).

The following result orders the Nash, Radner, and Pareto equilibria analytically.

Corollary 4.1. *In the setting of Theorem 3.3 we have:*

(i) The Nash, Radner, and Pareto equilibrium interest rates are ordered as

$$r(t) \leq r^{Radner} \leq r^{Pareto}, \quad t \in [0, T], \tag{4.3}$$

where the first inequality becomes an equality if and only if the initial stock endowments are equal with $\theta_{i,0} = \frac{L}{I}$, and the second inequality becomes equality if and only if $\rho^2 = 1$.

(ii) The Nash, Radner, and Pareto equilibrium stock-price processes $\hat{S}_t$ in (3.12)*, S^{Radner} in* (C.2)*, and S^{Pareto} in* (D.5) *produce the volatility ordering measured by quadratic variation*

$$d\langle \hat{S} \rangle_t \geq d\langle S^{Radner} \rangle_t \geq d\langle S^{Pareto} \rangle_t, \tag{4.4}$$

with strict inequalities whenever the inequalities in (4.3) *are strict.*

(iii) The Nash, Radner, and Pareto equilibrium instantaneous Sharpe ratios are identically

given by λ *in* (3.16)*; that is,*

$$\lim_{t \downarrow 0} \frac{SR(t)}{\sqrt{t}} = \frac{a}{I}(L\sigma_D + I\sigma_Y \rho). \tag{4.5}$$

First, consider the risk-free rate puzzle of Weil (1989). Pareto-efficient equilibrium models predict interest rates that are too high compared to empirical evidence. From in (3.11) we see that Nash equilibrium interest rate $r(t)$ is less than both the analogous competitive interest rate r^{Radner} in (3.18), and the analogous Pareto-efficient interest rate r^{Pareto} in (D.4) in Appendix D. Whenever there is unspanned income risk (i.e., when $\rho^2 < 1$), Christensen, Larsen, and Munk (2012) show that $r^{\text{Radner}} < r^{\text{Pareto}}$ due to a precautionary saving effect. Here, we find $r(t) < r^{\text{Radner}}$ whenever there is stock-endowment inequality in that $\theta_{i,0} \neq L/I$ for some trader $i \in \{1, ..., I\}$.[10] The intuition is that price-impact costs cause investors to rebalance more slowly, which exacerbates risk-bearing inefficiency, which, in turn, magnifies stock risk and increases bond demand.

Second, consider the volatility puzzle of LeRoy and Porter (1981) and Shiller (1981). Pareto-efficient models predict a stock-price volatility that is too low compared to empirical evidence. The ordering (4.3) leads to a reversed annuity ordering:

$$F(t) \geq F^{\text{Radner}}(t) \geq F^{\text{Pareto}}(t), \tag{4.6}$$

where $F(t)$ is given by the ODE (3.2) and

$$\frac{d}{dt}F^{\text{Radner}}(t) = F^{\text{Radner}}(t)r^{\text{Radner}} - 1, \quad F^{\text{Radner}}(T) = 1, \tag{4.7}$$

$$\frac{d}{dt}F^{\text{Pareto}}(t) = F^{\text{Pareto}}(t)r^{\text{Pareto}} - 1, \quad F^{\text{Pareto}}(T) = 1. \tag{4.8}$$

Consequently, the ordering (4.6) and the equilibrium stock-price processes (3.12), (C.2), and (D.5) produce the volatility ordering measured by quadratic variation in (4.4). The intuition is that the multiplication of the current dividend D_t by $F(t)$ in (3.12) represents

[10] Even without idiosyncratic income risks (i.e., $\rho^2 = 1$ so that $r^{\text{Radner}} = r^{\text{Pareto}}$), we have $r(t) < r^{\text{Radner}}$.

an annuity-valuation effect for the stream of future dividends following D_t at time t. Thus, lower interest rates in the Nash equilibrium intensify this annuity effect.

Third, consider the equity premium puzzle of Mehra and Prescott (1985). Pareto-efficient models predict the stock's excess return over the risk-free rate to be too low compared to empirical evidence. The equity premium puzzle involves empirical Sharpe ratios estimated over discrete horizons (e.g., monthly or annually), so the finite-horizon $[0,t]$ Sharpe ratios $\frac{\mathrm{SR}(t)}{\sqrt{t}}$ is the relevant measure. Note here that the finite-horizon Sharpe ratio (4.2) is a ratio of integrals and not an integral of instantaneous Sharpe ratios. Thus, finite-horizon Sharpe ratios can differ across the three models (Nash, Radner, and Pareto) even though their instantaneous Sharpe ratios are identical. However, Section 4.2 below shows that, over small horizons $[0,t]$ for $t > 0$, the magnitudes of the finite-horizon Nash Sharpe ratio are quantitatively similar to the Radner and Pareto Sharpe ratios for a reasonable set of calibrated model parameters. The reason is that, as $t \downarrow 0$, the time-normalized discrete-horizon Sharpe ratios in all three models are anchored, by Corollary 4.1, to the same instantaneous Sharpe ratio λ in (4.5).[11] The fact that the Radner and Pareto instantaneous Sharpe ratios agree with exponential utilities, continuous-time consumption, and Brownian motions randomness is proved in Christensen and Larsen (2014, p.273) and is discussed in Cochrane (2008, p.310). The fact that this is also true for the Nash instantaneous Sharpe ratio is a new result. Over longer horizons, our Nash equilibrium model with price-impact can produce bigger Sharpe ratios (4.2) than the analogous Radner and Pareto equilibrium models but only for unreasonable parameters. Thus, the calibrated magnitude of the price-impact Sharpe ratio effect is small.

[11] There are several ways model incompleteness can produce a different instantaneous Sharpe ratio in the competitive (i.e., price-taking) Radner equilibrium model relative to the analogous efficient Pareto model: (i) Traders can be restricted to only consume discretely as in Constantinides and Duffie (1996), (ii) The underlying filtration can have jumps as in, e.g., Barro (2006) and Larsen and Sae-Sue (2016), and (iii) Non-time additive utilities as in, e.g., Bansal and Yaron (2004).

4.2 Numerics

This section presents calibrated numerics to illustrate the effect of price-impact on all three asset-pricing puzzles. In our numerics, time is measured on an annual basis (i.e., one year is $t = 1$). We normalize the outstanding stock supply to $L := 100$. As noted in Remark 3.1.3, the key quantity in explaining the asset pricing puzzles is the heterogeneity in investors' initial endowments as measured by the difference $\sum_{i=1}^{I} \theta_{i,0}^2 - \frac{L^2}{I} \geq 0$ which is a metric for the distance of the initial stock endowments from Pareto efficiency. To provide some intuition for this difference, we note that the cross-sectional average and standard deviation of a set of initial stock endowments $\vec{\theta}_0 := \{\theta_{1,0}, ..., \theta_{I,0}\}$ are

$$\begin{aligned} \text{mean}\big[\vec{\theta}_0\big] &:= \frac{1}{I}\sum_{i=1}^{I} \theta_{i,0} \\ &= \frac{L}{I}, \\ \text{SD}\big[\vec{\theta}_0\big] &:= \sqrt{\frac{1}{I}\Big(\sum_{i=1}^{I} \theta_{i,0}^2 - \frac{L^2}{I}\Big)} \\ &= \sqrt{\frac{1}{I}\big(\psi(0) - \frac{L^2}{I}\big)}, \end{aligned} \tag{4.9}$$

where $\psi(t)$ is the function from (3.1).

The utility parameters for (2.10) in our numerics are

$$\delta := 0.02, \quad a := 2. \tag{4.10}$$

The annual time-preference rate δ is consistent with calibrated time preferences in Bansal and Yaron (2004), and the level of absolute risk aversion a is from the numerics in Christensen, Larsen, and Munk (2012). The coefficients for the arithmetic Brownian motion for the stock dividends in (2.2) are

$$\mu_D := 0.0201672, \quad \sigma_D := 0.0226743, \quad D_0 := 1. \tag{4.11}$$

The parameterizations of μ_D and σ_D are the annualized mean and standard deviation of monthly percentage changes in aggregate real US stock market dividends from January 1970 through December 2019 (from Robert Shiller's website `http://www.econ.yale.edu/~shiller/data.htm`). The starting dividend rate $D_0 = 1$ in (4.11) is a normalization. The annualized income volatility and income-dividend correlation are from the numerics in Christensen, Larsen, and Munk (2012):

$$\sigma_Y = 0.1, \quad \rho = 0. \tag{4.12}$$

The drift μ_Y and number of investors $I \in \mathbb{N}$ are found by calibrating the Radner equilibrium model so that

$$\lambda = 0.302324, \quad r^{\text{Radner}} = 8.137\%, \tag{4.13}$$

which produces the remaining coefficients[12]

$$\mu_Y := -0.0709146, \quad I = 15. \tag{4.14}$$

We set the model horizon T to $T := 3$ years. In our analysis we found that our numerics are relatively insensitive to T once T is sufficiently large.

We illustrate that price-impact in the Nash equilibrium can have a material effect on asset pricing relative to the analogous Pareto-efficient equilibrium. Figure 1 shows interest-rate and stock return-volatility trajectories over a year $t \in [0, 1]$ for the Nash equilibrium with price-impact, the price-taking Radner equilibrium, and the corresponding Pareto-efficient equilibrium. For visibility, Figure 1 also shows differences in Sharpe ratios between the Nash and Radner equilibria since these numerical values are small.

The Nash model with price-impact has two additional parameters relative to the competitive Radner model: The transitory price-impact coefficient α in (3.10) and the difference

[12]The discount rate δ, dividend parameters μ_D and σ_D, and income parameters μ_Y and σ_Y are all quoted in decimal form where $0.01 = 1\%$.

$\sum_{i=1}^{I} \hat{\theta}_{i,t}^2 - L^2/I = \psi(t) - L^2/I$ for deviations of initial stock endowments from the equal stock holdings, which is related to the $\mathrm{SD}[\vec{\theta}_0]$ in (4.9). Figure 1 illustrates the sensitivity of asset pricing moments to these two parameters.

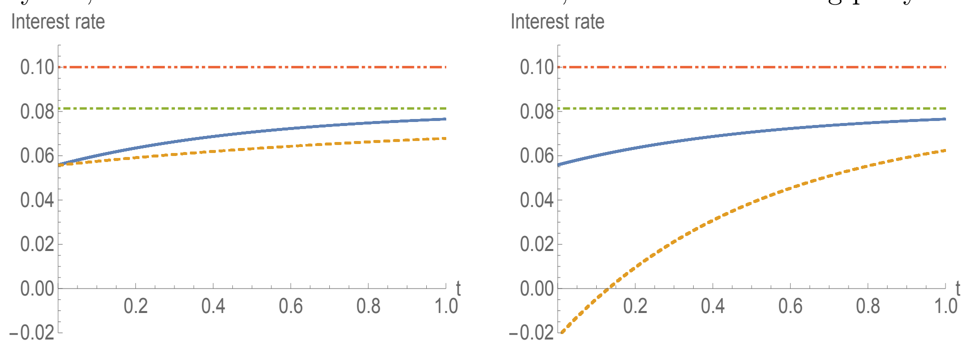
Figure 1, Plots A, C, and E show the effects of varying the temporary price-impact parameter $\alpha > 0$. Of course, when $\alpha > 0$ is close to zero, our Nash equilibrium is close to the Radner equilibrium. In our numerics, we consider two transitory price-impact parameters of $\alpha \in \{0.01, 0.002\}$. Appendix E shows that $\alpha = 0.002$ is roughly consistent with transitory price-impact estimates in Almgren et al. (2005). To put them in perspective, a price-impact of $\alpha = 0.002$ means if an investor trades at a constant rate $\theta_i' = 265$ to sell $\int_0^{\frac{1}{265}} \theta_i' dt = 1$ unit of the stock over a day (i.e., a large daily parent trade of 1 percent of $L = 100$ shares outstanding), the associated transitory price increase at each time t in the day would be $0.002 \times 265 = 0.53$. The stock (with $\alpha = 0.002$ and $\mathrm{SD}[\vec{\theta}_0] = 5$) has an endogenous initial equilibrium price of $\hat{S}_0 = 3.5737$, so this corresponds to a sustained percentage transitory price-impact of $\frac{0.002 \times 265}{3.5737} = 14.83\%$ over the day.

The price-impact feature in the Nash equilibrium can produce up to a 2% annual interest rate reduction (the reduction is biggest for shorter horizons where, from (3.17), holdings $\theta_{i,t}$ differ most from $\frac{L}{I}$). We see that the stock-return volatility increases by around 0.25% relative to the Radner volatility. The impact on the Sharpe ratio, while in the right direction qualitatively, is quantitatively negligible. The Sharpe ratio effects are biggest for longer horizons. Finally, from Plots A, C, and E, we see that all three asset-pricing impacts are increasing in the temporary price-impact coefficient $\alpha > 0$.

Figure 1, Plots B, D, and F consider the effect of different levels of stock-endowment inequality ($\mathrm{SD}[\vec{\theta}_0] \in \{5, 10\}$). As $\sum_{i=1}^{I} \theta_{i,0}^2$ approaches the calibrated lower bound $\frac{100^2}{15} \approx 666.67$ for $\frac{L^2}{I}$ from Cauchy-Schwarz's inequality, the Nash equilibrium converges to the Radner equilibrium. Plots B, D, and F show all three asset-pricing impacts are increasing in investor heterogeneity as measured by $\sum_{i=1}^{I} \theta_{i,0}^2 - \frac{L^2}{I}$.

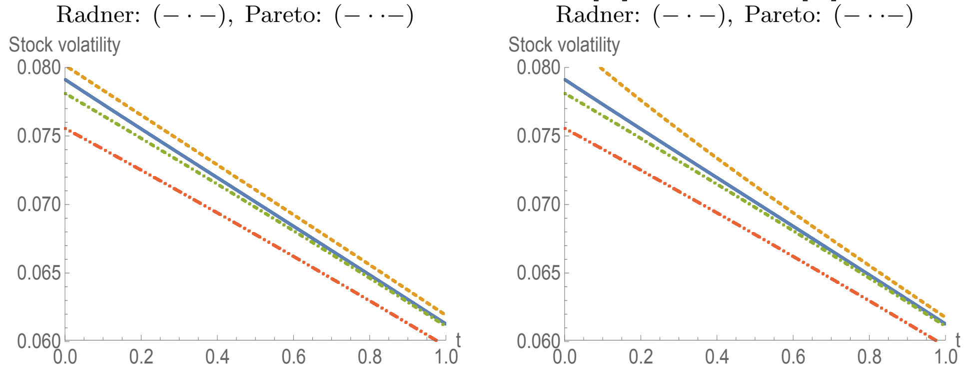
The small quantitative effect of price-impact on the equity premium in our analysis is not due to investor trading incentives being small. To the contrary, the incentive to trade is substantial in these numerical examples: The cross-investor endowment standard deviations are 5 and 10, relative to the Pareto-efficient stock holdings of $100/15 = 6.66$. This suggests stronger induced trading preferences given other utility specifications may be an important part of what is needed for price impact to affect the Sharpe Ratio. We leave this to future research, since exponential preferences are part of the modeling assumptions that make our current analysis tractable.

Figure 1: Trajectories of interest rates (Plots A and B), stock-price volatility (Plots C and D), and annualized Sharpe ratio differences $\frac{\mathrm{SR}(t)-\mathrm{SR}^{\mathrm{Radner}}(t)}{\sqrt{t}}$ (Plots E and F) for $t \in [0,1]$ over the first year. The model parameters are given in (4.10), (4.11), (4.12), (4.14), $L := 100$, $T := 3$ years, and the time discretization uses 250,000 rounds of trading per year.



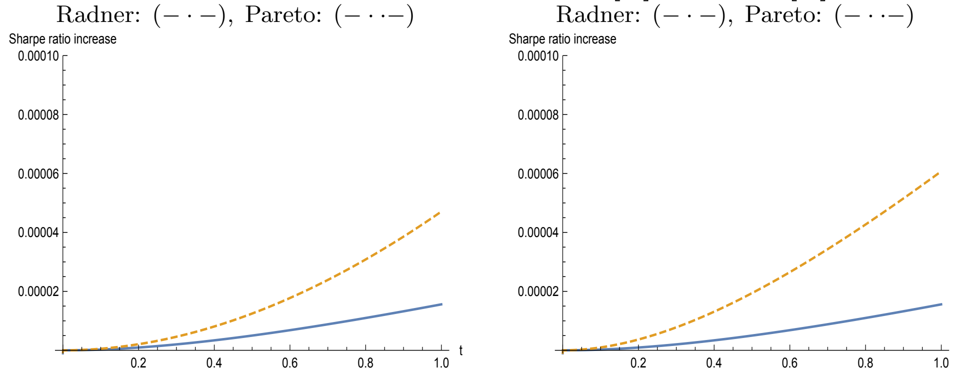
A: $\mathrm{SD}\left[\vec{\theta}_0\right] := 5$
Nash: $\alpha := 0.002$ (——), $\alpha := 0.01$ (−−),
Radner: (− · −), Pareto: (− · ·−)

B: $\alpha := 0.002$
Nash $\mathrm{SD}\left[\vec{\theta}_0\right] := 5$ (——), $\mathrm{SD}\left[\vec{\theta}_0\right] := 10$ (−−),
Radner: (− · −), Pareto: (− · ·−)



C: $\mathrm{SD}\left[\vec{\theta}_0\right] := 5$
Nash: $\alpha := 0.002$ (——), $\alpha := 0.01$ (−−),
Radner: (− · −), Pareto: (− · ·−)

D: $\alpha := 0.002$
Nash: $\mathrm{SD}\left[\vec{\theta}_0\right] := 5$ (——), $\left[\vec{\theta}_0\right] := 10$ (−−),
Radner: (− · −), Pareto: (− · ·−)



E: $\mathrm{SD}\left[\vec{\theta}_0\right] := 5$
Nash: $\alpha := 0.002$ (——), $\alpha := 0.01$ (−−)

F: $\alpha := 0.002$
Nash: $\mathrm{SD}\left[\vec{\theta}_0\right] := 5$ (——) , $\mathrm{SD}\left[\vec{\theta}_0\right] := 10$ (−−)

To summarize, Figure 1 shows that price-impact can have a significant impact on interest rates and a material impact on stock return volatility for reasonable model parameters. Moreover, we found that there are model parameterizations such that the quantitative effect of price-impact on the Sharpe ratio is material in our model. However, these other parameterizations involve unrealistically large price-impact and initial stock-holding heterogeneity and were also associated with unreasonable interest rates. Thus, we conclude that price-impact in our model only has a negligible impact on the finite-horizon equilibrium Sharpe ratios. Again, the intuition for this result is the common anchoring of the instantaneous Sharpe ratio in (4.5).

5 Model extensions

Our analysis has shown how to construct a parsimonious and tractable equilibrium model of price-impact in continuous-time. However, the following two model extensions illustrate our Nash equilibrium model's analytical robustness to variations.

5.1 Discrete-orders

We can allow traders to also place discrete orders (i.e., block trades) as well as consumption plans with lump sums. To illustrate this, we consider a simple case in which traders can place block trades and consume in lumps at time $t = 0$ after which they trade using order rates and consume using consumption rates for $t \in (0, T]$.

First, we start with block trades and use $\theta_{i,0-}$ to denote trader i's initial stock endowment so that $\Delta\theta_{i,0} := \theta_{i,0} - \theta_{i,0-}$ denotes the block trade at time $t = 0$. In addition to (2.11) for $t \in (0, T]$, we conjecture the response at time $t = 0$ for trader $j \neq i$ to be

$$\Delta\theta_{j,0} = \beta_0\big(F(0)D_0 - S_{i,0}\big) + \beta_1\theta_{j,0-} + \beta_2\theta_{i,0-} + \beta_3\Delta\theta_{i,0}, \tag{5.1}$$

where $(\beta_0, .., \beta_3)$ are constants (to be determined). The price-impact function trader i perceives is found using the stock-market clearing condition at time $t = 0$ when summing

(5.1):

$$
\begin{aligned}
0 &= (I-1)\beta_0\big(F(0)D_0 - S_{i,0}\big) + \beta_1(L-\theta_{i,0-}) \\
&\quad + (I-1)\big(\beta_2\theta_{i,0-} + \beta_3\Delta\theta_{i,0}\big) + \Delta\theta_{i,0}.
\end{aligned} \tag{5.2}
$$

Provided that $\beta_0 \neq 0$, we can solve (5.2) for trader i's perceived stock market-clearing price at time $t = 0$:

$$
S_{i,0} = D_0 F(0) + \frac{\beta_1 L}{\beta_0(I-1)} + \frac{\beta_2(I-1)-\beta_1}{\beta_0(I-1)}\theta_{i,0-} + \frac{\beta_3(I-1)+1}{\beta_0(I-1)}\Delta\theta_{i,0}. \tag{5.3}
$$

Second, we introduce time $t = 0$ lump sum consumption. Because stock prices are denoted ex dividend, the initial wealth is

$$
X_{i,0} = (D_0 + S_{i,0})\theta_{i,0-} + Y_{i,0} - C_i, \quad i \in \{1,...,I\}, \tag{5.4}
$$

where C_i is trader i's lump sum consumption at time $t = 0$ (to be determined). The expression for $X_{i,0}$ in (5.4) follows from the normalization that all strategic traders have zero endowments in the money market account. By using (5.4), the time $t = 0$ money market account balance of (2.15) for trader $i \in \{1,...,I\}$ is given by

$$
\begin{aligned}
M_{i,0} &:= X_{i,0} - S_{i,0}\theta_{i,0} \\
&= D_0\theta_{i,0-} - S_{i,0}\Delta\theta_{i,0} + Y_{i,0} - C_i.
\end{aligned} \tag{5.5}
$$

Next, we show how to modify to the objective in (2.10) to allow for both time $t = 0$ lump sum consumption C_i and block trades $\Delta\theta_{i,0}$. Trader i's optimization problem becomes:

$$
\begin{aligned}
&\inf_{(\Delta\theta_{i,0},C_i)\in\mathbb{R}^2,\,(\theta'_i,c_i)\in\mathcal{A}} \mathbb{E}\left[e^{-aC_i} + \int_0^T e^{-ac_{i,t}-\delta t}dt + e^{-a(X_{i,1}+Y_{i,T})-\delta T}\right] \\
&= \inf_{(\Delta\theta_{i,0},C_i)\in\mathbb{R}^2} \Big(e^{-aC_i} + v(0, M_{i,0}, D_0, \theta_{i,0}, Y_{i,0})\Big),
\end{aligned} \tag{5.6}
$$

where v is the value function defined below in (B.3) in Appendix B corresponding to the objective in (2.10). To minimize the objective in (5.6), we insert $M_{i,0}$ from (5.5) and $\theta_{i,0} = \theta_{i,0-} + \Delta\theta_{i,0}$ into the last line in (5.6) and minimize to produce the optimal initial block trade and lump sum consumption. For example, we have

$$\begin{aligned} \hat{S}_0 &= F(0)\Big(D_0 + \tfrac{LQ_{22}(0)}{I} + Q_2(0)\Big), \\ \hat{\theta}_{i,0} &= \theta_{i,0-} + \beta_1\Big(\theta_{i,0-} - \frac{L}{I}\Big), \end{aligned} \tag{5.7}$$

where β_1 is a free model parameter (similar to α in Theorem 3.3). From (5.7), we see that $\hat{S}_0$ matches the initial stock price in (3.12). Moreover, because of price-impact, we also see from (5.7) that trader i does not immediately jump to the Pareto efficient holdings $\frac{L}{I}$. Finally, we note that Corollary 4.1 about the asset pricing effects of price-impact continues to hold when $\sum_{i=1}^{I} \theta_{i,0}$ is replaced by

$$\sum_{i=1}^{I} \hat{\theta}_{i,0} = \sum_{i=1}^{I} \Big(\theta_{i,0-} + \beta_1\big(\theta_{i,0-} - \frac{L}{I}\big)\Big). \tag{5.8}$$

5.2 Penalties

In this section our model is extended to allow for inventory penalties and dissipative transaction costs. To do this, the objective (2.10) is replaced with

$$\inf_{(\theta_i', c_i)\in\mathcal{A}} \mathbb{E}\left[\int_0^T e^{-ac_{i,t}-\delta t}dt + e^{-a(X_{i,1}+Y_{i,T}-L_{i,T})-\delta T}\right], \quad i = 1, ..., I, \tag{5.9}$$

where $L_{i,T}$ is a penalty term. We consider two specifications of L_i. First, we can incorporate high-frequency traders (HFTs) who are incentivized to hold zero positions over time. We do this by defining the penalty processes:

$$L_{i,t} := \int_0^t \kappa(s)\theta_{i,s}^2 ds, \quad t \in [0, T], \quad i = 1, ..., I. \tag{5.10}$$

The deterministic function $\kappa : [0, T] \to [0, \infty)$ in (5.10) is a penalty-severity function. The strength of $\kappa(t)$ for $t \in [0, T]$ can vary periodically for times during overnight periods vs during trading days to give HFTs stronger incentive to hold no stocks overnight.

Second, we can approximate deadweight transaction costs by penalizing trading rates (as in, e.g., Gârleanu and Pedersen 2016). We do this by defining the penalty processes:

$$L_{i,t} := \frac{1}{2}\lambda \int_0^t (\theta'_{i,s})^2 ds, \quad t \in [0,T], \quad i = 1, ..., I. \tag{5.11}$$

The constant $\lambda > 0$ in (5.11) is interpreted as a transaction cost parameter.

By altering the ODEs, the existence result in Theorem 3.3 can be modified to include both penalties (5.10) and (5.11) and linear combinations of (5.10) and (5.11). For this model extension, the orderings in Corollary 4.1 also continue to hold.

6 Conclusion

This paper has shown, formally and in numerical examples, that price-impact can have material effects on asset pricing via an amplification effect on imperfect risk sharing. Calibrated price-impact helps resolve both the interest rate and volatility puzzles but has only a negligible quantitative effect on the Sharpe ratios for realistic model parameters.

Our analysis suggests many interesting directions for future work. These include extending our analysis to non-exponential utility, non-Gaussian dividends, and refinements of Nash equilibrium for off-equilibrium price perceptions and then investigating whether such variations relative to our baseline analysis can increase the quantitative effect of price impact on the Sharpe Ratio.

A Auxiliary ODE result

In the following ODE existence proof, there are no restrictions on the time horizon $T \in (0,\infty)$ and the constant $C_0 \in \mathbb{R}$. We note that the ODE (A.3) is quadratic in $g(t)$ and that the square coefficient $-\frac{2}{\alpha}$ is negative because $\alpha > 0$ is the temporary price-impact due to orders $\theta'_{i,t}$.

Proposition A.1. *For $I \in \mathbb{N}$, $C_0 \in \mathbb{R}$, and positive constants $T, a, \sigma_D, \alpha, k > 0$ there exists a unique constant $\hat{h}_0 \in (0, k)$ such that the ODE system:*

$$h'(t) = \frac{2g(t)}{\alpha} h(t), \quad h(0) = h_0, \tag{A.1}$$

$$f'(t) = 1 + f(t) \left(\tfrac{a^2 \sigma_D^2}{2I} h(t) - C_0 \right), \quad f(0) = 1, \tag{A.2}$$

$$g'(t) = a\sigma_D^2 f(t) - \frac{2}{\alpha} g(t)^2 + g(t) \left(\tfrac{a^2 \sigma_D^2}{2I} h(t) - C_0 \right), \quad g(0) = 0, \tag{A.3}$$

with initial condition $h_0 := \hat{h}_0$, has a unique solution for $t \in [0, T]$ that satisfies $h(T) = k$.

Proof.

Step 1/3 (h's range): Let $h_0 \in (0, k)$ be given. We evolve the ODEs (A.1)-(A.3) from $t = 0$ to the right ($t > 0$). The local Lipschitz property of the ODEs ensure that there exists a maximal interval of existence $[0, \tau)$ with $\tau \in (0, \infty]$ by the Picard-Lindelöf theorem (see, e.g., Theorem II.1.1 in Hartman 2002).

For a constant c, let $T_{f=c} \in [0, \tau]$ be defined as

$$T_{f=c} := \inf \{ t \in (0, \tau) : \, f(t) = c \} \wedge \tau, \tag{A.4}$$

where — as usual — the infimum over the empty set is defined as $+\infty$. We define $T_{g=c}$ and $T_{h=c}$ similarly. Suppose that $T_{f=0} < \tau$. Then, $f(0) = 1$ and the continuity of f imply that $f(t) > 0$ for $t \in [0, T_{f=0})$. Since $f(T_{f=0}) = 0$, we have $f'(T_{f=0}) \leq 0$, but (A.2) implies $f'(T_{f=0}) = 1 > 0$. Therefore, we conclude that

$$T_{f=0} = \tau \quad \text{and} \quad f(t) > 0 \text{ for } t \in [0, \tau). \tag{A.5}$$

Because $g(0) = 0$ and $g'(0) = a\sigma_D^2 > 0$, we have $T_{g=0} > 0$ and $g(t) > 0$ for $t \in (0, T_{g=0})$. The ODE (A.1) with $h(0) = h_0 > 0$ implies that $h(t)$ increases on the interval $[0, T_{g=0})$.

Therefore, the ODE (A.2) and the positivity of (f, h) produce

$$f'(t) > 1 - f(t)C_0, \quad t \in [0, T_{g=0}). \tag{A.6}$$

Then, Gronwall's inequality produces

$$f(t) \geq \begin{cases} \frac{1+(C_0-1)e^{-C_0 t}}{C_0} & \text{if } C_0 \neq 0, \\ 1+t & \text{if } C_0 = 0. \end{cases} \tag{A.7}$$

This inequality implies that

$$f(t) \geq C_1 \quad \text{for} \quad t \in [0, T_{g=0}) \quad \text{where} \quad C_1 := \begin{cases} 1, & \text{if } C_0 \leq 1, \\ \frac{1}{C_0}, & \text{if } C_0 > 1. \end{cases} \tag{A.8}$$

Suppose that $T_{g=0} < \tau$. Since $g(t) > 0$ for $t \in (0, T_{g=0})$ and $g(T_{g=0}) = 0$, we have $g'(T_{g=0}) \leq 0$. However, this is a contradiction because (A.3) and (A.8) imply $g'(T_{g=0}) \geq a\sigma_D^2 C_1 > 0$ where the positive constant C_1 is defined in (A.8).

Up to this point we have shown

$$T_{g=0} = \tau \quad \text{and} \quad \begin{cases} f(t) \geq C_1 > 0, \\ h(t) \geq 0, \\ g(t) \geq 0, \end{cases} \quad \text{for } t \in [0, \tau). \tag{A.9}$$

To proceed, the positive constant

$$C_2 := \begin{cases} -\frac{\alpha C_0}{2}, & \text{if } C_0 < 0, \\ \frac{\alpha\left(-C_0 + \sqrt{C_0^2 + \frac{4a\sigma_D^2 C_1}{\alpha}}\right)}{4}, & \text{if } C_0 \geq 0 \end{cases} \tag{A.10}$$

satisfies

$$-\frac{2}{\alpha}x^2 - C_0 x \geq -\frac{a\sigma_D^2 C_1}{2} \quad \text{for} \quad x \in [0, C_2]. \tag{A.11}$$

Because $0 \leq g(t) < C_2$ for $t \in [0, T_{g=C_2})$, we can bound (A.3) from below using (A.9) and (A.11) to see for $t \in [0, T_{g=C_2})$

$$\begin{aligned} g'(t) &\geq a\sigma_D^2 C_1 - \frac{2g(t)^2}{\alpha} - g(t)C_0 \\ &\geq \frac{1}{2}a\sigma_D^2 C_1. \end{aligned} \tag{A.12}$$

By integrating (A.12) and using the initial condition $g(0) = 0$ we see $g(t) \geq \frac{1}{2}a\sigma_D^2 C_1 t$ for $t \in [0, T_{g=C_2})$. Therefore,

$$T_{g=C_2} \leq \tfrac{2C_2}{a\sigma_D^2 C_1}. \tag{A.13}$$

Suppose that $T_{g=C_2} = \tau$. Then, for $t \in [0, \tau)$, we have $0 \leq g(t) < C_2$ and the ODE (A.1) produces

$$\begin{aligned} h'(t) &\leq \frac{2C_2}{\alpha} h(t), \\ h(t) &\leq h_0 \, e^{\frac{2C_2}{\alpha} t}, \end{aligned} \tag{A.14}$$

where the second inequality uses Gronwall's inequality. Similarly, for $t \in [0, \tau)$, the ODE (A.2) and Gronwall's inequality imply

$$\begin{aligned} f'(t) &\leq 1 + f(t)\left(\tfrac{a^2\sigma_D^2}{2I} h(t) + |C_0|\right) \\ &\leq 1 + f(t)\left(\tfrac{a^2\sigma_D^2 h_0}{2I} e^{\frac{2C_2}{\alpha} t} + |C_0|\right), \\ f(t) &\leq (1+t)\exp\left(|C_0|t + \tfrac{a^2\sigma_D^2 h_0 \alpha}{4IC_2}(e^{\frac{2C_2}{\alpha} t} - 1)\right). \end{aligned} \tag{A.15}$$

The boundedness properties $g(t) < C_2$, (A.14), and (A.15) imply that h, f, and g do not blow up for t finite. Then, Theorem II.3.1 in Hartman (2002) ensures $\tau = \infty$ which contradicts

(A.13). Consequently, we cannot have $T_{g=C_2} = \tau$ and it must be the case that

$$T_{g=C_2} < \tau. \tag{A.16}$$

Let $\hat{T}_{g=C_2}$ be defined as the first time g reaches C_2 strictly after time $t = T_{g=C_2}$; that is,

$$\hat{T}_{g=C_2} := \inf\left\{t \in (T_{g=C_2}, \tau) : \; g(t) = C_2\right\} \wedge \tau. \tag{A.17}$$

Because $g'(T_{g=C_2}) \geq \frac{a\sigma_D^2 C_1}{2} > 0$ by (A.12), we have

$$T_{g=C_2} < \hat{T}_{g=C_2} \text{ and } g(t) > C_2 \text{ for } t \in (T_{g=C_2}, \hat{T}_{g=C_2}). \tag{A.18}$$

Suppose that $\hat{T}_{g=C_2} < \tau$. Then, $g(\hat{T}_{g=C_2}) = C_2$ and (A.18) imply that $g'(\hat{T}_{g=C_2}) \leq 0$, but (A.3), (A.9), and (A.11) produce the contradiction:

$$\begin{aligned} g'(\hat{T}_{g=C_2}) &= a\sigma_D^2 f(\hat{T}_{g=C_2}) - \frac{2C_2^2}{\alpha} + C_2\left(\tfrac{a^2\sigma_D^2}{2I} h(\hat{T}_{g=C_2}) - C_0\right) \\ &\geq a\sigma_D^2 C_1 - \frac{2C_2^2}{\alpha} - C_2 C_0 \\ &\geq \tfrac{a\sigma_D^2 C_1}{2} \\ &> 0. \end{aligned} \tag{A.19}$$

Therefore, it must be the case that $\hat{T}_{g=C_2} = \tau$, which implies the lower bound

$$g(t) \geq C_2 > 0 \quad \text{for} \quad t \in [T_{g=C_2}, \tau). \tag{A.20}$$

Combining (A.12) and (A.20) gives the following global lower bound:

$$g(t) \geq \tfrac{a\sigma_D^2 C_1}{2} t \wedge C_2 \quad \text{for} \quad t \in [0, \tau). \tag{A.21}$$

In turn, using the ODE (A.1), the bound (A.21) produces the global lower bound for h via

Gronwall's inequality:

$$h(t) \geq h_0 \exp\left(\frac{2}{\alpha}\int_0^t \frac{a\sigma_D^2 C_1}{2} s \wedge C_2 ds\right) \quad \text{for} \quad t \in [0,\tau). \tag{A.22}$$

Next, we suppose $T_{h=k} = \tau$. Then, for $t \in [0,\tau)$, we have $0 \leq h(t) < k$, and the ODEs (A.2)-(A.3) and Gronwall's inequality imply

$$\begin{aligned} f'(t) &\leq 1 + f(t)C_3, \\ f(t) &\leq (1+t)e^{C_3 t}, \\ g'(t) &\leq a\sigma_D^2 f(t) + g(t)C_3 \\ &\leq a\sigma_D^2(1+t)e^{C_3 t} + g(t)C_3, \\ g(t) &\leq a\sigma_D^2 e^{C_3 t}(t + \tfrac{1}{2}t^2), \end{aligned} \tag{A.23}$$

where $C_3 := \frac{a^2\sigma_D^2}{2I}k + |C_0|$. The inequalities in (A.23) and $0 \leq h(t) \leq k$ imply that h, f, and g do not blow up for t finite. Then, Theorem II.3.1 in Hartman (2002) ensures $\tau = T_{h=k} = \infty$. This is a contradiction because (A.22) implies that $h(t)$ reaches k in finite time. Therefore, it must be the case that

$$T_{h=k} < \tau. \tag{A.24}$$

Step 2/3 (Monotonicity): Let $0 < h_0 < \tilde{h}_0 < k$, and denote the solution of the ODE system (A.1)-(A.3) with the initial condition $h(0) = \tilde{h}_0$ by $\tilde{f}, \tilde{h}$, and $\tilde{g}$. The corresponding maximal existence interval is denoted by $\tilde{\tau}$. We define $T_{g=\tilde{g}}$ as

$$T_{g=\tilde{g}} := \inf\{t \in (0, \tau \wedge \tilde{\tau}) : g(t) = \tilde{g}(t)\} \wedge \tau \wedge \tilde{\tau}. \tag{A.25}$$

Because $g(0) = \tilde{g}(0) = 0$, the ODEs (A.1)-(A.3) have the properties $g'(0) = \tilde{g}'(0) = a\sigma_D^2$

and

$$\begin{aligned} g''(0) &= a\sigma_D^2 \left(1 + \tfrac{a^2\sigma_D^2}{I} h_0 - 2C_0\right) \\ &< a\sigma_D^2 \left(1 + \tfrac{a^2\sigma_D^2}{I} \tilde{h}_0 - 2C_0\right) \\ &= \tilde{g}''(0). \end{aligned}$$

Therefore,

$$0 < g(t) < \tilde{g}(t) \quad \text{for} \quad t \in (0, T_{g=\tilde{g}}). \tag{A.26}$$

Suppose that $T_{g=\tilde{g}} < \tau \wedge \tilde{\tau}$. The inequality (A.26) and the ODEs (A.1) and (A.2) imply that

$$\begin{cases} h(t) < \tilde{h}(t) \\ f(t) < \tilde{f}(t) \end{cases} \quad \text{for} \quad t \in (0, T_{g=\tilde{g}}]. \tag{A.27}$$

Also, (A.26) and $g(T_{g=\tilde{g}}) = \tilde{g}(T_{g=\tilde{g}})$ produce $g'(T_{g=\tilde{g}}) \geq \tilde{g}'(T_{g=\tilde{g}})$. However, this contradicts

$$\begin{aligned} g'(T_{g=\tilde{g}}) &= a\sigma_D^2 f(T_{g=\tilde{g}}) - \frac{2g(T_{g=\tilde{g}})^2}{\alpha} + g(T_{g=\tilde{g}}) \left(\tfrac{a^2\sigma_D^2}{2I} h(T_{g=\tilde{g}}) - C_0\right) \\ &< a\sigma_D^2 \tilde{f}(T_{g=\tilde{g}}) - \frac{2\tilde{g}(T_{g=\tilde{g}})^2}{\alpha} + \tilde{g}(T_{g=\tilde{g}}) \left(\tfrac{a^2\sigma_D^2}{2I} \tilde{h}(T_{g=\tilde{g}}) - C_0\right) \\ &= \tilde{g}'(T_{g=\tilde{g}}), \end{aligned} \tag{A.28}$$

where we used (A.3) and (A.27). Therefore, we conclude that $T_{g=\tilde{g}} = \tau \wedge \tilde{\tau}$ and

$$\begin{cases} h(t) < \tilde{h}(t) \\ f(t) < \tilde{f}(t) \\ g(t) < \tilde{g}(t) \end{cases} \quad \text{for} \quad t \in (0, \tau \wedge \tilde{\tau}). \tag{A.29}$$

Step 3/3 (Existence): To emphasize the dependence on the initial condition $h(0) = h_0$, we write $\tau(h_0)$ and $T_{h=k}(h_0)$. For example,

$$T_{h=k}(h_0) := \inf\big\{t \in \big(0, \tau(h_0)\big) : \, h(t) = k\big\} \wedge \tau(h_0). \tag{A.30}$$

Inequality (A.24) in Step 1 implies that $T_{h=k}(h_0) < \infty$ for $h_0 \in (0, k)$. Step 2 implies that the map $(0, k) \ni h_0 \mapsto T_{h=k}(h_0)$ is strictly decreasing. Therefore, the following three statements and the Intermediate Value Theorem complete the proof in the sense that we can choose a unique $\hat{h}_0 \in (0, k)$ such that $T_{h=k}(\hat{h}_0) = T$ (recall that $T \in (0, \infty)$ is the model time horizon):

(i) $\lim_{h_0 \uparrow k} T_{h=k}(h_0) = 0$.

(ii) $\lim_{h_0 \downarrow 0} T_{h=k}(h_0) = \infty$.

(iii) The map $(0, k) \ni h_0 \mapsto T_{h=k}(h_0)$ is continuous.

Here are the proofs of these three statements:

(i) Inequality (A.22) implies (i).

(ii) The inequalities in (A.23) and Gronwall's inequality produce

$$\begin{aligned} h(t) &= h_0 \exp\left(\int_0^t \tfrac{2g(s)}{\alpha} ds\right) \\ &\leq h_0 \exp\left(\int_0^t \tfrac{2a\sigma_D^2 e^{C_3 s}(s+\frac{1}{2}s^2)}{\alpha} ds\right). \end{aligned} \tag{A.31}$$

Obviously, the function $[0, \infty) \ni t \to \exp\big(\int_0^t \frac{2a\sigma_D^2 e^{C_3 s}(s+\frac{1}{2}s^2)}{\alpha} ds\big)$ is increasing. Therefore, for any $t_0 > 0$, we can choose $h_0 > 0$ such that

$$h_0 < k \exp\left(-\int_0^t \frac{2a\sigma_D^2 e^{C_3 s}(s+\frac{1}{2}s^2)}{\alpha} ds\right), \quad t \in [0, t_0],$$

and use (A.31) to see $T_{h=k}(h_0) > t_0$. This shows (ii).

(iii) Let $h_0 \in (0, k)$ be fixed. To emphasize the dependence on the initial condition, we write $\big(h(t), g(t)\big)$ as $\big(h(t, h_0), g(t, h_0)\big)$. The local Lipschitz structure of the ODEs (A.1)-

(A.3) gives us the continuous dependence of their solutions on the initial condition h_0 (see, e.g., Theorem V.2.1 in Hartman 2002); that is,

$$\lim_{x\to h_0} h(t,x) = h(t,h_0), \quad t \in \big[0, \tau(h_0)\big). \tag{A.32}$$

For $0 < x < h_0$ we have $T_{h=k}(h_0) < T_{h=k}(x)$, and the ODE (A.1) and the Fundamental Theorem of Calculus produce:

$$\begin{aligned}
k &= h\big(T_{h=k}(x), x\big) \\
&= h\big(T_{h=k}(h_0), x\big) + \int_{T_{h=k}(h_0)}^{T_{h=k}(x)} \tfrac{\partial}{\partial t} h(t,x)dt \\
&= h\big(T_{h=k}(h_0), x\big) + \frac{2}{\alpha}\int_{T_{h=k}(h_0)}^{T_{h=k}(x)} g(t,x)h(t,x)dt \\
&\geq h\big(T_{h=k}(h_0), x\big) + \frac{2x}{\alpha}\int_{T_{h=k}(h_0)}^{T_{h=k}(x)} \big(\tfrac{a\sigma_D^2 C_1}{2} t \wedge C_2\big) e^{\frac{2}{\alpha}\int_0^t \frac{a\sigma_D^2 C_1}{2} s\wedge C_2 ds} dt \\
&\geq h\big(T_{h=k}(h_0), x\big) + xC_4\big(T_{h=k}(x) - T_{h=k}(h_0)\big),
\end{aligned} \tag{A.33}$$

where the second last line uses the bounds (A.21) and (A.22) and $C_4 > 0$ is an irrelevant constant independent of x. Letting $x \uparrow h_0$ and using (A.32) produce

$$\lim_{x\uparrow h_0} T_{h=k}(x) \leq T_{h=k}(h_0). \tag{A.34}$$

The opposite inequality trivially holds because $T_{h=k}(x)$ is strictly decreasing. Therefore, (A.34) holds with equality. Similarly, for $x \in (h_0, \frac{k+h_0}{2})$, we have $T_{h=k}(\frac{k+h_0}{2}) < T_{h=k}(x) < T_{h=k}(h_0)$ and

$$\begin{aligned}
h\big(T_{h=k}(h_0), x\big) &= h\big(T_{h=k}(x), x\big) + \int_{T_{h=k}(x)}^{T_{h=k}(h_0)} \tfrac{\partial}{\partial t} h(t,x)dt \\
&= k + \int_{T_{h=k}(x)}^{T_{h=k}(h_0)} g(t,x)h(t,x)dt \\
&\geq k + xC_5\big(T_{h=k}(h_0) - T_{h=k}(x)\big),
\end{aligned} \tag{A.35}$$

for a constant C_5 independent of x. Letting $x \downarrow h_0$ and using (A.32) produce

$$\lim_{x \downarrow h_0} T_{h=k}(x) \geq T_{h=k}(h_0). \tag{A.36}$$

Again, the opposite inequality trivially holds because $T_{h=k}(x)$ is strictly decreasing. Therefore, (A.36) holds with equality and the continuity property follows.

$\diamondsuit$

Proposition A.2. *Let* $h_0 = 0$ *in* (A.1)*. Then, the ODEs* (A.1)-(A.3) *have unique solutions on* $t \in [0, \infty)$ *with* $h(t) = 0$ *for all* $t \geq 0$*.*

Proof. As in the proof of Proposition A.1, denote the maximal interval of existence by $(0, \tau)$ for $\tau \in (0, \infty]$. For $t \in [0, \tau)$, the solutions to (A.1) and (A.2) are

$$\begin{aligned} &h(t) = 0, \\ &f(t) = \begin{cases} \frac{1+(C_0-1)e^{-C_0 t}}{C_0} & \text{if } C_0 \neq 0 \\ 1 + t & \text{if } C_0 = 0 \end{cases}. \end{aligned} \tag{A.37}$$

As in the proof of Proposition A.1, we can check that

$$g(t) \geq 0 \quad \text{for} \quad t \in [0, \tau). \tag{A.38}$$

Then (A.3), (A.37), and (A.38) imply that for $t \in [0, \tau)$,

$$\begin{aligned} g'(t) &= a\sigma_D^2 f(t) - \tfrac{2g(t)^2}{\alpha} - C_0 g(t) \\ &\leq a\sigma_D^2 f(t) + \tfrac{\alpha C_0^2}{8}. \end{aligned} \tag{A.39}$$

Gronwall's inequality implies that g cannot blow up in finite time. Therefore, we conclude that $\tau = \infty$.

$\diamondsuit$

B Proof of Lemma 3.2, Theorem 3.3, and Corollary 4.1

Proof of Lemma 3.2. We prove that the coupled ODEs (3.1), (3.2), and (3.4) have unique solutions for $t \in [0, T]$. We apply Proposition A.1 and Proposition A.2 with

$$\begin{aligned} C_0 &:= \delta - \frac{a(-2L\mu_D - 2I\mu_Y + 2aL\rho\sigma_D\sigma_Y + aI\sigma_Y^2)}{2I} - \frac{a^2\sigma_D^2 L^2}{2I^2}, \\ k &:= \sum_{i=1}^{I} \theta_{i,0}^2 - \frac{L^2}{I}, \end{aligned} \tag{B.1}$$

where k is non-negative by Cauchy-Schwarz's inequality. The functions

$$\psi(t) := h(T-t) + \tfrac{L^2}{I}, \quad F(t) := f(T-t), \quad Q_{22}(t) := -\frac{g(T-t)}{f(T-t)}, \tag{B.2}$$

solve (3.1), (3.2), and (3.4) for $t \in [0, T]$.

From (A.8) in the proof of Proposition A.1, we know that $f(t)$ is bounded away from zero for $t \in [0, T]$. Therefore, the solutions to the linear ODEs for $Q(t)$ and $Q_2(t)$ in (3.5) and (3.3) can be found by integration.

$\diamondsuit$

Proof of Theorem 3.3.

Step 1/2 (Individual optimality): In this step, we define the function

$$v(t, M_i, D, \theta_i, Y_i) := e^{-a\left(\frac{M_i}{F(t)} + D\theta_i + Y_i + Q(t) + Q_2(t)\theta_i + \frac{1}{2}Q_{22}(t)\theta_i^2\right)}, \tag{B.3}$$

for $t \in [0, T]$ and $M_i, D, \theta_i, Y_i \in \mathbb{R}$. In (B.3), the deterministic functions are defined in (3.2)-(3.4). We note the terminal ODE conditions produce

$$v(T, M_i, D, \theta_i, Y_i) = e^{-a(M_i + D\theta_i + Y_i)}. \tag{B.4}$$

Consequently, because $S_{i,T} = D_T$, we have

$$e^{-\delta T} v(T, M_{i,T}, D_T, \theta_{i,T}, Y_{i,T}) = e^{-a(X_{i,T}+Y_{i,T})-\delta T}, \tag{B.5}$$

which is the terminal condition in (2.10). Next, we show that the function $e^{-\delta t}v$ with v defined in (B.3) is the value function for (2.10). To see this, let $(\theta'_i, c_i) \in \mathcal{A}$ be arbitrary. Itô's lemma shows that the process $e^{-\delta t}v + \int_0^t e^{-ac_{i,u}-\delta u}du$ — with v being shorthand notation for the process $v(t, M_{i,t}, D_t, \theta_{i,t}, Y_{i,t})$ — has dynamics

$$\begin{aligned}
&d\big(e^{-\delta t}v\big) + e^{-ac_{i,t}-\delta t}dt \\
&= e^{-\delta t}v\Big(e^{a(-c_{i,t}+D_t\theta_{i,t}+\frac{M_{i,t}}{F(t)}+Q(t)+\theta_{i,t}Q_2(t)+\frac{1}{2}\theta_{i,t}^2Q_{22}(t)+Y)} \\
&- \frac{-a\alpha(\theta'_{i,t})^2 - ac_{i,t} + aD_t\theta_{i,t} + aQ(t) + a\theta_{i,t}Q_2(t) + \frac{1}{2}a\theta_{i,t}^2Q_{22}(t) + aY - \log\Big(\frac{1}{F(t)}\Big) + 1}{F(t)} \\
&+ \frac{aF(t)Q_{22}(t)^2(L-\theta_{i,t}I)^2}{\alpha I^2} - \frac{aM_{i,t}}{F(t)^2} + \frac{2a\theta'_{i,t}Q_{22}(t)(L-\theta_{i,t}I)}{I}\Big)dt \\
&- ae^{-\delta t}v\Big(\theta_{i,t}\sigma_D dB_t + \sigma_Y\big(\rho dB_t + \sqrt{1-\rho^2}dW_{i,t}\big)\Big),
\end{aligned} \tag{B.6}$$

where we have used the ODEs (3.1)-(3.4) and the interest rate (3.11). The local martingale on the last line in (B.6) can be upgraded to a martingale. To see this, we note that $\theta_{i,t}$ is bounded and v is square integrable by (2.9) so we can use Cauchy-Schwarz's inequality to obtain the needed integrability. Furthermore, to see that the drift in (B.6) is non-negative, we note the second-order conditions for the HJB equation are (there are no cross terms)

$$\begin{aligned}
\theta'_{i,t}: &\quad \frac{a\alpha}{F(t)} > 0, \\
c_{i,t}: &\quad a^2 e^{-ac_{i,t}} > 0.
\end{aligned} \tag{B.7}$$

This first inequality in (B.7) holds because $F(t)$ in (3.2) is the annuity (> 0). Consequently, the drift in (B.6) is minimized to zero by the controls (3.13) and (3.14). This implies that $e^{-\delta t}v + \int_0^t e^{-ac_{i,u}-\delta u}du$ is a submartingale for all admissible order-rate and consumption processes $\theta'_{i,t}$ and $c_{i,t}$.

It remains to verify admissibility of the controls (3.13) and (3.14). The explicit solution

(3.17) is deterministic and uniformly bounded. Inserting the controls (3.13) and (3.14) into the money market account balance dynamics (2.16) produces

$$
\begin{aligned}
dM_{i,t} &= \Big(r(t)M_{i,t} + \hat{\theta}_{i,t}D_t - \hat{S}_t\hat{\theta}'_{i,t} + (Y_{i,t} - \hat{c}_{i,t})\Big)dt \\
&= \Big(\frac{\log(\frac{1}{F(t)})}{a} + M_{i,t}\big(r(t) - \tfrac{1}{F(t)}\big) - Q(t) \\
&\quad - \frac{1}{2}\hat{\theta}_{i,t}\big(2Q_2(t) + \hat{\theta}_{i,t}Q_{22}(t)\big) - \hat{S}_t\hat{\theta}'_{i,t}\Big)dt.
\end{aligned}
\tag{B.8}
$$

The linear SDE (B.8) has a unique well-defined (Gaussian) solution that satisfies (2.9). All in all, this shows the admissibility requirements in Definition 2.1 and, hence, optimality of (3.13) and (3.14) follows from the martingale property of $e^{-\delta t}v + \int_0^t e^{-a\hat{c}_{i,u}-\delta u}du$.

Step 2/2 (Clearing): Clearly, summing the optimal orders in (3.13) and using $\sum_{i=1}^I \theta_{i,0} = L$ show that the stock market clears for all $t \in [0,T]$. Summing (3.14) gives us

$$
\sum_{i=1}^I \hat{c}_{i,t} = I\frac{\log\big(F(t)\big)}{a} + D_tL + IQ(t) + LQ_2(t) + \frac{1}{2}Q_{22}(t)\sum_{i=1}^I \hat{\theta}^2_{i,t} + \sum_{i=1}^I Y_{i,t}. \tag{B.9}
$$

Because $\psi(0) = \sum_{i=1}^I \theta^2_{i,0}$ and $\sum_{i=1}^I \hat{\theta}^2_{i,t}$ satisfies the ODE (3.1), we have $\psi(t) = \sum_{i=1}^I \hat{\theta}^2_{i,t}$ for all $t \in [0,T]$. Therefore, the real good market clears if and and only if

$$
0 = I\frac{\log\big(F(t)\big)}{a} + IQ(t) + LQ_2(t) + \frac{1}{2}Q_{22}(t)\psi(t). \tag{B.10}
$$

The terminal conditions in the ODEs (3.5)-(3.4) ensure clearing holds at time $t = T$. By computing time derivatives in (B.10) and using $r(t)$ defined in (3.11), we see that clearing holds for all $t \in [0,T]$.

Finally, the terminal stock-price condition (2.3) for the equilibrium stock-price process $\hat{S}_t$ in (3.12) holds by the terminal conditions in the ODEs (3.2), (3.3), and (3.4).

$\diamondsuit$

Proof of Corollary 4.1:

(i): The expressions in (3.11) and (3.18) imply that the inequality $r(t) \leq r^{\text{Radner}}$ is equivalent to the inequality $\psi(t) \geq \frac{L^2}{I}$. From the proof of Proposition A.1 and Proposition A.2, we know that the function h is either strictly positive or identically equal to zero on the interval $[0, T]$. Therefore, the expression for ψ in (B.2) implies that $\psi(t) \geq \frac{L^2}{I}$. Furthermore, $\psi(t) = \frac{L^2}{I}$ if and only if $\sum_{i=1}^{I} \theta_{i,0}^2 = \psi(0) = \frac{L^2}{I}$ which — by Cauchy-Schwarz's inequality — is equivalent to $\theta_{i,0} = \frac{L}{I}$ for all $1 \leq i \leq I$.

By using the expressions in (3.18) and (D.4), we obtain

$$r^{\text{Pareto}} - r^{\text{Radner}} = \frac{a^2\sigma_Y^2(1-\rho^2)}{2}\left(1 - \frac{1}{I}\right) \geq 0. \tag{B.11}$$

(ii): The equilibrium stock-price processes in (3.12), (C.2), and (D.5) have quadratic variation processes

$$\begin{aligned} d\langle \hat{S}\rangle_t &= F(t)^2\sigma_D^2 dt,\\ d\langle S^{\text{Radner}}\rangle_t &= F^{\text{Radner}}(t)^2\sigma_D^2 dt,\\ d\langle S^{\text{Pareto}}\rangle_t &= F^{\text{Pareto}}(t)^2\sigma_D^2 dt, \end{aligned} \tag{B.12}$$

where the Nash annuity is given in (3.2) and the Radner and Pareto annuities are defined as

$$F^{\text{Radner}}(t) := \frac{(r^{\text{Radner}} - 1)e^{r^{\text{Radner}}(t-T)} + 1}{r^{\text{Radner}}}, \tag{B.13}$$

$$F^{\text{Pareto}}(t) := \frac{(r^{\text{Pareto}} - 1)e^{r^{\text{Pareto}}(t-T)} + 1}{r^{\text{Pareto}}}. \tag{B.14}$$

The Radner and Pareto annuities have derivatives (4.7)-(4.8). Therefore, inserting (3.11) into the Nash annuity (3.2) and using the interest rate ordering (4.3) produce the ordering (4.6).

(iii): To see that (4.5) holds, we write the equity premium (4.1) as

$$\begin{aligned}
\mathrm{EP}(t) &= \frac{1}{\hat{S}_0}\Big(\mathbb{E}[\hat{S}_t] + \int_0^t \mathbb{E}[D_u] e^{\int_u^t r(s)ds} du\Big) - e^{\int_0^t r(u)du} \\
&= \frac{F(t)\Big(Q_2(t) + \frac{L}{I}Q_{22}(t)\Big) + F(t)(D_0 + \mu_D t) + S_t^{(0)} \int_0^t \frac{(D_0+\mu_D u)}{S_u^{(0)}} du}{\hat{S}_0} - e^{\int_0^t r(u)du},
\end{aligned} \tag{B.15}$$

where $S_t^{(0)} := e^{\int_0^t r(u)du}$. L'Hopital's rule and the ODEs (3.1)-(3.4) produce the limit

$$\lim_{t\downarrow 0} \frac{\mathrm{EP}(t)}{t} = \frac{a\sigma_D(L\sigma_D + I\rho\sigma_Y)F(0)}{I\hat{S}_0}. \tag{B.16}$$

To compute the numerator in (4.2), we define the function

$$G(t) := \int_0^t \frac{1}{S_u^{(0)}} du = O(t) \text{ as } t \downarrow 0. \tag{B.17}$$

Then, we apply Itô's Lemma to compute the dynamics $d\big(B_t G(t)\big)$ which produces the representation

$$F(t)B_t + S_t^{(0)} \int_0^t \frac{B_u}{S_u^{(0)}} du = \int_0^t \Big(F(t) + S_t^{(0)}\big(G(t) - G(u)\big)\Big) dB_u. \tag{B.18}$$

Therefore, Itô's Isometry gives us the variance

$$\begin{aligned}
&\mathbb{V}\Big[F(t)B_t + S_t^{(0)} \int_0^t \frac{B_u}{S_u^{(0)}} du\Big] \\
&= \int_0^t \Big(F(t) + S_t^{(0)}\big(G(t) - G(u)\big)\Big)^2 du \\
&= \int_0^t \Big(F(t)^2 + 2F(t)S_t^{(0)}\big(G(t) - G(u)\big) + (S_t^{(0)})^2\big(G(t)^2 - 2G(t)G(u) + G(u)^2\big)\Big) du \\
&= F(t)^2 t + 2F(t)S_t^{(0)}\Big(G(t)t - \int_0^t G(u)du\Big) + (S_t^{(0)})^2\Big(G(t)^2 t - 2G(t)\int_0^t G(u)du + \int_0^t G(u)^2 du\Big) \\
&= F(t)^2 t + 2F(t)S_t^{(0)}\big(O(t)t - O(t^2)\big) + (S_t^{(0)})^2\big(O(t^3) - 2O(t)O(t^2) + O(t^3)\big) \\
&= F(t)^2 t + O(t^2).
\end{aligned} \tag{B.19}$$

Based on (B.16) and (B.19), the limit in (4.5) equals

$$\begin{aligned}
\lim_{t \downarrow 0} \frac{\mathrm{SR}(t)}{\sqrt{t}} &= \lim_{t \downarrow 0} \frac{\hat{S}_0 \mathrm{EP}(t)}{\sigma_D \sqrt{t \mathbb{V}\Big[F(t) B_t + S_t^{(0)} \int_0^t \frac{B_u}{S_u^{(0)}} du\Big]}} \\
&= \lim_{t \downarrow 0} \frac{\hat{S}_0 \mathrm{EP}(t)}{\sigma_D t \sqrt{F(t)^2 + O(t)}} \\
&= \lambda,
\end{aligned} \tag{B.20}$$

where λ is the market price of risk defined in (3.16).

The Radner and Pareto instantaneous Sharpe ratios can be computed similarly.

$\diamondsuit$

C Competitive Radner equilibrium

Theorem 2 in Christensen, Larsen, and Munk (2012) shows that there exists a competitive Radner equilibrium in which the equilibrium interest rate is given by

$$r^{\mathrm{Radner}} = \delta + \frac{a}{I}(L\mu_D + I\mu_Y) - \frac{1}{2}\frac{a^2}{I^2}\big(I^2\sigma_Y^2 + 2IL\rho\sigma_D\sigma_Y + L^2\sigma_D^2\big), \tag{C.1}$$

and the equilibrium stock-price process is given by

$$\begin{aligned}
S_t^{\mathrm{Radner}} = &\frac{(r^{\mathrm{Radner}} - 1)e^{r^{\mathrm{Radner}}(t-T)} + 1}{r^{\mathrm{Radner}}} D_t \\
&- \frac{\Big(e^{r^{\mathrm{Radner}}(t-T)}((r^{\mathrm{Radner}}-1)r^{\mathrm{Radner}}(t-T)+1)-1\Big)\Big(\mu_D - \frac{a\sigma_D}{I}(I\rho\sigma_Y + L\sigma_D)\Big)}{(r^{\mathrm{Radner}})^2}.
\end{aligned} \tag{C.2}$$

Itô's lemma and (C.2) produce the competitive Radner equilibrium stock-price volatility coefficient of S_t^{Radner} to be

$$\frac{(r^{\mathrm{Radner}} - 1)e^{r^{\mathrm{Radner}}(t-T)} + 1}{r^{\mathrm{Radner}}}\sigma_D. \tag{C.3}$$

Equivalently, we can write (C.3) as $F^{\mathrm{Radner}}(t)\sigma_D$ where the Radner annuity $F^{\mathrm{Radner}}(t)$ is given by (4.7).

D Pareto efficient equilibrium

The following analysis uses the C-CAPM analysis from Breeden (1979). The utilities (2.10) produce the representative agent's utility function as

$$-e^{-\frac{a}{I}c-\delta t}, \quad c \in \mathbb{R}, \quad t \in [0,T]. \tag{D.1}$$

Because the economy's aggregate consumption is $LD_t + \sum_{i=1}^{I} Y_{i,t}$, the Pareto efficient equilibrium model's unique state-price density $\xi^{\text{Pareto}} = (\xi_t^{\text{Pareto}})_{t\in[0,T]}$ is proportional to the process

$$e^{-\frac{a}{I}(LD_t+\sum_{i=1}^{I} Y_{i,t})-\delta t}, \quad t \in [0,T]. \tag{D.2}$$

Itô's lemma produces the relative state-price dynamics to be:

$$\begin{aligned}
&\frac{d\xi_t^{\text{Pareto}}}{\xi_t^{\text{Pareto}}} \\
&= -\delta dt - \frac{a}{I}\Big(LdD_t + \sum_{i=1}^{I} dY_{i,t}\Big) + \frac{1}{2}\frac{a^2}{I^2} d\langle LD + \sum_{i=1}^{I} Y_i\rangle_t \\
&= -\delta dt - \frac{a}{I}\Big((L\mu_D + I\mu_Y)dt + (L\sigma_D + I\sigma_Y\rho)dB_t + \sigma_Y\sqrt{1-\rho^2}\sum_{i=1}^{I} dW_{i,t}\Big) \\
&+ \frac{1}{2}\frac{a^2}{I^2}\Big((L\sigma_D + I\sigma_Y\rho)^2 + I\sigma_Y^2(1-\rho^2)\Big)dt.
\end{aligned} \tag{D.3}$$

From (D.3), the Pareto efficient equilibrium's interest rate (i.e., the dt term in $-\frac{d\xi_t^{\text{Pareto}}}{\xi_t^{\text{Pareto}}}$) and the market price of risk related to the Brownian motion B_t (i.e., the dB_t volatility term in $-\frac{d\xi_t^{\text{Pareto}}}{\xi_t^{\text{Pareto}}}$) are

$$\begin{aligned}
r^{\text{Pareto}} &= \delta + \frac{a}{I}(L\mu_D + I\mu_Y) - \frac{1}{2}\frac{a^2}{I^2}\big((L\sigma_D + I\sigma_Y\rho)^2 + I\sigma_Y^2(1-\rho^2)\big), \\
\lambda &= \frac{a}{I}(L\sigma_D + I\sigma_Y\rho).
\end{aligned} \tag{D.4}$$

In turn, (D.4) produces the stock-price process in the Pareto efficient equilibrium to be

$$
\begin{aligned}
S_t^{\text{Pareto}} &= \frac{1}{\xi_t^{\text{Pareto}}}\mathbb{E}_t\Big[\int_t^T D_u \xi_u^{\text{Pareto}} du + D_T \xi_T^{\text{Pareto}}\Big] \\
&= -\frac{\left(e^{r^{\text{Pareto}}(t-T)}((r^{\text{Pareto}}-1)r^{\text{Pareto}}(t-T)+1)-1\right)\left(\mu_D - \frac{a\sigma_D}{I}(I\rho\sigma_Y + L\sigma_D)\right)}{(r^{\text{Pareto}})^2} \\
&+ \frac{(r^{\text{Pareto}}-1)e^{r^{\text{Pareto}}(t-T)}+1}{r^{\text{Pareto}}} D_t.
\end{aligned}
\tag{D.5}
$$

Itô's lemma and (D.5) produce the Pareto efficient equilibrium stock-price volatility coefficient of S_t^{Pareto} to be

$$
\frac{(r^{\text{Pareto}}-1)e^{r^{\text{Pareto}}(t-T)}+1}{r^{\text{Pareto}}}\sigma_D. \tag{D.6}
$$

Equivalently, we can write (D.6) as $F^{\text{Pareto}}(t)\sigma_D$ where the annuity $F^{\text{Pareto}}(t)$ is given by (4.8).

E Transitory price-impact calibration

The challenge in calibrating the transitory price-impact parameter α in (3.10) is that, α in our model is a measure of the perceived price-impact of fundamental trading imbalances for the aggregate stock market due to frictions in accessing asset-holding capacity from other natural end-counterparties (e.g., large pensions and mutual funds) and not transactional bid-ask bounce and market-maker inventory effects. In contrast, most empirical research measures transitory price effects for individual orders for individual stocks (e.g., as in Hasbrouck (1991) and Hendershott and Menkveld (2014)). The two concepts are related but there are some differences: First, α represents the transitory price effects of sustained trading programs associated with underlying parent orders rather than with isolated child orders (see, e.g., O'Hara (2015)) and one-off single orders. Second, sustained trading occurs in practice both via liquidity-making limit orders as well as via liquidity-taking market orders. From a transactional perspective, market and limit orders have opposite prices of liquidity since one is paying for liquidity and the other is being compensated for providing liquidity.

However, limit buying and market buying both create fundamental asset-holding pressure on the available ultimate (i.e., non-market-maker) asset sellers. It is the latter that α measures in our model. Third, stock in our model represents the aggregate stock market as an asset class and, thus, differs from individual stocks both in terms of its scale and as being a source of systematic risk rather than also including idiosyncratic stock-specific randomness. As a result, it seems natural, for example, to measure aggregate trading imbalances relative to market capitalization (as a measure of fundamental distortions in aggregate asset supply and demand) rather than in terms of shares (as in a transactional market-maker inventory model).

Our calibration involves adjusting empirical estimates of transitory price-impact for individual stocks into an estimate of the transitory price-impact of trading demand imbalances for the aggregate market. We proceed as follows: First, rather than using price-impact measures for individual trades (e.g., as in Hasbrouck (1991)) or market-maker inventory changes (e.g., as in Hendershott and Menkveld (2014)), we use estimates of the daily transitory price-impact of parent orders in Almgren, Thum, Hauptmann, and Li (2005). One advantage of the Almgren et al. (2005) estimation for our purposes is that it measures transitory price-impacts at the parent order level rather than at the child order level. Another advantage is that there is a natural way to rescale estimated transitory price-impact for individual stocks into a price-impact for the aggregate market. In particular, the Almgren et al. (2005) estimation is an industry-standard approach in which daily price-impact is estimated given panel data for a sample of parent orders over time for a cross-section of actively traded stocks. In doing so, the transitory price-impact (TPI) is scaled relative to a stock's individual price and daily return volatility and by scaling the underlying parent order size $\Delta\theta$ as a percentage relative to a stock's average daily trading volume (ADV):

$$\frac{\text{TPI}}{\text{stock price} \times \text{daily stock return volatility}} = \eta \times \Big(\frac{\Delta\theta}{\text{ADV}} \times 100\Big)^{\beta}. \tag{E.1}$$

The coefficient η is estimated in Almgren et al. (2005) to be 0.141, and the exponent

β is estimated to be 0.6 (i.e., slightly larger than the standard square-root model). One final advantage is that these estimates are average effects for all stocks rather than being driven by stock-specific differences in the trading environment for a particular stock (e.g., price level, bid-ask spread, institutional vs. retail ownership, market-maker inventory risk due to idiosyncratic stock returns). This gives a "dimensionless" standardized measure of transitory price-impact that can then be rescaled for the aggregate market.

Hence, a preliminary estimate of α in our model is:

$$\begin{aligned} \text{TPI} &\approx \text{market value} \times \text{daily market return volatility} \times \eta \times \frac{Q_3^\beta - Q_1^\beta}{Q_3 - Q_1} \times \frac{\text{SO}}{\text{ADV}} \times \frac{100}{L} \times \frac{\theta_i'}{265} \\ &\approx 3.5 \times 0.2\sqrt{\frac{1}{265}} \times 0.141 \times \frac{1.36^{0.6} - 0.38^{0.6}}{1.36 - 0.38} \times 121.36 \times \frac{100}{100} \times \frac{\theta_i'}{265} \\ &\approx 0.0018 \times \theta_i'. \end{aligned} \tag{E.2}$$

The following steps were used to derive (E.2): First, the market value (\$3.50) is set so that the calibrated absolute (dollar) price-impact is roughly consistent with the stock prices our asset-pricing model produces. Second, the daily return volatility is set to a ballpark 20% annual return volatility for the aggregate stock market deannualized for one trading day. Third, the power function in (E.1) is linearized using its slope between the empirical interquartile values Q_1 and Q_3 reported in Almgren et al (2005) for the percentage parent-size/ADV ratio. Fourth, the ratio $\frac{\Delta\theta}{\text{ADV}}$ is factored for our model as $\frac{\text{SO}}{\text{ADV}}\frac{\Delta\theta}{\text{SO}} = 121.36\frac{\Delta\theta}{L}$ where 121.36 is the empirical average ratio of shares outstanding to ADV for the NYSE and Nasdaq for 2009-2018,[13] and where shares outstanding $\text{SO} = L = 100$ in our model. This rescaling measures parent order size relative to shares outstanding, which, as discussed above, is a natural measure of trade size in our asset-pricing model. Fifth, the parameter α in (3.10) in our model measures the transitory price-impact relative to the trading rate $\theta_{i,t}'$ (i.e., where $\theta_{i,t}'dt$ is the instantaneous child order flow). Thus, we write the daily parent order $\Delta\theta$ as

$$\Delta\theta = \int_0^{\frac{1}{265}} \theta_i' dt = \theta_i' \frac{1}{265} \tag{E.3}$$

in terms of a constant child flow rate θ_i' over a trading day (i.e., $\frac{1}{265}$ of a year).

[13]From the *2019 SIFMQ Capital Market Fact Book.*

References

[1] Acharya, V. V., and L. H. Pedersen (2005): *Asset pricing with liquidity risk*, Journal of Financial Economics **77**, 375–410.

[2] Almgren, R. (2003): *Optimal execution with nonlinear impact functions and trading-enhanced risk*, Applied Mathematical Finance **10**, 1–18.

[3] Almgren, R., C. Thum, E. Hauptmann, and E. Li (2005): *Direct estimation of equity market impact*, Risk **18**, 58–62.

[4] Amihud, Y., and H. Mendelson (1986): *Asset pricing and the bid-ask spread*, Journal of Financial Economics **17**, 223–249.

[5] Amihud, Y., H. Mendelson, and L. H. Pedersen (2006): *Liquidity and asset prices*, Foundations and Trends in Finance **1**, 269–364.

[6] Back, K. (1992): *Insider trading in continuous time*, Review of Financial Studies **5**, 387–409.

[7] Back, K., H. Cao, and G. Willard (2000): *Imperfect competition among informed traders*, Journal of Finance **55**, 2117–2155.

[8] Baker, M., and J. Wurgler (2006): *Investor sentiment and the cross-section of stock returns*, Journal of Finance **61**, 1645–1680.

[9] Bansal, R., and A. Yaron (2004): *Risks for the long run: A potential resolution of asset pricing puzzles*, Journal of Finance **59**, 1481–1509.

[10] Barro, R. J. (2006): *Rare disasters and asset markets in the twentieth century*, Quarter Journal of Economics **121**, 823–866.

[11] Basak, S. (1996): *Dynamic consumption-portfolio choice and asset pricing with non-price-taking agents*, working paper.

[12] Basak, S. (1997): *Consumption choice and asset pricing with a non-price-taking agent*, Economic Theory **10**, 437–462.

[13] Basak, S., and D. Cuoco (1998): *An equilibrium model with restricted stock market participation*, Review of Financial Studies **11**, 309-341.

[14] Bouchard, B., M. Fukasawa, M. Herdegen, and J. Muhle-Karbe (2018): *Equilibrium returns with transaction costs*, Finance and Stochastics **22**, 569–601.

[15] Breeden, D. (1979): *An inter temporal asset pricing model with stochastic consumption and investment opportunities*, Journal of Financial Economics **7**, 265–296.

[16] Brunnermeier, M. K., and L. H. Pedersen (2005): *Predatory trading*, Journal of Finance **60**, 1825–1863.

[17] Calvet, L. E. (2001): *Incomplete markets and volatility*, Journal of Economic Theory **98**, 295–338.

[18] Campbell, J. Y., and J. H. Cochrane (1999): *By force of habit: A consumption-based explanation of aggregate stock market behavior*, Journal of Political Economy **107**, 205–251.

[19] Campbell, J. Y., S. J. Grossman, and J. Wang (1993): *Trading volume and serial correlation in stock returns*, Quarterly Journal of Economics **108**, 905–939.

[20] Chien, Y., H. Cole, and H. Lustig (2011): *A multiplier approach to understanding the macro implications of household finance*, Review of Economic Studies **78**, 199–234.

[21] Choi, J. H., and K. Larsen (2015): *Taylor approximation of incomplete Radner equilibrium models*, Finance and Stochastics **19**, 653–679.

[22] Choi, J. H., K. Larsen, and D. Seppi (2020): *Equilibrium effects of intraday order-splitting benchmarks*, submitted, `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3146658`.

[23] Christensen, P. O., K. Larsen, and C. Munk (2012): *Equilibrium in securities markets with heterogeneous investors and unspanned income risk*, Journal of Economic Theory **147**, 1035–1063.

[24] Christensen, P. O., and K. Larsen (2014): *Incomplete continuous-time securities markets with stochastic income volatility*, Review of Asset Pricing Studies **4**, 247–285.

[25] Cochrane, J. H. (2005): *Asset pricing*. Princeton University Press.

[26] Cochrane, J. H. (2008): *Financial markets and the real economy*, Handbook of the Equity Risk Premium, 237–325.

[27] Constantinides, G. M., (1986): *Capital market equilibrium with transaction costs*, Journal of Political Economy **94**, 842–862.

[28] Constantinides, G. M. (1990): *Habit formation: A resolution of the equity premium puzzle*, Journal of Political Economy, **98**, 519–543.

[29] Constantinides, G. M. (2008): *Understanding the equity risk premium puzzle*, Handbook of the Equity Risk Premium, 331–376.

[30] Constantinides, G. M., and A. Gosh (2017): *Asset pricing with countercyclical household consumption risk*, Journal of Finance, **72**, 415–460.

[31] Constantinides, G. M., and D. Duffie (1996): *Asset pricing with heterogeneous consumers*, Journal of Political Economy **104**, 219–240.

[32] Coval, J., and E. Stafford (2007): *Asset fire sales (and purchases) in equity markets*, Journal of Financial Economics **86**, 479–512.

[33] Cvitanić, J., and D. Cuoco (1998): *Optimal consumption choices for a large investor*, Journal of Economic Dynamics and Control **22**, 401-436.

[34] Duffie, D., Gârleanu, N., and L. H. Pedersen (2005): *Over the counter markets*, Econometrica **73**, 1815–1847.

[35] Easley, D., and M. O'Hara (2003): *Microstructure and asset pricing*, Handbook of the Economics of Finance **1B**, 1021–1051.

[36] Feldman, M., and C. Gilles (1985): *An expository note on individual risk without aggregate uncertainty*, Journal of Economic Theory **35**, 26–32.

[37] Gârleanu, N., and L. H. Pedersen (2016): *Dynamic portfolio choice with frictions*, Journal of Economic Theory **165**, 487–516.

[38] Grossman, S. J., and M. H. Miller (1988): *Liquidity and market structure*, Journal of Finance **43**, 617–633.

[39] Grossman, S. J., and J. E. Stiglitz (1980): *On the impossibility of informationally efficient markets*, American Economic Review **70**, 393-408.

[40] Hartman, P. (2002): *Ordinary differential equations*, 2nd Ed., SIAM Classics in Applied Mathematics.

[41] Hasbrouck, J. (1991): *Measuring the information content of stock trades*, Journal of Finance **46**, 179–207.

[42] Heaton, J., and D. Lucas (1992): *The effects of incomplete insurance markets and trading costs in a consumption-based asset pricing model*, Journal of Economic Dynamics and Control **16**, 601–620.

[43] Heaton, J., and D. Lucas (1996): *Evaluating the effects of incomplete markets on risk sharing and asset pricing*, Journal of Political Economy **104**, 443–487.

[44] Hellwig, M. F. (1980): *On the aggregation of information in competitive markets*, Journal of Economic Theory **22**, 477–498.

[45] Hendershott, T., and A. J. Menkveld (2014): *Price pressures*, Journal of Financial Economics **114**, 405–423.

[46] Hugonnier, J. (2012): *Rational asset pricing bubbles and portfolio constraints*, Journal of Economic Theory **147**, 2260–2302.

[47] Judd, K. L. (1985): *The law of large numbers with a continuum of iid random variables*, Journal of Economic Theory **35**, 19–25.

[48] Kardaras, C., H. Xing, and G. Žitković (2015): *Incomplete stochastic equilibria with exponential utilities close to Pareto optimality*, working paper.

[49] Krueger, D., and H. Lustig (2010): *When is market incompleteness irrelevant for the price of aggregate risk (and when is it not)?*, Journal of Economic Theory **145**, 1–41.

[50] Kyle, A. (1985): *Continuous auctions and insider trading*, Econometrica **53**, 1315-1336.

[51] Larsen, K., and T. Sae-Sue (2016): *Radner equilibrium in incomplete Lévy models*, Mathematics and Financial Economics **10**, 321–337.

[52] LeRoy, S. F., and R. D. Porter (1981): *The present value relation: Tests based on implied variance bounds*, Econometrica **49**, 555–574.

[53] Noh, E., and K. Weston (2020): *Price impact equilibrium with transaction costs and TWAP trading*, working paper.

[54] Mehra, R., and E. C. Prescott (1985): *The equity premium: A puzzle*, Journal of Monetary Economics **15**, 145–162.

[55] O'Hara, M. (2015): *High frequency market microstructure*, Journal of Financial Economics, **116**, 257–270.

[56] Ohasi, K. (1991): *A note on the terminal date security prices in a continuous time trading model with dividends*, Journal of Mathematical Economics **20**, 219–224.

[57] Ohasi, K. (1992): *Corrigendum: A note on the terminal date prices in a continuous time trading model with dividends*, Journal of Mathematical Economics **21**, 199.

[58] Pritsker, M. (2009): *Large investors: Implications for equilibrium asset returns, shock absorption, and liquidity*, working paper.

[59] Sannikov, Y., and A. Skrzypacz (2016): *Dynamic trading: Price inertia and front-running*, working paper, Stanford University Graduate School of Business.

[60] Schied, A., and T. Schöneborn (2009): *Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets*, Finance Stochastics **13**, 181–204.

[61] Shiller, R. J. (1981): *Do stock prices move too much to be justified by subsequent changes in dividends?*, American Economic Review **71**, 421–436.

[62] Stoll, H. R. (1978): *The pricing of security dealer services: An empirical study of NASDAQ stocks*, Journal of Finance **33**, 1153–1172.

[63] Storesletten, K., C. Telmer, and A. Yaron (2007): *Asset pricing with idiosyncratic risk and overlapping generations*, Review of Economic Dynamics **10**, 519–548.

[64] Storesletten, K., C. Telmer, and A. Yaron (2008): *Asset prices and intergenerational risk sharing: The role of idiosyncratic earnings shocks.* Handbook of the Equity Risk Premium, 565–590.

[65] Townsend, R. M. (1983): *Forecasting the forecasts of others*, Journal of Political Economy **91**, 546–588.

[66] Uhlig, H. (1996): *A law of large numbers for large economies*, Economic Theory **8**, 41–50.

[67] Vayanos, D. (1999): *Strategic trading and welfare in a dynamic market*, Review of Economic Studies **66**, 219–254.

[68] Vayanos, D. (2001): *Strategic trading in a dynamic noisy market*, Journal of Finance **56**, 131–171.

[69] Vayanos, D., and J.-L. Vila (1999): *Equilibrium interest rate and liquidity premium with transaction costs*, Economic Theory **13**, 509–539.

[70] Wachter, J. A. (2013): *Can time-varying risk of rare disasters explain aggregate stock market volatility?*, Journal of Finance **68**, 987–1035.

[71] Wang, N. (2003): *Caballero meets Bewley: The permanent-income hypothesis in general equilibrium*, American Economic Review **93**, 927–936.

[72] Weil, P. (1989): *The equity premium puzzle and the risk-free rate puzzle*, Journal of Monetary Economics **24**, 401–421.

[73] Weston, K. (2018): *Existence of a Radner equilibrium in a model with transaction costs*, Mathematics and Financial Economics **12**, 517–539.

[74] Weston, K., and G. Žitković (2020): *An incomplete equilibrium with a stochastic annuity*, Finance and Stochastics **24**, 359–382.

[75] Žitković, G. (2012): *An example of a stochastic equilibrium with incomplete markets*, Finance and Stochastics **16**, 177–206.