

Optimal Transport and Informed Trading in Securities Markets

13th October 2026

Abstract

We establish connections between optimal transport theory and the dynamic version of the ? model of informed trading. We use the connections to extend the model to risk-averse market makers and multiple assets with general distributions. Risk aversion generates excess volatility, excess co-movement, short-term reversals, a variance risk premium, and a non-monotone pricing kernel. We identify the separate effects of adverse selection and inventory risk on market illiquidity. We apply the model to show that option-implied volatility predicts stock returns when there is informed trading in stocks and options and market makers are risk averse.

Keywords: Optimal transport, option pricing, insider trading, Kyle model, asymmetric information, market microstructure, filtering theory.

1 Introduction

The ? model has been instrumental for understanding the role of asymmetric information and liquidity in financial markets. The dynamic version of the model reflects the reality that, in most markets, large investors split their orders into small pieces and trade dynamically to minimize price impacts. In this paper, we make substantial extensions to the dynamic Kyle model by establishing a connection with optimal transport theory. The existence of an equilibrium in the model is ensured by the existence of an optimal transport map, which “transports” trades to prices. Using this result, we characterize informed trading equilibria with multiple non-normal assets (including derivative securities) and when market makers are averse to inventory risk. The use of optimal transport theory and Monge-Kantorovich duality produces characterizations of the gains from informed trading that are new even for the single-asset/risk-neutral model.

We model risk averse market makers (also called dealers) in the same way as, for example, ?, ?, and ?. One interpretation is that market makers have access to an inter-dealer market that is competitive in the Walrasian sense. Implicitly, this assumes that there are many market makers. We discuss this further in Sections ?? and ??.

By including risk-averse market makers in the Kyle model, we merge the two main theories of market liquidity: adverse selection and dealer aversion to inventory risk. We can separate the contribution each makes to market illiquidity. With either risk-neutral or risk-averse market makers, price changes are driven by orders, and the stochastic matrix (Kyle’s λ) that relates orders to price changes is always symmetric and positive semidefinite. We show that, when market makers are more risk averse, market liquidity is lower (λ s are larger in the partial order of positive definiteness). This is consistent with models of the bid-ask spread based on dealer aversion to inventory risk (?). Furthermore, risk aversion produces mean reversion in market makers’ inventories, whereas they are a random walk in the risk-neutral model. The mean reversion in inventories correspond to expected price changes (risk premia) via the λ matrix.

Our model exhibits excess volatility and short-term reversals, as in other models of inventory risk (??). When the informed trader and noise traders have been selling in aggregate, prices fall due to both the informational component of orders and due to the rising risk premia that result from rising dealer inventories. The reverse is true when the informed trader and noise traders have been buying. Thus, price changes exceed the changes that are induced by information alone. The ‘excess’ price changes are reversed on average as risk diminishes over time and risk premia are realized. The precise expression of the excess volatility phenomenon that we derive is that the quadratic variation (cumulative minute-to-minute variation) of the price process exceeds the unconditional variance (long-run variation) of terminal prices. The quadratic variation is higher because it equals the risk-neutral variance. The risk-neutral variance is higher than the variance under the real probability measure of the economy, because the risk-neutral distribution puts high weights on ‘bad’ states. For market makers, ‘bad’ states are states in which the informed trader has extreme information, so the tails get more weight under the risk-neutral distribution.

As an application, we examine informed trading in an underlying asset and its corresponding

options. The early evidence of a trade-off for informed traders between the spot and its related option market goes back to [? and ?](#). Since then, many studies use quantities measured in the option market to predict returns and other quantities in the underlying market.¹ Surprisingly, as empirical evidence has accumulated, theoretical guidance has remained scarce. Our extension of the Kyle model to multiple non-normal assets allows inclusion of optionality in the payoffs, so we can explore theoretical relationships between the options and underlying markets. Our extension also permits more realistic underlying asset distributions, e.g., restricting prices to be non-negative and featuring skewness.

In our theoretical model, we show that option-implied volatility predicts the return of the underlying asset – high implied volatilities predict high returns. This is consistent with empirical evidence presented by [?](#). This type of predictability cannot arise with risk-neutral market makers, because expected returns always equal the risk-free rate when market makers are risk neutral. We also show that, regardless of whether uninformed trades in an option and underlying asset are positively or negatively correlated, risk-averse market makers usually end up with hedged positions – for example, long the underlying asset if they are short a call. This is again contrary to the risk-neutral model, and it is consistent with known empirical strategies of market makers.

The rest of the paper is organized as follows. Section [??](#) reviews the existing literature. Section [??](#) applies optimal transport theory to extend the Kyle model to multiple non-normal assets when market makers are risk neutral. Section [??](#) extends the model to risk-averse market makers; Section [??](#) studies option markets, and Section [??](#) concludes. Appendices contain proofs of the results stated in the main text, as well as a self-contained description of the numerical method used in Section [??](#).

2 Literature Review

Optimal transport theory is our main tool. Optimal transport theory has been applied in economics ([?](#)), but its applications to asymmetric information in financial markets have been limited. We are aware of only two existing papers that apply optimal transport theory to informed trading in financial markets, both of which are in single-period models. [?](#) study a version of the [?](#) model, which is a single-period variant of the Kyle model in which the informed trader can condition her order on realized noise trades. [?](#) contains as special cases both [?](#) and [?](#). They show that optimal policies for the single-period persuasion problem are necessarily derived from optimal transport maps. After the first draft of this manuscript was circulated, two papers closely related to our work were made available: [?](#) and [?](#). [?](#) finds an equilibrium in the continuous-time Kyle model in which the market maker is risk-averse and one normally distributed risky asset is traded. They recover our Theorem [??](#). They also compute an equilibrium with a different assumption regarding market-maker behavior and compare the two. Those assumptions are discussed further below. [?](#) studies a version of the Kyle model in which the informed trader faces quadratic temporary transaction

¹For recent studies as well as a thorough perspective on this literature see [?, ?, ?, ?, ? and ?](#).

costs in addition to adverse selection. In this context, an equilibrium can be constructed from an entropy-regularized optimal transport problem.

The continuous-time Kyle model with risk-neutral market makers has been applied and extended many times. A short list would include [?, ?, ?, ?, ?, ?, ?, ?, ?](#) and [?](#). Especially relevant to our work are the papers that study multiple assets. [?](#) analyzes multiple assets in what is essentially a Gaussian model. He assumes that the vector of asset values is homeomorphic to a Gaussian vector, so market makers can filter for the Gaussian vector and then compute the conditional distribution of the asset values via the homeomorphism. This is a special case of our model. The homeomorphism assumption excludes options on an underlying asset, which we study. [?](#) and [?](#) also study multi-asset versions of the continuous-time Kyle model. The first of these studies a normally distributed asset value and a call option on the asset and makes a special parametric assumption to establish the existence of an equilibrium. The second studies informed trading in a stock and bond but assumes that the informed trader’s signal has only two possible values. Our results nest both of those papers as special cases.

Risk aversion is crucial to generate time-varying equilibrium risk premia and hence predictability, and intermediary asset pricing theory ([??](#)) explains why it should be included at the level of market makers. We envision an inter-dealer market in which dealers are price takers and in which there is a representative investor (dealer). The representative dealer’s marginal utility evaluated at aggregate dealer wealth serves as a stochastic discount factor in the inter-dealer market. Competition between dealers ensures that the marginal-utility-based price at which they can trade in the inter-dealer market is also the price that they offer to non-dealer traders. Empirical investigations by, among others, [?, ?](#) and [?](#), provide supportive evidence relating intermediaries’ marginal utility (intermediary factors) to asset returns through intermediaries’ Euler equations. [?](#) document a direct empirical connection between intermediaries’ risk-bearing capacity and risk premia.

[?](#) and [?](#) introduce risk-averse market makers into the dynamic Kyle model. [Çetin and Danilova](#) assume that a representative market maker always fulfills the net demand of the informed and noise traders at a price such that the market maker is indifferent about trading. On the other hand, we assume that the market maker fulfills the demand at a price such that, after fulfilling it, the market maker is indifferent at the margin about making any further trades at that price. As stated above, our assumption can be motivated by the idea that market makers are price takers in an inter-dealer market and is the same assumption made by [?, ?, and ?](#) and in the intermediary asset pricing literature ([?](#)). A competitive inter-dealer market is modeled explicitly by [?](#) and [?](#). We remark that the terms ‘dealer’ and ‘market maker’ seem to be used interchangeably in this literature. We follow [?](#) and use the term ‘market maker.’ We discuss the assumptions further and give a small example in Section [??](#).

[?](#) takes a standard representative investor approach to pricing. His representative investor consumes the aggregate dividend of the economy. Apparently, gains and losses from trading with informed and noise traders are shared broadly across investors and have only a negligible effect on the market’s pricing kernel. Instead of evaluating the representative dealer’s marginal utility at the

economy's aggregate dividend as Ying does, we evaluate the representative dealer's marginal utility at aggregate dealer profits/wealth. Again, our approach is consistent with the intermediary asset pricing literature, in that it is dealer wealth, rather than the aggregate dividend, that determines asset prices.

The concepts we tackle have been analyzed to some extent in single-period models. ? combine inventory risk and adverse selection in a model of bid-ask spreads, treating inventory cost as the price of an option on an asset valued based on whether the dealer trades with an informed or uninformed trader. ? analyzes risk aversion on the part of the informed trader and market makers in a single-period Kyle model with a normally distributed asset. Gaussian risk-neutral multi-asset single-period Kyle models are studied by ?, ? and ?. Single-period models of informed trading in options include ? and ?. The first of these considers an underlying asset that can take only three possible values and assumes, as in the Kyle model, that competitive market makers set prices after seeing an order from either an informed or uninformed trader. The second considers an underlying asset that can take only two possible values and assumes competitive market makers quote bid and ask prices at which informed and uninformed traders can trade.

3 Risk-Neutral Market Makers

In this section, we describe the dynamic Kyle model with risk-neutral market makers and establish the connection to optimal transport theory. In case any reader might be unfamiliar with the model, we provide here a brief informal description. There is a single trader with private information. Other market participants are aware that some trades come from a better informed investor. They attempt to discern the investor's information from the flow of orders. However, there are other trades made for other reasons, and they provide camouflage for the informed trader. Orders are anonymous, and prices react to all orders, because they could potentially have come from the informed trader. The liquidity of the market depends on how much prices are moved by trades, and the informational efficiency of the market depends on the degree to which uninformed market participants can discern the information in orders.

3.1 Model

The formal model is as follows. There is a risk-free asset with risk-free rate normalized to zero. There are n risky assets. The assets are traded on the time horizon $[0, 1]$.² The values of the risky assets at date 1 is an exogenous random vector $\tilde{v}$. Each element of $\tilde{v}$ has a finite variance. At date 1, the vector $\tilde{v}$ is publicly revealed.

There is a single trader who observes $\tilde{v}$ at date 0 and is risk neutral. The trades that provide camouflage for the informed trader are called noise or liquidity trades. Rather than modeling the

²We do not specify the unit in which time is measured, so $[0, 1]$ could be 1 year, 3 months, or whatever the appropriate horizon is in any given application. In other words, there is no loss of generality in calling the end of the model "date 1."

motivations for the trades, it is simply assumed that the cumulative trades form a vector Brownian motion Z . Assume Z has zero drift and a constant instantaneous covariance matrix Σ_z .³ Let X_t denote the vector of share holdings of the informed trader in the risky assets at date t . While we could allow X to be a quite general stochastic process, we show in the appendix that it is without loss of generality to assume that X is absolutely continuous and satisfies $dX_t = \theta_t dt$ for some stochastic process θ . The differential dX_t is the market order submitted by the informed trader at date t , and θ_t is the rate of informed orders at date t . Set $Y = X + Z$. The differential dY_t is the net market order of the informed trader and noise traders at date t .

The rational market participants who trade with the informed and noise traders are called market makers (or dealers). In this section, we assume they are risk neutral. They observe the net order process Y and compete to fill the orders. Competition and risk neutrality force prices to equal expected values conditional on the information in orders, which we write as

$$P_t = \mathbb{E}[\tilde{v} \mid \mathcal{F}_t^Y]. \quad (3.1)$$

We look for an equilibrium in which Y_t is a sufficient statistic for P_t and denote the pricing rule by $P_t = H(t, Y_t)$, for $t \in [0, 1)$.

We assume the informed trader observes Z . There is little if any loss of generality in making this assumption rather than assuming the informed trader observes the price P .⁴ Thus, her information is $\tilde{v}$ and the history of Z . In addition to (??), the other equilibrium condition is that the informed trader's strategy is optimal, given the pricing rule H . The informed trader's realized profit is

$$\int_0^1 (\tilde{v} - P_t)' \theta_t dt. \quad (3.2)$$

The equilibrium condition is that the informed trader maximizes the expected value of (??), conditional on $\tilde{v}$ and recognizing that $P_t = H(t, Y_t)$ with $Y = X + Z$. Thus, the informed trader is aware that her trades move prices and takes this into account when deciding how much and when to trade. Hence, the model involves both filtering (by the market makers) and control (by the informed trader) and each depends on the other.

3.2 The Role of Optimal Transport Theory

Let F denote the distribution function of $\tilde{v}$, and let G denote the distribution function of Z_1 , which is normal $(0, \Sigma_z)$. It can be intuited from ? that a key to constructing an equilibrium is to find a map Φ from $\mathbb{R}^n$ to the support of F that transports G to F , meaning that $\Phi(Z_1)$ has distribution F . In

³We can extend our results to noise trades satisfying $dZ_t = \sigma(t) dW_t$ where W is a vector of independent Brownian motions and σ is a continuous bounded matrix-valued function of time that is nonsingular for all t and is such that σ^{-1} is also a bounded function of t . However, to economize on notation, we take σ to be constant. This produces the constant instantaneous covariance matrix $\Sigma_z = \sigma\sigma'$.

⁴In the examples below, this assumption is not necessary, as Kyle's Lambda in Theorem ?? is invertible and the informed trader can infer the increments of Y from the increments of P . Theorem ?? shows that Kyle's Lambda is the expectation of the Hessian of a convex function, which gives simple criteria for its invertibility.

the univariate case studied in ?, this is simply constructed as $F^{-1} \circ G$, assuming F is invertible. In the multivariate case, assuming F is absolutely continuous with respect to Lebesgue measure, the map that is needed is provided by Brenier's theorem in optimal transport theory (?, Proposition 3.1)-(?, Main Theorem).

Theorem 3.1 (Brenier's Theorem). *For any probability distribution F , there exists a convex function $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\nabla\Gamma(Z_1)$ has distribution function F . The function Γ is unique up to an additive constant. If F is absolutely continuous, then $\nabla\Gamma$ is invertible (on the support of F) and $\tilde{\zeta} := (\nabla\Gamma)^{-1}(\tilde{v})$ has distribution G .*

The function Γ in this result is called a Brenier potential. The gradient $\nabla\Gamma$ is called a transport map. Since Γ is finite and convex, the set $\mathcal{D}_\Gamma$ of points at which it is differentiable is a Borel set of full Lebesgue measure, $\nabla\Gamma$ is Borel measurable and continuous on $\mathcal{D}_\Gamma$ (?, Theorems 25.2 and 25.5), and we set $\nabla\Gamma := 0$ on $\mathbb{R}^n \setminus \mathcal{D}_\Gamma$. This convention does not impact our analysis. Indeed, $\nabla\Gamma$ is only integrated against Gaussian densities or evaluated at random variables with absolutely continuous distributions, which lie in $\mathcal{D}_\Gamma$ almost surely. $\nabla\Gamma$ transports the source distribution G to the target distribution F . The transport map $\nabla\Gamma$ is an optimizer for the Monge problem with quadratic objective (?, Theorem 6.5) i.e., maximizing the total expected liquidation value $Y_1'\tilde{v}$ among all couplings of Y_1 with distribution G and $\tilde{v}$ with distribution F . Although we rely only on the existence of the Brenier potential when G is Gaussian, this map exists under very mild conditions on F and G . For example, (?, Main Theorem) states the existence of the Brenier potential when G is absolutely continuous without any additional assumption on F .

The equilibrium of the Kyle model can be characterized in terms of $\tilde{\zeta}$ from Brenier's theorem, as we will show, when F is absolutely continuous. Specifically, the informed trader controls the net order process Y such that $Y_1 = \tilde{\zeta}$, so that the unconditional distribution function of Y_1 is G , the same as for Z_1 . When F is not absolutely continuous, then $\nabla\Gamma$ may not be invertible. In other words, the map $v \mapsto (\nabla\Gamma)^{-1}(v)$ will be a correspondence instead of a function. This will be the case, for example, when some of the securities are options written on others, because then the support of F is a lower dimensional subset of $\mathbb{R}^n$. We address this situation by allowing the informed trader to play a mixed strategy, randomly choosing a point in $(\nabla\Gamma)^{-1}(v)$ for each realization v of $\tilde{v}$. We model the randomization by allowing the informed trader to generate a random variable $\tilde{u} \in \mathbb{R}^n$ at date 0 with distribution G (in fact any absolutely continuous distribution would work here), which is independent of $\tilde{v}$ and Z , and to base her trades on $\tilde{u}$ as well as on $\tilde{v}$ and the history of Z . The following shows that, even when F is not absolutely continuous, there is a random variable $\tilde{\zeta}$ depending on $\tilde{u}$ and $\tilde{v}$ such that (i) $\tilde{\zeta} \in (\nabla\Gamma)^{-1}(\tilde{v})$ with probability one and (ii) $\tilde{\zeta}$ has distribution function G . This allows construction of a mixed strategy equilibrium such that it is again true that the informed trader controls Y to satisfy $Y_1 = \tilde{\zeta}$ as in the case of an absolutely continuous F .⁵ The proof of the following Lemma is in Section ??.

⁵We focus on the mixed strategy equilibrium for simplicity. It may also be possible to derive a pure strategy equilibrium using h -transforms. An example of such a construction is in ?. See ? for a general discussion of the h -transform method.

Lemma 3.1. *In the context of Brenier's theorem, even if F is not absolutely continuous, there exists an $\mathbb{R}^n$ -valued random variable $\tilde{\zeta}$ depending on $(\tilde{u}, \tilde{v})$ which has joint distribution $G \otimes F$ such that $\tilde{\zeta}$ has distribution G and $\nabla\Gamma(\tilde{\zeta}) = \tilde{v}$.*

The reason that the informed trader controls Y so that $Y_1 = (\nabla\Gamma)^{-1}(\tilde{v})$ in the case of absolutely continuous F and so that $Y_1 = \tilde{\zeta}$ in general is that, given the equilibrium pricing rule, her profit (??) turns out to equal

$$\tilde{v}'Y_1 - \Gamma(Y_1) - \int_0^1 (\tilde{v} - P_t)' dZ_t. \quad (3.3)$$

This is shown in the proof of Theorem ?? (for even more general trading strategies than we have defined here). The last term in (??) has a zero mean due to a mild admissibility restriction that we impose on X in Theorem ?. Therefore, the informed trader maximizes her expected profit by controlling Y to maximize $\tilde{v}'Y_1 - \Gamma(Y_1)$. Because of the convexity of Γ , a sufficient condition for optimality of an admissible continuous finite-variation strategy is that $Y_1 \in \mathcal{D}_\Gamma$ and $\nabla\Gamma(Y_1) = \tilde{v}$ almost surely. This gives us $Y_1 = (\nabla\Gamma)^{-1}(\tilde{v})$ or in general $Y_1 = \tilde{\zeta}$.

To preview Theorem ?? a bit more, note that, for $Y_1 \in \mathcal{D}_\Gamma$, the first-order condition $\nabla\Gamma(Y_1) = \tilde{v}$ is equivalent to $\tilde{v}'Y_1 - \Gamma(Y_1) = \Gamma^*(\tilde{v})$, where Γ^* is the convex conjugate (Legendre-Fenchel transform) of the convex function Γ , defined as $\Gamma^*(v) = \sup \{v'y - \Gamma(y) \mid y \in \mathbb{R}^n\}$. Thus, Brenier's theorem allows us to characterize expected informed trading profits in terms of the convex conjugate of the Brenier potential (Part (D) of Theorem ??).

3.3 Equilibrium

We can now prove the existence of equilibrium, largely by following the reasoning in ? and the subsequent developments summarized in ?. The mild restriction on trading strategies in the theorem below is described as ‘no doubling strategies’ in ?. Even though most of the proof of the following is similar to that in ?, the characterization of the equilibrium in terms of Γ , $\tilde{\zeta}$, and Γ^* is new even in the univariate case.

Theorem 3.2. *Recall that F is the distribution of $\tilde{v}$ which has finite variance. Let Γ denote the Brenier potential such that $\nabla\Gamma$ transports G to F and select the additive constant in Brenier's theorem so that $\mathbb{E}[\Gamma(Z_1)] = 0$.⁶ Let $k(t, y, \cdot)$ denote the normal $(y, (1-t)\Sigma_z)$ density function, and set*

$$K(t, y) = \int_{\mathbb{R}^n} \Gamma(z) k(t, y, z) dz. \quad (3.4)$$

Assume informed trading strategies are constrained to satisfy

$$\mathbb{E} \int_0^1 \|H(t, X_t + Z_t)\|^2 dt < \infty. \quad (3.5)$$

where $H := \nabla_y K$. There is an equilibrium with the following properties.

⁶We show in Lemma ?? that $\Gamma(Z_1)$ is integrable. Therefore, we can choose the additive constant to ensure $\mathbb{E}[\Gamma(Z_1)] = 0$.

A. The equilibrium price is $P_t = \nabla_y K(t, Y_t)$ for $t \in [0, 1)$ and satisfies

$$dP_t = \Lambda(t, Y_t) dY_t, \quad (3.6)$$

where $\Lambda = \nabla_y^2 K$.

B. The market is eventually informationally efficient in the sense that $\lim_{t \rightarrow 1} P_t = \tilde{v}$ with probability one.

C. The equilibrium informed trading strategy is

$$\theta_t := \frac{\tilde{\zeta} - Y_t}{1 - t}, \quad (3.7)$$

where $\tilde{\zeta}$ is defined in Brenier's theorem or Lemma ??.

D. The informed trader's value function is

$$\begin{aligned} J(t, v, y) &:= \sup_{\theta} \mathbb{E} \left[\int_t^1 (v - H(s, Y_s))' \theta_s ds \mid \tilde{v} = v, Y_t = y \right] \\ &= \Gamma^*(v) + K(t, y) - y'v, \end{aligned} \quad (3.8)$$

and the informed trader's expected profit at time 0 conditional on $\tilde{v} = v$ is $J(0, v, 0) = \Gamma^*(v)$.

E. The cumulative order process Y is a $(0, \Sigma_z)$ Brownian motion on $\mathcal{F}^Y$ and satisfies $Y_1 = \tilde{\zeta}$ with probability one.

In the univariate Gaussian model studied by ?, the price evolves as $dP = \lambda dY$ for a constant λ that is universally known as Kyle's lambda. Theorem ?? shows that price changes are also linearly related to orders in multivariate non-Gaussian models, via a possibly time-varying matrix Λ . Notice that Λ must be everywhere symmetric and positive semidefinite, because it is the Hessian of $y \mapsto K(t, y)$, which inherits the convexity of the Brenier potential Γ .

3.4 Informed Trading Profits

The connection to optimal transport theory provides additional characterizations of informed trading profits beyond those in Part (D) of the theorem. Monge-Kantorovich duality implies

$$\sup_{\Phi \in \Pi(F, G)} \int v' y \Phi(dv, dy) = \mathbb{E}[\Gamma^*(\tilde{v})] = \inf_{f, g} \mathbb{E}[f(\tilde{v})] + \mathbb{E}[g(Y_1)], \quad (3.9)$$

where $\Pi(F, G)$ is the set of probability measures Φ on $\mathbb{R}^n \times \mathbb{R}^n$ for which the marginal distributions are the exogenously given F and G , where the infimum is taken over all functions f and g with the property that $f(x) + g(y) \geq x'y$ for all (x, y) , and where the expectations are taken over the distributions F for $\tilde{v}$ and G for Y_1 . For this result, see Sections 2.3 and 6.2 of ?. The first equality

in (??) provides the interpretation that it is as if the informed trader earns $v'y$ and can choose any joint distribution for v and y with the given marginals. Such a joint distribution is called a coupling. A map $v \mapsto y$ in concert with the distribution of $\tilde{v}$ determines a coupling, provided the induced distribution of y is G . Thus, it is as if the informed trader chooses Y_1 as a function of $\tilde{v}$ to maximize $\mathbb{E}[\tilde{v}'Y_1]$ subject to the constraint that Y_1 has the same distribution as Z_1 . The distributional constraint on Y stems from the classical Lagrange multiplier interpretations of f, g in the dual problem (??). Indeed, (??), can also be written as

$$\sup_{\Phi} \inf_{f, g} \int (v'y - f(v) - g(y)) \Phi(dv, dy) \quad (3.10)$$

where the supremum is taken over the set of all probability measures Φ on $\mathbb{R}^n \times \mathbb{R}^n$ (without the distributional constraints on the marginals) and f, g satisfies

$$\int f(v)F(dv) = 0 \quad \text{and} \quad \int g(y)G(dy) = 0.$$

In (??), $g(Y_1) = \int_0^1 P'_t dY_t$ ensures that the distribution of y is G . Accordingly, the market maker observes Y and chooses the price to control the law of Y_1 to match G . Thus, the equilibrium forces the informed trader to optimize his expected liquidation value $\mathbb{E}[\tilde{v}'X_1] = \mathbb{E}[\tilde{v}'Y_1]$ while keeping Y_1 Gaussian.

The second equality in (??) provides another characterization of the informed trader's expected profit. The equality shows that the expected profit is what the informed trader would achieve if her profit were additively separable as $f(v) + g(y)$ with the given distributions F and G for v and y , and if market makers could choose f and g to minimize the expected profits, with the proviso that the profit $f(v) + g(y)$ could never be smaller than $v'y$. It turns out that the functions f and g that achieve this minimum are $\Gamma^*(v)$ and $\Gamma(y)$ (??, Proposition 6.4).

Furthermore, the expected profit of the informed trader is related to the Wasserstein-2 distance between F and G . Indeed, the unconditional expected profit (??) can also be written as

$$\frac{\|\tilde{v}\|^2 + \|Z_1\|^2 - W_2^2(F, G)}{2}, \quad (3.11)$$

where $\|\cdot\|$ denotes the $\mathcal{L}^2$ norm and where

$$W_2^2(F, G) := \inf_{\Phi \in \Pi(F, G)} \int |v - y|^2 \Phi(dv, dy)$$

is the square of the Wasserstein-2 distance between F and G . Thus, the expected profit depends on the amount of noise trading as measured by $\|Z_1\|^2$ and, given the distribution G of noise trading, depends on the amount of private information as measured by $\|\tilde{v}\|^2 - W_2^2(F, G)$.

In summary, we see that five important financial quantities in Kyle's model are obtained via important quantities from optimal transport theory. First, by (A), the optimal transport map gives

the pricing rule of the market maker. Second, by (C), the optimal transport map from F to G (or its generalization $\tilde{\zeta}$) determines the terminal target of the informed trader's strategy, whose drift is that of a Brownian bridge toward this target. Third, by (D), the expected profit of the informed trader is given by the convex conjugate of the Brenier potential. Fourth, by (A), Kyle's lambda matrix is given by the Hessian of the Brenier potential. Finally, by (??), the unconditional expected profit of the informed trader can be computed via the Wasserstein-2 distance between F and G .

3.5 Market Maker Profits

To set the stage for the next section, we compute the aggregate equilibrium market maker profits. Aggregate market maker profits equal the negative of the sum of the informed and noise traders' profits. The profits of noise traders are

$$\int_0^1 (\tilde{v} - P_t)' dZ_t - \langle P, Z \rangle_1, \quad (3.12)$$

where $\langle \cdot, \cdot \rangle$ denotes the sharp bracket process of the continuous vector semimartingales P and X (heuristically, $\langle P, Z \rangle_1 = \int_0^1 (dP_t)'(dZ_t)$ – see below for more). This formulation of the noise traders' profits follows from the intertemporal budget constraint of ? via integration by parts, as shown by ? in the univariate case. The analogous formula applies to the informed trader, but taking $dX = \theta dt$ implies $\langle P, X \rangle = 0$; hence, we arrive at (??). The formulas (??) and (??) imply that realized market maker profits are

$$\langle P, Y \rangle_1 - \int_0^1 (\tilde{v} - P_t)' dY_t. \quad (3.13)$$

Extending our earlier discussion of expected informed trader profits in terms of the Brenier potential Γ and its conjugate, we obtain the following.

Corollary 3.1. *Given the pricing rule $H = \nabla_y K$, market maker profits (??) equal*

$$\langle P, Y \rangle_1 + \Gamma(Y_1) - \tilde{v}' Y_1 \quad (3.14)$$

for any informed trading strategy $dX = \theta_t dt$. When the informed trader follows the optimal strategy (??), market maker profits are

$$\langle P, Y \rangle_1 - \Gamma^*(\tilde{v}). \quad (3.15)$$

3.6 Transaction Costs

The sharp bracket quantity in (??) and in Corollary ?? equals

$$\langle P, Y \rangle_1 = \text{tr} \left(\Sigma_z \int_0^1 \nabla_y^2 K(t, Y_t) dt \right), \quad (3.16)$$

where tr denotes the trace of a square matrix. This can be calculated as⁷

$$\begin{aligned}\langle P, Y \rangle_1 &= \sum_{i=1}^n \int_0^1 (dP_i)(dY_i) \\ &= \sum_{i=1}^n \int_0^1 \left(\sum_{j=1}^n \Lambda_{ij}(t, Y_t) dZ_j \right) (dZ_i) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} \int_0^1 \Lambda_{ij}(t, Y_t) dt.\end{aligned}$$

This implies (??) after substituting $\Lambda = \nabla_y^2 K$. The term $(dP_i)(dY_i)$ is the continuous time version of $(P_{t+1} - P_t)(Y_{t+1} - Y_t)$. In this expression, the factor $Y_{t+1} - Y_t$ is the net order and $P_{t+1} - P_t$ is the price change induced by the order. Thus, the product of differentials captures the higher cost paid on buy orders and lower revenue received on sell orders due to the price impacts of trades. The transaction cost is zero for the informed trader in continuous time, because her orders are very small (of order dt) but nonzero for noise traders (and for market makers) because noise trades are relatively large (of order $\sqrt{dt}$). ? show that the different transactions costs of informed and noise traders can alternatively be interpreted as the limit of a discrete-trade model in which orders of informed and noise traders are the same size but informed trades are optimally made at a much lower frequency than noise trades.

It is noteworthy that market maker profits in (??) depend on the path of noise trading only via the transaction costs. If noise traders make bad trades, then the informed trader will make more money by subsequently reversing them, but there is no net gain or loss to market makers. Noise traders are as likely to make good trades as bad trades, so on average this is also a wash for the informed trader, whose unconditional expected profit is simply $E[\Gamma^*(\tilde{v})]$, as shown in Theorem ??.

4 Risk Averse Market Makers

In the previous section, we assumed that competition among market makers drives prices to their expected values. In this section, market makers remain competitive, but we now allow them to be risk-averse. To clarify our equilibrium concept and to explain how it relates to the concept employed by ?, consider the following simple example. The example is chosen for simplicity and is not based on either our model or Çetin and Danilova's, each of which is set in continuous time with Itô demand processes. Naturally, our assumption of a Walrasian inter-dealer market as described in this example requires in practice that there be more than two market makers. Implicitly, we are assuming that there is a 'large' number of market makers.

Example. Suppose there are two market makers who each have utility for shares x and cash y equal to $u(x) + y$ for a continuously differentiable, increasing, concave function u . Suppose each market

⁷The notation $(dP_i)(dY_i)$ represents the differential of the joint variation process, also written as $d\langle P_i, Y_i \rangle$. Likewise, $(dZ_j)(dZ_i) := d\langle Z_j, Z_i \rangle$.

maker currently owns n shares. Suppose a market order to buy two shares arrives. In a model like ?, without an inter-dealer market, each market maker sells one share at a price of $p^\dagger := u(n) - u(n-1)$, because that is the price at which the utility gain for each is zero. In this paper, we assume the presence of an inter-dealer market. The inter-dealer market clears at a price such that neither trader wants to make further trades, given her current position. When the market order arrives, each market maker rationally anticipates that the equilibrium price in the inter-dealer market will be $p^* := u'(n-1)$. In both our model and the Çetin-Danilova model, dealers get zero utility gain from trading with the market orders, due to price competition between the dealers. In both cases, the utility gain should be measured relative to the next-best opportunity. In Çetin and Danilova, the next-best opportunity is to hold the initial position of n shares. Our model differs in that the next-best opportunity is to trade in the inter-dealer market. Because dealers assume they can sell all they want in the inter-dealer market at price p^* , they get zero utility gain from selling to the market orders at price p^* . Notice that the dealers do not have an incentive to undercut p^* , because if they wished to sell more then they assume they can do so at p^* in the inter-dealer market, which is better than selling at a discounted price to the market orders. However, they do have an incentive to undercut any price $p > p^*$, because they would profit from buying at p^* in the inter-dealer market and selling at $p - \epsilon > p^*$ to the market orders. This is the standard Bertrand result of competing price down to marginal cost, here the cost of buying in the inter-dealer market. In our model, each of the two market makers gets a positive utility gain ($p^* > p^\dagger$ due to decreasing marginal utility), but the utility gain comes from the ability to trade in the inter-dealer market, which is a normal consequence of competitive markets (see Theorem 1, known as The First Welfare Theorem, in ?).

Assume the market makers have constant absolute risk aversion (CARA) utility. It is well known that a group of competitive CARA agents can be aggregated into a single representative agent with CARA preferences. Asset prices are determined by the marginal utility of the representative market maker, evaluated at the aggregate wealth of all market makers.

More formally, the stochastic discount factor (SDF) in the economy is proportional to the representative market maker's marginal utility. Because we have normalized the risk-free rate to zero, the SDF is

$$\tilde{m} := \frac{e^{-\alpha \tilde{w}}}{E[e^{-\alpha \tilde{w}}]} \quad (4.1)$$

where α denotes the absolute risk aversion of the representative market maker, and $\tilde{w}$ denotes aggregate market maker wealth at date 1. The random variable $\tilde{w}$ depends on the informed trader's strategy and the pricing rule via the Merton intertemporal budget constraint. Specifically, market maker wealth is the sum of market maker gains from trade (??) plus the value of any initial inventories. Let β denote the aggregate position of market makers in the asset at date 0. Then, aggregate market maker wealth at date 1 is

$$\tilde{w} := \beta' \tilde{v} + \langle P, Y \rangle_1 - \int_0^1 (\tilde{v} - P_t)' dY_t. \quad (4.2)$$

Instead of pricing directly via the SDF, we can as usual equivalently price via the risk-neutral probability determined by the SDF. Denote the risk-neutral probability by $\mathbb{Q}$ and the physical (actual) probability of the economy by $\mathbb{P}$. The risk-neutral probability is given by the change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \tilde{m}. \quad (4.3)$$

We replace the equilibrium condition (??) with

$$P_t = \mathbb{E}^{\mathbb{Q}}[\tilde{v} \mid \mathcal{F}_t^Y]. \quad (4.4)$$

The model is otherwise unchanged. In particular, we continue to assume that the informed trader is risk-neutral (maximizing expected profits under $\mathbb{P}$). We continue to set $Y = X + Z$, where X is the number of shares purchased by the informed trader and Z is the number of shares purchased by noise traders. This implies that the number of shares held by market makers at any date t is $\beta - Y_t$. We assume that noise trades are a Brownian motion with zero drift and covariance matrix Σ_z under the physical distribution $\mathbb{P}$. As is well known, a change of measure such as (??) does not change the instantaneous covariance matrix of a Brownian motion, so Σ_z is also the instantaneous covariance matrix of Z under the risk-neutral probability $\mathbb{Q}$.

There are now three (information set, probability measure) pairs in play. For the informed trader, $\tilde{u}$, $\tilde{v}$, and Z are observable, and the relevant probability measure is the physical probability $\mathbb{P}$. The rest of the market only observes Y (until the announcement at date 1) so the relevant filtration is $\mathcal{F}^Y$. On this filtration, we have both the physical probability and the risk-neutral probability. We will do various calculations under the physical probability, but pricing is under the risk-neutral probability as described in (??).

For clarity, assume, without loss of generality, that the underlying probability space is a product space $\Omega := \mathbb{R}^n \times \mathbb{R}^n \times C[0, 1]^n$, with generic element (u, v, z) , where $u \in \mathbb{R}^n$ denotes the variable used for mixing, $v \in \mathbb{R}^n$ denotes the vector of asset values, and $z \in C[0, 1]^n$ denotes the path of cumulative noise trades. We continue to let G denote the normal $(0, \Sigma_z)$ distribution. Let ν denote the $(0, \Sigma_z)$ -Wiener measure on $C[0, 1]^n$ —i.e., the distribution of a Brownian motion with zero drift and instantaneous covariance matrix Σ_z . We take the physical probability measure on Ω to be the product of three measures: the distribution G for the mixing variable $\tilde{u}$, the distribution F of $\tilde{v}$, and the Wiener measure ν on $C[0, 1]^n$. The random vectors $\tilde{u}$ and $\tilde{v}$ and the vector Brownian motion Z are defined on this product space as the projection maps: $\tilde{u}(u, v, z) = u$, $\tilde{v}(u, v, z) = v$, and $Z(u, v, z) = z$. Thus, they are all independent under the physical probability. We will indicate the measures under which expectations are taken with superscripts; for example, $\mathbb{E}^\nu$ denotes expectation with respect to the Wiener measure ν . In the case of normal distributions (Sections ?? and ??), mixing plays no role, and we can equivalently take the probability space to be $\mathbb{R}^n \times C[0, 1]^n$, with generic element (v, z) .

4.1 A Single Normally Distributed Asset

We are generalizing the previous literature on the continuous-time Kyle model in two directions: by considering multiple possibly non-normally distributed assets and by allowing market makers to be risk-averse. The previous section describes the former extension in isolation. To clearly demonstrate the effects of the latter extension, we consider here a single normally distributed asset, which is the case studied by ? under risk neutrality. The implications of risk aversion are especially transparent in this case, so we describe it first.

In the next subsection, we address the n -asset model with normal distributions. Both the single-asset model and the n -asset model with normal distributions are special cases of the general model discussed in Section ?. Optimal transport theory does not directly appear in this subsection or the next, because we have substituted an analytic formula for the Brenier potential; however, the results are derived as corollaries of the general result in Section ?, which builds on Theorem ?.

Theorem 4.1. *Assume that $\tilde{v}$ is a scalar and is normally distributed with mean μ and variance σ_v^2 under the physical probability. Let σ_z denote the instantaneous standard deviation of the noise trades. Assume the informed trader is constrained to choose a strategy such that*

$$\mathbb{E}^{\mathbb{P}} \int_0^1 X_t^2 dt < \infty. \quad (4.5)$$

There is an equilibrium in which the risk-neutral distribution of $\tilde{v}$ is also normal. Furthermore, the following hold:

A. The unconditional risk-neutral mean of $\tilde{v}$ is

$$\mu^* := \mu - \alpha\beta\sigma_v^2.$$

B. The equilibrium price satisfies $P_0 = \mu^*$ and $dP_t = \lambda dY_t$, where

$$\lambda := \frac{\alpha\sigma_v^2}{2} + \sqrt{\frac{\alpha^2\sigma_v^4}{4} + \frac{\sigma_v^2}{\sigma_z^2}}. \quad (4.6)$$

C. The risk-neutral variance of $\tilde{v}$ is $\lambda^2\sigma_z^2 > \sigma_v^2$.

D. The market is eventually informationally efficient in the sense that $P_t \rightarrow \tilde{v}$ with probability one (under both $\mathbb{P}$ and $\mathbb{Q}$).

E. The equilibrium trading strategy is

$$\theta_t := \frac{1}{1-t} \left(\frac{\tilde{v} - \mu^*}{\lambda} - Y_t \right). \quad (4.7)$$

F. Under $\mathbb{P}$, the expected profit of the informed trader conditional on $\tilde{v}$ is

$$\frac{\lambda\sigma_z^2}{2} + \frac{(\tilde{v} - \mu^*)^2}{2\lambda}. \quad (4.8)$$

The unconditional expected profit of the informed trader is

$$\lambda\sigma_z^2 - \frac{\alpha\sigma_z^2\sigma_v^2}{2} + \frac{\alpha^2\beta^2\sigma_v^4}{2\lambda} \quad (4.9)$$

and the unconditional expected loss of noise traders is $\lambda\sigma_z^2$.

G. The SDF is

$$\tilde{m} = e^{\frac{\alpha(\tilde{v}-\mu^*)^2}{2\lambda} - \alpha\beta\tilde{v}} \bigg/ \mathbb{E} \left[e^{\frac{\alpha(\tilde{v}-\mu^*)^2}{2\lambda} - \alpha\beta\tilde{v}} \right]. \quad (4.10)$$

H. The cumulative order process Y is a $(0, \sigma_z^2)$ Brownian motion on $(\mathcal{F}^Y, \mathbb{Q})$ and satisfies

$$dY_t = \delta_t(\beta - Y_t) dt + \sigma_z dB_t, \quad (4.11)$$

where B is a standard $\mathcal{F}^Y$ Brownian motion under $\mathbb{P}$ and

$$\delta_t = \frac{\alpha\lambda\sigma_z^2}{1 + (1-t)\alpha\lambda\sigma_z^2}. \quad (4.12)$$

All aspects of the equilibrium are the same under the risk-neutral probability as in the risk-neutral model of ?. For example, Part (C) shows that Kyle's lambda equals the ratio of the risk-neutral standard deviation of $\tilde{v}$ to the standard deviation of Z . It is unsurprising that we recover Kyle's results under the risk-neutral probability. What is important in the theorem is how the parameters of the physical distribution determine the equilibrium processes.

We see in Part (H) that market maker inventories are mean reverting under the physical probability. This is a natural consequence of market maker aversion to inventory risk. Via Part (B), the mean reversion in market maker inventories corresponds to mean reversion in the equilibrium price – indeed, as we discuss further below, it is by adjusting the price that market makers induce mean reversion in their inventories. Because the price is mean reverting under the physical probability but not under the risk-neutral probability, the variance of P_1 is higher under $\mathbb{Q}$ than under $\mathbb{P}$. From Part (D), we see that this means that the variance of $\tilde{v}$ must be higher under $\mathbb{Q}$ than under $\mathbb{P}$. This is shown explicitly in the formula in Part (C). The higher risk-neutral variance means that there is a variance risk premium, as has been documented empirically. The mean reversion of prices also implies that there is excess volatility, in the sense that cumulative squared price changes exceed the physical variance of the asset on average. This follows from the fact that

$$(dP)^2 = (\lambda\sigma_z dB^*)^2 = \lambda^2\sigma_z^2 dt > \sigma_v^2 dt.$$

Part (B) of the theorem shows that the illiquidity parameter λ depends on two ingredients:

adverse selection (σ_v/σ_z) and inventory costs ($\alpha\sigma_v^2$). In the absence of inventory costs ($\alpha = 0$), the price impact parameter λ coincides with the adverse selection component, σ_v/σ_z , as in the standard Kyle (1985) model. In the absence of adverse selection ($\sigma_z = \infty$), λ coincides with the inventory cost component, $\alpha\sigma_v^2$, as in a CARA/normal competitive equilibrium with CARA investors and noise traders. Notice that λ is not a simple sum of the two ingredients. In terms of $x := \alpha\sigma_v^2$ and $y := \sigma_v/\sigma_z$, it equals $x/2 + \sqrt{x^2/4 + y^2}$. This is an increasing convex function of x and y with both partial derivatives being less than 1. Thus, the contribution of each component to λ is smaller than it would be if the other were not present.

Part (H) of the theorem shows that Y reverts towards β under the physical probability, which means that market maker inventories $\beta - Y$ revert towards zero. Combining this with Part (B), we obtain

$$dP_t = \lambda\delta_t(\beta - Y_t)dt + \lambda\sigma_z dB_t \quad (4.13)$$

When market makers have positive inventories, they quote lower prices in order to shed the inventories, with the consequence that prices are expected to be higher in the future; hence, the expected return (= risk premium) is positive. The positive risk premium is compensation for bearing the inventory risk. The reverse is true when market makers have short positions. This is shown explicitly for P_0 in Parts (A) and (B). Notice, however, that illiquidity (λ) is invariant with respect to market maker inventories. Market makers manage inventories by incentivizing the trades of the informed trader through their choice of price levels, rather than through their choice of spreads. This generalizes the result of ? to a dynamic model with adverse selection.

By substituting $P_t = \mu^* + \lambda Y_t$ from Part (B) in (??), we obtain

$$dP_t = \delta_t(\mu^* + \lambda\beta - P_t)dt + \lambda\sigma_z dB_t. \quad (4.14)$$

Thus, given the information of market makers, the price tends to revert to $\mu^* + \lambda\beta$ (implying in particular that it is expected to rise starting at date 0 when inventories are positive). The mean reversion implies short-term reversals. The reversion to $\mu^* + \lambda\beta$ from the market makers' point of view contrasts with the expected price change given the information of the informed trader. Part (E) implies that the informed trader sees the evolution of the price as

$$dP_t = \frac{1}{1-t}(\tilde{v} - P_t)dt + \lambda dZ_t. \quad (4.15)$$

Thus, she sees the price as reverting to $\tilde{v}$ (as a result of her own trades). Of course, the drift and the stochastic part of (??) are not separately observable to market makers.

In the risk-neutral model, competition between market makers forces expected market maker profits to zero. However, in the current model, market makers must be compensated for bearing inventory risk. It follows from Part (F) that expected market maker profits are

$$\frac{\alpha\sigma_z^2\sigma_v^2}{2} - \frac{\alpha^2\beta^2\sigma_v^4}{2\lambda}. \quad (4.16)$$

This shows that market makers expect to earn positive profits when $\beta = 0$. The magnitude of expected profits depends on inventory costs $\alpha\sigma_v^2$ and the volatility of inventories σ_z . However, when $\beta \neq 0$, market makers will accept losses to shed the initial inventory risk. Their willingness to accept losses in this circumstance contributes positively to informed trading profits, as shown in (??). In fact, expected informed trading profits are

$$\left(1 - \frac{\alpha\sigma_v^2}{2\lambda}\right) \lambda\sigma_z^2 + \left(\frac{\alpha\sigma_v^2}{2\lambda}\right) \alpha\beta^2\sigma_v^2 \quad (4.17)$$

which is a weighted average of the expected losses $\lambda\sigma_z^2$ of noise traders and the quantity $\alpha\beta^2\sigma_v^2$, which can be interpreted as the required risk premium per unit risk ($\alpha\beta\sigma_v^2$) and the quantity of risk β initially held by market makers.

Part (G) shows that the SDF is the exponential of a quadratic in the asset value $\tilde{v}$ and hence is non-monotone in $\tilde{v}$. In fact, it is high when $\tilde{v}$ is either very low or very large. This is because market makers lose money whenever the informed trader has extreme information. Thus, “bad states of the world” for the representative investor (representative market maker in this model) are states in which $\tilde{v}$ takes extreme values, either high or low, rather than states in which $\tilde{v}$ takes low values, which is normally the case in consumption-based asset pricing models. This implies that deep out-of-the-money call options will be expensive, so expected call returns may be non-monotone in the strike price. This is in contrast to the Black-Scholes model, in which call options with higher strikes have higher expected returns because they are more highly levered and the SDF is log-linear.

4.2 Multiple Normally Distributed Assets

Before considering the general model with possibly non-normally distributed assets, we first explain how Theorem ?? extends to multiple normally distributed assets. Assume there are n assets and $\tilde{v}$ is normally distributed under the physical probability with mean vector μ and nonsingular covariance matrix Σ_v . Diagonalize the symmetric positive-definite matrix $\Sigma_z^{1/2}\Sigma_v\Sigma_z^{1/2}$ as VEV' , where V is the orthogonal matrix whose columns are the eigenvectors of $\Sigma_z^{1/2}\Sigma_v\Sigma_z^{1/2}$ and $E := \text{diag}(e_1, \dots, e_n)$ is the diagonal matrix of eigenvalues. For each i , define

$$d_i = e_i + \frac{1}{2}\alpha e_i \left(\alpha e_i + \sqrt{\alpha^2 e_i^2 + 4e_i} \right). \quad (4.18)$$

For each t and each i , define

$$c_{it} = \frac{\alpha\sqrt{d_i}}{1 + (1-t)\alpha\sqrt{d_i}}.$$

Set $D = \text{diag}(d_1, \dots, d_n)$ and $C_t = \text{diag}(c_{1t}, \dots, c_{nt})$.

Theorem 4.2. *Assume that $\tilde{v}$ is normally distributed under the physical probability with mean vector μ and nonsingular covariance matrix Σ_v . Assume the informed trader is constrained to choose a*

strategy such that

$$\mathbb{E}^{\mathbb{P}} \int_0^1 \|X_t\|^2 dt < \infty. \quad (4.19)$$

There is an equilibrium in which the risk-neutral distribution of $\tilde{v}$ is also normal. Furthermore, the following hold:

A. The risk-neutral mean of $\tilde{v}$ is

$$\mu^* := \mu - \alpha \Sigma_v \beta. \quad (4.20)$$

B. The equilibrium price satisfies $P_0 = \mu^*$ and $dP_t = \Lambda dY_t$, where

$$\Lambda := \Sigma_z^{-1/2} V D^{1/2} V' \Sigma_z^{-1/2} \quad (4.21)$$

C. The risk-neutral covariance matrix of $\tilde{v}$ and the covariance matrix of the price changes dP is

$$\Sigma_v^* := \Sigma_z^{-1/2} V D V' \Sigma_z^{-1/2} \quad (4.22)$$

and

$$\Sigma_v^* - \Sigma_v = \Sigma_z^{-1/2} V (D - E) V' \Sigma_z^{-1/2}, \quad (4.23)$$

which is positive definite.

D. The market is eventually informationally efficient in the sense that $P_t \rightarrow \tilde{v}$ with probability one (under both $\mathbb{P}$ and $\mathbb{Q}$).

E. The equilibrium trading strategy is

$$\theta_t := \frac{1}{1-t} [\Lambda^{-1}(\tilde{v} - \mu^*) - Y_t]. \quad (4.24)$$

F. Under $\mathbb{P}$, the expected profit of the informed trader conditional on $\tilde{v}$ is

$$\frac{1}{2} \text{tr}(\Sigma_z \Lambda) + \frac{1}{2} (\tilde{v} - \mu^*)' \Lambda^{-1} (\tilde{v} - \mu^*). \quad (4.25)$$

The unconditional expected profit of the informed trader is

$$\text{tr}(\Sigma_z \Lambda) - \frac{\alpha}{2} \text{tr}(\Sigma_z \Sigma_v) + \frac{\alpha^2}{2} \beta' \Sigma_v \Lambda^{-1} \Sigma_v \beta \quad (4.26)$$

and the unconditional expected loss of liquidity traders is $\text{tr}(\Sigma_z \Lambda)$.

G. The SDF is

$$\tilde{m} = e^{\frac{\alpha}{2} (\tilde{v} - \mu^*)' \Lambda^{-1} (\tilde{v} - \mu^*) - \alpha \beta' \tilde{v}} / \mathbb{E} \left[e^{\frac{\alpha}{2} (\tilde{v} - \mu^*)' \Lambda^{-1} (\tilde{v} - \mu^*) - \alpha \beta' \tilde{v}} \right]. \quad (4.27)$$

H. The cumulative order process Y is a $(0, \Sigma_z)$ Brownian motion on $(\mathcal{F}^Y, \mathbb{Q})$ and satisfies

$$dY_t = \Delta_t(\beta - Y_t) dt + \Sigma_z^{1/2} dB_t \quad (4.28)$$

where B is a standard n -dimensional $\mathcal{F}^Y$ Brownian motion under $\mathbb{P}$ and

$$\Delta_t := \Sigma_z^{1/2} V C_t V' \Sigma_z^{-1/2}. \quad (4.29)$$

The implications we saw in the single asset case hold for multiple assets as well: there is mean reversion in market maker inventories and prices, there is a variance risk premium, there is excess volatility, market makers make money on average as compensation for bearing inventory risk but may accept losses to shed initial inventories, illiquidity depends on adverse selection and inventory costs, and the log of the SDF is quadratic in $\tilde{v}$. Note that (??) implies that the elements of the diagonal matrix $D^{1/2}$ in the formula for Λ are equal to

$$\sqrt{d_i} = \frac{\alpha e_i}{2} + \sqrt{\frac{\alpha^2 e_i^2}{4} + e_i}. \quad (4.30)$$

From this, it is easy to see that the formula for Λ in Part (B) simplifies to the formula for λ in Part (B) of Theorem ?? when $n = 1$.

To see the excess volatility in the multivariate model, observe that Equations (??) and (??) imply that

$$\Lambda \Sigma_z \Lambda = \Sigma_v^*. \quad (4.31)$$

This is the multivariate analogue of Part (C) of Theorem ?. In conjunction with $dP = \Lambda dY$, Equation (??) implies that the instantaneous covariance matrix of the price changes is the risk-neutral covariance matrix of $\tilde{v}$, which exceeds the physical covariance matrix of $\tilde{v}$ by a positive definite matrix, as shown in Part (C) of Theorem ?.

In addition to excess volatility, there is also excess co-movement, in the sense that correlations of price changes exceed correlations of asset values. For a simple example, suppose $n = 2$ and the two assets are uncorrelated and have common variance σ_v^2 under the physical probability. Suppose the noise trades have common variance σ_z^2 and correlation $\rho > 0$. Using (??), we can calculate that the correlation of the price changes is

$$\frac{(1 - \rho)d_2 - (1 + \rho)d_1}{(1 - \rho)d_2 + (1 + \rho)d_1}, \quad (4.32)$$

where d_i depends on α and the eigenvalue e_i as in (??) and in this example $e_1 = \sigma_v^2 \sigma_z^2 (1 - \rho)$ and $e_2 = \sigma_v^2 \sigma_z^2 (1 + \rho)$. From these facts, we can calculate that the correlation of the price changes is increasing in α and lies in the interval $(0, \rho)$. Thus, the price changes inherit some but not all of the correlation of the noise trades when market makers are risk averse, even when the asset values are independent under the physical measure. Like excess volatility, excess co-movement stems from

the fact that trades move prices both because of informational effects and because of risk premia associated with inventory concerns. The risk premia comove when inventories comove.

4.3 The General Model

We now consider the n asset model without the normality assumption. The risk-neutral probability is defined on the sigma-field $\mathcal{F}_1^Y$. It will again be true that the market is eventually informationally efficient, so $\tilde{v}$ will be measurable in equilibrium with respect to $\mathcal{F}_1^Y$. Thus, the risk-neutral probability will determine a distribution for $\tilde{v}$. Denote it by F^* . We will implicitly define a class of physical distributions F for $\tilde{v}$ by imposing joint conditions on (F, F^*) . This class includes all normal distributions, but it is much larger than just the normal distributions.

Definition 4.1. *Consider any pair (F, F^*) of distribution functions on $\mathbb{R}^n$. Let Γ denote the Brenier potential such that $\nabla\Gamma$ transports G to F^* and $\mathbb{E}^G[\Gamma(Z_1)] = 0$. Define K as in Theorem ?? . Define*

$$\tilde{\chi} = \exp \left\{ -\alpha(Z_1 - \beta)' \nabla\Gamma(Z_1) + \alpha\Gamma(Z_1) + \alpha \operatorname{tr} \left(\Sigma_z \int_0^1 \nabla_z^2 K(t, Z_t) dt \right) \right\}. \quad (4.33)$$

Define (F, F^*) to be admissible if $\tilde{\chi}$ has a finite ν -mean and F is the $\hat{\nu}$ distribution of $\nabla\Gamma(Z_1)$, where $d\hat{\nu}/d\nu = \tilde{\chi}/\mathbb{E}^\nu[\tilde{\chi}]$.

Definition ?? provides a recipe for starting with an arbitrary distribution F^* on $\mathbb{R}^n$ and then computing F such that (F, F^*) is admissible (provided $\tilde{\chi}$ has a finite ν -mean). This is analogous to the standard practice in derivative security pricing, which is to start by specifying the risk-neutral distribution of the underlying assets. The change of measure $d\hat{\nu}/d\nu$ in the definition is the equilibrium change of measure from the risk-neutral distribution of Y on $\mathcal{F}^Y$ to the physical distribution. We show in the appendix that if one starts with a normal F^* , then one obtains a normal F . Moreover, every normal F can be obtained in this way. This produces Theorems ?? and ??. We also show in the appendix how one can numerically compute F starting from an arbitrary F^* (satisfying the mild regularity condition that $\tilde{\chi}$ has a finite ν -mean).

We need a few other minor sufficient conditions to construct an equilibrium such that F^* is the risk-neutral distribution corresponding to the physical distribution F . Here, we give easily interpretable such conditions for F^* .

Theorem 4.3. *Let F^* be a distribution function on $\mathbb{R}^n$. Assume that either (i) F^* has bounded support or (ii) F^* is absolutely continuous with respect to Lebesgue measure and its density is $e^{-U(v)}$ for some U satisfying $\nabla^2 U \geq \kappa I$ for some $\kappa > 0$ and I the identity matrix of dimension n . Then, $\tilde{\chi}$ defined in (??) has a finite mean, and the pair (F, F^*) is admissible, where F is the $\hat{\nu}$ distribution of $\nabla\Gamma(Z_1)$ as in Definition ??. Define $\mathbb{P}$ to be the product measure $G \otimes F \otimes \nu$ over $(u, v, Z) \in \mathbb{R}^n \times \mathbb{R}^n \times C[0, 1]^n$. Then, $\mathbb{E}^\mathbb{P}[\tilde{v}]$ is finite. Define Γ as in Definition ??, and define K and H from Γ as in Theorem ??. Assume the informed trader is constrained to choose a trading strategy X satisfying (??) under $\mathbb{P}$. There is an equilibrium where the risk-neutral distribution $\mathbb{Q}$ is*

given by (??) and in which F^* is the risk-neutral distribution of $\tilde{v}$. Furthermore, the following hold for this equilibrium:

A. The equilibrium price is $P_t = \nabla_y K(t, Y_t)$ for $t \in [0, 1)$ and satisfies, on this interval,

$$dP_t = \Lambda(t, Y_t) dY_t \quad (4.34)$$

where $\Lambda = \nabla_y^2 K$.

B. The market is eventually informationally efficient, in the sense that $P_t \rightarrow \tilde{v}$ with probability one (under both $\mathbb{P}$ and $\mathbb{Q}$).

C. The equilibrium trading strategy is

$$\theta_t := \frac{1}{1-t}(\tilde{\zeta} - Y_t) + \Sigma_z \nabla_y \psi(t, Y_t, \tilde{\zeta}) \quad (4.35)$$

where $\tilde{\zeta} = f(\tilde{u}, \tilde{v})$ is constructed in Lemma ??(1), so that $\nabla \Gamma(\tilde{\zeta}) = \tilde{v}$ almost surely and the $\mathbb{P}$ -distribution of $\tilde{\zeta}$ is the $\hat{\nu}$ -distribution of Z_1 and for $b \in \mathcal{D}_\Gamma$,

$$\begin{aligned} \psi(t, z, b) &:= -\alpha(b - \beta)' \nabla \Gamma(b) + \alpha \Gamma(b) \\ &+ \log \mathbb{E}^\nu \left[\exp \left\{ \alpha \operatorname{tr} \left(\Sigma_z \int_t^1 \nabla_z^2 K(u, Z_u) du \right) \right\} \middle| Z_t = z, Z_1 = b \right]. \end{aligned} \quad (4.36)$$

D. Under the physical probability, the expected profit of the informed trader conditional on $\tilde{v}$ is $\Gamma^*(\tilde{v})$.

E. The cumulative order process Y is a $(0, \Sigma_z)$ Brownian motion on $(\mathcal{F}^Y, \mathbb{Q})$ and satisfies

$$dY_t = \Sigma_z \nabla_y \phi(t, Y_t) dt + \Sigma_z^{1/2} dB_t \quad (4.37)$$

where B is a standard n -dimensional $\mathcal{F}^Y$ Brownian motion under the physical probability and

$$\begin{aligned} \phi(t, z) &:= \log \mathbb{E}^\nu \left[\exp \left\{ -\alpha(Z_1 - \beta)' \nabla \Gamma(Z_1) + \alpha \Gamma(Z_1) \right. \right. \\ &\quad \left. \left. + \alpha \operatorname{tr} \left(\Sigma_z \int_t^1 \nabla_z^2 K(u, Z_u) du \right) \right\} \middle| Z_t = z \right]. \end{aligned} \quad (4.38)$$

Also, $Y_1 = \tilde{\zeta}$ with probability one (under both $\mathbb{P}$ and $\mathbb{Q}$).

The term $\Sigma_z \nabla_y \phi(t, Y_t) dt$ in (??) is the Girsanov drift term in the change of measure from $\mathbb{P}$ to $\mathbb{Q}$. It takes the form $\Delta_t(\beta - Y_t) dt$ in the case of a normally distributed $\tilde{v}$, where Δ_t is the deterministic matrix (??). The second term in the informed trading strategy (??) is related to this Girsanov drift. That term disappears when $\tilde{v}$ is normally distributed, because the Brenier potential is quadratic in that case, which makes the Hessian of K a constant matrix and causes $\psi(t, z, b)$ to be independent of z .

The formula $\Gamma^*(\tilde{v})$ for the conditional expected profit of the informed trader in Part (D) is valid in all of our models – with and without risk aversion and with and without normality. It implies a dependence of the conditional expected profit on risk aversion and initial market maker inventories when market makers are risk averse, as discussed earlier. The dependence is seen more explicitly in Theorems ?? and ??.

5 Informed Trading in Options

We now apply Theorem ?? to study informed trading in options. In particular, we investigate the extent to which information from the options market can be used to predict the return of the underlying asset. We consider an asset and a European call option on the asset that matures at the announcement date 1. This is the model studied by ?, except that we allow market makers to be risk averse—and hence for the return of the underlying asset to be predictable—and we do not need to make special parametric assumptions to solve the model. By put-call parity, the model is equivalent to one in which an underlying asset and a put option are traded or to one in which an underlying asset and a straddle are traded. For the sake of brevity, we call the underlying asset a stock. We use subscripts s and o to denote ‘stock’ and ‘option.’

We take the initial endowment of the market makers to be $\beta = 0$, and we set absolute risk aversion to be $\alpha = 0.2$.⁸ We assume the stock is truncated lognormally distributed: $\log \tilde{v}_s = \mu + \sigma \tilde{x}$, where $\tilde{x}$ is a standard normal conditioned on $|\tilde{x}| \leq 3.3$ under the risk-neutral distribution. We set $\sigma = 0.2$ and $\mu = \log(100) - \sigma^2/2$. The truncation is the hypothesis that makes Theorem ?? applicable. It removes less than 0.1% of the lognormal mass, and the grid $\mu \pm 3.3\sigma$ used in Appendix ?? is the support of the truncated distribution. The risk-neutral mean of $\tilde{v}_s$ is approximately 100, and we set the strike of the option to be 100. We consider four different Σ_z matrices—varying the sign of the correlation between stock and option uninformed order imbalances and varying the relative standard deviations of stock and option orders—and solve the model for each. We then simulate 10,000 sample paths for each choice of Σ_z . Figures ?? and ?? present aspects of the solution for one choice of Σ_z ; Figures ?? and ?? present the solution for two values of Σ_z ; and Table ?? presents results for all four values of Σ_z . The qualitative features shown in the Figures are very similar for the other values of Σ_z .

Figure ?? illustrates the solution of the model at $t = 0.5$. The state vector is Y_t , which is the vector of stock and option order imbalances. The option is in the money (the stock price is above 100) in the top-right part of each panel in Figure ?. Panels (a) and (b) show that, when the option is in the money, the stock and option prices are approximately functions of the sum of stock and option imbalances. This reflects the fact that the assets are informationally equivalent when the option is certain to finish in the money. On the other hand, in the bottom left part of the panels,

⁸The qualitative results do not depend on the magnitude of risk aversion. Our choice of $\alpha = 0.2$ is prompted by the following reasoning. We have not specified the units of wealth, but since we will take the standard deviation of stock noise trading to be 2 in all examples in this section, a reasonable unit would be \$100 million. An investor with an absolute risk aversion of 0.2 when wealth is measured in units of \$100 million would pay \$1,000 to avoid a coin toss for \$1,000,000. This seems like a not unreasonable level of risk aversion.

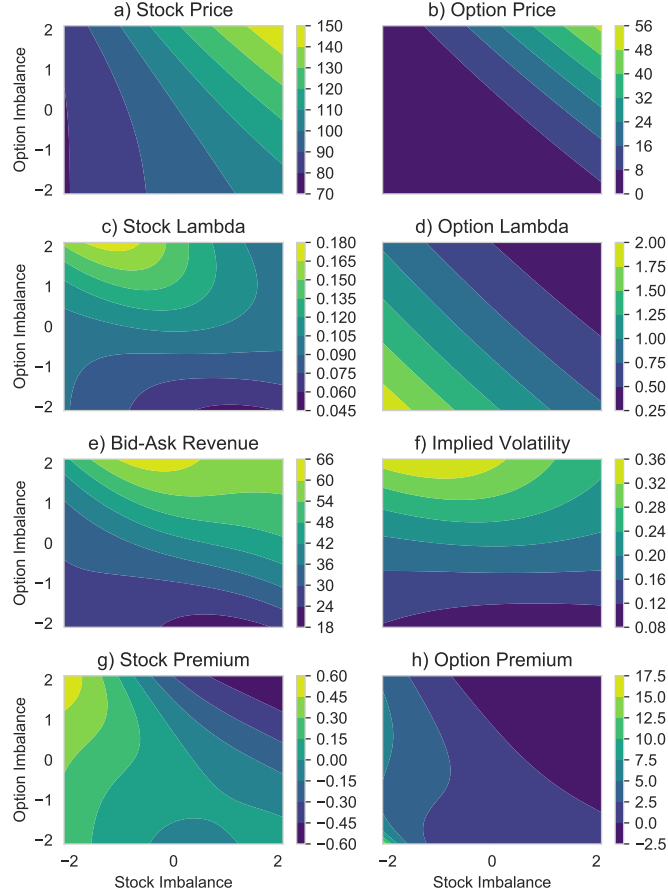


Figure 1: Solution of the Model at $t = 0.5$

The figure presents properties of the model at the midpoint $t = 0.5$ as a function of the state variables $Y_s = \text{Stock Imbalance}$ and $Y_o = \text{Option Imbalance}$. The plots are for $\Sigma_z = [[4, -2], [-2, 4]]$.

the option is out of the money. When the option is deeply out of the money, the stock price is virtually independent of the option imbalance, reflecting the fact that option orders are almost certainly pure noise in that circumstance. These features become more prominent as $t \rightarrow 1$, and at $t = 1$, the stock price depends solely on $y_s + y_o$ when the option is in the money, depends solely on y_s when the option is out of the money, and is discontinuous in (y_s, y_o) at the locus of points such that the option is exactly at the money.

Information flow in the market is affected in a qualitative way by the presence of option trading, which introduces a strong nonlinearity. For example, conditional variances can rise over time in nonlinear models, unlike linear Gaussian models. In this model, the conditional variance of $\tilde{v}_s$ increases at many state/dates. Uncertainty can also shrink very fast, depending on stock and option orders. This phenomenon influences many of the variables plotted in Figure ???. When the call has been purchased and the stock sold, market makers become very uncertain about the value of the stock. It is quite likely that the trades in one of the assets are noise trades, but market makers do not know which is which. On the other hand, when the call has been sold and the stock purchased,

it is quite likely that the call will finish out of the money, but the stock value is not too low. This narrows the range of possibilities considerably. These two alternatives correspond to the top left and bottom right parts of the panels. This phenomenon also becomes more prominent as $t \rightarrow 1$ and mirrors the discontinuity of the stock price in (y_s, y_o) at $t = 1$, which occurs roughly along the off-diagonal and is larger in the top left than in the bottom right in the (y_s, y_o) plane.

The stock and option lambdas in Panels (c) and (d) of Figure ?? are the diagonal elements of the Λ matrix, each divided by its respective price, so they are in relative terms. The option lambda is considerably larger than the stock lambda, consistent with the empirical fact that liquidity is generally lower in options markets than in markets for the underlying assets⁹. The option lambda is decreasing in moneyness. The stock lambda is influenced by the phenomenon discussed in the previous paragraph. It is high in the top left where market makers are very uncertain about the stock value, and it is low in the bottom right where uncertainty is much lower. Moreover, it is generally increasing in the option order imbalance.

Panel (e) in Figure ?? plots the price-impact revenue received by market makers from noise traders per unit time. This is $\text{tr}(\Sigma_z \nabla_y^2 K(t, y))$. Panel (f) plots the Black-Scholes implied volatility of the option. Both plots have a similar pattern to the stock lambda: they are generally increasing in the option imbalance and in particular are high in the upper left and low in the bottom right.

Panels (g) and (h) in Figure ?? plot the risk premia of the stock and option. These are the elements of $\Lambda(t, y) \Sigma_z \nabla \phi(t, y)$, each divided by its respective price, so they are the expected rates of return per unit time. Both are generally inversely related to the stock price, though the patterns are complex. An interesting feature of the stock premium is that it is especially high in the upper left, where the implied volatility is also high. This induces a positive correlation between the implied volatility and subsequent stock returns, as we will see.

Figure ?? presents some data from a simulation of the model, using the same Σ_z as in Figure ?. We simulate the path of Y and $dP_t = \Lambda(t, Y_t) dY_t$, using the physical dynamics of Y given in Theorem ?. The terminal value of P is $P_1 = (\tilde{v}_s, \tilde{v}_o)$, where $\tilde{v}_o = (\tilde{v}_s - K)^+$. Panel (a) presents the realized values of the SDF $e^{-\alpha \tilde{w}} / E[e^{-\alpha \tilde{w}}]$, replacing the expectation with the mean across simulations, plotted against the realized values of $\tilde{v}_s$. We expect the SDF to be high for extreme values of $\tilde{v}_s$, because the informed trader makes more money and market makers typically lose money when the informed trader has unexpected information. This same phenomenon appears in Theorem ?. It is somewhat surprising that the SDF is also often high for values of $\tilde{v}_s$ near the option strike. This occurs when there are large buy orders and option sell orders early during the trading period. These push the market to the bottom right in the panels in Figure ?, where uncertainty is low. Competition between market makers is then much like risk-neutral competition, pushing down the lambdas and transaction cost revenue from noise traders. Consequently, the revenue is insufficient to cover the losses to the informed trader from the initially overvalued option.

Panel (b) of Figure ?? shows the assumed lognormal risk-neutral density and a kernel estimate of the physical density of $\tilde{v}_s$, based on the simulated values of P_1 . The risk-neutral density is

⁹Unlike (?, Theorem 3) where lambda's are in absolute terms, lambda's considered here are in relative terms.

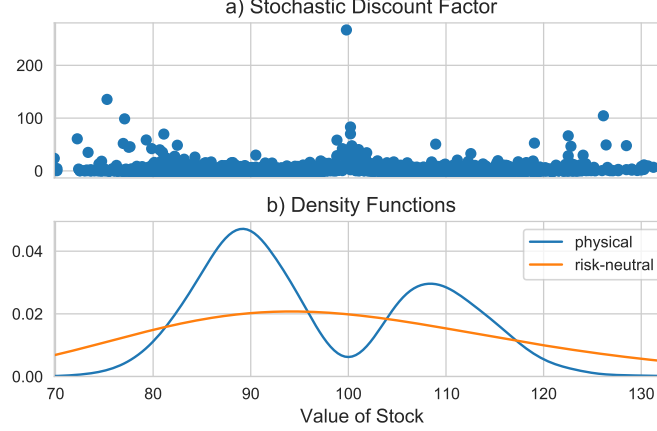


Figure 2: Risk Neutral and Physical Densities

Panel (a) is a scatter plot of the SDF versus realized values of $\tilde{v}_s$. Panel (b) presents a kernel estimate of the physical density based on realized values of $\tilde{v}_s$ and also presents the assumed lognormal risk-neutral density. The scatter plot and kernel estimate are based on 10,000 simulations with $\Sigma_z = [[4, -2], [-2, 4]]$.

the physical density multiplied by the SDF, so high values of the SDF produce high risk-neutral probabilities. Because the SDF is high for extreme values of $\tilde{v}_s$, the risk-neutral distribution has extra weight in the tails and a larger variance than the physical distribution, as in Theorem ???. The high SDF near the option strike also adds weight to that part of the distribution. If the physical distribution were unimodal, this would produce a spike in the risk-neutral distribution. However, we start with a truncated lognormal risk-neutral distribution, so the added weight in the risk-neutral distribution is possible only if the physical density is small near the option strike. Consequently, the physical density that is consistent with the given risk-neutral density is bimodal. This particular feature of the model seems to be an artifact of assuming there is only a single option traded. Trading multiple strikes should smooth out the spikes in the risk-neutral density or equivalently the troughs in the physical density.

Figure ?? shows the physical distribution of Y_1 for two choices of Σ_z . Panel (a) uses the same value of Σ_z as in Figures ?? and ??. Panel (b) uses the same Σ_z except that the correlation of Z_s and Z_o is switched from negative to positive. The figure shows that the correlation of net imbalances Y_s and Y_o from date 0 to date 1 is negative, regardless of the correlation of noise trades. This means that market makers usually end up with hedged positions—long the stock if they are short calls and long calls if they are short the stock. This is an intuitive consequence of risk aversion. One can see from the figure that market makers are short the call and long the stock at the modes of the distributions. The same is true for the means of the distributions, so on average market makers in this model sell options and use the stock to hedge, consistent with what actual option market makers usually do (for example, ? show that option market makers on average hold portfolios with negative gammas, meaning that they are net short options).

Figure ?? plots market maker profits as a function of Y_1 for the same two values of Σ_z as in

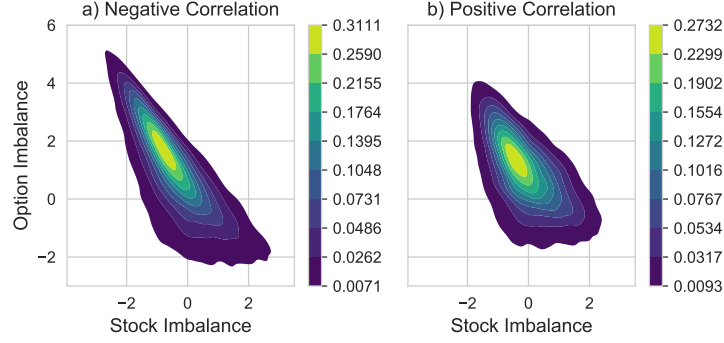


Figure 3: Physical Density of Order Imbalances at date $t = 1$

The plots are kernel estimates of the bivariate density of $Y_s = \text{Stock Imbalance}$ and $Y_o = \text{Option Imbalance}$ at the terminal date $t = 1$. Panel (a) is based on 10,000 simulations with $\Sigma_z = [[4, -2], [-2, 4]]$, and Panel (b) is based on 10,000 simulations with $\Sigma_z = [[4, 2], [2, 4]]$.

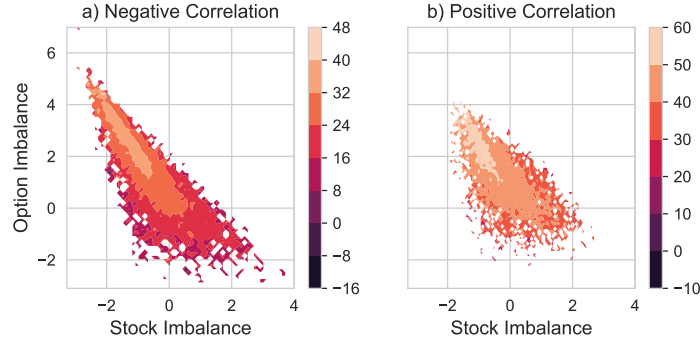


Figure 4: Market Maker Profits

Realized market maker profits are averaged within bins in a 100×100 grid. Panel (a) is based on 10,000 simulations with $\Sigma_z = [[4, -2], [-2, 4]]$, and Panel (b) is based on 10,000 simulations with $\Sigma_z = [[4, 2], [2, 4]]$.

Figure ?? shows that market maker profits depend on the path of Y , not just on Y_1 , so the figure shows an estimate of expected market maker profits conditional on Y_1 . The region in which profits are highest is the upper left, which, as discussed in connection with Figure ??, is the area in which market makers are most uncertain about the value of the stock. The combination of stock sells and option buys amplifies uncertainty for market makers. To compensate for this extra risk, price impacts are higher, as shown in Figure ??, and market maker profits are higher on average.

We can use the simulated model to estimate moments and conditional moments for which we do not have analytic expressions. As an illustration, we look at the correlation between the implied volatility and future stock returns. ? document empirically that higher implied volatilities predict higher future stock returns. We find the same correlation in our model: the conditional risk premium is higher when the implied volatility is higher. This is a robust finding across all values of Σ_z that we examine. It is driven by the fact that implied volatilities are highest in the upper left corners of the panels in Figure ??, where the stock has been sold and the option purchased. The stock risk

Table 1: Regression of Stock Risk Premium on Implied Volatility at $t = 0.5$

The stock risk premium (defined as the stock-element of the vector $\Lambda(t, y)\Sigma_z \nabla \phi(t, y)$ divided by the stock price) is regressed at the midpoint $t = 0.5$ on the change in the implied volatility from $t = 0$ to $t = 0.5$. The data points are independent simulations of the model. Regressions (2), (4), (6), and (8) include a control for the continuously compounded stock return from $t = 0$ to $t = 0.5$. Under each coefficient is the 95% confidence interval (between square brackets) and standard error.

	$\Sigma_z = [4, -2; -2, 4]$		$\Sigma_z = [4, -1; -1, 1]$	
	(1)	(2)	(3)	(4)
Intercept	-0.003 [-0.011, 0.005] s.e. = 0.004	0.075 [0.072, 0.078] s.e. = 0.001	0.007 [0.002, 0.012] s.e. = 0.003	0.037 [0.036, 0.038] s.e. = 0.001
Δ Implied vol	0.444 [0.384, 0.503] s.e. = 0.030	0.491 [0.472, 0.511] s.e. = 0.010	0.354 [0.277, 0.430] s.e. = 0.039	0.547 [0.527, 0.568] s.e. = 0.010
Stock return		-2.187 [-2.202, -2.172] s.e. = 0.008		-1.774 [-1.783, -1.764] s.e. = 0.005
Obs.	10000	10000	10000	10000
R ²	0.021	0.894	0.008	0.932
	$\Sigma_z = [4, 2; 2, 4]$		$\Sigma_z = [4, 1; 1, 4]$	
	(5)	(6)	(7)	(8)
Intercept	0.041 [0.035, 0.048] s.e. = 0.003	0.346 [0.343, 0.348] s.e. = 0.001	0.037 [0.032, 0.041] s.e. = 0.002	0.183 [0.182, 0.184] s.e. = 0.001
Δ Implied vol	0.381 [0.280, 0.482] s.e. = 0.052	-0.603 [-0.630, -0.575] s.e. = 0.014	0.253 [0.135, 0.372] s.e. = 0.060	-0.188 [-0.217, -0.160] s.e. = 0.014
Stock return		-2.360 [-2.373, -2.347] s.e. = 0.007		-1.830 [-1.839, -1.821] s.e. = 0.004
Obs.	10000	10000	10000	10000
R ²	0.005	0.928	0.002	0.943

premium is a decreasing function of stock orders, and is also especially high in the upper left corner in Figure ??; hence, the implied volatility and risk premium are positively correlated. Table ?? reports regressions of the risk premium on the implied volatility at $t = 0.5$ for all four values of Σ_z .¹⁰ The coefficient is positive and significant in all of the univariate regressions.

Table ?? also reports regressions of the risk premium on the implied volatility and the prior stock return. This is a partial replication of Table VI of ?, which controls for prior returns (‘short term reversal’) and other stock return predictors studied in the literature. The coefficient on the prior return is consistently negative in Table ??, as it is in An et al.’s Table VI. This is a consequence of excess volatility and mean reversion as discussed previously. The sign of the coefficient on the

¹⁰We look only at options with time remaining to maturity equal to 0.5 to mirror An et al.’s study of options with a fixed time to maturity (30 days). The implied volatility is initially 20% for all observations, so regressing on the change is equivalent to regressing on the level.

implied volatility, unlike the univariate coefficient, is sensitive to the model parameters. It is positive (as in the data) when the correlation between stock and option noise trades is negative (Columns 2 and 4) but negative when the correlation between stock and option noise trades is positive (Columns 6 and 8). The reason that the sign varies across these cases in the multivariate regressions is that the correlation between the regressors changes sign across the cases. The prior return and implied volatility are positively correlated in Columns 2 and 4 (thus, the multivariate regression coefficient is larger than the univariate coefficient) but negatively correlated in Columns 6 and 8 (thus, the multivariate regression coefficient is smaller than the univariate coefficient and in fact is negative). Additional empirical investigation of this phenomenon would be useful.

6 Conclusion

The solution of the Kyle model is much simpler when described in terms of optimal transport theory. By recognizing this fact, we are able to extend the solution to a much broader class of models. The extension to risk-averse market makers combines two important literatures that have heretofore developed primarily in parallel: the adverse selection and the inventory risk explanations of market illiquidity. We are able to identify the contribution each makes to illiquidity. As one example of the applicability of this new approach, we derived a novel result regarding the predictive power of implied volatilities for stock returns.

We anticipate that many additional applications will be made in the future. The multiasset setting provides guidance for understanding spillover effects of informed trading. For example, ? provide evidence that private information can potentially benefit rival firms. General equilibrium extensions of our setting can provide guidance for further empirical work on such topic. Price manipulation is also worth investigation as per the recent findings in ? and ?. Manipulators may naturally leverage the derivatives market to increase their gains. Our setting can also be adapted to this kind of problem.

? consider a multi-period model of a market maker with significant market power collecting profits from bid-ask spreads. Market power can generate an equilibrium featuring simultaneously high liquidity and large bid-ask spread at market opening and closing as observed in the markets. Although challenging, our setting could also be extended to include market makers with market power.

Appendix

Appendix A Theorem ?? and Corollary ??

A.1 Lemmas

Lemma A.1. *Let Γ be a Brenier potential as described in Theorem ??, where $\|\tilde{v}\|^2$ is integrable. The random variables $|\Gamma(Z_1)|^{3/2}$ and $\|\nabla\Gamma(Z_1)\|^2$ are integrable. Furthermore, for each $y \in \mathbb{R}^n$ and each $t \in [0, 1)$, the random variables $|\Gamma(y + Z_1 - Z_t)|^{5/4}$ and $\|\nabla\Gamma(y + Z_1 - Z_t)\|^{3/2}$ are integrable.*

Proof. By the definition of Γ , $\nabla\Gamma(Z_1)$ has the distribution of $\tilde{v}$. Hence, $\|\nabla\Gamma(Z_1)\|^2$ is integrable. By the convexity of Γ , we have

$$\Gamma(0) + g'_0 Z_1 \leq \Gamma(Z_1) \leq \Gamma(0) + \nabla\Gamma(Z_1)' Z_1 \quad (\text{A.1})$$

where g_0 is an element of the subdifferential of Γ at 0. Because all moments of Z_1 exist and $\|\nabla\Gamma(Z_1)\|^2$ is integrable, Young's inequality implies that $|\nabla\Gamma(Z_1)' Z_1|^{3/2} \leq \frac{3\|\nabla\Gamma(Z_1)\|^2}{4} + \frac{\|Z_1\|^6}{4}$ is integrable. Thus, $|\Gamma(Z_1)|^{3/2}$ is bounded above and below by integrable random variables and hence is itself integrable. Fix t , and recall that $k(t, y, \cdot)$ denotes the normal $(y, (1-t)\Sigma_z)$ density function. For fixed $t \in [0, 1)$, we now want to show that $\|\nabla\Gamma(y + Z_1 - Z_t)\|^{3/2}$ and $|\Gamma(y + Z_1 - Z_t)|^{5/4}$ are integrable for all y , meaning

$$\int \|\nabla\Gamma(z)\|^{3/2} k(t, y, z) dz < \infty \quad (\text{A.2a})$$

and

$$\int |\Gamma(z)|^{5/4} k(t, y, z) dz < \infty \quad (\text{A.2b})$$

for all y . For every $t \in [0, 1)$ and $y, z \in \mathbb{R}^n$, since $(1-t)^{-1} \geq 1$,

$$k(t, y, z) \leq (1-t)^{-n/2} \exp\left(y' \Sigma_z^{-1} z - \frac{1}{2} y' \Sigma_z^{-1} y\right) k(0, 0, z).$$

Since G has exponential moments of all orders, Hölder's inequality with exponents $(4/3, 4)$ and the G -integrability of $\|\nabla\Gamma(Z_1)\|^2 = (\|\nabla\Gamma(Z_1)\|^{3/2})^{4/3}$ give (??), and Hölder's inequality with exponents $(6/5, 6)$ and the G -integrability of $|\Gamma(Z_1)|^{3/2} = (|\Gamma(Z_1)|^{5/4})^{6/5}$ give (??), for all $t \in [0, 1)$ and $y \in \mathbb{R}^n$. $\square$

Lemma A.2. *Let L, F be two distribution functions on $\mathbb{R}^n$, set $\mu = G \otimes F$ where G is the multivariate normal $(0, \Sigma_z)$ distribution and assume that there exists a map $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that transports L to F . Then, there exists a function $f : \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$ such that $\mu(\{(u, v) \mid g(f(u, v)) = v\}) = 1$, and $\mu(\{(u, v) \mid f(u, v) \leq a\}) = L(a)$ for all $a \in \mathbb{R}^n$.*

Proof. Denote $\hat{\Phi} \in \Pi(L, F)$ the joint distribution of $(Y, g(Y))$ where Y has distribution L . There exists a disintegration of the measure $\hat{\Phi}$ on $g(Y)$, meaning there exists a family of probability measures $\hat{\phi}$ such that $\hat{\Phi}(dy, dv) = \hat{\phi}(dy, v) dF(v)$, and $v \in \mathbb{R}^n \mapsto \hat{\phi}(A, v)$ is measurable for any Borel measurable subset A of $\mathbb{R}^n$.

$\hat{\Phi}$ is supported on $\{(y, g(y)) : y \in \mathbb{R}^n\}$. Thus, we also have that for F almost every $v \in \mathbb{R}^n$, the measure $\hat{\phi}(dy, v)$ is supported on $(g)^{-1}(v) := \{y \in \mathbb{R}^n : g(y) = v\}$. For each $v \in \mathbb{R}^n$, we use Brenier's theorem for the transport of the absolutely continuous distribution G to the probability measure $\hat{\phi}(dy, v)$ to obtain the existence of a function $f_v : \mathbb{R}^n \mapsto \mathbb{R}^n$ so that $f_v(\tilde{u})$ has distribution $\hat{\phi}(dy, v)$ where $\tilde{u}$ has distribution G . Set $f(u, v) = f_v(u)$ which is jointly measurable thanks to (??, Theorem

1.1). Since $\hat{\phi}(\mathrm{d}y, v)$ is supported in $(g)^{-1}(v)$ we have that $\mu(g(f(u, v)) = v) = 1$. Moreover, for each Borel A ,

$$\begin{aligned} G \otimes F(\{(u, v) \mid f(u, v) \in A\}) &= \int G(f_v(\tilde{u}) \in A) \mathrm{d}F(v) = \int \hat{\phi}(A, v) \mathrm{d}F(v) \\ &= \int \hat{\Phi}(A, \mathrm{d}v) = \int_A \mathrm{d}L. \end{aligned}$$

The last equality follows from the fact that the first marginal of $\hat{\Phi}$ is L . $\square$

Note that the Lemma ?? allows us to claim that $K(t, y)$ in (??) is well-defined for $(t, y) \in [0, 1] \times \mathbb{R}^n$.

Lemma A.3. *$K(t, y)$ is continuously differentiable in t and twice continuously differentiable in y on $[0, 1] \times \mathbb{R}^n$, with the derivative in t at $t = 0$ understood from the right, and*

$$\nabla_y K(t, y) = \int_{\mathbb{R}^n} \nabla \Gamma(z) k(t, y, z) \mathrm{d}z, \quad \nabla_y^2 K(t, y) = \int_{\mathbb{R}^n} \nabla \Gamma(z) \frac{(z - y)' \Sigma_z^{-1}}{1 - t} k(t, y, z) \mathrm{d}z \quad (\text{A.3})$$

for $t \in [0, 1)$ and $y \in \mathbb{R}^n$.

Proof. Let $z \mapsto k(t, y, z)$ be the normal $(y, (1 - t)\Sigma_z)$ density function. The definition of $K(t, y)$ is $K(t, y) = \int \Gamma(z) k(t, y, z) \mathrm{d}z$. For all $t \in [0, 1)$ and y , let $y_1, \dots, y_N$ be the vertices of a simplex containing a neighborhood of y . Since $\log(k(t, \tilde{y}, z)/k(t, y_i, z))$ is affine in z with slope $\Sigma_z^{-1}(\tilde{y} - y_i)/(1 - t)$, choosing for each z the index i that maximizes $z' \Sigma_z^{-1} y_i$ gives a constant C such that $k(t, \tilde{y}, z) \leq C \max\{k(t, y_i, z) \mid i = 1, \dots, N\}$ for all z and $\tilde{y}$ in this neighborhood of y . Since the derivatives of k are polynomials in z and $\int |\Gamma(z)|^{5/4} k(t, y_i, z) \mathrm{d}z < \infty$ for all i by Lemma ??, we can differentiate under the integral by Hölder's inequality and dominated convergence to have that derivatives of $K(\cdot, \cdot)$ of all order exist and are equal to integrals of $\Gamma(z)$ multiplied by the corresponding derivatives of k . The domination can be chosen uniformly for (t, y) in compact subsets of $[0, 1) \times \mathbb{R}^n$: the Gaussian kernel estimate in Lemma ??, together with Hölder's inequality, also controls the polynomial factors in $\partial_t k$ and its spatial derivatives. Differentiation under the integral therefore gives a continuous time derivative on $[0, 1)$, including the right derivative at $t = 0$, and $\partial_t K = -\frac{1}{2} \mathrm{tr}(\Sigma_z \nabla_y^2 K)$. Thus, $K(t, y)$ has the smoothness claimed. An equivalent definition of $K(t, y)$ is

$$K(t, y) = \int \Gamma(y + z) k(t, 0, z) \mathrm{d}z.$$

We wish to show that we can pass differentiation with respect to y through the integral in this formulation, producing

$$\nabla K(t, y) = \int \nabla \Gamma(y + z) k(t, 0, z) \mathrm{d}z \equiv \mathbb{E}[\nabla \Gamma(y + Z_1 - Z_t)].$$

This holds if, for each y , there exists $\epsilon > 0$ such that the family of functions

$$\{z \mapsto \nabla \Gamma(y' + z) k(t, 0, z) \mid \|y' - y\| < \epsilon\}$$

is uniformly integrable. Noting that $k(t, 0, z) = k(t, y', y' + z)$ and making the change of variables $u = y' + z$, we see that this is equivalent to the family of functions

$$\{u \mapsto \nabla \Gamma(u) k(t, y', u) \mid \|y - y'\| < \epsilon\}$$

being uniformly integrable. By the bound above and (??), this family is dominated by the integrable function $C\|\nabla\Gamma(u)\|\max_{i=1,\dots,N}k(t,y_i,u)$, which gives the uniform integrability.

To prove the expression of the Hessian, we differentiate (??) under the integral sign in y . The integrand is $\nabla\Gamma(z)k(t,y,z)$, in which only k depends on y , with $\nabla_y k(t,y,z) = \frac{\Sigma_z^{-1}(z-y)}{1-t}k(t,y,z)$. For $\tilde{y}$ in the neighborhood of y considered above, the bound $k(t,\tilde{y},z) \leq C\max_i k(t,y_i,z)$ gives $\|\nabla\Gamma(z)\|\|\nabla_{\tilde{y}}k(t,\tilde{y},z)\| \leq C'\|\nabla\Gamma(z)\|(1+\|z\|)\max_i k(t,y_i,z)$, which is integrable by (??) and Hölder's inequality with exponents $(3/2, 3)$. Dominated convergence therefore justifies the differentiation and yields the expression for the Hessian. $\square$

We need the following lemma for the proof of Theorem ??.

Lemma A.4. *Let H and $\tilde{v}$ be as in Theorem ??, and let Y be the total order process associated with an admissible strategy satisfying (??). Then*

$$M_t := \int_0^t (\tilde{v} - H(s, Y_s))' dZ_s, \quad t \in [0, 1],$$

is a square-integrable martingale in the informed trader's filtration and satisfies $\mathbb{E}[M_1 \mid \tilde{v}] = 0$.

Proof. Since $(\tilde{u}, \tilde{v})$ is independent of Z , the process Z remains Brownian in the informed trader's filtration. Moreover,

$$\mathbb{E} \int_0^1 (\tilde{v} - H(s, Y_s))' \Sigma_z (\tilde{v} - H(s, Y_s)) ds \leq 2\|\Sigma_z\| \left(\mathbb{E}[|\tilde{v}|^2] + \mathbb{E} \int_0^1 |H(s, Y_s)|^2 ds \right) < \infty.$$

Thus M is a square-integrable martingale starting at zero. Since $\tilde{v}$ is measurable at time zero, $\mathbb{E}[M_1 \mid \tilde{v}] = 0$. $\square$

A.2 Proof of Lemma ??

This follows directly from Lemmas A.1 and A.2 by choosing $L = G$.

A.3 Proof of Theorem ??

As promised in Section ??, we prove here a more general version of Theorem ?? in which the informed trading strategy is allowed to be a continuous semimartingale satisfying (??). In other words, we do not impose that $dX_t = \theta_t dt$ for some θ . In this more general setting, the profit of the informed trader is, in analogy with (??),

$$\int_0^1 (\tilde{v} - P_t)' dX_t - \langle P, X \rangle_1. \quad (\text{A.4})$$

As remarked before, this follows from the Merton intertemporal budget constraint after integrating by parts, but here we can regard it simply as a definition.

As stated in Section ??, our chief contribution is to identify the gradient of the Brenier potential as the map Φ such that there is a random variable $\tilde{\zeta}$ with distribution G that identifies $\tilde{v}$ in the sense that $\Phi(\tilde{\zeta}) = \tilde{v}$. From this map, the proof of the theorem follows familiar lines, as first laid out by ? and then further developed by other authors, notably L. Campi, U. Çetin, and A. Danilova in a series of papers culminating in ?.

We will show that the pricing rule defined in Part (A) and the trading strategy defined in Part (C) constitute an equilibrium. Then, we will derive the other statements in the theorem. First, we will show that the trading strategy is optimal given the pricing rule.

Assume the pricing rule is $H = \nabla_y K$ and consider any trading strategy X of the informed trader. Because X is a continuous semimartingale, we have $X = A + L$ where A is a finite-variation and continuous process, and L is a continuous local martingale. Moreover, $dL_t = C_t dZ_t$ for some matrix-valued process C . Thus, $dY_t = dA_t + (C_t + I) dZ_t$. By Itô's formula, substituting $H = \nabla_y K$, we have

$$dK(t, Y_t) = \frac{\partial K(t, Y_t)}{\partial t} dt + H(t, Y_t)' dY_t + \frac{1}{2} \text{tr} \left((C_t + I) \Sigma_z (C_t + I)' \nabla_y^2 K(t, Y_t) \right) dt.$$

Because K is defined via the expectation (??) and is smooth thanks to Lemma ??, it solves the heat equation

$$\frac{\partial K}{\partial t} + \frac{1}{2} \text{tr} \left(\Sigma_z \nabla_y^2 K \right) = 0. \quad (\text{A.5})$$

Furthermore, a direct computation yields

$$d \sum_{i=1}^n \langle X^i, H^i(\cdot, Y) \rangle_t = \text{tr} (C_t \Sigma_z (I + C_t)' \nabla_y^2 K(t, Y_t)) dt.$$

Hence,

$$\begin{aligned} \frac{1}{2} \text{tr} \left((C + I) \Sigma_z (C + I)' \nabla_y^2 K \right) dt &= \text{tr} (C \Sigma_z \nabla_y^2 K) dt + \frac{1}{2} \text{tr} (C \Sigma_z C' \nabla_y^2 K) dt \\ &\quad + \frac{1}{2} \text{tr} (\Sigma_z \nabla_y^2 K) dt \\ &= d \sum_{i=1}^n \langle X^i, H^i(\cdot, Y) \rangle_t - \frac{1}{2} \text{tr} (C \Sigma_z C' \nabla_y^2 K) dt - \frac{\partial K}{\partial t} dt \end{aligned}$$

Making this substitution, we obtain

$$dK(t, Y_t) = H(t, Y_t)' dY_t + d \sum_{i=1}^n \langle X^i, H^i(\cdot, Y) \rangle_t - \frac{1}{2} \text{tr} (C_t \Sigma_z C_t' \nabla_y^2 K(t, Y_t)) dt.$$

Since K is $C^{1,2}$ on $[0, 1] \times \mathbb{R}^n$ by Lemma ??, we may integrate over $[0, 1 - \epsilon]$ for $\epsilon \in (0, 1)$. Letting $\epsilon \rightarrow 0$, using that $K(1 - \epsilon, Y_{1-\epsilon}) \rightarrow \Gamma(Y_1)$ by the continuity of K on $[0, 1] \times \mathbb{R}^n$ with $K(1, \cdot) = \Gamma$, rearranging, and using the facts that $K(0, 0) = 0$ and $H(t, Y_t) = P_t$ yields

$$\int_0^1 (\tilde{v} - P_t)' dY_t - \langle P, X \rangle_1 = \tilde{v}' Y_1 - \Gamma(Y_1) - \frac{1}{2} \text{tr} \left(\int_0^1 C_t \Sigma_z C_t' \nabla_y^2 K(t, Y_t) dt \right). \quad (\text{A.6})$$

By Lemma ?? and (??),

$$\mathbb{E} \left[\int_0^1 (\tilde{v} - H(t, Y_t))' dZ_t \middle| \tilde{v} \right] = 0.$$

Therefore, taking conditional expectations given $\tilde{v}$ in (??) yields

$$\begin{aligned} \mathbb{E} \left[\int_0^1 (\tilde{v} - P_t)' dX_t - \langle P, X \rangle_1 \middle| \tilde{v} \right] \\ = \mathbb{E} \left[\tilde{v}' Y_1 - \Gamma(Y_1) - \frac{1}{2} \operatorname{tr} \left(\int_0^1 C_t \Sigma_z C_t' \nabla_y^2 K(t, Y_t) dt \right) \middle| \tilde{v} \right]. \quad (\text{A.7}) \end{aligned}$$

The left-hand side is the definition of the informed trader's expected profit. On the right-hand side, the trace is nonnegative due to the convexity of Γ . Therefore, the expected profit of the informed trader is bounded above by $\mathbb{E}[\tilde{v}' Y_1 - \Gamma(Y_1) | \tilde{v}]$ and the bound is achieved by any finite-variation and continuous strategy such that, almost surely, $Y_1 \in \arg \max_{y \in \mathbb{R}^n} \tilde{v}' y - \Gamma(y)$, in particular by any such strategy with $Y_1 \in \mathcal{D}_\Gamma$ and $\nabla \Gamma(Y_1) = \tilde{v}$ almost surely. The strategy (??) implies

$$dY_t = \frac{\tilde{\zeta} - Y_t}{1-t} dt + dZ_t,$$

which implies that Y is a Brownian bridge ending at $Y_1 = \tilde{\zeta}$; hence $\nabla \Gamma(Y_1) = \tilde{v}$. The strategy achieves the maximum expected profit of

$$\mathbb{E}[\sup \{ \tilde{v}' y - \Gamma(y) \mid y \in \mathbb{R}^n \}] = \mathbb{E}[\Gamma^*(\tilde{v})]. \quad (\text{A.8})$$

To show that the strategy is optimal, it remains only to show that the restriction (??) is satisfied. First, we turn to the other equilibrium condition (??), taking the trading strategy (??) as given.

The strategy (??) implies that Y is a $(0, \Sigma_z)$ -Brownian bridge conditional on $\tilde{\zeta}$. Because $\tilde{\zeta}$ has the normal $(0, \Sigma_z)$ distribution, Y is a $(0, \Sigma_z)$ -Brownian motion on its own filtration $(\mathcal{F}_t^Y, 5.6.B)$. Thus, conditionally on $\mathcal{F}_t^Y$, $\tilde{\zeta} = Y_1$ has the normal $(Y_t, (1-t)\Sigma_z)$ distribution. Using this fact and (??), we obtain

$$H(t, Y_t) = \mathbb{E}[\nabla \Gamma(Y_1) | \mathcal{F}_t^Y] = \mathbb{E}[\tilde{v} | \mathcal{F}_t^Y].$$

This establishes the equilibrium condition (??).

The equilibrium condition (??) and the fact that the elements of $\tilde{v}$ have finite variances implies that $H(t, Y_t)$ is a square-integrable martingale on $\mathcal{F}^Y$. Therefore, the trading strategy (??) and pricing rule H satisfy (??). This completes the proof of optimality of the strategy (??) and completes the proof that both equilibrium conditions hold.

Equation (??) implies that the heat equation holds for each element H_i on $[0, 1] \times \mathbb{R}^n$. Combining this with Itô's formula and the fact that $H(t, y) = \nabla_y K(t, y)$, we obtain $dP_t = \Lambda_t dY_t$ for $t \in [0, 1]$, where $\Lambda_t = \nabla_y^2 K(t, Y_t)$. This verifies the last statement of Part (A). As noted above, Y is a Brownian bridge ending at $\tilde{\zeta}$ given the informed trader's information and is a Brownian motion on $\mathcal{F}^Y$. This verifies Part (E). Part (B) follows from the equilibrium condition (??) and the martingale convergence theorem: since Y is continuous, $\mathcal{F}_1^Y = \bigvee_{t < 1} \mathcal{F}_t^Y$ up to null sets, so $P_t = \mathbb{E}[\tilde{v} | \mathcal{F}_t^Y] \rightarrow \mathbb{E}[\tilde{v} | \mathcal{F}_1^Y] = \tilde{v}$ almost surely as $t \rightarrow 1$. To verify Part (D), note that, by following the same reasoning as before but starting the integration of dK at t , we obtain

$$\begin{aligned} \int_t^1 (\tilde{v} - H(t, y_t))' dY_t - \int_t^1 d \sum_{i=1}^n \langle X^i, H^i(\cdot, Y) \rangle_t \\ = K(t, Y_t) + \tilde{v}'(Y_1 - Y_t) - \Gamma(Y_1) - \frac{1}{2} \operatorname{tr} \left(\int_t^1 C_t \Sigma_z C_t' \nabla_y^2 K(t, Y_t) dt \right). \end{aligned}$$

Taking expectations as before, we see that the conditional expected profit is bounded above by

$$K(t, Y_t) - \tilde{v}' Y_t + \mathbb{E} [\tilde{v}' Y_1 - \Gamma(Y_1) \mid \mathcal{F}_t^Y] .$$

Substituting $\sup \{\tilde{v}' y - \Gamma(y) \mid y \in \mathbb{R}^n\} = \Gamma^*(\tilde{v})$ yields (??). The rest of Part (D) is shown in (??).

A.4 Proof of Corollary ??

When X has finite variation, (??) implies

$$\int_0^1 (\tilde{v} - P_t)' dY_t = \tilde{v}' Y_1 - \Gamma(Y_1) .$$

Substituting this into (??) yields the first formula for dealer profits. Optimization by the informed trader and the definition of the convex conjugate then yields the second formula.

Appendix B Theorem ??

We will prove the theorems in Section ?? in reverse order, beginning with Theorem ?. Throughout this appendix, we take F^* to be a distribution function on $\mathbb{R}^n$ satisfying (i) or (ii) of Theorem ?. Since both assumptions imply the finiteness of the second (in fact any polynomial) moments of F^* , we take Γ as the Brenier potential to transport G to F^* such that $\mathbb{E}^G[\Gamma(Z_1)] = 0$, and we define K by (??) where $k(t, y, \cdot)$ denotes the normal density function with mean y and covariance matrix $(1-t)\Sigma_z$. We define $\tilde{\chi}$ by (??) and ϕ, ψ by (??)-(??) and denote

$$h(t, z, b) = k(t, z, b) e^{-\phi(t, z) + \psi(t, z, b)} . \quad (\text{B.1})$$

We split the proof of Theorem ?? in two parts. In the first part, we prove the following lemmas that contain preliminary results.

Lemma B.1. *Under the assumptions of Theorem ?? the following results hold.*

- (1) $\tilde{\chi}^2$ has finite ν -mean, and ϕ and $\psi(\cdot, \cdot, b)$, for $b \in \mathcal{D}_\Gamma$, are finite and locally $C^{1,2}$ on $[0, 1) \times \mathbb{R}^n$, with time derivatives at $t = 0$ understood from the right. Moreover, ψ satisfies

$$\sup_{b \in \mathcal{D}_\Gamma} \sup_z \sup_{0 \leq s < 1} (|\nabla_z \psi(s, z, b)| + \sqrt{1-s} \|\nabla_z^2 \psi(s, z, b)\|) \leq C \quad (\text{B.2})$$

for some constant C .

- (2) $d\hat{\nu}/d\nu = \tilde{\chi}/\mathbb{E}^\nu[\tilde{\chi}]$ defines a probability measure $\hat{\nu}$ equivalent to ν so that $dZ_t - \Sigma_z \nabla \phi(t, Z_t) dt$ is a $\hat{\nu}$ -Brownian motion and

$$\begin{aligned} & \mathbb{E}^{\hat{\nu}} \left[\exp \left(-\alpha \beta' \nabla \Gamma(Z_1) + \alpha \Gamma^*(\nabla \Gamma(Z_1)) - \alpha \int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(s, Z_s)) ds \right) \mid \mathcal{F}_t^Z \right] \\ &= e^{-\phi(t, Z_t) - \alpha \int_0^t \text{tr}(\Sigma_z \nabla_z^2 K(u, Z_u)) du} . \end{aligned} \quad (\text{B.3})$$

- (3) For all $t < 1$, h is differentiable in z , $\sup_{|z| \leq R} \sup_{0 \leq s \leq t} \int_{\mathbb{R}^n} |\nabla_z h(s, z, b)| db < \infty$ for every $R < \infty$, and

$$dh(t, Z_t, b) = \nabla_z h(t, Z_t, b)' (dZ_t - \Sigma_z \nabla \phi(t, Z_t) dt) . \quad (\text{B.4})$$

Proof. We first prove the statement under assumption (i). Because the support of F^* is bounded and $\nabla\Gamma(Z_1)$ has distribution F^* , $\nabla\Gamma \in \text{supp } F^*$ Lebesgue-almost everywhere; since $\nabla\Gamma$ is continuous on the full-measure set $\mathcal{D}_\Gamma$ and $\text{supp } F^*$ is closed, $\nabla\Gamma$ is bounded on $\mathcal{D}_\Gamma$, hence on $\mathbb{R}^n$ by our convention. Using (??) and the convexity of $K(t, \cdot)$ we have

$$\begin{aligned} 0 \leq \text{tr}(\Sigma_z \nabla^2 K(t, z)) &= \int_{\mathbb{R}^n} \frac{(y-z)'}{1-t} \nabla\Gamma(y) k(t, z, y) dy \\ &= \mathbb{E}^\nu \left[\frac{(Z_1 - Z_t)'}{1-t} \nabla\Gamma(z + Z_1 - Z_t) \right]. \end{aligned}$$

Thus,

$$0 \leq \text{tr}(\Sigma_z \nabla_z^2 K(t, z)) \leq \mathbb{E}^\nu \left[\frac{|Z_1 - Z_t|}{1-t} |\nabla\Gamma(z + Z_1 - Z_t)| \right] \leq \frac{C}{\sqrt{1-t}} \quad (\text{B.5})$$

for some constant C depending only on the bound of the price. Thus, for every path of Z ,

$$\int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(t, Z_t)) dt \leq \int_0^1 \frac{C}{\sqrt{1-t}} dt =: C_0.$$

Additionally, the right hand side of (??) implies that $\tilde{\chi}$, defined in (??), satisfies

$$\begin{aligned} \tilde{\chi} &= \exp \left\{ \alpha (\beta' \nabla\Gamma(Z_1) + \Gamma(Z_1) - Z_1' \nabla\Gamma(Z_1)) + \alpha \text{tr} \left(\Sigma_z \int_0^1 \nabla_z^2 K(t, Z_t) dt \right) \right\} \\ &\leq \exp \left\{ \alpha (\beta' \nabla\Gamma(Z_1) + \Gamma(0)) + \alpha C_0 \right\} \end{aligned}$$

which is uniformly bounded since $\nabla\Gamma(Z_1)$ has distribution F^* which is assumed to have bounded support. Thus, $\tilde{\chi}$ is bounded.

For fixed t, z, b , using the distribution of the conditioned Brownian motion, we can write

$$\begin{aligned} &\mathbb{E}^\nu \left[\exp \left\{ \alpha \text{tr} \left(\Sigma_z \int_t^1 \nabla_z^2 K(u, Z_u) du \right) \right\} \middle| Z_t = z, Z_1 = b \right] \\ &= \mathbb{E}^\nu \left[\exp \left\{ \alpha \text{tr} \left(\Sigma_z \int_t^1 \nabla_z^2 K \left(u, \frac{1-u}{1-t} z + \frac{u-t}{1-t} b + \mathcal{Z}_u^{0,t,1} \right) du \right) \right\} \right] \end{aligned}$$

where $(\mathcal{Z}_u^{0,t,1})_{u \in [t,1]}$ is the Brownian motion that is conditioned to be 0 at times t and 1 and does not depend on z .

Since $\nabla\Gamma$ is bounded dominated convergence allows us to differentiate (??) under the integral sign in z , which yields

$$\nabla_z^j K(u, z) = \int_{\mathbb{R}^n} \nabla\Gamma(y) \otimes \nabla_z^{j-1} k(u, z, y) dy, \quad j = 3, 4.$$

and

$$\| \nabla_z^3 K(u, z) \| \leq \frac{C}{1-u}, \quad \| \nabla_z^4 K(u, z) \| \leq \frac{C}{(1-u)^{3/2}} \quad (\text{B.6})$$

for all $(u, z) \in [0, 1) \times \mathbb{R}^n$, where C depends only on n, Σ_z and $\sup_{y \in \mathbb{R}^n} \|\nabla\Gamma(y)\|$.

Set $c(u, x) := \alpha \operatorname{tr}(\Sigma_z \nabla_z^2 K(u, x))$ and, for fixed b ,

$$\ell_u(t, z) := \frac{1-u}{1-t}z + \frac{u-t}{1-t}b + \mathcal{Z}_u^{0,t,1}, \quad A(t, z) := \int_t^1 c(u, \ell_u(t, z)) du,$$

so that

$$\psi(t, z, b) = -\alpha(b - \beta)' \nabla \Gamma(b) + \alpha \Gamma(b) + \log \mathbb{E}^\nu[e^{A(t, z)}],$$

and, by (??), $0 \leq A(t, z) \leq \alpha C_0$. Let $J^{(i)}(u, x) := \partial_{x_i} c(u, x)$ and $J^{(i,j)}(u, x) := \partial_{x_i x_j}^2 c(u, x)$. By (??), after enlarging C ,

$$|J^{(i)}(u, x)| \leq \frac{C}{1-u} \quad \text{and} \quad |J^{(i,j)}(u, x)| \leq \frac{C}{(1-u)^{3/2}}.$$

Since $\partial_{z_i} \ell_u(t, z) = \frac{1-u}{1-t} e_i$, dominated convergence allows us to differentiate A under the du -integral, and we obtain, for every path of $\mathcal{Z}^{0,t,1}$,

$$\begin{aligned} \Psi^{(i)}(t, z) &:= \partial_{z_i} A(t, z) = \int_t^1 \frac{1-u}{1-t} J^{(i)}(u, \ell_u(t, z)) du, \\ \Psi^{(i,j)}(t, z) &:= \Psi^{(i)}(t, z) \Psi^{(j)}(t, z) + \int_t^1 \frac{(1-u)^2}{(1-t)^2} J^{(i,j)}(u, \ell_u(t, z)) du, \end{aligned}$$

so that $\partial_{z_i} e^{A(t, z)} = e^{A(t, z)} \Psi^{(i)}(t, z)$ and $\partial_{z_i z_j}^2 e^{A(t, z)} = e^{A(t, z)} \Psi^{(i,j)}(t, z)$. These functions are continuous in z and satisfy the bounds

$$|\Psi^{(i)}(t, z)| \leq C \int_t^1 \frac{1-u}{1-t} \frac{du}{1-u} = C, \tag{B.7}$$

$$|\Psi^{(i,j)}(t, z)| \leq C^2 + C \int_t^1 \frac{(1-u)^2}{(1-t)^2(1-u)^{3/2}} du = C^2 + \frac{2C}{3}(1-t)^{-1/2}, \tag{B.8}$$

which are uniform in z and in the path of $\mathcal{Z}^{0,t,1}$. For each fixed $t \in [0, 1)$ and every path of $\mathcal{Z}^{0,t,1}$, the random variables $e^{A(t, z)}$ and $e^{A(t, z)} \Psi^{(i)}(t, z)$ are continuously differentiable in z , with derivatives $e^{A(t, z)} \Psi^{(i)}(t, z)$ and $e^{A(t, z)} \Psi^{(i,j)}(t, z)$, which are bounded by the deterministic constants $e^{\alpha C_0} C$ and $e^{\alpha C_0} (C^2 + \frac{2C}{3}(1-t)^{-1/2})$ thanks to $0 \leq A(t, z) \leq \alpha C_0$, (??) and (??). Both random variables are therefore Lipschitz continuous in z , uniformly in the path, so that their difference quotients are bounded and bounded convergence allows us to differentiate $\mathbb{E}^\nu[e^{A(t, z)}]$ twice under the expectation. Hence ψ is twice continuously differentiable in z with

$$\begin{aligned} \frac{\partial \psi}{\partial z_i}(t, z, b) &= \frac{\mathbb{E}^\nu[e^{A(t, z)} \Psi^{(i)}(t, z)]}{\mathbb{E}^\nu[e^{A(t, z)}]}, \\ \frac{\partial^2 \psi}{\partial z_i \partial z_j}(t, z, b) &= \frac{\mathbb{E}^\nu[e^{A(t, z)} \Psi^{(i,j)}(t, z)]}{\mathbb{E}^\nu[e^{A(t, z)}]} - \frac{\partial \psi}{\partial z_i}(t, z, b) \frac{\partial \psi}{\partial z_j}(t, z, b). \end{aligned}$$

Since the first terms on the right-hand sides are averages of $\Psi^{(i)}(t, z)$ and $\Psi^{(i,j)}(t, z)$ under the probability measure with density $e^{A(t, z)}/\mathbb{E}^\nu[e^{A(t, z)}]$ with respect to ν , (??) and (??) give

$$\left| \frac{\partial \psi}{\partial z_i}(t, z, b) \right| \leq C, \quad \left| \frac{\partial^2 \psi}{\partial z_i \partial z_j}(t, z, b) \right| \leq 2C^2 + \frac{2C}{3}(1-t)^{-1/2}.$$

In particular, after enlarging C , (??) holds, and $\nabla_z \psi$ and $\nabla_z^2 \psi$ are bounded on $[0, 1 - \varepsilon] \times \mathbb{R}^n \times \mathcal{D}_\Gamma$ for all $\varepsilon \in (0, 1)$. Note that convexity gives

$$\Gamma(b) + (\beta - b)' \nabla \Gamma(b) \leq \Gamma(\beta) \quad \text{for all } b \in \mathcal{D}_\Gamma.$$

Together with the deterministic bound on the integrated trace term, this implies

$$e^{\psi(t,z,b)} \leq C \quad \text{for all } (t, z, b) \in [0, 1] \times \mathbb{R}^n \times \mathcal{D}_\Gamma. \quad (\text{B.9})$$

The identity

$$e^{\phi(t,z)} = \int_{\mathbb{R}^n} k(t, z, b) e^{\psi(t,z,b)} db,$$

the uniform bound (??), which suffices since $\mathbb{R}^n \setminus \mathcal{D}_\Gamma$ is Lebesgue-null, the bounds on $\nabla_z \psi$ and $\nabla_z^2 \psi$ established above, and the Gaussian bounds on the derivatives of k allow us to differentiate twice under the integral in z . Since $e^\phi > 0$, this gives the local C^2 regularity of ϕ in z . The corresponding backward equation gives its time regularity. The time regularity of ψ follows similarly, which concludes the proof of (1).

By definition of ϕ and the Markov property of Z , for $t \in [0, 1)$,

$$M_t := e^{\phi(t, Z_t) + \alpha \int_0^t \text{tr}(\Sigma_z \nabla_z^2 K(u, Z_u)) du} = \mathbb{E}^\nu[\tilde{\chi} | \mathcal{F}_t^Z]$$

is a positive, uniformly integrable ν -martingale, closed by $\tilde{\chi}$. Since ϕ is $C^{1,2}$ on $[0, 1) \times \mathbb{R}^n$, Itô's formula gives, for $t \in [0, 1)$,

$$dM_t = M_t \left(\nabla_z \phi(t, Z_t)' dZ_t + D(t, Z_t) dt \right), \quad D := \partial_t \phi + \frac{1}{2} \text{tr}(\Sigma_z \nabla_z^2 \phi) + \frac{1}{2} \nabla_z \phi' \Sigma_z \nabla_z \phi + \alpha \text{tr}(\Sigma_z \nabla_z^2 K).$$

As M is a martingale, the uniqueness of the decomposition of a continuous semimartingale implies $\int_0^t M_u D(u, Z_u) du = 0$ for all $t \in [0, 1)$ almost surely; since $M > 0$ and D is continuous, $D(u, Z_u) = 0$ for all $u \in [0, 1)$ almost surely. Because Z_u has a positive density on $\mathbb{R}^n$ for $u \in (0, 1)$ and D is continuous on $[0, 1) \times \mathbb{R}^n$, ϕ solves the PDE

$$\partial_t \phi + \frac{1}{2} \text{tr}(\Sigma_z \nabla_z^2 \phi) + \frac{1}{2} \nabla_z \phi' \Sigma_z \nabla_z \phi + \alpha \text{tr}(\Sigma_z \nabla_z^2 K) = 0 \quad \text{on } [0, 1) \times \mathbb{R}^n. \quad (\text{B.10})$$

In particular, $dM_t = M_t \nabla_z \phi(t, Z_t)' dZ_t$, and since $M_0 = e^{\phi(0,0)} = \mathbb{E}^\nu[\tilde{\chi}]$, we obtain for $t \in [0, 1)$

$$e^{\phi(t, Z_t) + \alpha \int_0^t \text{tr}(\Sigma_z \nabla_z^2 K(u, Z_u)) du} = e^{\phi(0,0) + \int_0^t \nabla_z \phi(u, Z_u)' dZ_u - \frac{1}{2} \int_0^t \nabla_z \phi(u, Z_u)' \Sigma_z \nabla_z \phi(u, Z_u) du}. \quad (\text{B.11})$$

Consequently, $d\hat{\nu}/d\nu = \tilde{\chi}/\mathbb{E}^\nu[\tilde{\chi}]$ defines a probability measure equivalent to ν whose density on $\mathcal{F}_t^Z$ is M_t/M_0 , which by (??) is the stochastic exponential of $\int_0^\cdot \nabla_z \phi(u, Z_u)' dZ_u$. This stochastic exponential is a true martingale on $[0, T]$ for every $T < 1$, so no integrability condition on $\nabla_z \phi(u, Z_u)$ (such as Novikov's) is needed, and Girsanov's theorem shows that $dZ_t - \Sigma_z \nabla \phi(t, Z_t) dt$ is a $\hat{\nu}$ -Brownian motion on $[0, 1)$.

Given that a Brownian motion has finite second moments, its time marginals are uniformly integrable. Thus, by martingale convergence theorem, this Brownian motion admits an almost sure

and L^1 limit which allows us to extend it to $[0, 1]$. To prove (??), we use (??) to have

$$\begin{aligned} & \mathbb{E}^{\hat{\nu}} \left[\exp \left(-\alpha \beta' \nabla \Gamma(Z_1) + \alpha \Gamma^*(\nabla \Gamma(Z_1)) - \alpha \int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(s, Z_s)) \, ds \right) | \mathcal{F}_t^Z \right] \\ &= \mathbb{E}^{\hat{\nu}} \left[\frac{1}{\tilde{\chi}} | \mathcal{F}_t^Z \right] = \frac{\mathbb{E}^{\nu} \left[\frac{d\hat{\nu}/d\nu}{\tilde{\chi}} | \mathcal{F}_t^Z \right]}{\mathbb{E}^{\nu} [d\hat{\nu}/d\nu | \mathcal{F}_t^Z]} = \frac{1}{e^{\phi(t, Z_t) + \alpha \int_0^t \text{tr}(\Sigma_z \nabla_z^2 K(u, Z_u)) \, du}} \end{aligned}$$

which concludes the proof of (2).

Let f be a continuous bounded function, for $t < 1$, we have

$$\begin{aligned} \int f(b) h(t, z, b) \, db &= \frac{\int f(b) k(t, z, b) e^{\psi(t, z, b)} \, db}{e^{\phi(t, z)}} \\ &= \frac{\mathbb{E}^{\nu} \left[f(Z_1) e^{-\alpha(Z_1 - \beta)' \nabla \Gamma(Z_1) + \alpha \Gamma(Z_1) + \alpha \text{tr}(\Sigma_z \int_t^1 \nabla_z^2 K(u, Z_u) \, du)} \mid Z_t = z \right]}{\mathbb{E}^{\nu} \left[e^{-\alpha(Z_1 - \beta)' \nabla \Gamma(Z_1) + \alpha \Gamma(Z_1) + \alpha \text{tr}(\Sigma_z \int_t^1 \nabla_z^2 K(u, Z_u) \, du)} \mid Z_t = z \right]} \\ &= \mathbb{E}^{\hat{\nu}} [f(Z_1) \mid Z_t = z]. \end{aligned}$$

Thus, $b \mapsto h(t, z, b)$ is the $\hat{\nu}$ conditional density of Z_1 . To prove the integrability of $\nabla_z h$, we use (??) and (??) to obtain

$$\begin{aligned} \int_{\mathbb{R}^n} |\nabla_z h(t, z, b)| \, db &\leq C e^{-\phi(t, z)} \left(\int_{\mathbb{R}^n} |\nabla_z k(t, z, b)| \, db + 1 \right) + |\nabla_z \phi(t, z)| \\ &\leq C e^{-\phi(t, z)} \left((1 - t)^{-1/2} + 1 \right) + |\nabla_z \phi(t, z)|. \end{aligned}$$

This bound is uniform on compact subsets of $[0, 1) \times \mathbb{R}^n$. The same estimates provide an integrable Gaussian domination locally in (t, z) , justifying differentiation under the integral in b .

Since $b \mapsto h(t, Z_t, b)$ is a conditional density under $\hat{\nu}$, $h(t, Z_t, b)$ is a $\hat{\nu}$ -local-martingale. Thus, (??) is a consequence of smoothness of h for $t < 1$, Ito's formula, and the martingality of $dZ_t - \Sigma_z \nabla \phi(t, Z_t) \, dt$.

We now prove the results under assumption (ii). The inequality $\nabla^2 U(v) \geq \kappa I_n$ implies that there exists a constant $C > 0$ so that

$$e^{-U(v)} \leq C e^{-\frac{\kappa \|v\|^2}{4}}.$$

Thus, F^* has finite second moments. The version of Caffarelli's contraction theorem in ?, Corollary 6.1 implies that $\nabla \Gamma$ is Lipschitz continuous with some constant L ; in particular, $\mathcal{D}_{\Gamma} = \mathbb{R}^n$. Since $\nabla_z K(t, z) = \int_{\mathbb{R}^n} \nabla \Gamma(y) k(t, z, y) \, dy$ is then Lipschitz in z with the same constant, $\|\nabla_z^2 K(t, z)\| \leq L$ on $[0, 1) \times \mathbb{R}^n$, so that (??) holds with $C/\sqrt{1-t}$ replaced by C , and $\int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(t, Z_t)) \, dt \leq C_0$ for a deterministic constant C_0 . The bound on $\tilde{\chi}$ is no longer uniform, but $\|\nabla \Gamma(z)\| \leq \|\nabla \Gamma(0)\| + L\|z\|$, so that $\tilde{\chi} \leq \exp\{\alpha(\beta' \nabla \Gamma(Z_1) + \Gamma(0)) + \alpha C_0\}$ has finite polynomial moments under ν because $\|Z_1\|$ has exponential moments of all orders; this gives the first claim of (1). In the derivation of (??), the boundedness of $\nabla \Gamma$ is replaced by its Lipschitz continuity: the linear growth of $\nabla \Gamma$ suffices to differentiate (??) under the integral sign, and since $\int_{\mathbb{R}^n} \nabla_z^{j-1} k(u, z, y) \, dy = 0$ for $j \geq 2$, we have

$$\nabla_z^j K(u, z) = \int_{\mathbb{R}^n} (\nabla \Gamma(y) - \nabla \Gamma(z)) \otimes \nabla_z^{j-1} k(u, z, y) \, dy \quad \text{and} \quad \|\nabla \Gamma(y) - \nabla \Gamma(z)\| \leq L\|y - z\|,$$

which again yields (??). The rest of the proof of (1) and the proofs of (2) and (3) are unchanged. $\square$

Lemma B.2. (1) Assume $\tilde{\chi}$ has a finite ν -mean and let F denote the $\hat{\nu}$ distribution of $\nabla\Gamma(Z_1)$, where $d\hat{\nu}/d\nu = \tilde{\chi}/\mathbb{E}^\nu[\tilde{\chi}]$. Then, there exists a function $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that, setting $\tilde{\zeta} = f(\tilde{u}, \tilde{v})$, where the distribution of $(\tilde{u}, \tilde{v})$ is $G \otimes F$, we have (i) $\tilde{\zeta} \in \mathcal{D}_\Gamma$ and $\nabla\Gamma(\tilde{\zeta}) = \tilde{v}$ almost surely, and (ii) the distribution of $\tilde{\zeta}$ is the $\hat{\nu}$ -distribution of Z_1 .

(2) The stochastic differential equation

$$dY_t = \Sigma_z \nabla_z \phi(t, Y_t) dt + \Sigma_z \nabla_y \log h(t, Y_t, \tilde{\zeta}) dt + dZ_t, \quad Y_0 = 0, \quad (\text{B.12})$$

admits a unique strong solution on $[0, 1]$, and this strong solution satisfies $Y_1 = \tilde{\zeta}$.

Proof. (1) follows directly from Lemma ?? by taking L to be the $\hat{\nu}$ -distribution of Z_1 and $g = \nabla\Gamma$, which transports L to F by definition of F . The function f constructed in Lemma ?? satisfies conditions (i) and (ii) by Lemma ??, the first part of (i) following from (ii) since $\hat{\nu} \sim \nu$ and $\mathbb{R}^n \setminus \mathcal{D}_\Gamma$ is Lebesgue-null.

For (2), by (??), the dynamics (??) is equivalent to

$$dY_t = \frac{1}{1-t}(\tilde{\zeta} - Y_t) dt + \Sigma_z \nabla_y \psi(t, Y_t, \tilde{\zeta}) dt + dZ_t.$$

For each fixed $a \in \mathcal{D}_\Gamma$ and $T < 1$, the drift $(a - y)/(1 - t) + \Sigma_z \nabla_y \psi(t, y, a)$ is globally Lipschitz in y and has linear growth, uniformly on $[0, T]$, by (??). Thus, conditional on $\tilde{\zeta} = a$, (??) admits a unique nonexplosive strong solution on $[0, T]$. Since $T < 1$ is arbitrary, this gives a unique strong solution on $[0, 1)$. To extend it to $[0, 1]$, we show that $Y_t \rightarrow \tilde{\zeta}$ as $t \uparrow 1$. We have $d\frac{\tilde{\zeta} - Y_t}{1-t} = -\frac{\Sigma_z \nabla_y \psi(t, Y_t, \tilde{\zeta})}{1-t} dt - \frac{dZ_t}{1-t}$, and thanks to (??) and $Y_0 = 0$, for all $t \in [0, 1)$,

$$\left| \frac{\tilde{\zeta} - Y_t}{1-t} - \tilde{\zeta} + \int_0^t \frac{dZ_s}{1-s} \right| \leq C \int_0^t \frac{1}{1-s} ds = -C \ln(1-t).$$

Since $t \mapsto (1-t) \int_0^t \frac{dZ_s}{1-s}$ is a Brownian bridge and converges to 0 as $t \rightarrow 1$, and $(1-t) \ln(1-t) \rightarrow 0$ as $t \rightarrow 1$, the solution of (??) satisfies $Y_1 := \lim_{t \rightarrow 1} Y_t = \tilde{\zeta}$. $\square$

We are now ready to prove our main result.

Proof of Theorem ??. Let F denote the $\hat{\nu}$ -distribution of $\nabla\Gamma(Z_1)$ given in Lemma ?? and define on the probability space $\Omega := \mathbb{R}^n \times \mathbb{R}^n \times C[0, 1]^n$ the physical distribution $\mathbb{P}$ as the product measure $G \otimes F \otimes \nu$ for $(\tilde{u}, \tilde{v}, Z)$. Assume the informed trader is constrained to choose a strategy satisfying (??) where $H(t, y) := \nabla_y K(t, y)$.

Step 1: The physical second moment of $\tilde{v}$. From the definitions, we have

$$\mathbb{E}^F[|\tilde{v}|^2] = \mathbb{E}^{\hat{\nu}}[|\nabla\Gamma(Z_1)|^2] = \frac{\mathbb{E}^\nu[\tilde{\chi} |\nabla\Gamma(Z_1)|^2]}{\mathbb{E}^\nu[\tilde{\chi}]},$$

which is finite because $\tilde{\chi}$ is ν -square integrable by Lemma ??, and because $\nabla\Gamma$ transports G to F^* which has finite polynomial moments by assumption.

Step 2: Optimality of (??). Take the pricing rule to be $H = \nabla_y K$. Because K solves the heat equation with final condition Γ , proceeding similarly as in the proof of Theorem ??, we obtain (??) and (??). The second moment established in Step 1 and (??) allow us to apply the argument of

Lemma ?? under $\mathbb{P}$, which yields (??), so the maximum expected profit conditional on $\tilde{v}$ is $\Gamma^*(\tilde{v})$ and a sufficient condition for optimality of a continuous finite-variation strategy is that $Y_1 \in \mathcal{D}_\Gamma$ and $\nabla\Gamma(Y_1) = \tilde{v}$ a.s. The informational efficiency of the equilibrium price, Part (B), is established in Step 5 below.

In the remainder of the proof, we assume that the informed trader uses the strategy (??) and market makers use the pricing rule $\nabla_y K(t, y)$. A combination of (1)-(2) in Lemma ?? leads to $Y_1 = \tilde{\zeta} \in \mathcal{D}_\Gamma$ and $\nabla\Gamma(Y_1) = \tilde{v}$ a.s. and as in Corollary ??, the condition $\nabla\Gamma(Y_1) = \tilde{v}$ implies that market maker wealth is

$$\tilde{w} = \beta' \tilde{v} - \Gamma^*(\tilde{v}) + \langle P, Y \rangle_1 = \beta' \nabla\Gamma(Y_1) - \Gamma^*(\nabla\Gamma(Y_1)) + \int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(s, Y_s)) ds. \quad (\text{B.13})$$

Thus, to establish the optimality of (??), we need only establish that the restriction (??) holds. We return to this condition below.

Step 3: Filtering problem. If the informed trader uses the strategy (??), then

$$dY_t = \Sigma_z \nabla_z \phi(t, Y_t) dt + \Sigma_z \nabla_y \log h(t, Y_t, \tilde{\zeta}) dt + dZ_t$$

which is equivalent to

$$dY_t = \frac{1}{1-t}(\tilde{\zeta} - Y_t) dt + \Sigma_z \nabla_y \psi(t, Y_t, \tilde{\zeta}) dt + dZ_t.$$

By (??), for each fixed $a \in \mathcal{D}_\Gamma$, the combined drift is globally Lipschitz in y on $[0, 1 - \epsilon]$, with a Lipschitz constant independent of a , and satisfies

$$\left| \frac{a - y}{1 - t} + \Sigma_z \nabla_y \psi(t, y, a) \right| \leq C_\epsilon(1 + |a| + |y|).$$

We now consider the problem of filtering the $\mathbb{P}$ -conditional distribution of $\tilde{\zeta}$ conditional on Y on $[0, 1)$. The Kushner-Stratonovich equation for this filtering problem is

$$\begin{aligned} \frac{d\hat{p}(t, a)}{\hat{p}(t, a)} &= \left(\nabla_z \log h(t, Y_t, a) - \int \nabla_z \log h(t, Y_t, \tilde{a}) \hat{p}(t, \tilde{a}) d\tilde{a} \right)' \\ &\times \left(dY_t - \Sigma_z \nabla_z \phi(t, Y_t) dt - \Sigma_z \int \nabla_z \log h(t, Y_t, \tilde{a}) \hat{p}(t, \tilde{a}) d\tilde{a} dt \right) \end{aligned}$$

for $t \in [0, 1)$. Thus, thanks to (?, Theorem 3.3 and Corollary 2.5), the distribution of $\tilde{\zeta}$ conditional on Y on $[0, 1)$ admits a unique solution in the set of probability measure valued process satisfying

$$\int_0^{1-\varepsilon} \left| \int \nabla_y \log h(t, Y_t, a) \hat{p}(t, a) da \right|^2 dt < \infty \text{ for all } \varepsilon \in (0, 1).$$

Note that $\int_0^{1-\varepsilon} \left| \int \nabla_y \log h(t, Y_t, a) h(t, Y_t, a) da \right|^2 dt = \int_0^{1-\varepsilon} \left| \int \nabla_y h(t, Y_t, a) da \right|^2 dt < \infty$ thanks to Lemma ??. Due to the choice of $\tilde{\zeta}$, the initial condition of this equation is $\hat{p}(0, a) = h(0, 0, a)$. Note that (??) holds for any Ito process whose quadratic variation is Σ_z , which is the case for Y . Thus, we have

$$dh(t, Y_t, a) = \nabla_z h(t, Y_t, a)' [dY_t - \Sigma_z \nabla_z \phi(t, Y_t) dt].$$

Additionally, by the integrability of $\nabla_z h$ established in (3) of Lemma ??, we have

$$0 = \nabla_z 1 = \nabla_z \int h(t, z, a) da = \int \nabla_z h(t, z, a) da = \int \nabla_z (\log h)(t, z, a) h(t, z, a) da.$$

Injecting these identities to the Kushner-Stratonovich equation, we can now conclude that $h(t, Y_t, a)$ is the unique solution to this equation on $[0, 1)$ and is therefore the density of $\tilde{\zeta}$ conditional on $\mathcal{F}_t^Y$. Finally, integrating against this density, we conclude that on its own filtration, Y satisfies

$$\begin{aligned} dY_t &= \left(\int_{\mathbb{R}^n} \Sigma_z [\nabla_z (\log h)(t, Y_t, a) + \nabla_z \phi(t, Y_t)] h(t, Y_t, a) da \right) dt + d\hat{Y}_t \\ &= \Sigma_z \nabla_z \phi(t, Y_t) dt + d\hat{Y}_t \end{aligned} \quad (\text{B.14})$$

where $\hat{Y}_t$ is a $(0, \Sigma_z)$ Brownian motion on $(\mathbb{P}, \mathcal{F}^Y)$. The drift $\Sigma_z \nabla_z \phi$ is locally Lipschitz in space. Uniqueness in law follows by localization, and the equality $Y_1 = \tilde{\zeta}$ established in Lemma ?? supplies a solution on $[0, 1]$. Consequently, the $(\mathbb{P}, \mathcal{F}^Y)$ -distribution of Y is $\hat{\nu}$, which proves (??).

Step 4: Defining the risk neutral probability measure. Set

$$M_t = \exp \left(-\frac{1}{2} \int_0^t \nabla_z \phi(s, Y_s)' \Sigma_z \nabla_z \phi(s, Y_s) ds - \int_0^t \nabla_z \phi(s, Y_s)' d\hat{Y}_s \right)$$

where $\hat{Y}$ is the $(\mathbb{P}, \mathcal{F}^Y)$ Brownian motion with covariance matrix Σ_z defined by the equality (??). M is clearly a $(\mathbb{P}, \mathcal{F}^Y)$ -local martingale. Applying Itô's formula to the $C^{1,2}$ function ϕ along Y , whose quadratic variation is $\Sigma_z dt$ by (??), and using the PDE (??), we obtain

$$\begin{aligned} M_t &= \exp \left(-\frac{1}{2} \int_0^t \nabla_z \phi(s, Y_s)' \Sigma_z \nabla_z \phi(s, Y_s) ds - \int_0^t \nabla_z \phi(s, Y_s)' d\hat{Y}_s \right) \\ &= \exp \left(\frac{1}{2} \int_0^t \nabla_z \phi(s, Y_s)' \Sigma_z \nabla_z \phi(s, Y_s) ds - \int_0^t \nabla_z \phi(s, Y_s)' dY_s \right) \\ &= \exp \left(\phi(0, 0) - \phi(t, Y_t) - \alpha \int_0^t \text{tr}(\Sigma_z \nabla_z^2 K(s, Y_s)) ds \right). \end{aligned}$$

Recall that $\nabla \Gamma$ is continuous on $\mathcal{D}_\Gamma$. Fix $y \in \mathcal{D}_\Gamma$ and write $\xi := z + Z_1 - Z_t$, so that

$$\phi(t, z) = \log \mathbb{E}^\nu \left[e^{-\alpha(\xi - \beta)' \nabla \Gamma(\xi) + \alpha \Gamma(\xi) + \alpha \text{tr}(\Sigma_z \int_t^1 \nabla_z^2 K(u, z + Z_u - Z_t) du)} \right].$$

As $(t, z) \rightarrow (1, y)$, $\xi \rightarrow y$ ν -a.s., and $\xi \in \mathcal{D}_\Gamma$ ν -a.s. because its law is Gaussian; hence $\nabla \Gamma(\xi) \rightarrow \nabla \Gamma(y)$ and $\Gamma(\xi) \rightarrow \Gamma(y)$ ν -a.s., while the trace term is bounded by $\alpha C \sqrt{1-t} \rightarrow 0$ by (??). By convexity, $\Gamma(\xi) + (\beta - \xi)' \nabla \Gamma(\xi) \leq \Gamma(\beta)$, so the integrand is bounded by the constant $e^{\alpha \Gamma(\beta) + \alpha C}$, as in (??). Dominated convergence therefore gives

$$\phi(t, z) \rightarrow -\alpha(y - \beta)' \nabla \Gamma(y) + \alpha \Gamma(y) = \alpha \beta' \nabla \Gamma(y) - \alpha \Gamma^*(\nabla \Gamma(y)),$$

the last equality being the Fenchel identity at the differentiability point y . Finally, $Y_1 = \tilde{\zeta}$ has an absolutely continuous distribution by Lemma ??(1), so $Y_1 \in \mathcal{D}_\Gamma$ almost surely, and $Y_t \rightarrow Y_1$ by Lemma ??(2). Hence $\phi(t, Y_t) \rightarrow \alpha \beta' \nabla \Gamma(Y_1) - \alpha \Gamma^*(\nabla \Gamma(Y_1))$ almost surely, and we can extend M by

continuity to

$$\begin{aligned} M_1 &= \exp \left(\phi(0,0) - \alpha \beta' \nabla \Gamma(Y_1) + \alpha \Gamma^*(\nabla \Gamma(Y_1)) - \alpha \int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(s, Y_s)) \, ds \right) \\ &= \exp(\phi(0,0) - \alpha \tilde{w}). \end{aligned}$$

where we used (??) to obtain the last equality. The pricing rule is $\nabla_y K(t, Y_t)$. We want to show that M is a martingale and

$$M_1 = \frac{e^{-\alpha \tilde{w}}}{\mathbb{E}[e^{-\alpha \tilde{w}}]}. \quad (\text{B.15})$$

Since the $(\mathbb{P}, \mathcal{F}^Y)$ distribution of Y is $\hat{\nu}$, (??) implies that

$$\begin{aligned} \mathbb{E}[M_1 | \mathcal{F}_t^Y] &= e^{\phi(0,0)} \mathbb{E} \left[\exp \left(-\alpha \beta' \nabla \Gamma(Y_1) + \alpha \Gamma^*(\nabla \Gamma(Y_1)) - \alpha \int_0^1 \text{tr}(\Sigma_z \nabla_z^2 K(s, Y_s)) \, ds \right) | \mathcal{F}_t^Y \right] \\ &= e^{\phi(0,0)} e^{-\phi(t, Y_t) - \alpha \int_0^t \text{tr}(\Sigma_z \nabla_z^2 K(u, Y_u)) \, du} = M_t. \end{aligned}$$

Thus, M is a martingale and we can define $\mathbb{Q}$ by $d\mathbb{Q}/d\mathbb{P} = M_1$ as in (??). Thanks to the martingality of M , (??) and Girsanov's theorem the $(\mathbb{Q}, \mathcal{F}^Y)$ -distribution of Y is a $(0, \Sigma_z)$ -Brownian motion, which is the ν -distribution of Z .

Step 5: $\mathbb{Q}$ dynamics of the price process and equilibrium. We have (i) $\nabla \Gamma$ transports the normal $(0, \Sigma_z)$ distribution to F^* , (ii) Y is a $(0, \Sigma_z)$ Brownian motion on $(\mathbb{Q}, \mathcal{F}^Y)$, (iii) $\nabla \Gamma(Y_1) = \tilde{v}$, and (iv) $\nabla_y K(t, y)$ solves the heat equation. Facts (i)–(iii) imply that the $\mathbb{Q}$ -distribution of $\tilde{v}$ is F^* , as claimed. Facts (ii)–(iv) imply that the pricing rule $H = \nabla_y K$ and trading strategy (??) satisfy the equilibrium condition (??), as in the proof of Theorem ?? . By martingale convergence, $P_t = \mathbb{E}^\mathbb{Q}[\tilde{v} | \mathcal{F}_t^Y] \rightarrow \tilde{v}$ $\mathbb{Q}$ -a.s. as $t \rightarrow 1$, hence $\mathbb{P}$ -a.s. since $\mathbb{P} \sim \mathbb{Q}$; this proves Part (B). It remains only to complete the proof that (??) is an optimal strategy. Thanks to Step 2, $Y_1 \in \mathcal{D}_\Gamma$ and $\nabla \Gamma(Y_1) = \tilde{v}$, we only need to show that the strategy satisfies the restriction (??). Since $(\mathbb{Q}, \mathcal{F}^Y)$ distribution of Y is the ν distribution of Z , we have

$$\begin{aligned} \mathbb{E}^\mathbb{P} \left[\int_0^1 \|H(s, Y_s)\|^2 \, ds \right] &= \mathbb{E}^\mathbb{Q} \left[M_1^{-1} \int_0^1 \|H(s, Y_s)\|^2 \, ds \right] = e^{-\phi(0,0)} \mathbb{E}^\nu \left[\tilde{\chi} \int_0^1 \|H(s, Z_s)\|^2 \, ds \right] \\ &\leq e^{-\phi(0,0)} \mathbb{E}^\nu [\tilde{\chi}^2]^{1/2} \mathbb{E}^\nu \left[\left(\int_0^1 \|H(s, Z_s)\|^2 \, ds \right)^2 \right]^{1/2} \end{aligned}$$

and the factor $\mathbb{E}^\nu [\tilde{\chi}^2]$ is finite by Lemma ?? . We also have

$$\begin{aligned} \mathbb{E}^\nu \left[\left(\int_0^1 \|H(s, Z_s)\|^2 \, ds \right)^2 \right] &= \mathbb{E}^\mathbb{Q} \left[\left(\int_0^1 \|H(s, Y_s)\|^2 \, ds \right)^2 \right] = \mathbb{E}^\mathbb{Q} \left[\left(\int_0^1 \|\mathbb{E}^\mathbb{Q}[\tilde{v} | \mathcal{F}_s^Y]\|^2 \, ds \right)^2 \right] \\ &\leq \int_0^1 \mathbb{E}^\mathbb{Q} [\|\mathbb{E}^\mathbb{Q}[\tilde{v} | \mathcal{F}_s^Y]\|^4] \, ds \leq \mathbb{E}^\mathbb{Q}[\|\tilde{v}\|^4] < \infty \end{aligned}$$

where we successively used the Cauchy-Schwartz inequality, Jensen inequality, and the fact that F^* has finite polynomial moments with both assumptions (i) and (ii) to obtain the finiteness. $\square$

Appendix C Theorems ?? and ??

Since ?? is a special case of ??, we prove ?? first.

C.1 Proof of Theorem ??

Take F to be the normal (μ, Σ_v) distribution for $\tilde{v}$ and take F^* to be the normal (μ^*, Σ_v^*) distribution, where μ^* is defined in Part (A) and Σ_v^* is defined in Part (C). We will show that (F, F^*) is admissible in the sense of Definition ??.

A direct calculation confirms that the map $y \mapsto \mu^* + \Lambda y$ transports G to F^* , where $\Lambda = \Sigma_z^{-1/2} V D^{1/2} V' \Sigma_z^{-1/2}$ as in Part (B). The vector $\mu^* + \Lambda y$ is the gradient of the convex function $\Gamma(y) := a + (1/2)y' \Lambda y + y' \mu^*$, which is the Brenier potential. The constant a is determined by the condition that $\mathbb{E}^G[\Gamma(Z_1)] = 0$. To compute a , observe that

$$\mathbb{E}^G \left[\frac{1}{2} \tilde{z}' \Lambda \tilde{z} + \tilde{z}' \mu^* \right] = \frac{1}{2} \text{tr}(\Sigma_z \Lambda) = \frac{1}{2} \text{tr}(D^{1/2}).$$

Thus,

$$\Gamma(y) = \frac{1}{2} y' \Lambda y + y' \mu^* - \frac{1}{2} \text{tr}(D^{1/2}). \quad (\text{C.1})$$

The definition (??) of K now gives us

$$K(t, y) = \frac{1}{2} y' \Lambda y + y' \mu^* - \frac{t}{2} \text{tr}(D^{1/2})$$

Set

$$\pi(b) = \alpha \beta' \Lambda b - \frac{\alpha}{2} b' \Lambda b$$

Then, $\tilde{\chi}$ in Definition ?? is

$$\tilde{\chi} = e^{\alpha \beta' \mu^* + \frac{\alpha}{2} \text{tr}(D^{1/2}) + \pi(Z_1)}.$$

Furthermore, ψ defined in (??) is

$$\psi(t, z, b) = \alpha \beta' \mu^* + \alpha \left(\frac{1}{2} - t \right) \text{tr}(D^{1/2}) + \pi(b),$$

which is independent of z , and ϕ defined in (??) satisfies

$$e^{\phi(t, z)} = e^{\alpha(\frac{1}{2}-t) \text{tr}(D^{1/2}) + \alpha \beta' \mu^*} \mathbb{E}^\nu \left[e^{\pi(Z_1)} \middle| Z_t = z \right].$$

It follows that

$$\tilde{\xi} := \frac{\tilde{\chi}}{\mathbb{E}^\nu[\tilde{\chi}]} = \frac{e^{\pi(Z_1)}}{\mathbb{E}^\nu[e^{\pi(Z_1)}]}$$

and

$$\xi_t := \mathbb{E}^\nu[\tilde{\xi} | Z_t] = \frac{\mathbb{E}^\nu[e^{\pi(Z_1)} | Z_t]}{\mathbb{E}^\nu[e^{\pi(Z_1)}]} = e^{-\alpha(\frac{1}{2}-t) \text{tr}(D^{1/2}) - \alpha \beta' \mu^* + \phi(t, Z_t)} / \mathbb{E}^\nu[e^{\pi(Z_1)}].$$

Using this fact, Itô's formula, and the fact that ξ is a ν -martingale, we have

$$\frac{d\xi_t}{\xi_t} = \nabla_z \phi(t, Z_t)' dZ_t$$

for the $(0, \Sigma_z)$ Brownian motion Z under ν . By Girsanov's theorem, $\hat{Z}$ defined by

$$dZ_t = \Sigma_z \nabla_z \phi(t, Z_t) dt + d\hat{Z}_t \quad (\text{C.2})$$

is a $(0, \Sigma_z)$ Brownian motion under $\hat{\nu}$, where $d\hat{\nu}/d\nu = \tilde{\chi}/E^\nu[\tilde{\chi}]$ as in Definition ???. We will use (??) to verify that F is the $\hat{\nu}$ distribution of $\nabla\Gamma(Z_1)$.

It is convenient to make the change of variables $U_t = V' \Sigma_z^{-1/2} Z_t$, so U is a vector of independent standard Brownian motions under ν . Also, set $\hat{U} = V' \Sigma_z^{-1/2} \hat{Z}_t$ so $\hat{U}$ is a vector of independent standard Brownian motions under $\hat{\nu}$. From (??), we have

$$dU_t = V' \Sigma_z^{1/2} \nabla_z \phi(t, Z_t) dt + d\hat{U}_t. \quad (\text{C.3})$$

We will show in a moment that

$$V' \Sigma_z^{1/2} \nabla_z \phi(t, Z_t) = -C_t V' \Sigma_z^{-1/2} (Z_t - \beta) = -C_t (U_t - V' \Sigma_z^{-1/2} \beta). \quad (\text{C.4})$$

Because C_t is diagonal, this implies that the system (??) is a system of n independent linear stochastic differential equations in the processes U_i . Let b denote the vector $V' \Sigma_z^{-1/2} \beta$. Then, given (??), the system (??) is equivalent to: for $i = 1, \dots, n$,

$$dU_{it} = -c_{it}(U_{it} - b_i) dt + d\hat{U}_{it},$$

which has solution

$$U_{it} = \frac{\alpha t \sqrt{d_i}}{1 + \alpha \sqrt{d_i}} b_i + \frac{1 + \alpha(1-t)\sqrt{d_i}}{1 + \alpha \sqrt{d_i}} \int_0^t \frac{1 + \alpha \sqrt{d_i}}{1 + \alpha(1-s)\sqrt{d_i}} d\hat{U}_{is}.$$

It follows that U_{it} for $t = 1$ is unconditionally normally distributed under $\hat{\nu}$ with mean equal to

$$\frac{\alpha \sqrt{d_i}}{1 + \alpha \sqrt{d_i}} b_i$$

and variance equal to

$$\int_0^1 \left(\frac{1}{1 + \alpha(1-s)\sqrt{d_i}} \right)^2 ds = \frac{1}{1 + \alpha \sqrt{d_i}}.$$

Therefore, the vector $U_1 = (U_{11}, \dots, U_{n1})'$ is unconditionally normally distributed under $\hat{\nu}$ with mean

$$\alpha D^{1/2} (I + \alpha D^{1/2})^{-1} V' \Sigma_z^{-1/2} \beta$$

and covariance matrix equal to

$$(I + \alpha D^{1/2})^{-1}.$$

We have

$$\nabla\Gamma(Z_1) = \mu^* + \Lambda Z_1 = \mu^* + \Sigma_z^{-1/2} V D^{1/2} U_1,$$

so $\nabla\Gamma(Z_1)$ is unconditionally normally distributed under $\hat{\nu}$ with mean equal to

$$\mu^* + \alpha \Sigma_z^{-1/2} V D (I + \alpha D^{1/2})^{-1} V' \Sigma_z^{-1/2} \beta = \mu^* + \alpha \Sigma_v \beta$$

and covariance matrix equal to

$$\Sigma_z^{-1/2} V D^{1/2} (I + \alpha D^{1/2})^{-1} D^{1/2} V' \Sigma_z^{-1/2} = \Sigma_v.$$

For these calculations, we use the facts that $D(I + \alpha D^{1/2})^{-1} = E$ and $VEV' = \Sigma_z^{1/2} \Sigma_v \Sigma_z^{1/2}$. Hence, we have shown that the $\hat{\nu}$ distribution of $\nabla \Gamma(Z_1)$ is F , implying that (F, F^*) is admissible, subject to verifying (??).

To verify (??), note that

$$\mathbb{E}^\nu[e^{\pi(Z_1)} | Z_t = z] = \mathbb{E}^\nu \left[\exp \left(\alpha \beta' \Sigma_z^{-1/2} V D^{1/2} U_1 - \frac{\alpha}{2} U_1' D^{1/2} U_1 \right) \middle| U_t = V' \Sigma_z^{-1/2} z \right]$$

Because U_1 is normal $(U_t, (1-t)I)$ under ν conditional on U_t , this implies that

$$\phi(t, z) = A_t - \frac{1}{2} (z - \beta)' \Sigma_z^{-1/2} V C_t V' \Sigma_z^{-1/2} (z - \beta),$$

for some deterministic A_t . Differentiating this expression gives (??). This completes the proof that (F, F^*) is admissible.

The normal distribution F^* with nonsingular covariance matrix satisfies condition (ii) of Theorem ??, so the conclusions of Theorem ?? all hold here. It remains to show that the conclusions of Theorem ?? imply conditions (B) and (D)–(H).

Part (B) follows from Part (A) of Theorem ?? and the fact that $\nabla_y^2 K(t, y) = \Lambda$. Part (D) is Part (B) of Theorem ???. Part (E) follows from Part (C) of Theorem ?? and the facts that $(\nabla \Gamma)^{-1}(v) = \Lambda^{-1}(v - \mu^*)$ and $\nabla_z \psi(t, z, b) = 0$. Part (H) follows from Part (E) of Theorem ??, using (??), which gives us

$$\Sigma_z \nabla_y \phi(t, y) = -\Sigma_z^{1/2} V C_t V' \Sigma_z^{-1/2} (y - \beta) = \Delta_t (\beta - y).$$

Part (D) of Theorem ?? and (??) show that the informed trader's expected profit conditional on $\tilde{v}$ is

$$\Gamma^*(\tilde{v}) = \frac{1}{2} \text{tr}(D^{1/2}) + \sup_y y' (\tilde{v} - \mu^*) - \frac{1}{2} y' \Lambda y = \frac{1}{2} \text{tr}(D^{1/2}) + \frac{1}{2} (\tilde{v} - \mu^*)' \Lambda^{-1} (\tilde{v} - \mu^*).$$

This implies the first part of Part (F) after substituting $\text{tr}(D^{1/2}) = \text{tr}(\Sigma_z \Lambda)$. Noise trader losses are $\langle P, Z \rangle_1 = \text{tr}(\Sigma_z \Lambda)$. It remains to calculate the unconditional expected profit of the informed trader. The unconditional expected profit is

$$\frac{1}{2} \text{tr}(D^{1/2}) + \frac{1}{2} (\mu - \mu^*)' \Lambda^{-1} (\mu - \mu^*) + \frac{1}{2} \text{tr}(\Lambda^{-1} \Sigma_v).$$

Direct substitution shows that

$$(\mu - \mu^*)' \Lambda^{-1} (\mu - \mu^*) = \alpha^2 \beta' \Sigma_v \Lambda^{-1} \Sigma_v \beta,$$

and we can calculate that

$$\frac{1}{2} \text{tr}(D^{1/2}) + \frac{1}{2} \text{tr}(\Lambda^{-1} \Sigma_v) = \text{tr}(\Sigma_z \Lambda) - \frac{\alpha}{2} \text{tr}(\Sigma_z \Sigma_v)$$

using the fact that $D^{1/2} - D^{-1/2} E = \alpha E$. This completes the proof of Part (F).

From Corollary ?? and calculations above, the realized wealth of market makers is

$$\beta' \tilde{v} + \langle P, Z \rangle_1 - \Gamma^*(\tilde{v}) = \beta' \tilde{v} + \text{tr}(\Sigma_z \Lambda) - \frac{1}{2} \text{tr}(\Sigma_z \Lambda) - \frac{1}{2} (\tilde{v} - \mu^*)' \Lambda^{-1} (\tilde{v} - \mu^*).$$

This implies Part (G).

C.2 Proof of Theorem ??

This follows directly from Theorem ?? by taking $n = 1$.

Appendix D Numerical Methods

To solve the risk-neutral model, we need to compute the transport map $\nabla \Gamma$ and the pricing rule $H(t, y) = \nabla_y K(t, y)$, which is the conditional expectation of $\nabla \Gamma(Z_1)$. In the risk-averse case, we also need to compute the vector of prices of risk $\nabla_y \phi(t, y)$. Due to the fact that the underlying space for order imbalances is two dimensional, our methods are based on discretization of this space. Higher dimensional problems would have required more sophisticated tools.

D.1 Transport Map and Potential

The transport map can be written explicitly for any pair of one-dimensional distributions and in the Gaussian-to-Gaussian case in any dimension. The map may occasionally be constructed geometrically in special cases, such as in ?. Thankfully, where an explicit optimal transport map is not known, numerical estimation is possible. Many algorithms exist, and efficient implementations are widely available. As observed in ?, with options we expect discontinuities of the transport map $\nabla \Gamma(y)$. After exploring alternative methods including the Sinkhorn algorithm (?) and input convex neural networks (?), we chose to use classical linear programming approach for the computation of the transport map, because this method handles the discontinuities of the transport map better than alternatives. This methodology is based on generating a grid of size $N \times N$ (for $N \sim 151$) in the order imbalance space and a grid of size M (with $M \sim 1001$) for the price of the underlying so that denoting $y_{i,j} \in \mathbb{R}^2$ and $v_k \in \mathbb{R}$ the grid points, for appropriate choice of a and b , the measures

$$\sum_{i=1, j=1}^N a_{i,j} \delta_{y_{i,j}}, \quad \sum_{k=1}^M b_k \delta_{(v_k, (v_k - K)^+)}$$

approximate the two dimensional distributions of Z_1 and $\tilde{v}$. Then, the optimization problem with constraint

$$\inf_{\pi} \sum_{i,j=1}^N \sum_{k=1}^M \pi_{i,j,k} (|y_{i,j,1} - v_k|^2 + |y_{i,j,2} - (v_k - K)^+|^2)$$

is solved,¹¹ where π is a probability distribution that has to satisfy positivity constraints and marginal distribution constraints given by $a_{i,j}$ and b_k . Thanks to Brenier's theorem the optimizer π^* is supported on the graph of $\nabla \Gamma$.

In order to obtain the potential Γ from the values of $\nabla \Gamma(y_{i,j})$, we train a feed-forward neural network whose gradient at point $y_{i,j}$ is approximately $\nabla \Gamma(y_{i,j})$. This training step is done by

¹¹The POT (Python Optimal Transport) library contains an implementation of this procedure (??).

minimizing the loss function

$$\min_f \sum_{i=1}^N \sum_{j=1}^2 \left| \frac{f(y_{i,j} + \varepsilon e_k) - f(y_{i,j} - \varepsilon e_k)}{2\varepsilon} - \partial_k \Gamma(y_{i,j}) \right|^2 + \text{Conv}(f) \quad (\text{D.1})$$

where the minimization problem is over the parameters of the feed-forward neural network f and $(e_k)_{k=1,2}$ is the canonical basis of $\mathbb{R}^2$. The term $\text{Conv}(f)$ is a cost for lack of convexity of the neural network that is defined in (?). The minimizer f^* of this problem is an approximation of the potential function $\Gamma(y)$ and in this low dimensional setting we can plot ∇f and compare it to $\nabla \Gamma(y_{i,j})$. If needed, we can compute $\mathbb{E}[\Gamma(Z_1)]$ by a Monte Carlo scheme and subtract it from Γ to obtain the normalization of Γ in Theorem ??.

D.2 Backward integration of PDEs

For time step $\Delta t > 0$ fixed, the functions $K(t, y)$, $\nabla K(t, y)$ and $e^{\phi(t,y)}$ satisfy the dynamic programming principles

$$K(t, y) = \mathbb{E}[K(t + \Delta t, y + Z_{\Delta t})], \quad \nabla_y K(t, y) = \mathbb{E}[\nabla_y K(t + \Delta t, y + Z_{\Delta t})]$$

and

$$\begin{aligned} e^{\phi(t,y)} &= \mathbb{E} \left[e^{\phi(t+\Delta t, y+Z_{\Delta t}) + \alpha \text{tr}(\Sigma_z \int_t^{t+\Delta t} \nabla_y^2 K(u, Z_u) du)} \right] \\ &\sim (1 + \alpha \Delta t \text{tr}(\Sigma_z \nabla_y^2 K(t, y))) \mathbb{E} \left[e^{\phi(t+\Delta t, y+Z_{\Delta t})} \right] \end{aligned}$$

where the approximation is valid for small Δt . We choose $\Delta t \sim 1/100$. Then, using the transition densities of $Z_{\Delta t}$ on the grid $y_{i,j}$, we use the dynamic programming principles to compute these functions on the space-time grid $t = k\Delta t$ and $y = y_{i,j}$ backward starting from the final conditions

$$\Gamma(1, y_{i,j}) = \Gamma(y_{i,j}), \quad \nabla \Gamma(1, y_{i,j}) = \nabla \Gamma(y_{i,j}), \quad \text{and} \quad e^{\phi(1, y_{i,j})} = e^{-\alpha y'_{i,j} \nabla \Gamma(y_{i,j}) + \alpha \Gamma(y_{i,j})}.$$

D.3 Computing the Physical Distribution

We can take the logarithm of $e^{\phi(k\Delta t, y_{i,j})}$ and differentiate numerically to obtain $\nabla \phi(k\Delta t, y_{i,j})$. Then, we simulate Y under the physical measure as (??) via an Euler scheme

$$Y_{k+1} = Y_k + \Sigma_z \nabla \phi(k\Delta t, Y_k) \Delta t + \Delta Z_{k+1}$$

where $\Delta Z_{k+1} = Z_{(k+1)\Delta t} - Z_{k\Delta t}$ and $\nabla \phi(k\Delta t, Y_k)$ is evaluated at $y_{i,j}$ closest to Y_k . The empirical distribution of the resulting random vector $\nabla \Gamma(Y_N)$ is approximately $\hat{F}$.