Cross impact in derivative markets

Mehdi Tomas[1,2,3], Iacopo Mastromatteo[3,4], and Michael Benzaquen[1,3,4]

[1]LadHyX UMR CNRS 7646, Ecole Polytechnique, 91128 Palaiseau Cedex, France
[2]CMAP UMR CNRS 7641, Ecole Polytechnique, 91128 Palaiseau Cedex, France
[3]Chair of Econophysics & Complex Systems, Ecole Polytechnique, 91128 Palaiseau Cedex, France
[4]Capital Fund Management, 23-25, Rue de l'Université 75007 Paris, France

February 4, 2021

Abstract

We introduce a linear cross-impact framework in a setting in which the price of some given financial instruments (*derivatives*) is a deterministic function of one or more, possibly tradeable, stochastic factors (*underlying*). We show that a particular cross-impact model, the multivariate Kyle model, prevents arbitrage and aggregates (potentially non-stationary) traded order flows on derivatives into (roughly stationary) liquidity pools aggregating order flows traded on both derivatives and underlying. Using E-Mini futures and options along with VIX futures, we provide empirical evidence that the price formation process from order flows on derivatives is driven by cross-impact and confirm that the simple Kyle cross-impact model is successful at capturing parsimoniously such empirical phenomenology. Our framework may be used in practice for estimating execution costs, in particular hedging costs.

Contents

Electronic copy available at: https://ssrn.com/abstract=3779461

1 Introduction

Understanding the relation that connects trade imbalances and price changes is arguably one of the main goals of market microstructure theory. By leveraging a *data deluge* that increased both quantity and quality of financial data over the past 20 years, many empirical results have shed light on market impact (see e.g. [2, 4, 17]). Further, it lead to an adjustment of the theoretical framework underpinning the foundations of market microstructure [7, 9], and it drove a more data-driven community to build novel settings in which to accommodate these findings (see e.g. [4, 15]). On the other hand, only sparse attention has been given in the literature to the related problem of establishing a relation between order flows and price changes in a multi-instrument setting (*cross-impact*). Compared to its univariate counterpart, cross-impact is much harder to characterize empirically, due to the larger number of degrees of freedom involved, and the consequently smaller signal-to-noise ratio [8, 12, 14, 19]. The challenge of cross-impact modeling also raises from genuinely new effects that only appear in the multivariate setting, such as the possible presence of cross-sectional arbitrages. Accounting for these problems requires building some dedicated theoretical infrastructure [1, 11], and it led us to investigate in [16] the principles by which a cross-impact model should abide in order to be free from basic inconsistencies and arbitrages.

This works aims precisely at illustrating the consequences of such a principled approach in a limit case that exposes both the flaws of inconsistently formulated models and the benefits of having strong theoretical guarantees on the behavior of a cross-impact model. In fact, we focus our study on cross-impact in a universe of instruments that comprises potentially illiquid derivatives. This is the prototypical case in which a univariate approach to modeling impact and liquidity falls short, due to two main reasons respectively related to *prices* and *liquidity*. Indeed, in an efficient market the *prices* of derivatives should be locked by non-arbitrage, and hence they are not expected to respond to trades independently of their underlying. Regarding *liquidity*, it is intuitive that in presence of many strongly correlated and individually illiquid financial instruments (e.g. options), it is necessary to aggregate the liquidity of multiple products into common liquidity pools in order to have a satisfactory description of the price response, which would otherwise appear anomalously strong. Both points indicate that in order to have a viable model to describe impact on derivative markets in presence of fragmented liquidity it is unavoidable to take a multivariate perspective on the system, namely one that is able to single out the relevant liquidity factors.

Our approach allows to map this problem onto a dimensionality reduction one, showing that one can replace a high dimensional space of instruments (say, option surfaces) with a lower dimensional representation of underlying factors (spot, implied volatility), which are the only degrees of freedom allowed to respond to order flow imbalances.

It is worth emphasizing at this point that, although other models for cross-impact of derivatives (in particular for options) have already appeared in the literature, our focus is different in several respects. A first stream of works aims at the *characterization* of non-arbitrage properties in cross-impact models [1], aiming at defining necessary conditions that should be satisfied in the general case. Here, we study a specific instance of such class of models, which leads to a very rich phenomenology that cannot be fully appreciated in a completely generic setting. A second series of works focuses on the *implications* that some specific cross-impact models have in the context of option replication and hedging [3, 10], due to the fact that in presence of cross-impact non-linear effects arise, and potentially dangerous feedback loops emerge. Finally, in [18] the emphasis is given to the *empirical determination* of cross-impact; one of its findings of interest to us is the identification of impact along a non-trivial underlying factor (level of implied volatility) that does not mechanically correspond to any individual option, and yet emerges from the aggregation of the whole volatility surface. An important contribution that pushes this approach even further is [13], which indicates that other factors (skew of the volatility surface) are also necessary to accurately describe cross-impact on options. Even though the last two references strongly relate to our work, there are several differences that characterize our approach. First, we try to infer from data the rules to be used to aggregate liquidities, rather than postulating them. Second, our approach can easily accommodate multi-factor cross-impact models, whereas in such references only one factor at the time was considered. Third and last, the theoretical foundations our our approach are strongly grounded in the classical market microstructure literature [5, 12], and in particular in [6], which provides a solid micro-foundation of our model (see also [16]).

The paper is organized as follows. In Section 2 we introduce the notations used. Section 3 presents our modeling framework. Section 4 provides illustrative examples of its applications. Section 5 presents the empirical results of cross-impact on options. In Section 6 we conclude on the contributions of the paper, open questions, and directions for future work.

Electronic copy available at: https://ssrn.com/abstract=3779461

2 Notations

Throughout the paper, we write scalars in roman lower cases, vectors in bold lower cases and matrices in roman upper cases. The set of n by n real-valued square matrices is denoted by $\mathscr{M}_n(\mathbb{R})$, the set of orthogonal matrices by $\mathscr{O}(n)$, the set of real non-singular n by n matrices by $\mathrm{GL}_n(\mathbb{R})$, the set of real n by n symmetric positive semi-definite matrices by $\mathscr{S}_n^+(\mathbb{R})$, and the set of real n by n symmetric positive definite matrices by $\mathscr{S}_n^{++}(\mathbb{R})$. Further, given a matrix A in $\mathscr{M}_n(\mathbb{R})$, $A^\top$ denotes its transpose. Given A in $\mathscr{S}_n^+(\mathbb{R})$, we write $A^{1/2}$ for a matrix such that $A^{1/2}(A^{1/2})^\top = A$ and $\sqrt{A}$ for the matrix square root, the unique positive semi-definite symmetric matrix such that $(\sqrt{A})^2 = A$. We write $\ker(M)$ for the null space of a matrix $M \in \mathscr{M}_n$, Π_V for the projector on a linear subspace of $V \in \mathbb{R}^n$ and $\bar{\Pi}_V = \mathbb{I} - \Pi_V$ for the orthogonal projector. Finally, given a vector $\boldsymbol{v} \in \mathbb{R}^n$, we write $\boldsymbol{v} = (v_1, \dots, v_n)$ and $\mathrm{diag}(\boldsymbol{v})$ for the diagonal matrix with diagonal components the components of $\boldsymbol{v}$.

All stochastic processes in the text are defined on a probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{R}}, \mathbb{P})$ and will be adapted to the filtration $(\mathscr{F}_t)_{t\in\mathbb{R}}$ unless stated otherwise. Standard Brownian motions are defined with respect to the probability measure $\mathbb{P}$. All stochastic differential equations introduced will be assumed to have a unique strong solution and correspondingly the functions appearing in these equations will be assumed to be sufficiently regular for this to be true. We denote by $\mathbb{E}$ the expectation with respect to the probability measure $\mathbb{P}$ and $\mathbb{E}_t[\dots]$ will denote the conditional expectation $\mathbb{E}[\dots|\mathscr{F}_t]$.

3 The model

We consider a universe comprising two classes of financial instruments, that we will refer to as *underlying* and *derivatives*. With some abuse of language, the notion of *underlying* will apply to a set of N stochastic processes, that might describe indistinctly a set of tradable financial instruments or an ensemble of stochastic factors. The prices of these N instruments will be denoted $\boldsymbol{p}_t = (p_t^1, \dots, p_t^N)$. We define as *derivatives* a set of M instruments, whose prices $\boldsymbol{P}(\boldsymbol{p}_t, t) = (P^1(\boldsymbol{p}_t, t), \dots, P^M(\boldsymbol{p}_t, t))$ are deterministic functions of the underlying price process $\boldsymbol{p}_t$.

We assume that impact is linear in the traded order flows and we denote by $\boldsymbol{q}_t = (q_t^1, \dots, q_t^N)$ the stochastic process corresponding to the net traded order flows on the underlying and by $\boldsymbol{Q}_t = (Q_t^1, \dots, Q_t^M)$ the stochastic process corresponding to the net traded order flows on derivatives. Such flows are not associated to any specific agent, and rather denote aggregate market order flow. Since only the non-predictable component of order flows contributes to impact, up to replacing $\boldsymbol{q}_t$ by $\boldsymbol{q}_t - \mathbb{E}_{t-}[\boldsymbol{q}_t]$ and $\boldsymbol{Q}_t$ by $\boldsymbol{Q}_t - \mathbb{E}_{t-}[\boldsymbol{Q}_t]$ in the following, we assume that $\mathbb{E}_{t-}[\boldsymbol{q}_t] = 0, \mathbb{E}_{t-}[\boldsymbol{Q}_t] = 0$.

Section 3.1 introduces the underlying and derivative price dynamics with cross-impact. Section 3.2 presents the proposed cross-impact model and Section 3.3 discusses the properties of this cross-impact model.

3.1 Price dynamics

We assume that the dynamics of the underlying are

$$\mathrm{d}\boldsymbol{p}_t = \boldsymbol{\mu}_p(\boldsymbol{p}_t, t)\mathrm{d}t + \mathscr{G}_p(\boldsymbol{p}_t, t)\mathrm{d}\boldsymbol{w}_t + \Lambda_{pq}(\boldsymbol{p}_t, t)\mathrm{d}\boldsymbol{q}_t + \Lambda_{pQ}(\boldsymbol{p}_t, t)\mathrm{d}\boldsymbol{Q}_t, \tag{1}$$

where $\boldsymbol{w}$ is a standard N-dimensional Brownian motion, $\boldsymbol{\mu}_p \colon \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ is the drift, $\mathscr{G}_p \colon \mathbb{R}^N \times \mathbb{R} \to \mathrm{GL}_N(\mathbb{R})$ is the diffusion matrix, $\Lambda_{pq} \colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{N,N}(\mathbb{R})$ is the underlying-underlying cross-impact matrix and $\Lambda_{pQ} \colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{N,M}(\mathbb{R})$ is the underlying-derivative cross-impact matrix. The underlying dynamics of Equation (1) could be enriched by jump processes, but for the sake of simplicity we chose a simple continuous framework which covers several well-known models of derivatives pricing.

Since derivative prices are deterministic functions of the underlying, without loss of generality, the derivative dynamics can be written as

$$\mathrm{d}\boldsymbol{P}_t = \boldsymbol{\mu}_P(\boldsymbol{p}_t, t)\mathrm{d}t + \mathscr{G}_P(\boldsymbol{p}_t, t)\mathrm{d}\boldsymbol{w}_t + \Lambda_{Pq}(\boldsymbol{p}_t, t)\mathrm{d}\boldsymbol{q}_t + \Lambda_{PQ}(\boldsymbol{p}_t, t)\mathrm{d}\boldsymbol{Q}_t, \tag{2}$$

where $\boldsymbol{\mu}_P \colon \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^M$ is the derivative drift, $\mathscr{G}_P \colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{M,N}(\mathbb{R})$ is the derivative diffusion matrix, $\Lambda_{Pq} \colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{M,N}(\mathbb{R})$ is the derivative-underlying cross-impact matrix and $\Lambda_{PQ} \colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{N,M}(\mathbb{R})$ is the derivative-derivative cross-impact matrix.

For convenience cross-impact matrices can be compactly rearranged into a single matrix, Λ, which we refer to as the cross-impact matrix since it describes the cross-impact of the complete system

$$\Lambda(\boldsymbol{p}_t, t) := \begin{pmatrix} \Lambda_{pq} & \Lambda_{pQ} \\ \Lambda_{Pq} & \Lambda_{PQ} \end{pmatrix}(\boldsymbol{p}_t, t). \tag{3}$$

Electronic copy available at: https://ssrn.com/abstract=3779461

Similarly drift and diffusion terms can be grouped as

$$\boldsymbol{\mu}(\boldsymbol{p}_t, t) = (\boldsymbol{\mu}_p(\boldsymbol{p}_t, t), \boldsymbol{\mu}_P(\boldsymbol{p}_t, t)) \qquad \mathcal{G}(\boldsymbol{p}_t, t) = (\mathcal{G}_p(\boldsymbol{p}_t, t), \mathcal{G}_P(\boldsymbol{p}_t, t))\,. \tag{4}$$

Obviously, not all the terms appearing in Equations (3) and (4) are independent, given that the function $\boldsymbol{P}(\boldsymbol{p}_t, t)$ fixes the derivative prices as a function of the underlying. In Sec. 3.3.1 we will discuss the conditions required to prevent arbitrage.

We also draw the reader's attention to the fact that one may directly obtain an equation for the derivatives' prices by applying Ito's formula and using the underlying dynamics of Equation (1). This would yield expressions for Λ_{Pq} and Λ_{PQ} as a function of Λ_{pq} and Λ_{pQ}. However, not all cross-impact models give expressions of Λ_{Pq} and Λ_{PQ} consistent with Ito's formula and it is thus more convenient to choose a cross-impact model which, by construction, yields this result instead of imposing it *a priori*. Our choice of cross-impact model is discussed in the next section.

3.2 Impact model

The impact model we propose involves two parameters, the return covariance matrix and the order flow covariance matrix. The underlying-underlying return covariance matrix $\Sigma_{pp}\colon \mathbb{R}^N \times \mathbb{R} \to \mathcal{S}_N^+(\mathbb{R})$ is defined as

$$\Sigma_{pp}(\boldsymbol{p}_t, t)\mathrm{d}t := \mathbb{E}_t[\mathrm{d}\boldsymbol{p}_t \mathrm{d}\boldsymbol{p}_t^\top] - \mathbb{E}_t[\mathrm{d}\boldsymbol{p}_t]\mathbb{E}_t[\mathrm{d}\boldsymbol{p}_t^\top]\,, \tag{5}$$

and we similarly denote $\Sigma_{pP}, \Sigma_{Pp} = \Sigma_{pP}^\top, \Sigma_{PP}$ for the underlying-derivative, derivative-underlying and derivative-derivative return covariance matrices. Naturally, since derivative prices are deterministic function of the underlying, these matrices are all related to Σ_{pp}. We denote by $\Omega_{qq}\colon \mathbb{R}^N \times \mathbb{R} \to \mathcal{S}_N^+(\mathbb{R})$ the underlying-underlying order flow covariance matrix

$$\Omega_{qq}(\boldsymbol{p}_t, t)\mathrm{d}t := \mathbb{E}_t[\mathrm{d}\boldsymbol{q}_t \mathrm{d}\boldsymbol{q}_t^\top] - \mathbb{E}_t[\mathrm{d}\boldsymbol{q}_t]\mathbb{E}_t[\mathrm{d}\boldsymbol{q}_t^\top]\,, \tag{6}$$

and we denote Ω_{qQ}, $\Omega_{Qq} = \Omega_{qQ}^\top$ and Ω_{QQ} for the underlying-derivative, derivative-underlying and derivative-derivative order flow covariances. Contrary to return covariance matrices, there are no constraints betweeen these order flow covariance matrices and Ω_{qq}. The covariance structure of returns and flows for the whole system can be arranged compactly as

$$\Sigma(\boldsymbol{p}_t, t) = \begin{pmatrix} \Sigma_{pp} & \Sigma_{pP} \\ \Sigma_{pP}^\top & \Sigma_{PP} \end{pmatrix}(\boldsymbol{p}_t, t) \qquad \Omega(\boldsymbol{p}_t, t) = \begin{pmatrix} \Omega_{qq} & \Omega_{qQ} \\ \Omega_{qQ}^\top & \Omega_{QQ} \end{pmatrix}(\boldsymbol{p}_t, t)\,.$$

The impact model that we propose, first derived in [5], has been analyzed in this context in [6, 16], where it was referred as the *Kyle cross-impact* model. The model prescribes using a Λ of the form

$$\Lambda = \sqrt{Y}\Lambda_{\text{kyle}}(\Sigma, \Omega) := \sqrt{Y}(\Omega^{-1/2})^\top \sqrt{(\Omega^{1/2})^\top \Sigma \Omega^{1/2}}\,\Omega^{-1/2}\,, \tag{7}$$

where we have omitted the dependence on $(\boldsymbol{p}_t, t)$ for convenience, and where we have introduced $\Lambda_{\text{kyle}}\colon \mathcal{S}_{N+M}^+(\mathbb{R}) \times \mathcal{S}_{N+M}^{++}(\mathbb{R}) \to \mathcal{M}_{N+M,N+M}(\mathbb{R})$. We introduced $0 < Y < 1$ whose interpretation will appear clearly in the next section. The next section emphasizes the properties that motivate the usage of the Kyle cross-impact model.

3.3 Properties

The Kyle cross-impact model (i) ensures derivatives are priced consistently with the underlying dynamics, (ii) reduces the dynamics of the system to a classic SDE when flow degrees of freedom are integrated inside volatility terms, (iii) allows for dimensionality reduction, (iv) properly aggregates liquid and illiquid instruments and (v) does not need to know which instruments belong to the set of underlyings and which ones are derivatives. We show these properties below and discuss their implications.

3.3.1 Price efficiency

In order for the model to be self-consistent (i.e., for the derivative to be efficiently priced at all times), one should make sure that the dynamics that we have postulated in Equation (2) are consistent with the one obtained by applying Ito's formula to the dynamics of the underlying via the relation $\boldsymbol{P}(\boldsymbol{p}_t, t)$. This requirement prescribes that the following conditions should hold.

Electronic copy available at: https://ssrn.com/abstract=3779461

Diffusion In order for the diffusion dynamics to be consistent, one should have

$$\mathscr{G}_P(\boldsymbol{p}_t, t) = \Xi(\boldsymbol{p}_t, t)\mathscr{G}_p(\boldsymbol{p}_t, t)\,, \tag{8}$$

where $\Xi\colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{M,N}(\mathbb{R})$ is a *sensitivity matrix* defined as

$$\Xi_{ij}(\boldsymbol{p}_t, t) = \frac{\partial P_i}{\partial p_j}(\boldsymbol{p}_t, t)\,.$$

Impact By the same token, the cross-impact matrix should satisfy the condition

$$\Lambda(\boldsymbol{p}_t, t) = \begin{pmatrix} \Lambda_{pq} & \Lambda_{pQ} \\ \Xi\Lambda_{pq} & \Xi\Lambda_{pQ} \end{pmatrix}(\boldsymbol{p}_t, t)\,, \tag{9}$$

indicating that the price response of the derivative should be completely fixed by the one on the underlying.

Drift Finally, the drift term should also match the one obtained via Ito's formula, so that

$$\boldsymbol{\mu}_P(\boldsymbol{p}_t, t) = \boldsymbol{\Theta}(\boldsymbol{p}_t, t) + \Xi(\boldsymbol{p}_t, t)\boldsymbol{\mu}_p(\boldsymbol{p}_t, t) + \frac{1}{2}\sum_{jk} \boldsymbol{\chi}_{jk}(\boldsymbol{p}_t, t)(\Sigma_{pp}(\boldsymbol{p}_t, t))_{jk}\,, \tag{10}$$

where

$$\Theta_i(\boldsymbol{p}_t, t) = \frac{\partial P_i}{\partial t}(\boldsymbol{p}_t, t)$$
$$\chi_{i,jk}(\boldsymbol{p}_t, t) = \frac{\partial P_i}{\partial p_j \partial p_k}(\boldsymbol{p}_t, t)\,,$$

are sensitivities, of nature similar to Ξ defined above.

We will assume in the following that the conditions for drift and diffusion hold by construction, taking Equations (8) and (10) as definitions of respectively $\mathscr{G}_P$ and $\boldsymbol{\mu}_P$. On the other hand, since Equation (9) is not guaranteed *a priori* under our choice of the cross-impact model, we will be required to prove that our model satisfies Equation (9) above, and thus that our construction is consistent.

In order to do so, we need to discuss a crucial property of the Kyle model, that is referred in [16] as *strong fragmentation invariance*. With strong fragmenation invariance, we denote the fact that for any $\emptyset \subset V \subseteq \ker(\Sigma)$ we have

$$\Pi_V \Lambda_{\text{kyle}}(\Sigma, \Omega) = 0 \tag{11}$$
$$\Lambda_{\text{kyle}}(\Sigma, \Omega)\Pi_V = 0 \tag{12}$$
$$\Lambda_{\text{kyle}}(\Sigma, \bar{\Pi}_V \Omega \bar{\Pi}_V) = \Lambda_{\text{kyle}}(\Sigma, \Omega)\,. \tag{13}$$

Equation (11) shows that linear combinations of instruments with constant price (zero volatility) are not impacted by the order flow. In our setting, this guarantees that derivatives are always priced efficiently even in the presence of order flow pressure. Equation (12) maintains that order flow pressure on non-fluctuating modes should not be able to influence the price of any other combination of products. This prevents pushing the price of fluctuating instruments by trading zero-volatility (free) linear combination of instruments. Finally, Equation (13) shows that the way non-fluctuating modes are traded has no influence the price of any instrument.

In our context, strong fragmentation invariance and symmetry imply that

$$\Lambda(\boldsymbol{p}_t, t) = \begin{pmatrix} \Lambda_{pq} & \Lambda_{pq}\Xi^\top \\ \Xi\Lambda_{pq} & \Xi\Lambda_{pq}\Xi^\top \end{pmatrix}(\boldsymbol{p}_t, t)\,. \tag{14}$$

Thus the Kyle model satisfies the condition of Equation (9) above, thus guaranteeing that (along with the conditions of Equations (8) and (10) which we impose and are independent of the impact model) the derivative is efficiently priced.

At this moment, it is worth emphasizing a central point of our approach: we are postulating that derivatives are efficiently priced at all times despite the presence of finite liquidity. This hides the implicit assumption that some market actors are able to arbitrage away the spread between underlying and derivatives very quickly, and that those market directions are effectively frictionless for such actors. Inefficiencies can be implemented in this framework by relaxing the assumption that $\boldsymbol{P}_t = \boldsymbol{P}(\boldsymbol{p}_t, t)$, rather assuming that derivative prices mean-revert with a finite velocity to their theoretical value. On long enough time scales, we expect the empirical estimate for Σ to nevertheless be close to the value given by no-arbitrage so that these considerations can be ignored. For pedagogical clarity, hereafter we will stick to the idealized case in which derivatives are efficiently priced.

Electronic copy available at: https://ssrn.com/abstract=3779461

3.3.2 Consistency of frictionless and impacted dynamics

We now want to show that the dynamics with impact are consistent with the *frictionless* dynamics that would be observed when disregarding the flows degrees of freedom. The term frictionless in this context does not indicate absence of impact, but rather identifies the effective system in which the degrees of freedom related to flows ($\boldsymbol{q}_t$ and $\boldsymbol{Q}_t$) are integrated inside volatility contributions.

To prove this, we first remark that the Kyle model is the only linear, symmetric positive definite impact model that is consistent with the covariance structure of the system (see [16]), namely the fact that

$$Y\Sigma(\boldsymbol{p}_t, t) = \Lambda(\boldsymbol{p}_t, t)\Omega(\boldsymbol{p}_t, t)\Lambda^\top(\boldsymbol{p}_t, t)\,. \tag{15}$$

This property is not surprising, given ubiquity of the *inconspicous equilibrium* property in theoretical frameworks in which market efficiency is achieved by rational agents that attempt to forecast future returns [5, 6]. In these frameworks, predictions are optimal when Equation (15) is verified. The prefactor Y expresses that a fraction Y of the total covariance is due to trading activity while the remainder comes from shocks unrelated to the order flow.

Now, let us relate a frictionless dynamics to the one that we have postulated in our model thanks to the covariance consistency condition (Equation (15)). To do so, consider the dynamics without impact

$$\mathrm{d}\tilde{\boldsymbol{p}}_t = \tilde{\boldsymbol{\mu}}_p(\tilde{\boldsymbol{p}}_t, t)\mathrm{d}t + \tilde{\mathscr{G}}_p(\tilde{\boldsymbol{p}}_t, t)\mathrm{d}\tilde{\boldsymbol{w}}_t \tag{16}$$

$$\mathrm{d}\tilde{\boldsymbol{P}}_t = \tilde{\boldsymbol{\mu}}_P(\tilde{\boldsymbol{p}}_t, t)\mathrm{d}t + \tilde{\mathscr{G}}_P(\tilde{\boldsymbol{p}}_t, t)\mathrm{d}\tilde{\boldsymbol{w}}_t \tag{17}$$

where $\tilde{\boldsymbol{w}}$ is a standard Brownian motion, $\tilde{\boldsymbol{\mu}}_p \colon \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ is the frictionless underlying drift, $\tilde{\boldsymbol{\mu}}_P \colon \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^M$ is the frictionless derivative drift, $\tilde{\mathscr{G}}_p \colon \mathbb{R}^N \times \mathbb{R} \to \mathrm{GL}_N(\mathbb{R})$ and $\tilde{\mathscr{G}}_P \colon \mathbb{R}^N \times \mathbb{R} \to \mathscr{M}_{M,N}(\mathbb{R})$ are the frictionless diffusion matrices. Then it is simple to verify that thanks to Equation (15) one can recover the equivalence between frictionless and impacted dynamics if $\mathscr{G} = \sqrt{1-Y}\tilde{\mathscr{G}}$ and $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}$. In intuitive terms, our price dynamics reduces to the vanilla stochastic evolution of derivatives and underlying for any observer that does not possess information about the order flows. This stems from the Gaussian nature of both order flow processes and underlying increments, along with the covariance consistency condition (Equation (15)) that is implicitly contained in the definition of the Kyle model.

3.3.3 Dimensionality reduction

The strong fragmentation property introduced in Sec. 3.3.1 has another important implication in our setting: by imposing the structure of the impact matrix to the form of Equation (14), it explicitly shows that $\mathrm{rank}(\Lambda) = \mathrm{rank}(\Lambda_{pq}) = N$. This has the consequence of making explicit that no more than N distinct liquidity pools are necessary in order to compute impact. In particular, we will show that it is enough to estimate $\bar{\Pi}_V\Omega\bar{\Pi}_V$ instead of the (potential) rank $N+M$ matrix Ω in order to fully determine the cross-impact matrix, where with V we denote the subspace spanning the direction $\boldsymbol{P}_t - \boldsymbol{P}(\boldsymbol{p}_t, t)$, which obviously belongs to $\ker(\Sigma)$ if we assume that derivatives are consistently priced. In intuitive terms, it is only the liquidity in the direction of the fluctuating degrees of freedom that contributes to the overall liquidity pool, whereas any trading in the direction of the mispricing has no effect on the system.

To show this, remark that Equation (14) allows us to express the full cross-impact matrix Λ once an expression for the lower-rank Λ_{pq} object is available. We will thus derive the expression of Λ_{pq}. Writing $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{P}}$ for the underlying and derivative impact component of the price, one has $\mathrm{d}\hat{\boldsymbol{P}}_t = \Xi\mathrm{d}\hat{\boldsymbol{p}}_t$ and

$$\begin{pmatrix}\mathrm{d}\hat{\boldsymbol{p}}_t\\ \mathrm{d}\hat{\boldsymbol{P}}_t\end{pmatrix} = \begin{pmatrix}I & 0\\ \Xi & I\end{pmatrix}\begin{pmatrix}\mathrm{d}\hat{\boldsymbol{p}}_t\\ 0\end{pmatrix} = \begin{pmatrix}I & 0\\ \Xi & I\end{pmatrix}\begin{pmatrix}\Lambda_{pq} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}I & \Xi^\top\\ 0 & I\end{pmatrix}\begin{pmatrix}\mathrm{d}\boldsymbol{q}_t\\ \mathrm{d}\boldsymbol{Q}_t\end{pmatrix},$$

so that the impacted underlying price is

$$\mathrm{d}\hat{\boldsymbol{p}}_t = \Lambda_{pq}(\mathrm{d}\boldsymbol{q}_t + \Xi^\top\mathrm{d}\boldsymbol{Q}_t)\,.$$

Since Λ is symmetric and positive-definite, one can easily show that Λ_{pq} is symmetric and positive-definite. Furthermore, Λ_{pq} is covariance-consistent and

$$Y\Sigma_{pp} = \Lambda_{pq}\Omega_\Xi\Lambda_{pq}^\top,$$

where we have introduced the aggregated order flow covariance matrix

$$\Omega_\Xi := \mathbb{E}_t[(\mathrm{d}\boldsymbol{q}_t + \Xi^\top\mathrm{d}\boldsymbol{Q}_t)(\mathrm{d}\boldsymbol{q}_t + \Xi^\top\mathrm{d}\boldsymbol{Q}_t)^\top] = \Omega_{qq} + \Xi^\top\Omega_{QQ}\Xi + \Xi^\top\Omega_{Qq} + \Omega_{qQ}\Xi\,.$$

Electronic copy available at: https://ssrn.com/abstract=3779461

Therefore, since these properties imply that Λ_{pq} is the Kyle cross-impact model with return covariance Σ_{pp} and order flow covariance Ω_Ξ one has

$$\Lambda_{pq} = \sqrt{Y}\Lambda_{\text{kyle}}(\Sigma_{pp}, \Omega_\Xi) = \sqrt{Y}(\Omega_\Xi^{-1/2})^\top \sqrt{(\Omega_\Xi^{1/2})^\top \Sigma_{pp} \Omega_\Xi^{1/2}}\, \Omega_\Xi^{-1/2}.$$

This property is very useful in practice since it gives a recipe for computing Λ_{pq}. It also tells us that the aggregated order flow covariance matrix combines direct trading of underlying and indirect trading through derivatives exposures to the underlying. In particular, even if an underlying is not tradeable, as long as derivatives which are exposed to that factor are tradeable its aggregated liquidity will be non-zero.

3.3.4 Cross-stability and spurious liquidity effects

The Kyle-model is also *cross-stable* (see [16]), meaning that trading an illiquid combination of products cannot move liquid products by a disproportionate amount. To make this property explicit, let us introduce a subset of illiquid instruments W, on which we assume to measure a modified covariance matrix

$$\Omega_\epsilon := (\bar{\Pi}_W + \epsilon \Pi_W)\Omega(\bar{\Pi}_W + \epsilon \Pi_W),$$

where $\Pi_W = \mathbb{I} - \bar{\Pi}_W$ is a projector on the space W. The covariance matrix above is the one we would have observed if liquidities of instruments belonging to W were multiplied by ϵ. Then, for any $(\Sigma, \Omega) \in (\mathscr{S}_n^+(\mathbb{R}) \times \mathscr{S}_n^{++}(\mathbb{R}))$, the Kyle model satisfies

$$\bar{\Pi}_W \Lambda_{\text{kyle}}(\Sigma, \Omega_\epsilon)\Pi_W \underset{\epsilon\to 0}{=} O(1) \tag{18}$$

$$\bar{\Pi}_W \Lambda_{\text{kyle}}(\Sigma, \Omega_\epsilon)\bar{\Pi}_W \underset{\epsilon\to 0}{\to} \bar{\Pi}_W \Lambda_{\text{kyle}}\left(\bar{\Pi}_W \Sigma \bar{\Pi}_W, \bar{\Pi}_W \Omega \bar{\Pi}_W\right)\bar{\Pi}_W. \tag{19}$$

Equation (18) indicates that volumes traded on illiquid products cannot move by a disproportionate (diverging in ϵ) the price of liquid instruments, whereas Equation (19) indicates that the impact of the flows of liquid instruments is not affected by the presence of other, illiquid, instruments.

These properties allow one to include non-tradeable instruments (such as implied volatilities) within the definition of the underlying because even though the quantities $\Omega^{-1/2}$ appearing in Equation (7) are divergent whenever an illiquid product appears, this property implies that the divergence is restricted to the subspace of zero-liquidity products. This is convenient also for cases in which liquidity is fragmented in a large set of products that might display weak liquidity (e.g., out-of-the money options) as this property guarantees that such behavior won't induce spurious features on the overall liquidity pool of the system.

3.3.5 Universality

Throughout our approach, we have split instruments between underlyings and derivatives. However, a compelling property of the chosen cross-impact model is that Equation (7) does not need to know which is which in the universe of instruments. This may be convenient when dealing with a large number of instruments where it would be time-consuming to make explicit which instruments are derivatives and the associated sensitivity matrix Ξ.

4 Examples

To illustrate the flexibility of our setup and the usefulness in practice of the properties of the Kyle cross-impact model, we discuss examples in increasing complexity below.

4.1 Futures

For our first example, we consider a universe of $N = M = 1$ instruments, consisting in a spot with price p_t and a futures contract expiring at a later time T, quoting a price $P(p_t, t)$. By assuming a constant, continuously compounded, deterministic interest rate r one has

$$P(p_t, t) = e^{r(T-t)} p_t.$$

In this case $\Xi(p_t, t) = \partial_p P(p_t, t) = e^{r(T-t)}$ and Equation (14) yields

$$\Lambda(p_t, t) = \sqrt{Y} p_t \frac{\sigma(p_t, t)}{\omega(p_t, t)} \begin{pmatrix} 1 & e^{r(T-t)} \\ e^{r(T-t)} & e^{2r(T-t)} \end{pmatrix},$$

Electronic copy available at: https://ssrn.com/abstract=3779461

where

$$\sigma^2(p_t,t) := \frac{\mathbb{E}_t[p_t^2]-\mathbb{E}_t[p_t]^2}{p_t^2}$$
$$\omega^2(p_t,t) := (1, e^{r(T-t)})^\top \Omega(p_t,t)(1, e^{r(T-t)})\,.$$

The meaning of this formula is rather simple: when dealing with a spot and a future market, there is a single relevant liquidity pool. Such liquidity pool should mix the flow traded on the future and the one traded on the spot. Volumes traded on the futures market should be properly adjusted for the interest rate.

4.2 Black-Scholes model

We now consider a system with a single underlying $N = 1$ and M derivatives. The underlying is a spot with price p_t, whereas the derivatives are a set of European call or put options labeled by $i = 1,\dots,M$, differing for either their strike or their maturity. We assume the price p_t follows the usual log-normal dynamics, with risk-free rate r and implied volatility σ

$$\mathrm{d}p_t = r p_t \mathrm{d}t + \sigma p_t \mathrm{d}w_t\,.$$

Then, with the usual notation for the Black-Scholes Δ, we have

$$\Xi^i(p_t,t) = \partial_p P^i(p_t,t) := \Delta^i_t(p_t,t)\,,$$

and, writing $\boldsymbol{\Delta} := (\partial_p P^1(p_t,t),\cdots,\partial_p P^M(p_t,t))$, Equation (14) yields

$$\Lambda(p_t,t) = \sqrt{Y} p_t \frac{\sigma}{\omega(p_t,t)} \begin{pmatrix} 1 & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \boldsymbol{\Delta}\boldsymbol{\Delta}^\top \end{pmatrix}(p_t,t)\,,$$

where

$$\omega^2(p_t,t) = (1,\boldsymbol{\Delta}(p_t,t))^\top \Omega(p_t,t)(1,\boldsymbol{\Delta}(p_t,t))\,.$$

Thus, as in the previous example, there is a single liquidity pool, with volumes traded on options adjusted for the options' Δ. Volume traded on deep in-the-money options ($\Delta^i \approx 1$) contribute to the overall liquidity pool as if it was the spot itself that was traded, whereas deeply out-of-the-money options ($\Delta^i \approx 0$) give negligible contributions.

4.3 Volatility factors

Building on our previous example, we want to focus on the same class of instruments (one spot and a strip of M European options) in the case in which it is necessary to add an implied volatility term to the Black-Scholes formula to obtain their price. One is required to use a pricing formula $P^i(p_t,\hat{\sigma}^i_t,t)$ in order to accurately describe the price of the options, where $\hat{\sigma}^i_t$ is an *implied volatility*.

In order to reduce the dimensionality of the M implied volatilities $\hat{\sigma}^i_t$, we assume that they lie on a low dimensional surface, so that we are allowed to write

$$\hat{\sigma}^i_t = F^i(\boldsymbol{\varsigma}_t)\,,$$

where $\boldsymbol{\varsigma}_t = (\varsigma^1_t,\dots,\varsigma^Q_t)$ is a set of volatility factors that completely describe the volatility surface through a set of M functions $(F^i(\boldsymbol{\varsigma}))_{i=1}^M$. Note that with some abuse of notation, we will often write

$$P^i(p_t,\hat{F}^i(\boldsymbol{\varsigma}_t),t) = P^i(p_t,\boldsymbol{\varsigma}_t,t)\,,$$

and we will employ a similar notation for other functions of the implied volatility $\hat{\sigma}^i_t$. We then have an underlying consisting of $N = 1+Q$ instrument, of which only one is tradeable (the spot), and where the other Q factors correspond to non-tradeable volatility factors. The sensitivities of the system in this case correspond to

$$\Xi^{i1}_t(p_t,\boldsymbol{\varsigma}_t,t) = \Delta^i_t(p_t,\boldsymbol{\varsigma}_t,t) := \frac{\partial P^i(p_t,\boldsymbol{\varsigma}_t,t)}{\partial p_t}$$
$$\Xi^{i(q+1)}_t(p_t,\hat{\sigma}^i_t,t) = \frac{\partial P^i(p_t,\hat{\sigma}^i_t,t)}{\partial \hat{\sigma}^i_t}\frac{\partial F^i(\boldsymbol{\varsigma}_t)}{\partial \varsigma^q} := \mathcal{V}^i_t(p_t,\boldsymbol{\varsigma}_t,t)\beta^{iq}(p_t,\varsigma_t,t) =: \Upsilon^{iq}\,,$$

Electronic copy available at: https://ssrn.com/abstract=3779461

where $q = 1,\dots,Q$ and where, as it is customary in the literature on option pricing, we have introduced the *vega*

$$\mathcal{V}_t^i(p_t, \hat{\sigma}_t^i, t) = \frac{\partial P_t^i(p_t, \hat{\sigma}_t^i, t)}{\partial \hat{\sigma}_t^i},$$

and the sensitivities of the volatility surface to $\boldsymbol{\varsigma}_t$

$$\beta^{iq}(p_t, \boldsymbol{\varsigma}_t, t) = \frac{\partial F^i(\boldsymbol{\varsigma}_t)}{\partial \varsigma^q}.$$

For convenience, we write the sensitivity matrix in block

$$\Xi = \left(\boldsymbol{\Delta} \mid \Xi^{\cdot 1} \mid \cdots \mid \Xi^{\cdot Q}\right) =: \left(\boldsymbol{\Delta} \mid \Upsilon\right).$$

At this point, the most general expression for the cross-impact matrix that we can write is

$$\Lambda(p_t, \boldsymbol{\varsigma}_t, t) = \begin{pmatrix} \Lambda_{pp} & \Lambda_{p\varsigma} & \Lambda_{pp}\boldsymbol{\Delta}^\top + \Lambda_{p\varsigma}\Upsilon^\top \\ \Lambda_{p\varsigma}^\top & \Lambda_{\varsigma\varsigma} & \Lambda_{\varsigma\varsigma}\Upsilon^\top + \Lambda_{p\varsigma}^\top\boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta}\Lambda_{pp} + \Upsilon\Lambda_{p\varsigma}^\top & \Upsilon\Lambda_{\varsigma\varsigma} + \boldsymbol{\Delta}\Lambda_{p\varsigma} & \boldsymbol{\Delta}\Lambda_{pp}\boldsymbol{\Delta}^\top + \Upsilon\Lambda_{\varsigma\varsigma}\Upsilon^\top + \boldsymbol{\Delta}\Lambda_{p\varsigma}\Upsilon^\top + (\boldsymbol{\Delta}\Lambda_{p\varsigma}\Upsilon^\top)^\top \end{pmatrix}.$$

Single-factor model It is instructive to understand the behavior of the system in the case in which $Q = 1$, and the volatility surface is parametrized by a single level factor:

$$\hat{\sigma}_t^i(\varsigma_t) = F(\varsigma_t) = \varsigma_t,$$

so that one has

$$\Xi^{i1}(p_t, \varsigma_t, t) = \mathcal{V}_t^i(p_t, \varsigma_t, t).$$

In this case the rank two cross-impact matrix can be written as

$$\Lambda(p_t, \varsigma_t, t) = \begin{pmatrix} \Lambda_{pp} & \Lambda_{p\varsigma} & \Lambda_{pp}\boldsymbol{\Delta}^\top + \Lambda_{p\varsigma}\boldsymbol{\mathcal{V}}^\top \\ \Lambda_{p\varsigma} & \Lambda_{\varsigma\varsigma} & \Lambda_{\varsigma\varsigma}\boldsymbol{\mathcal{V}}^\top + \Lambda_{p\varsigma}\boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta}\Lambda_{pp} + \boldsymbol{\mathcal{V}}\Lambda_{p\varsigma} & \boldsymbol{\mathcal{V}}\Lambda_{\varsigma\varsigma} + \boldsymbol{\Delta}\Lambda_{p\varsigma} & \Lambda_{pp}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top + \Lambda_{\varsigma\varsigma}\boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{V}}^\top + \Lambda_{p\varsigma}(\boldsymbol{\Delta}\boldsymbol{\mathcal{V}}^\top + \boldsymbol{\mathcal{V}}^\top\boldsymbol{\Delta}) \end{pmatrix}.$$

In the case where the Δ order flow and $\mathcal{V}$ order flow are not correlated, i.e. writing $\boldsymbol{\Delta}_\mathrm{c} := (1, 0, \frac{\partial P^1}{\partial p}, \cdots, \frac{\partial P^M}{\partial p})$, $\boldsymbol{\mathcal{V}}_\mathrm{c} := (0, 1, \frac{\partial P^1}{\partial \varsigma^1}, \cdots, \frac{\partial P^M}{\partial \varsigma^1})$ we have $\boldsymbol{\Delta}_\mathrm{c}^\top \Omega \boldsymbol{\mathcal{V}}_\mathrm{c} = 0$, we can obtain the expression of the kyle cross-impact matrix for the underlying:

$$\begin{pmatrix} \Lambda_{pp} & \Lambda_{p\varsigma} \\ \Lambda_{p\varsigma} & \Lambda_{\varsigma\varsigma} \end{pmatrix}(p_t, \varsigma_t, t) = \frac{1}{\sqrt{\sigma^2\omega_\Delta^2 + \xi^2\omega_\mathcal{V}^2 + 2\sigma\xi\rho\omega_\Delta\omega_\mathcal{V}}} \begin{pmatrix} \sigma^2 + \frac{\omega_\mathcal{V}}{\omega_\Delta}\sigma\xi\sqrt{1-\rho^2} & \sigma\xi\rho \\ \sigma\xi\rho & \xi^2 + \frac{\omega_\Delta}{\omega_\mathcal{V}}\sigma\xi\sqrt{1-\rho^2} \end{pmatrix}(p_t, \varsigma_t, t), \tag{20}$$

where $\omega_\Delta^2 := \boldsymbol{\Delta}_\mathrm{c}^\top \Omega \boldsymbol{\Delta}_\mathrm{c}$ is the delta-aggregated liquidity, $\omega_\mathcal{V}^2 := \boldsymbol{\mathcal{V}}_\mathrm{c}^\top \Omega \boldsymbol{\mathcal{V}}_\mathrm{c}$ is the vega-aggregated liquidity, $\sigma^2(p_t, \varsigma_t, t) := \mathbb{E}_t[p_t^2] - \mathbb{E}_t[p_t]^2$ is the spot volatility, $\xi^2(p_t, \varsigma_t, t) := \mathbb{E}_t[\varsigma_t^2] - \mathbb{E}_t[\varsigma_t]^2$ is the volatility of volatility and ρ is the spot-vol correlation.

5 Empirical Results

We now illustrate Section 3 with an empirical analysis of cross-impact on derivatives markets. Section 5.1 describes the universe of instruments and the chosen derivative modeling. Section 5.2 shows the empirical observables Σ_{pp} and Ω_Ξ used in Section 5.3 to compute the resulting cross-impact matrix Λ_{pp}. Finally, Section 5.4 stress-tests the fit of cross-impact models and Section 5.5 examines non-parametric evidence of cross-impact.

5.1 Setup

The universe of instruments is made up of (i) the front-month E-mini future, (ii) the two front-month VIX futures, (iii) a set of $M-2$ call and put options on the E-mini. We thus have M derivatives. We model the implied volatility surface with Q volatility factors, as discussed in Section 4.3. We aggregate returns and order flows on a time window of five minutes, so that

Electronic copy available at: https://ssrn.com/abstract=3779461

$\mathrm{d}\boldsymbol{p}_t$ is approximated by the price change on the timeframe of five minutes and $(\mathrm{d}\boldsymbol{q}_t, \mathrm{d}\boldsymbol{Q}_t)$ is approximated by the signed traded order flow during this time window. To avoid confusion, we write $\delta\boldsymbol{p}_t$, $\delta\boldsymbol{q}_t$ and $\delta\boldsymbol{Q}_t$ for these empirical quantities in the rest of this section. Prices and order flows for these instruments are taken from trades and quotes data and more detail on the dataset is provided in Appendix A.

We consider a linear approximation of the implied volatility surface with volatility factors, so that using the notations of Section 4.3, we have $F^i(\boldsymbol{\varsigma}) = \sum_{q=1}^{Q} \beta^{iq}\varsigma^q$ where $i = 1, \ldots, M$. To fit surfaces, we choose $M = 3$ and describe volatility factors in Figure 1. The first factor is a classic implied volatility level factor and we make the rough approximation that VIX futures are solely explained by such *level* factor. The second factor corresponds to the skew of the implied volatility surface, referred to as the *skew* factor hereafter. The third factor explains the term structure of the implied volatility, hence the name the *term* factor in the following.

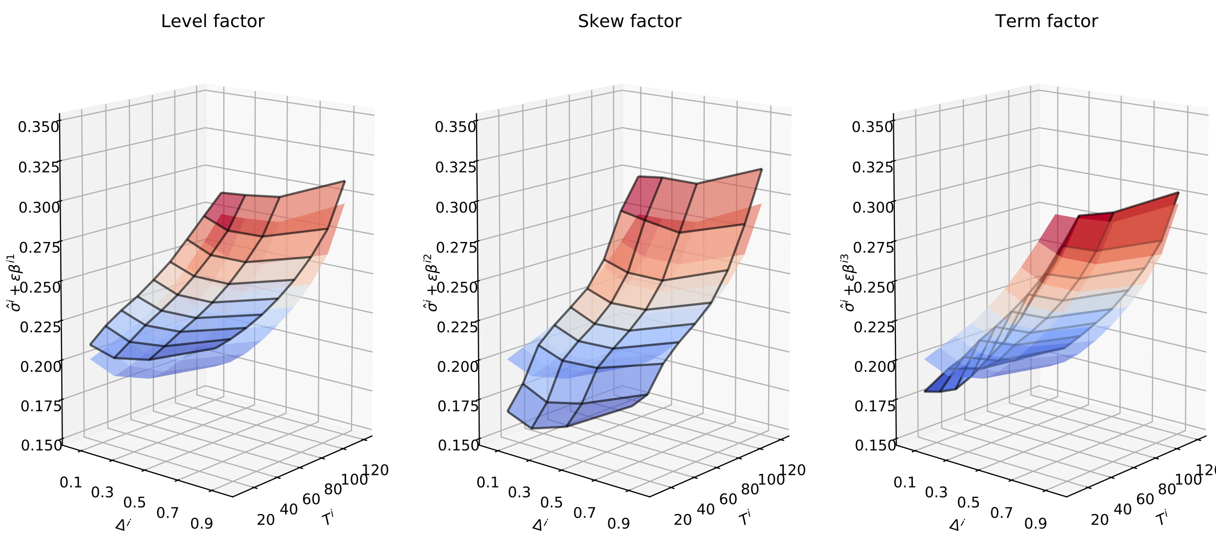

Figure 1: Effect of the different volatility factors on the implied volatility surface.
Starting from a historical implied volatility $\hat{\sigma}^i$, we show the modified implied volatility surface after with a small contribution from the factor q: $\hat{\sigma}^i + \epsilon\beta^{iq}$. The original (non-modified) implied volatility surface is shown in light opacity for reference.

5.2 Relevant observables

Using Section 3.3.3, it suffices to compute the Kyle model associated to the underlying return covariance matrix $\Sigma_{pp}(\boldsymbol{p}_t, t)$ and the aggregated order flow covariance matrix $\Omega_\Xi(\boldsymbol{p}_t, t)$ to obtain the full cross-impact matrix. As we have approximated the behaviour of the system with 4 underlyings, these observables are 4 by 4 matrices. To estimate them, we make the additional assumption that Σ_{pp} and Ω_Ξ are stationary and independent of $\boldsymbol{p}_t$. Figure 2 displays the underlying return correlation matrix $\varrho_{pp} := \mathrm{diag}(\boldsymbol{\sigma})^{-1}\Sigma_{pp}\mathrm{diag}(\boldsymbol{\sigma})^{-1}$ and the risk order flow covariance matrix $\Omega_\Xi^{\mathrm{risk}} := \mathrm{diag}(\boldsymbol{\sigma})\Omega_\Xi\mathrm{diag}(\boldsymbol{\sigma})$ where $\boldsymbol{\sigma} = ((\Sigma_{pp}^{11})^{1/2}, \cdots, (\Sigma_{pp}^{NN})^{1/2})$ is the underlying volatility.

The traded risk (volatility times liquidity) is concentrated on the spot and level directions. This justifies approximating cross-impact on options using solely spot and level underlyings, which we delve in more detail in Section 5.3. The traded risk in the skew direction is much smaller than all other directions and skew order flow is thus expected to contribute less to cross-impact. The underlying return correlation matrix correlation matrix shows strong negative correlation between the spot and level mode. This is a well-known stylised fact, sometimes referred to as the "leverage effect". This will play an important role in the form of the cross-impact model, as highlighted in Equation (24). Unsurprisingly, the correlation between spot and level order flow is much smaller, although still noticeable (around -0.15%).

5.3 Cross-impact models

We can now use the empirical estimates of Σ_{pp} and Ω_Ξ from the previous section to compute the cross-impact matrix Λ. For comparison purposes, we also introduce other cross-impact models. The first cross-impact model used for comparison

Electronic copy available at: https://ssrn.com/abstract=3779461

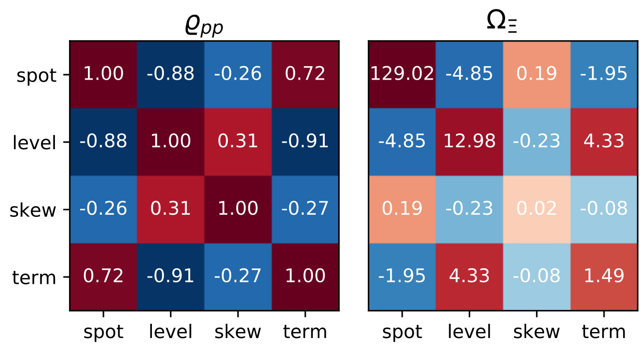

Figure 2: Empirical estimates of return correlation matrix ϱ and order flow covariance Ω.
The return correlation matrix ϱ (left) and the order flow covariance matrix Ω (right) estimates on our dataset. The order flow is reported in thousands of dollars of risk.

is the Black-Scholes cross-impact model introduced in Section 4.2 which has a single underlying: the spot. It is defined as

$$\Lambda_{\mathrm{bs}}(\boldsymbol{p}_t, t) := \frac{\sigma}{\sqrt{\boldsymbol{\Delta}_{\mathrm{c}}^\top \Omega \boldsymbol{\Delta}_{\mathrm{c}}}} \boldsymbol{\Delta}_{\mathrm{c}} \boldsymbol{\Delta}_{\mathrm{c}}^\top . \tag{21}$$

The Black-Scholes model coincides with the Kyle cross-impact model if all the liquidity is concentrated on the spot. In particular, this model is unable to account for changes in the volatility factors. We thus introduce the two-dimensional direct model $\Lambda_{\text{direct-2d}}$ which accounts for the spot and implied volatility factor but ignores cross-sectional effects, defined as

$$\Lambda_{\text{direct-2d}}(\boldsymbol{p}_t, t) := \frac{\sigma}{\sqrt{\boldsymbol{\Delta}_{\mathrm{c}}^\top \Omega \boldsymbol{\Delta}_{\mathrm{c}}}} \boldsymbol{\Delta}_{\mathrm{c}} \boldsymbol{\Delta}_{\mathrm{c}}^\top + \frac{\xi}{\sqrt{\boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top \Omega \boldsymbol{\mathcal{V}}_{\mathrm{c}}}} \boldsymbol{\mathcal{V}}_{\mathrm{c}} \boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top . \tag{22}$$

To account for all underlyings without correcting for cross-sectional effects, we introduce the four-dimensional direct model

$$\Lambda_{\text{direct-4d}}(\boldsymbol{p}_t, t) := \frac{\sigma}{\sqrt{\boldsymbol{\Delta}_{\mathrm{c}}^\top \Omega \boldsymbol{\Delta}_{\mathrm{c}}}} \boldsymbol{\Delta}_{\mathrm{c}} \boldsymbol{\Delta}_{\mathrm{c}}^\top + \frac{\xi}{\sqrt{\boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top \Omega \boldsymbol{\mathcal{V}}_{\mathrm{c}}}} \boldsymbol{\mathcal{V}}_{\mathrm{c}} \boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top + \sum_{i=3}^{Q+1} \sqrt{\frac{\Sigma_{pp}^{ii}}{\Omega_{\Xi}^{ii}}} \Xi^{\cdot i}(\boldsymbol{p}_t, t) (\Xi^{\cdot i}(\boldsymbol{p}_t, t))^\top . \tag{23}$$

Direct models ignore the off-diagonal structure of Σ_{pp} and Ω_Ξ. In particular they do not account for the leverage effect, which is an essential characteristic of the underlying return covariance matrix Σ_{pp}. To fix this, we introduce the two-dimensional Kyle cross-impact model Λ_{2d} which captures the two dominating underlyings of the system: the spot and level factor. Since Figure 2 shows that the delta and vega order flow correlation is small (around -0.15%) and $\xi\omega_{\mathcal{V}} \ll \sigma\omega_\Delta$, we can use Equation (20) to obtain the approximation

$$\Lambda_{\mathrm{2d}}(p_t, \varsigma_t, t) \approx \frac{\sigma}{\sqrt{\boldsymbol{\Delta}_{\mathrm{c}}^\top \Omega \boldsymbol{\Delta}_{\mathrm{c}}}} \boldsymbol{\Delta}_{\mathrm{c}} \boldsymbol{\Delta}_{\mathrm{c}}^\top + \frac{\xi\sqrt{1-\rho^2}}{\sqrt{\boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top \Omega \boldsymbol{\mathcal{V}}_{\mathrm{c}}}} \boldsymbol{\mathcal{V}}_{\mathrm{c}} \boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top + \frac{\xi\rho}{\sqrt{\boldsymbol{\Delta}_{\mathrm{c}}^\top \Omega \boldsymbol{\Delta}_{\mathrm{c}}}} (\boldsymbol{\mathcal{V}}_{\mathrm{c}} \boldsymbol{\Delta}_{\mathrm{c}}^\top + \boldsymbol{\Delta}_{\mathrm{c}} \boldsymbol{\mathcal{V}}_{\mathrm{c}}^\top). \tag{24}$$

The two-dimensional Kyle cross-impact model predicts that when trading options, one pushes the price in the amount of notional $\mathcal{V}$ traded divided by the typical $\mathcal{V}$ liquidity, which is compatible with findings from the meta-order study [18]. The prefactor $\sqrt{1-\rho^2} \approx 0.4$ is close to the empirical value of $0.33-0.4$ reported in [18]. Further, because of the leverage effect, trading delta-hedged portfolio impacts the spot price.

The full, four-dimensional Kyle cross-impact model Λ_{4d} is shown in Figure 3. Compared to the two-dimensional Kyle cross-impact model, it decouples the contribution of options on the level mode depending on the direction. This increases the explanatory power of the model, as shall be evidenced in Table 1.

5.4 Explanatory power of cross-impact models

For practical applications, a good cross-impact model should explain realized price changes from order flows. Thus to compare the models previously introduced, we now examine their explanatory power on empirical data. Given a realization

Electronic copy available at: https://ssrn.com/abstract=3779461

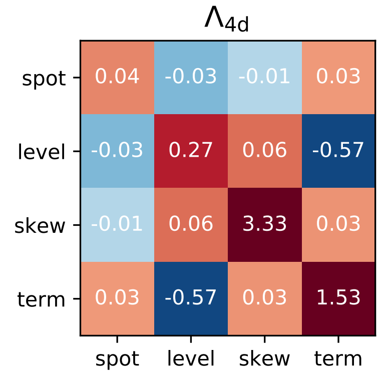

Figure 3: Four dimensional Kyle cross-impact model on options.
We report the four dimensional Kyle model estimated using empirical estimates of the covariances of Figure 2. The cross-impact matrix are reported in units of risk and in basis points so that Λ_{ij} encodes by how many basis points of volatility Asset i is pushed by trading one dollar of risk on Asset j.

of the underlying price process $(\delta \boldsymbol{p}_t)_{1\leq t\leq T}$ of length T, a corresponding series of predictions $(\widehat{\delta \boldsymbol{p}}_t)_{1\leq t\leq T}$ and a symmetric positive semi-definite matrix M, we introduce the generalized $R^2(M)$ error as

$$R^2(M) := 1 - \frac{\sum_{1\leq t\leq T}(\delta \boldsymbol{p}_t - \widehat{\delta \boldsymbol{p}}_t)^\top M(\delta \boldsymbol{p}_t - \widehat{\delta \boldsymbol{p}}_t)}{\sum_{1\leq t\leq T}\delta \boldsymbol{p}_t^\top M \delta \boldsymbol{p}_t}.$$

The matrix M is used to examine a model's predictive power for different portfolios. As the underlyings of our system are natural directions to consider, we report the $R^2(M)$ in Table 1 for $\Pi_{(1,0,0,0)} =: \Pi_{\text{spot}}$, $\Pi_{(0,1,0,0)} =: \Pi_{\text{level}}$, $\Pi_{(0,0,1,0)} =: \Pi_{\text{skew}}$ and $\Pi_{(0,0,0,1)} =: \Pi_{\text{term}}$.

Model	Scores			
	$R^2(\Pi_{\text{spot}})$	$R^2(\Pi_{\text{level}})$	$R^2(\Pi_{\text{skew}})$	$R^2(\Pi_{\text{term}})$
Λ_{bs}	0.18 ± 0.01	-0.00 ± 0.02	-0.00 ± 0.01	-0.00 ± 0.02
$\Lambda_{\text{direct-2d}}$	0.18 ± 0.01	-0.03 ± 0.02	-0.01 ± 0.01	0.00 ± 0.02
$\Lambda_{\text{direct-4d}}$	0.18 ± 0.01	-0.03 ± 0.02	-0.14 ± 0.02	-0.26 ± 0.02
Λ_{2d}	0.20 ± 0.01	0.12 ± 0.01	-0.01 ± 0.01	0.01 ± 0.02
Λ_{4d}	0.20 ± 0.01	0.14 ± 0.01	-0.12 ± 0.02	0.04 ± 0.01

Table 1: Scores of different cross-impact models.
All scores were computed in-sample using the same data used for the calibration of the cross-impact models.

Looking at $R^2(\Pi_{\text{spot}})$, we see that all models show similar scores on the spot, with cross-impact models being slightly better. Furthermore, there is no difference between Λ_{2d} and Λ_{4d}. This is consistent with the liquidity reported in Figure 6. Indeed, most of the liquidity is placed on the spot and the order flow traded on other factors is small in comparison. Models which only take into account the spot thus capture most of the order flow explanatory power. There is also a small advantage in using order flow on the level mode since Λ_{2d} and Λ_{4d} score better, but using term and skew order flow provides no improvement.

While using solely spot liquidity to explain spot returns is a good approximation, $R^2(\Pi_{\text{level}})$ shows the same is not true for the level underlying. Indeed, only models with spot-level cross-impact are able to account for returns of the implied volatility level. Therefore, to explain price changes on the level underlying we need to use order flow traded on the spot. This is natural as most of the traded order flow is on the spot but there is a high negative correlation between spot and level underlying (see Figure 6). Unfortunately, all models fail to explain skew returns. We suspect this comes from the low signal to noise ratio and low liquidity (in risk terms) of the skew underlying (see Figure 6).

On all metrics, Λ_{4d} performs at least as well as Λ_{2d}, which shows that the model is able to combine additional factors without suffering from noise. The additional underlying also help weigh trades appropriately on the implied volatility surface, which improves the $R^2(\Pi_{\text{level}})$ score.

Finally, we report the expected realized return conditional on the prediction of Λ_{4d} in Figure 4. This shows that, skew aside, Λ_{4d} provides a good fit for the realized returns of the different underlying as $\mathbb{E}_t[\delta \boldsymbol{p}_t \mid \delta \hat{\boldsymbol{p}}_t] \approx \delta \boldsymbol{p}_t$.

Electronic copy available at: https://ssrn.com/abstract=3779461

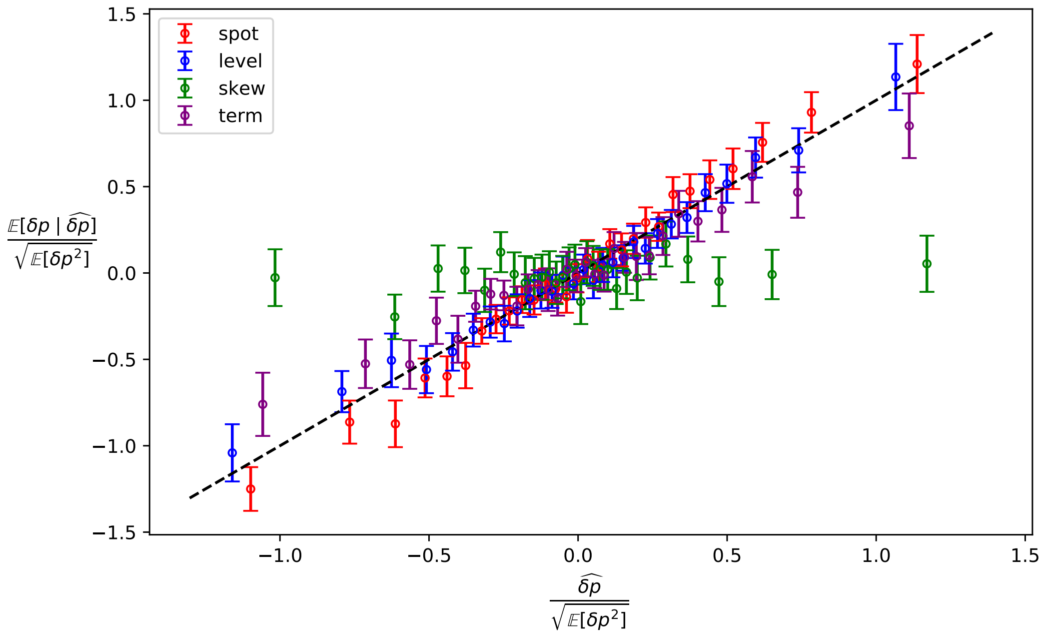

Figure 4: Predictions of the four-dimensional Kyle model on the main directions of the system.
We report the expected price change conditional on the predicted price change of the four-dimensional Kyle model for the four main directions of the system: in red for the spot, blue for the level, green for the skew and purple for the term structure. Predicted price changes and conditional averages are both normalized by the standard deviation of price changes along the given direction.

5.5 Evidence of cross-impact on options

Section 5.4 showed that only cross-impact models are able to explain returns for the level and term underlying. Aside from this explanatory power, this section tests their ability to explain other features of our data. To do so, we introduce the cross aggregate impact metric. The cross aggregate impact induced from the portfolio $\boldsymbol{u} \in \mathbb{R}^{N+M}$ on the return of the portfolio $\boldsymbol{v} \in \mathbb{R}^{N+M}$ is

$$\mathrm{Agg}_{\boldsymbol{u},\boldsymbol{v}}(x) := \mathbb{E}_t[\boldsymbol{v}^\top(\delta \boldsymbol{p}_t, \delta \boldsymbol{P}_t) \mid \boldsymbol{u}^\top(\delta \boldsymbol{q}_t, \delta \boldsymbol{Q}_t) = x]\,.$$

If returns are given by a linear cross-impact model Ψ and if we further assume $(\delta \boldsymbol{q}_t, \delta \boldsymbol{Q}_t)$ is a zero-mean Gaussian, then

$$\mathrm{Agg}_{\boldsymbol{u},\boldsymbol{v}}(x) = \mathbb{E}_t[\boldsymbol{v}^\top \Psi(\delta \boldsymbol{q}_t, \delta \boldsymbol{Q}_t) \mid \boldsymbol{u}^\top(\delta \boldsymbol{q}_t, \delta \boldsymbol{Q}_t)] := \mathrm{Agg}^{\Psi}_{\boldsymbol{u},\boldsymbol{v}}(x) = a_\Psi x\,,$$

where the slope a_Ψ depends on the cross-impact model Ψ and on the order flow covariance. Even in the absence of cross-impact, the presence of order flow correlations between two portfolios $\boldsymbol{u}$ and $\boldsymbol{v}$ may lead to a non-zero cross aggregate impact. Thus, to test whether there is cross-impact, we compare the empirically measured $\mathrm{Agg}_{\boldsymbol{u},\boldsymbol{v}}$ to the prediction $\mathrm{Agg}^{\Psi}_{\boldsymbol{u},\boldsymbol{v}}$ for different cross-impact models Ψ. We differentiate models between those which have no off-diagonal contributions $(\Lambda_{\text{bs}}, \Lambda_{\text{direct-4d}}, \Lambda_{\text{direct-2d}})$ and thus ignore cross-impact and those that take it into account $(\Lambda_{\text{2d}}, \Lambda_{\text{4d}})$.

We report $\mathrm{Agg}_{\boldsymbol{u},\boldsymbol{v}}$ in Figure 4 for different portfolios $\boldsymbol{u}, \boldsymbol{v}$ described in Table 2. Diagonal plots show aggregate direct impact. As expected, buying the E-Mini increases, on average, the price of the E-Mini as shown by the $\boldsymbol{u}, \boldsymbol{v}$ = spot plot (first row, first column). We see from the $\boldsymbol{u}, \boldsymbol{v}$ = level plot (second row, second column) that buying options and VIX futures increases, on average, the implied volatility. Furthermore, buying options and VIX futures decreases, on average, the E-Mini price as shown by the $\boldsymbol{u}$ = level, $\boldsymbol{v}$ = spot plot (first row, second column). This same plot, among others of Figure 4, shows that direct models provide a poor fit for cross aggregate impact. This suggests that the cross aggregate impact can only be explained by using a cross-impact model with off-diagonal elements, such as Λ_{4d}. Further, the fit is noticeably better for Λ_{4d} than Λ_{2d} which highlights the importance of taking into account the skew and term factors.

6 Conclusion

Let us summarize what we have achieved. Our main objective was to examine cross-impact when some instruments, derivatives, are deterministic functions of others, underlying. This posed modeling challenges on two fronts. First,

Electronic copy available at: https://ssrn.com/abstract=3779461

Name	Components			
	spot	VIX_0	VIX_1	options
spot	(1,	0,	0,	$0, \cdots, 0)$
level	(0,	β^{11},	β^{21},	$\beta^{31}, \cdots, \beta^{M1})$
skew	(0,	β^{12},	β^{22},	$\beta^{32}, \cdots, \beta^{M2})$
term	(0,	β^{13},	β^{23},	$\beta^{33}, \cdots, \beta^{M3})$

Table 2: Description of different directions used in this section.

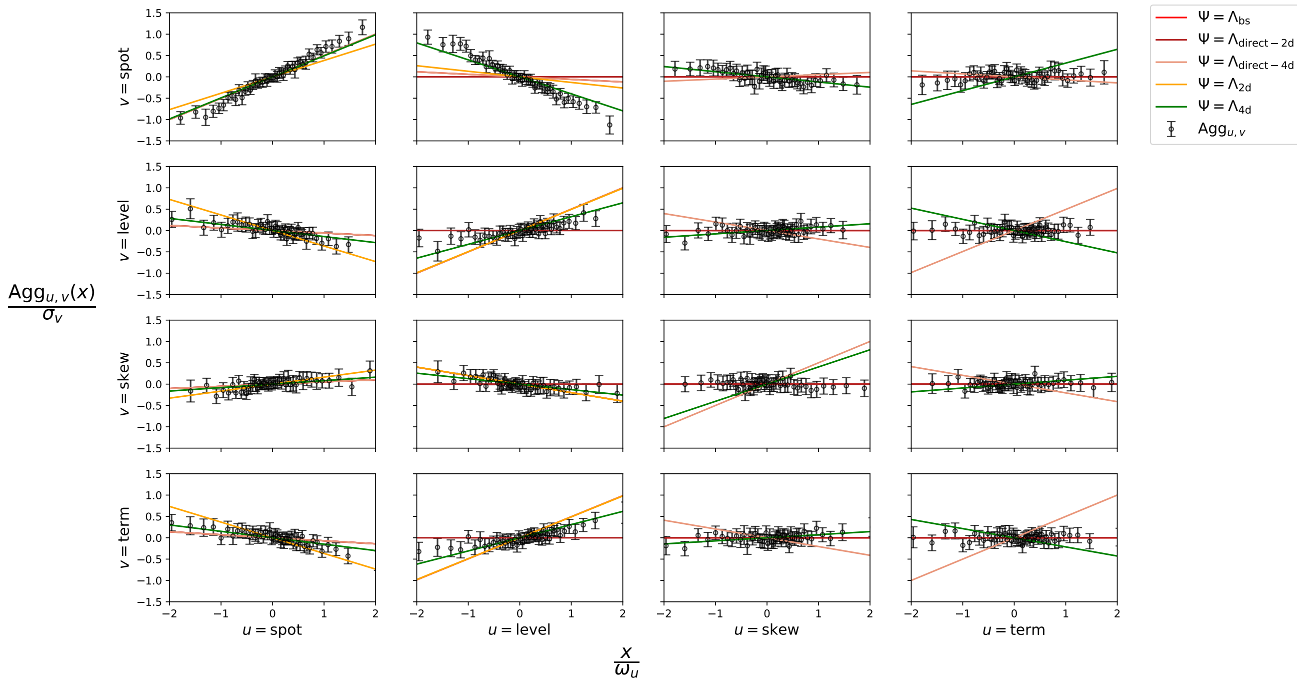

Figure 5: Normalized cross aggregate impact curves.
We report the cross aggregate cross impact curves for the spot, level, skew and term structure directions. Aggregate traded volumes are normalized by the typical deviations $\omega_u^2 := \mathbb{E}_t[(\boldsymbol{u}^\top(\delta \boldsymbol{q}_t, \delta \boldsymbol{Q}_t))^2]$ and portfolio returns by the typical deviations $\sigma_v^2 := \mathbb{E}_t[(\boldsymbol{v}^\top(\delta \boldsymbol{p}_t, \delta \boldsymbol{P}_t))^2]$. Estimated cross aggregate impact $\text{Agg}_{\boldsymbol{u},\boldsymbol{v}}$ is reported along with predicted cross aggregate impact $\text{Agg}^{\Psi}_{\boldsymbol{u},\boldsymbol{v}}$ for different choices of linear cross-impact models Ψ.

derivative prices with impact must be locked-in by no-arbitrage. Second, the liquidity of derivative instruments may be very small and heavily non-stationary.

Leveraging the results of [16], we introduced the Kyle cross-impact model on derivatives. We showed in Section 3.3 that this model (i) prevents arbitrage, (ii) provides impact dynamics which can be factored to recover frictionless dynamics, (iii) aggregates traded order flow to a few liquidity pools, (iv) is well-behaved even if some instruments are highly illiquid, and (v) can be applied without specifying which instruments are derivatives. The Kyle cross-impact model is thus theoretically satisfying and practical for applications. This justifies its use compared to other cross-impact models (see [16] for examples of other models).

To stress-test our framework on empirical data, we used the front-month E-Mini future, E-Mini vanilla options and VIX futures. Despite our simplistic approach to model implied volatility dynamics, we showed (see Figure 5 and Table 1) that the Kyle model better explains returns than models which ignored cross-impact, and that the Kyle model can be improved by more precise modeling of the implied volatility surface. This points at the effectiveness of the proposed framework to aggregate liquidity and suggests more sophisticated implied volatility modeling may further improve results. Aside from explanatory power, cross-impact models are consistent with some empirical observations noted in [13, 18]. We provided evidence of cross-impact in derivative markets by studying price responses conditional on traded order flow and showing models which ignore cross-impact are unable to provide an adequate fit (see Figure 5).

On both theoretical and empirical grounds, the simple static, linear framework presented here thus accounts for important properties of impact on derivative markets. The methodology is flexible and can be readily adapted to handle

Electronic copy available at: https://ssrn.com/abstract=3779461

complex price dynamics and exotic derivatives. The cross-impact estimates may be used in practice for estimating execution costs, in particular hedging costs.

While this work provides a first proof of concept which captures many features of cross-impact on derivatives, it fails to capture the auto-correlations of order flows. A multi-period cross-impact model, apart from providing more accurate descriptions of cross-impact, would yield insight into the cross-impact of meta-orders, another topic left unexplored to this date and particularly difficult to measure on options. We leave this extension to future work since it raises new challenges. For example, multi-period, cross-sectional arbitrages yield new constraints for the chosen cross-impact model. Furthermore, to remain tractable, such models need to find the proper way to aggregate traded order flow in liquidity pools.

7 Acknowledgments

We warmly thank J.-P. Bouchaud, Z. Eisler and B. Toth and M. Rosenbaum for fruitful discussions. This research was conducted within the *Econophysics & Complex Systems* Research Chair under the aegis of the *Fondation du Risque*, a joint initiative by the *Fondation de l'École polytechnique, l'École polytechnique* and Capital Fund Management.

References

[1] Aurélien Alfonsi, Florian Klöck, and Alexander Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Mathematics of operations research*, 41(3):914–934, 2016.

[2] Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.

[3] Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Hedging of covered options with linear market impact and gamma constraint. *SIAM Journal on Control and Optimization*, 55(5):3319–3348, 2017.

[4] Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, Quotes and Prices*. Cambridge University Press, 3 2018.

[5] Jordi Caballe and Murugappa Krishnan. Imperfect competition in a multi-security market with risk neutrality. *Econometrica (1986-1998)*, 62(3):695, 1994.

[6] Luis Carlos Garcia del Molino, Iacopo Mastromatteo, Michael Benzaquen, and Jean-Philippe Bouchaud. The multivariate kyle model: More is different. *SIAM Journal on Financial Mathematics*, 11(2):327–357, 2020.

[7] Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

[8] Joel Hasbrouck and Duane J Seppi. Common factors in prices, order flows, and liquidity. *Journal of financial Economics*, 59(3):383–411, 2001.

[9] Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.

[10] Gregoire Loeper and others. Option pricing with linear market impact and nonlinear Black–Scholes equations. *The Annals of Applied Probability*, 28(5):2664–2726, 2018.

[11] Iacopo Mastromatteo, Michael Benzaquen, Zoltán Eisler, and Jean-Philippe Bouchaud. Trading Lightly: Cross-Impact and Optimal Portfolio Execution. 2017.

[12] Paolo Pasquariello and Clara Vega. Strategic cross-trading in the US stock market. *Review of Finance*, 19(1):229–282, 2015.

[13] Emilio Said. How option hedging shapes market impact. *Available at SSRN 3470915*, 2019.

[14] Michael Schneider and Fabrizio Lillo. Cross-impact and no-dynamic-arbitrage. *Quantitative Finance*, 19(1):137–154, 2019.

[15] Damian Eduardo Taranto, Giacomo Bormetti, Jean-Philippe Bouchaud, Fabrizio Lillo, and Bence Tóth. Linear models for the impact of order flow on prices. I. History dependent impact models. *Quantitative Finance*, 18(6):903–915, 2018.

Electronic copy available at: https://ssrn.com/abstract=3779461

[16] Mehdi Tomas, Iacopo Mastromatteo, and Michael Benzaquen. How to build a cross-impact model from first principles: Theoretical requirements and empirical results. *arXiv preprint arXiv:2004.01624*, 2020.

[17] Nicolo Torre. BARRA market Impact model handbook. *BARRA Inc., Berkeley*, 1997.

[18] Bence Tóth, Zoltán Eisler, and J-P Bouchaud. The square-root impact law also holds for option markets. *Wilmott*, 2016(85):70–73, 2016.

[19] Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Price response in correlated financial markets: empirical results. *arXiv preprint arXiv:1510.03205*, 2015.

Electronic copy available at: https://ssrn.com/abstract=3779461

A Data

This section gives motivation about our choice of instruments, details on the data and methodology which were omitted in the main text for conciseness.

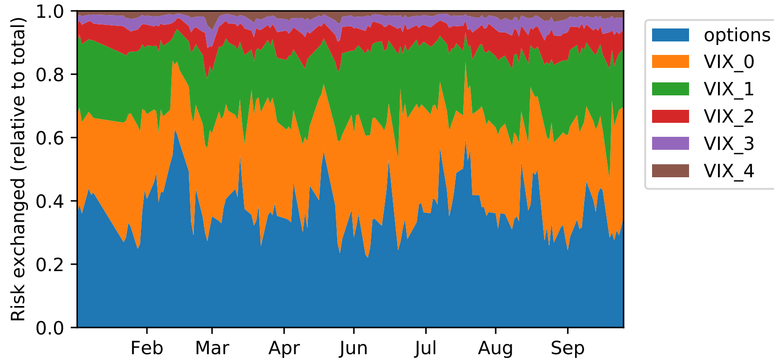

Figure 6: Distribution of liquidity among VIX futures and options.

Choice of instruments To stress-test our approach, we sought an actively traded derivative market with many derivatives. Thus we considered E-mini vanilla options and their underlying (both quoted on the CME), the front-month futures contract. However, a large fraction of the traded risk in derivatives comes from VIX futures (quoted on the CBOE) as shown in Figure 6. The VIX index is computed using options with maturities between 23 and 37 days and is meant to track the level of the implied volatility for options expiring in one month. Thus, because of their liquidity and close relationship with the implied volatility of options, order flow traded on VIX futures play an important role and should not be ignored.

Filtering instruments Given the very large number of options quoted on the market, we kept options within a given range of strikes and maturities to limit the size of the data set.

Resolution and time frame Our dataset contains the trades and quotes of all previously selected products, at the five minute time scale, from January 2019 to September 2019. This time frame was chosen because of the large level of noise on derivatives' prices and the size of the data set which encumbered analysis. In a given five minute bin, signed trades were aggregated on their volumes, so that we have the opening and closing prices of instruments along with the aggregated signed traded order flow. We considered hours where both options and their underlying are liquid, further removing 30 minutes around opening and closing for stationarity purposes. Doing so, data ranges between 3PM and 8:30PM UTC.

Implied volatility and greeks We now explain how implied volatility and Greeks were computed. For a given day and for a particular option, we have access to the opening bid and ask prices of that option for each five minute window. Furthermore, the bid and ask Black-Scholes implied volatilities are computed using the bid and ask price of the option and the price of the E-mini future contract with closest maturity. Correspondingly, the usual Black-Scholes Greeks Δ and $\mathcal{V}$ are computed for both the bid and ask sides. In our analysis, we use the mid of bid and ask quantities (option prices, implied volatilities and greeks) to perform computations.

Electronic copy available at: https://ssrn.com/abstract=3779461