Efficient Trading with Price Impact

Xavier Brokmann[1], Lukas Gonon[2,3], Guangyi He[2], David Itkin[4], and Johannes Muhle-Karbe*[2]

[1]Optiver
[2]Imperial College London
[3]University of St. Gallen
[4]London School of Economics and Political Science

November 2, 2026

Abstract

This paper studies how to compute trading strategies that trade off expected returns, risk, and the trading costs generated by price impact in an efficient manner. For a general model where impact is nonlinear and decays gradually, we show that the performance of linear feedback policies can still be computed in closed form. These are optimal for linear impact and linear decay dynamics only. However, when the linear impact parameters are optimized for the nonlinear impact model at hand, then the performance of the resulting policy is competitive with state-of-the-art machine-learning methods. The gap in performance becomes significant only when the market is so illiquid that neither of the strategies is attractive net of trading costs.

1 Introduction

A crucial challenge in the implementation of algorithmic trading strategies is to mitigate the adverse price impact that the corresponding trades generate. If this is not done in an efficient manner, then impact costs quickly turn seemingly profitable strategies into losing bets.

A rich empirical literature documents the subtle features of price impact. One key finding is that its magnitude does not depend linearly on order size, but is instead more accurately described by a "square-root law" (Loeb, 1983; Hasbrouck, 1991). Moreover, impact is also not permanent, but gradually decays as liquidity recovers following a large order (Hasbrouck, 1991; Biais et al., 1995). A sizeable fraction of this decay happens quickly, whereas another component is surprisingly persistent, a multiscale behaviour consistent with a power law (Bouchaud et al., 2009).[1]

Incorporating these empirical features into optimal trading is a challenging task for which analytical results are only available in special cases, e.g., when impact is linear (Gârleanu and Pedersen, 2013, 2016; Collin-Dufresne et al., 2020; Abi Jaber and Neuman, 2025), impact is small (Guasoni and Weber, 2020; Cayé et al., 2020), or when there are no risk constraints (Hey et al., 2025). In contrast, no efficient and interpretable trading rules are available that account for the full range of salient empirical properties.

As a consequence, researchers and practitioners alike have turned to numerical methods to account for price impact in more general optimal trading problems. Modern machine learning methods open the door to tackling complex problems of this kind, but the results are often difficult to interpret and lack simple, robust benchmarks. In the present paper, we propose to use the class of linear feedback controls as a benchmark. These strategies are optimal only when price impact is linear and decays on a single exponential time scale (Gârleanu and Pedersen, 2016). If one simply plugs in point estimates for the price impact parameters, then this leads to substantial performance losses relative to the optimal strategy that takes into account the nonlinear nature of price impact (Hey et al., 2024). Whence, we instead view the

*Corresponding author, j.muhle-karbe@imperial.ac.uk.

[1]For a detailed account of the more recent empirical literature, we refer to the textbooks Bouchaud et al. (2018); Webster (2023) and the references therein.

linear impact level and decay rate as tuning parameters that need to be chosen to optimize performance in a more realistic price impact model.

Moallemi and Sağlam (2017) propose an approach in this spirit in a very general context, enabling one to deploy efficient methods from convex optimization. Brokmann et al. (2025) focus on models with impact that decays instantaneously as in Almgren et al. (2005), where each trade only affects its own execution price but not subsequent ones. They show that the performance of linear strategies can then be computed in closed form, so that no numerical optimization is required. In this context, linear policies perform surprisingly well, but it remains unclear to what extent this translates to state-of-the-art models with gradual impact decay.

In the present study, we develop results of this kind for general price impact models, where impact can be of an arbitrary concave form and a general decay kernel describes impact decay. This impact model nests virtually all specifications from the empirical literature as special cases, and does not admit price manipulation for any log-convex price impact decay kernel. Even at this level of generality, it turns out that the performance of linear feedback strategies can still be computed explicitly. With this closed-form expression for the performance of an arbitrary linear policy at hand, the tuning parameters can be optimized in a straightforward manner to obtain optimal "effective" values of the linear price impact and decay parameters.

The resulting trading rules are available in closed form, making them easy to analyze and interpret. However, this analysis raises the key question: how much performance is sacrificed by restricting to the subclass of linear policies? To assess this, their performance needs to be compared to numerical methods that solve for the globally optimal trading policy. In Brokmann et al. (2025), the Viterbi algorithm of Kolm and Ritter (2014) is used for this purpose in models with instantaneous impact. This approach can also be adapted to the present context with gradual impact decay, but achieving the precision necessary to improve on linear policies typically leads to infeasibly long runtimes. We therefore, instead, focus on strategies parametrized by a suitable family of neural networks, which we optimize using stochastic gradient descent-type algorithms and backpropagation, akin to classical policy gradient methods in reinforcement learning (Sutton and Barto, 2018).[2] Using a neural network architecture specifically tailored to the price impact setting, the resulting algorithm matches the results of the Viterbi algorithm run at very high accuracy in a small fraction of the runtime and comes close to an upper duality bound for our goal function (cf. Appendix E for more details).

In a numerical case study, we then compare the performance of the optimized linear and neural network strategies. As a baseline case, we first focus on the model of Alfonsi et al. (2010), where impact has a square-root form and decays at an exponential rate. The parameters for this experiment are calibrated to resemble realistic market regimes faced by mid-to-high capacity trading firms. We find that neural network strategies capture about 48.7% of a gross Sharpe ratio of 3 after accounting for realistic trading costs, whereas the corresponding figure for the optimal linear strategy is 47.0%. Put differently, neural networks improve the Sharpe ratio net of costs by about 3.5%. This is a larger improvement compared to the 1-2% observed for models with instantaneous impact by Brokmann et al. (2025), but still modest compared to substantial estimation errors for trading signals and price impact parameters(cf., e.g., (Bouchaud and Potters, 2000, Chapter 10) or Harvey and Liu (2015)).[3] Indeed, a common baseline is to apply a haircut of 50% when converting in-sample to out-of-sample Sharpe ratios, potentially adjusted depending on signal strength and size of the training dataset.

To illustrate this point, we also compare the performance of the linear and neural network strategies optimized for a single decay timescale in a model that combines three different exponential decay kernels fit using the same dataset (Hey et al., 2025). We observe that out-of-model, the already modest performance gap is reduced further and drops from 3.5% to well below 1%, illustrating the robustness of the simple and interpretable linear policies.

Due to the long time horizons, training neural network strategies for the fully path-dependent dynamics implied by power-law decay kernels requires significantly more effort. In contrast, linear strategies

[2]Similar methods have been used to optimize hedging strategies in the presence of instantaneous trading costs by Buehler et al. (2019); Kolm and Ritter (2019), for example. Other machine learning methods for optimal trading with price impact have been proposed by Moallemi and Wang (2022); Leal et al. (2022); Cartea et al. (2023); Chen et al. (2024), for example. The contemporaneous learning of impact parameters and optimal trading strategies is studied for linear price impact models by Park and Van Roy (2015); Neuman and Zhang (2026), for example.

[3]Bigger improvements can be obtained using neural networks when the strategy is near capacity because i) trading costs are high, ii) impact decay is slow, or iii) price forecasts ("alpha signals") are not accurate or persistent enough. However, in most of these scenarios, trading costs degrade the performance so severely that even optimized strategies are not attractive anymore. Conversely, in most scenarios where the strategy does remain attractive net of costs, the extra performance that can be obtained by adding nonlinear features to the optimal trading rules is small. The only notable exception is the case of very slow impact decay.

can still be tuned in the same manner as before. We find that shifted power law and multi-exponential decay kernels with comparable decay profiles lead to similar optimal linear trading strategies, although the Sharpe ratios they achieve are influenced by the precise choice of decay kernel. This again highlights the robustness of the linear feedback strategies.

In summary, our analytical results for linear feedback policies are straightforward to apply and lead to simple trading rules, whose performance almost exactly matches those of state-of-the-art machine-learning methods in profitable trading environments. Incorporating nonlinear features is computationally feasible with modern architectures and computing power but, in our experiments, only leads to significant performance gains when liquidity is too low to render any strategy attractive net of trading costs.

These conclusion corroborate recent results from the machine-learning literature, where parametric families of linear policies have also been found to be competitive for a number of continuous-time control tasks (Fazel et al., 2018; Rajeswaran et al., 2017; Mania et al., 2018). This builds on a long tradition in the automation and control literature, ranging from the stochastic linearization of nonlinear system components (Booton, 1954) to using linear policies as approximately optimal controls (Levine and Athans, 1970; Makila and Toivonen, 1987).

The remainder of this article is organized as follows. Section 2 introduces our general price impact model and shows how it nests the typical specifications from the literature. Section 3, in turn, provides sufficient conditions for the absence of price manipulation and formulates the tradeoff between expected returns, trading costs, and risk. For linear price impact and exponential impact decay, the model simplifies to the setup of Gârleanu and Pedersen (2016) and the linear feedback policies that are optimal in this special case are summarized in Section 4. The performance of linear strategies in our general model with nonlinear impact and impact decay is analyzed in Section 5, and compared to the performance of neural network strategies in Section 6. Section 7 concludes and Appendices A and B contain the derivations of the goal functional and the performance of linear strategies, respectively. Appendix C describes the neural-network approach that we use for the numerical optimization of our goal functional. Finally, Appendix D compares the neural network strategies to strategies optimized using the Viterbi algorithm, as in Brokmann et al. (2025), while Appendix E establishes upper duality bounds for the risk-adjusted P&L and uses these to demonstrate near optimality of the neural-network strategies.

2 Price Impact Model

2.1 Unaffected Prices

We consider two assets: the first one is safe (and its price is normalized to one); the second one is risky in that its *unaffected price* has dynamics

$$dS_t = \alpha f_t \, dt + \sigma dB_t. \tag{2.1}$$

Here, B_t is a standard Brownian motion driving the price shocks with volatility $\sigma > 0$. Future returns are partially predictable through the *trading signal* f_t; the parameter $\alpha > 0$ regulates the accuracy of these predictions. For tractability, we assume as in many related studies that the predictor has linear mean-reverting dynamics:

$$df_t = -\phi f_t \, dt + \sqrt{2\phi} dW_t, \tag{2.2}$$

for another standard Brownian motion W_t independent of B_t.[4] The mean reversion and volatility parameters of the signal are chosen so that its stationary distribution is normalized to a Gaussian with mean zero and unit variance; the parameter $\phi > 0$ in turn controls the half-life of the signal.

2.2 Nonlinear Price Impact and Impact Decay

When large trades are executed quickly, they are not settled at the unaffected price (2.1) but instead incur price impact. Empirically, this price impact is a nonlinear function of the trade size and decays over several different timescales – a sizeable part vanishes rapidly, while another component lingers for a surprisingly long time. We capture both of these effects in a general price impact model, which nests the typical specifications from the empirical literature as special cases; see Section 2.3 for a range of examples.

[4] This means that the predictor f_t is optimal in that it is independent from the noise B_t. If the signal and noise were cross-correlated in time then one could construct a better predictor f_t.

To wit, we denote by Q_t the quantity of the risky asset held by a large trader at time t, and write $P_t = S_t + I_t$ for the *market price* of the risky asset. It is composed of the *unaffected price* S_t (which models price changes caused by other traders or news, for example) and the *price impact* generated by the trader's current and past transactions dQ_s, $s \leq t$:

$$I_t = \sum_{n=1}^{N} h^n(J_t^n), \quad \text{where} \quad J_t^n = \int_0^t K_n(t-s)dQ_s. \tag{2.3}$$

Here, each *impact state* J_t^n, $n = 1, \ldots, N$ is a different weighted average of current and past trades, for which the impact of past transactions decays according to a *propagator kernel* $K_n(t-s)$. As in Alfonsi et al. (2010), price impact is a nonlinear function $h^n(J_t^n)$ of these weighted averages, capturing concave relationships.[5] Considering linear combinations of several such models in (2.3) allows us to incorporate different impact concavities at different decay timescales. For example, a permanent impact component can be modelled linearly, whereas impact states with faster decay can enter in a nonlinear fashion.

2.3 Examples

We now illustrate with a range of examples how the general framework from the previous section naturally includes the typical specifications from the empirical and optimal trading literatures.[6]

The linear model of Obizhaeva and Wang (2013) The analytically most tractable specification is to assume that the impact function

$$h(x) = \lambda x$$

is linear and the impact decay kernel is exponential:

$$K(t) = e^{-\beta t}.$$

In this case, J_t is an exponentially weighted moving average satisfying $dJ_t = -\beta J_t dt + dQ_t$, and the corresponding impact state dynamics are linear. This allows one to solve a wide range of optimal trading problems in closed form using the machinery of linear-quadratic stochastic control (see Section 4 for a summary and, e.g., Gârleanu and Pedersen (2013, 2016) or the textbook Cartea et al. (2015) for more details).

As the impact decay β tends to zero, price impact in the model of Obizhaeva and Wang (2013) becomes permanent ($I_t = \lambda Q_t$). Conversely, when the impact decay rate tends to infinity (and the impact level λ is rescaled accordingly), then the price impact becomes instantaneous (i.e., $I_t = \lambda \dot{Q}_t$ only depends on the current trading rate for smooth strategies $dQ_t = \dot{Q}_t dt$ and blows up otherwise, see Gârleanu and Pedersen (2016) for more details). Whence, these two limiting regimes correspond to the model of Almgren and Chriss (2000), another workhorse of the price impact literature.

The nonlinear model of Alfonsi, Fruth, and Schied (2010) Retaining the exponential decay kernel of Obizhaeva and Wang (2013) but considering a general impact function h leads to the model of Alfonsi et al. (2010). This allows the model to incorporate the nonlinear dependence of impact on trade sizes, with a square-root function, corresponding to $p = 1/2$ below, typically providing the best fit among all power laws (cf., e.g., Hey et al. (2024, 2025) and the references therein):

$$h(x) = \lambda \operatorname{sign}(x)|x|^p, \qquad \lambda > 0, \quad p \in (0,1).$$

When impact decay β becomes faster and faster (and the impact level λ is rescaled accordingly), price impact in the model of Alfonsi et al. (2010) converges to a nonlinear function $h(\dot{Q}_t)$ of the current trading speed. This limiting regime in turn recovers the model of Almgren (2003).

The model of Alfonsi et al. (2010) retains analytical tractability for risk-neutral optimization problems (Alfonsi et al., 2010; Hey et al., 2025, 2024), but there are no known analytical results for traders that are not only capacity but also risk constrained.

[5] A similar but different effect is the concavity of the price impact of individual trades. Bouchaud et al. (2009) provide an overview of the empirical literature on such "local concavities"; optimal trading in this context is studied by Curato et al. (2017); Muhle-Karbe et al. (2023). The present study instead focuses on "global concavities" at the metaorder level.

[6] To ease notation, we simply write h, J, and K without the superscript n when dealing with a single impact function, state variable, or kernel, respectively.

General decay kernels Just like impact depends on trade sizes in a nonlinear fashion, its decay also does not happen on a single exponential time scale (Bouchaud et al., 2009; Bacry et al., 2015; Brokmann et al., 2015; Bucci et al., 2018). One way to incorporate this is to consider several exponential kernels

$$K_n(t) = e^{-\beta_n t}, \quad n = 1, \dots, N,$$

with different decay rates $\beta_n > 0$ as in Hey et al. (2025). For example, with one very fast and one very slow timescale, this specification nests, as a limiting case, the model of Almgren et al. (2005) with nonlinear instantaneous and linear permanent impact. Exponential kernels have the advantage of retaining Markovian price impact dynamics (albeit with each J^n becoming an additional state variable), but require fitting additional parameters for each extra decay timescale considered. A more parsimonious parametric model class is the class of *shifted power law kernels* of Bouchaud et al. (2004),

$$K(t) = \frac{1}{(\epsilon + t)^\nu},$$

for a *decay elasticity* parameter $\nu \in (0, 1)$ and a *shift* parameter $\epsilon > 0$. Compared to exponential kernels, a single power law can capture several fast and slow decay timescales. However, shifted power law kernels induce non-Markovian impact dynamics where the evolution of the current impact state truly depends on the entire history of trades and not just a few sufficient statistics. For linear price impact functions, some optimal trading results can still be obtained (Gatheral et al., 2012; Abi Jaber and Neuman, 2025), but optimal trading with nonlinear impact and such general decay kernels requires sophisticated numerical methods (Abi Jaber et al., 2025).

3 Goal Functional for Optimal Trading

We now formulate a suitable goal functional that trades off expected returns, risk, and trading costs due to price impact. To this end, we first discuss the trader's P&L net of trading costs for smooth trading strategies[7] and how to complement this with a suitable risk penalty. Then, we show how to reformulate the resulting goal functional in a form that easily extends to general strategies, and will form the basis of our subsequent analytical and numerical results.

Formulating the goal functional in a consistent manner for general strategies is not just a technical subtlety but crucially important to obtain stable numerical results. To wit, if bulk trades are not penalized consistently with their smooth approximations, then naive discretizations of the model typically admit "price manipulation". Sufficiently flexible numerical solvers (e.g., based on neural networks) in turn learn to exploit these inconsistencies in the model rather than the trading signal they are supposed to implement.

3.1 The Trader's PnL

We now consider the trader's gains and losses from trading when price impact is given by (2.3). For simplicity, we assume all strategies satisfy $Q_0 = 0$. We start by deriving the PnL of smooth strategies $Q_t = \int_0^t \dot{Q}_s ds$. As these only execute small orders of order $O(dt)$ at each time t, the corresponding execution costs are $P_t \dot{Q}_t dt + O(dt^2)$, because the additional impact of each trade on the current market price $P_t = S_t + I_t$ is of order $O(dt)$.[8] On a finite time horizon T, the balance on the trader's cash account is then $-\int_0^T (S_t + I_t) dQ_t$.

To value the corresponding terminal position Q_T, we use the unaffected price S_T.[9] The trader's PnL from trading until time T is $\mathrm{PnL}_T = Q_T S_T - \int_0^T (S_t + I_t) dQ_t$. With the integration by parts $Q_T S_T = \int_0^T S_t dQ_t + \int_0^T Q_t dS_t$ this can be rewritten as

$$\mathrm{PnL}_T = \int_0^T Q_t dS_t - \int_0^T I_t \dot{Q}_t dt.$$

[7] That is, holdings that do not change rapidly due to block trades or diffusive fluctuations.

[8] In contrast, for holdings with a nontrivial Brownian motion component, both trade sizes and additional price impact are of order $O(dt^{1/2})$ and therefore lead to an additional correction term of order $O(dt)$ as in Gârleanu and Pedersen (2016). We sidestep dealing with these terms and how to specify where block trades are executed between the pre- and post-trade prices by focusing on smooth strategies first. We then reformulate the resulting goal functional in a form in which it can readily be extended to general strategies.

[9] In contrast to using the market price $P_T = S_T + I_T$, this avoids "illusory profits" based on inflating the market price with one's own impact.

Here, the first term is the usual stochastic integral that describes the gains and losses from changes of the unaffected price. The second term, in turn, records the additional effects of price impact on the trader's PnL. To extend this representation to strategies that are not smooth, we rewrite it in a form that does not depend on the trading rate $\dot{Q}_t = dQ_t/dt$ anymore. To precisely state the resulting PnL equation we first introduce the following mild assumption on the impact functions and decay kernels:

Assumption 3.1. *For each $n = 1, \dots, N$ we assume that:*

(i) *The price impact function $h^n : \mathbb{R} \to \mathbb{R}$ is continuous, nondecreasing and satisfies $h^n(0) = 0$;*

(ii) *The decay kernel $K_n : [0, \infty) \to (0, \infty)$ is twice continuously differentiable, nonincreasing and is such that $\log K_n$ is convex.*

All of the examples in Subsection 2.3 satisfy Assumption 3.1.

Proposition 3.2 (PnL Equation)**.** *Under Assumption 3.1, the investor's PnL when trading a smooth strategy $dQ_t = \dot{Q}_t dt$ over the time interval $[0, T]$ is given by*

$$\mathrm{PnL}_T(Q) = \int_0^T Q_t dS_t - \sum_{n=1}^N \frac{H^n(J_T^n)}{K_n(0)} + \sum_{n=1}^N \int_0^T h^n(J_t^n) \left(\frac{K_n'(0)}{K_n(0)} Q_t + \int_0^t \frac{K_n''(t-s)}{K_n(0)} Q_s ds \right) dt, \tag{3.1}$$

where $H^n(x) = \int_0^x h^n(y)dy$ is the antiderivative of $h^n(x)$.

The proof of this result is delegated to Appendix A. Here, we just note that no integrals against the trades dQ_t appear in (3.1) anymore, so this representation readily extends to any strategy that can be approximated by smooth strategies; in particular, we define via (3.1) the PnL of any strategy in the class

$$\mathcal{Q} = \left\{ Q : \begin{array}{c} Q \text{ is } \mathcal{F}^{B,W}\text{-predictable, } \int_0^T \mathbb{E}[Q_t^2] dt < \infty \text{ for every } T > 0 \\ \text{and } \lim_{T \to \infty} \frac{1}{T} \mathbb{E}[Q_T^2 + \sum_{n=1}^N |H^n(J_T^n)|] = 0 \end{array} \right\}, \tag{3.2}$$

where $\mathcal{F}^{B,W}$ is the augmentation of the filtration generated by the Brownian motions B and W. In Appendix B we verify that the linear feedback controls considered in Sections 4 and 5 are members of $\mathcal{Q}$ under the mild Assumptions 3.1 and 5.1 on the kernel and impact functions.

3.2 Absence of Price Manipulation

Price manipulation refers to round trip trades with negative expected trading costs. If such trading opportunities exist in a given model, then profits can be generated not by exploiting forecasts about future exogenous price changes, but by suitable combinations of purchases and sales that benefit rather than suffer from the price impact they generate. Even when they exist, such manipulation strategies typically depend very sensitively on the model at hand and are unlikely to be effective when actually deployed in live trading. Ruling them out therefore is an important consistency condition that any price impact model should satisfy to be used for optimization.

We now discuss how Assumption 3.1 (and, in particular, the log convexity of the decay kernel) rules out price manipulation in the context of the present paper via a classical result of Gripenberg (1979). Here, we say that a trading strategy Q achieves *price manipulation* if $Q_0 = Q_T = 0$ and $\mathbb{E}[\int_0^T Q_t dS_t - \mathrm{PnL}_T(Q)] < 0$ on some time horizon $[0, T]$. Since $\int_0^T Q_t dS_t$ represents gains and losses due to fundamental price changes, the difference between this term and the PnL (net of costs) precisely picks up the expected cost paid when executing a round trip trading strategy. From (3.1), we see that this difference is explicitly given in terms of the holdings Q, impact state processes J^n and the price impact kernel K_n and impact function h^n.

Proposition 3.3 (Absence of price manipulation)**.** *Suppose Assumption 3.1 holds. Then for any time horizon $T \geq 0$ and any $Q \in \mathcal{Q}$ satisfying $Q_0 = Q_T = 0$ we have $\mathbb{E}[\int_0^T Q_t dS_t - \mathrm{PnL}_T(Q)] \geq 0$.*

The proof of this result is provided in Appendix A. As an illustrative example, we now consider a parametric family of shifted power law kernels. For this specification, we can obtain a necessary and sufficient condition for the absence of price manipulation.

Example 3.4 (Power impact and generalized power-law decay)**.** *Consider*

$$h(x) = \lambda \operatorname{sign}(x)|x|^p, \qquad K(t) = (\epsilon + t^q)^{-\nu}, \qquad \lambda, \epsilon, \nu, p, q > 0.$$

The kernel is twice continuously differentiable on $[0,\infty)$ *precisely when* $q = 1$ *or* $q \geq 2$. *Among these kernels, there is no price manipulation if and only if*

$$q = 1 \qquad or \qquad q = 2 \text{ and } p = 1. \tag{3.3}$$

Thus the boundary case $q = 2$ *excludes manipulation under linear impact, but admits manipulation for every nonlinear power impact function, including square-root impact. In contrast, the specification* $q = 1$ *used in our numerical examples in Section 6 is manipulation free for any choice of price impact function. A derivation of the condition* (3.3) *is provided in Appendix A.*

3.3 Inventory Costs and Risk Constraints

To simplify the exposition as much as possible, we focus on the long-run average PnL per unit time $\lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\mathrm{PnL}_T(Q)]$. After plugging in the dynamics (2.1) of the unaffected price, we see by the transversality conditions in the definition of $\mathcal{Q}$ that this is given by

$$\begin{aligned}\mathrm{PnL}(Q) &= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\mathrm{PnL}_T(Q)] \\ &= \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}\left[\alpha f_t Q_t + \sum_{n=1}^N h^n(J_t^n)\left(\frac{K_n'(0)}{K_n(0)}Q_t + \int_0^t \frac{K_n''(t-s)}{K_n(0)}Q_s ds\right)\right] dt.\end{aligned} \tag{3.4}$$

The most tractable way to penalize large inventories when optimizing the expected PnL (3.4) is to include a running cost $-\frac{\gamma\sigma^2}{2}Q_t^2 dt$ on the trader's squared positions as in Gârleanu and Pedersen (2013, 2016), where $\gamma > 0$ is a risk aversion parameter. The corresponding mean-variance goal functional is

$$\mathcal{J}(Q) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}\left[\alpha f_t Q_t + \sum_{n=1}^N h^n(J_t^n)\left(\frac{K_n'(0)}{K_n(0)}Q_t + \int_0^t \frac{K_n''(t-s)}{K_n(0)}Q_s ds\right) - \frac{\gamma\sigma^2}{2}Q_t^2\right] dt. \tag{3.5}$$

As in classical mean-variance optimization, this objective can be viewed as the Lagrangian formulation of maximizing the PnL per unit time net of costs subject to the constraint that the average *dollar risk* per unit time,

$$\mathcal{R}(Q) = \sqrt{\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E}[\sigma^2 Q_t^2]dt}, \tag{3.6}$$

matches a given risk limit $\mathrm{DR} > 0$. Another equivalent formulation is to maximize the Sharpe ratio $\mathrm{SR}(Q) = \mathrm{PnL}(Q)/\mathcal{R}(Q)$ (net of trading costs) for a given risk budget $\mathcal{R}(Q) = \mathrm{DR}$.

4 Solution with Linear Impact and Exponential Decay

With nonlinear price impact and general decay kernels, the optimization problem (3.5) is intractable and can only be tackled with numerical methods such as the ones outlined in Appendix C. However, for linear price impact decaying at a single exponential time scale, the corresponding optimization problem is linear-quadratic and has a simple explicit solution. Indeed, in this case the goal functional (3.5) simplifies as follows:

Lemma 4.1. *For linear price impact* $h(x) = \lambda x$ *and a single exponential decay kernel* $K(t) = e^{-\beta t}$, *the goal functional* (3.5) *can be rewritten for* $Q \in \mathcal{Q}$ *as*

$$\mathcal{J}(Q) = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \left(\alpha Q_t f_t - \beta\lambda J_t Q_t - \frac{\gamma\sigma^2}{2}Q_t^2\right) dt\right]. \tag{4.1}$$

This goal functional is equivalent to the one studied in (Gârleanu and Pedersen, 2016, Equation (2.1)) in the limit where the time-discount rate tends to zero.

The solution for this problem has been derived by Gârleanu and Pedersen (2016) in terms of an algebraic Riccati equation. For easy reference, we provide a short self-contained derivation in Appendix B:

Theorem 4.2 (Optimal Strategy with Linear Impact and Exponential Decay)**.** *The optimal strategy* Q_t^* *for* (4.1) *is a linear feedback function of the current signal and an exponential moving average of current and past trades:*

$$Q_t^* = C_f f_t - C_J J_t, \tag{4.2}$$

where

$$C_f = \frac{\alpha(1+\beta/\phi)}{\gamma\sigma^2\left(\sqrt{1+\frac{2\lambda\beta}{\gamma\sigma^2}}+\beta/\phi\right)}, \quad C_J = \sqrt{1+\frac{2\lambda\beta}{\gamma\sigma^2}} - 1. \tag{4.3}$$

The corresponding risk is given by

$$\mathcal{R}^2(Q^*) = \frac{\alpha^2\sigma^2}{(\gamma\sigma^2)^2}\frac{(\beta/\phi+1)^2\left(\beta/\phi\sqrt{1+\frac{2\lambda\beta}{\gamma\sigma^2}}+1\right)}{\sqrt{1+\frac{2\lambda\beta}{\gamma\sigma^2}}\left(\sqrt{1+\frac{2\lambda\beta}{\gamma\sigma^2}}+\beta/\phi\right)^3}. \tag{4.4}$$

From the closed-form expressions of the optimal strategy in Theorem 4.2, it is straightforward to explore its comparative statics. Indeed, the coefficients C_f and C_J are both positive, so that the trader's holdings in the risky asset are increasing in the strength of the predictive signal and decreasing in the accumulated impact. Moreover, we see that C_f is decreasing in λ, while C_J is increasing in this parameter. This is intuitive since larger price impact affects the investor adversely. As such, the trader exploits a given signal less aggressively, in particular, if impact has already been accumulated in the same direction.

The corresponding risk $\mathcal{R}(Q^*)$ can be shown to be decreasing in both the impact and risk aversion parameters. This is again intuitive as risk averse traders avoid large positions due to the associated risk, whereas the capacity constraints induced by large impact costs also lead to smaller positions. The risk (4.4) ranges from infinity to zero as the risk aversion parameter varies from 0 to ∞. Together with the monotonicity in γ it follows that for a fixed price impact model (λ, β), there is a unique value $\gamma(\lambda, \beta)$ for which the optimal policy matches the given risk constraint.

5 Performance of Linear Feedback Policies

We now turn our attention back to the general impact model (2.3) and the corresponding objective function (3.5). With nonlinear price impact and general decay kernels, no analytical solutions are available. One way forward is to tackle the problem using numerical methods, such as the algorithm of Kolm and Ritter (2014) inspired by the classical Viterbi algorithm, or by approximating the optimal feedback controls using neural networks and optimizing these using stochastic gradient descent as in Buehler et al. (2019), for example.

In this section, we present a simple alternative in the spirit of Moallemi and Sağlam (2017); Brokmann et al. (2025), namely to use the *linear* feedback controls from Theorem 4.2. These are available in closed form as a function of only two tuning parameters, impact level and impact decay (when the corresponding risk aversion is fixed by a risk limit). Of course, these policies cannot be expected to perform well in models with nonlinear impact and impact decay if the level and decay rate are chosen arbitrarily or as the statistical point estimates for the linear model (Hey et al., 2024). Instead, akin to Brokmann et al. (2025), one needs to choose "effective" linear model parameters which take into account the nonlinearities in a suitable manner, tailored to the target risk levels and trading signals in the problem at hand.

To implement this program, we consider linear models with arbitrary risk aversion $\boldsymbol{\gamma}$, impact level $\boldsymbol{\lambda}$, impact decay $\boldsymbol{\beta}$ and the corresponding exponential moving averages $d\boldsymbol{J}_t = -\boldsymbol{\beta}\boldsymbol{J}_t dt + dQ_t$ of the trader's current and past transactions.[10] For each of these models, we then consider the corresponding optimal linear policy $Q_t^* = C_f f_t - C_J \boldsymbol{J}_t$ with coefficients from (4.3) (with λ, β, and γ replaced by $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively). More specifically, we compute the performance of these linear policies for the nonlinear goal functional (3.5). Using the properties of Gaussian processes, the performance of the linear strategies can be computed in closed form up to some integrals of the price impact function against the standard Normal distribution and integrals of the decay kernel and its derivatives. We now state the mild technical assumption needed and the PnL result for linear strategies:

Assumption 5.1. *For all $\kappa > 0$ and every $n = 1, \dots, N$ we have $\int_{-\infty}^{\infty} |xh^n(x)|e^{-\kappa x^2}dx < \infty$ and $\int_0^\infty |K_n''(t)|dt < \infty$.*

Theorem 5.2 (Performance of Linear Strategies)**.** *Suppose Assumptions 3.1 and 5.1 hold and consider the linear strategy $Q_t^* = C_f f_t - C_J \boldsymbol{J}_t$ with coefficients* (4.3) *that are optimal in a linear model with impact level $\boldsymbol{\lambda}$, impact decay $\boldsymbol{\beta}$, and risk aversion $\boldsymbol{\gamma}$.*

[10] Here, and in the sequel, we use bold to differentiate the tunable variables from the true underlying impact variables.

Then, the corresponding risk is given by (4.4) *(with* $\boldsymbol{\lambda}$*,* $\boldsymbol{\beta}$*,* $\boldsymbol{\gamma}$ *in place of* λ*,* β*,* γ*) and the expected PnL per unit time net of nonlinear impact costs is*

$$\begin{aligned}\mathrm{PnL}(\boldsymbol{\lambda},\boldsymbol{\beta};\boldsymbol{\gamma}) = \frac{\alpha^2\boldsymbol{\gamma}\sigma^2(\boldsymbol{\beta}+\phi)^2}{(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2+\xi\phi)^2} + \sum_{n=1}^{N}\frac{K_n'(0)}{K_n(0)}\frac{\tau_n}{\sigma_n}\mathbb{E}[h^n(\sigma_n Z)Z] \\ + \sum_{n=1}^{N}\frac{\mathbb{E}[h^n(\sigma_n Z)Z]}{\sigma_n}\Big(\int_0^\infty K_n''(t)\rho(t)dt + \int_0^\infty\int_0^\infty \frac{K_n''(t)K_n'(s)}{K_n(0)}\rho(|t-s|)dsdt\Big),\end{aligned} \tag{5.1}$$

where Z *is a standard normally distributed random variable and*

$$\xi = \sqrt{\boldsymbol{\gamma}\sigma^2(\boldsymbol{\gamma}\sigma^2+2\boldsymbol{\lambda}\boldsymbol{\beta})}, \tag{5.2}$$

$$\rho(t) = C_1 e^{-\phi t} - C_2 e^{-\frac{\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2}{\xi}t}, \tag{5.3}$$

$$C_1 = \frac{\alpha^2\boldsymbol{\gamma}^2\sigma^4(\boldsymbol{\beta}-\phi)(\boldsymbol{\beta}+\phi)^3}{(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2-\xi\phi)(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2+\xi\phi)^3}, \tag{5.4}$$

$$C_2 = \frac{2\alpha^2\boldsymbol{\gamma}^2\boldsymbol{\lambda}\boldsymbol{\beta}^2\sigma^4\phi(\boldsymbol{\beta}+\phi)^2}{\xi(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2-\xi\phi)(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2+\xi\phi)^3}, \tag{5.5}$$

$$\tau_n = K_n(0)\rho(0) + \int_0^\infty K_n'(t)\rho(t)dt,$$

$$\sigma_n^2 = K_n^2(0)\rho(0) + 2K_n(0)\int_0^\infty K_n'(t)\rho(t)dt + \int_0^\infty\int_0^\infty K_n'(t)K_n'(s)\rho(|t-s|)dsdt.$$

Remark 5.3. (i) *It is straightforward to check that* $C_1 > C_2$ *so that* $\rho(0) > 0$*;*

(ii) *When* $\phi = \boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2/\xi$*,* ρ *in* (5.3) *should be understood as the limit* $\phi \to \boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2/\xi$*.*

When the impact function is of power form and the decay kernel is a linear combination of exponentials, then the integrals appearing in Theorem 5.2 can all readily be computed in closed form, leading to a fully explicit formula for the performance of each linear strategy.

Corollary 5.4 (Power Law Impact and Exponential Decay Kernels)**.** *For* $n = 1,\dots,N$ *suppose that* $h^n(x) = \lambda_n \mathrm{sign}(x)|x|^{p_n}$ *for some* $p_n \in (0,1)$ *and* $\lambda_n > 0$*, and let* $K_n(t) = e^{-\beta_n t}$ *for some* $\beta_n > 0$*. Then for any linear feedback strategy* $Q_t = C_f f_t - C_J \boldsymbol{J}_t$ *with coefficients* C_f, C_J *from* (4.3) *with impact level* $\boldsymbol{\lambda}$*, impact decay* $\boldsymbol{\beta}$ *and risk aversion* $\boldsymbol{\gamma}$*, we have*

$$\begin{aligned}\mathrm{PnL}(\boldsymbol{\lambda},\boldsymbol{\beta};\boldsymbol{\gamma}) = \frac{\alpha^2\boldsymbol{\gamma}\sigma^2(\boldsymbol{\beta}+\phi)^2}{(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2+\xi\phi)^2} \\ - \sum_{n=1}^{N}\Gamma\Big(\frac{p_n}{2}\Big)\frac{p_n\beta_n\lambda_n}{2\sqrt{\pi}}\left(\frac{2\alpha^2\boldsymbol{\gamma}^2\sigma^4\phi(\boldsymbol{\beta}+\phi)^2(\boldsymbol{\beta}^2\xi+\beta_n\phi\xi+\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2(\beta_n+\phi))}{\xi(\beta_n+\phi)(\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2+\beta_n\xi)(\phi\xi+\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2)^3}\right)^{\frac{p_n+1}{2}},\end{aligned} \tag{5.6}$$

where $\Gamma(x)$ *is the gamma function and we recall that* $\xi = \xi(\boldsymbol{\lambda},\boldsymbol{\beta};\boldsymbol{\gamma})$ *is given by* (5.2)*.*

Similarly, for shifted power law kernels the corresponding integrals can still be expressed in terms of special functions implemented in many coding languages:

Corollary 5.5 (Power Law Impact and Shifted Power Law Decay Kernel)**.** *Let* $N = 1$ *and suppose that (i)* $h(x) = \lambda\mathrm{sign}(x)|x|^p$ *for some* $p \in (0,1)$ *and* $\lambda > 0$*, and (ii)* $K(t) = (t+\epsilon)^{-\nu}$ *for some* $\nu \in (0,1)$ *and* $\epsilon > 0$*. Then for any linear feedback strategy* $Q_t = C_f f_t - C_J \boldsymbol{J}_t$ *with coefficients* C_f, C_J *from* (4.3) *with impact level* $\boldsymbol{\lambda}$*, impact decay* $\boldsymbol{\beta}$ *and risk aversion* $\boldsymbol{\gamma}$ *we have*

$$\begin{aligned}&\mathrm{PnL}(\boldsymbol{\lambda},\boldsymbol{\beta};\boldsymbol{\gamma}) \\ &= \frac{\alpha^2\boldsymbol{\gamma}\sigma^2(\boldsymbol{\beta}+\phi)^2}{(\boldsymbol{\beta}\boldsymbol{\gamma}\sigma^2+\xi\phi)^2} - \lambda\nu\frac{2^{\frac{p+1}{2}}\Gamma(1+\frac{p}{2})}{\epsilon\sqrt{\pi}}\frac{\tau}{\sigma_{\text{power}}^{1-p}} \\ &\quad + \lambda\nu(\nu+1)\frac{2^{\frac{p+1}{2}}\Gamma(1+\frac{p}{2})}{\sqrt{\pi}\sigma_{\text{power}}^{1-p}}\Big(C_1\phi^{\nu+1}e^{\phi\epsilon}\Gamma(-\nu-1,\phi\epsilon) - C_2\Big(\frac{\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2}{\xi}\Big)^{\nu+1}e^{\frac{\epsilon\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2}{\xi}}\Gamma\Big(-\nu-1,\frac{\epsilon\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2}{\xi}\Big)\Big) \\ &\quad - \lambda\epsilon^\nu\nu^2(\nu+1)\frac{2^{\frac{p+1}{2}}\Gamma(1+\frac{p}{2})}{\sqrt{\pi}\sigma_{\text{power}}^{1-p}}\Big(C_1\Big(\mathcal{I}(\phi,\nu+1,\nu+2)+\mathcal{I}(\phi,\nu+2,\nu+1)\Big) \\ &\qquad\qquad - C_2\Big[\mathcal{I}\Big(\frac{\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2}{\xi},\nu+1,\nu+2\Big)+\mathcal{I}\Big(\frac{\boldsymbol{\gamma}\boldsymbol{\beta}\sigma^2}{\xi},\nu+2,\nu+1\Big)\Big]\Big),\end{aligned} \tag{5.7}$$

where

$$\tau = \frac{C_1 - C_2}{\epsilon^\nu} - \nu C_1 \phi^\nu e^{\phi\epsilon}\Gamma(-\nu, \phi\epsilon) + \nu C_2 \Big(\tfrac{\gamma\beta\sigma^2}{\xi}\Big)^\nu e^{\frac{\epsilon\gamma\beta\sigma^2}{\xi}}\Gamma\Big(-\nu, \tfrac{\epsilon\gamma\beta\sigma^2}{\xi}\Big),$$

$$\begin{aligned}\sigma^2_{\text{power}} = \frac{C_1 - C_2}{\epsilon^{2\nu}} - \frac{2\nu}{\epsilon^\nu} C_1 \phi^\nu e^{\phi\epsilon}\Gamma(-\nu, \phi\epsilon) + \frac{2\nu}{\epsilon^\nu} C_2 \Big(\tfrac{\gamma\beta\sigma^2}{\xi}\Big)^\nu e^{\frac{\epsilon\gamma\beta\sigma^2}{\xi}}\Gamma\Big(-\nu, \tfrac{\epsilon\gamma\beta\sigma^2}{\xi}\Big) \\ + 2\nu^2 C_1\, \mathcal{I}(\phi, \nu+1, \nu+1) - 2\nu^2 C_2\, \mathcal{I}\Big(\tfrac{\gamma\beta\sigma^2}{\xi}, \nu+1, \nu+1\Big),\end{aligned}$$

$$\mathcal{I}(x, q, r) = \frac{\epsilon^{\,2-q-r}}{\Gamma(r)} G^{3,1}_{2,3}\left(\epsilon x \,\middle|\, \begin{matrix} 0,\ q+r-1 \\ 0,\ r-1,\ q+r-2 \end{matrix}\right), \qquad x > 0, \quad q, r > 1.$$

Here, ξ, C_1, C_2 are given by (5.2), (5.4) *and* (5.5) *respectively, $\Gamma(\cdot)$ is the Gamma function, $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function and $G^{3,1}_{2,3}(\cdot\,|:)$ is the Meijer G-function.*

Using the explicit expressions (5.1) and especially (5.6) and (5.7), the linear policy with the maximal Sharpe ratio net of trading costs can be determined in a straightforward manner. Indeed, to this end, we consider a grid of values $(\boldsymbol{\lambda}, \boldsymbol{\beta})$ of impact levels and decay rates. For each of these, we then use the formula (4.4) for the risk of the corresponding strategy to find the value of the risk aversion parameter $\boldsymbol{\gamma}(\boldsymbol{\lambda}, \boldsymbol{\beta})$ for which a given risk constraint binds. Using the formulas for the PnL (Equation (5.1), resp. (5.6), (5.7)), we then compute the net Sharpe ratio of each linear strategy on the grid and finally select the value for which the optimal value is attained. As a byproduct, this analysis not only yields the optimal linear policy, but also provides a sensitivity analysis with respect to changes of the tuning parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. The performance of this approach under realistic market conditions is investigated in the next section.

6 Numerical Results

The linear feedback policies from the previous sections offer a simple and interpretable approach to trade scheduling for general price impact dynamics. Using our closed-form expressions for their performance, comparative statics can be computed directly and the tuning parameters can readily be optimized to take into account nonlinearities of the impact function and decay dynamics.

As in Moallemi and Sağlam (2017), this leads to an optimal policy in the chosen parametric class, but leaves open the key question how much performance is sacrificed relative to the global optimum. For a model with instantaneous impact, Brokmann et al. (2025) compared the performance of optimal linear policies to policies computed using the Viterbi algorithm of Kolm and Ritter (2014), run at very high accuracy. In the present study, we instead benchmark the performance of the linear policies against feedback controls parametrized by neural networks, which are in turn optimized via stochastic gradient descent and backpropagation methods. Appendix C.2 describes the network architecture, implementation and related methods in the financial literature. In Appendix D we verify that this approach recovers the results of Brokmann et al. (2025) but, at the same time, significantly reduces computation time and readily generalizes to models with impact decay as well.[11] Moreover, in Appendix E we establish near optimality of our trained neural networks by deriving a theoretical upper bound for the objective (3.5) using duality methods and numerically comparing them to the trained neural network strategies. This allows us to consider complex price impact models with gradual impact decay and nevertheless assess the suboptimality of linear strategies across a range of parameter regimes.

6.1 Parameter Calibration

A key step for a meaningful comparison between the different computational methods is to choose reasonable model parameters that match impact dynamics and trading costs observed in real markets. In our numerical experiments, we measure time in trading days, and holdings as well as trade sizes are measured as fractions of average daily volume (ADV); that is, a 1% ADV order has size $\Delta Q = 0.01$.

For the trading signals and risk constraints, we follow Brokmann et al. (2025) and choose the model parameters to resemble realistic market regimes faced by mid-to-high capacity trading firms. To calibrate the market parameters, we assume that there are $d = 500$ uncorrelated homogeneous assets with daily volatility $\sigma = 0.02$. Then, the multivariate optimization problem decouples into d individual optimization problems of the form studied in Section 3. We suppose each asset has an independent predictor with a

[11] We have also adapted and implemented the Viterbi algorithm for models with impact decay, but do not report the results here since it does not outperform the neural network strategies and is computationally far more expensive.

half-life of 5 days (corresponding to $\phi = 0.139$) and set the maximal (yearly) frictionless portfolio Sharpe ratio to SR = 3 by choosing $\alpha = \sigma \times \mathrm{SR}/\sqrt{256d} = 1.67 \times 10^{-4}$ (in daily units). More or less persistent trading signals and lower or higher frictionless Sharpe ratios are considered in a robustness check below.

For the price impact model, we focus on an impact function of square-root form ($p = 0.5$), as corroborated by many empirical studies in different contexts. There is less agreement about the speed of impact decay and estimates vary between more and less liquid securities. We first focus on a single decay timescale ($N = 1$) for simplicity. As a baseline case, we use an exponential kernel with impact half-life of approximately a third of a trading day ($\beta = 2$), in line with estimates from CFM meta order data reported in Hey et al. (2025). The impact of slower or faster decay is, in turn, explored in another one of our robustness checks below. We also consider the three-timescale model from Hey et al. (2025), which corresponds here to $K_n(t) = e^{-\beta_n t}$ and $h_n(x) = \lambda w_n \mathrm{sign}(x)\sqrt{|x|}$ for timescale weights w_n and decay rates β_n estimated to be

$$\beta_1 = 2, \quad \beta_2 = 1/65, \quad \beta_3 = 1/7, \qquad\qquad w_1 = 1.2, \quad w_2 = 0.6, \quad w_3 = 0.2.$$

In addition to these exponential decay profiles, we also illustrate that our methods readily extend to (shifted) power law kernels $K(t) = (\epsilon+t)^{-\nu}$, which capture multiple timescales of decay in a parsimonious manner. We take $\nu = 0.8$ as estimated in Brokmann et al. (2015) on proprietary CFM meta order data. Finally, we set the shift parameter equal to $\epsilon = 0.5$, as this choice leads to a similar one-day accumulated impact trajectory as the three-exponential model for a TWAP strategy trading 1% of ADV, roughly in line with the meta orders studied in Hey et al. (2025).

In all specifications, the push factor λ is chosen to obtain a realistic level of trading costs. More specifically, we assume the ADV for each of the traded securities is \$15.8M in line with US equity markets. We then choose the parameter λ (for all of the kernel choices above) so that the execution of a TWAP day order accounting for 1% of ADV induces an average price impact of 1.8bps in line with Brokmann et al. (2025). In the baseline setting of a single exponential decay kernel this led us to setting $\lambda = 0.0035$, while in the three-timescale model it was $\lambda = 0.0016$. The impact of higher trading costs is explored in another robustness check below.

6.2 One Decay Timescale

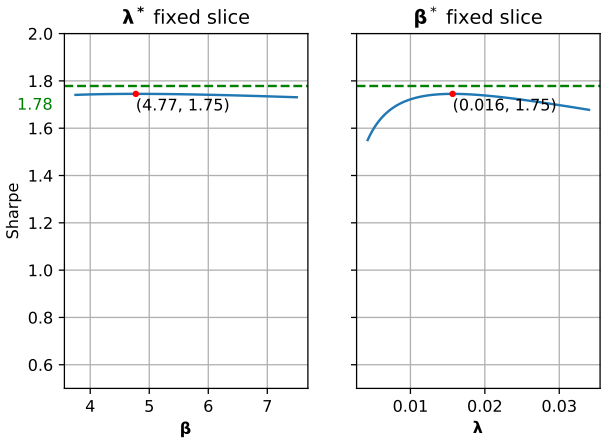
Performance comparison We now compare the performance of the linear strategies from Section 5 and the neural network strategies. For our baseline parameters with one exponential decay timescale, Figure 1 displays the optimal effective parameters $(\boldsymbol{\lambda}^*, \boldsymbol{\beta}^*)$ and the performance of the corresponding trading strategies for three different risk constraints on the portfolio of 500 uncorrelated assets.[12]

In all three cases, we see that the performance of the linear policies is relatively insensitive to the choice of the impact decay β, whereas varying the effective impact level has a considerably bigger effect. At the optimum, we see that the outperformance of the neural networks increases as the impact costs of the strategy grow for higher risk levels. With gradual impact decay, the gap between the optimal linear policy and the global optimum generally increases compared to the models with instantaneous impact studied in Brokmann et al. (2025). For example, in our baseline case with a daily risk of 1.3M, the optimal neural network policy captures a fraction 1.46/3=48.7% of the optimal frictionless Sharpe ratio, whereas the corresponding figure for the optimal linear strategy is 1.41/3=47.0%. Put differently, adding nonlinear features improves the performance net of costs by about 3.5% in this case, whereas the corresponding figure with instantaneous impact is 1.3% in Brokmann et al. (2025).

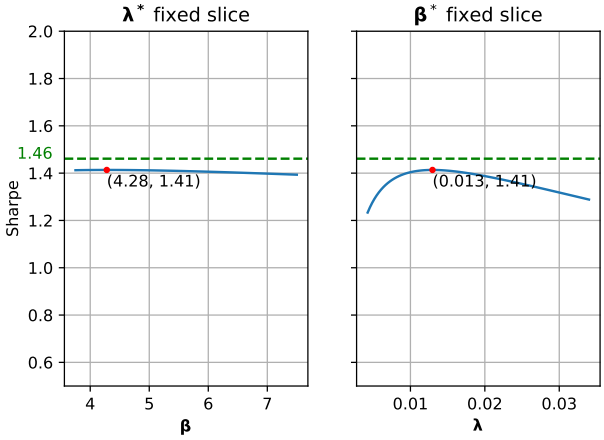
However, unless the strategy is severely capacity constrained as in the lower panel of Figure 1 (where trading costs drastically degrade the Sharpe ratio by more than two thirds to less than one), the relative loss is still very modest, in particular, when compared to the challenges in fitting predictive models and price impact parameters.

Robustness checks Table 1 assesses the robustness of the baseline results from the previous section with respect to variations of the model parameters. Each row corresponds to our baseline model from Section 6.1, with only the stated parameter perturbed. The outperformance of the neural network generally increases when the trading costs have a bigger effect because i) the level of price impact is higher, ii) it decays more slowly, or iii) the trading signal is weaker or less persistent.

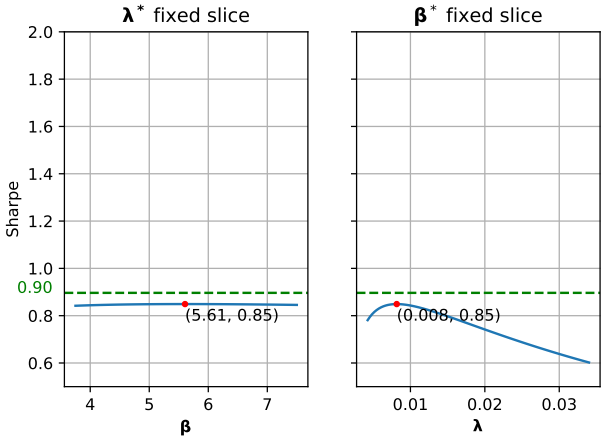
[12] For d uncorrelated homogeneous assets, the daily risk associated with holding Q shares split evenly between assets is $\mathrm{DR} = \sigma Q/\sqrt{d}$, so the smallest risk level we use corresponds to $Q = 560$M shares overall. The largest risk level corresponds to a position that is about ten times as large.



(a) DR=0.5M



(b) DR=1.3M



(c) Daily risk DR = 5.3M

Figure 1: Performance of neural networks (dashed green) and linear policies (solid blue) when one of the effective parameters $\boldsymbol{\beta}$,$\boldsymbol{\lambda}$ is fixed to its optimal value and the other varies. The daily risk (DR) is increasing from the upper to the lower panels.

Perturbed parameter value	NN Sharpe	Linear Sharpe	NN outperformance
$\phi \sim$ Half-life = 3 days	0.9	0.85	5.9%
$\phi \sim$ Half-life = 10 days	2.04	2.01	1.5%
Frictionless SR = 2	0.76	0.72	5.6%
Frictionless SR = 4	2.21	2.17	1.8%
$\lambda \sim$ 5bp per 1% ADV	0.65	0.61	6.6%
$\lambda \sim$ 10bp per 1% ADV	0.235	0.215	9.3%
$\beta \sim$ Half-life = 0.5 days	1.37	1.33	3%
$\beta \sim$ Half-life = 1 day	1.26	1.21	4.1%

Table 1: Performances with perturbed parameters at daily risk DR $\approx$ 1.3M. In the table "$\sim$" means "corresponds to".

Crucially, we see that as in the previous section, the neural networks typically only yield significant improvements in the regimes where the net Sharpe ratio is reduced too much to render any strategy attractive. The only notable exception is the case of very slow impact decay, which does not degrade performance too much but nevertheless allows nonlinearities in the optimal policy to improve performance by about 4%.

Strategy comparison To gain a better understanding of the differences between linear and neural network strategies, we now compare the corresponding trades and holdings. By construction, the linear strategy leads to trades dQ_t which are normally distributed. The network strategy, by contrast, learns to trade taking into account the true nonlinear underlying price impact leading to a different, non-Gaussian, distribution of the trades. Figure 2 displays histograms of the trades dQ_t of the two strategies, obtained by simulation on a discrete time grid with 100 trades per day. We see that the network trades have both a higher peak and fatter tails than the linear strategy, so that both very small and very large trades occur more frequently, and mid-size trades less frequently. This allows the strategy to benefit from the concavity of price impact, which marginally (i.e., per share) penalizes large trades less than it penalizes mid-size trades.

To understand this effect better, in Figure 3 we display the behaviour of the strategies for fixed signal values $f_t \in \{-1, 0, 1\}$ corresponding to negative, zero and positive future price returns of the risky asset. For the linear strategy, at any fixed slice, holdings depend on the accumulated distortion $\boldsymbol{J}_t$ in a linear fashion. The network strategy is a feedback function of (f_t, J_t^0). For each fixed signal value, the mapping $J_t^0 \mapsto J_t^0 + Q_t = J_t$ is numerically one-to-one over the plotted range, allowing us to express the network holdings as a function of J_t. This allows to compare with the linear strategy's impact state variable $\boldsymbol{J}_t$, although the two impact states need not perfectly coincide. The network shows a nonlinear dependence on $\boldsymbol{J}_t$, leading to larger positions than the linear strategy when the respective impact states are large in absolute magnitude. In Brokmann et al. (2025), a systematic nonlinear relation was observed between the current holdings and trade size for instantaneous impact. Such a relationship is not present here, as shown in Figure 4, which displays a scatter plot of the trades dQ_t against the current holdings Q_t. Unlike in Figure 3, there is no discernible relationship between the two quantities either for the linear or for the network strategies. This highlights the significance of closely monitoring the accumulated impact state and demonstrates a key new and important feature of this study due to gradual impact decay.

6.3 Multiple Decay Timescales

So far, we have compared the performance of linear and neural network strategies in a model where square-root impact decays at a single exponential timescale. However, empirical data suggests a more complex multiscale behaviour, where some fraction of the price impact dissipates very quickly while another fraction remains visible for a surprisingly long time.

To take this into account while retaining the Markovian structure of the model, one can consider a combination of several exponential decay kernels as in Hey et al. (2025). The price impact of each order then depends on several exponential moving averages with different smoothing parameters, which in turn appear as additional state variables for the optimal trading problem. Figure 5 compares the performance of the optimal neural network strategy that learns the dependency on all these variables to a number of competitors: (i) the linear strategy optimized for the one-exponential model of Section 6.2,

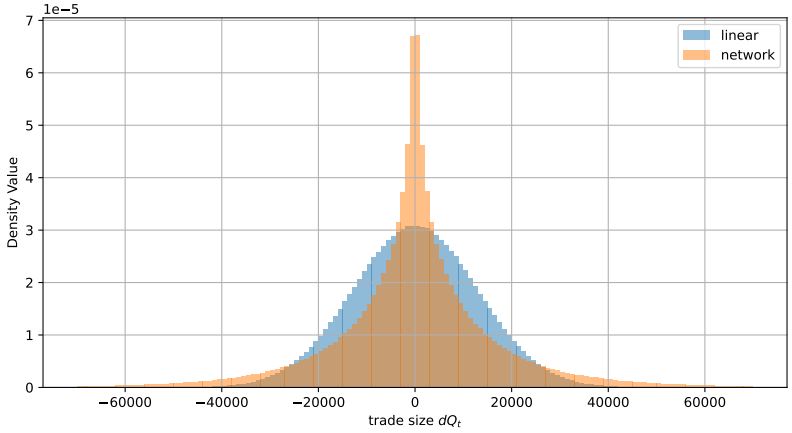


Figure 2: Trade histogram at daily risk level DR = 1.3M

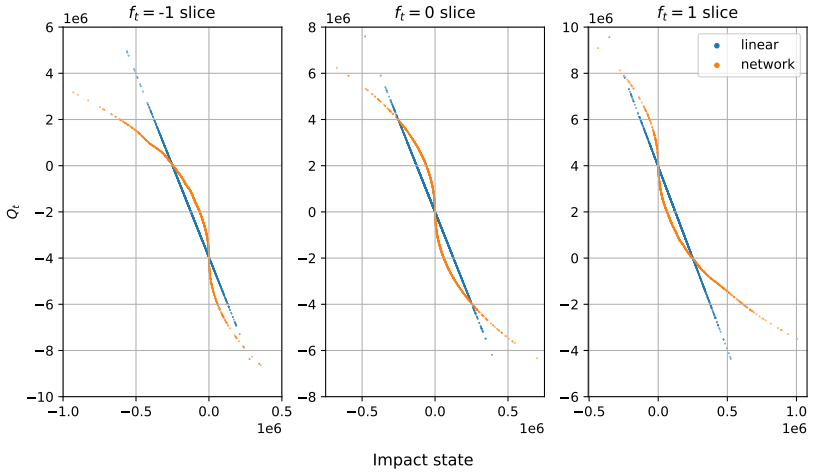


Figure 3: Feedback strategy slices at representative levels of the trading signal f_t.The x-axis corresponds to $\boldsymbol{J}$ for the linear strategy, and it corresponds to J for the network strategy.

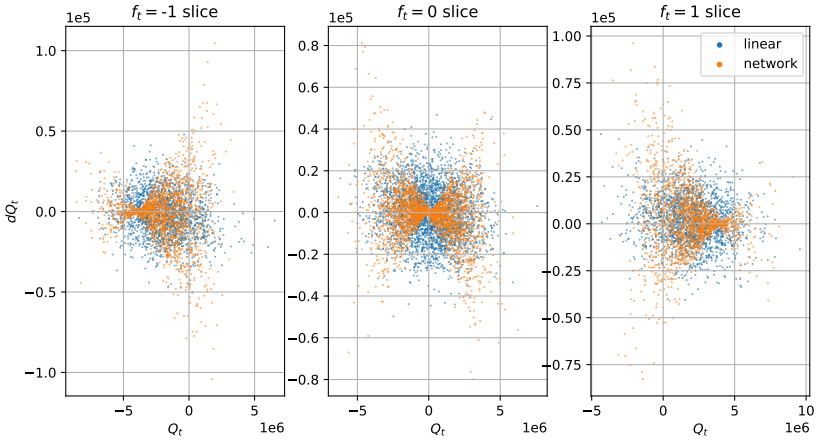


Figure 4: Scatter plot of trades dQ_t against current position Q_t at different signal values f_t.

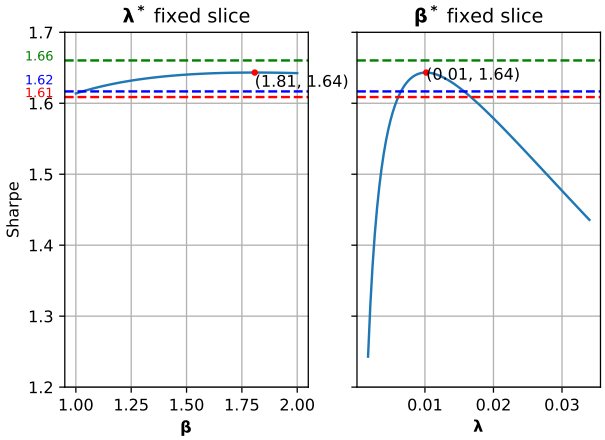


Figure 5: Performance of linear policies for three exponential decay timescales at daily risk DR = 1.3M when one of the effective parameters $\boldsymbol{\beta}$,$\boldsymbol{\lambda}$ is fixed to its optimal value and the other varies. The dashed lines indicate the performances of the optimized NN strategy (green), the NN strategy trained for one exponential timescale (blue), and the best linear strategy for one exponential timescale (red).

(ii) the neural network strategy optimized for the one-exponential model of Section 6.2, and (iii) a linear strategy which still only uses a single decay parameter but now tunes it to the multiscale impact decay as in Corollary 5.4.

A first observation is that when the strategies optimized for the simpler one-timescale model are applied out-of-model, the performance gap between them narrows substantially. This suggests that the simplicity of the linear strategies offers additional robustness. Second, we see that the performance of both strategies is markedly different from the single-exponential model, but rather close to the performances of the strategies that are optimized for the multiscale decay. This suggests that the decay kernel crucially influences the performance of trading strategies, which can be determined by simulation without the need for numerical optimization. In contrast, the additional gains that can be achieved by optimizing for the multiple timescales are relatively modest, in particular, when recalling that these figures apply in the simulated model but may well be reduced out-of-model just like when passing from one to three exponential decay kernels.

A more parsimonious approach to capture multiscale impact decay is to replace the exponential decay kernel by a (shifted) power law. This substantially reduces the number of model parameters, but makes the model fully path dependent. Optimizing neural network strategies becomes much more challenging, as there no longer is a natural finite set of state variables in this context, which is beyond our scope here. In contrast, using the explicit formulas from Corollary 5.5, the optimal linear strategies can readily be determined in the same straightforward manner as before, see Figure 6. Encouragingly, we see that the optimal linear strategy does not change too much when replacing the three-exponential kernel by a shifted power law with a similar decay profile. In contrast, the performance of the strategies is much more sensitive to the precise specification of the decay kernel.[13]

7 Conclusion

In this paper, we show that the performance of linear feedback controls remains tractable even in price impact models that account for the nonlinearities observed empirically. When the parameters of the corresponding "effective" linear model are tuned appropriately, the resulting strategies turn out to be competitive with modern machine learning methods, where the optimal feedback controls are approximated using neural networks. In particular, the already modest performance gap is reduced even further if the strategies' performances are tested outside the model on which they were trained. This suggests

[13]This parallels results for option hedging with various frictions, where the Black-Scholes delta hedge also turns out to be surprisingly robust and often close to optimal, but the corresponding hedging error is typically far from zero and crucially depends on the actual market dynamics (Denkl et al., 2013).

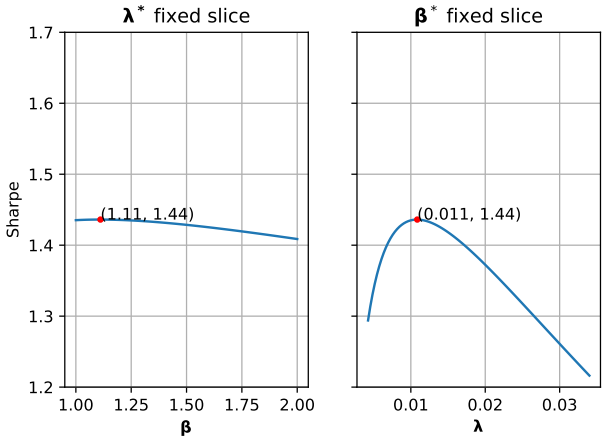


Figure 6: Performance of linear policies for power-law decay at daily risk DR = 1.3M when one of the effective parameters $\boldsymbol{\beta}$,$\boldsymbol{\lambda}$ is fixed to its optimal value and the other varies.

that while it is crucial to accurately measure the nonlinear features of price impact, taking these into account for optimal trading can be done in a straightforward and efficient manner by focusing on a simple parametric class of linear strategies.

An important direction for further research is to extend the analysis of the present paper to settings with several correlated assets. The starting point for such an analysis is a suitable model for nonlinear cross impact, that does not lead to degenerate optimization problems due to price manipulation. Such an extension of the model of Alfonsi et al. (2010) has recently been put forward by Hey et al. (2026). In the baseline case where impact decay is assumed to be decoupled and the same across assets, the approach of the present paper has a natural extension where the tuning parameter for the linear strategies is the symmetric positive-definite matrix that governs the immediate price impact in the linear proxy model. However, the number of tuning parameters then still grows quickly with the number of assets. Whence, the natural starting point to tackling multi-asset models is to assume as in Gârleanu and Pedersen (2013, 2016) that the impact matrix is a scalar multiple of the assets' covariance matrix or constructed from it with one multiplier for each asset (Hey et al., 2026).

Acknowledgements

The authors gratefully acknowledge pertinent remarks by Hans Buehler, Natascha Hey and Peter Schmidt as well as by an anonymous associate editor and an anonymous referee.

Conflicts of interest

No conflicts of interest to declare.

Funding

Financial support from Qube Research and Technologies is gratefully acknowledged.

Data availability statement

No datasets were generated or analyzed during the present study.

A The Goal Functional and Absence of Price Manipulation

Proof of Proposition 3.2. For a smooth strategy $dQ_t = \dot{Q}_t dt$ the investor's cash balance C_T from trading on $[0, T]$ is

$$C_T = -\int_0^T P_t dQ_t = -\int_0^T S_t dQ_t - \sum_{n=1}^{N} \int_0^T h^n(J_t^n) dQ_t. \tag{A.1}$$

We now rewrite both integrals using integration by parts. For the first one, we have

$$\int_0^T S_t dQ_t = S_T Q_T - \int_0^T Q_t dS_t. \tag{A.2}$$

For the second one, we recall the kernel in the definition of J^n in (2.3). We integrate by parts to see that for any $0 \le u \le t < \infty$,

$$K_n(t-u)Q_u = \int_0^u K_n(t-s)dQ_s - \int_0^u K_n'(t-s)Q_s ds.$$

Evaluating at $u = t$ yields

$$J_t^n = K_n(0)Q_t + \int_0^t K_n'(t-s)Q_s ds. \tag{A.3}$$

Isolating for Q_t and writing $A_t^n = \int_0^t K_n'(t-s)Q_s ds$ we obtain that

$$dQ_t = \frac{1}{K_n(0)}\left(dJ_t^n - dA_t^n\right).$$

As a consequence, the fundamental theorem of calculus gives

$$\begin{aligned}\int_0^T h^n(J_t^n)dQ_t &= \frac{1}{K_n(0)}\left(\int_0^T h^n(J_t^n)dJ_t^n - \int_0^T h_n(J_t^n)dA_t^n\right)\\ &= \frac{H^n(J_T^n)}{K_n(0)} - \frac{1}{K_n(0)}\int_0^T h^n(J_t^n)dA_t^n,\end{aligned} \tag{A.4}$$

where we used that $H^n(J_0^n) = 0$ since $J_0^n = 0$. Now plugging (A.2) and (A.4) into (A.1) yields

$$\begin{aligned}C_T &= -S_T Q_T - \sum_{n=1}^N \frac{H^n(J_T^n) - H^n(J_0^n)}{K_n(0)} + \int_0^T Q_t dS_t\\ &\quad + \sum_{n=1}^N \frac{1}{K_n(0)}\int_0^T h^n(J_t^n)dA_t^n\\ &= -S_T Q_T - \sum_{n=1}^N \frac{H^n(J_T^n) - H^n(J_0^n)}{K_n(0)} + \int_0^T Q_t dS_t\\ &\quad + \sum_{n=1}^N \frac{1}{K_n(0)}\int_0^T \left(h^n(J_t^n)K_n'(0)Q_t + h^n(J_t^n)\int_0^t K_n''(t-s)Q_s ds\right)dt,\end{aligned}$$

where in the final equality we used the identity $\frac{dA_t^n}{dt} = K_n'(0)Q_t + \int_0^t K_n''(t-s)Q_s ds$. As such, the investor's PnL is given by

$$\begin{aligned}\mathrm{PnL}_T &= C_T + S_T Q_T\\ &= \int_0^T Q_t dS_t - \sum_{n=1}^N \frac{H^n(J_T^n) - H^n(J_0^n)}{K_n(0)}\\ &\quad + \sum_{n=1}^N \frac{1}{K_n(0)}\int_0^T \left(h^n(J_t^n)K_n'(0)Q_t + h^n(J_t^n)\int_0^t K_n''(t-s)Q_s ds\right)dt,\end{aligned}$$

which completes the proof. □

Proof of Proposition 3.3. It suffices to prove the result when there is one impact component ($N = 1$). By (Gripenberg, 1979, Lemma 2.1), since K is log-convex, there exists a nonnegative, nonincreasing function r satisfying

$$\frac{K(t)}{K(0)} + \int_0^t K(t-s)r(s)ds = 1, \qquad t \ge 0.$$

We now set $B(t) = (r * J)(t) = \int_0^t r(t-s)J_s ds$. By (A.3), we have $J_t = K(0)Q_t + (K' * Q)(t)$. Differentiating the defining identity for r gives $K(0)r(t) + (r * K')(t) = -K'(t)/K(0)$. Hence, by associativity of the convolution operator we have

$$B(t) = (r * J)(t) = \Big(\big(K(0)r + r * K'\big) * Q\Big)(t) = -\frac{(K' * Q)(t)}{K(0)}.$$

Differentiating now gives

$$\dot{B}_t = -\frac{K'(0)}{K(0)} Q_t - \int_0^t \frac{K''(t-s)}{K(0)} Q_s ds.$$

As such, we have from (3.1) that

$$\mathbb{E}\Big[\int_0^T Q_t dS_t - \mathrm{PnL}_T\Big] = \mathbb{E}\Big[\frac{H(J_T)}{K(0)} + \int_0^T h(J_t)\dot{B}_t dt\Big]. \tag{A.5}$$

On the other hand, differentiating the convolution term directly gives the alternative expression:

$$\dot{B}_t = \frac{d}{dt}\Big(\int_0^t r(t-s)J_s ds\Big) = r(0)J_t + \int_0^t r'(t-s)J_s ds = r(t)J_t + \int_0^t [-r'(t-s)](J_t - J_s)ds.$$

We note that H is convex, since $H' = h$ is nondecreasing, so that we have the inequalities $h(x)x \geq H(x)$ and $h(x)(x-y) \geq H(x) - H(y)$. Since $r \geq 0$ and $-r' \geq 0$, integration, Fubini's theorem and these inequalities involving H yield

$$\begin{aligned}
\int_0^T h(J_t)\dot{B}_t dt &= \int_0^T r(t)h(J_t)J_t dt + \int_0^T \int_0^t [-r'(t-s)]h(J_t)(J_t - J_s)ds dt \\
&\geq \int_0^T r(t)H(J_t)dt + \int_0^T \int_0^t [-r'(t-s)]\big(H(J_t) - H(J_s)\big)ds dt \\
&= \int_0^T r(T-t)H(J_t)dt \geq 0.
\end{aligned}$$

Substituting into (A.5) we see that the cost term is nonnegative ruling out price manipulation. □

Remark A.1. *Note that this proof actually establishes the stronger claim that under Assumption 3.1 any strategy starting with flat initial holdings $Q_0 = 0$ has pathwise nonnegative costs regardless of whether or not its terminal holdings are zero.*

We conclude this section by deriving the condition (3.3) in the context of Example 3.4. Observe that

$$(\log K)''(t) = \frac{\nu q\, t^{q-2}\big[t^q - (q-1)\epsilon\big]}{(\epsilon + t^q)^2}, \qquad t > 0.$$

Hence K is log-convex for $q = 1$, but not for $q \geq 2$. In particular, $q = 1$ satisfies Assumption 3.1, and Proposition 3.2 applies for every $p > 0$.

For linear impact ($p = 1$), absence of manipulation holds also for $q = 2$. Indeed, in this case it is easy to check that the impact cost of a smooth strategy satisfies

$$\frac{\lambda}{2}\int_0^T \int_0^T K(|t-s|)\dot{Q}_t \dot{Q}_s ds dt \geq 0.$$

By continuous extension to semimartingale strategies Q it is clear that price manipulation is not possible in this case.

For the remaining twice continuously differentiable kernels, consider three trades of sizes $-1, 1+b, -b$ at times $t_0, t_0+\delta, t_0+2\delta$, for some $b, t_0, \delta > 0$. These trades form a round trip. With $H(x) = \lambda|x|^{p+1}/(p+1)$, the cost of a block trade changing the impact state from J_- to J_+ is $[H(J_+) - H(J_-)]/K(0)$. Using

$$K(t) = \epsilon^{-\nu} - \nu\epsilon^{-\nu-1}t^q + O(t^{2q}),$$

the total cost therefore satisfies

$$C_\delta = \lambda\nu\epsilon^{-\nu p - 1}\big[1 + b^{p+1} - (2^q - 2)b^p\big]\delta^q + o(\delta^q).$$

Setting $b = p^{1/(p+1)}$ we then see that the sign of the leading coefficient is the same as the sign of

$$(p+1)p^{-p/(p+1)} - (2^q - 2).$$

Since

$$(p+1)p^{-p/(p+1)} = \min_{b>0}\big(b^{-p} + b\big) \leq 2,$$

with equality only when $p = 1$, the leading coefficient is strictly negative whenever $q = 2$ and $p \neq 1$, or whenever $q > 2$. For sufficiently small δ, these round trips thus have negative costs. Approximating the block trades by smooth trading preserves the strictly negative cost, completing the classification (3.3).

B Proofs for Linear Strategies

Proof of Lemma 4.1. Plugging the linear impact function and exponential kernel into (3.5) and computing the integral leads to the expression

$$\mathcal{J}(Q) = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \left(\alpha Q_t f_t - \lambda\beta J_t\left(Q_t - \beta\int_0^t e^{-\beta(t-s)}Q_s ds\right) - \frac{\gamma\sigma^2}{2}Q_t^2\right)dt\right]. \tag{B.1}$$

We note $J_t = Q_t - \beta\int_0^t e^{-\beta(t-s)}Q_s ds$ (see Equation (A.3)) so that the middle term in the right hand side of (B.1) is simply $-\lambda\beta J_t^2$. To put this into the form (4.1) we again use integration by parts to write

$$(Q_T - J_T)^2 = 2\beta\int_0^T (J_t Q_t - J_t^2)dt.$$

Taking expectation, dividing by T and sending $T\to\infty$ we see that

$$\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[(Q_T - J_T)^2] \le 2\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[Q_T^2] + 2\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[J_T^2] = 0, \tag{B.2}$$

where we used Young's inequality and the transversality condition in the definition of $\mathcal{Q}$; note that in the linear impact case considered here we have that $H(x)\propto x^2$. Hence, we have the identity $\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbb{E}[J_t^2]dt = \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbb{E}[J_t Q_t]dt$. Making these substitutions in (B.1) and recalling that $I_t = \lambda J_t$ gives precisely (4.1). □

Proof of Theorem 4.2; Optimal Strategy. To derive the optimality of (4.2), we first make a change of variables to put the problem at hand into a standard stochastic control framework. To this end we define $J_t^0 = J_t - Q_t$ and note that

$$dJ_t^0 = -\beta J_t dt = -\beta(J_t^0 + Q_t)dt. \tag{B.3}$$

Plugging this in for J in the objective function (4.1), we obtain

$$\mathcal{J}(Q) = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \left(\alpha Q_t f_t - \lambda\beta J_t^0 Q_t - \lambda\beta Q_t^2 - \frac{\gamma\sigma^2}{2}Q_t^2\right)dt\right].$$

This is now a standard ergodic stochastic control problem with one uncontrolled state variable f and one controlled state variable J^0. As such, the HJB equation for the corresponding value function $V(J^0, f)$ is

$$\begin{aligned}\ell = \sup_Q\Big\{&(\alpha f - \lambda\beta J^0)Q - \left(\lambda\beta + \frac{\gamma\sigma^2}{2}\right)Q^2 - \beta(J^0+Q)\partial_{J^0}V(J^0,f)\\ &-\phi f\partial_f V(J^0,f) + \phi\partial_{ff}V(J^0,f)\Big\},\end{aligned}$$

where ℓ is the optimal long-run average value of the objective. We now make a quadratic ansatz for V,

$$V(J^0, f) = -\frac{a}{2}(J^0)^2 + bJ^0 f + \mathcal{F}(f)$$

for some constants a and b to be determined. The functional $\mathcal{F}(\cdot)$ will be determined by an exogenous ODE but does not impact the optimal strategy so we do not pursue this further here. Plugging in and maximizing over Q yields

$$Q^* = \frac{\alpha - b\beta}{2\beta\lambda + \gamma\sigma^2}f - \frac{\lambda\beta - a\beta}{2\beta\lambda + \gamma\sigma^2}J^0. \tag{B.4}$$

Now we plug this back into the HJB equation and collect terms. Comparison of coefficients for the terms proportional to $(J^0)^2$ and $J^0 f$ leads to the following two equations for the two unknown parameters a and b:

$$\begin{aligned}a^2\beta + a\left(2\gamma\sigma^2 + 2\lambda\beta\right) + \lambda^2\beta = 0,\\ b\left(a\beta^2 + \lambda\beta^2 + \phi\left(\gamma\sigma^2 + 2\lambda\beta\right) + \gamma\beta\sigma^2\right) - a\alpha\beta + \alpha\lambda\beta = 0.\end{aligned}$$

The first equation involves only a and is quadratic in it, while the second equation is linear in b. Hence solving the equation for a (and taking the "+" solution) and then solving for b yields

$$\begin{aligned}a &= \frac{-\beta\lambda - \gamma\sigma^2 + \sqrt{\gamma\sigma^2(2\beta\lambda + \gamma\sigma^2)}}{\beta},\\ b &= \frac{\alpha(\sqrt{\gamma\sigma^2\left(\gamma\sigma^2 + 2\beta\lambda\right)} - \left(\gamma\sigma^2 + 2\beta\lambda\right))}{\gamma\sigma^2\phi + 2\beta\lambda\phi + \sqrt{\gamma\beta^2\sigma^2\left(\gamma\sigma^2 + 2\beta\lambda\right)}}.\end{aligned}$$

Inserting these expressions into (B.4) in turn pins down the optimal feedback trading strategy:

$$Q^* = \frac{\alpha(\beta+\phi)}{\gamma\sigma^2\phi + 2\beta\lambda\phi + \sqrt{\gamma\beta^2\sigma^2\left(\gamma\sigma^2+2\beta\lambda\right)}} f - \frac{\gamma\sigma^2 + 2\beta\lambda - \sqrt{\gamma\sigma^2\left(\gamma\sigma^2+2\beta\lambda\right)}}{\gamma\sigma^2+2\beta\lambda} J^0.$$

Recalling that $J^0 = J - Q^*$, isolating for Q^* and simplifying now finally yields the expression (4.2) for the optimal feedback strategy with coefficients (4.3). □

To complete the proof of Theorem 4.2 it now remains to establish formula (4.4) for the risk of the strategy. To accomplish this we first establish Lemma B.1 below, which is also used in the proof of Theorem 5.2, which follows. Throughout, we use the notation of Theorem 5.2.

Lemma B.1. *Consider the linear policy $Q_t = C_f f_t - C_J \boldsymbol{J}_t$ with coefficients from* (4.3)*. Then there exists $\kappa > 0$ such that:*

(i) *$\{(f_t, \boldsymbol{J}_t)\}_{t\geq 0}$ is a bivariate Gaussian process satisfying, as $t \to \infty$,*

$$\mathbb{E}[f_t^2] = 1 + o(e^{-\kappa t}), \tag{B.5}$$

$$\mathbb{E}[f_t \boldsymbol{J}_t] = \frac{\phi C_f}{\boldsymbol{\beta} + \phi(1+C_J)} + o(e^{-\kappa t}), \tag{B.6}$$

$$\mathbb{E}[\boldsymbol{J}_t^2] = \frac{\phi C_f^2}{(1+C_J)(\boldsymbol{\beta}+\phi(1+C_J))} + o(e^{-\kappa t}); \tag{B.7}$$

(ii) *$\{Q_t\}_{t\geq 0}$ is a Gaussian process satisfying*

$$\mathbb{E}[Q_t Q_s] = \rho(|t-s|) + o(e^{-\kappa \max\{t,s\}}); \tag{B.8}$$

(iii) *$\{(Q_t, J_t^n)\}_{t\geq 0}$ is a bivariate Gaussian process satisfying for $0 \leq t \leq T < \infty$,*

$$\mathbb{E}[J_T^n Q_t] = K_n(0)\rho(T-t) + \int_0^T K_n'(T-s)\rho(|t-s|)ds + o(1), \tag{B.9}$$

$$\mathbb{E}[(J_T^n)^2] = K_n^2(0)\rho(0) + 2K_n(0)\int_0^T K_n'(t)\rho(t)dt + \int_0^T\int_0^T K_n'(t)K_n'(s)\rho(|t-s|)dsdt + o(1), \tag{B.10}$$

where the remainder in (B.9) *is uniform over $0 \leq t \leq T$.*

(iv) *J_T^n converges in distribution as $T \to \infty$ to a centered Gaussian random variable with variance σ_n^2.*

Proof. Throughout we assume for simplicity and without loss of generality that $f_0 = \boldsymbol{J}_0 = J_0^n = 0$, consistent with $Q_0 = 0$. We also note that if $g(t) = o(e^{-\kappa_1 t})$ then $e^{-\kappa_2 t}\int_0^t g(s)ds = o(e^{-\kappa_3 t})$ for any $\kappa_1, \kappa_2 > 0$ and some $\kappa_3 > 0$. We don't keep track of this constant and just generically write κ in the computations that follow.

Proof of (i) Using the given form of Q we see that

$$d\boldsymbol{J}_t = -\boldsymbol{\beta}\boldsymbol{J}_t dt + C_f df_t - C_J d\boldsymbol{J}_t \implies d\boldsymbol{J}_t = -\frac{\boldsymbol{\beta}}{1+C_J}\boldsymbol{J}_t dt + \frac{C_f}{1+C_J} df_t. \tag{B.11}$$

From here it is clear that $\{(f_t, \boldsymbol{J}_t)\}_{t\geq 0}$ is jointly Gaussian, which can be most easily seen using the fact that f is an Ornstein-Uhlenbeck process and that $\boldsymbol{J}_t = \frac{C_f}{1+C_J}\int_0^t e^{-\frac{\boldsymbol{\beta}}{1+C_J}(t-s)} df_s$.

Equation (B.5) follows immediately from well-known properties of the Ornstein-Uhlenbeck process. To obtain (B.6) we integrate by parts,

$$\begin{aligned} d(f_t\boldsymbol{J}_t) &= f_t d\boldsymbol{J}_t + \boldsymbol{J}_t df_t + d[\boldsymbol{J}, f]_t \\ &= \left(-\frac{\boldsymbol{\beta}+\phi(1+C_J)}{1+C_J} f_t\boldsymbol{J}_t - \frac{\phi C_f}{1+C_J} f_t^2 + \frac{2\phi C_f}{1+C_J}\right) dt + \left(\frac{\sqrt{2\phi}C_f}{1+C_J} f_t + \sqrt{2\phi}\boldsymbol{J}_t\right) dW_t. \end{aligned}$$

Taking expectation we obtain the ODE

$$\frac{\partial}{\partial t}\mathbb{E}[f_t\boldsymbol{J}_t] = -\frac{\boldsymbol{\beta}+\phi(1+C_J)}{1+C_J}\mathbb{E}[f_t\boldsymbol{J}_t] - \frac{\phi C_f}{1+C_J}\mathbb{E}[f_t^2] + \frac{2\phi C_f}{1+C_J},$$

which has solution

$$\mathbb{E}[f_t \boldsymbol{J}_t] = \int_0^t \exp\left(-\frac{\boldsymbol{\beta} + \phi(1+C_J)}{1+C_J}(t-s)\right)\left(\frac{2\phi C_f}{1+C_J} - \frac{\phi C_f}{1+C_J}\mathbb{E}[f_s^2]\right) ds.$$

Using (B.5) to replace $\mathbb{E}[f_s^2]$ and integrating yields (B.6). For (B.7) we proceed similarly by computing

$$d\boldsymbol{J}_t^2 = \left(-\frac{2\boldsymbol{\beta}}{1+C_J}\boldsymbol{J}_t^2 - \frac{2\phi C_f}{1+C_J} f_t \boldsymbol{J}_t + \frac{2\phi C_f^2}{(1+C_J)^2}\right) dt + \frac{2\sqrt{2\phi} C_f \boldsymbol{J}_t}{1+C_J} dW_t.$$

Taking expectation yields the ODE

$$\frac{\partial}{\partial t}\mathbb{E}[\boldsymbol{J}_t^2] = -\frac{2\boldsymbol{\beta}}{1+C_J}\mathbb{E}[\boldsymbol{J}_t^2] - \frac{2\phi C_f}{1+C_J}\mathbb{E}[f_t \boldsymbol{J}_t] + \frac{2\phi C_f^2}{(1+C_J)^2},$$

which has solution

$$\mathbb{E}[\boldsymbol{J}_t^2] = \int_0^t \exp\left(-\frac{2\boldsymbol{\beta}}{1+C_J}(t-s)\right)\left(\frac{2\phi C_f^2}{(1+C_J)^2} - \frac{2\phi C_f}{1+C_J}\mathbb{E}[f_s \boldsymbol{J}_s]\right) ds.$$

Now using (B.6) to replace $\mathbb{E}[f_s \boldsymbol{J}_s]$ and integrating yields (B.7).

Proof of (ii) The Gaussianity of Q follows immediately from that of f and $\boldsymbol{J}$. To prove (B.8) we assume without loss of generality that $s \le t$ and note that

$$\mathbb{E}[Q_t Q_s] = C_f^2 \mathbb{E}[f_t f_s] - C_f C_J \mathbb{E}[f_t \boldsymbol{J}_s] - C_f C_J \mathbb{E}[f_s \boldsymbol{J}_t] + C_J^2 \mathbb{E}[\boldsymbol{J}_t \boldsymbol{J}_s]. \tag{B.12}$$

For the OU process, we have

$$\mathbb{E}[f_t | \mathcal{F}_s] = f_s e^{-\phi(t-s)}. \tag{B.13}$$

Similarly, using the same approach as in the proof of part (i) above, we can derive an ODE for $\mathbb{E}[\boldsymbol{J}_t|\mathcal{F}_s]$, which leads to the expression

$$\begin{aligned}\mathbb{E}[\boldsymbol{J}_t|\mathcal{F}_s] &= \boldsymbol{J}_s e^{-\frac{\boldsymbol{\beta}}{1+C_J}(t-s)} - \frac{\phi C_f}{1+C_J} f_s \int_s^t e^{-\phi(u-s)} e^{-\frac{\boldsymbol{\beta}}{1+C_J}(t-u)} du \\ &= \boldsymbol{J}_s e^{-\frac{\boldsymbol{\beta}}{1+C_J}(t-s)} + \frac{\phi C_f}{1+C_J} f_s \frac{e^{-\phi(t-s)} - e^{-\frac{\boldsymbol{\beta}}{1+C_J}(t-s)}}{\phi - \frac{\boldsymbol{\beta}}{1+C_J}}.\end{aligned} \tag{B.14}$$

Hence, we (a) use the tower property in (B.12) to condition on $\mathcal{F}_s$, (b) sub in the expressions (B.13), (B.14) for the conditional expectations, (c) use the identities (B.5)-(B.7) and collect like terms and finally (d) express C_f and C_J in terms of the model parameters (4.3) to obtain (B.8).

Proof of (iii) The joint Gaussianity of Q and J^n is immediate from the fact that Q is Gaussian and the integral expression (2.3) for J^n. By taking the differential of (2.3), as we did in the proof of Proposition 3.2, we recall that J^n has the representation (A.3). Hence

$$\begin{aligned}\mathbb{E}[J_T^n Q_t] &= K_n(0)\mathbb{E}[Q_T Q_t] + \int_0^T K_n'(T-s)\mathbb{E}[Q_t Q_s] ds \\ &= K_n(0)\rho(T-t) + \int_0^T K_n'(T-s)\rho(|t-s|) ds + O\left((e^{-\kappa T}) + \int_0^T |K_n'(T-s)| e^{-\kappa s} ds\right),\end{aligned}$$

where the second equality followed from (B.8) and is uniform over $0 \le t \le T$. Since K_n' is integrable as a consequence of K_n being nonincreasing, we have by a change of variables and the Dominated Convergence Theorem that

$$\int_0^T |K_n'(T-s)| e^{-\kappa s} ds = \int_0^T |K_n'(u)| e^{-\kappa(T-u)} du \longrightarrow 0 \quad \text{as } T \to \infty, \tag{B.15}$$

establishing (B.9) uniformly in t.

To prove (B.10) we again use (A.3) to infer that

$$\mathbb{E}[(J_T^n)^2] = K_n^2(0)\mathbb{E}[Q_T^2] + 2K_n(0)\int_0^T K_n'(T-s)\mathbb{E}[Q_T Q_s] ds + \int_0^T \int_0^T K_n'(T-t)K_n'(T-s)\mathbb{E}[Q_t Q_s] ds dt.$$

By (B.8), the errors from replacing the covariances in the first two terms by their stationary counterparts are $O(e^{-\kappa T})$. The error in the double integral is bounded by

$$C\int_0^T\int_0^T |K_n'(T-t)K_n'(T-s)|e^{-\kappa\max\{t,s\}}dsdt \le C\|K_n'\|_{L^1(0,\infty)}\int_0^T |K_n'(T-s)|e^{-\kappa s}ds \longrightarrow 0$$

for some $C>0$, where the convergence to zero is as in (B.15). Consequently,

$$\begin{aligned}\mathbb{E}[(J_T^n)^2] &= K_n^2(0)\rho(0) + 2K_n(0)\int_0^T K_n'(T-s)\rho(T-s)ds \\ &\quad + \int_0^T\int_0^T K_n'(T-t)K_n'(T-s)\rho(|t-s|)dsdt + o(1) \\ &= K_n^2(0)\rho(0) + 2K_n(0)\int_0^T K_n'(t)\rho(t)dt + \int_0^T\int_0^T K_n'(t)K_n'(s)\rho(|t-s|)dsdt + o(1),\end{aligned}$$

where the final equality follows by a change of variables. This proves (B.10).

Proof of (iv) Since J^n is a Gaussian process we just have to show that $\lim_{T\to\infty}\mathbb{E}[J_T^n]=0$ and that $\lim_{T\to\infty}\mathbb{E}[(J_T^n)^2]=\sigma_n^2$. The latter is immediate from (B.10). The former is also straightforward. Indeed, $\mathbb{E}[f_T]=0$ for every T and from the dynamics (B.11) it is straightforward to establish that the same is true for $\boldsymbol{J}$. Hence $\mathbb{E}[Q_T]=0$ for every T, from which it follows that $\mathbb{E}[J_T^n]=0$ for every T using (A.3). □

Proof of Theorem 4.2; Equation 4.4. Using L'Hôpital's rule we have that

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathbb{E}[Q_t^2]dt = \lim_{T\to\infty}\mathbb{E}[Q_T^2].$$

From Lemma B.1(ii) we have that $\lim_{T\to\infty}\mathbb{E}[Q_T^2]=\rho(0)=C_1-C_2$. Simplifying this expression and multiplying by σ^2 gives the formula (4.4) for $\mathcal{R}^2(Q^*)$. □

Proof of Theorem 5.2. We start by computing the return term $\lim_{T\to\infty}\frac{1}{T}\int_0^T \alpha\mathbb{E}[Q_t f_t]\,dt$. By L'Hôpital's rule this limit is given by $\alpha\lim_{T\to\infty}\mathbb{E}[Q_T f_T]$. Using Lemma B.1(i) we obtain that this can be computed as

$$\lim_{T\to\infty}\alpha\mathbb{E}[f_T Q_T] = \alpha C_f\lim_{T\to\infty}\mathbb{E}[f_T^2] - \alpha C_J\lim_{T\to\infty}\mathbb{E}[f_T\boldsymbol{J}_T] = \alpha C_f - \frac{\alpha\phi C_f C_J}{\boldsymbol{\beta}+\phi(1+C_J)}. \tag{B.16}$$

Next we study the impact term and note that again by L'Hôpital's rule,

$$\begin{aligned}&\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\int_0^T\sum_{n=1}^N h^n(J_t^n)\left(\frac{K_n'(0)}{K_n(0)}Q_t + \int_0^t\frac{K_n''(t-s)}{K_n(0)}Q_s ds\right)dt\right] \\ &= \sum_{n=1}^N\left(\frac{K_n'(0)}{K_n(0)}\lim_{T\to\infty}\mathbb{E}[h^n(J_T^n)Q_T] + \lim_{T\to\infty}\int_0^T\frac{K_n''(T-t)}{K_n(0)}\mathbb{E}[h^n(J_T^n)Q_t]dt\right)\end{aligned} \tag{B.17}$$

Since J_T^n and Q_t are jointly Gaussian random variables, they are independent if and only if uncorrelated. Hence:

$$\mathbb{E}[h^n(J_T^n)Q_t] = \mathbb{E}\left[h^n(J_T^n)\left(Q_t - \frac{\mathbb{E}[J_T^n Q_t]}{\mathbb{E}[(J_T^n)^2]}J_T^n\right)\right] + \frac{\mathbb{E}[J_T^n Q_t]}{\mathbb{E}[(J_T^n)^2]}\mathbb{E}[h^n(J_T^n)J_T^n] = \frac{\mathbb{E}[J_T^n Q_t]}{\mathbb{E}[(J_T^n)^2]}\mathbb{E}[h^n(J_T^n)J_T^n]. \tag{B.18}$$

Lemma B.1(iii) and (iv) yield

$$\lim_{T\to\infty}\mathbb{E}[J_T^n Q_T]=\tau_n,\quad \lim_{T\to\infty}\mathbb{E}[(J_T^n)^2]=\sigma_n^2,\quad \lim_{T\to\infty}\mathbb{E}[h^n(J_T^n)J_T^n]=\mathbb{E}[h^n(\sigma_n Z)\sigma_n Z], \tag{B.19}$$

where Z is a standard normal random variable. The final limit follows by the Dominated Convergence Theorem and Assumption 5.1, since the Gaussian densities of J_T^n converge pointwise and, for sufficiently large T, are uniformly bounded by Ce^{-cx^2} for some constants $c, C>0$ independent of T.

Hence, we can sub (B.18) into (B.17) and use the identities in (B.19) to obtain that (B.17) is given by

$$\sum_{n=1}^N\left(\frac{K_n'(0)}{K_n(0)}\frac{\tau_n}{\sigma_n}\mathbb{E}[h^n(\sigma_n Z)Z] + \frac{\mathbb{E}[h^n(\sigma_n Z)Z]}{\sigma_n K_n(0)}\lim_{T\to\infty}\int_0^T K_n''(T-t)\mathbb{E}[J_T^n Q_t]dt\right). \tag{B.20}$$

It therefore just remains to compute

$$\lim_{T\to\infty}\int_0^T K_n''(T-t)\mathbb{E}[J_T^n Q_t]dt.$$

To compute the final limit we note by (B.9) that $\mathbb{E}[J_T^n Q_t] = K_n(0)\rho(T-t)+\int_0^T K_n'(T-s)\rho(|t-s|)ds+o(1)$. From the integrability condition on K'' in Assumption 5.1 it follows that the $o(1)$ term does not contribute in the limit yielding

$$\begin{aligned}
&\lim_{T\to\infty}\int_0^T K_n''(T-t)\mathbb{E}[J_T^n Q_t]dt \\
&\quad= \lim_{T\to\infty}\left(K_n(0)\int_0^T K_n''(T-t)\rho(T-t)dt + \int_0^T\int_0^T K_n''(T-t)K_n'(T-s)\rho(|t-s|)dsdt\right) \\
&\quad= K_n(0)\int_0^\infty K_n''(t)\rho(t)dt + \int_0^\infty\int_0^\infty K_n''(t)K_n'(s)\rho(|t-s|)dsdt. \qquad \text{(B.21)}
\end{aligned}$$

Hence using (B.16), (B.20) and (B.21) and recalling the definitions of C_f, C_J given by (4.3) we obtain (5.1). Corollary 5.4 now follows by explicitly computing the integrals, which only involve exponential functions, and inserting the explicit formula for the absolute moment of the standard normal random variable into $\mathbb{E}[h^n(\sigma_n Z)Z] = \lambda_n|\sigma_n|^{p_n}\mathbb{E}[|Z|^{p_n+1}]$. Corollary 5.5 follows similarly by known integral representations that involve the incomplete Gamma and hypergeometric functions. □

We conclude this appendix by showing that linear strategies are members of $\mathcal{Q}$ under Assumptions 3.1 and 5.1. For any such strategy, from Lemma B.1(iv) it follows that $\mathbb{E}[H^n(J_T^n)] \to \mathbb{E}[H^n(\sigma_n Z)] < \infty$, which establishes the transversality condition $\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[H^n(J_T^n)] = 0$. Analogously, we have that $\int_0^T \mathbb{E}[Q_t^2]dt$ is finite by Lemma B.1(ii), establishing membership in $\mathcal{Q}$.

C Numerical Methods

In this appendix, we provide a detailed description of the numerical method we use to benchmark the performance of the linear policies: a machine learning-based approach where the feedback controls are parametrized by neural networks. To simplify the exposition, we focus on optimizing the objective function (3.5) under the assumption of a single exponential decay timescale:[14]

$$\begin{aligned}
\mathcal{J}(Q) &= \lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\int_0^T\left(\alpha Q_t f_t - \lambda\beta\,\mathrm{sign}(J_t)|J_t|^p\left(Q_t - \beta\int_0^t e^{-\beta(t-s)}Q_s ds\right) - \frac{\gamma\sigma^2}{2}Q_t^2\right)dt\right] \\
&= \lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\int_0^T\left(\alpha Q_t f_t - \lambda\beta|J_t|^{p+1} - \frac{\gamma\sigma^2}{2}Q_t^2\right)dt\right], \qquad \text{(C.1)}
\end{aligned}$$

where the last equality follows from (A.3). To optimize this numerically, we first derive a discretized version of the problem. Subsequently, we discuss the machine learning algorithm used to optimize it.

C.1 Discrete Setting

Discrete objective function To deploy numerical methods, we consider discrete strategies with a fixed trading frequency Δt. Each holding sequence $\{Q_n\}$ leads to a left-continuous holding process Q_t, where $t_0 = 0$ and for every $n = 1, 2\ldots$

$$\begin{aligned}
t_n &= n\Delta t, \\
Q_t &= Q_n \text{ for } t \in (t_{n-1}, t_n].
\end{aligned} \qquad \text{(C.2)}$$

With the exponential decay kernel $K(t) = e^{-\beta t}$ we then have

$$J_t = e^{-\beta(t-t_{n-1})}(J_{t_{n-1}} + Q_n - Q_{n-1}) \text{ for } t \in (t_{n-1}, t_n]. \qquad \text{(C.3)}$$

[14] We have also implemented a similar algorithm for multiple exponential decay timescales to obtain the results reported in Section 6.3. An extension to a fully path-dependent (shifted) power law kernel is more challenging and not pursued here, as the results from Section 6.3 suggest that the additional performance gains that can be achieved in this manner are rather modest.

In particular, we consider the sequence $\{J_n\}$ where

$$J_n = J_{t_n} = e^{-\beta\Delta t}(J_{n-1} + Q_n - Q_{n-1}). \tag{C.4}$$

The next step is to express the objective function (C.1) in terms of the sequences $\{Q_n\}$ and $\{J_n\}$. The expected asset return can be calculated as follows:

$$\begin{aligned}
\mathbb{E}\Big[\int_0^{N\Delta t} \alpha Q_t f_t dt\Big] &= \sum_{n=1}^{N} \mathbb{E}\Big[\mathbb{E}\Big[\int_{t_{n-1}}^{t_n} Q_t \alpha f_t dt \,\Big|\, \mathcal{F}_{t_{n-1}}\Big]\Big] \\
&= \sum_{n=1}^{N} \mathbb{E}\Big[Q_n \alpha \int_{t_{n-1}}^{t_n} \mathbb{E}\big[f_t \mid \mathcal{F}_{t_{n-1}}\big] dt\Big] \\
&= \sum_{n=1}^{N} \mathbb{E}\Big[Q_n \alpha \int_{t_{n-1}}^{t_n} e^{-\phi(t-t_{n-1})} f_{t_{n-1}} dt\Big] \\
&= \mathbb{E}\Big[\sum_{n=1}^{N} \alpha \frac{(1-e^{-\phi\Delta t})}{\phi} Q_n f_{n-1}\Big].
\end{aligned} \tag{C.5}$$

In the last equation, we denote $f_n = f_{t_n}$, thus simplifying the expression to handle the signal sequence $\{f_n\}$, which, recalling (2.2), is governed by the dynamics

$$f_n = e^{-\phi\Delta t} f_{n-1} + \sqrt{1 - e^{-2\phi\Delta t}} Z_n, \tag{C.6}$$

where $Z_n \sim N(0,1)$ are IID standard Normal random variables.

Regarding the impact term we have $J_t = e^{\beta(t_n - t)} J_{t_n}$ for $t \in (t_{n-1}, t_n]$. Hence, with $T = N\Delta t$,

$$\begin{aligned}
\mathbb{E}\Big[\int_0^T \beta\lambda |J_t|^{p+1} dt\Big] &= \mathbb{E}\Big[\sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} \beta\lambda |e^{\beta(t_n - t)} J_{t_n}|^{p+1} dt\Big] \\
&= \mathbb{E}\Big[\sum_{n=1}^{N} \beta\lambda \frac{(e^{\beta(p+1)\Delta t} - 1)}{\beta(p+1)} |J_{t_n}|^{p+1}\Big].
\end{aligned} \tag{C.7}$$

Lastly, the risk term is given by:

$$\mathbb{E}\Big[\int_0^T \frac{\gamma\sigma^2}{2} Q_t^2 dt\Big] = \mathbb{E}\Big[\sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} \frac{\gamma\sigma^2}{2} Q_t^2 dt\Big] = \mathbb{E}\Big[\sum_{n=1}^{N} \frac{\gamma\sigma^2}{2} Q_n^2 \Delta t\Big]. \tag{C.8}$$

In summary, the goal functional becomes

$$\begin{aligned}
\mathcal{J}(Q) &= \lim_{N\to\infty} \frac{1}{N\Delta t} \mathbb{E}\left[\sum_{n=1}^{N} \Big\{\alpha \frac{(1-e^{-\phi\Delta t})}{\phi} Q_n f_{n-1} - \beta\lambda \frac{(e^{\beta(p+1)\Delta t} - 1)}{\beta(p+1)} |J_{t_n}|^{p+1} - \frac{\gamma\sigma^2}{2} Q_n^2 \Delta t\Big\}\right] \\
&= \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{n=1}^{N} \Big\{\hat{\alpha} Q_n f_{n-1} - \beta\hat{\lambda} |J_{t_n}|^{p+1} - \frac{\gamma\sigma^2}{2} Q_n^2\Big\}\right],
\end{aligned} \tag{C.9}$$

where

$$\hat{\alpha} = \alpha \frac{(1-e^{-\phi\Delta t})}{\phi\Delta t}, \qquad \hat{\lambda} = \lambda \frac{(e^{\beta(p+1)\Delta t} - 1)}{\beta(p+1)\Delta t}. \tag{C.10}$$

With large N and small Δt, we obtain the discrete strategy $\{Q_n\}$ by optimizing the loss function $-\mathcal{J}$ determined by:

$$\mathcal{J} = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} \Big(\hat{\alpha} Q_n f_{n-1} - \frac{\gamma\sigma^2}{2} Q_n^2 - \hat{\lambda}\beta |J_n^0 + Q_n|^{p+1}\Big)\right], \tag{C.11}$$

where

$$J_n^0 = J_n - Q_n = e^{-\beta\Delta t}(J_{n-1} + Q_n - Q_{n-1}) - Q_n = e^{-\beta\Delta t} J_{n-1}^0 - (1 - e^{-\beta\Delta t}) Q_n. \tag{C.12}$$

Evaluating Sharpe ratios After obtaining a discrete trading strategy by optimizing the discretized objective (C.11), we aim to evaluate its Sharpe ratio (SR). This allows us to compare the trading performance of the discrete and continuous strategies. To obtain accurate results, this step can be run with a long sequence length $M > N$, as only Monte-Carlo simulations, rather than optimization steps, are performed here.

In view of (C.5), (C.7), and (C.8), we can express the profit and loss PnL_M and the risk $\mathcal{R}_M$ in terms of the sequence $\{Q_n\}$:

$$\mathrm{PnL}_M(\{Q_n\}) = \mathbb{E}\Big[\frac{1}{M\Delta t}\Big(\int_0^{M\Delta t} Q_t dS_t - \int_0^{M\Delta t} \beta\lambda|J_t|^{p+1}dt\Big)\Big] = \mathbb{E}\Big[\frac{1}{M}\sum_{n=1}^{M}\Big(\hat{\alpha}Q_n f_{n-1} - \hat{\lambda}\beta|J_n|^{p+1}\Big)\Big],$$

$$\mathcal{R}_M(\{Q_n\}) = \sqrt{\mathbb{E}\Big[\frac{1}{M\Delta t}\int_0^{M\Delta t}\sigma^2 Q_t^2 dt\Big]} = \sqrt{\mathbb{E}\Big[\frac{1}{M}\sum_{n=1}^{M}\sigma^2 Q_n^2\Big]}.$$

The Sharpe ratio for the discrete strategy $\{Q_n\}$ is then evaluated as follows:

$$\mathrm{SR}(\{Q_n\}) = \frac{\mathrm{PnL}_M(\{Q_n\})}{\mathcal{R}_M(\{Q_n\})}. \tag{C.13}$$

C.2 Deep Learning Algorithm

In this section, we describe our machine-learning method for optimizing the objective function (C.11). We employ a method akin to policy gradient methods from reinforcement learning (Sutton and Barto, 2018): we parametrize strategies by deep neural networks and then optimize the neural network parameters using stochastic gradient descent-type algorithms, with gradients computed through algorithmic differentiation by the backpropagation algorithm.

Algorithm overview Let us now describe in more detail the deep learning approach employed here. The dynamic programming principle suggests that the optimal holdings Q_n are a feedback function of the same state variables as for the linear version of the model discussed in Section 4, namely the current signal f_{n-1} and the transformed moving average of past trades J^0_{n-1}; see (C.12).

Motivated by this observation, we parametrize the current holdings Q_n by a neural network NN_θ with input (f_{n-1}, J^0_{n-1}). With the new holdings Q_n chosen, we then update the moving average of trades J^0_n according to (C.12). Once the next signal arrives, we compute the new holdings $\mathrm{NN}_\theta(f_n, J^0_n)$. Note that in this iterative procedure the neural network NN_θ is applied repeatedly and the input J^0_n depends on the previous neural network output. This can be summarized in the following general structure:

$$Q^\theta_n = \mathrm{NN}_\theta(f_{n-1}, J^{0,\theta}_{n-1}), \tag{C.14}$$

$$J^{0,\theta}_n = e^{-\beta\Delta t}J^{0,\theta}_{n-1} - (1 - e^{-\beta\Delta t})Q^\theta_n \tag{C.15}$$

$$J^\theta_n = Q^\theta_n + J^{0,\theta}_n, \tag{C.16}$$

where NN_θ, parameterized by θ, is the function we aim to learn during the training phase and the θ-superscripts indicate that the holdings and moving averages depend on the parameter θ.

For any given choice of neural network parameters θ, the procedure described in (C.14)–(C.16) thus takes the sequence $\{f_n\}$ as input and outputs $\{Q^\theta_n\}$ as well as $\{J^\theta_n\}$. Inserting these sequences in the discrete objective function (C.11) in turn leads to the following optimization problem:

$$\mathrm{argmax}_\theta \quad \mathbb{E}\left[\sum_{n=1}^{N}\left(\hat{\alpha}Q^\theta_n f_{n-1} - \frac{\gamma\sigma^2}{2}(Q^\theta_n)^2 - \hat{\lambda}\beta|J^\theta_n|^{p+1}\right)\right]. \tag{C.17}$$

We use stochastic gradient-type algorithms to iteratively update the parameter θ and compute an approximately optimal parameter θ^*. These algorithms require estimates of the gradient of (C.17) with respect to the parameters θ, which can be computed efficiently using algorithmic differentiation via the backpropagation algorithm. We refer, for example, to Goodfellow et al. (2016) and Buehler et al. (2019) for standard references that explain these algorithms in detail. While recent literature has developed various methods for making the iterative training process of the parameter θ faster (Mueller et al., 2024) or more robust against adversarial perturbations (He et al., 2026), we focus here on the most commonly used training algorithms.

Neural network architecture In our experiments, we tested different choices of the network NN_θ. One straightforward form of NN_θ is an ordinary feed-forward neural network. This is a standard choice that does not incorporate any domain knowledge about the problem. To wit,

$$\mathrm{NN}_\theta(x) = L_{8,1} \circ \sigma \circ L_{32,8} \circ \sigma \circ L_{128,32} \circ \sigma \circ L_{2,128}(x), \tag{C.18}$$

where $L_{d_1,d_2} : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ are affine functions and $\sigma(r) = \max(r, 0)$ is the ReLU activation function applied componentwise.

Alternatively, motivated by results for models without risk constraints (Hey et al., 2025), one can try to exploit that the optimal impact state $|I_t| \propto |Q_t + J_t^0|^p$ typically has a far more linear relationship to the state variables than the corresponding optimal position Q_t. This suggests that a ReLU neural network of the form (C.18), for example, can approximate the impact state more efficiently. Equivalently, one can reparametrize (C.14) by

$$Q_n^\theta = \mathrm{NL}(\mathrm{NN}_\theta(f_{n-1}, J_{n-1}^0)) - J_{n-1}^0, \tag{C.19}$$

where

$$\mathrm{NL}(x) = \begin{cases} \mathrm{sign}(x)|x|^{1/p}, & \text{if } |x| < B, \\ ax + \mathrm{sign}(x)b, & \text{otherwise,} \end{cases} \tag{C.20}$$

with the neural network NN_θ defined as in (C.18) and the state variable $J^0 = J^{0,\theta}$ updated according to (C.15). In this formulation, B represents a large threshold beyond which the nonlinearity $\mathrm{NL}(x)$ transitions to a linear function. The parameters a and b are uniquely determined by the condition that $\mathrm{NL}(x)$ is continuously differentiable at the boundary. This design ensures that within the bound B, the nonlinear relation between impact states and positions is captured; outside this region, $\mathrm{NL}(x)$ behaves linearly to prevent blow-up in the output when the network is applied iteratively.

Finally, we also evaluated a linear discrete strategy parametrized by a single linear layer $\mathrm{NN}_\theta = L_{2,1}$, and compared its performance against the linear continuous strategy detailed in Section 5 to explore differences between discrete and continuous approaches.

The network strategies mentioned in Section 6 were trained using a sequence length of $N = 5120$ and stepsize of $\Delta t = 0.01$ (i.e., 100 trades per day). For evaluating the Sharpe ratio in (C.13) the strategies were then evaluated using a longer sequence length corresponding to 10 years of trading ($M = 256,000$ for $\Delta t = 0.01$, averaged over $5,000$ sample paths).

Comparison of network architectures In our experiments, the second network choice (C.19)–(C.20) consistently achieved the best performance. Whence, in the discussion in Section 6 we only focus on this network architecture. For completeness, we also report here a detailed comparison of the three architectures.

As summarized above, three types of networks are trained. **NetSimple** is a simple network where NN_θ is defined as in (C.18) and the corresponding holdings are given by (C.14)–(C.16). **NetPower** incorporates a nonlinear transformation as in (C.19). Finally, we consider the linear network **NetLinear**, where $\mathrm{NN}_\theta = L_{2,1}$ is an affine function and the corresponding holdings are given by (C.14)–(C.16).

Figure 7 reports the dynamics of the objective function (C.17) on out-of-sample validation data during training. At the end of training, **NetPower** reaches a somewhat higher value of the objective function than the other two networks. This result aligns with the performance of the three strategies shown in Table 2. **NetPower** achieves a higher Sharpe Ratio (SR), at similar risk levels, compared to the other networks, while the performance of **NetSimple** lies between **NetPower** and **NetLinear**.

Table 2: Out-of-sample performance of different network architectures

Network	**SR**	**Daily Risk**
NetSimple	1.443	1,223,859
NetLinear	1.438	1,209,254
NetPower	1.461	1,252,336

Implementation details We now detail the hyper-parameters used in training. For each gradient update step we simulate a batch of 1024 signals with sequence length $N = 5120$ according to the dynamics described by equation (C.6). These signals are used as inputs to the neural network to obtain, for each

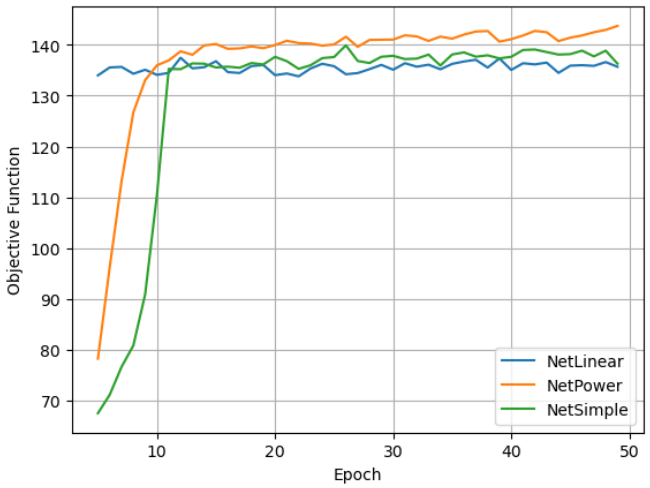


Figure 7: Evolution of the objective function during training, evaluated on out-of-sample validation data.

signal, the corresponding trading strategy according to (C.14)–(C.16). The objective function in (C.17) is then computed by a Monte Carlo approximation based on the 1024 samples and optimized using the ADAM optimizer (proposed in Kingma and Ba (2015)) across 50 batches per epoch for a total of 50 epochs. The learning rate is set to 10^{-4} (and 10^{-2} for the linear network) for the first 30 epochs, then multiplied by 1/4 every 5 epochs thereafter. The risk-aversion parameter γ in (C.17) is set to 0.9/ADV, which yields a DR of approximately 1.3M for the strategy produced by **NetPower**. The parameter γ can be adjusted to generate strategies with different risk levels.

A crucial training technique to approximate the ergodic problem at hand involves non-zero initialization of strategy states. To wit, the initial signal f_0, holding Q_0, and the price impact J_0 are not set to zero, but instead sampled from a three-dimensional normal distribution. The covariance matrix for this distribution is dynamically updated throughout training based on the strategies generated in the preceding gradient update step. Specifically, for each gradient update step, we compute the sample covariance matrix from "batch size $\times N$" triples (f_n, Q_n, J_n) and update the covariance matrix using a moving average:

$$\mathrm{Cov}_{\mathrm{distribution}} := 0.999\mathrm{Cov}_{\mathrm{distribution}} + 0.001\mathrm{Cov}_{\mathrm{sample}}. \tag{C.21}$$

This technique is crucial for several reasons. Firstly, a large price impact encourages the network to produce significant outputs, resulting in strategies with a high initial risk level during the early stages of training. Without this technique, transitioning to the desired risk level can be lengthy and sometimes unsuccessful, thus it effectively speeds up the training process. Secondly, we aim to evaluate the objective function in an ergodic state where the triplet (f_n, Q_n, J_n) follows a time-independent ergodic distribution. It is, therefore, reasonable to aim at initializing (f_0, Q_0, J_0) from this distribution, which we approximate as a normal distribution. This approach ensures that the initial conditions are representative of the expected long-term behaviour of the strategy and allows us to accurately train the networks for the ergodic problem with a rather short time series of simulated data.

An additional adjustment involves scaling the signal component of the neural network input by 10^6. This scaling balances the input magnitude between the signal and the price impact, which is essential for effective network training.

D Comparison of Neural Network and Viterbi Strategies

In this section, we compare the performance of the neural network strategies to the Viterbi algorithm of Kolm and Ritter (2014) used in Brokmann et al. (2025) for models with instantaneous impact. To obtain feasible runtimes for the Viterbi algorithm, we focus on the discrete-time setting with one trade

per day from Brokmann et al. (2025), where the objective is to maximize

$$J(Q) = \mathbb{E}\Big[\frac{1}{N}\sum_{n=1}^{N} \alpha Q_n f_n - \frac{\gamma\sigma^2}{2} Q_n^2 - \lambda|\Delta Q_n|^p\Big]$$

for discrete trades ΔQ_n and where the exponent $p > 1$ determines the form of the superlinear costs. Note that no adjustment to parameters α and λ is necessary in this case, as we directly work in the discrete-time case. For neural network training we use a sequence length of $N = 512$ days. For evaluating the Sharpe ratio, we then use $N = 25,600$ time steps, corresponding to around 100 years of trading. We employ model parameters as in Brokmann et al. (2025) and the same neural network hyperparameters as in Section C.2.

Similar to Section C.2, we use neural network strategies with the following general structure:[15]

$$Q_n = \mathrm{NN}_\theta(f_n, Q_{n-1}) \quad \text{for} \quad n = 1, 2, ... \tag{D.1}$$

For comparison we train three types of networks. **NetSimple** is a standard feed-forward network where NN_θ is defined as in (C.18). **NetPretrain** has the same architecture as **NetSimple**, but NN_θ is pretrained using the Viterbi strategy, aiming to minimize the mean-squared error between the neural network output and the Viterbi updating rule from Brokmann et al. (2025) by supervised learning. **NetPower** does not use any such information, but instead incorporates a nonlinear component given by (C.20), and trains the more linear residual component.

These three neural network strategies and the Viterbi strategy from Kolm and Ritter (2014); Brokmann et al. (2025) are evaluated by computing the corresponding Sharpe Ratios. For each of those strategies, we also generate the best linear strategy with the same risk level by using the procedure proposed in Brokmann et al. (2025) and compute its Sharpe ratio. The results are reported in Table 3.

Table 3: Out-of-sample SR comparison of neural network, Viterbi and linear strategies.

Network	**SR**	**Daily Risk**	**Linear SR**	**Improvement over Linear SR**
NetSimple	1.5814	1,236,161	1.5745	0.438%
NetPower	1.5900	1,245,040	1.5716	1.174%
NetPretrain	1.5919	1,247,775	1.5707	1.353%
Viterbi	1.5772	1,285,591	1.5584	1.209%

From the table, we observe that – unlike the simplest network **NetSimple** – **NetPretrain** and **NetPower** exhibit similar performances to the Viterbi strategy. Notably, **NetPower** achieves this without leveraging any information from the Viterbi strategy. These consistent results suggest that the outputs of the algorithms indeed approximate the optimizer of the problem well. In doing so, employing the neural network strategies yields a significant improvement in computational efficiency. Indeed, using an NVIDIA GeForce RTX 4060 Laptop GPU, training **NetSimple** and **NetPretrain** takes only 15 minutes, while **NetPower** requires 23 minutes. In contrast, it takes almost one day to generate the Viterbi strategy.

E Upper Duality Bound

In this appendix, we derive an upper duality bound for our primal goal function. By training a second class of neural networks to evaluate the latter, we can in turn assess whether the neural-networks strategies trained in the previous sections are indeed almost optimal in our concrete specifications. As in Appendices C and D we focus on the case of a single exponential decay timescale.

E.1 Dual upper bound

For a state (f, J^0) and control q, following the objective in (C.1), define the running reward

$$r(f, J^0, q) = \alpha f q - \beta\lambda|J^0 + q|^{1+p} - \frac{\gamma\sigma^2}{2} q^2. \tag{E.1}$$

[15] For this problem with instantaneous impact, dynamic programming suggests that the state variables parametrizing the optimal strategy are the signal f_n and the incoming holdings Q_{n-1}, which motivates (D.1).

Then the long-run objective is

$$\mathcal{J}(Q) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T r(f_t, J_t^0, Q_t)\, dt\right], \tag{E.2}$$

for strategies $Q \in \mathcal{Q}$ of (3.2). To establish the duality bound we will need to introduce a class of test functions $\mathcal{V}$ consisting of all functions $(f, J^0) \to V(f, J^0)$ that are continuously differentiable in (f, J^0) and twice continuously differentiable in f, and satisfy the polynomial growth condition

$$|V(f, J^0)| + |V_f(f, J^0)|^2 \le C\left(1 + |f|^m + |J^0|^2\right), \qquad (f, J^0) \in \mathbb{R}^2, \tag{E.3}$$

for some constant $C > 0$, $m \ge 1$. We then define, for fixed q, the controlled generator as

$$\mathcal{L}^q V = \phi\big(V_{ff} - fV_f\big) - \beta(q + J^0)V_{J^0}, \tag{E.4}$$

and the dual objective functionals

$$S_V(f, J^0) = \sup_{q\in\mathbb{R}}\{r(f, J^0, q) + \mathcal{L}^q V(f, J^0)\}, \qquad c(V) = \sup_{(f,J^0)\in\mathbb{R}^2} S_V(f, J^0). \tag{E.5}$$

Proposition E.1 (Weak duality)**.** *We have* $\sup_{Q\in\mathcal{Q}} \mathcal{J}(Q) \le \inf_{V\in\mathcal{V}} c(V)$.

Proof. Fixing arbitrary $Q \in \mathcal{Q}$, $V \in \mathcal{V}$, the definitions in (E.5) imply

$$r(f_t, J_t^0, Q_t) + \mathcal{L}^{Q_t} V(f_t, J_t^0) \le c(V). \tag{E.6}$$

By Itô's formula,

$$\begin{aligned} V(f_T, J_T^0) - V(f_0, J_0^0) &= \int_0^T \mathcal{L}^{Q_t} V(f_t, J_t^0)\, dt + \sqrt{2\phi}\int_0^T V_f(f_t, J_t^0)\, dW_t, \\ \mathbb{E}[V(f_T, J_T^0)] - \mathbb{E}[V(f_0, J_0^0)] &= \mathbb{E}\left[\int_0^T \mathcal{L}^{Q_t} V(f_t, J_t^0)\, dt\right], \\ \frac{1}{T}\mathbb{E}\left[\int_0^T r(f_t, J_t^0, Q_t)\, dt\right] &\le c(V) - \frac{\mathbb{E}[V(f_T, J_T^0)] - \mathbb{E}[V(f_0, J_0^0)]}{T}. \end{aligned} \tag{E.7}$$

Since $\mathbb{E}[|f_t|^{2m}]$ and $\mathbb{E}[|J_t^0|^2]$ are bounded uniformly on compacts in t, we can deduce from the estimate (E.3) that $\mathbb{E}[\int_0^T V_f(f_t, J_t^0)^2\, dt] < \infty$, so that the stochastic integral is a true martingale. Moreover, (E.3) gives

$$\frac{1}{T}\mathbb{E}\big[|V(f_T, J_T^0)|\big] \le \frac{C}{T}\left(1 + \mathbb{E}[|f_T|^m] + \mathbb{E}[|J_T^0|^2]\right). \tag{E.8}$$

The Gaussian dynamics of f ensure that $\frac{1}{T}\mathbb{E}[|f_T|^m] \to 0$, while $J_T^0 = J_T - Q_T = -\beta\int_0^T e^{-\beta(T-s)}Q_s ds$ gives

$$\mathbb{E}[(J_T^0)^2] = \mathbb{E}\left[\left(\int_0^T \beta e^{-\beta(T-s)}Q_s ds\right)^2\right] \le \beta\int_0^T e^{-\beta(T-s)}\mathbb{E}[Q_s^2] ds,$$

where the inequality follows from Cauchy–Schwarz applied to the measure $\mu(ds) = \beta e^{-\beta(T-s)}ds$ on $[0, T]$. By the transversality condition defining $\mathcal{Q}$ in (3.2), for any $\epsilon > 0$ there exists $T^* > 0$ such that $\mathbb{E}[Q_s^2] \le \epsilon s$ for $s \ge T^*$. As such, by splitting the integral at T^* we obtain

$$\frac{1}{T}\mathbb{E}[(J_T^0)^2] \le \frac{\beta e^{-\beta(T-T^*)}}{T}\int_0^{T^*} \mathbb{E}[Q_s^2] ds + \epsilon.$$

Now sending $T \to \infty$ and then $\epsilon \downarrow 0$ we obtain $\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[(J_T^0)^2] = 0$. Thus, from (E.8), the last term of (E.7) vanishes as $T \to \infty$ and we have $\mathcal{J}(Q) \le c(V)$, implying the weak duality bound. □

E.2 Training the dual value function

Following Appendix E.1, we train an optimal value function V that minimizes $c(V)$ and, thereby, gives an upper bound for the optimal risk-adjusted PnL $\sup_{Q\in\mathcal{Q}} \mathcal{J}(Q)$.

Fibre-loss form. To accomplish this we exploit some structure prevalent in the fibre-loss obtained by optimizing over only the variable J^0. For $V \in \mathcal{V}$, we define

$$M_V(f) = \sup_{J^0 \in \mathbb{R}} S_V(f, J^0), \qquad \mathcal{L}_{\text{fibre}}(V) = \mathbb{E}[M_V(Z)], \quad Z \sim N(0,1). \tag{E.9}$$

Then we have the following reduction for the dual problem.

Proposition E.2 (Fibre-loss reduction)**.** *For every $V \in \mathcal{V}$, such that M_V is continuous and at most of polynomial growth of order m in f for some $m > 0$, there exists a function $R(f)$ such that $V + R \in \mathcal{V}$ and*

$$c(V + R) = \mathcal{L}_{\text{fibre}}(V). \tag{E.10}$$

Proof. By assumption, M_V is integrable under the standard Gaussian measure. Set $C = \mathbb{E}_{Z\sim N(0,1)}[M_V(Z)]$ and consider a solution R to the ODE

$$\mathcal{A}R(f) = C - M_V(f), \tag{E.11}$$

where $\mathcal{A}R(f) = \phi(R'' - fR')$. One such solution is obtained by direct integration leading to

$$R'(f) = \frac{e^{f^2/2}}{\phi} \int_{-\infty}^{f} \big(C - M_V(u)\big) e^{-u^2/2}\, du = \frac{e^{f^2/2}}{\phi} \int_{f}^{\infty} (M_V(u) - C) e^{-u^2/2}\, du. \tag{E.12}$$

As $|f| \to \infty$, by the Gaussian-tail estimate we have

$$|R'(f)| \le c_M \frac{e^{f^2/2}}{\phi} \int_f^\infty (1 + u^m) e^{-u^2/2}\, du \sim \frac{1}{\phi} \left(\frac{1}{f} + f^{m-1} \right) \sim \frac{1}{\phi} f^{m-1}, \tag{E.13}$$

for some constant $c_M > 0$. This shows that $V + R \in \mathcal{V}$. Moreover, since R depends only on f, we have the relationship $S_{V+R}(f, J^0) = S_V(f, J^0) + \mathcal{A}R(f)$. Therefore,

$$\sup_{J^0 \in \mathbb{R}} S_{V+R}(f, J^0) = M_V(f) + \mathcal{A}R(f) = C \tag{E.14}$$

for every f, and hence $c(V + R) = C = \mathcal{L}_{\text{fibre}}(V)$.

□

Proposition E.2 suggests training V_θ by minimizing

$$\mathcal{L}_{\text{fibre}}(\theta) = \mathbb{E}_{Z\sim N(0,1)} \left[\sup_{J^0 \in \mathbb{R}} S_{V_\theta}(Z, J^0) \right]. \tag{E.15}$$

In practice, to cover large signal realizations more effectively during the training, instead of sampling Z from the standard normal law, we instead sample Z from the higher variance Gaussian distribution $N(0, 1.5)$, and then reweight the loss by the ratio of the two densities.

Architecture and tail structure. Within a bounded J^0 range, we use a scaled Softplus network N_θ with hidden widths $(128, 32, 8)$, subtracting its output at the origin so that $N_\theta(0, 0) = 0$, and apply the central-even projection

$$\widetilde{N}_\theta(f, J^0) = \frac{1}{2} \left\{ N_\theta(f, J^0) + N_\theta(-f, -J^0) \right\}, \tag{E.16}$$

which enforces the joint sign symmetry of the model. Let $b > 0$ be a tangent cap and set

$$a_+(f) = \partial_{J^0} \widetilde{N}_\theta(f, b), \qquad a_-(f) = \partial_{J^0} \widetilde{N}_\theta(f, -b). \tag{E.17}$$

We extend the full value function outside the central region via

$$V_\theta(f, J^0) = \begin{cases} \widetilde{N}_\theta(f, b) + a_+(f)(J^0 - b), & J^0 > b, \\ \widetilde{N}_\theta(f, J^0), & |J^0| \le b, \\ \widetilde{N}_\theta(f, -b) + a_-(f)(J^0 + b), & J^0 < -b. \end{cases} \tag{E.18}$$

This hard-tangent continuation preserves both the value and its J^0 derivative at $\pm b$, so V_θ is C^1 and $\partial_{J^0} V_\theta$ is constant in J^0 on each outer tail. In practice, we use a tangent cap of $b = 15$ million generally, and $b = 120$ million for high-risk scenarios. Additionally, the inner J_0 search is bounded to $[-2b, 2b]$ during the training only, which is computed by a grid search followed by a small refinement.

The following proposition verifies that the hard-tangent architecture defined in (E.18) has the required growth properties for the fibre-loss reduction. Consequently, the evaluated fibre loss can be interpreted as a bona fide dual upper bound on the primal objective.

Proposition E.3 (Polynomial growth of the fibre maximum)**.** *Let V_θ be the Softplus network with the hard-tangent continuation defined in* (E.18)*. Then M_{V_θ} is continuous and has polynomial growth in f.*

Proof. Since V_θ is a finite Softplus network and is affine in J^0 outside the central region, there exist $K_1, K_2 > 0$ such that

$$|(V_\theta)_{J^0}| \le K_1, \qquad |(V_\theta)_f| + |(V_\theta)_{ff}| \le K_2(1 + |J^0|). \tag{E.19}$$

In particular, $V_\theta \in \mathcal{V}$. Substituting these bounds into the definition of M_{V_θ} and using $|J^0| \le |J^0+q|+|q|$, there exists $K_3 > 0$ such that

$$M_{V_\theta}(f) \le \sup_{J^0,q} \Big\{ K_3(1+|f|) + K_3(1+|f|)\big(|q| + |J^0+q|\big) - \frac{\gamma\sigma^2}{2}q^2 - \beta\lambda|J^0+q|^{1+p} \Big\}.$$

By Young's inequality, we can separate $K_3(1+|f|)\big(|q| + |J^0+q|\big)$ into terms in $|f|$ and higher order terms,

$$K_3(1+|f|)|q| \le \varepsilon_1 q^2 + K_4(1+|f|)^2, \qquad K_3(1+|f|)|J^0+q| \le \varepsilon_2|J^0+q|^{1+p} + K_5(1+|f|)^{(1+p)/p}. \tag{E.20}$$

For sufficiently small ε_1 and ε_2, the coefficients of q^2 and $|J^0+q|^{1+p}$ remain negative. The remainder of the terms remains of polynomial growth in $|f|$. Hence, for some $K_6 > 0$ and finite m,

$$M_{V_\theta}(f) \le K_6(1+|f|^m). \tag{E.21}$$

On the other hand, by taking $J^0 = q = 0$, for some $K_7 > 0$, $M_{V_\theta}(f) \ge -K_7(1+|f|)$. Therefore, for some $K_8 > 0$, $|M_{V_\theta}(f)| \le K_8(1+|f|^m)$.

In addition, let $K \subset \mathbb{R}$ be compact. The preceding estimates imply that, uniformly for $f \in K$,

$$r(f, J^0, q) + \mathcal{L}^q V_\theta(f, J^0) \to -\infty \quad \text{as} \quad |q| + |J^0+q| \to \infty.$$

Hence the maximization defining $M_{V_\theta}(f)$ may be restricted to a compact set independent of $f \in K$. Since the objective is continuous in (f, J^0, q), M_{V_θ} is continuous on K, and therefore on $\mathbb{R}$.

□

Training scheme. We train the network using the Adam optimizer with a batch size of 512 and an initial learning rate of 10^{-3}, performing ten parameter updates per epoch. In the baseline configuration, training runs for 800 epochs: the initial learning rate is retained through epoch 200 and is subsequently halved every 50 epochs. To allow more extensive optimization in the high-risk scenario, we extend training to 3200 epochs and scale the duration of each learning-rate stage proportionally. Training is performed on a single NVIDIA GeForce RTX 4060 Laptop GPU. The 800-epoch runs take approximately 1 hours, while the 3200-epoch run takes approximately 4 hours.

E.3 Results

For each risk-aversion parameter γ, matched to that of the corresponding primal network policy, we train a value function V_θ. The reported numerical dual value $\mathcal{L}_{\text{fibre}}(V_\theta)$ is evaluated using a 512-node Gauss–Hermite rule in f. For the inner search over J_0, the search range is expanded until the score no longer increases. Table 4 reports these numerical dual values together with the primal policy values $\mathcal{J}(Q_{\text{NN}})$ and their ratio $\mathcal{J}(Q_{\text{NN}})/\mathcal{L}_{\text{fibre}}(V_\theta)$. Across all three risk-aversion levels, corresponding to DR levels of approximately 5.3M, 1.3M and 0.5M respectively, the ratio ranges from 97.7% to 98.7%. Thus, in view of our duality guarantees established in Proposition E.1, the value achieved by the neural-network strategy is close to the corresponding numerically approximated dual value, showing that the strategy indeed achieves near-optimal performance.

References

E. Abi Jaber and E. Neuman. Optimal liquidation with signals: the general propagator case. *Mathematical Finance*, 35(4):841–866, 2025.

E. Abi Jaber, A. Bondi, N. De Carvalho, E. Neuman, and S. Tuschmann. Fredholm approach to nonlinear propagator models. *Forthcoming in Finance and Stochastics*, 2025.

Table 4: Primal policy values and numerical dual values across risk-aversion levels. The values are in units of dollars per day per asset.

γ	$\mathcal{L}_{\text{fibre}}(V_\theta)$	$\mathcal{J}(Q_{\text{NN}})$	Ratio
0.1/ADV	428.9	419.2	97.7%
0.9/ADV	143.7	141.8	98.7%
2.9/ADV	67.37	66.48	98.7%

A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–39, 2000.

R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18 (7):57–62, 2005.

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

E. Bacry, A. Iuga, M. Lasnier, and C.-A. Lehalle. Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(2):1550009, 2015.

B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris bourse. *Journal of Finance*, 50(5):1655–1689, 1995.

R. C. Booton. Nonlinear control systems with random inputs. *IRE Transactions on Circuit Theory*, 1 (1):9–18, 1954.

J.-P.. Bouchaud and M. Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, UK, 2000.

J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance*, 4(2):176–190, 2004.

J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–160. North-Holland, Amsterdam, 2009.

J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould. *Trades, Quotes and Prices*. Cambridge University Press, Cambridge, UK, 2018.

X. Brokmann, E. Sérié, J. Kockelkoren, and J.-P. Bouchaud. Slow decay of impact in equity markets. *Market Microstructure and Liquidity*, 1(2):1550007, 2015.

X. Brokmann, D. Itkin, J. Muhle-Karbe, and P. Schmidt. Tackling nonlinear price impact with linear strategies. *Mathematical Finance*, 35(2):422–440, 2025.

F. Bucci, M. Benzaquen, F. Lillo, and J.-P. Bouchaud. Slow decay of impact in equity markets: insights from the ANcerno database. *Market Microstructure and Liquidity*, 4(3/4):1950006, 2018.

H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.

Á. Cartea, S. Jaimungal, and J. Penalva. *Algorithmic and High-Frequency Trading*. Cambridge University Press, Cambridge, UK, 2015.

Á. Cartea, S. Jaimungal, and L. Sánchez-Betancourt. Deep reinforcement learning for algorithmic trading. In *Machine Learning and Data Sciences for Financial Markets*, pages 230–250. Cambridge University Press, Cambridge, UK, 2023.

T. Cayé, M. Herdegen, and J. Muhle-Karbe. Trading with small nonlinear price impact. *Annals of Applied Probability*, 30(2):706–746, 2020.

T. Chen, M. Ludkovski, and M. Voß. On parametric optimal execution and machine learning surrogates. *Quantitative Finance*, 24(1):15–34, 2024.

P. Collin-Dufresne, K. Daniel, and M. Sağlam. Liquidity regimes and optimal dynamic asset allocation. *Journal of Financial Economics*, 136(2):379–406, 2020.

G. Curato, J. Gatheral, and F. Lillo. Optimal execution with non-linear transient market impact. *Quantitative Finance*, 17(1):41–54, 2017.

S. Denkl, M. Goy, J. Kallsen, J. Muhle-Karbe, and A. Pauwels. On the performance of delta hedging strategies in exponential Lévy models. *Quantitative Finance*, 13(8):1173–1184, 2013.

M. Fazel, R. Ge, S. Kakade, and M. Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1467–1476. Proceedings of Machine Learning Research, Stockholm, Sweden, 2018.

N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340, 2013.

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

J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3):445–474, 2012.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.

G. Gripenberg. An abstract nonlinear Volterra equation. *Israel Journal of Mathematics*, 34(3):198–212, 1979.

P. Guasoni and M. H. Weber. Nonlinear price impact and portfolio choice. *Mathematical Finance*, 30 (2):341–376, 2020.

C. R. Harvey and Y. Liu. Backtesting. Preprint, 2015.

J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207, 1991.

G. He, T. Sutter, and L. Gonon. Distributional Adversarial Attacks and Training in Deep Hedging. *Advances in Neural Information Processing Systems*, 38:131893–131925, 2026.

N. Hey, J.-P. Bouchaud, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. The cost of misspecifying price impact. *Risk*, 2024.

N. Hey, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. Trading with concave price impact and impact decay – theory and evidence. *Operations Research*, 73(3):1230–1247, 2025.

N. Hey, I. Mastromatteo, and J. Muhle-Karbe. Concave cross impact. Preprint, 2026.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

P. Kolm and G. Ritter. Multiperiod portfolio selection and Bayesian dynamic models. *Risk*, 18(7):50–54, 2014.

P. Kolm and G. Ritter. Dynamic replication and hedging: A reinforcement learning approach. *Journal of Financial Data Science*, 1:159–171, 2019.

L. Leal, M. Laurière, and C.-A. Lehalle. Learning a functional control for high-frequency finance. *Quantitative Finance*, 22(11):1973–1987, 2022.

W. Levine and M. Athans. On the determination of the optimal constant output feedback gains for linear multivariable systems. *IEEE Transactions on Automatic Control*, 15(1):44–48, 1970.

T. F. Loeb. Trading cost: The critical link between investment information and results. *Financial Analysts Journal*, 39(3):39–44, 1983.

P. Makila and H. Toivonen. Computational methods for parametric LQ problems–A survey. *IEEE Transactions on Automatic Control*, 32(8):658–671, 1987.

H. Mania, A. Guy, and B. Recht. Simple random search of static linear policies is competitive for reinforcement learning. *Advances in Neural Information Processing Systems*, 31, 2018.

C. Moallemi and M. Sağlam. Dynamic portfolio choice with linear rebalancing rules. *Journal of Financial and Quantitative Analysis*, 52(3):1247–1278, 2017.

C. Moallemi and M. Wang. A reinforcement learning approach to optimal execution. *Quantitative Finance*, 22(6):1051–1069, 2022.

K. Mueller, A. Akkari, L. Gonon, and B. Wood. Fast deep hedging with second-order optimization. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 319–327, 2024.

J. Muhle-Karbe, Z. Wang, and K. Webster. Stochastic liquidity as a proxy for nonlinear price impact. *Operations Research*, 72(2):444–458, 2023.

E. Neuman and Y. Zhang. Statistical learning with sublinear regret of propagator models. *The Annals of Applied Probability*, 36(4):2873 – 2915, 2026.

A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

B. Park and B. Van Roy. Adaptive execution: Exploration and learning of price impact. *Operations Research*, 63(5):1058–1076, 2015.

A. Rajeswaran, K. Lowrey, E. V. Todorov, and S. M. Kakade. Towards generalization and simplicity in continuous control. *Advances in Neural Information Processing Systems*, 30, 2017.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, second edition, 2018.

K. Webster. *Handbook of Price Impact Modeling*. CRC Press, Boca Raton, FL, 2023.