13

The Propagator Model

It is better to be roughly right than precisely wrong.

(John Maynard Keynes)

In the previous two chapters, we have considered how individual market orders and whole metaorders impact the mid-price. In both cases, our study of impact has primarily revolved around measuring the change in price that occurs contemporaneously with the specified market order or metaorder. However, as we briefly discussed in Sections 11.3.1 and 12.4, impact is not only felt immediately, but rather evolves dynamically. In fact, the "reactional" component of impact is expected to be mostly transient. In this chapter, we introduce a model that seeks to describe how impact evolves over time, in such a way to avoid the **efficiency paradox** created by the long memory of order signs.

Similarly to the previous chapters, we restrict our attention to market order time, in which we advance the clock by a single unit for each market order arrival. In this framework, we introduce and study a simple, linear model called the **propagator model**, which expresses price moves in terms of the influence of past trades. We demonstrate that to avoid creating statistical arbitrage opportunities and ensure **market efficiency**, the propagator model must possess certain properties, which we subsequently use to refine the model towards a more specific form. Despite its simplicity, the propagator model provides deep insight into the nature of price impact in financial markets. To conclude the chapter, we frame the propagator model in a wider context and discuss various conceptual points and extensions.

13.1 A Simple Propagator Model

In Chapter 11, we discussed how strongly, on average, the arrival of a market order of a given volume υ impacts the mid-price. After the arrival occurs, there is no reason why the impact of a market order should remain visible indefinitely, because that would mean that the order's arrival shifted the entire supply and demand curves forever. As we will show in this section, if that was the case, then the long-range autocorrelations of market order signs would create strong autocorrelations in returns. This is at odds with widespread empirical observations of signature plots (see the discussion in Section 10.3), which are nearly flat. Therefore, market efficiency imposes that a large fraction of a market order's impact must relax over time.

To illustrate why this is the case, we first consider a naive model of impact, in which every trade causes a fixed, constant lag-1 impact that persists permanently. To consider this in detail, let us consider the mathematical properties of the return series that such a process would generate. Let

$$r_t := m_{t+1} - m_t \tag{13.1}$$

denote the change in mid-price from immediately before the t^{th} event until immediately before the $(t+1)^{\text{th}}$ event. We assume that each trade has a mean permanent impact equal to G_1.[1] Using this notation, our simple model translates into

$$r_t = G_1 \varepsilon_t + \xi_t, \tag{13.2}$$

where ξ_t is a random noise term that models fluctuations around the average impact contribution and that also captures price changes not directly attributed to trading (e.g. quote changes that occur due to the release of public news, even without any trades occurring). We will assume here that ξ_t is independent of the order flow ε_t, with $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[\xi^2] = \Sigma^2$.

Given the mid-price m_{t_0} at some previous time $t_0 \leq t$, the mid-price m_t is given by

$$m_t = m_{t_0} + G_1 \sum_{t_0 \leq n < t} \varepsilon_n + \sum_{t_0 \leq n < t} \xi_n. \tag{13.3}$$

Equation (13.3) highlights the non-decaying nature of impact in this model: whatever happened at all previous times is permanently imprinted on the price at time t.

When the signs of the trades are independent random variables with zero mean, this simple model predicts that the response function (see Equation (11.6)) is simply constant:

$$\begin{aligned}\mathcal{R}(\ell) &= \mathbb{E}[\varepsilon_t \cdot (m_{t+\ell} - m_t)],\\ &= \mathbb{E}\left[\varepsilon_t \cdot [(m_{t+\ell} - m_{t+1}) + (m_{t+1} - m_t)]\right],\\ &= G_1.\end{aligned} \tag{13.4}$$

[1] In the present context, G_1 coincides with the quantity $\mathcal{R}(1)$ considered in the previous chapters, but this correspondence no longer holds in the presence of sign autocorrelations, which we will address in the subsequent sections.

The model also makes the following prediction for the mid-price variogram (see Equation (2.1)):

$$\begin{aligned}\mathcal{V}(\ell) &:= \mathbb{V}[m_{t+\ell} - m_t], \\ &= \left(G_1^2 + \Sigma^2\right)\ell. \end{aligned} \tag{13.5}$$

Therefore, the model predicts that the mid-price obeys a pure diffusion with lag-independent diffusion coefficient (or squared volatility)

$$\mathcal{D}(\ell) = \frac{\mathcal{V}(\ell)}{\ell} = G_1^2 + \Sigma^2, \tag{13.6}$$

as is observed empirically in liquid markets (except perhaps at small ℓ).

Despite the appealing simplicity of these results, this naive model ignores an important ingredient of real markets: the autocorrelation of the order-sign series ε_t. As we discussed in Chapter 10, real market order sign series have strong **long memory**, with an autocorrelation function $C(\ell)$ that decays approximately according to a power law. Incorporating this autocorrelation structure into the model, we arrive at the following expression for the autocovariance of price returns for $\ell > 0$:

$$\begin{aligned}\mathbb{E}[r_t r_{t+\ell}] &= G_1^2 \mathbb{E}[\varepsilon_t \varepsilon_{t+\ell}], \\ &= G_1^2 C(\ell).\end{aligned}$$

Therefore, if we incorporate the autocorrelation structure of market order signs, then our simple model predicts that price returns become strongly autocorrelated. This would violate market efficiency, because price returns would be strongly predictable. If prices are to be efficient, then we must conclude that the empirically observed long memory of order flow rules out the above assumption of impact being a simple constant (and thus permanent).

How might we resolve this conundrum? In the next section, we will consider two possibilities: relaxing the assumption that price impact is permanent, or relaxing the assumption that lag-1 impact is independent of the past order flow. As we will see from our analysis, these two seemingly different approaches are actually closely related.

13.2 A Model of Transient Impact and Long-Range Resilience

As we have noted several times, the long-range autocorrelation of market order signs is an interesting and puzzling property of financial markets. The fact that returns remain uncorrelated despite this long-range autocorrelation in market order signs suggests that markets are somehow resilient to these clusters of highly

 https://doi.org/10.1017/9781316659335.020

predictable market orders in the same direction. To understand this interesting property more clearly, let's consider a more general version of Equation (13.3), in which we regard the impact not as a constant G_1, but instead as a function $G(\ell)$ that describes how the impact of the trade propagates through time. We call the function $G(\cdot)$ the **propagator**.

Using this notation, we can generalise Equation (13.3) as:

$$m_t = m_{t_0} + \sum_{t_0 \le n < t} G(t-n)\varepsilon_n + \sum_{t_0 \le n < t} \xi_n, \tag{13.7}$$

where ξ_n again captures all price moves not directly induced by trades (e.g. due to quote revisions during news announcements).

Clearly, the functional form of the propagator $G(\cdot)$ influences the values of m_t in the model. For $G(\ell)$ to capture the transient component of impact, it seems reasonable that $G(\ell)$ should decrease with increasing ℓ. The long-time limit

$$G_\infty := \lim_{\ell \to \infty} G(\ell)$$

captures the permanent reaction impact of a market order. The simple, constant propagator that we considered in Equation (13.3) corresponds to the function $G(\ell) = G_1$ for all lags.

For the sake of simplicity, we will assume that impact is independent of trade volume, time and market conditions. All three of these assumptions are gross oversimplifications that do not accurately reflect the behaviour observed in real markets (see, e.g., Sections 11.3.1 and 11.3.2), but by adopting them we greatly simplify the mathematics associated with formulating a first model. We will revisit and relax these assumptions within a more general model in Chapter 14.

Equation (13.7) defines a family of linear, **transient impact model**s (TIM). Within this family of models, reaction impact is parameterised by the propagator $G(\ell)$, which encodes how a market order executed at time t impacts the price at a later time $t + \ell$.

We can use the model in Equation (13.7) to write an expression for the return series:

$$r_t = \underbrace{G_1 \varepsilon_t}_{\text{immediate impact}} + \underbrace{\sum_{n<t} K(t-n)\,\varepsilon_n}_{\text{impact decay}} + \underbrace{\xi_t}_{\text{noise}}, \tag{13.8}$$

where $G_1 := G(1)$, and for $\ell \ge 1$,

$$K(\ell) := G(\ell + 1) - G(\ell),$$

which is the discrete derivative of $G(\ell)$. Therefore, the model predicts the return at time t is due not only to the most recent trade (whose contribution is $G_1\varepsilon_t$), but also to *all* previous trades. Introducing the convention $G(0) = 0$, Equation (13.8) can be written in a more compact form:

$$r_t = \sum_{n \le t} K(t-n)\varepsilon_n + \xi_t. \tag{13.9}$$

As we will now demonstrate, the model allows us to compute the response function $\mathcal{R}(\ell)$ (or observed impact) and the price variogram $\mathcal{V}(\ell)$ for an arbitrary function $G(\ell)$ and an arbitrary correlation between trades $C(\ell)$.

13.2.1 The Response Function

By plugging Equation (13.7) into Equation (13.4), one readily obtains the following expression for the **response function**:

$$\mathcal{R}(\ell) = G(\ell) + \sum_{0<n<\ell} G(\ell-n)C(n) + \sum_{n>0} [G(\ell+n) - G(n)]\, C(n). \tag{13.10}$$

If the terms in the market order sign series were uncorrelated, it would follow that $C(n) = 0$ for all $n > 0$, so only the first term of the expression would be non-zero, leading to $\mathcal{R}(\ell) = G(\ell)$. In this case, the observed impact $\mathcal{R}(\ell)$ and the bare reaction impact $G(\ell)$ would be one and the same thing. In reality, however, market order signs are strongly autocorrelated, so we need to consider the full expression.

In this model, the lag-1 impact $\mathcal{R}(1)$ (which we measured in Chapter 11) includes the decay of the impact of all previous correlated trades, and is given by

$$\mathcal{R}(1) = \sum_{n \ge 0} K(n)C(n) = G_1 + \sum_{n>0} K(n)C(n). \tag{13.11}$$

Because the impact of an individual trade should decay with time, one expects $K(n > 0) \le 0$. If trade signs are positively autocorrelated, one concludes that $\mathcal{R}(1) \le G_1$.

The asymptotic response function is given by

$$\mathcal{R}_\infty = G_\infty + \lim_{\ell \to \infty} \left\{ \sum_{0<n<\ell} G(\ell-n)C(n) + \sum_{n>0} [G(\ell+n) - G(n)]\, C(n) \right\}, \tag{13.12}$$

which in fact requires some special conditions for the result to be finite (i.e. non-divergent). We return to this discussion in Section 13.2.3.

For any $\ell \ge 0$, Equation (13.10) relates the bare reaction impact $G(\ell)$, which is *not* directly observable, to two quantities that are: the observed impact $\mathcal{R}(\ell)$ and the autocorrelation of trade signs $C(\ell)$. One could therefore use this equation to estimate $G(\ell)$ from empirical data. However, this direct method is very sensitive to finite-size effects, and therefore provides poor estimates of $G(\ell)$: in practice, ℓ

must be smaller than the maximum lag L that allows a reasonable estimation of $\mathcal{R}$ and C, beyond which it becomes difficult to separate the signal from noise.

An alternative approach to this problem is to use the lagged **sign-return correlation**, which is defined as

$$S(\ell) := \mathbb{E}[r_{t+\ell} \cdot \varepsilon_t] = \mathcal{R}(\ell+1) - \mathcal{R}(\ell). \tag{13.13}$$

The function $S(\cdot)$ can be regarded as the discrete derivative of $\mathcal{R}(\cdot)$. It is straightforward to find the analogue of Equation (13.10) for $S(\cdot)$:

$$S(\ell) = \sum_{n \geq 0} C(|n-\ell|)K(n), \tag{13.14}$$

where we have again used the convention $G(0) = 0$. This equation can also be written in matrix form, as

$$\vec{S} = \mathbf{C}\vec{K}, \tag{13.15}$$

where the matrix $\mathbf{M}$ is given by

$$\mathbf{C}_{n,m} = C(|n-m|); \qquad 0 \leq n, m \leq L.$$

By inverting the matrix $\mathbf{C}$, Equation (13.15) can be rewritten as

$$K(\ell) = \sum_{n=0}^{L} (\mathbf{C}^{-1})_{\ell,n} S(n).$$

Observe that calculating $K(\ell)$ only requires knowledge of the values of S for positive lags.

In Figure 13.1, we plot $\mathcal{R}(\ell)$ and $G(\ell)$ for one large-tick stock and one small-tick stock. As expected, $G(\ell)$ decays with increasing ℓ. Upon closer scrutiny, one finds that the long-time behaviour of $G(\ell)$ decays approximately according to a power-law,

$$G(\ell) \approx_{\ell \gg 1} \frac{\Gamma_\infty}{\ell^\beta}, \qquad \beta < 1,$$

(see the inset of Figure 13.1). This is the precise definition of the "resilience" property that we discussed at the beginning of Section 13.2: the propagator $G(\ell)$ decays all the way to zero, but so slowly that its sum over all ℓ is divergent. We call this property **long-range resilience**, because it parallels the definition of long memory (see Section 10.1) when the decay exponent γ of $C(\ell)$ is less than 1.

In summary, the impact of a trade dissipates over time, but does so very slowly. As we will now show by considering the price variogram $\mathcal{V}(\ell)$, this slow decay turns out to play a crucial role in ensuring that the mid-price remains diffusive in the face of the long-range autocorrelations of market order signs.

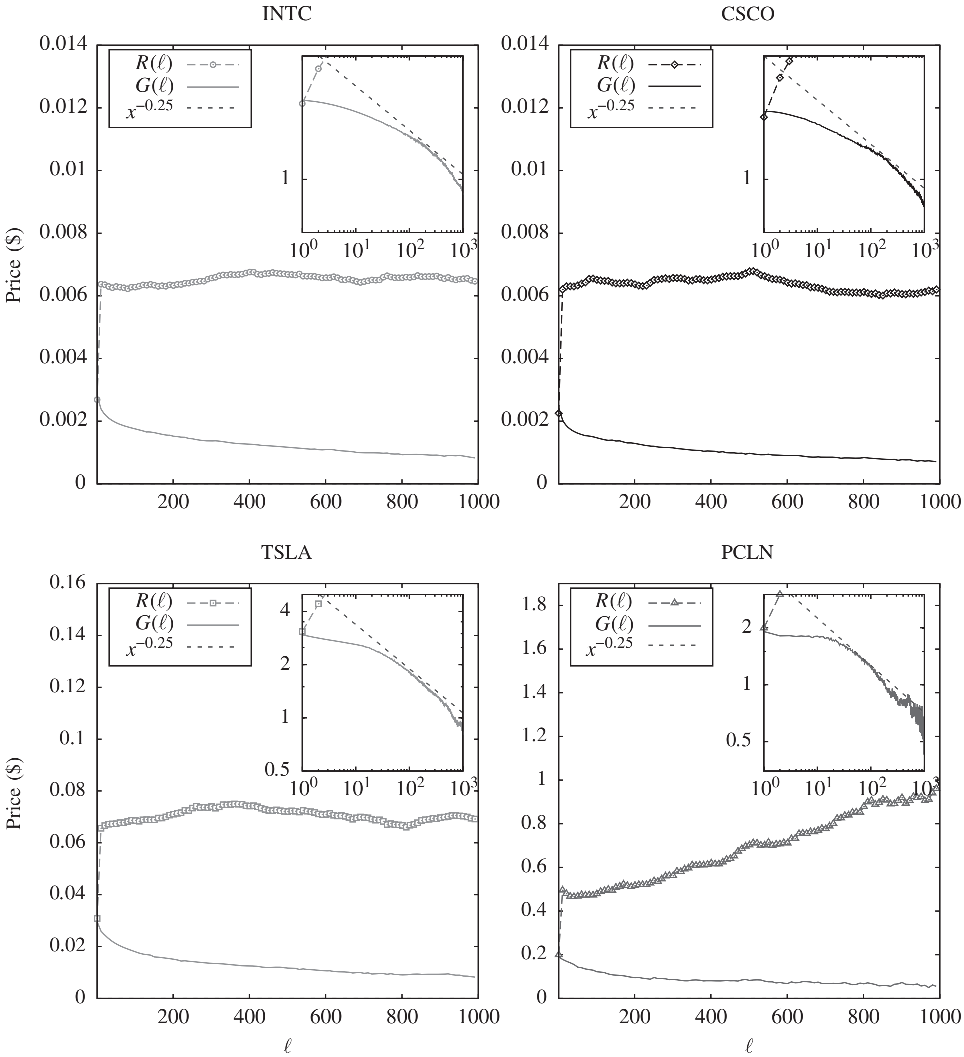

Figure 13.1. The empirical response function $\mathcal{R}(\ell)$ and the propagator $G(\ell)$, calculated from the empirical $\mathcal{R}(\ell)$ and $C(\ell)$, for (top left) INTC, (top right) CSCO, (bottom left) TSLA and (bottom right) PCLN. Inset: The same plots in doubly logarithmic axes. The dashed black lines depict a power-law with exponent −0.25.

13.2.2 The Variogram

Once $G(\ell)$ is determined, the propagator model provides a definite prediction for the **variogram**, shown in Figure 13.2. In full generality, the equation relating $\mathcal{V}(\ell)$ to $G(\ell)$ and $C(\ell)$ reads

$$\mathcal{V}(\ell) = \sum_{0 \le n < \ell} G^2(\ell - n) + \sum_{n > 0} [G(\ell + n) - G(n)]^2 + 2\Psi(\ell) + \Sigma^2 \ell, \tag{13.16}$$

where $\Psi(\ell)$ is the correlation-induced contribution to $\mathcal{V}$:

$$\Psi(\ell) = \sum_{0\le j<k<\ell} G(\ell-j)G(\ell-k)C(k-j)$$
$$+ \sum_{0<j<k} [G(\ell+j)-G(j)]\,[G(\ell+k)-G(k)]\,C(k-j)$$
$$+ \sum_{0\le j<\ell}\sum_{k>0} G(\ell-j)\,[G(\ell+k)-G(k)]\,C(k+j).$$

Let us analyse the large-ℓ behaviour of $\mathcal{V}(\ell)$, in the case where $C(\ell)$ decays as $c_\infty \ell^{-\gamma}$ with $\gamma < 1$ (which, as we saw in Chapter 10, is a good approximation of the large-ℓ behaviour of $C(\ell)$ for real markets). Motivated by the numerical result that we obtained for $G(\ell)$, we assume that for large ℓ, the propagator $G(\ell)$ decays as a power law $\Gamma_\infty \ell^{-\beta}$. When $0 < \beta < 1/2$ and $0 < \gamma < 1$, asymptotic analysis (with $\ell \gg 1$) of the first two terms of Equation (13.16) reveals that they both grow sub-linearly with ℓ:

$$\sum_{0\le n<\ell} G^2(\ell-n) \approx \sum_{0<n\le\ell} \frac{\Gamma_\infty^2}{\ell^{2\beta}} \approx \frac{\Gamma_\infty^2}{1-2\beta}\ell^{1-2\beta} \ll \Sigma^2\ell,$$

and

$$\sum_{n>0} [G(\ell+n)-G(n)]^2 \approx \Gamma_\infty^2 \ell^{1-2\beta}\int_0^\infty \mathrm{d}u\left(\frac{1}{(1+u)^\beta}-\frac{1}{u^\beta}\right)^2 = A(\beta)\Gamma_\infty^2\ell^{1-2\beta} \ll \Sigma^2\ell,$$

where A is a finite number,

$$A = \frac{1}{1-2\beta}\left(\frac{2\Gamma[2\beta]\Gamma[1-\beta]}{\Gamma[\beta]}-1\right); \qquad 0<\beta<\frac{1}{2}.$$

Therefore, only the contribution $\Psi(\ell)$ in Equation (13.16) may compete with the long-term diffusion behaviour of the price induced by the public news term ξ_t. Asymptotic analysis of $\Psi(\ell)$ yields the following contribution to the diffusion coefficient:

$$D_\Psi(\ell) := \frac{\Psi(\ell)}{\ell} \approx \Gamma_\infty^2 c_\infty I(\gamma,\beta)\,\ell^{1-2\beta-\gamma},$$

where $I(\gamma,\beta) > 0$ is a (complicated but finite) numerical integral.

We first examine what would happen if the propagator $G(\ell)$ did not decay at all ($\beta = 0$). This case corresponds to the assumption that impact is permanent. In this case, in the presence of long memory of the order flow $0 < \gamma < 1$, one finds that $D_\Psi(\ell) \propto \ell^{1-\gamma}$, which grows with increasing ℓ. This corresponds to a super-diffusive price – or, in financial language, the presence of persistent trends in the price series. This is the **efficiency paradox** created by the long-range correlation of market order signs.

As the propagator decays more quickly (i.e. as β increases), super-diffusion is less pronounced (i.e. $D_\Psi(\ell)$ grows more slowly), until

$$\beta = \beta_c := (1-\gamma)/2, \tag{13.17}$$

for which $D_\Psi(\ell)$ is a constant, independent of ℓ. Adding the variance of the "news" contribution Σ^2, the long-term volatility reads

$$\mathcal{D}_\infty = \Gamma_\infty^2 c_\infty I(\gamma,\beta_c) + \Sigma^2.$$

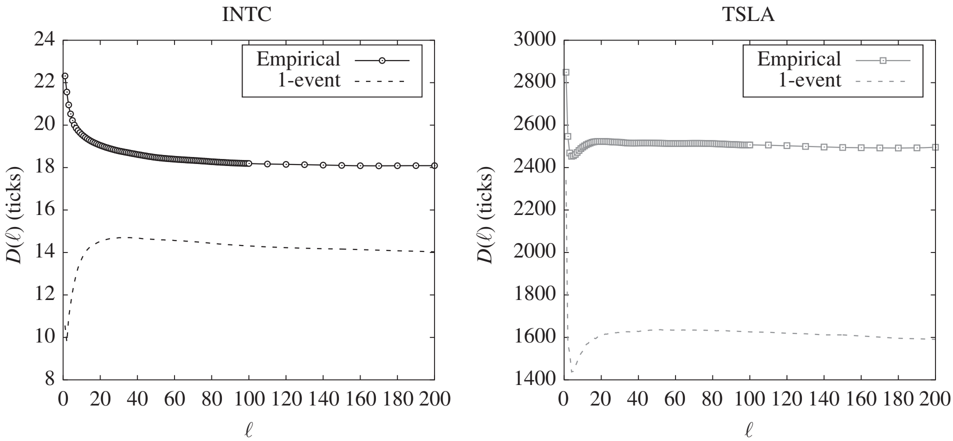

Figure 13.2. (Markers and solid curves) The empirical values of the function $\mathcal{D}(\ell)$ and (dashed curve) its approximation with the 1-event (market order) propagator model, Equation (13.16), with $\Sigma^2 = 0$. The left panel shows the results for INTC, which is a large-tick stock, and the right panel shows the results for TSLA, which is a small-tick stock. Remarkably, the trade-only contribution accounts for 0.65–0.8 of the long-term squared volatility. However, the model completely fails to capture the short-term structure of $\mathcal{D}(\ell)$, which suggests that one should also take into account some time-dependence of $G(\ell)$ (which might depend on event type and past history; see Chapter 14).

When $\beta > \beta_c$, $D_\Psi(\ell)$ decreases with ℓ, as do the contributions of the first two terms of Equation (13.16).[2] The lag-dependent volatility $\mathcal{D}(\ell)$ is in this case enhanced at short lags by the transient impact contribution, resulting in sub-diffusive prices at short lags.

Based on the calculations we have performed throughout this section, we can now reach an important conclusion: provided that the impact of single trades is transient (i.e. $\beta > 0$) and with a decay that precisely offsets the long-range autocorrelations in market order signs (i.e. $\beta = \beta_c := (1-\gamma)/2$), then the long-range autocorrelation of market order signs *is* compatible with a random-walk behaviour of the mid-price, and thus with statistical efficiency.

13.2.3 A Fine Balance

When the "fine balance" condition $\beta = \beta_c$ holds, we are in a rather odd situation where impact is not permanent (since in the long-time limit $\ell \to \infty$, the propagator $G(\ell)$ is zero) but is not really transient either, because the decay is extremely slow. The convolution of this **semi-permanent impact** balances exactly the slow decay of trade correlations to produce a finite contribution to the long-term volatility.

[2] Since $\gamma < 1$, it follows that $2 - 2\beta - \gamma > 1 - 2\beta$, and therefore at large ℓ the term $D_\Psi(\ell)$ always dominates the first two contributions of Equation (13.16) to $\mathcal{D}(\ell)$.

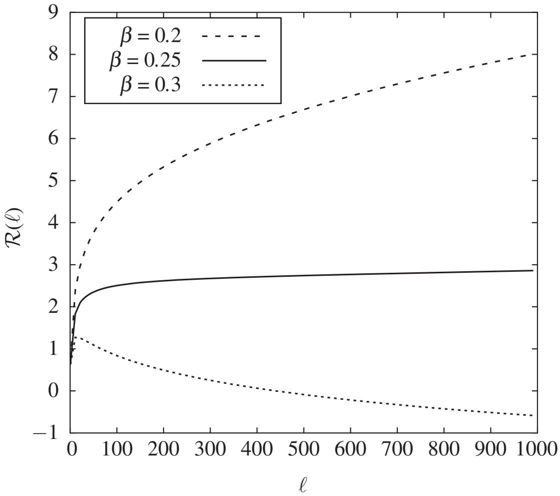

Figure 13.3. Calculation of $\mathcal{R}(\ell)$ for a synthetic propagator $G(\ell) \sim \ell^{-\beta}$ and a synthetic correlation function $C(\ell) \sim \ell^{-1/2}$, with $\beta \in \{0.2, 0.25, 0.35\}$. Observe that after an initial rapid increase, $\mathcal{R}$ is approximately flat for the parameter choice $\beta = \beta_c = 0.25$.

In fact, plugging a power law for $G(\ell)$ into Equation (13.10) leads to a response function $\mathcal{R}(\ell)$ that for large ℓ behaves as

$$\mathcal{R}(\ell) \approx_{\ell \gg 1} \Gamma_\infty c_\infty \frac{\Gamma[1-\gamma]}{\Gamma[\beta]\Gamma[2-\beta-\gamma]} \left[\frac{\pi}{\sin \pi\beta} - \frac{\pi}{\sin \pi(1-\beta-\gamma)} \right] \ell^{1-\beta-\gamma} \tag{13.18}$$

provided $\beta + \gamma < 1$. We display the explicit numerical prefactor to highlight that when $\ell \to \infty$, the value of $\mathcal{R}(\ell)$ diverges to $+\infty$ whenever $\beta < \beta_c$, and diverges to $-\infty$ whenever $\beta > \beta_c$. If $\beta = \beta_c$, then

$$\frac{\pi}{\sin \pi\beta_c} = \frac{\pi}{\sin \pi(1-\beta_c-\gamma)},$$

so the leading term in $\mathcal{R}(\ell)$ vanishes. In this case, one must also estimate the sub-leading term, which saturates to a finite value $\mathcal{R}(\infty)$. Hence, for $\beta = \beta_c$, the total asymptotic response to a trade (including all future correlated trades) is non-zero, even though the impact of each individual trade $G(\ell)$ decays to zero!

Figure 13.3 summarises this rather peculiar situation. Interestingly, we see that financial markets operate in a fragile regime where liquidity providers and liquidity takers offset each other, such that most of the predictable patterns that would otherwise be created by their predictable actions are removed from the price trajectory. Only an unpredictable contribution remains, even at the highest frequencies. We will return to the question of how this occurs in Chapter 14, where we will attempt to model the impact of all LOB events, not only that of market orders.

13.3 History-Dependent Impact Models

An alternative way to ensure the statistical efficiency of prices is to assume that the price impact of each order is permanent but history-dependent. Indeed, one way to rephrase the fact that price returns are difficult to predict is that most of the available information at time t is already included in the price, such that price moves only occur if something truly unexpected happens. In other words, since the signs of future trades are predictable, only the **surprise component** of the next sign should have an impact on the price. Why should this be the case? One possible explanation is that high-frequency participants act in such a way as to remove any predictability – e.g. by providing liquidity on the side where the next trade is most likely to happen (see Section 11.3.4).

13.3.1 The Impact of Surprises

We extend the expression in Equation (13.2) to formulate a version of the propagator model in which returns are unpredictable, but still linear in ε_t. There are two main ingredients to this approach. The first is to replace the (highly predictable) order-sign term ε_t term with the corresponding sign surprise $(\varepsilon_t - \widehat{\varepsilon_t})$, where $\widehat{\varepsilon_t}$ is the conditional expectation of ε_t:

$$\widehat{\varepsilon_t} := \mathbb{E}_{t-1}[\varepsilon_t].$$

In this way, we replace the ε_t term, whose behaviour depends only on the market order arrival at time t, with a term that also reflects the previous market order arrivals. The second ingredient to replace the constant term G_1 in Equation (13.2) with a term that may depend on t (to reflect, for example, the spread dynamics). Together, these ingredients lead to the following **history-dependent impact model** (HDIM):

$$r_t = G_{1,t} \times (\varepsilon_t - \widehat{\varepsilon_t}) + \xi_t. \tag{13.19}$$

Since neither the sign surprise $(\varepsilon_t - \widehat{\varepsilon_t})$ nor ξ_t can be predicted, it follows that for any immediate impact $G_{1,t}$,

$$\mathbb{E}_{t-1}[r_t] = G_{1,t}\mathbb{E}_{t-1}\left[\varepsilon_t - \mathbb{E}_{t_1}[\varepsilon_t]\right] + \mathbb{E}_{t-1}[\xi_t] = 0.$$

Therefore, prices in the model are a martingale, even when the sign of the next trade is highly predictable. This is a plausible first approximation to how prices behave in real markets, although one observes that real prices deviate from the martingale property at very high frequencies (see Figure 13.2 and Section 16.3).

Now, either the sign of the t^{th} transaction matches the sign of $\widehat{\varepsilon_t}$, or it does not. Let $\mathbb{E}_t^+[r_t]$ denote the expected ex-post value of the return of the t^{th} transaction, given that ε_t matches that of $\widehat{\varepsilon_t}$, and let $\mathbb{E}_t^-[r_t]$ denote the expected ex-post value of the return of the t^{th} transaction, given that ε_t does not match that of $\widehat{\varepsilon_t}$. By definition, the most likely outcome is that ε_t does match the sign of its predictor

$\widehat{\varepsilon}_t$, which happens with a probability given by $(1 + |\widehat{\varepsilon}_t|)/2$ (see Equation (13.21), below).

The absence-of-predictability criterion $\mathbb{E}_{t-1}[r_t] = 0$ can be rewritten as:

$$\frac{1+|\widehat{\varepsilon}_t|}{2}\mathbb{E}_t^+[r_t] + \frac{1-|\widehat{\varepsilon}_t|}{2}\mathbb{E}_t^-[r_t] = 0 \Rightarrow \left|\frac{\mathbb{E}_t^+[r_t]}{\mathbb{E}_t^-[r_t]}\right| = \frac{1-|\widehat{\varepsilon}_t|}{1+|\widehat{\varepsilon}_t|} \leq 1.$$

Therefore, the most likely outcome has the smallest impact. This generalises the result on $\mathcal{R}_\pm(1)$ given in Section 11.3.4, which was based on a qualitative argument. In the present picture, each trade has a permanent impact, but the impact depends on the past order flow and on its predictability.

13.3.2 The DAR Model

We now assume that the time series of market order signs is well modelled by a **discrete auto-regressive** process (see Section 10.4.2). In this framework, the best predictor $\widehat{\varepsilon}_t = \mathbb{E}_{t-1}[\varepsilon_t]$ of the next trade sign, just before it happens, can be written as

$$\widehat{\varepsilon}_t = \sum_{k=1}^{p} \mathbb{K}(k)\varepsilon_{t-k}. \tag{13.20}$$

The backward-looking kernel $\mathbb{K}(k)$ can be inferred from the sign autocorrelation function $C(\ell)$ using the Yule–Walker equation (10.6). The probability that the next sign is ε_t is then given by:

$$\mathbb{P}_{t-1}(\varepsilon_t) = \frac{1+\varepsilon_t\widehat{\varepsilon}_t}{2}. \tag{13.21}$$

Because ε_t is a binary random variable, Equation (13.21) is equivalent to $\widehat{\varepsilon}_t = \mathbb{E}_{t-1}[\varepsilon_t]$.

Inserting Equation (13.20) into Equation (13.19), we arrive at

$$r_t = G_{1,t}\varepsilon_t - G_{1,t}\sum_{k=1}^{p}\mathbb{K}(k)\varepsilon_{t-k} + \xi_t. \tag{13.22}$$

In the case where $G_{1,t}$ does not depend on time, and can therefore be written as a constant G_1, we see (by comparing this expression with that in Equation (13.8)) that this framework is equivalent to the linear propagator model, with

$$K(\ell) := G(\ell+1) - G(\ell) \equiv -G_1\mathbb{K}(\ell); \qquad (\ell \leq p); \qquad K(\ell > p) = 0. \tag{13.23}$$

Therefore, a slowly decaying $G(\ell) \sim \ell^{-\beta}$ can only be reproduced by a high-order DAR(k) model for which $\mathbb{K}(\ell)$ itself decays as a power-law (in fact, precisely as $\ell^{-\beta-1}$). The case where

$$\mathbb{K}(k) = \begin{cases} \rho, & \text{for } k = 1, \\ 0, & \text{for } k > 1 \end{cases}$$

leads to an exponential decay of the sign correlation, $C(\ell) = \rho^\ell$. This corresponds exactly to the MRR model that we will discuss in Section 16.2.1.

Throughout this section, we have greatly simplified all calculations by neglecting the influence of trade volume. We present an extended version of this model, including trade volumes, in Appendix A.3. Interestingly, even in cases where volumes fluctuate considerably, the final results and conclusions are qualitatively the same as those that we have presented above.

13.4 More on the Propagator Model

We now provide some advanced remarks about the propagator model, concerning in particular the impact of metaorders, the aggregate impact of volume imbalance, and its close kinship with Hasbrouck's VAR model.

13.4.1 Hasbrouck's VAR Model

In the wider literature on market microstructure, the so-called **vector autoregressive model** (VAR), as proposed by J. Hasbrouck in 1991, is a standard and widely used tool for predicting the present price return r_t and signed volume $v_t = \varepsilon_t \upsilon_t$ from their previous realisations. In this section, we illustrate how the propagator framework, as encapsulated in Equation (13.8), can be regarded as a special case of Hasbrouck's VAR model.

The VAR model is the joint linear regression:

$$\begin{aligned} r_t &= G_1 v_t + \sum_{-\infty<n<t} K(t-n)(t-n)v_n + \sum_{-\infty<n<t} \mathcal{H}_{rr}(t-n)r_n + \xi_t, \\ v_t &= \sum_{-\infty<n<t} \mathcal{H}_{vv}(t-n)v_n + \sum_{-\infty<n<t} \mathcal{H}_{rv}(t-n)r_n + \xi'_t. \end{aligned} \tag{13.24}$$

The propagator model can be seen as a special case of Equation (13.24), where:

- In the propagator model, the signed volume v_t in Equation (13.24) is replaced by the sign ε_t. This is justified by the fact that trade volumes carry much less information than a linear model would imply.
- In the propagator model, there is no direct impact of past returns on the present return (i.e. $\mathcal{H}_{rr} = 0$). This choice corresponds to a mechanistic interpretation of trade impact, where one assumes that past price changes by themselves do not influence present returns.
- In the propagator model, the dynamics of signed volumes (or signs) is summarised by the autocorrelation function $C_v(\ell)$ but is not described explicitly, although the DAR model is one possible specification, with the Yule–Walker relation:

$$C_v(\ell) = \sum_{k=1}^{\infty} \mathcal{H}_{vv}(k) C_v(|\ell - k|).$$

- In the propagator model, there is no direct feedback of the price change on the order flow (i.e. $\mathcal{H}_{rv} = 0$). The order flow is therefore assumed to be "rigid", independent of the price trajectory. This is clearly an approximation, which is partially addressed by the generalised propagator model (see Chapter 14).

In the VAR model, the **information content** I_H of a trade is defined as the long-term average price change, conditioned on a trade occurring at time $t = 0$:

$$I_H = \mathbb{E}[m_\infty - m_0 \mid v_0 = v].$$

This is very close to the asymptotic response function $\mathcal{R}_\infty$ in the propagator model, with the only difference being that volume fluctuations are neglected in the latter.

In other words, I_H is given by the same expression as $\mathcal{R}_\infty$ (see Equation (13.10)), with $C(\ell)$ replaced by $C_v(\ell)$ and with a prefactor proportional to the volume v. In Hasbrouck's initial specification, the information content of a trade is therefore proportional to its volume, whereas we know now that there is much less information in volume than initially anticipated (see Section 11.3.2).

13.4.2 The Propagator Model as a Reduced Description of All Order Flow

The propagator model that we have explored in this chapter considers price returns only at the times of market order arrivals. In a real LOB, by contrast, several other limit order arrivals and cancellations may occur between the arrivals of market orders. In Chapter 14, we extend the propagator model to include these other events explicitly. However, there are also good reasons for wishing to exclude these other events from the model specification, such as not having access to them in market data or simply wishing to keep the model formulation parsimonious. Therefore, we now explore the extent to which the basic propagator model that we have introduced in this chapter effectively includes the influence of these other LOB events.

Imagine that two types of events are important for price dynamics, but that we only observe one of them. Observable events (e.g. market order arrivals) are characterised by a random variable ε_t, which is simply the trade sign in the propagator model. Unobservable events (e.g. limit order arrivals and cancellations) are characterised by another random variable $\widehat{\xi}_t$. If all events were observed, the full propagator model for the mid-price, generalising Equation (13.7), would read

$$m_t = m_{t_0} + \sum_{t_0 \le n < t} G(t-n)\varepsilon_n + \sum_{t_0 \le n < t} H(t-n)\widehat{\xi}_n + \sum_{t_0 \le n < t} \xi_t. \qquad (13.25)$$

Assume for simplicity that the ε_t and $\widehat{\xi}_t$ are correlated Gaussian random variables. One can then always express the unobserved $\widehat{\xi}$ terms as a linear superposition of past ε terms, plus noise. Inserting this decomposition into the equation above, one finds an effective propagator model in terms of the ε only, plus an additional noise component ξ' coming from the unobserved events and from ξ:

$$m_t = m_{-\infty} + \sum_{-\infty < n < t} G(t-n)\varepsilon_n + \sum_{-\infty < n < t} H(t-n) \sum_{-\infty < m < n} \Xi(n-m)\varepsilon_m + \sum_{0 \le n < t} \xi'_t, \qquad (13.26)$$

where Ξ is the linear filter for expressing the $\widehat{\xi}$s in terms of the past εs. This kernel is the result of the classical *Wiener filter*, and can be expressed in terms of the correlation function between the two sets of variables.[3]

Equation (13.26) can in fact be recast in the form of the single-event propagator model, Equation (13.7), with an **effective propagator** $\widetilde{G}$ that contains both the direct effect modelled by G and the indirect influence of the unobserved events:

$$\widetilde{G}(\ell) = G(\ell) + \sum_{n=1}^{\ell} H(\ell')\Xi(\ell - \ell').$$

From this equation, it is clear that a non-trivial temporal dependence of $\widetilde{G}$ can arise, even when the true propagators G and H are lag-independent. In other words, the decay of a single market order's impact should in fact be interpreted a consequence of the interplay between market orders and limit orders. As a trivial example, suppose both propagators are equal and constant in time, such that $G(\ell) = H(\ell) = G_0$, but that $\varepsilon_n = -\widehat{\xi}_n$ for all n. This means that both types of events impact the price, but that

[3] See, e.g., Levinson, N. (1947). The Wiener RMS error criterion in filter design and prediction. *Journal of Mathematical Physics*, 25, 261–278, and Appendix A.3 for an explicit example.

they cancel each other out exactly. Then $\Xi(\ell) = -\delta_{\ell,0}$ leading to $\widetilde{G}(\ell) = 0$ for all ℓ. Therefore, the effective impact of events of the first type is zero in the effective model (and not G_0). In reality, limit orders only partially "screen" the impact of market orders, resulting in a non-trivial, time-decaying effective propagator.

13.4.3 Aggregate Impact

Within the propagator model, it is interesting to compute the **aggregate impact** of a given order-flow imbalance in a time interval containing T trades (see Section 11.4). For simplicity, we assume that the signed volume $v_n = \varepsilon_n \upsilon_n$ of each transaction n is a Gaussian variable with zero mean and unit variance, and long-range autocorrelations given by $C(\ell) \sim \ell^{-\gamma}$. In this framework, we aim to compute the average value of the mid-price change $m_{t+T} - m_t$, conditioned on a specified order-flow imbalance $\Delta V = \sum_{j=0}^{T-1} v_j$. In the following, we choose for convenience the origin of time such that $t = 0$.

Let us first compute the average value of v_n when such a condition is imposed:

$$\mathbb{E}[v_n|\Delta V] = \frac{1}{Z}\int \prod \mathrm{d}v_j\, v_n e^{-\frac{1}{2}\left(\sum_{j,k} v_j C^{-1}_{j,k} v_k\right)} \delta\left(\sum_{j=0}^{T-1} v_j - \Delta V\right),$$

where $C_{j,k} = C(|j-k|)$ is the correlation matrix, $\delta(\cdot)$ is the Dirac delta function, and Z is the same expression as in the numerator but without the v_n term.

Using the standard representation of the δ function in terms of its Fourier transform:

$$\mathbb{E}[v_n|\Delta V] = \frac{\partial}{\partial w_n} \ln\left(\int \mathrm{d}z e^{-iz\Delta V} \int \prod \mathrm{d}v_i e^{-\frac{1}{2}\sum_{ij} v_i C^{-1}_{ij} v_j + iz\sum_{j=0}^{T-1} v_j + w_n v_n}\right)\Bigg|_{w_n=0}.$$

Performing the Gaussian integral, and keeping only terms that survive when $w_n \to 0$:

$$\mathbb{E}[v_n|\Delta V] = \frac{\partial}{\partial w_n} \ln\left(\int \mathrm{d}z e^{-iz\Delta V} e^{-\frac{z^2}{2}\sum_{j,k=0}^{T-1} C_{jk} + izw_n \sum_{j=0}^{T-1} C_{nj}}\right)\Bigg|_{w_n=0}.$$

Performing the Gaussian integration over z:

$$\begin{aligned}\mathbb{E}[v_n|\Delta V] &= -\frac{\partial}{\partial w_n} \frac{\left(\Delta V - w_n \sum_{j=0}^{T-1} C_{nj}\right)^2}{2\sum_{j,k=0}^{T-1} C_{jk}}\Bigg|_{w_n=0}, \\ &= \frac{\sum_{j=0}^{T-1} C_{nj}}{\sum_{j,k=0}^{T-1} C_{jk}} \Delta V. \end{aligned} \tag{13.27}$$

In the case where there are no correlations (i.e. when $C_{i,j} = \delta_{i,j}$), one can check this expression directly:

$$\mathbb{E}[v_n|\Delta V] = \begin{cases} \frac{1}{T}\Delta V, & \text{for } 0 \le n \le T-1; \\ 0, & \text{otherwise.} \end{cases}$$

In the absence of correlations, the constraint that $\Delta V = \sum_{j=0}^{T-1} v_j$ is only effective when $n \in [1, T]$. Within this interval, all of the n's play a symmetric role.

The result is even more interesting in the limit of long-range autocorrelations $C(\ell) \sim_{\ell \gg 1} c_\infty \ell^{-\gamma}$, with $0 < \gamma < 1$. When $T \gg 1$, one finds

$$\mathbb{E}[v_n|\Delta V] \approx \Delta V \frac{2-\gamma}{2T}\left[\left(1-\frac{n}{T}\right)^{1-\gamma} + \left(\frac{n}{T}\right)^{1-\gamma}\right], \qquad 1 \le n \le T-1. \tag{13.28}$$

The propagator model predicts that the average change in mid-price between $t = 0$ and $t = T-1$ is given by

$$\begin{aligned}\mathbb{R}(\Delta V, T) &= \mathbb{E}[m_T - m_0|\Delta V], \\ &= \sum_{0\le n\le T-1} G(T-n)\mathbb{E}[\varepsilon_n|\Delta V] + \sum_{n>0} [G(T+n) - G(n)]\,\mathbb{E}[\varepsilon_n|\Delta V],\end{aligned}$$

where ε_n is the sign of v_n (i.e. the sign of a Gaussian variable with a given conditional mean and variance, as computed in Equation (13.27)).

If $\mathbb{E}[v_n|\Delta V] \ll 1$ and $T \gg 1$, then $\mathbb{E}[\varepsilon_n|\Delta V]$ simplifies to

$$\mathbb{E}[\varepsilon_n|\Delta V] \approx \sqrt{\frac{2}{\pi}}\mathbb{E}[v_n|\Delta V].$$

Using also that for $n \gg 1$, the propagator can be approximated as $G(n) \approx \Gamma_\infty/n^\beta$, one finally finds that in the large-T limit, the aggregated impact is given by

$$\mathbb{E}[m_T - m_0|\Delta V] \approx A(\beta,\gamma)\Gamma_\infty \frac{\Delta V}{T^\beta},$$

where $A(\beta,\gamma)$ is a numerical constant that depends on β and γ.

When ΔV reaches values of order T, one finds that $\mathbb{E}[\varepsilon_n|\Delta V] \to 1$ and $\mathbb{E}[m_T - m_0|\Delta V] \propto T^{1-\beta}$. The propagator model therefore yields the following result for the aggregate impact:

$$\mathbb{R}(\Delta V, T) = \Gamma_\infty T^{1-\beta} \times \mathscr{F}_{\rm prop}\left(\frac{\Delta V}{T}\right), \tag{13.29}$$

where $\mathscr{F}_{\rm prop}(u)$ is a scaling function that is linear for small arguments and that saturates for large arguments. Observe that Equation (13.29) has the same functional form as Equation (11.11), which describes empirical observations and that we recall here for convenience:

$$\mathbb{R}(\Delta V, T) \cong \langle s\rangle T^\chi \times \mathscr{F}\left(\frac{\Delta V}{\bar{V}_{\rm best}T^\varkappa}\right).$$

However, the propagator prediction makes incorrect predictions for the exponents χ and $\varkappa$. Empirical data suggests $\chi \approx 0.5 - 0.65$ and $\varkappa \approx 0.75 - 0.9$ (see Section 11.4), while the propagator model predicts $\chi = 1-\beta \approx 0.75$ and $\varkappa = 1$. Note, however, that the model correctly predicts the difference $\chi - \varkappa = 0.25$ which governs the scaling of Kyle's lambda with T (see Chapter 15).

13.4.4 Metaorder Impact

We now illustrate how the propagator model can also help to illuminate the peak impact and impact path of a metaorder. Throughout this section, we also make some additional assumptions about metaorder execution. Specifically, we assume that a metaorder of total volume Q is executed in N separate child orders, each of size υ, and that the (trade) time between each child order is equal to n. We describe such a metaorder as having a *participation rate* $f = 1/n$. In the present context, we assume that all trades have unit size, so the participation rate of a given metaorder is simply equal to the fraction of the total number of trades that it comprises. We also assume that the sign of the metaorder is uncorrelated with the sign of all other transactions.

We arbitrarily set the time of the first trade to be $t_0 = 0$. Due to our assumptions, trades occur at time $0, n, 2n, \ldots, T = (N-1)n$, such that the total execution time of the metaorder is $T = (N-1)n$. Conditional on the presence of a buy metaorder, the propagator model predicts that the average price change between t_0 and some $t \le T$ is given by

$$\mathfrak{I}^{\rm path}(t) = \mathbb{E}[m_t - m_0|\text{metaorder}] = \sum_{k=0}^{\lfloor t/n\rfloor} I_1(\upsilon)G(t-kn),$$

where $I_1(\upsilon)$ is the lag-1 impact of a child order of size υ. In what follows, we assume that lag-1 impact is constant, so we can absorb the $I_1(\upsilon)$ term into the definition of G, to lighten notation.

Approximating the discrete sum as an integral, and using the Euler–McLaurin formula, one finds:

$$\mathfrak{I}^{\text{path}}(t) \approx \int_0^{t/n} \mathrm{d}k\, G(t-kn) + \frac{1}{2}\left[G(1)+G(t)\right] + \ldots.$$

If we also assume that the propagator $G(t)$ decays as $\Gamma_\infty t^{-\beta}$ for large t, with $\beta < 1$, then to leading order

$$\mathfrak{I}^{\text{path}}(t) \approx \frac{\Gamma_\infty}{n(1-\beta)} t^{1-\beta}.$$

The impact of such a metaorder peaks at time $t = T$, with a value given by:

$$\mathfrak{I}^{\text{peak}}(Q,f) \approx \frac{\Gamma_\infty}{(1-\beta)} f^\beta \left(\frac{Q}{\upsilon}\right)^{1-\beta}, \qquad (N = Q/\upsilon \gg 1). \tag{13.30}$$

This **peak impact** is concave in the volume of the metaorder Q, and increases with the participation rate f (or, alternatively, decreases with the execution horizon $T \approx Q/(\upsilon f)$). In other words, expressing Equation (13.30) in terms of T instead of f:

$$\mathfrak{I}^{\text{peak}}(Q,T) \approx \frac{\Gamma_\infty}{(1-\beta)} T^{-\beta} \frac{Q}{\upsilon}, \qquad (N = Q/\upsilon \gg 1). \tag{13.31}$$

This corresponds to a linear impact, with a slope that decays as $T^{-\beta}$, in agreement with Equation (13.29). This result is not surprising: the propagator model is linear, so impacts add up.

Therefore, although for a fixed f it is qualitatively similar to the square-root impact law for $\beta < 1/2$, the metaorder impact law in the propagator framework is fundamentally *linear*. Furthermore, the model predicts that the impact of a metaorder scales like f^β or $T^{-\beta}$, which is not consistent with experimental data: as we emphasised in Section 12.3, the impact of a metaorder is in fact only weakly dependent on the time to completion T. Hence, while the propagator model captures some aspects of the square-root law, it needs to be amended in some way to reflect the behaviour of real markets, as observable in empirical data. We will examine two such amendments in Section 13.4.5 and in Chapter 19.

13.4.5 The LMF Surprise Model

In Chapter 10, we saw that a plausible explanation for the long-range autocorrelation of market order signs is the power-law distribution of metaorder volumes. In the context of the Lillo, Mike and Farmer (LMF) model, we showed that in order to account for an autocorrelation function $C(\ell)$ that behaves as $\ell^{-\gamma}$, the probability $\mathbb{P}(L)$ that a sequence of market order signs (chosen uniformly at random among all such sequences) has length exactly L should decay as $L^{-2-\gamma}$ for large L.

We will now show that in the extreme case where there is only a single active metaorder at each time step, it is straightforward to construct the best predictor for the next sign. This expression turns out to be very different from the linear combination of past signs that we previously obtained within a DAR description in Equation (13.20).

Consider a situation in which the previous ℓ market orders have all had sign $\varepsilon_t = +1$. If it is known that only one metaorder is active at this time, then the probability that the next market order sign is also $+1$ is given by

$$P_+(\ell) = \frac{\sum_{L=\ell+1}^{\infty} \mathbb{P}(L)}{\sum_{L=\ell}^{\infty} \mathbb{P}(L)} \approx_{\ell \gg 1} 1 - \frac{\gamma+1}{\ell}.$$

Within the LMF framework, the best predictor of the next sign is

$$\widehat{\varepsilon_t} = +\left(1 - \frac{\gamma+1}{\ell}\right) - \frac{\gamma+1}{\ell} = 1 - 2\frac{\gamma+1}{\ell}.$$

The history-dependent (or "surprise") model, which generalises the linear propagator model from Section 13.3, then reads

$$r_t = m_{t+1} - m_t = G_1 (\varepsilon_t - \widehat{\varepsilon}_t) + \xi_t, \tag{13.32}$$

with

$$r_t = \begin{cases} 2G_1 \left(\frac{\gamma+1}{\ell}\right) + \xi_t, & \text{if } \varepsilon = +1, \\ -2G_1 + 2G_1 \left(\frac{\gamma+1}{\ell}\right) + \xi_t, & \text{if } \varepsilon = -1. \end{cases}$$

The impact of a sequence of $\ell = Q$ consecutive trades in the same direction is then

$$\mathfrak{I}_{\text{LMF}}(Q) = \mathbb{E}[m_\ell - m_0] \approx_{Q \gg 1} 2G_1(1+\gamma) \ln Q.$$

Note that this behaviour is quite different from the prediction of the propagator model (see Equation (13.30)), which instead leads to $\mathfrak{I}_{\text{prop}}(Q) \sim Q^{1-\beta}$.

In the LMF model, the average price paid to execute the Q child orders is given by

$$p_{\text{ex.}} = m_0 + \frac{2G_1(1+\gamma)}{Q} \sum_{L=1}^{Q} \ln L.$$

For $Q \gg 1$, this expression can be approximated as

$$\begin{aligned} p_{\text{ex.}} &\approx m_0 + 2G_1(1+\gamma)(\ln Q - 1), \\ &\approx m_\ell - 2G_1(1+\gamma). \end{aligned}$$

At the end of a long metaorder, one does not need to wait long before a trade in the unexpected direction takes place. This trade has a large price impact of approximately $-2G_1$. Therefore, soon after the end of the metaorder, the price reverts to

$$m_\infty \approx m_\ell - 2G_1 > p_{\text{ex.}},$$

which exceeds the average paid price by an amount $2\gamma G_1$. This highlights a problem with this model: the *price is a martingale* thereafter, so the metaorder would be profitable on average. This represents an arbitrage opportunity that is absent from real markets.

How might we alter the surprise model in Equation (13.32) to address this weakness? One possible approach is to allow the impact parameter G_1 to be dependent on time and/or market conditions. Then, one way to remove the above arbitrage is to assume that the longer the metaorder has lasted, the larger the coefficient G_1:

$$G_1(\ell) := G_0 \ell^\delta, \qquad \delta > 0. \tag{13.33}$$

One now finds that the impact of a sequence of Q trades in the same direction is

$$\mathfrak{I}_\delta(Q) \approx 2G_0(1+\gamma)\frac{Q^\delta}{\delta},$$

and that the average price paid to execute the Q trades is

$$p_{\text{ex.}} \approx p_0 + 2G_0 \frac{(1+\gamma)}{\delta(1+\delta)} Q^\delta.$$

Long after the metaorder execution ends, the price reverts to

$$m_\infty \approx m_\ell - 2G_0 Q^\delta \qquad (Q \gg 1).$$

To make further progress with our study of this model, we now make the additional assumption that the **execution price is fair**, in the sense that the average execution price is equal to the long-term price. This leads to the condition

$$\begin{aligned} p_{\text{ex.}} = m_\infty &\Rightarrow 2G_0 \frac{(1+\gamma)}{\delta(1+\delta)} \ell^\delta = 2G_0(1+\gamma)\frac{\ell^\delta}{\delta} - 2G_0 \ell^\delta, \\ &\Rightarrow \frac{(1+\gamma)}{(1+\delta)}\left(1 - \frac{1}{\delta}\right) = 1, \\ &\Rightarrow \delta = \gamma. \end{aligned}$$

This idea, promoted by Farmer, Gerig, Lillo and Waelbrouck (FGLW), leads to a concave impact for metaorders, $\mathfrak{I}_\delta(Q) \sim Q^\delta$, with an exponent δ that is equal to the decay exponent of the autocorrelation of the trade-sign series, $\delta = \gamma < 1$. In this framework, the impact decay is nearly immediate, because the market is assumed to detect when the metaorder has ended. Interestingly, the FGLW calculation can be framed using slightly different language, in the context of the Glosten–Milgrom model (see Section 16.1.6).

The FGLW model also suggests a non-linear generalisation of the linear propagator model, where

$$G_1(\widehat{\varepsilon}_t) = G_0\left(1 + A\widehat{\varepsilon}_t^2 + ...\right),$$

where A is a constant and by symmetry the term linear in $\widehat{\varepsilon}_t$ is absent. In this set-up, the value of G_1 increases as the level of predictability increases, similarly to Equation (13.33). Setting $A = 0$ recovers the linear propagator model with $G_1 = G_0$. To date, the properties of this extended propagator model still remain to be explored. We present another (arguably better-motivated) non-linear generalisation of the propagator model in Chapter 19.

13.5 Conclusion

In this chapter, we have presented a simple linear model that expresses price moves in terms of the influence of all past trades. The model assumes that each trade should be treated on the same footing (independently of its size or how aggressive it is) and leads to a reaction impact in the direction of the trade, with dynamics encoded in a lag-dependent function $G(\cdot)$, called the propagator. In the model, the immediate reaction impact of a trade is given by $G(1)$ and the long-term reaction impact of a trade is given by $G(\infty)$.

To avoid creating statistical arbitrage opportunities (such as trends), the fact that real trade signs are long-ranged autocorrelated imposes that $G(\ell)$ has some special shape. In Section 13.2, we showed that if the autocorrelation of trade signs decays as $\ell^{-\gamma}$ with $\gamma < 1$, then $G(\ell)$ must decay to zero as $\ell^{-\beta}$, with $\beta = (1-\gamma)/2$. This very slow decay, which we call long-range resilience, means that the distinction between "transient" and "permanent" impact is necessarily fuzzy, at least empirically. Furthermore, even when the permanent impact of a single trade G_∞ is zero, the cumulative observed impact $\mathcal{R}_\infty$ of that trade and all of its correlated kinship can be non-zero, as a result of the subtle compensation between trade correlations and impact decay (see Equation (13.18)). Intuitively, this compensation reflects the way that liquidity providers and liquidity takers offset each other, to remove most of the linearly predictable patterns that their (predictable) actions would otherwise create.

Despite the fact that it only deals with market orders, we have argued that the propagator model is actually an effective reduced-form description of the full interplay between market orders, limit orders and cancellations. We have also noted that it is a simplified, and arguably more intuitive, specification of Hasbrouck's classic VAR model.

Although very illuminating, the propagator model suffers from several drawbacks. For example, the linearity of the model is at odds with empirical findings such as the square-root impact of metaorders. Resolving these problems requires a more sophisticated, non-linear version of the propagator model, which we introduce in Chapter 19.

Take-Home Messages

(i) The simple, linear propagator model assumes that all market orders have an identical, time-dependent impact on the price, consisting of both an immediate reaction and a lagged reaction.
(ii) Because market order signs are positively autocorrelated, if impact was constant and permanent, then the mid-price would be strongly super-diffusive, and therefore predictable. To ensure that the mid-price series is diffusive, we must assume that the impact of a market order decays over time.
(iii) To balance the power-law (long-range) autocorrelations of trade signs, the propagator must decay slowly as well. We call this behaviour, in which the effect of an action takes a long time to vanish, the *long-range resilience of markets*.
(iv) The propagator kernel G can be fitted empirically from the two-point correlation function of market order signs C and the response function $\mathcal{R}$, through a linear equation.
(v) In history-dependent impact models, impact is permanent but depends on the history in such a way that prices are unpredictable. In a DAR hypothesis for the order flow, one recovers exactly the propagator model.
(vi) Despite its many appealing properties, the linear propagator model is too simple to reproduce some empirical regularities observable in real markets. Addressing these deficiencies of the model involves considering a more complex, non-linear framework, which we introduce in Chapter 19.

13.6 Further Reading

See also Section 10.7.

The Propagator Model and History-Dependent Impact Models

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

Bouchaud, J. P., Kockelkoren, J., & Potters, M. (2006). Random walks, liquidity molasses and critical response in financial markets. *Quantitative Finance*, 6(02), 115–123.

Farmer, J. D., Gerig, A., Lillo, F., & Mike, S. (2006). Market efficiency and the long-memory of supply and demand: Is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6(02), 107–112.

Bouchaud, J. P., Farmer, J. D., & Lillo, F. (2009). How markets slowly digest changes in supply and demand. In Hens, T. & Schenke-Hoppe, K. R. (Eds.), *Handbook of financial markets: Dynamics and evolution*. North-Holland, Elsevier.

Eisler, Z., J. Kockelkoren, J., & Bouchaud, J. P. (2012). Models for the impact of all order book events. In Abergel, F., Bouchaud, J.-P., Foucault, T., Lehalle, C.-A., & Rosenbaum, M. (Eds.), *Market microstructure: Confronting many viewpoints*. Wiley

Taranto, D. E., Bormetti, G., Bouchaud, J. P., Lillo, F., & Tóth, B. (2016). Linear models for the impact of order flow on prices I. Propagators: Transient vs. history dependent impact. Available at SSRN: https://ssrn.com/abstract=2770352.

Tóth, B., Eisler, Z., & Bouchaud, J.-P. (2017). The short-term price impact of trades is universal. https://ssrn.com/abstract=2924029.

Hasbrouck's Model

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

Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Transactions, volume, and volatility. *Review of Financial Studies*, 7(4), 631–651.

Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*. Oxford University Press.

Metaorders and Propagators

Farmer, J. D., Gerig, A., Lillo, F., & Waelbroeck, H. (2013). How efficiency shapes market impact. *Quantitative Finance*, 13(11), 1743–1758.

Donier, J., Bonart, J., Mastromatteo, I., & Bouchaud, J. P. (2015). A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*, 15(7), 1109–1121.