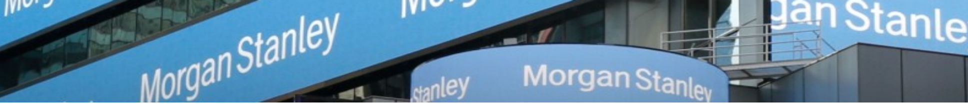

Do price trajectory data increase the efficiency of market impact estimation?

Fengpei Li

Morgan Stanley Machine Learning Research

The opinions expressed are those of the authors. They do not purport to reflect the opinions or views of Morgan Stanley

Sept. 2024

Main findings

- Standard estimation method on market Impact models could be "inefficient."
- For Almgren-Chriss model, there is an "optimal" calibration
- Insight: early price trajectory helps with calibrating impact functions for a general class of market impact models (recently [Eisler and Muhle-Karbe, 2024])

What is Market Impact

- Three key components of investment performance [Ferraris and AG, 2011]
  - Alpha
  - Risk
  - Cost
- Alpha and Risk subsume the vast majority of quant research
- Cost is "widely discussed but rarely measured" [Almgren et al., 2005]
- Yet it is "a large determinant of investment performance" :
  - realization of active investment strat
  - realization of liquidity

Transaction Cost

Transaction/Trade Cost falls in two categories:

- **Direct Cost**: can be explicitly stated and measured, e.g., Commissions, fees, taxes
- **Indirect Cost**: can not be explicitly measured. For example:
  - Impact of trader's own action $\rightarrow$ **Market Impact**
  - (less so) opportunity/timing cost, spread/delay risk etc

Market Impact: Why is it important

Widely recognized as a substantial factor in reducing investment strategy (notional performance) [Ferraris and AG, 2011]

- The average cost of a US large cap trade from 2003-2008 is

$$23bps = 9bps(commssions) + 14bps(market\ impact/slippage)$$

Market Impact: Why does it happen

- Liquidity consumption from Limit-Order Book (LOB); Fluctuations from supply-demand equilibrium
- Short term correlation between price changes and trades [Bouchaud, 2010]
- reveal of new/private information ([Kyle, 1985]) (optimal execution is incremental→ split of metaorders into child orders, modern EMM)

Market Impact: Observed robust/universal properties

"Quite remarkable that the square-root impact law appears to hold approximately in all cases" [Tóth et al., 2011]

- Square-root law: Price change $\propto \sigma\sqrt{\frac{X}{V_D}}$
- Power laws :Price change $\propto \sigma(\frac{X}{V_D})^{\delta}$ ("power-law best fit all points" : across mkt.cap size, asset class, a uniform price-impact curve [Lillo et al., 2003], typically $0.4 \leq \delta \leq 0.7$)
- Concave nature: "non-linear concave function of its size,... is robust, being observed for several markets and execution style" [Zarinelli et al., 2015],(they used log)

Market Impact Modeling

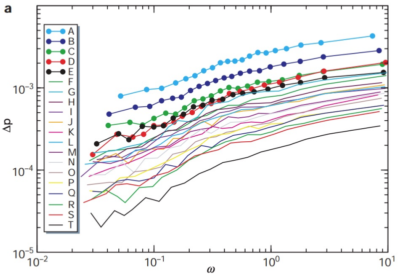

Figure 1: Price shift, Δp, plotted against normalized transaction size, ω for buyer initiated trades for 20 groups of stocks [Lillo et al., 2003].

- Research effort on designing market impact models
  - Consistent with observed properties
  - No arbitrage (trading cost non-negative)
  - Optimal execution strat derivation
- Modeling based on LOB ([Alfonsi et al., 2010])
- Modeling based on MFG (mean field game [Cardaliaguet and Lehalle, 2018]
- Modeling based on Price (most prevalent)

Price dynamic during execution

Most generally, consider a *volume weight average price* (VWAP) execution strategy (constant trading rate), the price follows

$$S_t = S_0 + \underbrace{\mu_\theta(t, v)}_{\text{price impact}} + \underbrace{\int_0^t \sigma dW_s}_{\text{randomness / volatility}} .$$

Here v is trading rate and t is trading time. θ is the model parameter.

- Almgren-Chriss Model [Almgren and Chriss, 2001]
- Propagator Models [Gatheral, 2010, Bouchaud et al., 2006]

Very few studies on estimation.

- Metaorder is private/proprietary
- Public data has inherent drawback (unknown info, "partial view"[Zarinelli et al., 2015])

Why is estimation important, [Ferraris and AG, 2011]

Accurate modeling of Market Impact $\rightarrow$ better execution strategy

- Under-estimate cost $\rightarrow$ trade too fast
- Over-estimate cost $\rightarrow$ trade too slow

[Schied and Schöneborn, 2009] : best liquidation $x_t = Xe^{-t\sqrt{\frac{\sigma^2 A}{2\eta}}}$

- η if off by 10% $\rightarrow$ off-target by 6.88% ($T = 0.2$)
- η if off by 5% $\rightarrow$ off-target by 3.36% ($T = 0.2$)

pre-trade cost estimation $\rightarrow$ investment decision/capacity

- Market model determine capacity of funds $\rightarrow$ protect alpha
- Larger trader $\rightarrow$ higher impact + low alpha

1. Introduction
2. Efficiency of Statistical Estimation
3. Main Results
4. Conclusion

How to judge the efficiency of estimation

Quick review, Example $X_1, ..., X_n \sim N(\theta, \sigma^2)$, estimate θ

- $\bar{X}$, the sample average
- Many other estimator (i.e., shrinkage)
- $\sqrt{n}(\bar{X} - \theta) \to N(0, \sigma^2)$ (CLT, consistency, asymptotic normality)
- asymptotic rate $n^{-\frac{1}{2}}$ with constant $(\frac{1}{\sigma^2})^{-1}$ (Cramer-Rao, can not do better)
- $\bar{X}$ is sufficient
- $\bar{X}$ is maximum likelihood estimator (MLE)

How to judge the efficiency of estimation

Regular statistical experiment

- Fisher information matrix $\mathbb{I}(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial l(\text{data}|\theta)}{\partial\theta}\right)\left(\frac{\partial l(\text{data}|\theta)}{\partial\theta}\right)^T\right]$.
- MLE performs best asymptotically, achieving the lowest possible factor $\mathbb{I}^{-1}(\theta_{\text{true}})$ with $n^{-\frac{1}{2}}$ rate
- $\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta_{\text{true}}) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{I}^{-1}(\theta_{\text{true}})\right)$

Design of experiment

MLE depends on what you observed $\rightarrow$ experimental design

- "bigger" fisher information matrix $\mathbb{I}_{exp1}(\theta) \succcurlyeq \mathbb{I}_{exp2}(\theta)$
- "smaller" asymptotic for any $c(\theta)$
  - $\sqrt{n}\big(c(\hat{\theta}_{\mathrm{MLE}}) - c(\theta^\star)\big) \rightarrow N\big(0, \nabla_\theta c^T(\theta^\star) \cdot \mathbb{I}^{-1}(\theta^\star) \cdot \nabla_\theta c(\theta^\star)\big)$
  - $\nabla_\theta c^T(\theta) \cdot \mathbb{I}^{-1}_{exp1}(\theta) \cdot \nabla c(\theta) \leq \nabla_\theta c^T(\theta) \cdot \mathbb{I}^{-1}_{exp2}(\theta) \cdot \nabla c(\theta)$

If you can obeserve all the data, then MLE based on any sufficient statistic of data $\phi(data)$ has the same Fisher formation

- $\mathbb{I}_{data}(\theta) = \mathbb{I}_{\phi(data)}(\theta)$ for sufficient $\phi(data)$
- $\mathbb{I}_{data}(\theta) \succcurlyeq \mathbb{I}_{\psi(data)}(\theta)$ for general function $\psi(data)$ (you lose information)

Almgren Chriss Model

The Almgren-Chriss model remains one of the most popular and influential model since introduction [Almgren and Chriss, 2001]:

$$
\begin{aligned}
S_t =& S_0 + S_0(g(v)t + h(v)) + S_0\sigma \int_0^t dW_s, \text{ when } t \leq T \\
S_t =& S_0 + S_0 g(v) T + S_0\sigma \int_0^t dW_s, \text{ when } t > T.
\end{aligned} \tag{1}
$$

- T is end-trading time (all scaled by vol time).
- g is the "permanent" impact $g(v;\theta) = \gamma v^\alpha$ (originally taken to be linear $\alpha = 1$)
- h is the "temporary" impact $h(v;\theta) = \eta v^\beta$

Power law is "extremely broad". (0.6 in favor of 0.5 for β). [Almgren et al., 2005]

Established method of estimation on Almgren Chriss

One of the few estimation paper, the method proposed in [Almgren et al., 2005] is based on statistic I, J

- $I = \frac{S_{T_{\text{post}}} - S_0}{S_0}$, "permanent impact" (price reverted a while after the trade)
- $J = \frac{\int_0^T S_t dt - S_0}{S_0}$, "realized impact" (average price for execution)

Estimation procedure

- Non-linear least square fitting, jointly on (I, J) across private data, by Gaussian-Newton
- Equivalent to MLE based on (I, J)

Sufficient Statistic for Almgren Chriss

Main theorems for Almgren-Chriss model

Theorem

- *The sufficient statistics (with most "information") is* $S_{\Delta t}, S_T, S_{T_{post}}$
- *Three points* $S_{t_{min}}, S_T, S_{T_{post}}$ *is sufficient for* $\{S_t\}_{t\in\mathbb{T}}$
- *Two points are not enough (inconclusive)*
- *For* $S_{\Delta t}, S_T, S_{T_{post}}$*, as long as* $\frac{t}{T} \leq \frac{1}{4}$*, it is strictly more efficient than* I, J.

Illustration [Ferraris and AG, 2011]

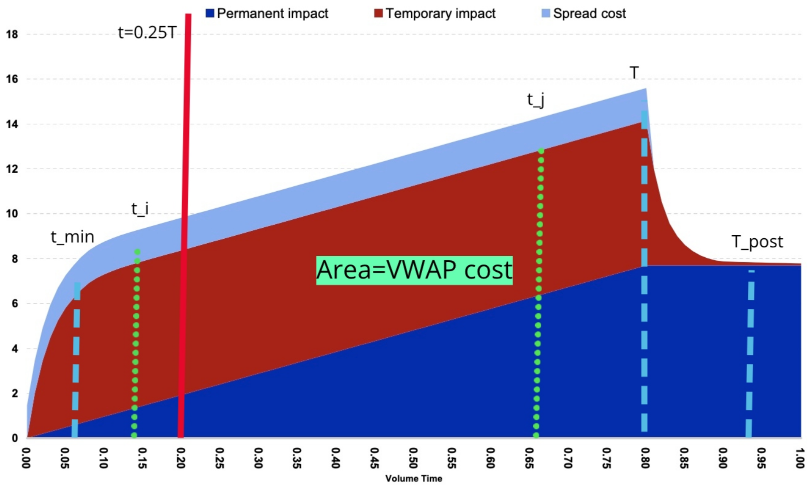

Figure 2: Simulation verification of theorem

Simulation

Method	Avg estimate $\hat{\theta}$ (average over 1000 simulations) α	β	γ	η	Theoretical $SE(\hat{\theta})$ (avg. of hessian implied SD) α	β	γ	η	Empirical $SE(\hat{\theta})$ (SD of estimate over 1000 sim) α	β	γ	η
$M = 50$												
Almgren	1.1227	0.6678	14.3141	0.7729	1.2663	0.5012	45.6219	1.2037	1.5315	0.6933	48.5203	3.8350
Two Point	1.1336	0.7353	12.0443	4.6894	1.7610	1.1619	61.8151	19.0205	1.4393	1.0526	42.5171	28.3513
3p Optimal	**1.0972**	**0.6384**	**10.5504**	**0.3011**	**1.6342**	**0.4542**	**43.9538**	**0.3884**	**1.4149**	**0.4618**	**30.1685**	**1.4073**
3p Late	1.1294	0.7185	11.8904	5.3670	1.7459	1.1094	60.8729	23.3794	1.4432	1.0632	43.7161	33.7297
Four Point	**1.0972**	**0.6384**	**10.5504**	**0.3011**	**1.6342**	**0.4542**	**43.9537**	**0.3884**	**1.4149**	**0.4618**	**30.1685**	**1.4073**

Figure 3: Simulation verification of theorem

Example: Suppose $(\gamma^\star, \eta^\star, \alpha^\star, \beta^\star) = (0.314, 0.142, 0.891, 0.600)$ and $(X, v, T, T_{\text{post}}, \sigma) = (0.1, 0.5, 0.2, 0.275, 1.57)$.

- Three point for $t = 0.1T$ more sample efficient over Almgren
- 21% for α, 51% for β
- 20.6% for γ, 51.5% for η
- 18.5% for cost estimation

The Propagator Model

Discrete [Bouchaud et al., 2003], continuous [Gatheral et al., 2012]

$$S_t = S_0 + \int_0^t f(v)G(t-s)ds + \sigma \int_0^t dWs.$$

- Impact is neither permanent nor temporary, but transient
- f: instantaneous impact, G: decay kernel
- G decreasing from 0 to ∞ with different tail

Properties of Propagator Models

Some peoperties of propagator models

- consistent with empirical properties (concavity,decay)
- Notable choice of f and G
  - power-law $f(v) \propto v^{\delta}$, power-law decay , $G(s) \propto s^{-\gamma}$ (square-root law: $\delta = \gamma = 0.5$) [Gatheral et al., 2012]
  - linear $f(v) \propto v$ and exponential decay $G(s) \propto e^{-\rho s}$. First transient model [Obizhaeva and Wang, 2013], links to LOB
  - logarithmic $f(v) \propto \log(v/v_0)$ and $G(s) \propto l_0(l_0 + s)^{-\gamma}$ or $G(s) \propto (l_0^2 + s^2)^{-\gamma/2}$. Here $\gamma \approx \frac{1-\alpha}{2}$, related to the exponent of auto-correlation among trade [Bouchaud et al., 2003].
- Many others (Gaussian kernels, etc), solving Fredholm equations for optimal execution (open),...

Calibration of Propagator Models

Theorem

The unique sufficient statistic is the full price path $\{S_t\}_{0\leq t\leq T}$

For just calibrate f, it is suggested in [Curato et al., 2017] that one should vwap $J = v\int_0^T S_t dt - XS_0$.

- How many points on the path is good enough? Two?

Theorem

For calibrating f*, we have* $\mathbb{I}_{S_t,S_T}(\theta) - \mathbb{I}_J(\theta) \geq 0$ *if*

$$\Big(\frac{(\int_0^t G(t)dt)^2}{t} + \frac{(\int_t^T G(t)dt)^2}{T-t}\Big) \geq \frac{3}{T^3}\Big(\int_0^T G(t)(T-t)dt\Big)^2.$$

Empirical Verification

Similar type result for "early" observation, in calibrating impact f

- Example1: For decay kernel $G(s) = s^{-\gamma}$ with $\gamma = 0.4$ [Bouchaud et al., 2003], we have $\mathbb{I}_{S_t,S_T}(\theta) \geq \mathbb{I}_J(\theta)$ when $2.11 \cdot 10^{-4} \leq \frac{t}{T} \leq 0.279$.
-

	$\gamma = 0.35$	$\gamma = 0.45$	$\gamma = 0.5$	$\gamma = 0.55$	$\gamma = 0.65$	$\gamma = 0.75$
$\tau = t/T$	$8.97 \cdot 10^{-4} \leq \tau \leq 0.369$	$9.41 \cdot 10^{-7} \leq \tau \leq 0.252$	$\tau \leq \frac{1}{4}$	$\tau \leq 0.257$	$\tau \leq 0.279$	$\tau \leq 0.301$

- Example 2: For $G(s) = e^{-\rho s}$, the comparison depends on specific values of t and T, not just their ratio τ. However, $t, T \to \infty$ but $\frac{t}{T} \to \tau$, then $\mathbb{I}_{S_t,S_T}(\theta) \geq \mathbb{I}_J(\theta)$ as long as

$$\tau \leq \frac{1}{3}$$

Sampling Strategy, more trajectory data

Empirical Studies: power-law kernel $G(s) = s^{-\gamma}$ with $\gamma = 0.4$ [Bouchaud et al., 2003, Busseti and Lillo, 2012] and power-law impact $f(v) = v^{\delta}$ with $\delta = 0.6$ [Almgren et al., 2005]

- $\frac{[\mathbb{I}_J]_{\delta,\delta}}{[\mathbb{I}_{\text{full data}}]_{\delta,\delta}} = 0.651$
- pick $t_1 = 0.125T$, $t_2 = 0.25T$, $t_3 = 0.625T$
-

	S_{t_1}, S_T	S_{t_2}, S_T	S_{t_3}, S_T	S_{t_1}, S_{t_2}, S_T	S_{t_1}, S_{t_3}, S_T	S_{t_2}, S_{t_3}, S_T	$S_{t_1}, S_{t_2}, S_{t_3}, S_T$
$[\mathcal{I}]_{\delta,\delta}/[\mathcal{I}_{\boldsymbol{S}_{\text{full}}}]_{\delta,\delta}$	**0.689**	0.657	0.595	**0.700**	**0.698**	**0.661**	**0.704**

Figure 4: Comparison of F.I. in terms of ratio for calibrating power-law impact

Miscellaneous result

- Seemingly some diminishing return effect
- early point not necessarily useful for calibrating kernel G:
  - pick $t_1 = 0.125T,\ t_2 = 0.25T,\ t_3 = 0.625T$
  -

	S_{t_1}, S_T	S_{t_2}, S_T	S_{t_3}, S_T	S_{t_1}, S_{t_2}, S_T	S_{t_1}, S_{t_3}, S_T	S_{t_2}, S_{t_3}, S_T	$S_{t_1}, S_{t_2}, S_{t_3}, S_T$
$\Vert\mathcal{I}^{1/2}_{\boldsymbol{S}_{\text{full}}} \cdot \mathcal{I}^{-1} \cdot \mathcal{I}^{1/2}_{\boldsymbol{S}_{\text{full}}}\Vert_2$	3.025	1.769	1.426	1.732	1.219	1.137	1.122
$[\mathcal{I}]_{c,c}/[\mathcal{I}_{\boldsymbol{S}_{\text{full}}}]_{c,c}$	**1-1.258e-5**	**1-8.184e-6**	**1-5.590e-6**	**1-7.962e-6**	**1-3.376e-6**	**1-2.274e-6**	**1-2.051e-6**

Figure 5: Comparison of F.I. in terms of ratio for calibrating of kernel $G(s) \propto e^{-\rho s}$ in [Obizhaeva and Wang, 2013]

- For calibrating square-root law:
  - $\mu(T, v) \propto (vT)^{\frac{1}{2}} = X^{\delta}$
  - $\frac{J}{X} \triangleq \frac{\mathbb{E}[v \int_0^T S_t dt - XS_0]}{X} \propto X^{\delta}$
  - $\mathbb{I}_{S_T}(\delta) \geq (\leq) \mathbb{I}_J(\delta)$ if $\delta \geq (\leq) \sqrt{3} - 1 \approx 0.732$

Limitations

Model Misspecification: No true model

- MLE minimizes the KL-divergence
  $\theta^{\star}_{\mathrm{KL}} = \arg\min_{\theta\in\Theta} D_{\mathrm{KL}}(F \| F(\theta))$
- $\theta^{\star} = \arg\min_{\theta\in\Theta} \int_0^T \mathbb{E}\left[\left(\frac{\partial \mu_\theta(t,v)}{\partial t} - \mu^{\star}(S_t;t,v)\right)^2\right] dt$
- Information matrix equivalence theorem no longer hold
  - $A(\theta) \triangleq \mathbb{E}\left[\left(\frac{\partial l(data|\theta)}{\partial \theta}\right)\left(\frac{\partial l(data|\theta)}{\partial \theta}\right)^T\right]$
  - $B(\theta) \triangleq -\mathbb{E}\left[\frac{\partial^2 l(data|\theta)}{\partial \theta^2}\right]$
  - $A(\theta) \neq B(\theta)$ during misspecification
  - asymptotic var (scaled by $n^{-0.5}$) $B^{-1}(\theta^{\star}_{\mathrm{KL}})A(\theta^{\star}_{\mathrm{KL}})B^{-1}(\theta^{\star}_{\mathrm{KL}})$

Thanks

Thanks! Paper link (Quantitative Finance volume 24, 2024):
https://www.tandfonline.com/doi/full/10.1080/14697688.2024.2351457

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

[Almgren and Chriss, 2001] Almgren, R. and Chriss, N. (2001).
Optimal execution of portfolio transactions.
Journal of Risk, 3:5–40.

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

[Bouchaud, 2010] Bouchaud, J.-P. (2010).
Price impact.
Encyclopedia of Quantitative Finance.

[Bouchaud et al., 2003] Bouchaud, J.-P., Gefen, Y., Potters, M., and Wyart, M. (2003).
Fluctuations and response in financial markets: the subtle nature ofrandom'price changes.
Quantitative finance, 4(2):176.

[Bouchaud et al., 2006] Bouchaud, J.-P., Kockelkoren, J., and Potters, M. (2006).
Random walks, liquidity molasses and critical response in financial markets.
Quantitative finance, 6(02):115–123.

[Busseti and Lillo, 2012] Busseti, E. and Lillo, F. (2012).
Calibration of optimal execution of financial transactions in the presence of transient market impact.
Journal of Statistical Mechanics: Theory and Experiment, 2012(09):P09010.

[Cardaliaguet and Lehalle, 2018] Cardaliaguet, P. and Lehalle, C.-A. (2018).
Mean field game of controls and an application to trade crowding.
Mathematics and Financial Economics, 12(3):335–363.

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

[Eisler and Muhle-Karbe, 2024] Eisler, Z. and Muhle-Karbe, J. (2024).
Optimizing broker performance evaluation through intraday modeling of execution cost.
arXiv preprint arXiv:2405.18936.

[Ferraris and AG, 2011] Ferraris, A. and AG, D. B. (2011).

Equity market impact models. mathematics at the interface between business and research, stifterverband für die deutsche wissenschaft. 4 december 2008, berlin.

[Gatheral, 2010] Gatheral, J. (2010).
No-dynamic-arbitrage and market impact.
Quantitative finance, 10(7):749–759.

[Gatheral et al., 2012] Gatheral, J., Schied, A., and Slynko, A. (2012).
Transient linear price impact and fredholm integral equations.
Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics, 22(3):445–474.

[Kyle, 1985] Kyle, A. S. (1985).
Continuous auctions and insider trading.
Econometrica: Journal of the Econometric Society, pages 1315–1335.

[Lillo et al., 2003] Lillo, F., Farmer, J. D., and Mantegna, R. N. (2003).
Master curve for price-impact function.
Nature, 421(6919):129–130.

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

[Schied and Schöneborn, 2009] Schied, A. and Schöneborn, T. (2009).
Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets.
Finance and Stochastics, 13(2):181–204.

[Tóth et al., 2011] Tóth, B., Lemperiere, Y., Deremble, C., De Lataillade, J., Kockelkoren, J., and Bouchaud, J.-P. (2011).

Anomalous price impact and the critical nature of liquidity in financial markets.
Physical Review X, 1(2):021006.

[Zarinelli et al., 2015] Zarinelli, E., Treccani, M., Farmer, J. D., and Lillo, F. (2015).
Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate.
Market Microstructure and Liquidity, 1(02):1550004.