Direct Estimation of Equity Market Impact

Reporters: Wang Yuke, Miao Jia

Date: 2021.12.23

Tsinghua University

CONTENTS

01
PART

Introduction

Transaction costs

✓ Transaction costs are widely recognized as a large determinant of investment performance.

Direct costs
are commissions and fees that are explicitly stated and easily measured. These are important and should be minimized, but are not the focus of this paper.

Indirect costs
are costs that are not explicitly stated. For large trades, the most important component of these is the impact of the trader's own actions on the market.

Market impact costs

a **permanent component** associated with information

a **temporary component** arising from the liquidity demands made by execution in a short time

Goal

is to use a large proprietary data set to **measure the indirect costs** experienced by **large institutional traders**, and to **characterize the dependence of these costs** on a few explanatory variables, so that these costs may be **estimated and controlled**.

02

PART

Data

2.1 Details

Each order
is broken into one or more transactions, each of which may generate one or more executions.
For each order, they have the following information:

- **The stock symbol**, requested order size (number of shares) and sign(buy or sell) of the entire order. Client identification is removed.
- **The times and methods** by which transactions were submitted by the Citigroup trader to the market. They take the time t_0 of the first transaction to be the start of the order. Some of these transactions are sent as market orders, some are sent as limit orders, and some are submitted to Citigroup's automated VWAP server.
- **The times, sizes, and prices** of execution corresponding to each transaction. They denote execution times by $t_1, \ldots, t_n$, sizes by $x_1, \ldots, x_n$, and prices by $S_1, \ldots, S_n$. Some transactions are cancelled or only partially executed; They use only the completed price and size.

In addition, they have various additional pieces of information, such as the instructions given by the client to the trader for the order, such as “over the day”, “market on close”, “market on open”, “VWAP”, or blank.

2.1 Filters

The total sample contains 682,562 orders, but for the data analysis they used only a subset.

1. only **orders on stocks in the Standard and Poor's 500 index.**
2. exclude about 400 orders for which the stock exhibits **more than 12.5% daily volatility** (200% annual).
3. exclude orders for which **the client requested “market on close” or “market on open”**.
4. exclude orders for which **the client requested VWAP execution**.
5. exclude orders for which any executions are **recorded after 4:10 PM.**
6. have **at least two completed transactions**.
7. are **at least 1000 shares**.
8. are **at least 0.25% of average daily volume** in that stock.

The results of the model are reasonably stable under changes in these criteria.

2.1 Description

After this filtering, they have **29,509 orders** in their data set.
The largest number of executions for any order is n=548; the median is around 5.
The median time is around one-half hour.

	Mean	Min	Q1	Median	Q3	Max
Total cost (%)	0.04	-3.74	-0.11	0.03	0.19	3.55
Permanent cost (I, %)	0.01	-3.95	-0.17	0.01	0.19	2.66
Temporary cost (J, %)	0.03	-3.57	-0.11	0.02	0.17	2.33
Shares/ADV ($\|X\|$, %)	1.51	0.25	0.38	0.62	1.36	88.62
Time (days)	0.39	0.00	0.10	0.32	0.65	1.01
Daily Vol. (%)	2.68	0.70	1.70	2.20	3.00	12.50
Mean spread (%)	0.14	0.03	0.08	0.11	0.16	2.37

Table: Summary statistics of orders in the sample: mean and quartile levels for each of several descriptive variables. The three cost variables are signed, and are very nearly symmetrically distributed about zero (IandJare defined in Section 3).

2.1 Description and filters

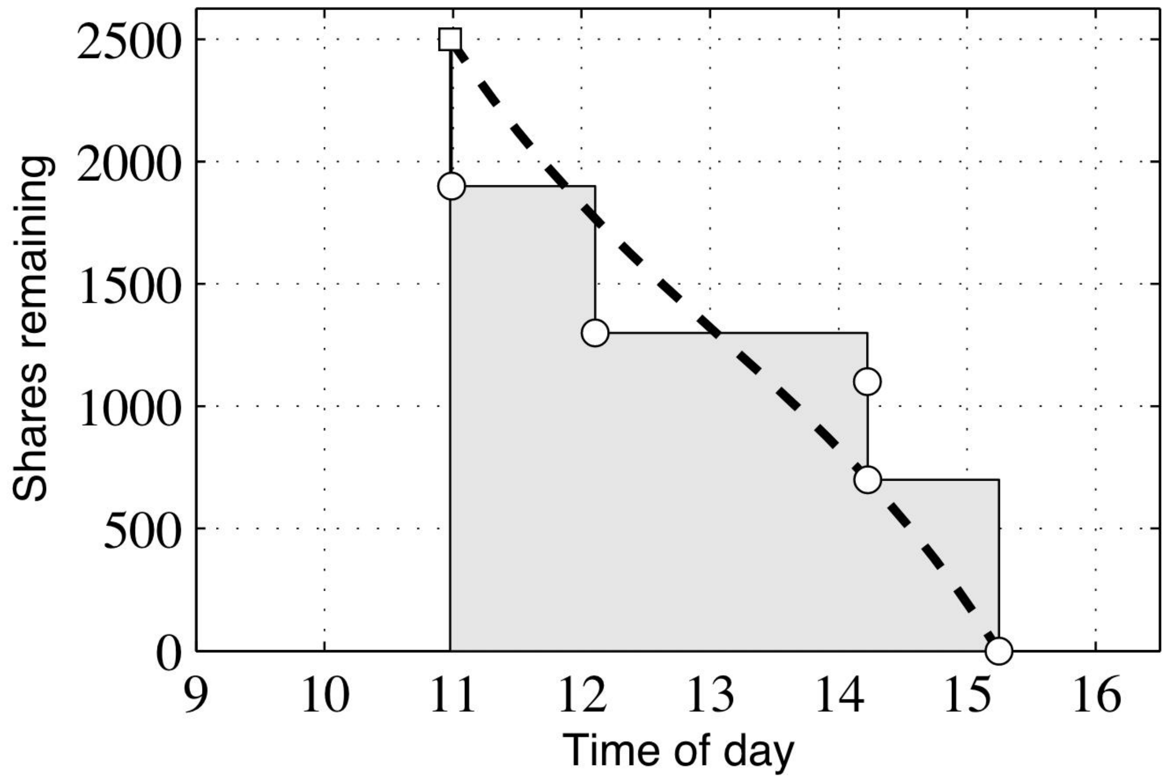


Figure: A typical trading trajectory.

The **vertical axis** represents shares remaining; each step downwards is one execution.
The trajectory **starts** at the first transaction (order submission) recorded in the system;
the program **ends** when the last execution has been completed.
The **dashed line** is the continuous-time approximation that we use in the modeling.

2.2 Variables

Observables

Let S(t) be the price of the asset being traded.
For each order, they define the following price points of interest:

$$\begin{aligned} S_0 &= \text{market price before this order begins executing} \\ S_{\text{post}} &= \text{market price after this order is completed} \\ \bar{S} &= \text{average realized price on the order} \end{aligned}$$

The pre-trade price S_0 is the price before the impact of the trade begins to be felt (this is an approximation, since some information may leak before any record enters the system).
The market prices S_0 and S_{post} are bid-ask midpoints from TAQ. They compute S_0 from the latest quote just preceding the first transaction.
The realized price $\bar{S}=\sum x_j S_j/\sum x_j$ is computed from the transaction data set (x_j, S_j are the sizes and prices of the individual executions).

2.2 Variables

Observables

The post-trade price S_{post} should capture the "permanent" effects of the trade program. That is, it should be taken long enough after the last execution that any effects of temporary liquidity effects have dissipated. For shorter time intervals the regressed values depend on the timelag, and at about this level the variation stops. That is, they define

$$t_{\text{post}} = t_n + \text{one half-hour.}$$

The price S_{post} is taken from the first quote following t_{post}. If t_{post} is after market close, then they carry over to the next morning.

2.2 Variables

Observables

Based on these prices, they define the following impact variables:

$$\text{Permanent impact:} \quad I = \frac{S_{\text{post}} - S_0}{S_0}$$

$$\text{Realized impact:} \quad J = \frac{\bar{S} - S_0}{S_0}.$$

The "effective" impact J is the quantity of most interest, since it determines the actual cash received or spent on the trade. In the model below, they will define temporary impact to be J minus a suitable fraction of I, and this temporary impact will be the quantity described by this paper.

2.2 Variables

Volume time

This paper map each of the clock times $t_0, \ldots, t_n$ in the data set to a corresponding volume time $\tau_0, \ldots, \tau_n$.

Since the stocks in sample are heavily traded, in this paper they use a nonparametric estimator that directly measures differences in τ: the shares traded during the period corresponding to the exection of each order.

2.2 Variables

Explanatory variables

This paper wants to describe the impacts I and J in terms of the input quantities

$$\begin{aligned} X &= \sum_{j=1}^{n} x_j &&= \text{Total executed size in shares} \\ T &= \tau_n - \tau_0 &&= \text{Volume duration of active trading} \\ T_{\text{post}} &= \tau_{\text{post}} - \tau_0 &&= \text{Volume duration of impact.} \end{aligned}$$

As noted above, **X** is **positive** for a **buy** order, **negative** for **sell**. They have explored defining **T** using a size-weighted average of execution times but the results are not substantially different. They make no use of the intermediate execution times $\tau_1, \ldots, \tau_{n-1}$, and make no use of the execution sizes except in computing the order size and the mean realised price.

2.2 Variables

Auxiliary variables

Although their goal is to explain the dependence of the impact costs I, J on order size X and trade time T, other market variables will influence the solution. The most important of these are

$$\begin{aligned} V &= \text{Average daily volume in shares, and} \\ \sigma &= \text{Daily volatility.} \end{aligned}$$

V is a ten-day moving average.
For volatility, they use an intraday estimator that makes use of every transaction in the day.
They find that it is important to track changes in these variables not only between different stocks but also across time for the same stock.

03

PART

Trajectory Cost Model

3. Trajectory Cost Model

(1) Based on the framework developed by previous literature

- the rate of trading is constant in volume time
- neglect cross-impact
- the asset under discussion is a single stock

(2) Price impact:

- **Permanent impact**: that reflects the information transmitted to the market by the imbalance.
- **Temporary impact**: the price concession needed to attract counterparties within a specified short time interval.

(3) Model:

Realized=Permanent+Temporary+Noise

3.1 Permanent impact & Temporary impact

Permanent impact

$$dS = S_0\, g(v)\, d\tau + S_0\, \sigma\, dB$$

$B(\tau)$ standard Brownian motion

$g(v)$ increasing $g(0) = 0$

$v = X/T \qquad 0 \le \tau \le T$

$$I = T\, g\left(\frac{X}{T}\right) + \sigma \sqrt{T_{\text{post}}}\, \xi$$

Temporary impact

$$\tilde{S}(\tau) = S(\tau) + S_0\, h\left(\frac{X}{T}\right)$$

constant liquidation rate

$$J - \frac{I}{2} = h\left(\frac{X}{T}\right) + \sigma \left(\sqrt{\frac{T}{12}\left(4 - 3\frac{T}{T_{\text{post}}}\right)}\, \chi - \frac{T_{\text{post}} - T}{2\sqrt{T_{\text{post}}}}\, \xi \right)$$

3.2 Choice of functional form

$$I = T\,g\left(\frac{X}{T}\right) + \sigma\sqrt{T_{\text{post}}}\,\xi$$

$$J - \frac{I}{2} = h\left(\frac{X}{T}\right) + \sigma\left(\sqrt{\frac{T}{12}\left(4 - 3\frac{T}{T_{\text{post}}}\right)}\,\chi - \frac{T_{\text{post}} - T}{2\sqrt{T_{\text{post}}}}\,\xi\right)$$

$$g(v) = \pm\gamma|v|^{\alpha}$$

$$h(v) = \pm\eta|v|^{\beta}$$

$$\alpha = 1$$

$$\beta = \frac{1}{2}$$

single stock fits

a universe of stocks

04 PART

Cross-Sectional Description

4.1 Parameter scaling

$X \longrightarrow X/VT$

Motion of Price $\longrightarrow$ the volatility σ

$$I = T\,g\left(\frac{X}{T}\right) + \sigma\sqrt{T_{\text{post}}}\,\xi$$

$$J - \frac{I}{2} = h\left(\frac{X}{T}\right) + \sigma\left(\sqrt{\frac{T}{12}\left(4 - 3\frac{T}{T_{\text{post}}}\right)}\,\chi - \frac{T_{\text{post}} - T}{2\sqrt{T_{\text{post}}}}\,\xi\right)$$

$$\Downarrow$$

$$I = \sigma\,T\,g\left(\frac{X}{VT}\right) + \langle\text{noise}\rangle$$

$$J - \frac{I}{2} = \sigma\,h\left(\frac{X}{VT}\right) + \langle\text{noise}\rangle$$

4.2 Model determination

$$\mathcal{L} = \left(\frac{\Theta}{V}\right)^{\delta}$$

$$I = \sigma T g\left(\frac{X}{VT}\right) + \langle\text{noise}\rangle$$

$$J - \frac{I}{2} = \sigma h\left(\frac{X}{VT}\right) + \langle\text{noise}\rangle$$

$$\alpha = 0.891 \pm 0.10$$
$$\delta = 0.267 \pm 0.22$$
$$\beta = 0.600 \pm 0.038$$

$$\alpha = 1$$
$$\delta = 1/4$$
$$\beta = 3/5$$

$$\frac{I}{\sigma} = \gamma T \operatorname{sgn}(X) \left|\frac{X}{VT}\right|^{\alpha} \left(\frac{\Theta}{V}\right)^{\delta} + \langle\text{noise}\rangle$$

$$\frac{1}{\sigma}\left(J - \frac{I}{2}\right) = \eta \operatorname{sgn}(X) \left|\frac{X}{VT}\right|^{\beta} + \langle\text{noise}\rangle$$

4.3 Determination of coefficients

$$\frac{I}{\sigma} = \gamma\, T\, \mathrm{sgn}(X) \left|\frac{X}{VT}\right|^{\alpha} \left(\frac{\Theta}{V}\right)^{\delta} + \langle\text{noise}\rangle$$

$$\frac{1}{\sigma}\left(J - \frac{I}{2}\right) = \eta\, \mathrm{sgn}(X) \left|\frac{X}{VT}\right|^{\beta} + \langle\text{noise}\rangle$$

$$\alpha = 1$$

$$\delta = 1/4$$

$$\beta = 3/5$$

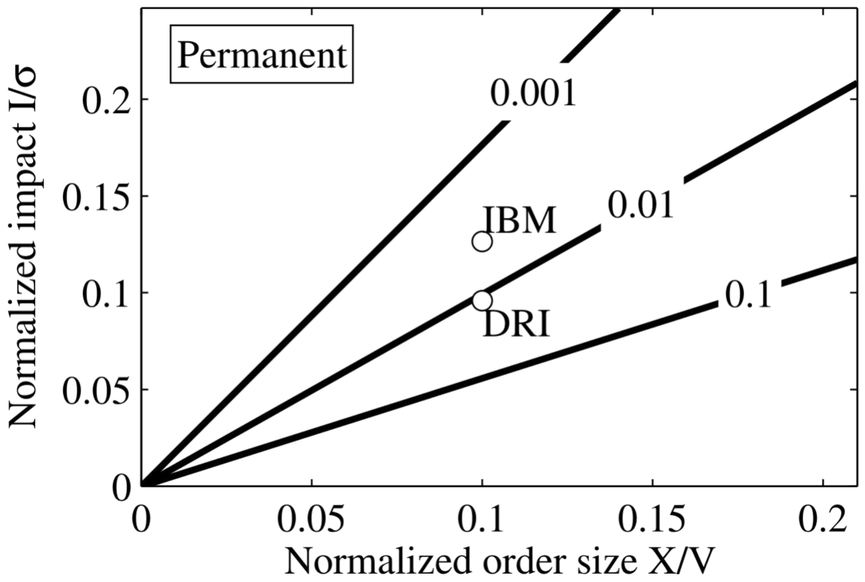
$$I = \gamma\, \sigma\, \frac{X}{V} \left(\frac{\Theta}{V}\right)^{1/4} + \langle\text{noise}\rangle$$

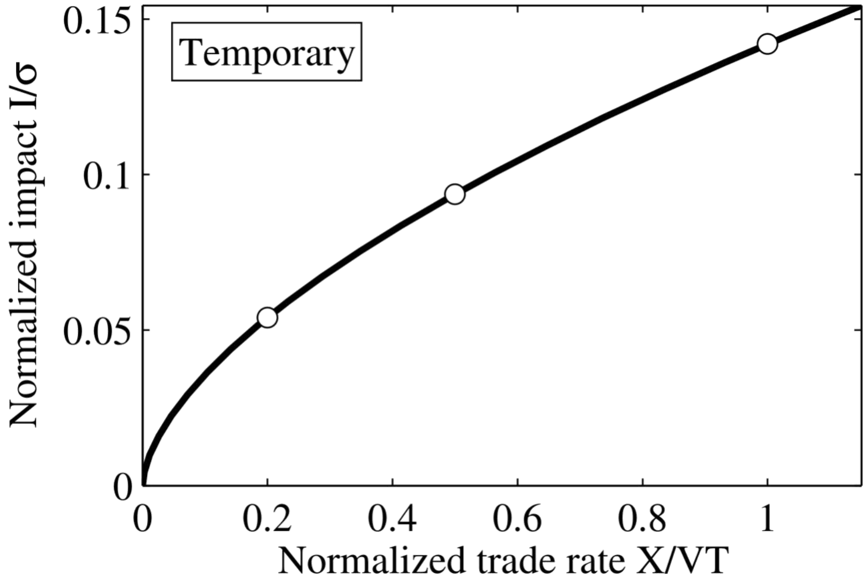
$$J = \frac{I}{2} + \mathrm{sgn}(X)\, \eta\, \sigma \left|\frac{X}{VT}\right|^{3/5} + \langle\text{noise}\rangle$$

4.3 Determination of coefficients

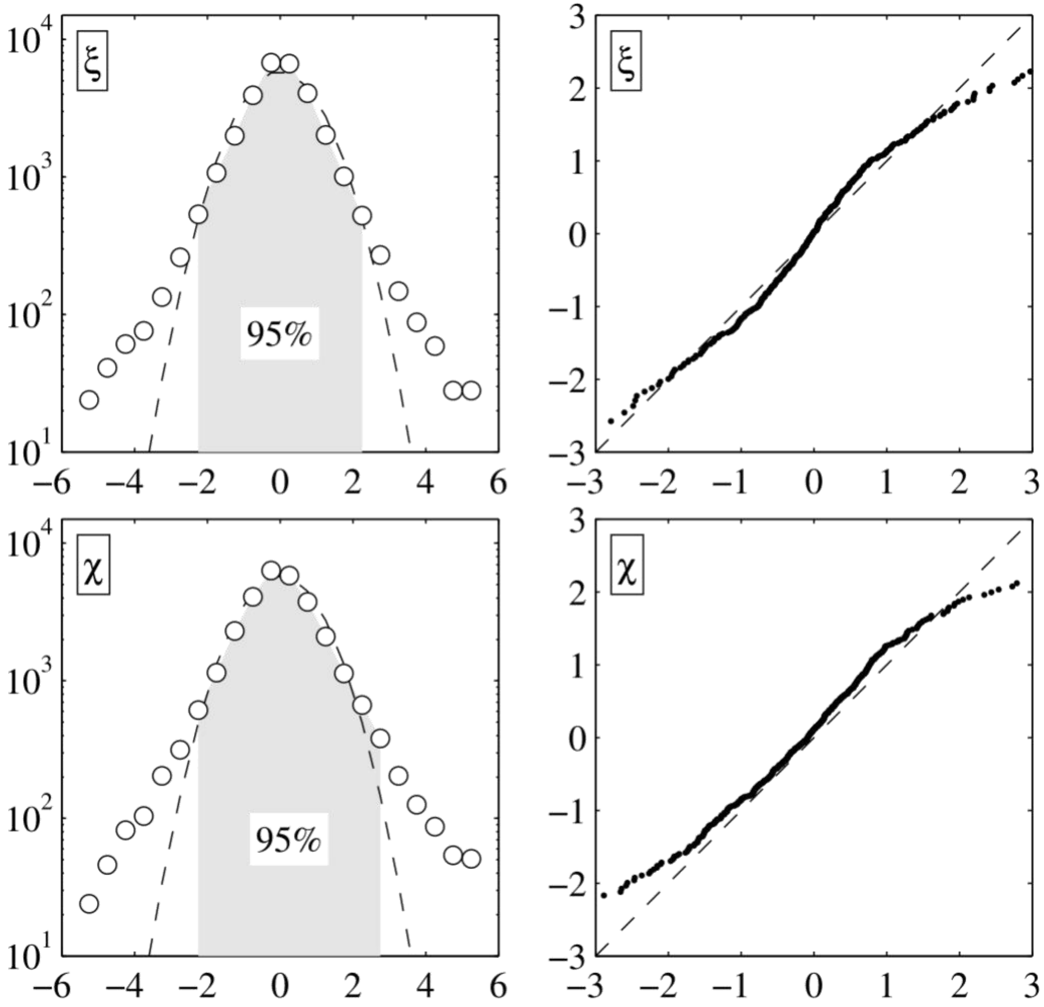
$$I = \gamma\sigma\frac{X}{V}\left(\frac{\Theta}{V}\right)^{1/4} + \langle\text{noise}\rangle$$

$$J = \frac{I}{2} + \operatorname{sgn}(X)\,\eta\,\sigma\left|\frac{X}{VT}\right|^{3/5} + \langle\text{noise}\rangle$$





4.4 Residual analysis



05

PART

Summary

5 Summary

- used a **large data sample** of US institutional orders
- model to **estimate price impact** functions for equity trades on large-cap stocks
- give **quantitatively accurate** pre-trade cost estimates
- the model is **being incorporated** into Citigroup's software

Thank you

清華大學