

Market Impact: A Vector Autoregressive with Exogenous Variables Approach

David Han

September 1, 2020

1 Introduction

In the algorithmic execution design and implementation, there are three major components:

- **Scheduler.** The scheduler decides an optimal schedule for the algorithms to follow and also provides a pair of upper and lower bands to make sure the algorithm follows the schedule closely enough.
- **Microtrader.** Given the target quantity to trade and the tolerance bands from scheduler, the microtrader decides how many child orders to trade and the aggressiveness of each child slice, taking into account the execution progress and short-term signals on market price movement.
- **SOR:** The Smart Order Router decides how to split each child order given by the microtrader and to which venue to send each slice.

Market impact plays a critical role in deciding the optimal trading trajectory and it is extensively studied in the last two decades, starting with the seminal paper *Optimal Execution of Portfolio Transactions* by Robert Almgren and Neil Chriss in 2000, which introduced a market impact model and proposed that optimal trade execution schedule is a trade-off between trading slowly to minimize market impact and trading rapidly to reduce timing risk relative to an arrival price.

2 Literature Review

There are three main streams of modeling market impact in academia and algorithmic trading practice:

- **Square root law.** This model is widely used before 2000. Roughly, the model has the following form:

$$\Delta P = \alpha \times \sigma \sqrt{\frac{S}{V}}$$

where σ is the daily volatility, S is the order size in shares, V is the daily volume and α is a constant to be estimated. This simple model assumes that market impact solely depends on the relative order size but fails to take into account the pace of execution, trading style, stock volatility, etc.

- **Almgren and Chriss framework.** This model is very popular among sell-side institutions as a basis for arrival price benchmark execution algorithms such as Implementation Shortfall. One of the popular functional form for market impact is

$$MI = \alpha \cdot \sigma^\beta \cdot \rho^\gamma$$

where α , β and γ are three parameters to be calibrated, σ is the annualized volatility and ρ is the participation rate. Taking the logarithmic operation on both sides of the above model yields

$$\log(MI) = \log\alpha + \beta\log\sigma + \gamma\log\rho$$

This model is very easy to calibrate using the standard *OLS*.

- **Propagator model.** In a continuous time propagation model, the stock price P_t at time t is given by

$$P_t = P_0 + P_0 \int_0^t f(v_s)g(t-s)ds + noise$$

where t is the volume time, v is the trading rate at time $s < t$, $f(\cdot)$ and $g(\cdot)$ are the impulse and decay functions. An appropriate functional form, *e.g.* power law function, can be assumed for f and g and a proprietary execution data can be used to calibrate the model.

In summary, all these models fail to address a few key aspects a modern trader is interested in:

- These models are linear or indirectly linear, *i.e.* linear after some transformation, so as to make it easy to calibrate in practice.
- These models focus on estimating the market impact of trading a single security, *i.e.* failing to exam the cross-asset impact.
- Market impact of portfolio trading and single stock trading.

This paper aims to provide a framework to model market impact and addresses the above-mentioned shortcomings in the current market impact research.

3 The Proposed Model

3.1 Vector Autoregressive (VAR)

The rapid development of modern trading technology has accelerated the integration of global financial markets and the price movement in one market/instrument can easily spread to other markets/instruments. Today's financial instruments are more dependent on each other and their dynamics must be jointly modeled. Vector autoregressive models provide an important research tool to analyze and predict the dynamics of prices of different instruments. A p -th order vector autoregressive, denoted VAR(p), is

$$\mathbf{y}_t = \alpha + \beta_1 \mathbf{y}_{t-1} + \beta_2 \mathbf{y}_{t-2} + \cdots + \beta_p \mathbf{y}_{t-p} + \epsilon_t$$

3.2 VAR with Exogenous Variables

A VAR process can be affected by other observable variables that are determined outside the system of interest. Such variables are called exogenous (independent) variables. Exogenous variables can be stochastic or non-stochastic. The process can also be affected by the lags of exogenous variables. A model used to describe this process is called a VARX(p, s) model which is written as

$$\mathbf{y}_t = \alpha + \sum_{i=1}^p \Phi_i \mathbf{y}_{t-i} + \sum_{i=1}^s \Theta_i \mathbf{x}_{t-i} + \epsilon_t$$

where $\mathbf{x}_t = (\mathbf{x}_{1t}, \dots, \mathbf{x}_{rt})'$ is a r -dimensional time series vector, Θ_i is a $k \times r$ matrix and ϵ_t is a sequence of *i.i.d.* random vectors with mean zero and positive-definite covariance Σ_ϵ .

The VARX model can explain the dynamics between endogenous and exogenous variables. It can also help us understand the impact of a variable or a set of variables on others through the impulse response function (IRF). Furthermore, VARX can be used to predict and forecast time series data.

3.3 The Model

Without loss of generality, a VARX(1, 0) model is proposed as follows:

$$\mathbf{y}_t = \alpha + \Phi_1 \mathbf{y}_{t-1} + \Theta_1 \mathbf{x}_t + \epsilon_t$$

where $\mathbf{y}_t = (y_{1t}, y_{2t})'$ and $\mathbf{x}_t = (x_{1t}, x_{2t}, x_{3t}, x_{4t})'$. For example,

- y_1 : the price return of interest rate futures of 10-year treasury bill
- y_2 : the price return of interest rate futures of 5-year treasury bill
- x_1 : the participation rate of trading interest rate futures of 10-year treasury bill
- x_2 : the participation rate of trading interest rate futures of 5-year treasury bill
- x_3 : the volatility of interest rate futures of 10-year treasury bill
- x_4 : the volatility of interest rate futures of 5-year treasury bill

More details are discussed in section 5 on examining:

- the optimal number of time lags
- unit root and stationary test
- Granger causality of the exogenous variables

4 Data

4.1 Tick and execution data

The following information is required to calibrate a VARX(1, 0) model specified in the previous section:

time: timestamp of each child execution

price: price of each child execution

size: size of each child execution

arrivalPrice: the arrival mid-quote price as of the start time of sampling period

volatility: intra-day price volatility in each sampling period

Some other information that are relevant to the model

4.2 Descriptive statistics

[\[Table inserted here\]](#)

5 Considerations for Model Fitting

5.1 Correlation within and among time series

Autocorrelation function (ACF)

See Python function `statsmodels.tsa.stattools.acf`.

[\[Charts inserted here\]](#)

Partial autocorrelation function (PACF)

See Python function `statsmodels.tsa.stattools.pacf`.

[\[Charts inserted here\]](#)

Cross-correlation function (CCF)

See Python function `statsmodels.tsa.stattools.ccf`.

[\[Charts inserted here\]](#)

5.2 Unit root and stationarity test

See Python function `statsmodels.tsa.stattools.adfuller` for unit root test and `statsmodels.tsa.stattools.kpss` to test trend stationarity.

5.3 Granger causality

See Python function `statsmodels.tsa.stattools.grangercausalitytests` for Granger causality test.

5.4 Information criteria

Multiple information criteria can be used to select the p and s in the VARX(p, s) model. For example,

- Bayesian information criteria (BIC)
- Akaike information criterion(AIC)
- Hannan-Quinn information criterion (HQ)

5.5 Model fitting

This mode can be fitted with VARMAX models in Python package statsmodels.

6 Application

This model is flexible enough to answer the following questions:

- how exogenous variables like participation rate, volatility, relative order size, affect market impact via the impulse response function.
- how to model the market impact of different horizons from minutes, hours to days by defining different lengths of time period and time lags.
- compare the market impact of trading one instrument vs multiple instruments

References

- [1] Robert Almgren and Neil Chriss. *Optimal Execution of Portfolio Transactions*. Journal of Risk, 5-30. 2000.
- [2] James Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ. 1994.
- [3] Ruey S. Tsay. *Multivariate Time Series Analysis with R and Financial Applications*. Wiley. Hoboken, NJ. 2014.
- [4] Statsmodels in Python, <https://www.statsmodels.org/stable/index.html>