The Amazing Power of Dimensional Analysis in Finance: Market Impact and the Intraday Trading Invariance Hypothesis

W. Schachermayer
joint work with M. Pohl, A. Ristig, L. Tangpi

University of Vienna
Faculty of Mathematics
Dimensional Analysis

The period of the pendulum

Functional relation:

\[ \text{period} = f(l, m, g). \]
The period of the pendulum

Functional relation:

\[ \text{period} = f(l, m, g). \]

Dimensions:

- the length \( l \) of the pendulum, measured in meters: dimension \( \mathbb{L} \).
The period of the pendulum

Functional relation:

\[ \text{period} = f(l, m, g). \]

Dimensions:
- the length \( l \) of the pendulum, measured in meters: dimension \( \mathbb{L} \).
- the mass \( m \) of the bob, measured in grams: dimension \( \mathbb{M} \).
Dimensional Analysis

The period of the pendulum

Functional relation:

$$\text{period} = f(l, m, g).$$

Dimensions:

- the length $l$ of the pendulum, measured in meters: dimension $\mathbb{L}$.
- the mass $m$ of the bob, measured in grams: dimension $\mathbb{M}$.
- the acceleration $g$ caused by gravity, measured in meters per second squared: dimension $\mathbb{L}/\mathbb{T}^2$. 
Dimensional Analysis

The period of the pendulum

Functional relation:

\[ \text{period} = f(l, m, g). \]

Dimensions:

- the length \( l \) of the pendulum, measured in meters: dimension \( \mathbb{L} \).
- the mass \( m \) of the bob, measured in grams: dimension \( \mathbb{M} \).
- the acceleration \( g \) caused by gravity, measured in meters per second squared: dimension \( \mathbb{L}/\mathbb{T}^2 \).

Basic assumption: The three variables \( l, m, g \) fully explain the period of the pendulum.
Dimensional Analysis: The period of the pendulum

Ansatz:

\[
\text{period} = p = f(l, m, g) = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}.
\]
Dimensional Analysis: The period of the pendulum

Ansatz:

\[ \text{period} = p = f(l, m, g) = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3} . \]

<table>
<thead>
<tr>
<th></th>
<th>( l )</th>
<th>( m )</th>
<th>( g )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l ) length</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( m ) mass</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( T ) time</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
y_1 + y_3 &= 0 \\
y_2 &= 0 \\
-2y_3 &= 1
\end{align*}
\]
Ansatz:

\[
\text{period} = p = f(l, m, g) = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}.
\]

<table>
<thead>
<tr>
<th></th>
<th>(l)</th>
<th>(m)</th>
<th>(g)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>mass</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>time</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>1</td>
</tr>
</tbody>
</table>

Unique solution: \(y_1 = \frac{1}{2}, y_2 = 0, y_3 = \frac{1}{2}\),

\[
\text{period} = \text{const} \cdot \sqrt{\frac{l}{g}}.
\]
Remark: The Ansatz

\[ \text{period} = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}, \]

does \textbf{not restrict the generality} of the relation

\[ \text{period} = f(l, m, g), \]
Remark: The Ansatz

\[ \text{period} = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}, \]

does not restrict the generality of the relation

\[ \text{period} = f(l, m, g), \]

or, equivalently

\[ \log(\text{period}) = F(\log(l), \log(m), \log(g)). \]
Remark: The Ansatz

\[ \text{period} = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}, \]

does \textbf{not restrict the generality} of the relation

\[ \text{period} = f(l, m, g), \]

or, equivalently

\[ \log(\text{period}) = F(\log(l), \log(m), \log(g)). \]

Indeed, the first row of the matrix translates into the requirement

\[ \log(\text{period}) = F(\log(l) + \log(x), \log(m), \log(g) + \log(x)), \forall x > 0. \]
Remark: The Ansatz

\[ \text{period} = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}, \]

does not restrict the generality of the relation

\[ \text{period} = f(l, m, g), \]

or, equivalently

\[ \log(\text{period}) = F(\log(l), \log(m), \log(g)). \]

Indeed, the first row of the matrix translates into the requirement

\[ \log(\text{period}) = F(\log(l) + \log(x), \log(m), \log(g) + \log(x)), \forall x > 0. \]

Hence \( F(\cdot, \cdot, \cdot) \) must be constant along any line in \( \mathbb{R}^3 \) parallel to the vector \( (1, 0, 1) \).
Remark: The Ansatz

\[ \text{period} = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}, \]

does not restrict the generality of the relation

\[ \text{period} = f(l, m, g), \]

or, equivalently

\[ \log(\text{period}) = F(\log(l), \log(m), \log(g)). \]

Indeed, the first row of the matrix translates into the requirement

\[ \log(\text{period}) = F(\log(l) + \log(x), \log(m), \log(g) + \log(x)), \forall x > 0. \]

Hence \( F(\cdot, \cdot, \cdot) \) must be constant along any line in \( \mathbb{R}^3 \) parallel to the vector \( (1, 0, 1) \).

As the row vectors of the above matrix span \( \mathbb{R}^3 \), this fully determines the function \( F \) (up to a constant).
Consider an agent who intends to buy/sell a large amount ("meta-order" or "bet") of some fixed stock.
Consider an agent who intends to buy/sell a large amount ("meta-order" or "bet") of some fixed stock. This bet will – ceteris paribus – move the market price to the disadvantage of the agent.
Market Impact

Consider an agent who intends to buy/sell a large amount ("meta-order" or "bet") of some fixed stock. This bet will – ceteris paribus – move the market price to the disadvantage of the agent. The agents are trying hard to minimize the market impact, but are unable to avoid it completely.
Consider an agent who intends to buy/sell a large amount ("meta-order" or "bet") of some fixed stock. This bet will – ceteris paribus – move the market price to the disadvantage of the agent. The agents are trying hard to minimize the market impact, but are unable to avoid it completely.

**Definition**

The market impact $G$ is the size of price change caused by a bet (in percentage of the price).
Functional relation:

market impact $G = g(Q, P, V, \sigma^2)$. 
Functional relation:

\[ G = g(Q, P, V, \sigma^2). \]

- \( Q \) the size of the meta-order, measured in units of shares \( S \),
Dimensional Analysis: The size of the market impact

Functional relation:

\[
G = g(Q, P, V, \sigma^2).
\]

- \( Q \): the size of the meta-order, measured in units of shares \( S \),
- \( P \): the price of the stock, measured in units of money per share \( U/S \),
- \( V \): the traded volume of the stock, measured in units of shares per time \( S/T \),
- \( \sigma^2 \): the squared volatility of the stock, measured in percentage of the stock price per unit of time \( T-1 \).
Functional relation:

\[ G = g(Q, P, V, \sigma^2). \]

- \( Q \) the size of the meta-order, measured in units of shares \( S \),
- \( P \) the price of the stock, measured in units of money per share \( U/S \),
- \( V \) the traded volume of the stock, measured in units of shares per time \( S/T \),
- \( \sigma^2 \) the squared volatility of the stock, measured in percentage of the stock price per unit of time \( T^{-1} \),
- \( G \) the market impact is a dimensionless quantity.

Basic assumption: the four variables \( Q, P, V, \sigma^2 \) fully explain the size of the market impact \( G \).
Dimensional Analysis: The size of the market impact

Functional relation:

\[ G = g(Q, P, V, \sigma^2). \]

- \( Q \) the size of the meta-order, measured in units of shares \( S \),
- \( P \) the price of the stock, measured in units of money per share \( U/S \),
- \( V \) the traded volume of the stock, measured in units of shares per time \( S/T \),
- \( \sigma^2 \) the squared volatility of the stock, measured in percentage of the stock price per unit of time \( T^{-1} \).
Functional relation:

market impact $G = g(Q, P, V, \sigma^2)$.

- $Q$ the size of the meta-order, measured in units of shares $S$,
- $P$ the price of the stock, measured in units of money per share $U/S$,
- $V$ the traded volume of the stock, measured in units of shares per time $S/T$,
- $\sigma^2$ the squared volatility of the stock, measured in percentage of the stock price per unit of time $T^{-1}$,
- $G$ the market impact is a dimensionless quantity.
Functional relation:

\[ G = g(Q, P, V, \sigma^2). \]

- \( Q \) the size of the meta-order, measured in units of shares \( S \),
- \( P \) the price of the stock, measured in units of money per share \( US \),
- \( V \) the traded volume of the stock, measured in units of shares per time \( S/T \),
- \( \sigma^2 \) the squared volatility of the stock, measured in percentage of the stock price per unit of time \( T^{-1} \),
- \( G \) the market impact is a dimensionless quantity.

Basic assumption: the four variables \( Q, P, V, \sigma^2 \) fully explain the size of the market impact \( G \).
Dimensional Analysis: The size of the market impact

Ansatz:

\[ G = g(Q, P, V, \sigma^2) = \text{const} \cdot Q^{y_1} P^{y_2} V^{y_3} \sigma^{2y_4}, \]

\( Q \) = size of bet, \( P \) = price of share, \( V \) = traded daily volume, \( \sigma \) = volatility.
Ansatz:

\[ G = g(Q, P, V, \sigma^2) = \text{const} \cdot Q^{y_1} P^{y_2} V^{y_3} \sigma^{2y_4}, \]

\(Q = \text{size of bet}, \ P = \text{price of share}, \ V = \text{traded daily volume}, \ \sigma = \text{volatility}.\)

<table>
<thead>
<tr>
<th></th>
<th>(Q)</th>
<th>(P)</th>
<th>(V)</th>
<th>(\sigma^2)</th>
<th>(G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>shares (\mathcal{S})</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>money (\mathcal{U})</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>time (\mathcal{T})</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>
Dimensional Analysis: The size of the market impact

Ansatz:

\[ G = g(Q, P, V, \sigma^2) = \text{const} \cdot Q^{y_1} P^{y_2} V^{y_3} \sigma^{2y_4}, \]

\( Q = \) size of bet, \( P = \) price of share, \( V = \) traded daily volume, \( \sigma = \) volatility.

<table>
<thead>
<tr>
<th></th>
<th>( Q )</th>
<th>( P )</th>
<th>( V )</th>
<th>( \sigma^2 )</th>
<th>( G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>shares ( S )</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>money ( U )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>time ( T )</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>

Leads to three linear equations in four unknowns \( y_1, y_2, y_3, y_4 \). The solution has one degree of freedom

\[ G = \text{const} \cdot \left( \frac{Q\sigma^2}{V} \right)^y, \]

where \( y \in \mathbb{R} \) and \( \text{const} > 0 \) are still free.
This time the ansatz **does** restrict the generality! The general solution for $G$, respecting the dimensional restrictions is

$$G = g\left(\frac{Q\sigma^2}{V}\right),$$

where $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is an *arbitrary* function.
This time the ansatz **does** restrict the generality! The general solution for $G$, respecting the dimensional restrictions is

$$G = g\left(\frac{Q\sigma^2}{V}\right),$$

where $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is an *arbitrary* function.

Can we find one more equation which will allow us to get a unique solution?
This time the ansatz **does** restrict the generality! The general solution for $G$, respecting the dimensional restrictions is

$$G = g \left( \frac{Q \sigma^2}{V} \right),$$

where $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is an *arbitrary* function.

Can we find one more equation which will allow us to get a unique solution?

Kyle, Obizhaeva (2016): YES
Leverage neutrality

Theorem of Modigliani-Miller (1958):

\[ A_t \] assets
\[ D_t \] debt
\[ E_t \] equity

\[ \begin{array}{c}
\text{Assets} \\
A_t \\
\end{array} \quad \text{Liabilities} \quad \begin{array}{c}
D_t \\
E_t \\
\end{array} \]

Basic assumption: \( (A_t) \geq 0 \) follows a stochastic process, e.g. Samuelson (1965):

\[
dA_t = (\sigma dW_t + \mu dt).
\]

Keeping the debt \( D_t \) constant, we therefore get

\[
dA_t = dE_t \quad \text{so that}
\]

\[
dE_t = A_t E_t (\sigma dW_t + \mu dt).
\]

Conclusion: Denoting by \( L_t = A_t E_t \) the leverage of the company, the relative dynamics of \( E_t \geq 0 \) are simply proportional to the leverage \( L_t \).
Leverage neutrality

Theorem of Modigliani-Miller (1958):

\[ A_t \text{ assets} \]
\[ D_t \text{ debt} \]
\[ E_t \text{ equity} \]

Basic assumption: \((A_t)_{t \geq 0}\) follows a stochastic process, e.g. Samuelson (1965):

\[
\frac{dA_t}{A_t} = (\sigma dW_t + \mu dt).
\]
Leverage neutrality

Theorem of Modigliani-Miller (1958):

\[ \begin{array}{c|c|c}
\text{Assets} & \text{Liabilities} \\
\hline
A_t & D_t \\
E_t & E_t \\
\end{array} \]

Basic assumption: \( (A_t)_{t \geq 0} \) follows a stochastic process, e.g. Samuelson (1965):

\[
\frac{dA_t}{A_t} = (\sigma dW_t + \mu dt).
\]

Keeping the debt \( D_t \) constant, we therefore get \( dA_t = dE_t \) so that

\[
\frac{dE_t}{E_t} = \frac{A_t}{E_t} (\sigma dW_t + \mu dt).
\]
Leverage neutrality

Theorem of Modigliani-Miller (1958):

\[ A_t \text{ assets} \]
\[ D_t \text{ debt} \]
\[ E_t \text{ equity} \]

Basic assumption: \((A_t)_{t \geq 0}\) follows a stochastic process, e.g. Samuelson (1965):

\[
\frac{dA_t}{A_t} = (\sigma dW_t + \mu dt).
\]

Keeping the debt \(D_t\) constant, we therefore get \(dA_t = dE_t\) so that

\[
\frac{dE_t}{E_t} = \frac{A_t}{E_t} (\sigma dW_t + \mu dt).
\]

Conclusion: Denoting by \(L_t = \frac{A_t}{E_t}\) the leverage of the company, the relative dynamics of \((E_t)_{t \geq 0}\) are simply proportional to the leverage \(L_t\).
What happens to the stock price \( (P_t)_{t \geq 0} \), if you change the leverage? Say, the leverage \( L \) is doubled by paying out half of the equity by dividends.
What happens to the stock price \((P_t)_{t \geq 0}\), if you change the leverage? Say, the leverage \(L\) is doubled by paying out half of the equity by dividends

- \(P\) is multiplied by \(\frac{1}{2}\).
What happens to the stock price \((P_t)_{t\geq 0}\), if you change the leverage? Say, the leverage \(L\) is doubled by paying out half of the equity by dividends

- \(P\) is multiplied by \(\frac{1}{2}\).
- \(\sigma\) is multiplied by 2.
What happens to the stock price \((P_t)_{t \geq 0}\), if you change the leverage? Say, the leverage \(L\) is doubled by paying out half of the equity by dividends

- \(P\) is multiplied by \(\frac{1}{2}\).
- \(\sigma\) is multiplied by 2.
- \(G\) is multiplied by 2.
What happens to the stock price \((P_t)_{t \geq 0}\), if you change the leverage? Say, the leverage \(L\) is doubled by paying out half of the equity by dividends

- \(P\) is multiplied by \(\frac{1}{2}\).
- \(\sigma\) is multiplied by 2.
- \(G\) is multiplied by 2.
- \(Q, V\) remain unchanged.
Leverage neutrality: (Kyle, Obizhaeva, 2017)
The value of a firm does not depend on its capital structure (Modigliani, Miller, 1958).
Leverage neutrality: (Kyle, Obizhaeva, 2017)
The value of a firm does not depend on its capital structure (Modigliani, Miller, 1958).
This no-arbitrage-type condition yields one more equation involving the *Modigliani-Miller* dimension $\mathbb{M}$ measuring the leverage of a company.
Leverage neutrality: (Kyle, Obizhaeva, 2017)
The value of a firm does not depend on its capital structure (Modigliani, Miller, 1958).
This no-arbitrage-type condition yields one more equation involving the *Modigliani-Miller* dimension $\mathbb{M}$ measuring the leverage of a company.
Mathematically speaking, the variation of the leverage (dimension $\mathbb{M}$) is analogous to the scalings of the dimensions $\mathbb{S}$, $\mathbb{U}$, and $\mathbb{T}$. 
Leverage neutrality: (Kyle, Obizhaeva, 2017)
The value of a firm does not depend on its capital structure (Modigliani, Miller, 1958).
This no-arbitrage-type condition yields one more equation involving the Modigliani-Miller dimension $M$ measuring the leverage of a company.
Mathematically speaking, the variation of the leverage (dimension $M$) is analogous to the scalings of the dimensions $S$, $U$, and $T$.

<table>
<thead>
<tr>
<th></th>
<th>$Q$</th>
<th>$P$</th>
<th>$V$</th>
<th>$\sigma^2$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$U$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$T$</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>$M$</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Theorem (Pohl, Ristig, S., Tangpi, '17 based on Kyle, Obizhaeva, '16):

Assume $G = g(Q, P, V, \sigma^2)$ is such that

- the variables $Q, P, V$, and $\sigma^2$ fully explain $G$,
- the function $g$ is invariant under scalings of the dimensions $S, U, T$,
Theorem (Pohl, Ristig, S., Tangpi, '17 based on Kyle, Obizhaeva, '16):

Assume $G = g(Q, P, V, \sigma^2)$ is such that

- the variables $Q, P, V$, and $\sigma^2$ fully explain $G$,
- the function $g$ is invariant under scalings of the dimensions $S, U, T$, and leverage neutrality holds true.
Theorem (Pohl, Ristig, S., Tangpi, ’17 based on Kyle, Obizhaeva, ’16):

Assume $G = g(Q, P, V, \sigma^2)$ is such that

- the variables $Q, P, V$, and $\sigma^2$ fully explain $G$,
- the function $g$ is invariant under scalings of the dimensions $S, U, T$, and leverage neutrality holds true.

Then there is a const $> 0$ such that

$$G = \text{const} \cdot \sigma \sqrt{\frac{Q}{V}}.$$
Theorem (Pohl, Ristig, S., Tangpi, ’17 based on Kyle, Obizhaeva, ’16):

Assume $G = g(Q, P, V, \sigma^2)$ is such that

- the variables $Q, P, V$, and $\sigma^2$ fully explain $G$,
- the function $g$ is invariant under scalings of the dimensions $S, U, T$, and leverage neutrality holds true.

Then there is a const $> 0$ such that

$$G = \text{const} \cdot \sigma \sqrt{\frac{Q}{V}}.$$ 

In particular, the market impact $G$ is proportional to the square root of the size $Q$ of the meta-order.
Does this relation hold true in the real world?
Empirics

- Does this relation hold true in the real world?
- Do we have to introduce more explanatory random variables (as analyzed by Kyle and Obizhaeva)?
Empirics

- Does this relation hold true in the real world?
- Do we have to introduce more explanatory random variables (as analyzed by Kyle and Obizhaeva)?

Unfortunately it is hard (if not impossible) to analyze empirically the “true” market impact $G$ of an order size $Q$.

We can hardly observe the meta-orders, however we can observe the actual orders.
We now apply the method of dimensional analysis to a different issue which has the advantage of being empirically observable:
We now apply the method of dimensional analysis to a different issue which has the advantage of being empirically observable:

\( N \): the number of trades (actual orders), measured per unit of time

\[ [N] = T^{-1}. \]
We now apply the method of dimensional analysis to a different issue which has the advantage of being empirically observable:

\( N \): the number of trades (actual orders), measured per unit of time

\[ [N] = T^{-1}. \]

What are the variables which might explain the quantity \( N \)?
The Intraday Trading Invariance Hypothesis

We now apply the method of dimensional analysis to a different issue which has the advantage of being empirically observable:

\( N \): the number of trades (actual orders), measured per unit of time

\[ [N] = T^{-1}. \]

What are the variables which might explain the quantity \( N \)? What are their dimensions?
Following Kyle and Obizhaeva (2017) and Bouchaud et al. (2016) the following quantities come into one’s mind.

- \( V \) traded volume (per day), \( [V] = ST^{-1} \)
- \( P \) price of a share, \( [P] = US^{-1} \)
- \( \sigma^2 \) squared volatility, \( [\sigma^2] = T^{-1} \)
Following Kyle and Obizhaeva (2017) and Bouchaud et al. (2016) the following quantities come into one’s mind.

- $V$ traded volume (per day), $[V] = ST^{-1}$
- $P$ price of a share, $[P] = US^{-1}$
- $\sigma^2$ squared volatility, $[\sigma^2] = T^{-1}$
- $C$ cost per trade, $[C] = U$. 
Proposition:
Assume that the number of trades $N$ depends only on the 3 quantities $\sigma^2$, $P$ and $V$, i.e.,

$$N = g(\sigma^2, P, V),$$

where the function $g : \mathbb{R}_+^3 \rightarrow \mathbb{R}_+$ is dimensionally invariant.
First attempt of explanatory variables: $P, V, \sigma^2$

Proposition:

Assume that the number of trades $N$ depends only on the 3 quantities $\sigma^2$, $P$, and $V$, i.e.,

$$N = g(\sigma^2, P, V),$$

where the function $g : \mathbb{R}_+^3 \to \mathbb{R}_+$ is dimensionally invariant.

Then, there is a constant $c > 0$ such that the number of trades $N$ obeys the relation

$$N = c \cdot \sigma^2.$$
First attempt of explanatory variables: $P, V, \sigma^2$

Proposition:
Assume that the number of trades $N$ depends only on the 3 quantities $\sigma^2, P$ and $V$, i.e.,

$$N = g(\sigma^2, P, V),$$

where the function $g : \mathbb{R}_+^3 \to \mathbb{R}_+$ is dimensionally invariant.
Then, there is a constant $c > 0$ such that the number of trades $N$ obeys the relation

$$N = c \cdot \sigma^2.$$

This relation was investigated e.g. in Jones et al. (1994).
Proposition:
Assume that the number of trades $N$ depends only on the 3 quantities $\sigma^2, P$ and $V$, i.e.,

$$N = g(\sigma^2, P, V),$$

where the function $g : \mathbb{R}_+^3 \rightarrow \mathbb{R}_+$ is dimensionally invariant. Then, there is a constant $c > 0$ such that the number of trades $N$ obeys the relation

$$N = c \cdot \sigma^2.$$ 

This relation was investigated e.g. in Jones et al. (1994).

Too simplistic!
Second attempt of explanatory variables: \( P, V, \sigma^2, C \)

\[ C: \text{cost per trade} \]
Second attempt of explanatory variables: $P, V, \sigma^2, C$

$C$: cost per trade

$C = \langle Q \rangle \cdot S = \text{average order size} \cdot \text{bid-ask spread}$
Second attempt of explanatory variables: $P, V, \sigma^2, C$

$C$: cost per trade

$C = \langle Q \rangle \cdot S = \text{average order size} \cdot \text{bid-ask spread}$

$[C] = \mathbb{U} = \text{money}$
Second attempt of explanatory variables: \( P, V, \sigma^2, C \)

Theorem [(3/2)-law] (Benzaquen, Bouchaud, Donier, 2016):

Suppose that the number of trades \( N \) depends only on the four quantities \( \sigma^2, P, V, C \) and \( N \), i.e.,

\[
N = g(\sigma^2, P, V, C),
\]
Second attempt of explanatory variables: \( P, V, \sigma^2, C \)

Theorem [(3/2)-law] (Benzaquen, Bouchaud, Donier, 2016):
Suppose that the number of trades \( N \) depends only on the four quantities \( \sigma^2, P, V, C \) and \( N \), i.e.,

\[
N = g(\sigma^2, P, V, C),
\]

where the function \( g : \mathbb{R}^4_+ \to \mathbb{R}_+ \) is dimensionally invariant.
Theorem [(3/2)-law] (Benzaquen, Bouchaud, Donier, 2016):
Suppose that the number of trades $N$ depends only on the four quantities $\sigma^2, P, V, C$ and $N$, i.e.,

$$N = g(\sigma^2, P, V, C),$$

where the function $g : \mathbb{R}_+^4 \to \mathbb{R}_+$ is *dimensionally invariant* and *leverage neutral*. 
Theorem [(3/2)-law] (Benzaquen, Bouchaud, Donier, 2016):

Suppose that the number of trades $N$ depends only on the four quantities $\sigma^2, P, V, C$ and $N$, i.e.,

$$N = g(\sigma^2, P, V, C),$$

where the function $g : \mathbb{R}^4_+ \rightarrow \mathbb{R}_+$ is dimensionally invariant and leverage neutral.

Then, there is a constant $c > 0$ such that the number of trades $N$ obeys the relation

$$N^{3/2} = c \cdot \frac{\sigma PV}{C}.$$ 

<table>
<thead>
<tr>
<th></th>
<th>$P$</th>
<th>$V$</th>
<th>$\sigma^2$</th>
<th>$C$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$U$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$T$</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>$M$</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table: A labelled overview of the dimensions of $P, V, \sigma^2, C$ and $N$. 

Empirical Results

Our empirical analysis is based on limit order book data provided by the LOBSTER database (https://lobsterdata.com). The considered sampling period begins on January 2, 2015 and ends on August 31, 2015, leaving 167 trading days. Among all NASDAQ stocks, \( d = 128 \) sufficiently liquid stocks with high market capitalizations are chosen.
Empirical Results

Our empirical analysis is based on limit order book data provided by the LOBSTER database (https://lobsterdata.com). The considered sampling period begins on January 2, 2015 and ends on August 31, 2015, leaving 167 trading days. Among all NASDAQ stocks, \( d = 128 \) sufficiently liquid stocks with high market capitalizations are chosen.

Let us fix an interval length \( T \in \{30, 60, 120, 180, 360\} \) min for which a developed hypothesis is tested.
Empirical Results

Our empirical analysis is based on limit order book data provided by the LOBSTER database (https://lobsterdata.com). The considered sampling period begins on January 2, 2015 and ends on August 31, 2015, leaving 167 trading days. Among all NASDAQ stocks, \( d = 128 \) sufficiently liquid stocks with high market capitalizations are chosen.

Let us fix an interval length \( T \in \{30, 60, 120, 180, 360\} \) min for which a developed hypothesis is tested.

\( N_j \) denotes the number of trades in the interval \( j \),

\[
Q_j = N_j^{-1} \sum_{k=1}^{N_j} Q_{t_k}
\]
denotes the average size of the trades in the interval \( j \), where \( Q_{t_k} \) denotes the number of shares traded at time \( t_k \),

\( V_j = N_j \times Q_j \) is the traded volume in the interval \( j \),

\[
P_j = N_j^{-1} \sum_{k=1}^{N_j} P_{t_k}
\]
denotes the average midquote price in the interval \( j \), where \( P_{t_k} = (A_{t_k} + B_{t_k}) / 2 \) and \( A_{t_k} \) (resp. \( B_{t_k} \)) denotes the best ask (resp. bid) price after the transaction at time \( t_k \).
\( \hat{\sigma}^2_j \) denotes the estimated squared volatility in the interval \( j \),

\[ S_j = N_j^{-1} \sum_{k=1}^{N_j} S_{t_k} \]
denotes the average bid-ask spread in the interval \( j \), where \( S_{t_k} = A_{t_k} - B_{t_k} \) is the bid-ask spread after the transaction at time \( t_k \), and

\[ C_j = Q_j \times S_j \]
is the spread cost per trade in the interval \( j \).
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left( \frac{\sigma PV}{C} \right)^{2/3}. \]
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left( \frac{\sigma PV}{C} \right)^{2/3}. \]

General form:

\[ N \sim (\sigma^2)^\beta \left( \frac{PV}{C} \right)^\gamma \]
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left( \frac{\sigma PV}{C} \right)^{2/3}. \]

General form:

\[ N \sim (\sigma^2)^{\beta} \left( \frac{PV}{C} \right)^{\gamma} \]

\( N \sim \sigma^2 \) corresponds to \( \beta = 1, \gamma = 0 \),

\( N \sim \left( \frac{\sigma PV}{C} \right)^{2/3} \) corresponds to \( \beta = 1/3, \gamma = 2/3 \),

Linear constraint:

\[ \beta + \gamma = 1. \]
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left( \frac{\sigma PV}{C} \right)^{2/3} \]

General form:

\[ N \sim (\sigma^2)^\beta \left( \frac{PV}{C} \right)^\gamma \]

\( N \sim \sigma^2 \) corresponds to \( \beta = 1, \gamma = 0 \),

\( N \sim \left( \frac{\sigma PV}{C} \right)^{2/3} \) corresponds to \( \beta = 1/3, \gamma = 2/3 \),
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left(\frac{\sigma PV}{C}\right)^{2/3}. \]

General form:

\[ N \sim (\sigma^2)^\beta \left(\frac{PV}{C}\right)^\gamma \]

\( N \sim \sigma^2 \) corresponds to \( \beta = 1, \gamma = 0 \),

\( N \sim \left(\frac{\sigma PV}{C}\right)^{2/3} \) corresponds to \( \beta = 1/3, \gamma = 2/3 \),

Linear constraint: \( \beta + \gamma = 1 \).
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left( \frac{\sigma PV}{C} \right)^{2/3}. \]

General form:

\[ N \sim (\sigma^2)^\beta \left( \frac{PV}{C} \right)^\gamma \]

\( N \sim \sigma^2 \) corresponds to \( \beta = 1, \gamma = 0 \),

\( N \sim \left( \frac{\sigma PV}{C} \right)^{2/3} \) corresponds to \( \beta = 1/3, \gamma = 2/3 \),

Linear constraint: \( \beta + \gamma = 1 \).

**Multiplicative model:**

\[ N_{ij} \sim \left( \hat{\sigma}^2_{ij} \right)^{\beta_i} \left( \frac{P_{ij}V_{ij}}{C_{ij}} \right)^{\gamma_i} \exp(\epsilon_{ij}), \]
Statistical analysis of the hypotheses

\[ N \sim \sigma^2 \quad \text{versus} \quad N \sim \left( \frac{\sigma PV}{C} \right)^{2/3}. \]

General form:

\[ N \sim (\sigma^2)^\beta \left( \frac{PV}{C} \right)^\gamma \]

\( N \sim \sigma^2 \) corresponds to \( \beta = 1, \gamma = 0, \)
\( N \sim \left( \frac{\sigma PV}{C} \right)^{2/3} \) corresponds to \( \beta = 1/3, \gamma = 2/3, \)

Linear constraint: \( \beta + \gamma = 1. \)

**Multiplicative model:**

\[ N_{ij} \sim (\hat{\sigma}_{ij}^2)^{\beta_i} \left( \frac{P_{ij}V_{ij}}{C_{ij}} \right)^{\gamma_i} \exp(\epsilon_{ij}), \]

**Linear model:**

\[ \log(N_{ij}) \sim \beta_i \log(\hat{\sigma}_{ij}^2) + \gamma_i \log \left( \frac{P_{ij}V_{ij}}{C_{ij}} \right) + \epsilon_{ij}. \]
**Figure:** The panels show kernel density estimates across the estimated parameters $\hat{\gamma}_i$ for different interval lengths $T \in \{30, 60, 120, 180, 360\}$ min.
**Figure:** The dependent variable $\log N$ is plotted versus the explanatory variables $\log \hat{\sigma}$ resp. $\log(\hat{\sigma}PV/C)$ for fixed interval $T = 60$ min and the AAL stock. The lines indicate the estimated linear relations between the considered quantities.
Figure: The dependent variable $\log N$ is plotted versus the explanatory variables $\log \hat{\sigma}$ resp. $\log(\hat{\sigma}PV/C)$ for fixed interval $T = 60$ min and the AAPL stock. The lines indicate the estimated linear relations between the considered quantities.
An Afterthought: the dimension of volatility

Definition of volatility per time $T$:

$$\sigma^2 = \text{Var}(\log(P_{t+T}) - \log(P_t)).$$
An Afterthought: the dimension of volatility

Definition of volatility per time $T$:
$$\sigma^2 = \text{Var}(\log(P_{t+T}) - \log(P_t)).$$

Example of price process (Black-Scholes):
$$dP_u = P_u(\sigma dW_u + \mu du).$$
Definition of volatility per time $T$:

$$\sigma^2 = \text{Var}(\log(P_{t+T}) - \log(P_t)).$$

Example of price process (Black-Scholes):

$$dP_u = P_u(\sigma dW_u + \mu du).$$

Obvious consequence: $[\sigma^2] = T^{-1}$. 
An Afterthought: the dimension of volatility

Definition of volatility per time $T$:

$$\sigma^2 = \text{Var}(\log(P_{t+T}) - \log(P_t)).$$

Example of price process (Black-Scholes):

$$dP_u = P_u(\sigma dW_u + \mu du).$$

Obvious consequence: $[\sigma^2] = T^{-1}$.

But what we really plug into a formula like

$$N^{3/2} = c \cdot \frac{\sigma PV}{C}$$

is an estimate $\hat{\sigma}^2$ of the “true” volatility $\sigma^2$ (whatever this is).
An Afterthought: the dimension of volatility

Definition of volatility per time $T$:

$$\sigma^2 = \text{Var}(\log(P_{t+T}) - \log(P_t)).$$

Example of price process (Black-Scholes):

$$dP_u = P_u(\sigma dW_u + \mu du).$$

Obvious consequence: $[\sigma^2] = T^{-1}$.

But what we really plug into a formula like

$$N^{3/2} = c \cdot \frac{\sigma PV}{C}$$

is an estimate $\hat{\sigma}^2$ of the “true” volatility $\sigma^2$ (whatever this is).

What does the empirical data tell us on this issue?
Typical estimator for $\sigma^2$:

$$\hat{\sigma}^2 := \sum_{k=1}^{n} \left( \log(P_{t_k}) - \log(P_{t_{k-1}}) \right)^2.$$ 

where $t = t_0 < t_1 < \ldots t_n = t + T$ are the points in $[t, t + T]$ where $P_t$ jumps.
Typical estimator for $\sigma^2$: 

$$\hat{\sigma}^2 := \sum_{k=1}^{n} \left( \log(P_{tk}) - \log(P_{tk-1}) \right)^2.$$ 

where $t = t_0 < t_1 < \ldots < t_n = t + T$ are the points in $[t, t + T]$ where $P_t$ jumps.

But we could also consider, for $H \in ]0, 1[$,

$$\hat{\sigma}^2(H) := \left( \sum_{k=1}^{N} \left| \log(P_{tk}) - \log(P_{tk-1}) \right|^{1/H} \right)^{2H}$$

Possible reasons for $H \neq \frac{1}{2}$:
Typical estimator for $\sigma^2$:

$$\hat{\sigma}^2 := \sum_{k=1}^{n} \left( \log(P_{t_k}) - \log(P_{t_{k-1}}) \right)^2.$$  

where $t = t_0 < t_1 < \ldots t_n = t + T$ are the points in $[t, t+T]$ where $P_t$ jumps.

But we could also consider, for $H \in ]0, 1[$,

$$\hat{\sigma}^2(H) := \left( \sum_{k=1}^{N} \left| \log(P_{t_k}) - \log(P_{t_{k-1}}) \right|^{1/H} \right)^{2H}$$

Possible reasons for $H \neq \frac{1}{2}$:

- fractional Brownian motion $W^H$ instead of $W$ (Mandelbrot, 1961, . . . )
Typical estimator for $\sigma^2$:

$$\hat{\sigma}^2 := \sum_{k=1}^{n} \left( \log(P_{tk}) - \log(P_{tk-1}) \right)^2.$$ 

where $t = t_0 < t_1 < \ldots t_n = t + T$ are the points in $[t, t + T]$ where $P_t$ jumps.

But we could also consider, for $H \in ]0, 1[$,

$$\hat{\sigma}^2(H) := \left( \sum_{k=1}^{N} \left| \log(P_{tk}) - \log(P_{tk-1}) \right|^{1/H} \right)^{2H}$$

Possible reasons for $H \neq \frac{1}{2}$:

- fractional Brownian motion $W^H$ instead of $W$ (Mandelbrot, 1961,\ldots)
- rough volatility (Bayer, Gatheral, Rosenbaum, \ldots)
Typical estimator for $\sigma^2$:

$$\hat{\sigma}^2 := \sum_{k=1}^{n} \left( \log(P_{t_k}) - \log(P_{t_{k-1}}) \right)^2.$$ 

where $t = t_0 < t_1 < \ldots t_n = t + T$ are the points in $[t, t + T]$ where $P_t$ jumps.

But we could also consider, for $H \in ]0, 1[$,

$$\hat{\sigma}^2(H) := \left( \sum_{k=1}^{N} \left| \log(P_{t_k}) - \log(P_{t_{k-1}}) \right|^{1/H} \right)^{2H}.$$ 

Possible reasons for $H \neq \frac{1}{2}$:

- fractional Brownian motion $W^H$ instead of $W$ (Mandelbrot, 1961, . . . )
- rough volatility (Bayer, Gatheral, Rosenbaum, . . . )
- market micro structure effects (Bouchaud, Rosenbaum, . . . )
\[ \mathbb{E}[(W_{t+T} - W_t)^2]^{1/2} = T^{1/2}, \]

but \( \mathbb{E}[(\lceil W_{t+T} \rceil - \lceil W_t \rceil)^2]^{1/2} \sim T^{1/4}, \) for \( T \rightarrow 0. \)
Theorem \([(1 + H)\text{-law}] (Pohl, Ristig, S., Tangpi, 2018)\) : Suppose that
\[
\hat{\sigma}^2(H) = T^{-2H}
\]
Theorem [(1 + H)-law] (Pohl, Ristig, S., Tangpi, 2018) : Suppose that \( [\hat{\sigma}^2(H)] = T^{-2H} \) and suppose again that the number of trades \( N \) depends \textit{only} on the four quantities \( \hat{\sigma}^2(H), P, V \) and \( C \), i.e.,

\[
N = g(\hat{\sigma}^2(H), P, V, C),
\]

where the function \( g : \mathbb{R}_+^4 \to \mathbb{R}_+ \) is \textit{dimensionally invariant} and \textit{leverage neutral}. 

\[
\begin{array}{ccc}
T & 1 & 0 \ 0 & -1 & -2H \ 0 & -1 & 0
\end{array}
\]

Table: An overview of the dimensions of quantities \( P, V, \hat{\sigma}^2(H), C \) and \( N \).
Theorem [(1 + H)-law] (Pohl, Ristig, S., Tangpi, 2018) : Suppose that
\[
\hat{\sigma}^2(H) = T^{-2H}
\] and suppose again that the number of trades \( N \) depends only on the four quantities \( \hat{\sigma}^2(H), P, V \) and \( C \), i.e.,

\[
N = g(\hat{\sigma}^2(H), P, V, C),
\]

where the function \( g : \mathbb{R}_+^4 \rightarrow \mathbb{R}_+ \) is dimensionally invariant and leverage neutral.

Then, there is a constant \( c > 0 \) such that the number of trades \( N \) obeys the relation

\[
N^{1+H} = c \cdot \frac{\hat{\sigma}(H)PV}{C}.
\]
Theorem [(1 + $H$)-law] (Pohl, Ristig, S., Tangpi, 2018): Suppose that $[\hat{\sigma}^2(H)] = T^{-2H}$ and suppose again that the number of trades $N$ depends only on the four quantities $\hat{\sigma}^2(H), P, V$ and $C$, i.e.,

$$N = g(\hat{\sigma}^2(H), P, V, C),$$

where the function $g : \mathbb{R}_+^4 \rightarrow \mathbb{R}_+$ is dimensionally invariant and leverage neutral.

Then, there is a constant $c > 0$ such that the number of trades $N$ obeys the relation

$$N^{1+H} = c \cdot \frac{\hat{\sigma}(H) PV}{C}.$$

<table>
<thead>
<tr>
<th></th>
<th>$P$</th>
<th>$V$</th>
<th>$\hat{\sigma}^2(H)$</th>
<th>$C$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S$</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$U$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$T$</td>
<td>0</td>
<td>-1</td>
<td>-2$H$</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>$\bar{M}$</td>
<td>-1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table: An overview of the dimensions of quantities $P, V, \hat{\sigma}^2(H), C$ and $N$. 
Empirical Analysis

Which estimator $\hat{\sigma}^2(H)$ gives the best fit to the empirical data?
Empirical Analysis

Which estimator $\hat{\sigma}^2(H)$ gives the best fit to the empirical data?

**Optimality Criterion:** The constant $c = c(H)$ in front of the relation

$$N_{ij}^{1+H} = c_{ij}(H)\frac{\hat{\sigma}_{ij}(H)P_{ij}V_{ij}}{C_{ij}}$$

should vary as little as possible.
Empirical Analysis

Which estimator $\hat{\sigma}^2(H)$ gives the best fit to the empirical data?

**Optimality Criterion:** The constant $c = c(H)$ in front of the relation

$$N_{ij}^{1+H} = c_{ij}(H) \frac{\hat{\sigma}_{ij}(H) P_{ij} V_{ij}}{C_{ij}}$$

should vary as little as possible.

The left panel illustrates the Gini-coefficient in dependence of $H$ for $T = 30\text{min}$ (solid), $T = 60\text{min}$ (long-dashed), $T = 120\text{min}$ (dashed), $T = 180\text{min}$ (dashed-dotted) and $T = 360\text{min}$ (dotted).
M. Benzaquen, J. Donier, and J.-P. Bouchaud.

C.M. Jones, G. Kaul and M. Lipson.

A.S. Kyle and A.A. Obizhaeva.

A.S. Kyle and A.A. Obizhaeva.

M. Pohl, A. Ristig, W. Schachermayer, and L. Tangpi.

M. Pohl, A. Ristig, W. Schachermayer, and L. Tangpi.


M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo.