Modelling Round-the-Clock Price Discovery for Cross-Listed Stocks using State Space Methods

Albert J. Menkveld  Siem-Jan Koopman  André Lucas

Vrije Universiteit Amsterdam
time line

Amsterdam

<table>
<thead>
<tr>
<th>τ</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>EST</td>
<td>4:00</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
<td>15:30</td>
</tr>
</tbody>
</table>

New York
agenda

1. literature review
2. a state space approach
3. estimation and signal extraction
4. NYSE-listed Dutch stocks
5. summary
We propose state space model to study round-the-clock price discovery for (partially) overlapping markets. It deals naturally with: (i) simultaneous quotes in overlap (ii) missing observations in non-overlap (iii) transient price changes due to “microstructure” effects

Findings:

- “NYSE Open” very informative, primarily stock-specific
- “NYSE Only” least informative and strong temporary effects
- overlap midquotes noisier for NYSE
- return persistence and noisy quotes for overlap indicate order-splitting
- results differ from “variance ratio” results
1. literature review
literature

(i) Price discovery in fragmented markets

- Methodology for measuring contribution to price discovery was developed for *simultaneous* trading (see Hasbrouck (1995), Gonzalo and Granger (1995)), Harris, McInish, and Wood (2002), J of Financial Markets 2002(3))
(i) Price discovery in fragmented markets

- Methodology for measuring contribution to price discovery was developed for simultaneous trading (see Hasbrouck (1995), Gonzalo and Granger (1995)), Harris, McInish, and Wood (2002), J of Financial Markets 2002(3))

- For non-U.S. stocks cross-listed in U.S. the evidence is
  - NYSE contributes at most 30% for Dutch, German, and Spanish stocks (Hupperets and Menkveld (2002), Grammig, Melvin, and Schlag (2001), Pascual, Pascual-Fuste, and Climent (2001))
  - U.S. prices adjust more to Canadian prices than vice versa (Eun and Sabherwal (2003))
(ii) Round-the-clock price discovery

- **Single-market** studies on 24-hour price discovery study variance ratios of open-to-close and close-to-open returns (Oldfield and Rogalski (1980), French and Roll (1986), Harvey and Huang (1991), Jones, Kaul, and Lipson (1994), and George and Hwang (2001))
(ii) Round-the-clock price discovery

- **Single-market** studies on 24-hour price discovery study variance ratios of open-to-close and close-to-open returns (Oldfield and Rogalski (1980), French and Roll (1986), Harvey and Huang (1991), Jones, Kaul, and Lipson (1994), and George and Hwang (2001))

- **Multiple-market** studies “without” overlap
  1. Regress home market overnight returns on return in foreign market(s) (Craig, Dravid, and Richardson (1995))
  2. Calculate Weighted Price Contributions (WPCs) as the foreign market return divided by overnight return in home market (Barclay and Hendershott (2003), Barclay and Warner (1993))
(iii) Issues that arise in applying existing methodologies for studying round-the-clock price discovery in multiple, partially overlapping markets

- prefer not to, *ex-ante*, choose a “home market”
- transient price changes due to microstructure effects
- missing observations in the non-overlap
- allow parameters to depend on time of day and market
- commonality in returns (Ronen (1997))
2. a state space approach
We propose a state space model, that arises naturally after

- we assume the unobserved, efficient price is characterized by a random walk
We propose a state space model, that arises naturally after

- we assume the unobserved, efficient price is characterized by a random walk

- we allow for non-trivial intraday seasonality in volatility
We propose a state space model, that arises naturally after

- we assume the unobserved, efficient price is characterized by a random walk

- we allow for non-trivial intraday seasonality in volatility

- we assume the observed midquotes equal the efficient price plus transient "microstructure" effects
We propose a state space model, that arises naturally after

- we assume the unobserved, efficient price is characterized by a random walk
- we allow for non-trivial intraday seasonality in volatility
- we assume the observed midquotes equal the efficient price plus transient “microstructure” effects

\[ \alpha_{t,\tau} = \alpha_{t,\tau-1} + \eta_{t,\tau} \]  
(state equation)
We propose a state space model, that arises naturally after

- we assume the unobserved, efficient price is characterized by a random walk
- we allow for non-trivial intraday seasonality in volatility
- we assume the observed midquotes equal the efficient price plus transient “microstructure” effects

\[
\alpha_{t,\tau} = \alpha_{t,\tau-1} + \eta_{t,\tau} \quad \text{(state equation)}
\]

\[
P_{k,t,\tau} = \alpha_{t,\tau} + \epsilon_{k,t,\tau} \quad \text{(observation equation market } k)\]
state equations

The (unobserved) efficient price process is defined as:

\[ \alpha_{t,\tau} = \alpha_{t,\tau-1} + \beta \xi_{t,\tau} + \eta_{t,\tau} \]

\[ \xi_{t,\tau} \sim N(0, \sigma^2_{\xi,\tau}) \quad \eta_{t,\tau} \sim N(\mu_{\tau}, \sigma^2_{\eta,\tau} C) \]

with \( \alpha_{t,0} = \alpha_{t-1,T} \) and \( C = \text{diag}(c_1, \ldots, c_n) \).
state equations

The (unobserved) efficient price process is defined as:

\[
\alpha_{t,T} = \alpha_{t,T-1} + \beta \xi_{t,T} + \eta_{t,T}
\]

\[
\xi_{t,T} \sim \mathcal{N}(0, \sigma_{\xi,T}^2) \quad \eta_{t,T} \sim \mathcal{N}(\mu_T, \sigma_{\eta,T}^2 C)
\]

with \(\alpha_{t,0} = \alpha_{t-1,T}\) and \(C = \text{diag}(c_1, \ldots, c_n)\).

To ensure identification of the model, we impose the parameter restrictions

\[
\frac{1}{n} \sum_{i=1}^{n} \beta_i^2 = 1 \quad \frac{1}{n} \sum_{i=1}^{n} c_i = 1
\]
state equations

The (unobserved) efficient price process is defined as:

\[
\alpha_{t,\tau} = \alpha_{t,\tau-1} + \beta \xi_{t,\tau} + \eta_{t,\tau}
\]

\[
\xi_{t,\tau} \sim N(0, \sigma^2_{\xi,\tau}) \quad \eta_{t,\tau} \sim N(\mu_\tau, \sigma^2_{\eta,\tau} C)
\]

with \( \alpha_{t,0} = \alpha_{t-1,T} \) and \( C = \text{diag}(c_1, \ldots, c_n) \).

To ensure identification of the model, we impose the parameter restrictions

\[
\frac{1}{n} \sum_{i=1}^{n} \beta_i^2 = 1 \quad \frac{1}{n} \sum_{i=1}^{n} c_i = 1
\]

Round-the-clock price discovery is then determined by

\[
\sigma^2_{E,\tau} = \sigma^2_{\xi,\tau} + \sigma^2_{\eta,\tau}
\]
The observation equation for market $k$ is defined as:

$$ p_{k,t,\tau} = \alpha_{t,\tau} + \varepsilon_{k,t,\tau} \quad \varepsilon_{k,t,\tau} \sim \mathcal{N}(0, \sigma_{\varepsilon,k,\tau}^2 \cdot I_n) $$
observation equations

The observation equation for market $k$ is defined as:

$$p_{k,t,\tau} = \alpha_{t,\tau} + \varepsilon_{k,t,\tau} \quad \varepsilon_{k,t,\tau} \sim N(0, \sigma^2_{\varepsilon,k,\tau} \cdot I_n)$$

We extend to account for potential “lagged” market response (due to e.g. strategic trading, inventory control by liquidity suppliers):

$$p_{k,t,\tau} = \alpha_{t,\tau} + \theta(\alpha_{t,\tau} - \alpha_{t,\tau-1}) + \varepsilon_{k,t,\tau}$$

$$= \alpha_{t,\tau} + \theta \beta \xi_{t,\tau} + \theta \eta_{t,\tau} + \varepsilon_{k,t,\tau}$$
The observation equation for market $k$ is defined as:

$$ p_{k,t,\tau} = \alpha_{t,\tau} + \varepsilon_{k,t,\tau} \quad \varepsilon_{k,t,\tau} \sim N(0, \sigma_{\varepsilon,k,\tau}^2 \cdot I_n) $$

We extend to account for potential “lagged” market response (due to e.g. strategic trading, inventory control by liquidity suppliers):

$$ p_{k,t,\tau} = \alpha_{t,\tau} + \theta(\alpha_{t,\tau} - \alpha_{t,\tau-1}) + \varepsilon_{k,t,\tau} $$

$$ = \alpha_{t,\tau} + \theta \beta \xi_{t,\tau} + \theta \eta_{t,\tau} + \varepsilon_{k,t,\tau} $$

And, we allow for time-of-day and innovation-specific $\theta$

$$ p_{k,t,\tau} = \alpha_{t,\tau} + \theta \xi_{t,\tau} \beta \xi_{t,\tau} + \theta \eta_{t,\tau} \eta_{t,\tau} + \varepsilon_{k,t,\tau} $$
Therefore, the model we will estimate is

\[
\alpha_{t,\tau} = \alpha_{t,\tau-1} + \beta \xi_{t,\tau} + \eta_{t,\tau}
\]

\[
p_{k, t, \tau} = \alpha_{t, \tau} + \theta \xi,\tau \beta \xi_{t,\tau} + \theta \eta,\tau \eta_{t,\tau} + \varepsilon_{k, t, \tau}
\]

with parameter vector

\[
\left[\sigma^{2'}, \theta', \sigma^{2'}, \theta', \sigma^{2'}, \sigma^{2'}, \sigma^{2'}, \beta', c'\right]'
\]
3. estimation and signal extraction
state space representation

The standard state space model is formulated for a vector of observations $y_s$ with a single time index $s$:

$$
\delta_{s+1} = T_s \delta_s + R_s \chi_s \quad \text{(state equation)}
$$

$$
y_s = Z_s \delta_s + \nu_s \quad \text{(observation equation)}
$$

for $s = 1, \ldots, M$ and disturbances $\chi_s \sim \mathcal{N}(0, Q_s)$ and $\nu_s \sim \mathcal{N}(0, H_s)$ are mutually and serially uncorrelated. The initial state vector $\delta_1 \sim \mathcal{N}(a, P)$ is uncorrelated with the disturbances. The system matrices or vectors $Z_s, T_s, R_s, H_s$ and $Q_s$, together with the initial mean $a$ and variance $P$, are assumed as fixed and known for all $s$. This general state space model is explored further in textbooks of Harvey (1989) and Durbin and Koopman (2001), amongst others.
state space representation (cont)

Our model can be represented as a state space model by choosing:

\[ y_s = (p'_{1,t,\tau}, \cdots, p'_{K,t,\tau})' \]
\[ \delta_s = (\alpha'_{t,\tau}, \eta'_{t,\tau-1}, \xi_{t,\tau-1})' \]
\[ \chi_s = (\eta'_{t,\tau}, \xi_{t,\tau})' \]
\[ s = (t - 1) \cdot T + \tau \]

The state space matrices are:

\[ Z_s = \iota_K \otimes [I_n, \theta_{\eta,\tau}I_n, \theta_{\xi,\tau}\beta] \]

\[ T_s = \begin{bmatrix} I_n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \]
\[ R_s = \begin{bmatrix} I_n & \beta \\ I_n & 0 \\ 0 & 1 \end{bmatrix} \]
\[ Q_s = \begin{bmatrix} \sigma^2_{\eta,\tau}C & 0 \\ 0 & \sigma^2_{\xi,\tau} \end{bmatrix} \]
estimation and signal extraction

The Kalman filter evaluates the conditional mean and variance of the state vector $\delta_s$ given the past observations, that is

$$a_{s|s-1} = \mathbb{E}(\delta_s | Y_{s-1}) \quad P_{s|s-1} = \text{var}(\delta_s | Y_{s-1}) \quad s = 1, \ldots, M$$

where $a_{1|Y_0} = a$ and $P_{1|Y_0} = P$. 
estimation and signal extraction

The Kalman filter evaluates the conditional mean and variance of the state vector \( \delta_s \) given the past observations, that is

\[
a_{s|s-1} = E(\delta_s|Y_{s-1}) \quad P_{s|s-1} = \text{var}(\delta_s|Y_{s-1}) \quad s = 1, \ldots, M
\]

where \( a_{1|Y_0} = a \) and \( P_{1|Y_0} = P \).

The recursive equations are given by

\[
a_{s+1|s} = T_s a_{s|s-1} + K_s v_s
\]

\[
P_{s+1|s} = T_s P_{s|s-1} T_s' + R_s Q_s R_s' - K_s F_s^{-1} K_s'
\]

with one-step ahead prediction error vector \( v_s = y_s - Z_s a_{s|s-1} \), its variance matrix \( F_s = Z_s P_{s|s-1} Z_s' + H_s \) and Kalman gain matrix \( K_s = T_s P_{s|s-1} Z_s' F_s^{-1} \) for \( s = 1, \ldots, M \).
estimation and signal extraction (cont)

- The parameters in the state space model are estimated by maximizing the loglikelihood that can be evaluated by the Kalman filter as a result of the prediction error decomposition. The loglikelihood function is given by

\[
\log L = -\frac{nKM}{2} \log 2\pi - \frac{1}{2} \sum_{s=1}^{M} \log |F_s| - \frac{1}{2} \sum_{s=1}^{M} v_s' F_s^{-1} v_s
\]

- Estimation was done in Ox (see Doornik (2001)) using the SsfPack state space routines (see Koopman, Shephard, and Doornik (1999), www.ssfpack.com).

- We use the quasi-Newton method by Broyden, Fletcher, Goldfarb, and Shanno (BFGS) for the optimization of the loglikelihood.
estimation and signal extraction (cont)

- An important feature of state space methods is their ability to deal with missing values, when all elements in $y_s$ are missing, the recursive equation reduces to

$$a_{s+1|s} = T_s a_{s|s-1} + K_s v_s$$

$$P_{s+1|s} = T_s P_{s|s-1} T_s' + R_s Q_s R_s' - K_s F_{s-1} K_s'$$

- Signal extraction refers to the estimation of the unobserved efficient price given all observations $Y_M$. The conditional mean vector

$$\hat{\delta}_s = E(\delta_s | Y_M)$$

and conditional variance matrix

$$V_s = \text{var}(\delta_s | Y_M)$$

can be computed by a smoothing algorithm.
4. NYSE-listed Dutch stocks
Dataset consists of

- NYSE TAQ data
- Euronext-Amsterdam trades and quotes
- Olsen&Associates exchange rate quotes

for

- July 1997 through June 1998
- seven Dutch stocks: Aegon, Ahold, KLM, KPN, Philips, Royal Dutch, Unilever
time line

Amsterdam

New York

Input

1 2 3 4 5 6

τ

EST

4:00 8:00 9:00 10:00 11:00 15:30
Time points far apart to ensure independence of transient effects.

Hansen and Lunde (2004, p. 2): “In fact, our empirical analysis shows that the noise process has a time-dependence that persists for up to about two minutes.”
the model

To recap, the model we will estimate is

\[
\begin{align*}
\alpha_{t,\tau} &= \alpha_{t,\tau-1} + \beta \xi_{t,\tau} + \eta_{t,\tau} \\
p_{k,t,\tau} &= \alpha_{t,\tau} + \theta_\xi \beta_\xi t_{,\tau} + \theta_\eta \eta t_{,\tau} + \varepsilon_{k,t,\tau}
\end{align*}
\]

with parameter vector

\[
\begin{bmatrix}
\sigma'_{\xi}, \theta'_{\xi}, \sigma'_{\eta}, \theta'_{\eta}, \sigma'_{\epsilon,A}, \sigma'_{\epsilon, NY}, \beta', c'
\end{bmatrix}'
\]

with sample length 220*6=1320, state vector dimension 7+7+1=15, observation vector dimension 2*7=14, number of parameters 6+6+6+6+4+3+7+7=45
<table>
<thead>
<tr>
<th>Time Period</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:00–8:00</td>
<td>AMS Only</td>
</tr>
<tr>
<td>8:00–9:00</td>
<td>NY PreOpen</td>
</tr>
<tr>
<td>9:00–10:00</td>
<td>NY Open</td>
</tr>
<tr>
<td>10:00–11:00</td>
<td>AMS Close</td>
</tr>
<tr>
<td>11:00–15:30</td>
<td>NY Only</td>
</tr>
<tr>
<td>15:30–4:00</td>
<td>Overnight</td>
</tr>
</tbody>
</table>

Intraday Variance Pattern (Hourly, 1/10000)

AMS Only

NY PreOpen

NY Open

AMS Close

NY Only

Overnight
Intraday Variance Pattern (Hourly, 1/10000)

- 4:00–8:00: AMS Only
- 8:00–9:00: NY PreOpen
- 9:00–10:00: NY Open
- 10:00–11:00: AMS Close
- 11:00–15:30: NY Only
- 15:30–4:00: Overnight

Intraday Variance Pattern As Estimated Using State Space Model (Hourly, 1/10000)
Intraday Variance Pattern (Hourly, 1/10000)

AMS Only

8:00−9:00
NY PreOpen

9:00−10:00
NY Open

10:00−11:00
AMS Close

11:00−15:30
NY Only

15:30−4:00
Overnight

Intraday Variance Pattern As Estimated Using State Space Model (Hourly, 1/10000)
### state innovations

<table>
<thead>
<tr>
<th>Event</th>
<th>Start</th>
<th>NY</th>
<th>NY</th>
<th>AMS</th>
<th>NY</th>
<th>Over-night</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMS PreOpen</td>
<td>4:00</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
<td>15:30</td>
</tr>
<tr>
<td>Close Only</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
<td>15:30</td>
<td>4:00</td>
</tr>
</tbody>
</table>

| $\sigma^2_{\xi,\tau}$ | 0.13  | 0.18     | 0.16     | 0.15      | 0.05     | 0.05       |
|                       | (0.01)| (0.02)   | (0.02)   | (0.02)    | (0.01)   | (0.01)     |
| $\theta_{\xi,\tau}$  | -0.35 |          | -0.16    |           |          |            |
|                       | (0.04)|          | (0.04)   |           |          |            |

| $\sigma^2_{\eta,\tau}$ | 0.22  | 0.15     | 0.27     | 0.23      | 0.05     | 0.05       |
|                         | (0.01)| (0.01)   | (0.01)   | (0.02)    | (0.01)   | (0.00)     |
| $\theta_{\eta,\tau}$  | -0.34 | -0.30    |          | 0.87      |          |            |
|                         | (0.02)| (0.08)   |          | (0.13)    |          |            |
## Measurement Errors

<table>
<thead>
<tr>
<th></th>
<th>4:00</th>
<th>8:00</th>
<th>9:00</th>
<th>10:00</th>
<th>11:00</th>
<th>15:30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\left( \sigma_{\tau}^{\epsilon,A} \right)^2$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\left( \sigma_{\tau}^{\epsilon,NY} \right)^2$</td>
<td></td>
<td></td>
<td></td>
<td>0.11</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
efficient price estimates

Royal Dutch: Full Sample 7/1/97–6/30/98

Log Midquote AMS
Log Efficient Price (Estimate)
Log Midquote NY

### correlation common factor and index

<table>
<thead>
<tr>
<th>Event</th>
<th>Start</th>
<th>NY</th>
<th>NY</th>
<th>AMS</th>
<th>NY</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMS</td>
<td></td>
<td>PreOpen</td>
<td>Open</td>
<td>Close</td>
<td>Only</td>
</tr>
<tr>
<td>Start (EST)</td>
<td>4:00</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
</tr>
<tr>
<td>End</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
<td>15:30</td>
</tr>
</tbody>
</table>

\( \rho(\text{Common Factor, AEX}) \)

\( \rho(\text{Common Factor, S&P500}) \)
<table>
<thead>
<tr>
<th>Event</th>
<th>Start</th>
<th>NY</th>
<th>NY</th>
<th>AMS</th>
<th>NY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AMS</td>
<td>PreOpen</td>
<td>Open</td>
<td>Close</td>
<td>Only</td>
</tr>
<tr>
<td>Start (EST)</td>
<td>4:00</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
</tr>
<tr>
<td>End</td>
<td>8:00</td>
<td>9:00</td>
<td>10:00</td>
<td>11:00</td>
<td>15:30</td>
</tr>
<tr>
<td>$\rho$(Common Factor, AEX)</td>
<td>0.57</td>
<td>0.38</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$(Common Factor, S&amp;P500)</td>
<td></td>
<td></td>
<td></td>
<td>0.21</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
</tbody>
</table>
robustness

We performed several robustness checks:

- sub-periods
- stock-specific measurement error variances
- correlated price innovations and measurement errors (see, e.g, George and Hwang (2001))

Results available from corresponding author’s website
5. summary
summary

We propose state space model to study round-the-clock price discovery for (partially) overlapping markets. It deals naturally with: (i) simultaneous quotes in overlap (ii) missing observations in non-overlap (iii) transient price changes due to “microstructure” effects

Findings:

- “NYSE Open” very informative, primarily stock-specific
- “NYSE Only” least informative and strong temporary effects
- overlap midquotes noisier for NYSE
- return persistence and noisy quotes for overlap indicate order-splitting
- results differ from “variance ratio” results
### intraday return autocorrelations

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Event</th>
<th>Lag 1</th>
<th>Lag 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:00-8:00</td>
<td>AMS Only</td>
<td>-0.077</td>
<td></td>
</tr>
<tr>
<td>8:00-9:00</td>
<td>NY PreOpen</td>
<td>0.056</td>
<td>-0.020</td>
</tr>
<tr>
<td>9:00-10:00</td>
<td>NY Open</td>
<td>-0.125*</td>
<td>-0.005</td>
</tr>
<tr>
<td>10:00-11:00</td>
<td>AMS Close</td>
<td>0.251*</td>
<td>-0.170*</td>
</tr>
<tr>
<td>11:00-15:30</td>
<td>NY Only</td>
<td>-0.050</td>
<td>0.039</td>
</tr>
<tr>
<td>15:30-4:00(+1)</td>
<td>Overnight</td>
<td>-0.165*</td>
<td>-0.022</td>
</tr>
</tbody>
</table>

*: Significant at a 95% confidence level.
References


