Modeling and Forecasting Intraday Volatility with Unobserved Component Structures

Gabriele Fiorentini* and Christian Macaro**

* University of Florence
** University of Roma Tre and NYU

Preliminary and Incomplete

December 5, 2007
Motivation

- The availability of UHF data for prices of financial assets has fostered a bulk of research that makes use of the continuously recorded data to deal with different issues.

- A non-trivial portion of this research focused on the modeling and forecasting of intraday volatility.

- Forecasts of intraday volatility should assist trading desk operators and market makers in placing limit orders and rebalancing portfolios.
Motivation

- The availability of UHF data for prices of financial assets has fostered a bulk of research that makes use of the continuously recorded data to deal with different issues.

- A non trivial portion of this research focused on the modeling and forecasting of intraday volatility

- Forecasts of intraday volatility should assist trading desk operators and market makers in placing limit orders and rebalancing portfolios
Motivation

- The availability of UHF data for prices of financial assets has fostered a bulk of research that makes use of the continuously recorded data to deal with different issues.

- A non trivial portion of this research focused on the modeling and forecasting of intraday volatility.

- Forecasts of intraday volatility should assist trading desk operators and market makers in placing limit orders and rebalancing portfolios.
Motivation

- It has long been recognized that conventional GARCH models are unsatisfactory for modeling returns sampled at intraday frequencies mainly because they fail to control for the pronounced intraday (seasonal) patterns of volatility.

- Alternative models have been suggested. In this project we build on the previous literature and propose a new model for intraday volatility that largely borrows from existing proposals.
Motivation

- It has long been recognized that conventional GARCH models are unsatisfactory for modeling returns sampled at intraday frequencies mainly because they fail to control for the pronounced intraday (seasonal) patterns of volatility.

- Alternative models have been suggested. In this project we build on the previous literature and propose a new model for intraday volatility that largely borrows from existing proposals.
The salient features of our model are

1. The volatility of high frequency returns is assumed to decompose into the product of stochastic components.
2. The volatility of high frequency returns is estimated by using higher frequency tick by tick data.
3. The information brought by weakly exogenous variables is easily incorporated.
4. Focus is on stocks but extensions to other financial assets is straightforward.
5. Implementation is simple and the model is apt to be applied to large data sets.
The salient features of our model are

1. The volatility of high frequency returns is assumed to decompose into the product of stochastic components.

2. The volatility of high frequency returns is estimated by using higher frequency tick by tick data.

3. The information brought by weakly exogenous variables is easily incorporated.

4. Focus is on stocks but extensions to other financial assets is straightforward.

5. Implementation is simple and the model is apt to be applied to large data sets.
The salient features of our model are

1. The volatility of high frequency returns is assumed to decompose into the product of stochastic components.
2. The volatility of high frequency returns is estimated by using higher frequency tick by tick data.
3. The information brought by weakly exogenous variables is easily incorporated.
4. Focus is on stocks but extensions to other financial assets is straightforward.
5. Implementation is simple and the model is apt to be applied to large data sets.
The salient features of our model are

1. The volatility of high frequency returns is assumed to decompose into the product of stochastic components.
2. The volatility of high frequency returns is estimated by using higher frequency tick by tick data.
3. The information brought by weakly exogenous variables is easily incorporated.
4. Focus is on stocks but extensions to other financial assets is straightforward.
5. Implementation is simple and the model is apt to be applied to large data sets.
The salient features of our model are

1. The volatility of high frequency returns is assumed to decompose into the product of stochastic components.
2. The volatility of high frequency returns is estimated by using higher frequency tick by tick data.
3. The information brought by weakly exogenous variables is easily incorporated.
4. Focus is on stocks but extensions to other financial assets is straightforward.
5. Implementation is simple and the model is apt to be applied to large data sets.
Proposal

The salient features of our model are

1. The volatility of high frequency returns is assumed to decompose into the product of stochastic components.
2. The volatility of high frequency returns is estimated by using higher frequency tick by tick data.
3. The information brought by weakly exogenous variables is easily incorporated.
4. Focus is on stocks but extensions to other financial assets is straightforward.
5. Implementation is simple and the model is apt to be applied to large data sets.
Related Literature

- Andersen and Bollerslev 1997 JoEF, 1997 JoF, 1998 JoF
- Ghose and Kroner 1996
- Taylor and Xu 1997 JoEF
- Andersen, Bollerslev and Das 2001 JoF
- Giot 2005 EJoF
- Engle, Sokalska and Chanda 2006
- Deo, Hurvich and Lu 2006 JoE
- Dacorogna, Gencay, Muller, Olsen and Pictet 2001
- Beltratti and Morana 2001 EN

- Mian and Adam 2001 AFE
- Martens, Chang and Taylor 2002 JoFR
- Worthington 2003
- Aradhyula and Ergun 2004 AFE
- McMillan and Speight 2004 AFE
- Wongswan 2006 RoFS
Related Literature

- Andersen and Bollerslev 1997 JoEF, 1997 JoF, 1998 JoF
- Ghose and Kroner 1996
- Taylor and Xu 1997 JoEF
- Andersen, Bollerslev and Das 2001 JoF
- Giot 2005 EJoF
- Engle, Sokalska and Chanda 2006
- Deo, Hurvich and Lu 2006 JoE
- Dacorogna, Gencay, Muller, Olsen and Pictet 2001
- Beltratti and Morana 2001 EN

- Mian and Adam 2001 AFE
- Martens, Chang and Taylor 2002 JoFR
- Worthington 2003
- Aradhyula and Ergun 2004 AFE
- McMillan and Speight 2004 AFE
- Wongswan 2006 RoFS
Notation

Let \( t = 1, \ldots, T \) denote days in the sample.
Each day is divided into \( M \) intervals.

Return in interval \( j \) of day \( t \) is defined as
\[
R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)
\]

E.g. for NYSE stocks

- if \( M = 12 \) \( \Rightarrow \) \( R_{t,j} \) are 30-minute returns
- if \( M = 36 \) \( \Rightarrow \) \( R_{t,j} \) are 10-minute returns

Overnight returns are dropped and the total number of observations is \( T \times N \).
Notation

Let $t = 1, \ldots, T$ denote days in the sample. Each day is divided into $M$ intervals. Return in interval $j$ of day $t$ is defined as:

$$R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)$$

E.g. for NYSE stocks:
- if $M = 12 \Rightarrow R_{t,j}$ are 30-minute returns
- if $M = 36 \Rightarrow R_{t,j}$ are 10-minute returns

Overnight returns are dropped and the total number of observations is $T \times N$
Notation

Let $t = 1, \ldots, T$ denote days in the sample.
Each day is divided into $M$ intervals.

Return in interval $j$ of day $t$ is defined as

$$R_{t,j} = \log \text{Prc}_{t,j} - \log \text{Prc}_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)$$

E.g. for NYSE stocks
- if $M = 12 \Rightarrow R_{t,j}$ are 30-minute returns
- if $M = 36 \Rightarrow R_{t,j}$ are 10-minute returns

Overnight returns are dropped and the total number of observations is $T \times N$
Notation

Let $t = 1, \ldots, T$ denote days in the sample
Each day is divided into $M$ intervals

Return in interval $j$ of day $t$ is defined as

$$R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)$$

E.g. for NYSE stocks

- if $M = 12 \Rightarrow R_{t,j}$ are 30-minute returns
- if $M = 36 \Rightarrow R_{t,j}$ are 10-minute returns

Overnight returns are dropped and the total number of observations is $T \times N$
Let $t = 1, \ldots, T$ denote days in the sample. Each day is divided into $M$ intervals. Return in interval $j$ of day $t$ is defined as $R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1}$ $(t = 1, \ldots, T)$ $(j = 1, \ldots, M)$

E.g. for NYSE stocks
- if $M = 12 \Rightarrow R_{t,j}$ are 30-minute returns
- if $M = 36 \Rightarrow R_{t,j}$ are 10-minute returns

Overnight returns are dropped and the total number of observations is $T \times N$
Let $t = 1, \ldots, T$ denote days in the sample.
Each day is divided into $M$ intervals.

Return in interval $j$ of day $t$ is defined as

$$R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)$$

E.g. for NYSE stocks

- if $M = 12 \Rightarrow R_{t,j}$ are 30-minute returns
- if $M = 36 \Rightarrow R_{t,j}$ are 10-minute returns

Overnight returns are dropped and the total number of observations is $T \times N$
Notation

Let \( t = 1, \ldots, T \) denote days in the sample
Each day is divided into \( M \) intervals

Return in interval \( j \) of day \( t \) is defined as

\[
R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)
\]

E.g. for NYSE stocks

- if \( M = 12 \Rightarrow R_{t,j} \) are 30-minute returns
- if \( M = 36 \Rightarrow R_{t,j} \) are 10-minute returns

Overnight returns are dropped and the total number of observations is \( T \times N \)
Let $t = 1, \ldots, T$ denote days in the sample. Each day is divided into $M$ intervals. Return in interval $j$ of day $t$ is defined as

$$R_{t,j} = \log Prc_{t,j} - \log Prc_{t,j-1} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M)$$

E.g. for NYSE stocks
- if $M = 12 \Rightarrow R_{t,j}$ are 30-minute returns
- if $M = 36 \Rightarrow R_{t,j}$ are 10-minute returns

Overnight returns are dropped and the total number of observations is $T \times N$
We assume that \( R_{t,j} = \sigma_{t,j} Z_{t,j} \) \( Z_{t,j} \) i.i.d. \( \sim (0, 1) \)

We further assume that \( \sigma_{t,j} \) can be decomposed into the product of components. Specifically

\[
\sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j}
\]

where

- \( P \) is the persistent component of volatility
- \( S \) is the intraday (seasonal) periodic factor
- \( U \) is the short-term irregular intraday factor
We assume that \( R_{t,j} = \sigma_{t,j} Z_{t,j} \quad Z_{t,j} \text{ i.i.d. } \sim (0, 1) \)

We further assume that \( \sigma_{t,j} \) can be decomposed into the product of components. Specifically

\[
\sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j}
\]

where

- \( P \) is the persistent component of volatility
- \( S \) is the intraday (seasonal) periodic factor
- \( U \) is the short-term irregular intraday factor
Model

We assume that \[ R_{t,j} = \sigma_{t,j} Z_{t,j} \quad Z_{t,j} \text{ i.i.d.} \sim (0, 1) \]

We further assume that \( \sigma_{t,j} \) can be decomposed into the product of components. Specifically

\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

where

- \( P \) is the persistent component of volatility
- \( S \) is the intraday (seasonal) periodic factor
- \( U \) is the short-term irregular intraday factor
We assume that $R_{t,j} = \sigma_{t,j} Z_{t,j}$ where $Z_{t,j}$ is i.i.d. $\sim (0, 1)$.

We further assume that $\sigma_{t,j}$ can be decomposed into the product of components. Specifically

$$\sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j}$$

where

- $P$ is the persistent component of volatility
- $S$ is the intraday (seasonal) periodic factor
- $U$ is the short-term irregular intraday factor
Model

We assume that $R_{t,j} = \sigma_{t,j} Z_{t,j}$, $Z_{t,j}$ i.i.d. $\sim (0, 1)$

We further assume that $\sigma_{t,j}$ can be decomposed into the product of components. Specifically

$$\sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j}$$

where

- $P$ is the persistent component of volatility
- $S$ is the intraday (seasonal) periodic factor
- $U$ is the short-term irregular intraday factor
Model

We assume that \( R_{t,j} = \sigma_{t,j} Z_{t,j} \) \( Z_{t,j} \) i.i.d. \( \sim (0, 1) \)

We further assume that \( \sigma_{t,j} \) can be decomposed into the product of components. Specifically

\[
\sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j}
\]

where

- \( P \) is the persistent component of volatility
- \( S \) is the intraday (seasonal) periodic factor
- \( U \) is the short-term irregular intraday factor
Remarks

\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

- Notice that each component is indexed by \( t \) and \( j \)
- Many popular models for high frequency volatility can be written in this unobserved component framework
- In A&B(97,98,...) and ESC(06) \( P_{t,j} \) depends only on \( t \).
  \[ P_{t,j}^2 = \frac{\sigma_t^2}{M}, \text{ where } \sigma_t^2 \text{ is daily volatility (e.g. from a GARCH model)} \]
- In G&K(96) and ESC(06) \( S_{t,j}^2 = S_j^2 = \frac{1}{T} \sum_{t=1}^{T} R_{t,j}^* \)
  where \( R_{t,j}^* = R_{t,j}/\sigma_t \)
- In the FFF approach of A&B \( S_{t,j} \) may either depend only on \( j \) or also on \( t \) via the interaction with the daily component.
- In T&X(97) \( S_{t,j} \) depends on \( j \) and also on \( t \) but only for the \textit{day of the week} effect
Remarks

\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

- Notice that each component is indexed by \( t \) and \( j \)

- Many popular models for high frequency volatility can be written in this unobserved component framework

- In A&B(97,98,...) and ESC(06) \( P_{t,j} \) depends only on \( t \).
  \[ P_{t,j}^2 = \frac{\sigma_t^2}{M} \], where \( \sigma_t^2 \) is daily volatility (e.g. from a GARCH model)

- In G&K(96) and ESC(06) \( S_{t,j}^2 = S_j^2 = \frac{1}{T} \sum_{t=1}^{T} R_{t,j}^* \),
  where \( R_{t,j}^* = R_{t,j}/\sigma_t \)

- In the FFF approach of A&B \( S_{t,j} \) may either depend only on \( j \) or also on \( t \) via the interaction with the daily component.

- In T&X(97) \( S_{t,j} \) depends on \( j \) and also on \( t \) but only for the day of the week effect.
Remarks

\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

- Notice that each component is indexed by \( t \) and \( j \)

- Many popular models for high frequency volatility can be written in this unobserved component framework

- In A&B(97,98,...) and ESC(06) \( P_{t,j} \) depends only on \( t \).

\[ P_{t,j}^2 = \frac{\sigma_t^2}{M}, \text{ where } \sigma_t^2 \text{ is daily volatility (e.g. from a GARCH model)} \]

- In G&K(96) and ESC(06) \( S_{t,j}^2 = S_j^2 = \frac{1}{T} \sum_{t=1}^{T} R_{t,j}^* \).

\[ \text{where } R_{t,j}^* = R_{t,j}/\sigma_t \]

- In the FFF approach of A&B \( S_{t,j} \) may either depend only on \( j \) or also on \( t \) via the interaction with the daily component.

- In T&X(97) \( S_{t,j} \) depends on \( j \) and also on \( t \) but only for the day of the week effect
Remarks

\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

- Notice that each component is indexed by \( t \) and \( j \)

- Many popular models for high frequency volatility can be written in this unobserved component framework

- In A&B(97,98,...) and ESC(06) \( P_{t,j} \) depends only on \( t \).
  \[ P_{t,j}^2 = \frac{\sigma_t^2}{M}, \text{ where } \sigma_t^2 \text{ is daily volatility (e.g. from a GARCH model)} \]

- In G&K(96) and ESC(06) \( S_{t,j}^2 = S_j^2 = \frac{1}{T} \sum_{t=1}^{T} R_{t,j}^{*2} \).
  where \( R_{t,j}^{*} = R_{t,j}/\sigma_t \)

- In the FFF approach of A&B \( S_{t,j} \) may either depend only on \( j \) or also on \( t \) via the interaction with the daily component.

- In T&X(97) \( S_{t,j} \) depends on \( j \) and also on \( t \) but only for the day of the week effect.
Remarks

\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

- Notice that each component is indexed by \( t \) and \( j \)

- Many popular models for high frequency volatility can be written in this unobserved component framework

- In A&B(97,98,...) and ESC(06) \( P_{t,j} \) depends only on \( t \).
  \[ P_{t,j}^2 = \frac{\sigma_t^2}{M}, \] where \( \sigma_t^2 \) is daily volatility (e.g. from a GARCH model)

- In G&K(96) and ESC(06) \( S_{t,j}^2 = S_j^2 = \frac{1}{T} \sum_{t=1}^{T} R_{t,j}^{*2} \).
  where \( R_{t,j}^{*} = R_{t,j}/\sigma_t \)

- In the FFF approach of A&B \( S_{t,j} \) may either depend only on \( j \) or also on \( t \) via the interaction with the daily component.

- In T&X(97) \( S_{t,j} \) depends on \( j \) and also on \( t \) but only for the day of the week effect
\[ \sigma_{t,j} = P_{t,j} S_{t,j} U_{t,j} \]

- Notice that each component is indexed by \( t \) and \( j \)
- Many popular models for high frequency volatility can be written in this unobserved component framework
- In A&B(97,98,...) and ESC(06) \( P_{t,j} \) depends only on \( t \).
  \[ P_{t,j}^2 = \frac{\sigma_t^2}{M} \], where \( \sigma_t^2 \) is daily volatility (e.g. from a GARCH model)
- In G&K(96) and ESC(06) \( S_{t,j}^2 = S_j^2 = \frac{1}{T} \sum_{t=1}^{T} R_{t,j}^* \),
  where \( R_{t,j}^* = R_{t,j}/\sigma_t \)
- In the FFF approach of A&B \( S_{t,j} \) may either depend only on \( j \) or also on \( t \) via the interaction with the daily component.
- In T&X(97) \( S_{t,j} \) depends on \( j \) and also on \( t \) but only for the day of the week effect
We assume that $\log \sigma_{t,j}$ and its components follow some stochastic linear processes (augmented with regressors).

The model can then be written in two related ways.

1. Arima Model Based (AMB) decomposition: Specify a general REG-AR(F)IMA model for $\log \sigma_{t,j}$ and derive compatible models for the components.

2. Structural Time Series (STS) model à la Harvey: Use a State-Space formulation in which the models for the components are specified directly.

The two approaches are equivalent and differ only for the estimation (filtering-smoothing) algorithm (W-K vs. Kalman) and, sometimes, for the identification assumptions.
We assume that log $\sigma_{t,j}$ and its components follow some stochastic linear processes (augmented with regressors).

The model can then be written in two related ways.

1. **Arima Model Based (AMB) decomposition**: Specify a general REG-AR(F)IMA model for log $\sigma_{t,j}$ and derive compatible models for the components.

2. **Structural Time Series (STS) model à la Harvey**: Use a State-Space formulation in which the models for the components are specified directly.

The two approaches are equivalent and differ only for the estimation (filtering-smoothing) algorithm (W-K vs. Kalman) and, sometimes, for the identification assumptions.
Specifications

We assume that $\log \sigma_{t,j}$ and its components follow some stochastic linear processes (augmented with regressors).

The model can then be written in two related ways.

1. **Arima Model Based (AMB) decomposition**: Specify a general REG-AR(F)IMA model for $\log \sigma_{t,j}$ and derive compatible models for the components.

2. **Structural Time Series (STS) model à la Harvey**: Use a State-Space formulation in which the models for the components are specified directly.

The two approaches are equivalent and differ only for the estimation (filtering-smoothing) algorithm (W-K vs. Kalman) and, sometimes, for the identification assumptions.
We assume that $\log \sigma_{t,j}$ and its components follow some stochastic linear processes (augmented with regressors).

The model can then be written in two related ways.

1. **Arima Model Based (AMB) decomposition**: Specify a general REG-AR(F)IMA model for $\log \sigma_{t,j}$ and derive compatible models for the components.

2. **Structural Time Series (STS) model à la Harvey**: Use a State-Space formulation in which the models for the components are specified directly.

The two approaches are equivalent and differ only for the estimation (filtering-smoothing) algorithm (W-K vs. Kalman) and, sometimes, for the identification assumptions.
Specification

We assume that $\log \sigma_{t,j}$ and its components follow some stochastic linear processes (augmented with regressors).

The model can then be written in two related ways.

1. Arima Model Based (AMB) decomposition: Specify a general REG-AR(F)IMA model for $\log \sigma_{t,j}$ and derive compatible models for the components.

2. Structural Time Series (STS) model à la Harvey: Use a State-Space formulation in which the models for the components are specified directly.

The two approaches are equivalent and differ only for the estimation (filtering-smoothing) algorithm (W-K vs. Kalman) and, sometimes, for the identification assumptions.
We assume that $\log \sigma_{t,j}$ and its components follow some stochastic linear processes (augmented with regressors).

The model can then be written in two related ways.

1. **Arima Model Based (AMB) decomposition**: Specify a general REG-AR(F)IMA model for $\log \sigma_{t,j}$ and derive compatible models for the components.

2. **Structural Time Series (STS) model à la Harvey**: Use a State-Space formulation in which the models for the components are specified directly.

The two approaches are equivalent and differ only for the estimation (filtering-smoothing) algorithm (W-K vs. Kalman) and, sometimes, for the identification assumptions.
AMB Specification

\[
\log \sigma_{t,j} = p_{t,j} + s_{t,j} + u_{t,j} = \log \sigma^*_{t,j} + f(X_{t,j}, \beta)
\]

\[
\log \sigma^*_{t,j} = p^*_{t,j} + s^*_{t,j} + u^*_{t,j}
\]

\[
\Phi(L) \left[ \log \sigma_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) \log \sigma^*_{t,j} = \Theta(L) a_{t,j}
\]

\[
\Phi_p(L)p^*_{t,j} = \Theta_p(L)a_P t,j; \quad \Phi_s(L)s^*_{t,j} = \Theta_s(L)a_S t,j; \quad \Phi_u(L)u^*_{t,j} = \Theta_u(L)a_U t,j;
\]

where \( f(X_{t,j}, \beta) \) is a known regression function of weakly exogenous variables \( X_{t,j} \) and unknown coefficients \( \beta \).

\( \Phi_\cdot(L) \) and \( \Theta_\cdot(L) \) are polynomials in the lag operator \( L \) with

\[
\Phi(L) = \Phi_p(L)\Phi_s(L)\Phi_u(L)
\]

\[
\Theta(L)a_t = \Theta_p(L)\Phi_s(L)\Phi_u(L)a_P t,j + \Theta_s(L)\Phi_p(L)\Phi_u(L)a_S t,j + \Theta_u(L)\Phi_p(L)\Phi_s(L)a_U t,j
\]
AMB Specification

\[
\log \sigma_{t, j} = p_{t, j} + s_{t, j} + u_{t, j} = \log \sigma^*_{t, j} + f(X_{t, j}, \beta)
\]

\[
\log \sigma^*_{t, j} = p^*_{t, j} + s^*_{t, j} + u^*_{t, j}
\]

\[
\Phi(L) \left[ \log \sigma_{t, j} - f(X_{t, j}, \beta) \right] = \Phi(L) \log \sigma^*_{t, j} = \Theta(L) a_{t, j}
\]

\[
\Phi_p(L)p^*_{t, j} = \Theta_p(L) a_{P t, j}; \quad \Phi_s(L)s^*_{t, j} = \Theta_s(L) a_{S t, j}; \quad \Phi_u(L)u^*_{t, j} = \Theta_u(L) a_{U t, j};
\]

where \( f(X_{t, j}, \beta) \) is a known regression function of weakly exogenous variables \( X_{t, j} \) and unknown coefficients \( \beta \).

\( \Phi_\bullet(L) \) and \( \Theta_\bullet(L) \) are polynomials in the lag operator \( L \) with

\[
\Phi(L) = \Phi_p(L)\Phi_s(L)\Phi_u(L)
\]

\[
\Theta(L) a_t = \Theta_p(L)\Phi_s(L)\Phi_u(L) a_{P t, j} + \Theta_s(L)\Phi_p(L)\Phi_u(L) a_{S t, j} + \Theta_u(L)\Phi_p(L)\Phi_s(L) a_{U t, j}
\]
AMB Specification

\[
\log \sigma_{t,j} = p_{t,j} + s_{t,j} + u_{t,j} = \log \sigma^*_{t,j} + f(X_{t,j}, \beta)
\]

\[
\log \sigma^*_{t,j} = p^*_{t,j} + s^*_{t,j} + u^*_{t,j}
\]

\[
\Phi(L) \left[ \log \sigma_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) \log \sigma^*_{t,j} = \Theta(L) a_{t,j}
\]

\[
\Phi_p(L)p^*_{t,j} = \Theta_p(L)a_{P_{t,j}}; \quad \Phi_s(L)s^*_{t,j} = \Theta_s(L)a_{S_{t,j}}; \quad \Phi_u(L)u^*_{t,j} = \Theta_u(L)a_{U_{t,j}};
\]

where \( f(X_{t,j}, \beta) \) is a known regression function of weakly exogenous variables \( X_{t,j} \) and unknown coefficients \( \beta \).

\( \Phi_\bullet(L) \) and \( \Theta_\bullet(L) \) are polynomials in the lag operator \( L \) with

\[
\Phi(L) = \Phi_p(L)\Phi_s(L)\Phi_u(L)
\]

\[
\Theta(L)a_t = \Theta_p(L)\Phi_s(L)\Phi_u(L)a_{P_{t,j}} + \Theta_s(L)\Phi_p(L)\Phi_u(L)a_{S_{t,j}} + \Theta_u(L)\Phi_p(L)\Phi_s(L)a_{U_{t,j}}
\]
AMB Specification

\[
\log \sigma_{t,j} = p_{t,j} + s_{t,j} + u_{t,j} = \log \sigma^*_{t,j} + f(X_{t,j}, \beta)
\]

\[
\log \sigma^*_{t,j} = p^*_{t,j} + s^*_{t,j} + u^*_{t,j}
\]

\[
\Phi(L) \left[ \log \sigma_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) \log \sigma^*_{t,j} = \Theta(L) a_{t,j}
\]

\[
\Phi_p(L)p^*_{t,j} = \Theta_p(L) a_{P\ t,j}; \quad \Phi_s(L)s^*_{t,j} = \Theta_s(L) a_{S\ t,j}; \quad \Phi_u(L)u^*_{t,j} = \Theta_u(L) a_{U\ t,j};
\]

where \( f(X_{t,j}, \beta) \) is a known regression function of weakly exogenous variables \( X_{t,j} \) and unknown coefficients \( \beta \).

\( \Phi_\bullet(L) \) and \( \Theta_\bullet(L) \) are polynomials in the lag operator \( L \) with

\[
\Phi(L) = \Phi_p(L)\Phi_s(L)\Phi_u(L)
\]

\[
\Theta(L)a_t = \Theta_p(L)\Phi_s(L)\Phi_u(L)a_{P\ t,j} + \Theta_s(L)\Phi_p(L)\Phi_u(L)a_{S\ t,j} + \Theta_u(L)\Phi_p(L)\Phi_s(L)a_{U\ t,j}
\]
AMB Specification

\[ \log \sigma_{t,j} = p_{t,j} + s_{t,j} + u_{t,j} = \log \sigma^*_{t,j} + f(X_{t,j}, \beta) \]

\[ \log \sigma^*_{t,j} = p^*_{t,j} + s^*_{t,j} + u^*_{t,j} \]

\[ \Phi(L) \left[ \log \sigma_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) \log \sigma^*_{t,j} = \Theta(L) a_{t,j} \]

\[ \Phi_p(L)p^*_{t,j} = \Theta_p(L)a_P t,j; \quad \Phi_s(L)s^*_{t,j} = \Theta_s(L)a_S t,j; \quad \Phi_u(L)u^*_{t,j} = \Theta_u(L)a_U t,j; \]

where \( f(X_{t,j}, \beta) \) is a known regression function of weakly exogenous variables \( X_{t,j} \) and unknown coefficients \( \beta \).

\( \Phi_\ast(L) \) and \( \Theta_\ast(L) \) are polynomials in the lag operator \( L \) with

\[ \Phi(L) = \Phi_p(L)\Phi_s(L)\Phi_u(L) \]

\[ \Theta(L)a_t = \Theta_p(L)\Phi_s(L)\Phi_u(L)a_P t,j + \Theta_s(L)\Phi_p(L)\Phi_u(L)a_S t,j + \Theta_u(L)\Phi_p(L)\Phi_s(L)a_U t,j \]
AMB Specification

\[
\log \sigma_{t,j} = p_{t,j} + s_{t,j} + u_{t,j} = \log \sigma_{t,j}^* + f(X_{t,j}, \beta)
\]

\[
\log \sigma_{t,j}^* = p_{t,j}^* + s_{t,j}^* + u_{t,j}^*
\]

\[
\Phi(L) \left[ \log \sigma_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) \log \sigma_{t,j}^* = \Theta(L) a_{t,j}
\]

\[
\Phi_p(L)p_{t,j}^* = \Theta_p(L)a_{P t,j}; \quad \Phi_s(L)s_{t,j}^* = \Theta_s(L)a_{S t,j}; \quad \Phi_u(L)u_{t,j}^* = \Theta_u(L)a_{U t,j};
\]

where \(f(X_{t,j}, \beta)\) is a known regression function of weakly exogenous variables \(X_{t,j}\) and unknown coefficients \(\beta\).

\(\Phi_\bullet(L)\) and \(\Theta_\bullet(L)\) are polynomials in the lag operator \(L\) with

\[
\Phi(L) = \Phi_p(L)\Phi_s(L)\Phi_u(L)
\]

\[
\Theta(L)a_t = \Theta_p(L)\Phi_s(L)\Phi_u(L)a_{P t,j} + \Theta_s(L)\Phi_p(L)\Phi_u(L)a_{S t,j} + \Theta_u(L)\Phi_p(L)\Phi_s(L)a_{U t,j}
\]
AMB Identification

Consider the following set of sufficient identification assumptions

1. Factorization of $\Phi(L)$ into $\Phi_p(L)\Phi_s(L)\Phi_u(L)$ is straightforward and is obtained through a simple root allocation.

2. Orthogonality of components $p^* \perp s^* \perp u^*$

3. Canonical decomposition: The model for the components are balanced and $\text{Var}(a_{P t,j})$ and $\text{Var}(a_{S t,j})$ are chosen as small as possible. In this way $p^*$ and $s^*$ are termed canonical components.

Remark: For any other admissible choice of, say, $\text{Var}(a_{S t,j})$ the corresponding periodic component can be written as the sum of the canonical one plus an uncorrelated white noise. (Hillmer and Tiao 1982)
Consider the following set of sufficient identification assumptions

1. Factorization of $\Phi(L)$ into $\Phi_p(L)\Phi_s(L)\Phi_u(L)$ is straightforward and is obtained through a simple root allocation.

2. Orthogonality of components $p^* \perp s^* \perp u^*$

3. Canonical decomposition: The model for the components are balanced and $\text{Var}(a_{P,t,j})$ and $\text{Var}(a_{S,t,j})$ are chosen as small as possible. In this way $p^*$ and $s^*$ are termed canonical components.

Remark: For any other admissible choice of, say, $\text{Var}(a_{S,t,j})$ the corresponding periodic component can be written as the sum of the canonical one plus an uncorrelated white noise. (Hillmer and Tiao 1982)
AMB Identification

Consider the following set of sufficient identification assumptions

1. **Factorization of** $\Phi(L)$ **into** $\Phi_p(L)\Phi_s(L)\Phi_u(L)$ **is straightforward and is obtained through a simple root allocation.**

2. **Orthogonality of components** $p^* \perp s^* \perp u^*$

3. **Canonical decomposition**: The model for the components are *balanced* and $\text{Var}(a_{P_{t,j}})$ and $\text{Var}(a_{S_{t,j}})$ are chosen as small as possible. In this way $p^*$ and $s^*$ are termed canonical components.

**Remark**: For any other admissible choice of, say, $\text{Var}(a_{S_{t,j}})$ the corresponding periodic component can be written as the sum of the canonical one plus an uncorrelated white noise. (Hillmeier and Tiao 1982)
AMB Identification

Consider the following set of sufficient identification assumptions

1. Factorization of $\Phi(L)$ into $\Phi_p(L)\Phi_s(L)\Phi_u(L)$ is straightforward and is obtained through a simple root allocation.

2. Orthogonality of components $p^* \perp s^* \perp u^*$

3. Canonical decomposition: The model for the components are balanced and $\text{Var}(a_{Pt,j})$ and $\text{Var}(a_{St,j})$ are chosen as small as possible. In this way $p^*$ and $s^*$ are termed canonical components.

Remark: For any other admissible choice of, say, $\text{Var}(a_{St,j})$ the corresponding periodic component can be written as the sum of the canonical one plus an uncorrelated white noise. (Hillmer and Tiao 1982)
Example: Consider the following typical model for quarterly data

\[(1 - L)(1 - L^4)Y_t = (1 + \theta_1 L)(1 + \theta_4 L^4)a_t\]

\[Y_t = \text{Trend} + \text{Seasonal} + \text{Noise}\]

- **TREND:** \((1 - 2L + L^2)p_t = (1 + \theta_p 1L + \theta_p 2L^2)a_{p,t}\)

- **SEASONAL:** \((1 + L + L^2 + L^3)s_t = (1 + \theta_s 1L + \theta_s 2L^2 + \theta_s 3L^3)a_{s,t}\)
Example: Consider the following typical model for quarterly data

\[ (1 - L)(1 - L^4)Y_t = (1 + \theta_1 L)(1 + \theta_4 L^4)a_t \]

\[ Y_t = \text{Trend} + \text{Seasonal} + \text{Noise} \]
AMB Example: Quarterly Airline Model

Example: Consider the following typical model for quarterly data

\[(1 - L)(1 - L^4)Y_t = (1 + \theta_1 L)(1 + \theta_4 L^4)a_t\]

\[Y_t = \text{Trend} + \text{Seasonal} + \text{Noise}\]

- TREND: \((1 - 2L + L^2)\rho_t = (1 + \theta_p 1L + \theta_p 2L^2)a_p t\)

- SEASONAL: \((1 + L + L^2 + L^3)s_t = (1 + \theta_s 1L + \theta_s 2L^2 + \theta_s 3L^3)a_s t\)
Example: Consider the following typical model for quarterly data

\[(1 - L)(1 - L^4)Y_t = (1 + \theta_1 L)(1 + \theta_4 L^4)a_t\]

\[Y_t = \text{Trend} + \text{Seasonal} + \text{Noise}\]

- **TREND:** \[(1 - 2L + L^2)p_t = (1 + \theta_p 1L + \theta_p 2L^2)a_p t\]

- **SEASONAL:** \[(1 + L + L^2 + L^3)s_t = (1 + \theta_s 1L + \theta_s 2L^2 + \theta_s 3L^3)a_s t\]
AMB Example: Quarterly Airline Model

\[ \theta_1 = -0.1 \quad \theta_4 = -0.3 \]
AMB Example: Quarterly Airline Model
Suppose that each one of the \( M \) intervals is further partitioned into \( N \) sub-intervals.

Return in sub-interval \( i \) of bin \( j \) of day \( t \) is defined as

\[
R_{t,j,i} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M), \quad (i = 1, \ldots, N)
\]

E.g. for NYSE

- if \( M = 12 \) and \( N = 30 \) \( \Rightarrow \) \( R_{t,j,i} \) are 1-minute returns
- if \( M = 36 \) and \( N = 20 \) \( \Rightarrow \) \( R_{t,j,i} \) are half-minute returns

We can, thus, define an estimator of \( \sigma_{t,j}^2 \) based on some realized measure that uses \( N \) observations.
Estimation of $\sigma^2_{t,j}$

- Suppose that each one of the $M$ intervals is further partitioned into $N$ sub-intervals.
- Return in sub-interval $i$ of bin $j$ of day $t$ is defined as

  $$R_{t,j,i} \ (t = 1, \ldots, T) \ (j = 1, \ldots, M), \ (i = 1, \ldots, N)$$

- E.g. for NYSE
  - if $M = 12$ and $N = 30$ ⇒ $R_{t,j,i}$ are 1-minute returns
  - if $M = 36$ and $N = 20$ ⇒ $R_{t,j,i}$ are half-minute returns

- We can, thus, define an estimator of $\sigma^2_{t,j}$ based on some realized measure that uses $N$ observations.
Estimation of $\sigma^2_{t,j}$

- Suppose that each one of the $M$ intervals is further partitioned into $N$ sub-intervals.
- Return in sub-interval $i$ of bin $j$ of day $t$ is defined as $R_{t,j,i} (t = 1, \ldots, T) \quad (j = 1, \ldots, M), \quad (i = 1, \ldots, N)$

- E.g. for NYSE
  - if $M = 12$ and $N = 30 \Rightarrow R_{t,j,i}$ are 1-minute returns
  - if $M = 36$ and $N = 20 \Rightarrow R_{t,j,i}$ are half-minute returns

- We can, thus, define an estimator of $\sigma^2_{t,j}$ based on some realized measure that uses $N$ observations.
Estimation of $\sigma^2_{t,j}$

- Suppose that each one of the $M$ intervals is further partitioned into $N$ sub-intervals.
- Return in sub-interval $i$ of bin $j$ of day $t$ is defined as $R_{t,j,i}$ ($t = 1, \ldots, T$) ($j = 1, \ldots, M$), ($i = 1, \ldots, N$)

- E.g. for NYSE
  - if $M = 12$ and $N = 30 \Rightarrow R_{t,j,i}$ are 1-minute returns
  - if $M = 36$ and $N = 20 \Rightarrow R_{t,j,i}$ are half-minute returns

- We can, thus, define an estimator of $\sigma^2_{t,j}$ based on some realized measure that uses $N$ observations.
Suppose that each one of the $M$ intervals is further partitioned into $N$ sub-intervals.

Return in sub-interval $i$ of bin $j$ of day $t$ is defined as

$$R_{t,j,i} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M), \quad (i = 1, \ldots, N)$$

E.g. for NYSE

if $M = 12$ and $N = 30 \Rightarrow R_{t,j,i}$ are 1-minute returns

if $M = 36$ and $N = 20 \Rightarrow R_{t,j,i}$ are half-minute returns

We can, thus, define an estimator of $\sigma_{t,j}^2$ based on some realized measure that uses $N$ observations.
Suppose that each one of the $M$ intervals is further partitioned into $N$ sub-intervals.

Return in sub-interval $i$ of bin $j$ of day $t$ is defined as

$$R_{t,j,i} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M), \quad (i = 1, \ldots, N)$$

E.g. for NYSE

- if $M = 12$ and $N = 30 \Rightarrow R_{t,j,i}$ are 1-minute returns
- if $M = 36$ and $N = 20 \Rightarrow R_{t,j,i}$ are half-minute returns

We can, thus, define an estimator of $\sigma^2_{t,j}$ based on some realized measure that uses $N$ observations.
Suppose that each one of the $M$ intervals is further partitioned into $N$ sub-intervals.

Return in sub-interval $i$ of bin $j$ of day $t$ is defined as

$$R_{t,j,i} \quad (t = 1, \ldots, T) \quad (j = 1, \ldots, M), \quad (i = 1, \ldots, N)$$

E.g. for NYSE

if $M = 12$ and $N = 30 \Rightarrow R_{t,j,i}$ are 1-minute returns

if $M = 36$ and $N = 20 \Rightarrow R_{t,j,i}$ are half-minute returns

We can, thus, define an estimator of $\sigma^2_{t,j}$ based on some realized measure that uses $N$ observations.
Estimation of $\sigma_{t,j}^2$

- For example to keep matters simple consider \emph{realized volatility}

\[ \hat{\sigma}_{t,j}^2 = \sum_{i=1}^{N} R_{t,j,i}^2 \]

- In principle we could take $M$ as large as desired and let $N \to \infty$

- In practice there are limits to the benefits attainable from UHF data and $M \times N$ cannot be too large
Estimation of $\sigma^2_{t,j}$

- For example to keep matters simple consider realized volatility

$$\hat{\sigma}^2_{t,j} = \sum_{i=1}^{N} R^2_{t,j,i}$$

- In principle we could take $M$ as large as desired and let $N \rightarrow \infty$

- In practice there are limits to the benefits attainable from UHF data and $M \times N$ cannot be too large
Estimation of $\sigma^2_{t,j}$

- For example to keep matters simple consider realized volatility

$$\hat{\sigma}^2_{t,j} = \sum_{i=1}^{N} R^2_{t,j,i}$$

- In principle we could take $M$ as large as desired and let $N \rightarrow \infty$

- In practice there are limits to the benefits attainable from UHF data and $M \times N$ cannot be too large
REGARIMA specification and estimation

Let $lr_{t,j} = \frac{1}{2} \log \hat{\sigma}_{t,j}^2$. We specify and estimate the model

$$\Phi(L) \left[ lr_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L)lr_{t,j}^* = \Theta(L)a_{t,j}$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with distributed effects
4) Detected outliers AO and TC
5) …

Model is identified with BIC and estimated by PML
Let $lrv_{t,j} = \frac{1}{2} \log \hat{\sigma}^2_{t,j}$. We specify and estimate the model

$$\Phi(L) \left[ lrv_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L)lrv^*_{t,j} = \Theta(L)a_{t,j}$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with distributed effects
4) Detected outliers AO and TC
5) …

Model is identified with BIC and estimated by PML
REGARIMA specification and estimation

Let $lrv_{t,j} = \frac{1}{2} \log \hat{\sigma}^2_{t,j}$. We specify and estimate the model

$$
\Phi(L) \left[ lrv_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L)lrv^*_{t,j} = \Theta(L)a_{t,j}
$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with distributed effects
4) Detected outliers AO and TC
5) …

Model is identified with BIC and estimated by PML
REGARIMA specification and estimation

Let $lr_{t,j} = \frac{1}{2} \log \hat{\sigma}_{t,j}^2$. We specify and estimate the model

$$\Phi(L) \left[ lr_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L)lr_{t,j}^* = \Theta(L)a_{t,j}$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with *distributed* effects
4) Detected outliers AO and TC
5) …

Model is identified with BIC and estimated by PML
REGARIMA specification and estimation

Let $lrv_{t,j} = \frac{1}{2} \log \hat{\sigma}_{t,j}^2$. We specify and estimate the model

$$
\Phi(L) \left[ lrv_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) lrv_{t,j}^* = \Theta(L) a_{t,j}
$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with distributed effects
4) Detected outliers AO and TC
5) ...

Model is identified with BIC and estimated by PML
REGARIMA specification and estimation

Let $lr_{t,j} = \frac{1}{2} \log \hat{\sigma}_{t,j}^2$. We specify and estimate the model

$$
\Phi(L) \left[ lr_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L)lr_{t,j}^* = \Theta(L)a_{t,j}
$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with distributed effects
4) Detected outliers AO and TC
5) ...

Model is identified with BIC and estimated by PML
REGARIMA specification and estimation

Let $lrv_{t,j} = \frac{1}{2} \log \hat{\sigma}_{t,j}^2$. We specify and estimate the model

$$\Phi(L) \left[ lrv_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L)lrv^*_t = \Theta(L)a_{t,j}$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with distributed effects
4) Detected outliers AO and TC
5) …

Model is identified with BIC and estimated by PML
REGARIMA specification and estimation

Let $lrv_{t,j} = \frac{1}{2} \log \hat{\sigma}^2_{t,j}$. We specify and estimate the model

$$\Phi(L) \left[ lrv_{t,j} - f(X_{t,j}, \beta) \right] = \Phi(L) lrv^*_t = \Theta(L) a_{t,j}$$

Regression variables $X_{t,j}$ may include:

1) Macro and sector/firm specific announcement variables
2) Day of the week dummies
3) (A non-negative function of) Overnight returns
   M regressors or 1 regressor with *distributed* effects
4) Detected outliers AO and TC
5) ...

Model is identified with BIC and estimated by PML
Next, we derive the (canonical) models for the components that can be estimated by linear projections with the Wiener-Kolmogorov symmetric filter

$$\hat{p}_{t,j}^* = WK_p(L) \ lrv_{t,j}^{*(E)}$$

where $lrv_{t,j}^{*(E)}$ is the series extended with forecasts and backcasts

$$WK_p(L) = \frac{\Theta_p(L)\Theta_p(L^{-1})}{\phi_p(L)\phi_p(L^{-1})} \frac{\Phi(L)\Phi(L^{-1})}{\Theta(L)\Theta(L^{-1})}$$

The frequency domain representation of the filter is given by the Ratio of Spectra. Thus the squared gain will always be less or equal than one and proportional to the relative power of the component spectrum.
Next, we derive the (canonical) models for the components that can be estimated by linear projections with the Wiener-Kolmogorov symmetric filter

$$\hat{p}_{t,j} = WK_p(L) \ lrv_{t,j}^{(E)}$$

where $lrv_{t,j}^{(E)}$ is the series extended with forecasts and backcasts

$$WK_p(L) = \frac{\Theta_p(L)\Theta_p(L^{-1})}{\Phi_p(L)\Phi_p(L^{-1})} \frac{\Phi(L)\Phi(L^{-1})}{\Theta(L)\Theta(L^{-1})}$$

The frequency domain representation of the filter is given by the Ratio of Spectra. Thus the squared gain will always be less or equal than one and proportional to the relative power of the component spectrum.
Estimation of Components

The estimated **regression effects** are assigned to the components according to their properties.

Forecasts of the components are also readily obtained.

Finally the multiplicative factors $\hat{P}_{t,j}$, $\hat{S}_{t,j}$ and $\hat{U}_{t,j}$ are computed.

**Remark:** When the intraday periodic component of volatility is pronounced and the “seasonal” factors are large, a bias correction is needed to correct the underestimation of the “seasonally adjusted” volatility caused by the fact that geometric means underestimate arithmetic means.
Estimation of Components

The estimated regression effects are assigned to the components according to their properties.

Forecasts of the components are also readily obtained.

Finally the multiplicative factors $\hat{P}_{t,j}$, $\hat{S}_{t,j}$ and $\hat{U}_{t,j}$ are computed.

Remark: When the intraday periodic component of volatility is pronounced and the “seasonal” factors are large, a bias correction is needed to correct the underestimation of the “seasonally adjusted” volatility caused by the fact that geometric means underestimate arithmetic means.
Estimation of Components

The estimated regression effects are assigned to the components according to their properties.

Forecasts of the components are also readily obtained.

Finally the multiplicative factors $\hat{P}_{t,j}$, $\hat{S}_{t,j}$ and $\hat{U}_{t,j}$ are computed.

Remark: When the intraday periodic component of volatility is pronounced and the “seasonal” factors are large, a bias correction is needed to correct the underestimation of the “seasonally adjusted” volatility caused by the fact that geometric means underestimate arithmetic means.
Estimation of Components

The estimated regression effects are assigned to the components according to their properties.

Forecasts of the components are also readily obtained.

Finally the multiplicative factors $\hat{P}_{t,j}$, $\hat{S}_{t,j}$ and $\hat{U}_{t,j}$ are computed.

Remark: When the intraday periodic component of volatility is pronounced and the “seasonal” factors are large, a bias correction is needed to correct the underestimation of the “seasonally adjusted” volatility caused by the fact that geometric means underestimate arithmetic means.
We consider alternative forecasting methods

1) Forecast directly $lrv_{t,j}$ with the REG-ARIMA model

2) Use the AMB decomposition to compute forecasts of the intraday periodic factor $S_{t,j}$ and of the persistent component $P_{t,j}$. Next, apply a GARCH type model to the standardized returns $R_{t,j}/(P_{t,j}S_{t,j})$.

3) Same as method 2) but the forecast of the persistent component is computed from:
   A) GARCH type model on daily data
   B) Forecasts of daily realized volatility
We consider alternative forecasting methods

1) Forecast directly $lrv_{t,j}$ with the REG-ARIMA model

2) Use the AMB decomposition to compute forecasts of the intraday periodic factor $S_{t,j}$ and of the persistent component $P_{t,j}$. Next, apply a GARCH type model to the standardized returns $R_{t,j}/(P_{t,j}S_{t,j})$.

3) Same as method 2) but the forecast of the persistent component is computed from:
   A) GARCH type model on daily data
   B) Forecasts of daily realized volatility
We consider alternative forecasting methods

1) Forecast directly $lrv_{t,j}$ with the REG-ARIMA model

2) Use the AMB decomposition to compute forecasts of the intraday periodic factor $S_{t,j}$ and of the persistent component $P_{t,j}$. Next, apply a GARCH type model to the standardized returns $R_{t,j}/(P_{t,j}S_{t,j})$.

3) Same as method 2) but the forecast of the persistent component is computed from:
   A) GARCH type model on daily data
   B) Forecasts of daily realized volatility
We consider alternative forecasting methods

1) Forecast directly $lrv_{t,j}$ with the REG-ARIMA model

2) Use the AMB decomposition to compute forecasts of the intraday periodic factor $S_{t,j}$ and of the persistent component $P_{t,j}$. Next, apply a GARCH type model to the standardized returns $R_{t,j}/(P_{t,j}S_{t,j})$.

3) Same as method 2) but the forecast of the persistent component is computed from:
   A) GARCH type model on daily data
   B) Forecasts of daily realized volatility
We consider alternative forecasting methods

1) Forecast directly $lrv_{t,j}$ with the REG-ARIMA model

2) Use the AMB decomposition to compute forecasts of the intraday periodic factor $S_{t,j}$ and of the persistent component $P_{t,j}$. Next, apply a GARCH type model to the standardized returns $R_{t,j}/(P_{t,j}S_{t,j})$.

3) Same as method 2) but the forecast of the persistent component is computed from:
   A) GARCH type model on daily data
   B) Forecasts of daily realized volatility
Method 3) is analogous to Engle Sokalska and Chanda (2006) (ESC). Differences are confined to the estimation and forecasting of the intraday periodic (diurnal) factors.

An FFF based alternative to Method 3) can also be formulated ESC and the FFF based variation are used as benchmarks against which we test the forecasting performance of our model.
Forecasting Comparisons

Method 3) is analogous to Engle Sokalska and Chanda (2006) (ESC). Differences are confined to the estimation and forecasting of the intraday periodic (diurnal) factors.

An FFF based alternative to Method 3) can also be formulated

ESC and the FFF based variation are used as benchmarks against which we test the forecasting performance of our model.
Method 3) is analogous to Engle Sokalska and Chanda (2006) (ESC). Differences are confined to the estimation and forecasting of the intraday periodic (diurnal) factors.

An FFF based alternative to Method 3) can also be formulated. ESC and the FFF based variation are used as benchmarks against which we test the forecasting performance of our model.
NYSE *quotes* for *Boeing*, *Exxon*, *General Electric* and IBM are obtained from the TAQ database.

Data spans a 16-month period in January 2006 - April 2007. Monday July 3rd and Friday November 24th are deleted from the sample.

Each day we remove the first and last 15-minute period and consider $M = 12$ and $N = 30$.

$R_{t,j,i}$ are then 1-minute returns on day $t$ between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as *size weighted bid/ask averages* (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
NYSE *quotes* for *Boeing*, *Exxon*, *General Electric* and *IBM* are obtained from the TAQ database.


Monday July 3rd and Friday November 24th are deleted from the sample.

Each day we remove the first and last 15-minute period and consider $M = 12$ and $N = 30$.

$R_{t,j,i}$ are then 1-minute returns on day $t$ between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as *size weighted bid/ask averages* (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
### Data

**NYSE quotes** for **BOEING**, **EXXON**, **GENERAL ELECTRIC** and **IBM** are obtained from the TAQ database.


Monday July 3\(^{rd}\) and Friday November 24\(^{th}\) are deleted from the sample.

Each day we remove the first and last 15-minute period and consider \(M = 12\) and \(N = 30\).

\(R_{t,j,i}\) are then 1-minute returns on day \(t\) between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as size weighted bid/ask averages (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
NYSE *quotes* for *Boeing*, *Exxon*, *General Electric* and IBM are obtained from the TAQ database.

Data spans a 16-month period in January 2006 - April 2007. Monday July 3rd and Friday November 24th are deleted from the sample.

Each day we remove the first and last 15-minute period and consider $M = 12$ and $N = 30$.

$R_{t,j,i}$ are then 1-minute returns on day $t$ between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as *size weighted bid/ask averages* (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
Data

NYSE *quotes* for **Boeing, Exxon, General Electric** and IBM are obtained from the TAQ database.

Data spans a 16-month period in January 2006 - April 2007. Monday July 3rd and Friday November 24th are deleted from the sample.

Each day we remove the first and last 15-minute period and consider $M = 12$ and $N = 30$.

$R_{t,j,i}$ are then 1-minute returns on day $t$ between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as *size weighted bid/ask averages* (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
Data

NYSE *quotes* for *Boeing*, *Exxon*, *General Electric* and *IBM* are obtained from the TAQ database.

Data spans a 16-month period in January 2006 - April 2007. Monday July 3rd and Friday November 24th are deleted from the sample.

Each day we remove the first and last 15-minute period and consider $M = 12$ and $N = 30$.

$R_{t,j,i}$ are then 1-minute returns on day $t$ between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as *size weighted bid/ask averages* (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
Data

NYSE quotes for **BOEING**, **EXXON**, **GENERAL ELECTRIC** and **IBM** are obtained from the TAQ database.

Data spans a 16-month period in January 2006 - April 2007. Monday July 3rd and Friday November 24th are deleted from the sample.

Each day we remove the first and last 15-minute period and consider $M = 12$ and $N = 30$.

$R_{t,j,i}$ are then 1-minute returns on day $t$ between 9:45 and 15:45. We aim at modeling and forecasting 30-minute volatility.

Prices are defined as size weighted bid/ask averages (other definitions have been considered).

We use four months for the “historical” analysis and retain twelve months for the rolling forecasting exercise.
The total sample length is 3948 observations on 30-minute returns (329 days).

In the period January 2006 - April 2006 there are 972 observations (81 days).

In the out-of-sample forecasting exercise we maintain the sample length equal to 972 and we evaluate 2976 $k$-periods ahead forecasts ($k = 1, 2, 3$).

Daily data on prices (since January 2003) are downloaded from http://finance.yahoo.com/

Overnight returns are computed as $\log(Open_t) - \log(Close_{t-1})$. Open prices are adjusted for dividends using an adjustment factor calculated from closing prices.
The total sample length is 3948 observations on 30-minute returns (329 days).

In the period January 2006 - April 2006 there are 972 observations (81 days)

In the out-of-sample forecasting exercise we maintain the sample length equal to 972 and we evaluate 2976 $k$-periods ahead forecasts ($k = 1, 2, 3$).

Daily data on prices (since January 2003) are downloaded from http://finance.yahoo.com/

Overnight returns are computed as $\log(Open_t) - \log(Close_{t-1})$. Open prices are adjusted for dividends using an adjustment factor calculated from closing prices.
The total sample length is 3948 observations on 30-minute returns (329 days).

In the period January 2006 - April 2006 there are 972 observations (81 days)

In the out-of-sample forecasting exercise we maintain the sample length equal to 972 and we evaluate 2976 $k$-periods ahead forecasts ($k = 1, 2, 3$).

Daily data on prices (since January 2003) are downloaded from http://finance.yahoo.com/

Overnight returns are computed as $\log(Open_t) - \log(Close_{t-1})$. Open prices are adjusted for dividends using an adjustment factor calculated from closing prices.
Data

The total sample length is 3948 observations on 30-minute returns (329 days).

In the period January 2006 - April 2006 there are 972 observations (81 days)

In the out-of-sample forecasting exercise we maintain the sample length equal to 972 and we evaluate 2976 $k$-periods ahead forecasts ($k = 1, 2, 3$).

Daily data on prices (since January 2003) are downloaded from http://finance.yahoo.com/

Overnight returns are computed as $\log(\text{Open}_t) - \log(\text{Close}_{t-1})$. Open prices are adjusted for dividends using an adjustment factor calculated from closing prices.
The total sample length is 3948 observations on 30-minute returns (329 days).

In the period January 2006 - April 2006 there are 972 observations (81 days)

In the out-of-sample forecasting exercise we maintain the sample length equal to 972 and we evaluate 2976 $k$-periods ahead forecasts ($k = 1, 2, 3$).


Overnight returns are computed as $\log(Open_t) - \log(Close_{t-1})$. Open prices are adjusted for dividends using an adjustment factor calculated from closing prices.
30-minute returns

(Boeing, Exxon, General Electric, IBM)

(January-April 2006)
30-minute realized volatilities (January-April 2006)
ACF plots    BOEING    (January-April 2006)

Squared Returns

Realized Volatilities

Log of Squared Returns

Log of Realized Volatilities
ACF plots GENERAL ELECTRIC (January-April 2006)

Squared Returns

Realized Volatilities

Log of Squared Returns

Log of Realized Volatilities
ACF plots          IBM          (January-April 2006)

Squared Returns

Realized Volatilities

Log of Squared Returns

Log of Realized Volatilities
ACF plots     BOEING     (January 2006 -April 2007) Full Sample

Squared Returns

Realized Volatilities

Log of Squared Returns

Log of Realized Volatilities
Summary: ARIMA \((p, d, q) \times (bp, bd, bq)\)

### ORDERS OF ESTIMATED MODELS

<table>
<thead>
<tr>
<th></th>
<th>NO REG.</th>
<th>DoW</th>
<th>DoW + OR</th>
<th>DoW + 12OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOEING</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
</tr>
<tr>
<td>EXXON</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((2, 0, 1) \times (0, 1, 1))</td>
<td>((2, 0, 1) \times (0, 1, 1))</td>
<td>((3, 0, 1) \times (0, 1, 1))</td>
</tr>
<tr>
<td>G.E.</td>
<td>((2, 0, 1) \times (0, 1, 1))</td>
<td>((2, 0, 1) \times (0, 1, 1))</td>
<td>((2, 0, 1) \times (0, 1, 1))</td>
<td>((2, 0, 1) \times (0, 1, 1))</td>
</tr>
<tr>
<td>IBM</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((1, 0, 1) \times (0, 1, 1))</td>
<td>((0, 1, 3) \times (0, 1, 1))</td>
</tr>
</tbody>
</table>

### DETECTED OUTLIERS

<table>
<thead>
<tr>
<th></th>
<th>NO REG.</th>
<th>DoW</th>
<th>DoW + OR</th>
<th>DoW + 12OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOEING</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>EXXON</td>
<td>3 AO + 2 TC</td>
<td>2 AO + 2 TC</td>
<td>2 AO + 2 TC</td>
<td>2 AO + 2 TC</td>
</tr>
<tr>
<td>G.E.</td>
<td>5 AO + 2 TC</td>
<td>6 AO + 2 TC</td>
<td>9 AO + 3 TC</td>
<td>6 AO + 1 TC</td>
</tr>
<tr>
<td>IBM</td>
<td>1 TC</td>
<td>—</td>
<td>—</td>
<td>1 AO + 4 TC</td>
</tr>
</tbody>
</table>
$\sqrt{RV}$, Long-Term and $\sqrt{RV_{DAILY}}$  

EXXON (January-April 2006)

WITHOUT REGRESSORS

DAYS

DAY OF THE WEEK

DAYS

DAYS

DAY OF THE WEEK and SINGLE OVERNIGHT RETURNS

DAY OF THE WEEK and MULTIPLE OVERNIGHT RETURNS

DAYS

DAYS
$\sqrt{RV}$, Long-Term and $\sqrt{RV_{DAILY}}$  
IBM (January-April 2006)
INTRADAY PERIODIC FACTORS  BOEING  (January-April 2006)

WITHOUT REGRESSORS

DAY OF THE WEEK

DAY OF THE WEEK and SINGLE OVERNIGHT RETURNS

DAY OF THE WEEK and MULTIPLE OVERNIGHT RETURNS
INTRADAY PERIODIC FACTORS  GENERAL ELECTRIC  (January-April 2006)

WITHOUT REGRESSORS

DAYS

DAY OF THE WEEK

DAYS

DAY OF THE WEEK and SINGLE OVERNIGHT RETURNS

DAYS

DAY OF THE WEEK and MULTIPLE OVERNIGHT RETURNS

DAYS
INTRADAY PERIODIC FACTORS COMPARISON IBM (January-April 2006)

WITHOUT REGRESSORS

DAY OF THE WEEK

AMB ESC FFF (no int.)

DAY OF THE WEEK and SINGLE OVERNIGHT RETURNS

DAY OF THE WEEK and MULTIPLE OVERNIGHT RETURNS
FORECASTING