Outline

The HAR model:


HAR extensions:

- F. Corsi, R. Renó (2009) “HAR volatility modelling with heterogeneous leverage and jumps”
- F. Audrino, F. Corsi “Modeling tick-by-tick realized correlations” (CSDA)

Application to Option Pricing:

- F. Corsi, N. Fusari, D. La Vecchia ”Realizing Smiles: Option pricing using realized volatility”
Stylized Facts I: Heteroskedasticity - Fat Tail

- USD/CHF 1989–2001
- Gaussian noise
Stylized Facts II: Volatility Persistence

Sample autocorrelation coefficients

- S&P500
- Brownian Motion
- USD/CHF

Confidence intervals are also shown for each series.
In standard GARCH and SV models volatility shocks decay with exponential rate:

\[ \rho_h \sim \gamma^h \quad \text{with} \quad 0 < \gamma < 1 \]

while empirical data show an hyperbolic decay rate:

\[ \rho_h \sim h^\gamma \quad \text{with} \quad 0 < \gamma < 1 \]

\[ \downarrow \]

**Fractional Integration**: generalize the usual differencing of \( I(1) \) series \( y_t \) to get an \( I(0) \) \( \epsilon_t \) as:

\[ (1 - L)^d y_t = \epsilon_t \]

\( I(d) \) gives an infinite MA representation for \( y_t = \sum_{k=0}^{\infty} a_k(d) \epsilon_{t-k} \) with \( a_k(d) = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \)

which displays long memory with \( \gamma = 2d - 1 \)

- Fractional Integration + ARMA \( \rightarrow \) ARFIMA
- Fractional Integration + GARCH \( \rightarrow \) FIGARCH
In std GARCH and SV models volatility shocks decay with exponential rate:

\[ \rho_h \sim \gamma^h \quad \text{with} \quad 0 < \gamma < 1 \]

while empirical data show an hyperbolic decay rate:

\[ \rho_h \sim h^{\gamma} \quad \text{with} \quad 0 < \gamma < 1 \]

**Fractional Integration**: generalize the usual differencing of \( I(1) \) series \( y_t \) to get an \( I(0) \) \( \epsilon_t \) as:

\[ (1 - L)^d y_t = \epsilon_t \]

\( I(d) \) gives an infinite MA representation for \( y_t = \sum_{k=0}^{\infty} a_k(d)\epsilon_{t-k} \) with

\[ a_k(d) = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \]

which displays long memory with \( \gamma = 2d - 1 \)

**Fractional Integration + ARMA** \( \longrightarrow \) **ARFIMA**

**Fractional Integration + GARCH** \( \longrightarrow \) **FIGARCH**
Self-Similar, Fractal or Scaling Process:

\[(Y_{t1}, Y_{t2}, Y_{t3}, \ldots) \overset{d}{=} c^{-H}(Y_{ct1}, Y_{ct2}, Y_{ct3}, \ldots)\]

which in terms of "generalized volatilities" implies:

\[E[|r(\Delta t)|^q] \sim \Delta t^{H(q)}\]

if \(H(q)\) is linear i.e. \(H(q) = Hq\) \(\rightarrow\) **Unifractal or Monofractal** process:

- Brownian Motion \((H = 0.5)\)
- Fractionally Integrated processes \((H = d - 0.5)\)

if \(H(q)\) is nonlinear \(\rightarrow\) **Multifractal** process: different scaling of different generalized volatility (Ding et al. 1993, Lux 1996, Andersen and Bollerslev 1997 + "econophysicists").

- Multifractal Model of Asset Returns (MMAR), Mandelbrot, Calvet and Fisher (1997):

\[X(t) \equiv B[\theta(t)] \quad \text{where } \theta(t) = \text{c.d.f. of multifractal measure}\]

Possible economic explanation: long term volatility matters for short-term traders while, short-term volatility does not affect long-term trading strategies.

induced some authors to propose analogies with energy cascades of turbulent fluids, borrowing from Kolmogorov model of hydrodynamic turbulence: the so called Stochastic Multiplicative Cascade (SMC)
Stylized Facts and Volatility Models

Standard volatility models are not able to reproduce all the stylized facts:

- GARCH and SV (one factor):
  - no long memory
  - no scaling
  - no volatility cascade

- Fractionally Integrated models:
  - no multi-scaling
  - no volatility cascade
## Models Summary

<table>
<thead>
<tr>
<th>Desired Properties of Volatility Models</th>
<th>GARCH</th>
<th>FIGARCH</th>
<th>ARFIMA</th>
<th>MSTAR</th>
<th>SMAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat tail</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tail cross-over</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Long memory</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Self-similarity</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multifractality</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Volatility Cascade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Economic Interpretation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Simplicity of estimation</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate Extendibility</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A different approach: Heterogeneity

- Additive processes with heterogeneous components can generate those stylized facts!

- Es, Le Baron (2001): combination of only 3 AR(1) can display apparent long memory.

- If the aggregation level is not \( \gg \) than the lowest frequency component
  \( \Rightarrow \) asymptotically short memory models can be mistaken for long memory
  i.e. they are empirically indistinguishable.
Heterogeneous Market Hypothesis (Müller et al. 1993): Main heterogeneity: difference in time horizons ⇒ agents perceive, react and cause different volatility components $\tilde{\sigma}_t^{(\cdot)}$.

Volatility Cascade: hierarchical process from Low to High Frequency.

Realized Volatility Measures: makes volatility observable.

Cascade of Few Heterogeneous Realized Volatility Components

we consider only 3 partial volatility components: daily $\tilde{\sigma}_t^{(d)}$, weekly $\tilde{\sigma}_t^{(w)}$, monthly $\tilde{\sigma}_t^{(m)}$.
1. **Heterogeneous Market Hypothesis** (Müller et al. 1993): Main heterogeneity: difference in time horizons \( \Rightarrow \) agents perceive, react and cause different volatility components \( \tilde{\sigma}_t^{(\cdot)} \)

2. **Volatility Cascade**: hierarchical process from Low to High Frequency.

3. **Realized Volatility Measures**: makes volatility observable.

\[ \downarrow \]

**Cascade of Few Heterogeneous Realized Volatility Components**

we consider only 3 partial volatility components: daily \( \tilde{\sigma}_t^{(d)} \), weekly \( \tilde{\sigma}_t^{(w)} \), monthly \( \tilde{\sigma}_t^{(m)} \)
The HAR-RV Model

- we work with **logs** to avoid negativity issues and get approximately Normal distributions.
- Consider the log RV aggregated, as follow:

\[
\log \text{RV}_t^{(n)} = \frac{1}{n} (\log \text{RV}_t + \ldots + \log \text{RV}_{t-n+1})
\]

at the 3 different horizons: **daily** \( d = 1 \), **weekly** \( w = 5 \), **monthly** \( m = 22 \)

Hence the model reads:

\[
\begin{align*}
\text{HAR-RV}_t &= \tilde{\sigma}_t(d) z_t \\
\log \tilde{\sigma}_{t+m}^{(m)} &= c^{(m)} + \phi^{(m)} \log \text{RV}_t^{(m)} + \tilde{\omega}_{t+m}^{(m)} \\
\log \tilde{\sigma}_{t+w}^{(w)} &= c^{(w)} + \phi^{(w)} \log \text{RV}_t^{(w)} + \gamma^{(w)} E_t[\log \tilde{\sigma}_{t+w}^{(m)}] + \tilde{\omega}_{t+w}^{(w)} \\
\log \tilde{\sigma}_{t+d}^{(d)} &= c^{(d)} + \phi^{(d)} \log \text{RV}_t^{(d)} + \gamma^{(d)} E_t[\log \tilde{\sigma}_{t+w}^{(w)}] + \tilde{\omega}_{t+d}^{(d)}
\end{align*}
\]

A possible economic interpretation

- each market component forms expectation for the next period volatility based on:
  - the current RV experienced at the same time scale ("AR(1) part")
  - the expectation for the next longer horizon partial volatility (Hierarchical part)
The HAR-RV Model

- we work with logs to avoid negativity issues and get approximately Normal distributions.

- Consider the log RV\(_t\) **aggregated**, as follow:

\[
\log \text{RV}\(_t^{(n)}\) = \frac{1}{n} (\log \text{RV}\(_t\) + \ldots + \log \text{RV}\(_{t-n+1}\))
\]

at the 3 different horizons: daily \(d = 1\), weekly \(w = 5\), monthly \(m = 22\)

Hence the model reads:

\[
\begin{align*}
\log \tilde{\sigma}_{t+m}^{(m)} &= c^{(m)} + \phi^{(m)} \log \text{RV}\(_t^{(m)}\) + \tilde{\omega}_{t+m}^{(m)} \\
\log \tilde{\sigma}_{t+w}^{(w)} &= c^{(w)} + \phi^{(w)} \log \text{RV}\(_t^{(w)}\) + \gamma^{(w)} E_t[\log \tilde{\sigma}_{t+m}^{(m)}] + \tilde{\omega}_{t+w}^{(w)} \\
\log \tilde{\sigma}_{t+d}^{(d)} &= c^{(d)} + \phi^{(d)} \log \text{RV}\(_t^{(d)}\) + \gamma^{(d)} E_t[\log \tilde{\sigma}_{t+w}^{(w)}] + \tilde{\omega}_{t+d}^{(d)}
\end{align*}
\]

A possible economic interpretation

each market components forms expectation for the next period volatility based on:
- the current RV experienced at the same time scale ("AR(1) part")
- the expectation for the next longer horizon partial volatility (Hierarchical part)
The HAR-RV Model

- we work with \textbf{logs} to avoid negativity issues and get approximately Normal distributions.

- Consider the log RV \( t \) \textbf{aggregated}, as follow:

\[
\log RV_t^{(n)} = \frac{1}{n} (\log RV_t + \ldots + \log RV_{t-n+1})
\]

at the 3 different horizons: \textit{daily} \( d = 1 \), \textit{weekly} \( w = 5 \), \textit{monthly} \( m = 22 \)

Hence the model reads:

\[
\begin{align*}
    r_t &= \tilde{\sigma}_t^{(d)} z_t \\
    \log \tilde{\sigma}_{t+m}^{(m)} &= c^{(m)} + \phi^{(m)} \log RV_t^{(m)} + \tilde{\omega}_{t+m}^{(m)} \\
    \log \tilde{\sigma}_{t+w}^{(w)} &= c^{(w)} + \phi^{(w)} \log RV_t^{(w)} + \gamma^{(w)} E_t[\log \tilde{\sigma}_{t+m}^{(m)}] + \tilde{\omega}_{t+w}^{(w)} \\
    \log \tilde{\sigma}_{t+d}^{(d)} &= c^{(d)} + \phi^{(d)} \log RV_t^{(d)} + \gamma^{(d)} E_t[\log \tilde{\sigma}_{t+w}^{(w)}] + \tilde{\omega}_{t+d}^{(d)}
\end{align*}
\]

\textbf{A possible economic interpretation}

each market components forms expectation for the next period volatility based on:

- the current RV experienced at the same time scale ("AR(1) part")

- the expectation for the next longer horizon partial volatility (Hierarchical part)
HAR-RV Model

By straightforward recursive substitution

\[
\log \sigma_{t+1d}^{(d)} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + \epsilon_{t+1d}
\]

A three factors Stochastic Volatility model where the factors are directly the past RV

Moreover, being:

\[
\log \sigma_{t+1d} = \log RV_{t+1d} + \tilde{\epsilon}_{t+1d}
\]

where \(\tilde{\epsilon}_t\) is the measurement errors of \(\log RV\), we get

\[
\log RV_{t+1d} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + \epsilon_{t+1d}
\]

a simple AR-type model in the RV with the feature of considering volatilities realized over different interval sizes.

\[\downarrow\]

Heterogeneous AR model in the RV (HAR-RV).
HAR-RV Model

By straightforward recursive substitution

\[
\log \sigma_{t+1d}^{(d)} = c + \beta^{(d)} \log \text{RV}_{t}^{(d)} + \beta^{(w)} \log \text{RV}_{t}^{(w)} + \beta^{(m)} \log \text{RV}_{t}^{(m)} + \epsilon_{t+1d}
\]

A three factors Stochastic Volatility model where the factors are directly the past RV.

Moreover, being:

\[
\log \sigma_{t+1d}^{(d)} = \log \text{RV}_{t+1d}^{(d)} + \tilde{\epsilon}_{t+1d}
\]

where \(\tilde{\epsilon}_t\) is the measurement errors of \(\log \text{RV}\), we get

\[
\log \text{RV}_{t+1d}^{(d)} = c + \beta^{(d)} \log \text{RV}_{t}^{(d)} + \beta^{(w)} \log \text{RV}_{t}^{(w)} + \beta^{(m)} \log \text{RV}_{t}^{(m)} + \epsilon_{t+1d}
\]

a simple AR-type model in the RV with the feature of considering volatilities realized over different interval sizes.

\[\Downarrow\]

Heterogeneous AR model in the RV (HAR-RV).
Simulation Results - daily returns

USD/CHF

Simulation
Simulation Results - daily RV

USD/CHF

Simulation

Fulvio Corsi
HAR Model for Realized Volatility: Extensions and Applications
SNS Pisa 3 March 2010 15 / 102
Kurtosis | daily returns | weekly returns | monthly returns
---|---|---|---
USD/CHF | 4.72 | 3.78 | 3.04
HAR-RV model | 4.89 | 3.90 | 3.50
Simulation Results - RV acf

Sample autocorrelation coefficients

sacf values

k-values
Sample partial autocorrelation coefficients
USD/CHF

Simulated process
Simul. Results - RV periodogram

log–log Periodogram
USD/CHF

Simulation
HAR volatility modelling with heterogeneous leverage and jumps

F. Corsi  R. Renó
“Leverage Effect” (Black 1976): asymmetry in the relationship between returns and volatility. Well-established for equity index through estimation of parametric GARCH and SV models.
Motivation: Volatility memory reloaded

\[ corr(RV_t, RV_{t+h}) \quad corr(r^+_t, RV_{t+h}) \quad corr(r^-_t, RV_{t+h}) \quad corr(J_t, RV_{t+h}) \]
Extending HAR: Leveraged HAR (LHAR)

idea: extend the heterogeneous structure to Leverage effect

With \( r_t = X_t - X_{t-1} \) define aggregated negative and positive returns as:

\[
\begin{align*}
  r_t^{(n)}+ &= \frac{1}{n} (r_t + \ldots + r_{t-n}) I\{ (r_t + \ldots + r_{t-n}) \geq 0 \} \\
  r_t^{(n)}- &= \frac{1}{n} (r_t + \ldots + r_{t-n}) I\{ (r_t + \ldots + r_{t-n}) < 0 \}
\end{align*}
\]

Suppose that now latent volatility is determined by the cascade:

\[
\begin{align*}
  \log \tilde{\sigma}_{t+n_1}^{(n_1)} &= c^{(n_1)} + \beta^{(n_1)} \log RV_t^{(n_1)} + \gamma^{(n_1)} - r_t^{(n_1)} - + \gamma^{(n_1)} + r_t^{(n_1)} + + \epsilon_t^{(n_1)} \\
  \log \tilde{\sigma}_{t+n_2}^{(n_2)} &= c^{(n_2)} + \beta^{(n_2)} \log RV_t^{(n_2)} + \gamma^{(n_2)} - r_t^{(n_2)} - + \gamma^{(n_2)} + r_t^{(n_2)} + + \delta^{(n_2)} \mathbb{E}_t \left[ \log \tilde{\sigma}_{t+n_1}^{(n_1)} \right] + \epsilon_t^{(n_2)}
\end{align*}
\]

which now gives:

\[
\begin{align*}
  \log RV_{t+n_2}^{(n_2)} &= c + \beta^{(n_2)} \log RV_t^{(n_2)} + \beta^{(n_1)} \log RV_t^{(n_1)} \\
  &+ \gamma^{(n_2)} - r_t^{(n_2)} - + \gamma^{(n_2)} + r_t^{(n_2)} + + \gamma^{(n_1)} - r_t^{(n_1)} - + \gamma^{(n_1)} + r_t^{(n_1)} + + \epsilon_t^{(n_2)}
\end{align*}
\]

leverage effects influence each market component separately, and then they appear aggregated at different horizons in the realized volatility dynamics.
Extending HAR: Jumps


\[ dX_t = \mu_t dt + \sigma_t dW_t + dJ_t \]

where: \( \sigma_t dW_t \) the continuous variation, \( dJ_t \) the discontinuous variations with \( J_t = \sum_{j=1}^{N_t} c_j \)

Realized Variance (Andersen and Bollerslev, 1998):

\[
RV_\delta(X)_t = \sum_{j=1}^{[t/\delta]} (\Delta_j X)^2 \longrightarrow \int_0^t \sigma_s^2 ds + \sum_{i=1}^{N_t} c_{\tau_i}^2 \tag{1}
\]

\[
BPV_\delta(X)_t = \mu_1^{-2} MPV_\delta(X)_{[1,1]} = \frac{2}{\pi} \sum_{j=2}^{[t/\delta]} |\Delta_{j-1} X| \cdot |\Delta_j X| \longrightarrow \int_0^t \sigma_s^2 ds \tag{2}
\]

\[
\hat{J}_t = RV_\delta(X)_t - BPV_\delta(X)_t \overset{p}{\longrightarrow} \sum_{j=1}^{N_t} c_j^2
\]

However, BPV has large finite sample bias in presence of jumps (Corsi, Pirino, Renó 2008)
Inspired by the results of Mancini (2007), Corsi, Pirino and Renó (2008) proposed the Threshold Bipower Variation:

\[
TBPV_t = \frac{2}{\pi} \sum_{j=0}^{n-2} |\Delta_{t,j}X| \cdot |\Delta_{t,j+1}X| I\{|\Delta_{t,j}X|^2 \leq \vartheta_{j-1}\} I\{|\Delta_{t,j+1}X|^2 \leq \vartheta_j\}
\]

where \( \vartheta_t \) is a threshold function proportional to the local variance.

It combines threshold estimation for large jumps and bipower variation for smaller jumps (twin blade approach). It is expected to be not too sensitive to the threshold.
The LHAR-CJ model

Putting all components together (heterogeneity, leverage and jumps) we propose the

**Leverage Heterogeneous AutoRegressive with Continuous volatility and Jumps (LHAR-CJ) model.**

considering 3 different horizons: monthly $n_1 = 22$, weekly $n_2 = 5$ and daily $n_3 = 1$:

$$\log \text{RV}^{(h)}_{t+h} = c + \beta^{(d)} \log C^{(d)}_t + \beta^{(w)} \log C^{(w)}_t + \beta^{(m)} \log C^{(m)}_t$$

$$+ \alpha^{(d)} \log (1 + J^{(d)}_t) + \alpha^{(w)} \log (1 + J^{(w)}_t) + \alpha^{(m)} \log (1 + J^{(m)}_t)$$

$$+ \gamma^{(d)} - r^{(d)}_t + \gamma^{(w)} - r^{(w)}_t + \gamma^{(m)} - r^{(m)}_t + \varepsilon^{(h)}_{t+h},$$

The model is estimated by OLS with Newey-West covariance correction for serial correlation.
Empirical Analysis: Data

- Long time span of 28 years of tick-by-tick data for S&P 500 futures from January 1982 to February 2009 (6,669 days), '87 crash excluded.

- In order to mitigate the impact of microstructure effects:
  - The daily RV are computed with the Two-Scales estimator proposed by Zhang Mykland and Ait Sahalia (2005),
  - While the TBPV for jump detection is computed at the sampling frequency of 5 minutes (84 returns per day).

<table>
<thead>
<tr>
<th>Variable</th>
<th>One day</th>
<th>One week</th>
<th>Two weeks</th>
<th>One month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>$0.442^*$</td>
<td>$0.549^*$</td>
<td>$0.662^*$</td>
<td>$0.858^*$</td>
</tr>
<tr>
<td></td>
<td>$(10.699)$</td>
<td>$(9.258)$</td>
<td>$(8.525)$</td>
<td>$(7.756)$</td>
</tr>
<tr>
<td>$C$</td>
<td>$0.307^*$</td>
<td>$0.201^*$</td>
<td>$0.154^*$</td>
<td>$0.116^*$</td>
</tr>
<tr>
<td>$C^{(5)}$</td>
<td>$0.369^*$</td>
<td>$0.359^*$</td>
<td>$0.332^*$</td>
<td>$0.286^*$</td>
</tr>
<tr>
<td></td>
<td>$(13.908)$</td>
<td>$(11.251)$</td>
<td>$(9.166)$</td>
<td>$(6.784)$</td>
</tr>
<tr>
<td>$C^{(22)}$</td>
<td>$0.222^*$</td>
<td>$0.319^*$</td>
<td>$0.370^*$</td>
<td>$0.415^*$</td>
</tr>
<tr>
<td></td>
<td>$(10.958)$</td>
<td>$(10.913)$</td>
<td>$(10.198)$</td>
<td>$(9.344)$</td>
</tr>
<tr>
<td>$J$</td>
<td>$0.043^*$</td>
<td>$0.020^*$</td>
<td>$0.017^*$</td>
<td>$0.012^*$</td>
</tr>
<tr>
<td></td>
<td>$(7.057)$</td>
<td>$(4.453)$</td>
<td>$(4.485)$</td>
<td>$(3.804)$</td>
</tr>
<tr>
<td>$J^{(5)}$</td>
<td>$0.011^*$</td>
<td>$0.013^*$</td>
<td>$0.011^*$</td>
<td>$0.010$</td>
</tr>
<tr>
<td></td>
<td>$(3.373)$</td>
<td>$(3.112)$</td>
<td>$(2.256)$</td>
<td>$(1.913)$</td>
</tr>
<tr>
<td>$J^{(22)}$</td>
<td>$0.005^*$</td>
<td>$0.008^*$</td>
<td>$0.010^*$</td>
<td>$0.014^*$</td>
</tr>
<tr>
<td></td>
<td>$(2.199)$</td>
<td>$(2.106)$</td>
<td>$(2.205)$</td>
<td>$(2.336)$</td>
</tr>
<tr>
<td>$r^-$</td>
<td>$-0.007^*$</td>
<td>$-0.005^*$</td>
<td>$-0.004^*$</td>
<td>$-0.003^*$</td>
</tr>
<tr>
<td></td>
<td>$(-9.669)$</td>
<td>$(-10.435)$</td>
<td>$(-8.298)$</td>
<td>$(-5.518)$</td>
</tr>
<tr>
<td>$r^{(5)}^-$</td>
<td>$-0.008^*$</td>
<td>$-0.006^*$</td>
<td>$-0.008^*$</td>
<td>$-0.007^*$</td>
</tr>
<tr>
<td></td>
<td>$(-4.412)$</td>
<td>$(-3.059)$</td>
<td>$(-4.012)$</td>
<td>$(-3.472)$</td>
</tr>
<tr>
<td>$r^{(22)}^-$</td>
<td>$-0.009^*$</td>
<td>$-0.012^*$</td>
<td>$-0.009$</td>
<td>$-0.004$</td>
</tr>
<tr>
<td></td>
<td>$(-2.845)$</td>
<td>$(-2.314)$</td>
<td>$(-1.481)$</td>
<td>$(-0.467)$</td>
</tr>
</tbody>
</table>

| $R^2$    | $0.7664$  | $0.8137$  | $0.8030$  | $0.7629$  |
| HRMSE    | $0.2168$  | $0.1692$  | $0.1699$  | $0.1796$  |
In-sample forecast


horizon (days)

0.74 0.76 0.78 0.8 0.82

HAR
HAR–CJ
LHAR–CJ
Out of sample forecast

Mincer-Zarnowitz $R^2$

- **HAR**
- **HAR−CJ**
- **LHAR−TCJ**

Fulvio Corsi ()
HAR Model for Realized Volatility: Extensions and Applications
SNS Pisa 3 March 2010
Is leverage effect induced by jumps

<table>
<thead>
<tr>
<th>HAR-CJ$^+$ regression</th>
<th>1-day</th>
<th>1-week</th>
<th>2-weeks</th>
<th>1-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.232*</td>
<td>0.377*</td>
<td>0.505*</td>
<td>0.747*</td>
</tr>
<tr>
<td></td>
<td>(5.774)</td>
<td>(6.217)</td>
<td>(6.418)</td>
<td>(6.736)</td>
</tr>
<tr>
<td>$C$</td>
<td>0.398*</td>
<td>0.265*</td>
<td>0.214*</td>
<td>0.165*</td>
</tr>
<tr>
<td>$C^{(5)}$</td>
<td>0.366*</td>
<td>0.368*</td>
<td>0.346*</td>
<td>0.291*</td>
</tr>
<tr>
<td></td>
<td>(13.889)</td>
<td>(11.697)</td>
<td>(9.750)</td>
<td>(7.327)</td>
</tr>
<tr>
<td>$C^{(22)}$</td>
<td>0.190*</td>
<td>0.291*</td>
<td>0.338*</td>
<td>0.390*</td>
</tr>
<tr>
<td></td>
<td>(9.470)</td>
<td>(9.743)</td>
<td>(9.059)</td>
<td>(8.875)</td>
</tr>
<tr>
<td>$J^+$</td>
<td>0.044*</td>
<td>0.018*</td>
<td>0.016*</td>
<td>0.013*</td>
</tr>
<tr>
<td></td>
<td>(6.099)</td>
<td>(3.264)</td>
<td>(3.000)</td>
<td>(2.538)</td>
</tr>
<tr>
<td>$J^-$</td>
<td>0.074*</td>
<td>0.040*</td>
<td>0.039*</td>
<td>0.027*</td>
</tr>
<tr>
<td>$J^{(5)}$</td>
<td>0.009*</td>
<td>0.012*</td>
<td>0.010*</td>
<td>0.010*</td>
</tr>
<tr>
<td></td>
<td>(2.645)</td>
<td>(2.724)</td>
<td>(2.028)</td>
<td>(1.799)</td>
</tr>
<tr>
<td>$J^{(22)}$</td>
<td>0.005</td>
<td>0.007</td>
<td>0.009*</td>
<td>0.014*</td>
</tr>
<tr>
<td></td>
<td>(1.845)</td>
<td>(1.875)</td>
<td>(2.026)</td>
<td>(2.242)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LHAR-CJ$^+$ regression</th>
<th>1-day</th>
<th>1-week</th>
<th>2-week</th>
<th>1-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.442*</td>
<td>0.549*</td>
<td>0.661*</td>
<td>0.858*</td>
</tr>
<tr>
<td></td>
<td>(10.724)</td>
<td>(9.277)</td>
<td>(8.531)</td>
<td>(7.778)</td>
</tr>
<tr>
<td>$C$</td>
<td>0.307*</td>
<td>0.201*</td>
<td>0.154*</td>
<td>0.116*</td>
</tr>
<tr>
<td></td>
<td>(16.972)</td>
<td>(14.185)</td>
<td>(13.007)</td>
<td>(10.608)</td>
</tr>
<tr>
<td>$C^{(5)}$</td>
<td>0.369*</td>
<td>0.359*</td>
<td>0.332*</td>
<td>0.286*</td>
</tr>
<tr>
<td>$C^{(22)}$</td>
<td>0.222*</td>
<td>0.319*</td>
<td>0.370*</td>
<td>0.415*</td>
</tr>
<tr>
<td></td>
<td>(10.914)</td>
<td>(10.905)</td>
<td>(10.183)</td>
<td>(9.336)</td>
</tr>
<tr>
<td>$J^+$</td>
<td>0.044*</td>
<td>0.018*</td>
<td>0.015*</td>
<td>0.012*</td>
</tr>
<tr>
<td></td>
<td>(6.176)</td>
<td>(3.182)</td>
<td>(2.819)</td>
<td>(2.395)</td>
</tr>
<tr>
<td>$J^-$</td>
<td>0.043*</td>
<td>0.019*</td>
<td>0.020*</td>
<td>0.011*</td>
</tr>
<tr>
<td></td>
<td>(4.598)</td>
<td>(3.387)</td>
<td>(3.633)</td>
<td>(2.285)</td>
</tr>
<tr>
<td>$J^{(5)}$</td>
<td>0.011*</td>
<td>0.013*</td>
<td>0.011*</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(3.372)</td>
<td>(3.110)</td>
<td>(2.254)</td>
<td>(1.914)</td>
</tr>
<tr>
<td>$J^{(22)}$</td>
<td>0.005*</td>
<td>0.008*</td>
<td>0.010*</td>
<td>0.014*</td>
</tr>
<tr>
<td></td>
<td>(2.200)</td>
<td>(2.104)</td>
<td>(2.203)</td>
<td>(2.337)</td>
</tr>
</tbody>
</table>

$R^2$: 0.7543 0.8060 0.7960 0.7582
HRMSE: 0.2201 0.1721 0.1722 0.1812

Fulvio Corsi
HAR Model for Realized Volatility: Extensions and Applications
SNS Pisa 3 March 2010
Modeling tick-by-tick realized correlations

F. Audrino   F. Corsi
Realized Covariances (RCov)

- Given a continuous time process for asset $j$
  \[ dp_j(t) = \mu_j(t)dt + \sigma_j(t)dW_j(t) \]
  and discrete price observations \( \{p_j(n_q)\}_{q=1,2,\ldots,M} \)

- Realized Volatility (RV) approach affords estimation of the variance of asset $j$ through
  \[ RV_{j,t} \equiv \sum_{q=1}^{M} r_{j,q}^2 \xrightarrow{M \to \infty} \int_{0}^{t} \sigma_j^2(\omega)d\omega \equiv IV_{j,t} \]
  with \( r_{j,q} = p_j(n_j,q) - p_j(n_j,q-1) \) the intraday high frequency returns.

- **Idea**: as for Realized Volatility, employing high frequency data for the computation of covariances between two asset $i$ and $j$.

- **Standard way**: choose a time interval, construct an artificially regularly spaced time series (say 5-min) and take the sum of contemporaneous cross products.

- But, as for the RV, the presence of market microstructure can induce significant bias in standard RCov. The so called "Epps effect".
The microstructure effects responsible for this bias in the standard RC is the so called \textit{non-synchronous trading} effect (Lo and MacKinlay 1990) which induces a bias toward zero in the standard interpolated measure:

\begin{center}
\textit{lost portion of covariance}
\end{center}

\[ HY_t = \sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} r_{i,s} r_{j,q} I(\delta_{q,s} > 0) \]

with \( I(\cdot) \) the indicator function and

\[ \delta_{q,s} = \max(0, \min(n_{i,s+1}, n_{j,q+1}) - \max(n_{i,s}, n_{j,q})) \]

being the overlap between any two tick-by-tick returns \( r_{i,s} \) and \( r_{j,q} \).
Realized Correlations - Empirical Results

Lag

ACF

full sample
90-94
94-98
98-July 07
July 07-March 08

Fulvio Corsi

HAR Model for Realized Volatility: Extensions and Applications
SNS Pisa 3 March 2010 37 / 102
Let $\{\tilde{RC}_{i} \}_{i \geq 1}$ be the daily Fisher-transformed (FT) series of the tick-by-tick realized correlations $\{RC_{i} \}_{i \geq 1}$,

We then model the series $\{\tilde{RC}_{i} \}_{i \geq 1}$ as:

$$\tilde{RC}_{t+1} = \mathbb{E}_{t}[\tilde{RC}_{t+1}] + \sigma_{t+1} U_{t+1}$$

where the conditional dynamics of the FT correlations are given by:

$$\mathbb{E}_{t}[\tilde{RC}_{t+1}] = \sum_{j=1}^{k} \left( a_{j} + b_{j}^{(d)} \tilde{RC}_{t} + b_{j}^{(w)} \tilde{RC}_{t}^{(w)} + b_{j}^{(m)} \tilde{RC}_{t}^{(m)} \right) I_{x_{t}^{\text{pred}} \in \mathcal{R}_{j}}$$

$$\sigma_{t+1}^{2} = \sum_{j=1}^{k} \sigma_{j}^{2} I_{x_{t}^{\text{pred}} \in \mathcal{R}_{j}}, \quad \sigma_{j}^{2} > 0, j = 1, \ldots, k$$

$\theta = (a_{j}, b_{j}^{(d)}, b_{j}^{(w)}, b_{j}^{(m)}, \sigma_{j}^{2} : j = 1, \ldots, k)$ is a parameter vector which parameterizes the local HAR dynamics in the different regimes,

$k$ is the number of regimes (endogenously estimated from the data)
Regimes are characterized by partition cells $\mathcal{R}_j$ of the relevant predictor space $G$ of $X_t^{\text{pred}}$:

$$G = \bigcup_{j=1}^{k} \mathcal{R}_j, \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset \ (i \neq j)$$

The relevant **predictor variables** in $X_t^{\text{pred}}$ are:

- past-lagged FT-realized correlations,
- past-lagged realized volatilities (daily, weekly, and monthly),
- past-lagged returns (daily, weekly, and monthly),
- time.

Similar to the standard classification and regression trees (CART) procedure we impose regimes to be characterized by **rectangular partition cells** with edges determined by thresholds on the predictor variables. The rectangular partition has two major advantages:

1. allows a clear interpretation of regimes in terms of relevant predictor variables,
2. allows model estimation also in large-dimensional predictor spaces $G$. 

in our empirical application of Tree-HAR on US stock–bond the optimal partition cells are:

<table>
<thead>
<tr>
<th>Regime structure</th>
<th>Local parameters</th>
<th>Local parameters</th>
<th>Local parameters</th>
<th>Local parameters</th>
<th>Local parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R ) ( S&amp;P 500 \leq -0.935, ) ( t \leq July 2007 )</td>
<td>( a_j )</td>
<td>( b_j^{(d)} )</td>
<td>( b_j^{(w)} )</td>
<td>( b_j^{(m)} )</td>
<td>( \sigma_j^2 )</td>
</tr>
<tr>
<td>( R ) ( S&amp;P 500 &gt; -0.935, ) ( t \leq December 1993 )</td>
<td>0.0649</td>
<td>0.1042</td>
<td>0.3419</td>
<td>0.2741</td>
<td>0.0219</td>
</tr>
<tr>
<td>( R ) ( S&amp;P 500 &gt; -0.935, ) ( December 1993 &lt; t \leq July 2007 )</td>
<td>0.0069</td>
<td>0.2125</td>
<td>0.4127</td>
<td>0.3290</td>
<td>0.0407</td>
</tr>
<tr>
<td>( t &gt; July 2007 )</td>
<td>-0.2432</td>
<td>0.3237</td>
<td>0.2643</td>
<td>-0.0416</td>
<td>0.1468</td>
</tr>
</tbody>
</table>
### Tree HAR In-sample estimation results

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>MAE</th>
<th>MSE</th>
<th>LB(5)</th>
<th>LB(20)</th>
<th>LB(50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>224.81</td>
<td>244.02</td>
<td>0.1794</td>
<td>0.0615</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>-919.18</td>
<td>-893.58</td>
<td>0.1551</td>
<td>0.0475</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARIMA(1,1,1)</td>
<td>-980.01</td>
<td>-954.40</td>
<td>0.1545</td>
<td>0.0469</td>
<td>0.0005</td>
<td>0</td>
<td>0.0014</td>
</tr>
<tr>
<td>Tree-AR(1)</td>
<td>-905.30</td>
<td>-828.49</td>
<td>0.1650</td>
<td>0.0522</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAR</td>
<td>-984.68</td>
<td>-952.67</td>
<td>0.1539</td>
<td>0.0468</td>
<td>0</td>
<td>0</td>
<td>0.0012</td>
</tr>
<tr>
<td>RS-HAR</td>
<td>-1876.9</td>
<td>-1787.4</td>
<td>0.1532</td>
<td>0.0464</td>
<td>0.1205</td>
<td>0.0254</td>
<td>0.2196</td>
</tr>
<tr>
<td>Tree-HAR</td>
<td>-1575.6</td>
<td>-1447.6</td>
<td><strong>0.1523</strong></td>
<td><strong>0.0455</strong></td>
<td>0.1081</td>
<td><strong>0.0608</strong></td>
<td><strong>0.4333</strong></td>
</tr>
</tbody>
</table>

→ For all measures considered, the tree-HAR or the RS-HAR models yield the best results.

→ Neglecting to incorporate long-memory or structural breaks lead to highly inaccurate estimates.
Forecasting results: single-period forecasts

<table>
<thead>
<tr>
<th>Model</th>
<th>Loglik.</th>
<th>MAE</th>
<th>MSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>418.20 (0)</td>
<td>0.2031 (0)</td>
<td>0.0811 (0)</td>
<td>0.3747</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>78.04 (0.014)</td>
<td><strong>0.1696 (0.137)</strong></td>
<td>0.0595 (0.011)</td>
<td>0.4758</td>
</tr>
<tr>
<td>ARIMA(1,1,1)</td>
<td>59.46 (0.033)</td>
<td><strong>0.1685 (0.469)</strong></td>
<td><strong>0.0585 (0.897)</strong></td>
<td>0.4814</td>
</tr>
<tr>
<td>Tree-AR(1)</td>
<td>168.37 (0)</td>
<td>0.1948 (0)</td>
<td>0.0801 (0)</td>
<td>0.3899</td>
</tr>
<tr>
<td>HAR</td>
<td>70.73 (0.009)</td>
<td><strong>0.1683 (0.658)</strong></td>
<td><strong>0.0589 (0.366)</strong></td>
<td>0.4779</td>
</tr>
<tr>
<td>RS-HAR</td>
<td>$-102.03 (0.813)$</td>
<td><strong>0.1689 (0.343)</strong></td>
<td><strong>0.0586 (0.685)</strong></td>
<td>0.4815</td>
</tr>
<tr>
<td>Tree-HAR</td>
<td>$-99.31 (0.617)$</td>
<td><strong>0.1676 (0.700)</strong></td>
<td><strong>0.0586 (0.433)</strong></td>
<td><strong>0.4823</strong></td>
</tr>
</tbody>
</table>

→ The tree-HAR is always among the best models and the best for $R^2$. 
Forecasting results: multi-period forecasts II

<table>
<thead>
<tr>
<th>Model</th>
<th>one week horizon</th>
<th>one month horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>MSE</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.3249</td>
<td>0.1777</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.1895</td>
<td>0.0753</td>
</tr>
<tr>
<td>ARIMA(1,1,1)</td>
<td><strong>0.1874</strong></td>
<td><strong>0.0721</strong></td>
</tr>
<tr>
<td>Tree-AR(1)</td>
<td>0.2629</td>
<td>0.1553</td>
</tr>
<tr>
<td>HAR</td>
<td>0.1874</td>
<td>0.0731</td>
</tr>
<tr>
<td>RS-HAR</td>
<td>0.1888</td>
<td>0.0743</td>
</tr>
<tr>
<td>Tree-HAR</td>
<td><strong>0.1844</strong></td>
<td><strong>0.0702</strong></td>
</tr>
</tbody>
</table>

→ Tree-HAR model consistently outperform alternative ones.
→ Gains are, in most cases, statistically significant (especially at one month).