HAR Model for Realized Volatility: Extensions and Applications

Fulvio Corsi

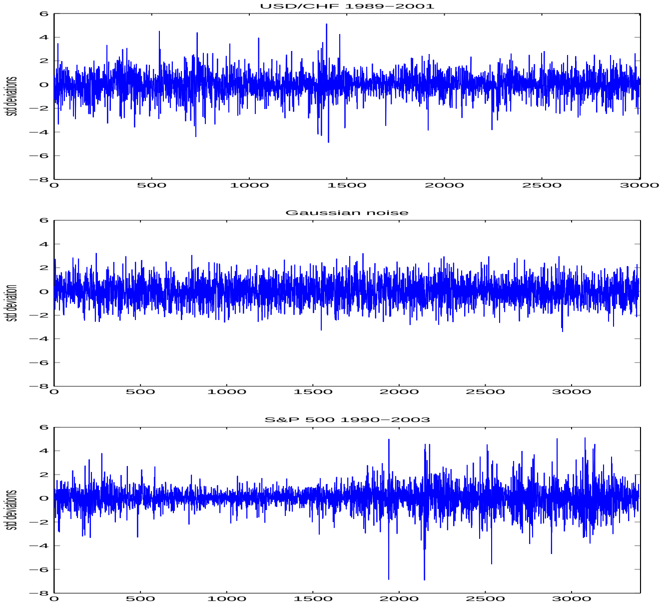
SNS Pisa

3 March 2010

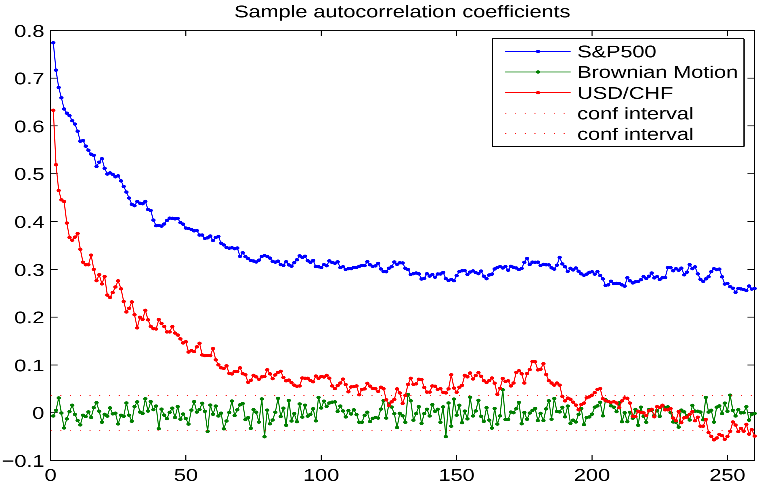
Outline

- The HAR model:
  - F. Corsi (2009). *"A Simple Approximate Long Memory Model of Realized Volatility"* (JFEC)
- 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



Stylized Facts II: Volatility Persistence



Long Memory Models

- 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:

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

$$\downarrow$$

- Fractional Integration: generalize the usual differencing of $I(1)$ series y_t to get an $I(0)$ ϵ_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 $\longrightarrow$ FIGARH

Long Memory Models

- 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:

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

$$\downarrow$$

- Fractional Integration: generalize the usual differencing of $I(1)$ series y_t to get an $I(0)$ ϵ_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 $\longrightarrow$ FIGARH

Stylized Facts III: Fractality & Multifractality

- Self-Similar, Fractal or Scaling Process:

$$(Y_{t1}, Y_{t2}, Y_{t3}, \ldots) \stackrel{d}{=} \mathbf{c}^{-H}(Y_{\mathbf{c}t1}, Y_{\mathbf{c}t2}, Y_{\mathbf{c}t3}, \ldots)$$

which in terms of "generalized volatilities" implies:

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

- if $\mathcal{H}(q)$ is linear i.e. $\mathcal{H}(q) = H\,q \longrightarrow$ Unifractal or Monofractal process:

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

- if $\mathcal{H}(q)$ is nonlinear $\longrightarrow$ 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):

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

Stylized Facts IV: Volatility Cascade and Scaling

- Asymmetric propagation of volatility: volatility over longer time intervals have stronger influence on those at shorter time intervals than conversely.
  (Müller et al. 1997, Arneodo et al. 1998, Lynch 2000, Gençay et al. 2002).

  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

Desired Properties of Volatility Models ↓	GARCH	SV	FIGARCH	ARFIMA	MMAR	SMC
Fat tail	√	√	√	√	√	√
Tail cross-over	√	√				√
Long memory			√	√	√	√
Self-similarity			√	√	√	√
Multifractality					√	√
Volatility Cascade						√
Economic Interpretation						√
Simplicity of estimation	√					
Multivariate Extendibility	√					

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.

Model Ingredients

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 $\tilde{\sigma}_t^{(m)}$

Model Ingredients

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 $\log \mathrm{RV}_t$ **aggregated**, as follow:

$$\log \mathrm{RV}_t^{(n)} = \frac{1}{n}\left(\log \mathrm{RV}_t + \ldots + \log \mathrm{RV}_{t-n+1}\right)$$

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

Hence the model reads:

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

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 **logs** to avoid negativity issues and get approximately Normal distributions.
- Consider the $\log \mathrm{RV}_t$ **aggregated**, as follow:

$$\log \mathrm{RV}_t^{(n)} = \frac{1}{n}\left(\log \mathrm{RV}_t + \ldots + \log \mathrm{RV}_{t-n+1}\right)$$

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

Hence the model reads:

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

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 **logs** to avoid negativity issues and get approximately Normal distributions.
- Consider the $\log \mathrm{RV}_t$ **aggregated**, as follow:

$$\log \mathrm{RV}_t^{(n)} = \frac{1}{n}\left(\log \mathrm{RV}_t + \ldots + \log \mathrm{RV}_{t-n+1}\right)$$

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

Hence the model reads:

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

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

$$\log \sigma_{t+1d}^{(d)} \quad = \quad c + \beta^{(d)} \log \mathrm{RV}_t^{(d)} + \beta^{(w)} \log \mathrm{RV}_t^{(w)} + \beta^{(m)} \log \mathrm{RV}_t^{(m)} + \epsilon_{t+1d}^{(d)}$$

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

- Moreover, being:

$$\log \sigma_{t+1d}^{(d)} = \log \mathrm{RV}_{t+1d}^{(d)} + \tilde{\epsilon}_{t+1d}$$

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

$$\log \mathrm{RV}_{t+1d}^{(d)} \quad = \quad c + \beta^{(d)} \log \mathrm{RV}_t^{(d)} + \beta^{(w)} \log \mathrm{RV}_t^{(w)} + \beta^{(m)} \log \mathrm{RV}_t^{(m)} + \epsilon_{t+1d}$$

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

$$\log \sigma_{t+1d}^{(d)} \quad = \quad c + \beta^{(d)} \log \mathrm{RV}_t^{(d)} + \beta^{(w)} \log \mathrm{RV}_t^{(w)} + \beta^{(m)} \log \mathrm{RV}_t^{(m)} + \epsilon_{t+1d}^{(d)}$$

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

- Moreover, being:

$$\log \sigma_{t+1d}^{(d)} = \log \mathrm{RV}_{t+1d}^{(d)} + \tilde{\epsilon}_{t+1d}$$

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

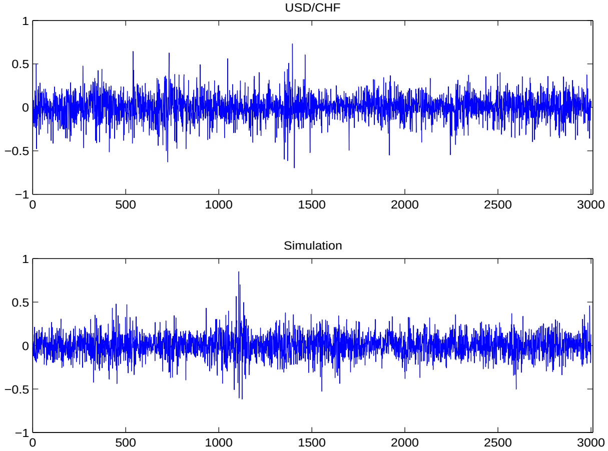
$$\log \mathrm{RV}_{t+1d}^{(d)} \quad = \quad c + \beta^{(d)} \log \mathrm{RV}_t^{(d)} + \beta^{(w)} \log \mathrm{RV}_t^{(w)} + \beta^{(m)} \log \mathrm{RV}_t^{(m)} + \epsilon_{t+1d}$$

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

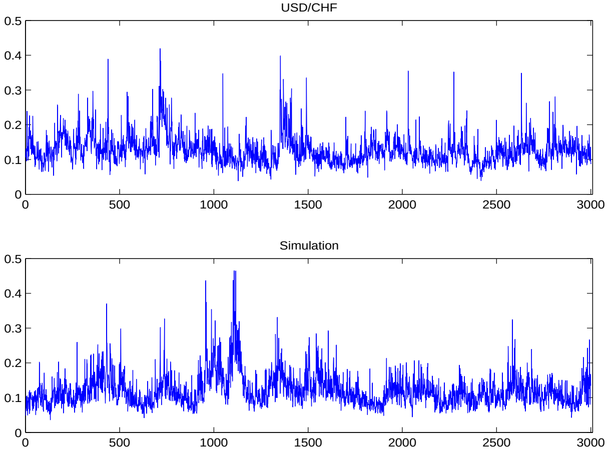
$$\Downarrow$$

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

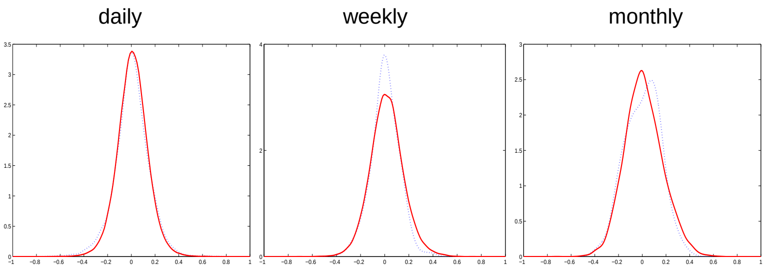
Simulation Results - daily returns



Simulation Results - daily RV

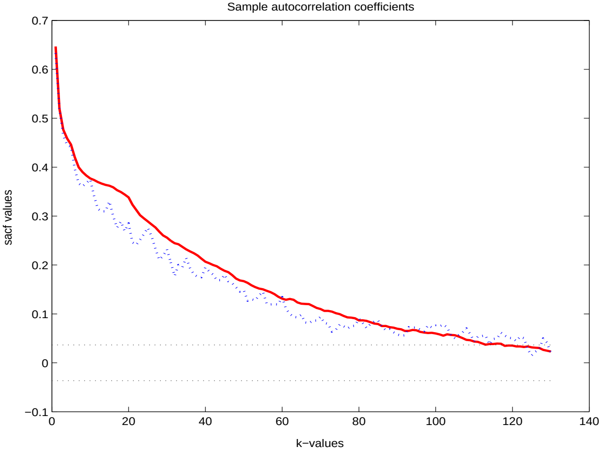


Simulation Results - return pdf

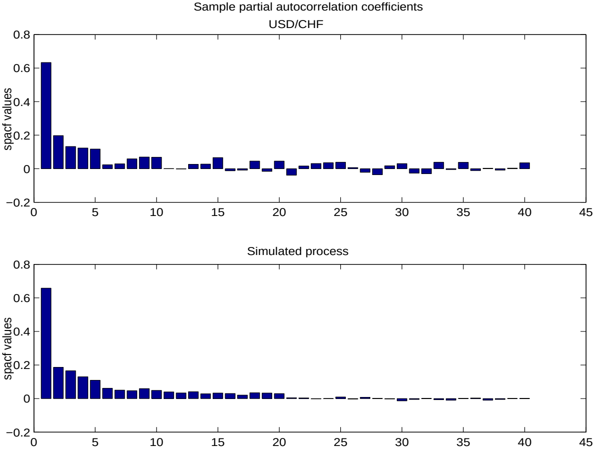


Kurtosis	daily returns	weekly returns	monthly returns
USD/CHF	4.72	3.78	3.04
HAR-RV model	4.89	3.90	3.50

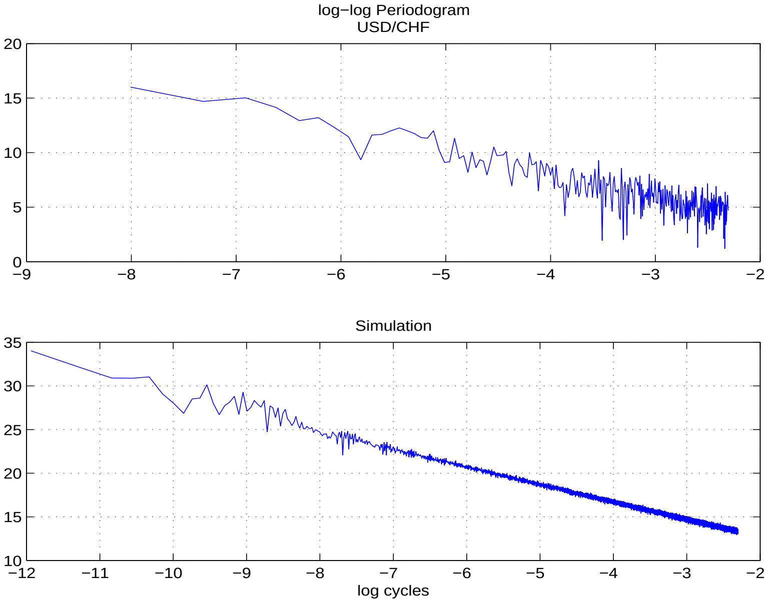
Simulation Results - RV acf



Simul. Results - RV partial-acf



Simul. Results - RV periodogram



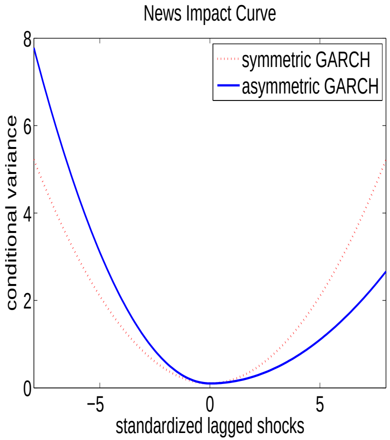
HAR volatility modelling with heterogeneous leverage and jumps

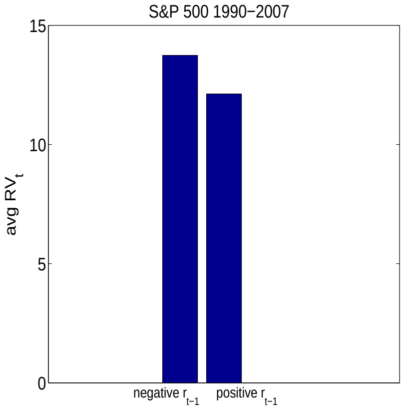
F. Corsi   R. Renó

The "Leverage Effect"

"Leverage Effect" (Black 1976): asymmetry in the relationship between returns and volatility.

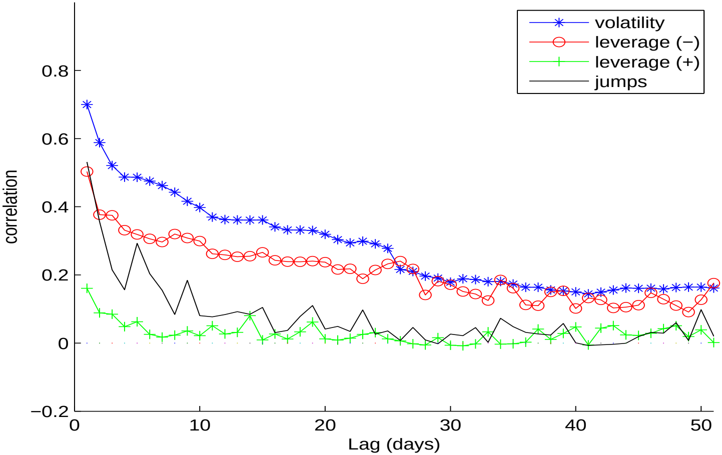
well-established for equity index through estimation of parametric GARCH and SV models





Motivation: Volatility memory reloaded

$corr(RV_t, RV_{t+h})$ $corr(r_t^+, RV_{t+h})$ $corr(r_t^-, RV_{t+h})$ $corr(J_t, RV_{t+h})$



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:

$$
\begin{aligned}
r_t^{(n)+} &= \frac{1}{n}\left(r_t + \ldots + r_{t-n}\right) I_{\{(r_t+\ldots+r_{t-n})\geq 0\}} \\
r_t^{(n)-} &= \frac{1}{n}\left(r_t + \ldots + r_{t-n}\right) I_{\{(r_t+\ldots+r_{t-n})< 0\}}
\end{aligned}
$$

Suppose that now latent volatility is determined by the cascade:

$$
\begin{aligned}
\log \tilde{\sigma}_{t+n_1}^{(n_1)} &= c^{(n_1)} + \beta^{(n_1)} \log \mathrm{RV}_t^{(n_1)} + \gamma^{(n_1)-} r_t^{(n_1)-} + \gamma^{(n_1)+} r_t^{(n_1)+} + \varepsilon_{t+n_1}^{(n_1)} \\
\log \tilde{\sigma}_{t+n_2}^{(n_2)} &= c^{(n_2)} + \beta^{(n_2)} \log \mathrm{RV}_t^{(n_2)} + \gamma^{(n_2)-} r_t^{(n_2)-} + \gamma^{(n_2)+} r_t^{(n_2)+} + \delta^{(n_2)} \mathbb{E}_t \left[\log \tilde{\sigma}_{t+n_1}^{(n_1)}\right] + \varepsilon_{t+n_2}^{(n_2)}
\end{aligned}
$$

which now gives:

$$
\begin{aligned}
\log \mathrm{RV}_{t+n_2}^{(n_2)} = c \quad &+ \quad \beta^{(n_2)} \log \mathrm{RV}_t^{(n_2)} + \beta^{(n_1)} \log \mathrm{RV}_t^{(n_1)} \\
&+ \quad \gamma^{(n_2)-} r_t^{(n_2)-} + \gamma^{(n_2)+} r_t^{(n_2)+} + \gamma^{(n_1)-} r_t^{(n_1)-} + \gamma^{(n_1)+} r_t^{(n_1)+} + \tilde{\varepsilon}_{t+n_2}^{(n_2)}
\end{aligned}
$$

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

Extending HAR: Jumps

Literature on jump-diffusion models: Andersen, Benzoni and Lund (2002), Bates (1996, 2000), Chernov, Gallant, Ghysels and Tauchen (2003), Eraker (2004), Eraker, Johannes and Polson (2003), Johannes (2004), Pan (2002). Simple jump-diffusion are:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t$$

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):

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

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

$$\hat{J}_t = \mathrm{RV}_\delta\,(X)_t - \mathrm{BPV}_\delta\,(X)_t \xrightarrow{p} \sum_{j=1}^{N_t} c_j^2$$

However, BPV has large finite sample bias in presence of jumps (Corsi, Pirino, Renó 2008)

Threshold Bipower Variation

- Inspired by the results of Mancini (2007), Corsi, Pirino and Renó (2008) proposed the Threshold Bipower Variation:

$$\text{TBPV}_t = \frac{2}{\pi} \sum_{j=0}^{n-2} |\Delta_{t,j}X| \cdot |\Delta_{t,j+1}X| I_{\{|\Delta_{t,j}X|^2 \leq \vartheta_{j-1}\}} I_{\{|\Delta_{t,j+1}X|^2 \leq \vartheta_j\}}$$

- where ϑ_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 $n_3 = 1$:

$$\begin{aligned}\log \mathrm{RV}_{t+h}^{(h)} = c \quad &+ \quad \beta^{(\mathbf{d})} \log \mathrm{C}_{\mathbf{t}}^{(\mathbf{d})} + \beta^{(\mathbf{w})} \log \mathrm{C}_{\mathbf{t}}^{(\mathbf{w})} + \beta^{(\mathbf{m})} \log \mathrm{C}_{\mathbf{t}}^{(\mathbf{m})} \\ &+ \quad \alpha^{(\mathbf{d})} \log(\mathbf{1} + \mathrm{J}_{\mathbf{t}}^{(\mathbf{d})}) + \alpha^{(\mathbf{w})} \log(\mathbf{1} + \mathrm{J}_{\mathbf{t}}^{(\mathbf{w})}) + \alpha^{(m)} \log(1 + \mathrm{J}_{t}^{(m)}) \\ &+ \quad \gamma^{(\mathbf{d})-} \mathbf{r}_{\mathbf{t}}^{(\mathbf{d})-} + \gamma^{(\mathbf{w})-} \mathbf{r}_{\mathbf{t}}^{(\mathbf{w})-} + \gamma^{(\mathbf{m})-} \mathbf{r}_{\mathbf{t}}^{(\mathbf{m})-} + \varepsilon_{t+h}^{(h)},\end{aligned}$$

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 A it Sahalya (2005),

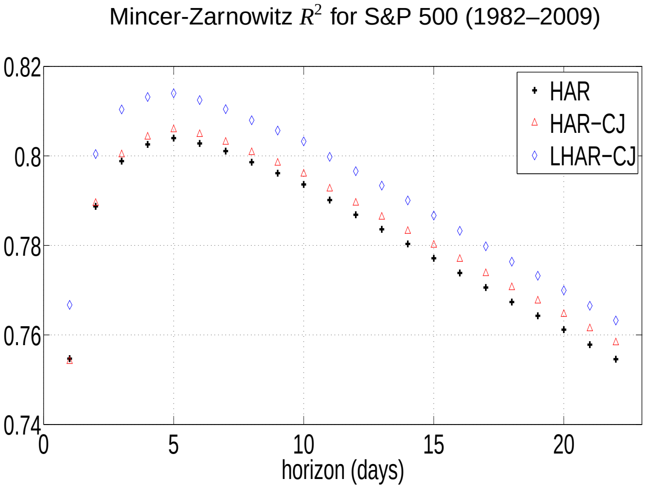
    - while the TBPV for jump detection is computed at the sampling frequency of 5 minutes (84 returns per day).

In-sample estimates

S&P500 LHAR-CJ in-sample estimates, period 1982–2009

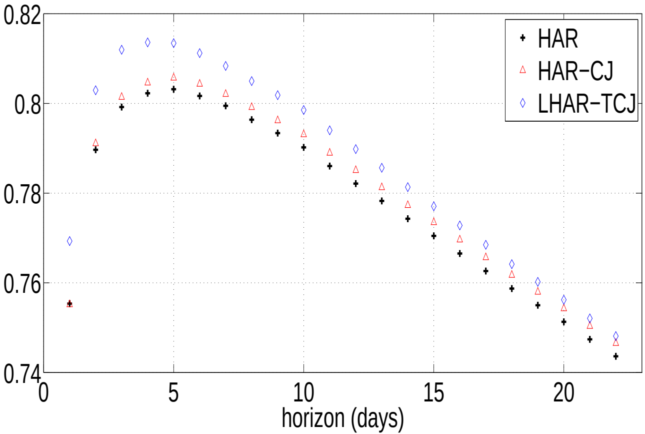
Variable	One day	One week	Two weeks	One month
c	0.442*	0.549*	0.662*	0.858*
	(10.699)	(9.258)	(8.525)	(7.756)
C	0.307*	0.201*	0.154*	0.116*
	(16.983)	(14.158)	(12.984)	(10.590)
$C^{(5)}$	0.369*	0.359*	0.332*	0.286*
	(13.908)	(11.251)	(9.166)	(6.784)
$C^{(22)}$	0.222*	0.319*	0.370*	0.415*
	(10.958)	(10.913)	(10.198)	(9.344)
J	0.043*	0.020*	0.017*	0.012*
	(7.057)	(4.453)	(4.485)	(3.804)
$J^{(5)}$	0.011*	0.013*	0.011*	0.010
	(3.373)	(3.112)	(2.256)	(1.913)
$J^{(22)}$	0.005*	0.008*	0.010*	0.014*
	(2.199)	(2.106)	(2.205)	(2.336)
r^{-}	-0.007*	-0.005*	-0.004*	-0.003*
	(-9.669)	(-10.435)	(-8.298)	(-5.518)
$r^{(5)-}$	-0.008*	-0.006*	-0.008*	-0.007*
	(-4.412)	(-3.059)	(-4.012)	(-3.472)
$r^{(22)-}$	-0.009*	-0.012*	-0.009	-0.004
	(-2.845)	(-2.314)	(-1.481)	(-0.467)
R^2	0.7664	0.8137	0.8030	0.7629
HRMSE	0.2168	0.1692	0.1699	0.1796

In-sample forecast



Out of sample forecast

Mincer-Zarnowitz R^2



Is leverage effect induced by jumps

HAR-CJ$^+$ regression

	1-day	1-week	2-weeks	1-month
c	0.232*	0.377*	0.505*	0.747*
	(5.774)	(6.217)	(6.418)	(6.736)
C	0.398*	0.265*	0.214*	0.165*
	(21.521)	(18.225)	(16.084)	(12.442)
C$^{(5)}$	0.366*	0.368*	0.346*	0.291*
	(13.889)	(11.697)	(9.750)	(7.327)
C$^{(22)}$	0.190*	0.291*	0.338*	0.390*
	(9.470)	(9.743)	(9.059)	(8.875)
J$^+$	0.044*	0.018*	0.016*	0.013*
	(6.099)	(3.264)	(3.000)	(2.538)
J$^-$	0.074*	0.040*	0.039*	0.027*
	(6.909)	(6.833)	(6.658)	(5.351)
J$^{(5)}$	0.009*	0.012*	0.010*	0.010
	(2.645)	(2.724)	(2.028)	(1.799)
J$^{(22)}$	0.005	0.007	0.009*	0.014*
	(1.845)	(1.875)	(2.026)	(2.242)
-				
-				
-				
R^2	0.7543	0.8060	0.7960	0.7582
HRMSE	0.2201	0.1721	0.1722	0.1812

LHAR-CJ$^+$ regression

	1-day	1-week	2-week	1-month
c	0.442*	0.549*	0.661*	0.858*
	(10.724)	(9.277)	(8.531)	(7.778)
C	0.307*	0.201*	0.154*	0.116*
	(16.972)	(14.185)	(13.007)	(10.608)
C$^{(5)}$	0.369*	0.359*	0.332*	0.286*
	(13.885)	(11.237)	(9.144)	(6.777)
C$^{(22)}$	0.222*	0.319*	0.370*	0.415*
	(10.914)	(10.905)	(10.183)	(9.336)
J$^+$	0.044*	0.018*	0.015*	0.012*
	(6.176)	(3.182)	(2.819)	(2.395)
J$^-$	0.043*	0.019*	0.020*	0.011*
	(4.598)	(3.387)	(3.633)	(2.285)
J$^{(5)}$	0.011*	0.013*	0.011*	0.010
	(3.372)	(3.110)	(2.254)	(1.914)
J$^{(22)}$	0.005*	0.008*	0.010*	0.014*
	(2.200)	(2.104)	(2.203)	(2.337)
r^-	-0.007*	-0.005*	-0.004*	-0.003*
	(-9.772)	(-10.057)	(-7.804)	(-5.341)
$r^{(5)-}$	-0.008*	-0.006*	-0.008*	-0.007*
	(-4.409)	(-3.068)	(-4.020)	(-5.341)
$r^{(22)-}$	-0.009*	-0.012*	-0.008	-0.004
	(-2.844)	(-2.315)	(-1.484)	(-0.467)
R^2	0.7664	0.8137	0.8030	0.7629
HRMSE	0.2168	0.1692	0.1698	0.1796

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 $\{p_j(n_q)\}_{q=1,2,\cdots,M}$

- Realized Volatility (RV) approach affords estimation of the variance of asset j through

$$RV_{j,t} \equiv \sum_{q=1}^{M} r_{j,q}^2 \xrightarrow{M\rightarrow\infty} \int_0^t \sigma_j^2(\omega)d\omega \equiv IV_{j,t}$$

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.

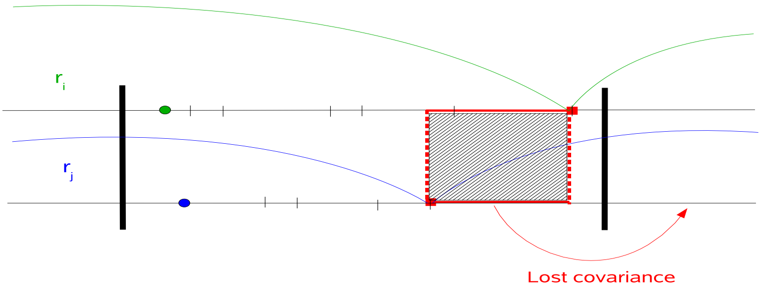
- Standard way: choose a time interval, construct an artificially regularly spaced time series (say 5-min) and take the sum of contemporaneous cross products.

- But, as for the RV, the presence of market microstructure can induce significant bias in standard RCov. The so called "**Epps effect**".

The Epps Effect

The microstructure effects responsible for this bias in the standard RC is the so called non-synchronous trading effect (Lo and MacKinlay 1990) which induces a bias toward zero in the standard interpolated measure:

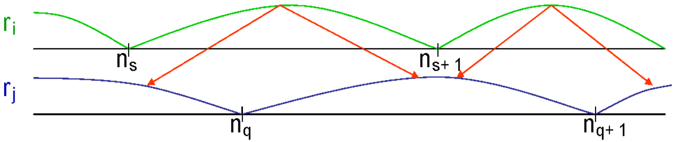
lost portion of covariance



RCov Tick-by-Tick: Hayashi & Yoshida

Hayashi and Yoshida (2005) estimator: does not resort on the construction of a regular grid.

Intuition



i.e. summing all the cross products of tick-by-tick returns with a non zero overlap

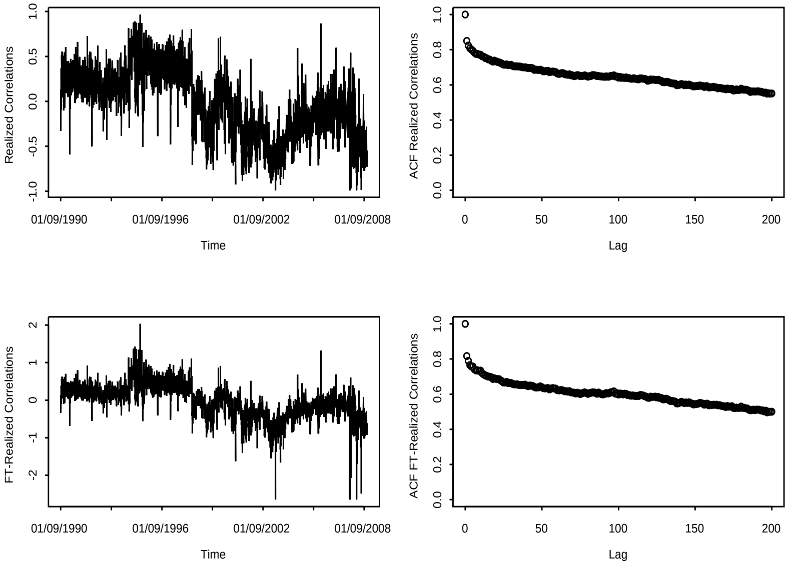
$$HY_t = \sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} r_{i,s}\, r_{j,q}\, I(\delta_{q,s} > 0)$$

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

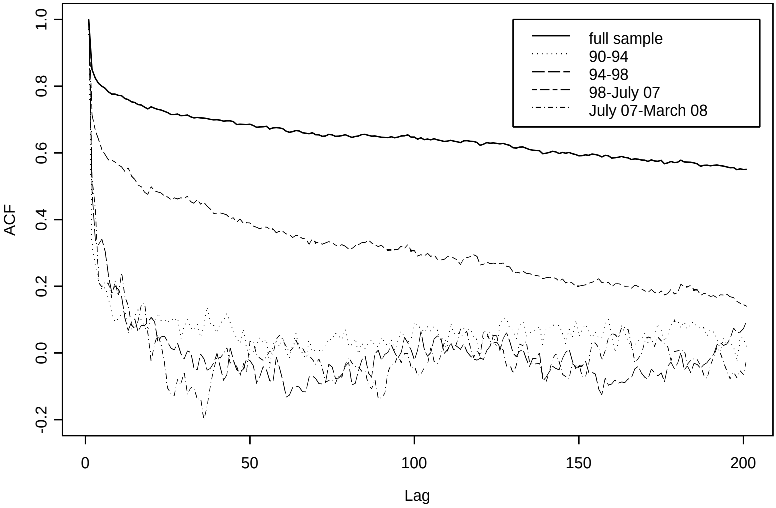
$$\delta_{q,s} = \max(0, \min(n_{i,s+1}, n_{j,q+1} - \max(n_{i,s}, n_{j,q}))$$

being the overlap between any two tick-by-tick returns $r_{i,s}$ and $r_{j,q}$.

Realized Correlations - Empirical Results



Realized Correlations - Empirical Results



Tree Structured HAR

- Let $\{\widetilde{RC}\}_{t\geq 1}$ be the daily Fisher-transformed (FT) series of the tick-by-tick realized correlations $\{RC\}_{t\geq 1}$,
- We then model the series $\{\widetilde{RC}\}_{t\geq 1}$ as:

$$\widetilde{RC}_{t+1} = \mathbb{E}_t[\widetilde{RC}_{t+1}] + \sigma_{t+1} U_{t+1} \tag{3}$$

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

$$\mathbb{E}_t[\widetilde{RC}_{t+1}] = \sum_{j=1}^{k} \left(a_j + b_j^{(d)} \widetilde{RC}_t + b_j^{(w)} \widetilde{RC}_t^{(w)} + b_j^{(m)} \widetilde{RC}_t^{(m)}\right) I_{[\mathbf{X}_t^{\text{pred}} \in \mathcal{R}_j]}$$

$$\sigma_{t+1}^2 = \sum_{j=1}^{k} \sigma_j^2 I_{[\mathbf{X}_t^{\text{pred}} \in \mathcal{R}_j]}, \ \sigma_j^2 > 0, j = 1, \ldots, k$$

  - $\theta = (a_j, b_j^{(d)}, b_j^{(w)}, b_j^{(m)}, \sigma_j^2 : j = 1, \ldots, k)$ 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)

Predictor Space

Regimes are characterized by partition cells $\mathcal{R}_j$ of the relevant predictor space G of $\mathbf{X}_t^{\text{pred}}$:

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

The relevant predictor variables in $\mathbf{X}_t^{\text{pred}}$ 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.

S&P-US Bond Correlation Regimes

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

Regime structure	Local parameters				
$\mathcal{R}_j$	a_j	$b_j^{(d)}$	$b_j^{(w)}$	$b_j^{(m)}$	σ_j^2
$R_{\text{S\&P 500}} \leq -0.935$, $t \leq$ July 2007	-0.0499 (0.0140)	0.3685 (0.0531)	0.3399 (0.0913)	0.2895 (0.1026)	0.0638 (0.0060)
$R_{\text{S\&P 500}} > -0.935$, $t \leq$ December 1993	0.0649 (0.0164)	0.1042 (0.0297)	0.3419 (0.0477)	0.2741 (0.0898)	0.0219 (0.0020)
$R_{\text{S\&P 500}} > -0.935$, December 1993 $< t \leq$ July 2007	0.0069 (0.0036)	0.2125 (0.0199)	0.4127 (0.0291)	0.3290 (0.0303)	0.0407 (0.0018)
$t >$ July 2007	-0.2432 (0.0627)	0.3237 (0.0546)	0.2643 (0.0647)	-0.0416 (0.1223)	0.1468 (0.0271)

Tree HAR In-sample estimation results

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

→ 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

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

$\longrightarrow$ The tree-HAR is always among the best models and the best for R^2.

Forecasting results: multi-period forecasts II

one week horizon

Model	MAE	MSE	R^2
AR(1)	0.3249 (0)	0.1777 (0)	0.1555
ARMA(1,1)	0.1895 (0.019)	0.0753 (0.004)	0.3603
ARIMA(1,1,1)	**0.1874** (**0.432**)	0.0721 (0.034)	0.3731
Tree-AR(1)	0.2629 (0)	0.1553 (0)	0.0057
HAR	0.1874 (0.479)	0.0731 (0.041)	0.3649
RS-HAR	0.1888 (0.025)	0.0743 (0.033)	0.3633
Tree-HAR	**0.1844** (**0.891**)	**0.0702** (**0.860**)	**0.4129**

one month horizon

Model	MAE	MSE	R^2
AR(1)	0.4514 (0)	0.3111 (0)	0.0199
ARMA(1,1)	0.2422 (0)	0.1175 (0)	0.1549
ARIMA(1,1,1)	0.2265 (0.002)	0.1012 (0.026)	0.1919
Tree-AR(1)	0.2962 (0)	0.1878 (0)	0.0049
HAR	0.2243 (0.049)	0.1023 (0.022)	0.1845
RS-HAR	0.2304 (0.001)	0.1072 (0.001)	0.1840
Tree-HAR	**0.2163** (**0.781**)	**0.0960** (**0.545**)	**0.2394**

$\longrightarrow$ Tree-HAR model consistently outperform alternative ones.

$\longrightarrow$ Gains are, in most cases, statistically significant (especially at one month).