Comparing Predictive Accuracy Under Long Memory - With an Application to Volatility Forecasting -

Robinson Kruse[a,b], Christian Leschinski[c] and Michael Will[c]

University of Groningen[a] and CREATES, Aarhus University[b], University Hannover[c]

December 8, 2015

Introduction

- The test of Diebold and Mariano [1995] (DM) is one of the standard tools used for forecast evaluation.
- Reduced form approach that only imposes assumptions on the loss differential series: **stationarity** and **weak dependence**.
- We focus on the behavior of DM tests under long memory and propose robust extensions.
- Monte Carlo comparison of the memory autocorrelation consistent (MAC) of Robinson [2005] and the extended fixed-b (EFB) approach of McElroy and Politis [2012], (cf. also Kiefer and Vogelsang [2005]).
- Re-evaluation of recent findings in literature on forecasting realized variance.

Outline

1. Concept of long memory
2. Conventional DM tests using HAC and fixed-b estimators
3. Long memory in forecast error loss differentials
    - Transmission of long memory from forecasts or the forecast objective to loss differentials
    - Performance of DM test under long memory in loss differential series
4. Long memory robust tests
    - MAC estimator
    - Extended fixed-b approach
5. Monte Carlo results
6. Application to volatility forecasting

Long Memory

- Long memory processes are characterized by strong intertemporal dependence.
- A time series x_t is a long memory series if it has a spectral density with power law $f(\lambda) \sim g_x \lambda^{-2d_x}$ as $\lambda \to 0$, or if its autocovariance function $\gamma_x(\tau)$ is $\gamma_x(\tau) \sim G_x \tau^{2d_x - 1}$, for $\tau \to \infty$. (The autocorrelation function of an AR(1) is $\rho(\tau) = \phi^{|\tau|}$.)
- Properties of the processes depend on the value of d.
- Many applications in economics and finance: implied and realized Volatility, squared and absolute returns, interest rate spreads, inflation rates, unemployment rates, trading volumes, sales numbers,...

Diebold-Mariano test

- Forecast error loss differentials are given by $z_t = g(e_{1t}) - g(e_{2t})$, where $g(\cdot) \geq 0$ is a loss function and $e_{it} = y_t - \widehat{y}_{it}$ are the forecast errors for $i = 1, 2$. Let $\mu_z = E(z_t)$ and $|\mu_z| < \infty$
- $H_0 : \mu_z = 0$ against $H_1 : \mu_z \neq 0$.
- Main assumption: z_t is a weakly stationary linear short memory process.
- Let $V = \lim_{T\to\infty} \mathrm{Var}\left(T^a \left[\bar{z} - \mu_z\right]\right)$ denote the long-run variance.

$$t_{DM} = T^a \frac{\bar{z}}{\sqrt{\widehat{V}}}$$

with $a = 1/2 - d$ and $0 \leq d < 1/2$.

Conventional approach: HAC

- Usually, V is estimated using HAC estimators (c.f. Diebold [2015]):

$$\widehat{V}_{HAC} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{B}\right)\widehat{\gamma}_j,$$

  where $k(\cdot)$ is a user-chosen kernel function, $\widehat{\gamma}_j$ denotes the j-th sample autocovariance, and B denotes the bandwidth.

- Under short memory $\widehat{V} \xrightarrow{p} V$, if $b = B/T \to 0$ as $T \to \infty$. Then we have

$$t_{HAC} = T^{\frac{1}{2}} \frac{\bar{\bar{z}}}{\sqrt{\widehat{V}_{HAC}}} \Rightarrow \mathcal{N}(0,1)\ .$$

Fixed-b approach

- Choi and Kiefer (2010) suggest the Fixed-b method (c.f. Kiefer and Vogelsang 2005) for the DM test (see also Patton 2015).

- Now assume instead that B/T approaches a fixed constant $b \in (0, 1]$ as $T \to \infty$ (c.f. Kiefer and Vogelsang 2005).

- This results in $\widehat{V}(k, b) \Rightarrow VQ(k, b)$ and in case of the *Bartlett* kernel:

$$
\begin{aligned}
t_{FB} &\Rightarrow \frac{W(1)}{\sqrt{Q(k, b)}}, \\
Q(k, b) &= \frac{2}{b}\left(\int_0^1 \widetilde{W}(r)^2 \mathrm{d}r - \int_0^{1-b} \widetilde{W}(r+b)\widetilde{W}(r)\mathrm{d}r\right),
\end{aligned}
$$

with $\widetilde{W}(r) = W(r) - rW(1)$ denoting a standard Brownian bridge.

Quadratic forecast error loss function

- Under quadratic loss functions we can rewrite the loss differential as follows:

$$
\begin{aligned}
z_t = e_{1t}^2 - e_{2t}^2 &= (y_t - \hat{y}_{1t})^2 - (y_t - \hat{y}_{2t})^2 \\
&= \hat{y}_{1t}^2 - \hat{y}_{2t}^2 - 2y_t(\hat{y}_{1t} - \hat{y}_{2t}).
\end{aligned} \tag{1}
$$

- Setting $y_t = y_t^* + \mu_y$ and $\hat{y}_{it} = \hat{y}_{it}^* + \mu_i$, we get

$$
\begin{aligned}
z_t = &- 2\left[y_t^*(\mu_1 - \mu_2) + \hat{y}_{1t}^*(\mu_y - \mu_1) + \hat{y}_{2t}^*(\mu_y - \mu_2)\right] \\
&- 2\left[y_t^*(\hat{y}_{1t}^* - \hat{y}_{2t}^*)\right] + \hat{y}_{1t}^{*2} - \hat{y}_{2t}^{*2} + const.
\end{aligned} \tag{2}
$$

- From (2) we can determine the memory of z_t using results of Chambers [1998] and Leschinski [2015].

Assumptions

Assumption 1 (Long memory)

The time series y_t, $\hat{y}_{1t}$, $\hat{y}_{2t}$, x_t have long memory according to Definition 1 of orders d_y, d_1, d_2 and d_x with finite expectations $E(y_t) = \mu_y$, $E(\hat{y}_{1t}) = \mu_1$, $E(\hat{y}_{2t}) = \mu_2$ and $E(x_t) = \mu_x$, respectively.

Assumption 2 (No common long memory)

If $a_t, b_t \sim LM(d_x)$, then $a_t - \psi_0 - \psi_1 b_t \sim LM(d_x)$ for all $\psi_0 \in \mathbb{R}, \psi_1 \in \mathbb{R}$ and $a_t, b_t \in \{y_t, \hat{y}_{1t}, \hat{y}_{2t}\}$.

Long memory in forecast error loss differentials

Proposition 1

Under Assumptions 1 and 2, we have for the quadratic forecast error loss differential: $z_t \sim LM(d_z)$,

$$d_z = \begin{cases} \max\{d_y,\ d_1,\ d_2\}, & \text{if } \mu_1 \neq \mu_2 \neq \mu_y \\ \max\{d_1,\ d_2\}, & \text{if } \mu_1 = \mu_2 \neq \mu_y \\ \max\{2d_1 - 1/2,\ d_2,\ d_y\}, & \text{if } \mu_1 = \mu_y \neq \mu_2 \\ \max\{2d_2 - 1/2,\ d_1,\ d_y\}, & \text{if } \mu_1 \neq \mu_y = \mu_2 \\ \max\big\{2\max\{d_1,\ d_2\} - 1/2,\ d_y + \max\{d_1,\ d_2\} - 1/2,\ 0\big\}, & \text{if } \mu_1 = \mu_2 = \mu_y. \end{cases}$$

- (Un)Biasedness plays an important role
- Only under unbiasedness of both forecasts, a short memory series z_t might result when $d_1, d_2 \leq 0.25$

Properties under common long memory

Assumption 3 (Linear common long memory)

If $a_t, b_t \sim CLM(d_x, d_x - b)$, then they can be represented as

$$\begin{aligned} y_t &= \beta_y + \xi_y x_t + \eta_t, \quad \text{for } \; a_t, b_t = y_t \\ \hat{y}_{it} &= \beta_i + \xi_i x_t + \varepsilon_{it}, \quad \text{for } \; a_t, b_t = \hat{y}_{it}, \end{aligned}$$

with $\xi_y, \xi_i \neq 0$, $E(\eta_t) = 0$, $\eta_t \sim LM(d_\eta)$ and $E(\varepsilon_{it}) = 0$, $\varepsilon_{it} \sim LM(d_{\varepsilon_i})$ and $0 \leq d_\eta, d_{\varepsilon_i} < d_x < 1/2$ for $i = 1, 2$.

(Simplifying assumption)

Remarks

- There are many possible CLM situations

  CLM between one of the forecasts and the forecast objective
  CLM between both forecasts but to the forecast objective
  CLM between both forecasts and the forecast objective

- In addition, one can consider in each case (un)biasedness

  One of the forecasts is biased
  Both are biased
  Both are unbiased

  $\Rightarrow$ all this leads to a (too) complex situation, look at special cases

Result for biased forecasts

Proposition 2

Under Assumptions 1 and 3, the forecast error loss differential is $z_t \sim LM(d_z)$, where

$$d_z \geq \begin{cases} d_1, & \text{if } \mu_1 \neq \mu_y \\ d_2, & \text{if } \mu_2 \neq \mu_y \\ d_y, & \text{if } \mu_1 \neq \mu_2. \end{cases}$$

- Long memory is transmitted whenever forecasts are biased.

Results under unbiasedness and further restrictions

Proposition 3

Let b_c denote a constant such that $d_x > b_c > 0$, for all $c = 1, ..., 5$. Under Assumptions 1 and 3, with $\mu_x \neq 0$, if $\mu_y = \mu_1 = \mu_2$ and $\xi_a = \xi_b$, then $z_t \sim LM(d_z)$, with

$$d_z = \begin{cases} \max\{d_x + d_2 - 1/2,\ 2d_2 - 1/2,\ 2d_x - 1/2,\ d_{\varepsilon_1}\}, & \text{if } y_t, \hat{y}_{1t} \sim CLM(d_x, d_x - b_1) \\ \max\{d_x + d_1 - 1/2,\ 2d_1 - 1/2,\ 2d_x - 1/2,\ d_{\varepsilon_2}\}, & \text{if } y_t, \hat{y}_{2t} \sim CLM(d_x, d_x - b_2) \\ \max\{d_{\varepsilon_1},\ d_{\varepsilon_2}\}, & \text{if } \hat{y}_{1t}, \hat{y}_{2t} \sim CLM(d_x, d_x - b_3) \\ \max\{d_{\varepsilon_1},\ d_{\varepsilon_2}\}, & \text{if } y_t, \hat{y}_{1t} \sim CLM(d_x, d_x - b_4) \\ & \text{and } y_t, \hat{y}_{2t} \sim CLM(d_x, d_x - b_5). \end{cases}$$

- Short memory only when $d_{\varepsilon_1} = d_{\varepsilon_2} = 0$.

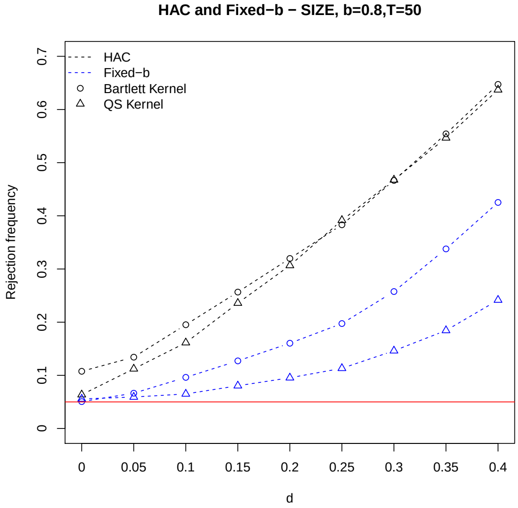
Asymptotic rejection frequency under long memory (z_t)

Proposition 4

For $z_t \sim LM(d_z)$ with $d_z \in (0, 1/4) \cup (1/4, 1/2)$, the asymptotic size of the t_{HAC}-statistic equals unity as $T \to \infty$.

- Distribution of t_{HAC} changes at $d = 1/4$ from normality ($0 \leq d < 1/4$) to non-normal Rosenblatt ($1/4 < d < 1/2$), see Abadir et al. [2009].

Small sample rejection frequency under long memory (z_t)



MAC estimator

- A key result for the MAC of Robinson [2005] estimator is that:

$$\mathrm{Var}\left(T^{1/2-d}\bar{z}\right) \to b_0 p(d)$$

$$p(d) = \begin{cases} \frac{2\Gamma(1-2d)\sin(\pi d)}{d(1+2d)} & \text{if } d \neq 0\ , \\ 2\pi & \text{if } d = 0\ . \end{cases}$$

- d is estimated with an estimator that fulfills $\widehat{d} - d = o_p(1/\log T)$ such as the local Whittle estimator in Robinson [1995] and b_0 can be estimated by:

$$\widehat{b}_m(\widehat{d}) \;=\; m^{-1}\sum_{j=1}^{m}\lambda_j^{2\widehat{d}} I_T(\lambda_j).$$

MAC estimator

- MAC estimator is then given by:

$$\widehat{V}(\widehat{d}, m_d, m) = \widehat{b}_m(\widehat{d})p(\widehat{d}).$$

- Under the necessary regularity conditions $\widehat{d} \xrightarrow{p} d$, $\widehat{b}_m(\widehat{d}) \xrightarrow{p} b_0$:

$$t_{MAC} = T^{1/2-\widehat{d}} \frac{\bar{z}}{\sqrt{\widehat{V}(\widehat{d}, m_d, m)}} \Rightarrow \mathcal{N}(0,1)\ .$$

- MSE-optimal bandwidth choice $m = [T^{4/5}]$ is independent of d.

Extended fixed-b approach

- McElroy and Politis [2012] extend the fixed-b approach to long memory. They obtain the following result:

$$t_{EFB} = T^{1/2}\frac{\bar{z}}{\sqrt{\widehat{V}(k,b)}} \Rightarrow \frac{W_d(1)}{\sqrt{Q(k,b,d)}} ,$$

  where $W_d(r)$ is a fractional Brownian motion.

- Note that $Q(k,b,d)$ depends on the first and second derivatives of k, the bandwidth fraction b and the memory parameter d.

- A plug-in estimator of d is needed to obtain critical values.

Monte Carlo: Size

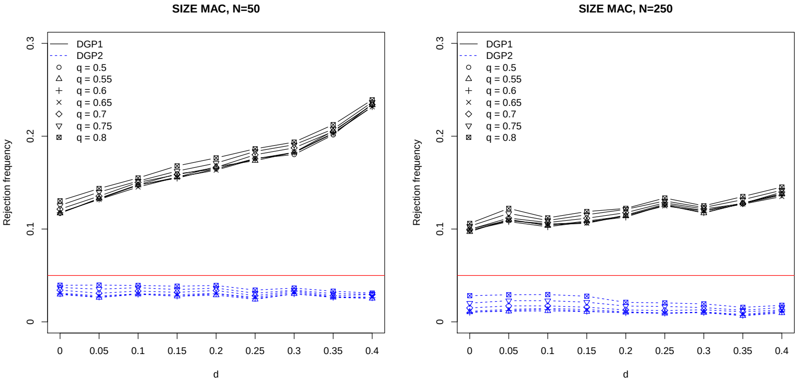


Figure: Size of the t_{MAC} statistic with bandwidth $q = m$. DGP1 is $FI(d)$ and DGP2 is $ARFI(0.6, d)$.

Monte Carlo: Size

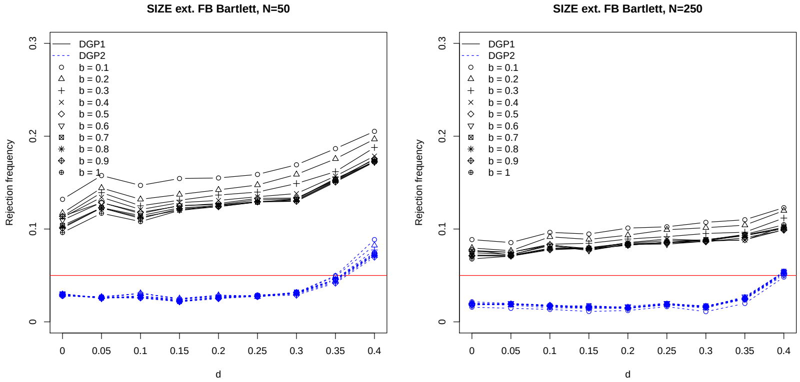


Figure: Size of the t_{EFB} statistic using the Bartlett kernel. DGP1 is $FI(d)$ and DGP2 is $ARFI(0.6, d)$.

Monte Carlo: Size

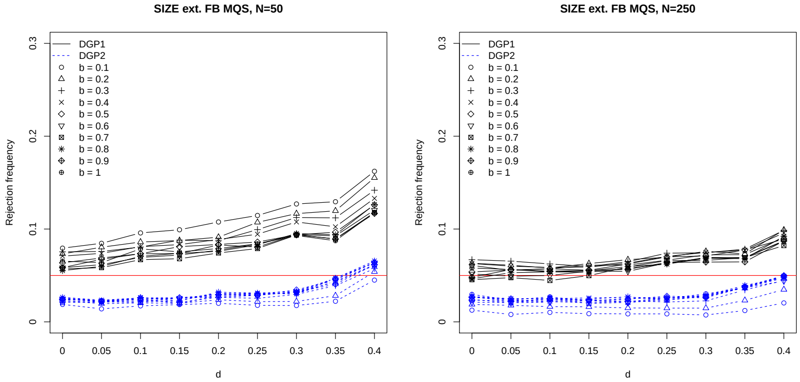


Figure: Size of the t_{EFB} statistic using the MQS kernel. DGP1 is $FI(d)$ and DGP2 is $ARFI(0.6, d)$.

Monte Carlo: Potential Power Loss

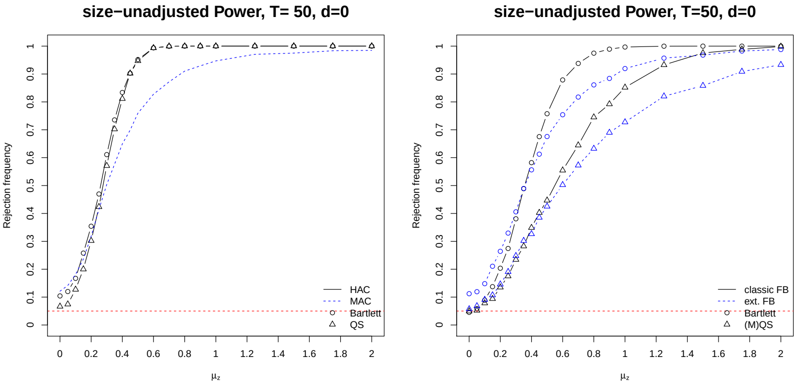


Figure: Power comparison between t_{HAC} and t_{MAC} (left) and t_{FB} and t_{EFB} (right) when $d = 0$ and $T = 50$.

Monte Carlo: Potential Power Loss

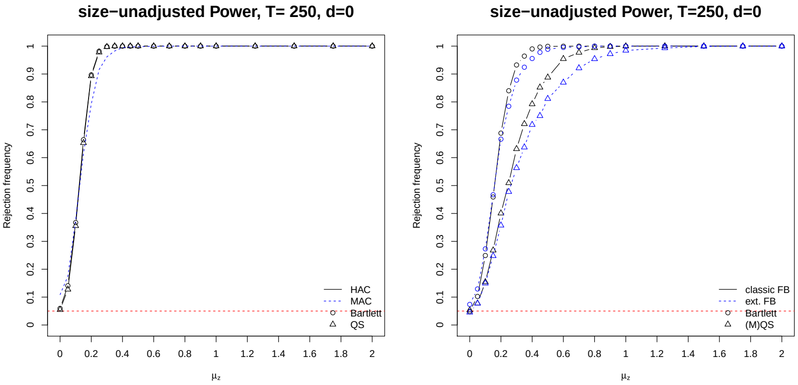
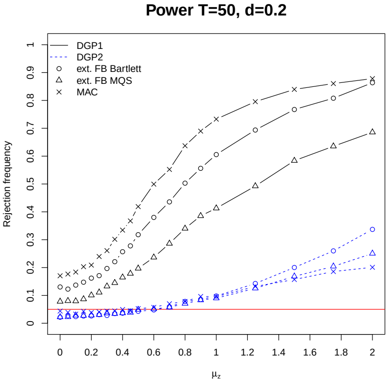


Figure: Power comparison between t_{HAC} and t_{MAC} (left) and t_{FB} and t_{EFB} (right) when $d = 0$ and $T = 250$.

Monte Carlo: Power



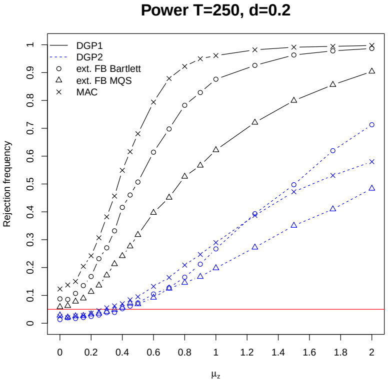


Figure: Power comparison of t_{MAC} and t_{EFB} with Bartlett kernel and with MQS kernel, when $d = 0.2$.

Monte Carlo: Size-adjusted Power (d known)

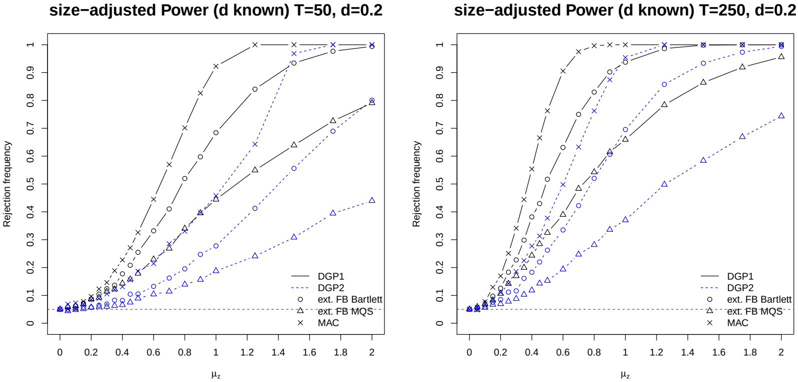
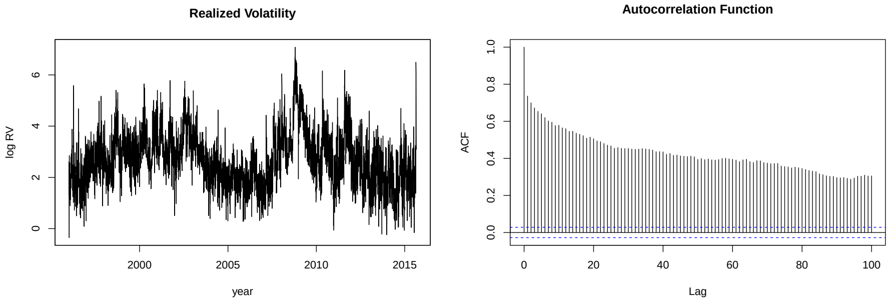


Figure: Size-adjusted Power comparison of t_{MAC} and t_{EFB} with Bartlett kernel and with MQS kernel, when $d = 0.2$ is known.

Volatility prediction

- Volatility forecasting is a typical application of long memory models (e.g. Deo et al. [2006], Martens et al. [2009] and Chiriac and Voev [2011]).
- Shift from GARCH to HAR-RV of Corsi [2009] and its extensions. We re-evaluate recent results from the HAR-RV literature.
- 5-minute log-returns of the S&P 500 index from January 2, 1996 to August 31, 2015 including close-to-open returns. From Thomson Reuters Tick History Database.
- Following Andersen et al. [2001] and Barndorff-Nielsen and Shephard [2002] the daily realized variance is defined as

$$RV_t = \sum_{j=1}^{N} r_{t,j}^2 \ .$$



- Typical features of a long memory time series. Local whittle estimates of the memory parameter between 0.5 and 0.6.
- Test of Qu [2011] does not indicate spurious long memory.
- RV_t contains a measurement error and multi-step forecasts of stock variables induce short memory dynamics. We therefore use the local polynomial Whittle plus noise (LPWN) estimator of Frederiksen et al. [2012].

Separation of Continuous and Jump Components

- Log-realized variance provides better approximation to the normal distribution (cf. Andersen et al. [2001]).
- HAR-RV-model of Corsi [2009] explains realized variance RV_t by an autoregression involving overlapping averages of past realized variances.

$$\ln RV_t^{(h)} = \alpha + \gamma_{22} \ln RV_{t-h}^{(22)} + \gamma_5 \ln RV_{t-h}^{(5)} + \gamma_1 \ln RV_{t-h}^{(1)} + \varepsilon_t \,, \quad (3)$$

where $RV_t^{(M)} = \frac{22}{M} \sum_{j=0}^{M-1} RV_{t-j}$ and $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

Separation of Continuous and Jump Components

- Andersen et al. [2007]: jump components in realized volatility.

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t) \ ,$$

  - $\mu(t)$ is the drift with locally bounded variation,
  - $\sigma(t)$ is a strictly positive stochastic volatility process,
  - $q(t)$ takes the value $dq(t) = 1$ if a jump is realized,
  - $\kappa(t)$ determines the size of discrete jumps.

- Quadratic variation of the cumulative return process decomposed into integrated volatility plus the sum of squared jumps.

$$[r]_t^{t+h} = \int_t^{t+h} \sigma^2(s)ds + \sum_{t<s\leq t+h} \kappa^2(s) \ .$$

Separation of Continuous and Jump Components

- Corsi et al. [2010] introduce TBPV to measure continuous volatility component

$$TBPV(r)_t = \mu_1^{-2} \sum_{j=2}^{N} |r_{t,j}||r_{t,j-1}| \mathbb{I}(|r_{t,j}|^2 \leq \zeta_j) \mathbb{I}(|r_{t,j-1}|^2 \leq \zeta_{j-1}) \; ,$$

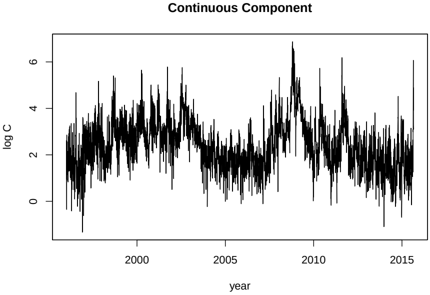
where ζ_j is a threshold function, $\mu_1 = \sqrt{2/\pi}$.

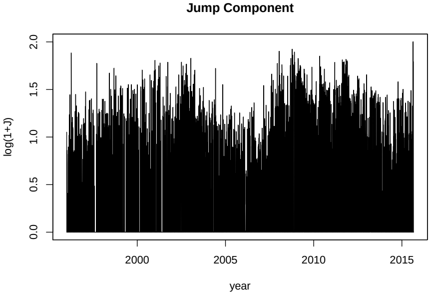
- For $N \to \infty$, realized volatility can be decomposed into a continuous integrated volatility component C_t and the jump component J_t.

$$TBPV(r)_t \to \int_t^{t+1} \sigma^2(s) ds$$

$$J_t = \max\left\{RV_t - TBPV_t, 0\right\} \mathbb{I}(\text{C-Tz} > 3.09)$$

Separation of Continuous and Jump Components





- HAR-RV-TCJ model of Corsi et al. [2010] is given by

$$\begin{aligned}\ln RV_t^{(h)} &= \alpha + \gamma_{22} \ln C_{t-h}^{(22)} + \gamma_5 \ln C_{t-h}^{(5)} + \gamma_1 \ln C_{t-h}^{(1)} \\ &+ \delta_{22} \ln\left(1 + J_{t-h}^{(22)}\right) + \delta_5 \ln\left(1 + J_{t-h}^{(5)}\right) \\ &+ \delta_1 \ln\left(1 + J_{t-h}^{(1)}\right) + \varepsilon_t \ .\end{aligned}$$

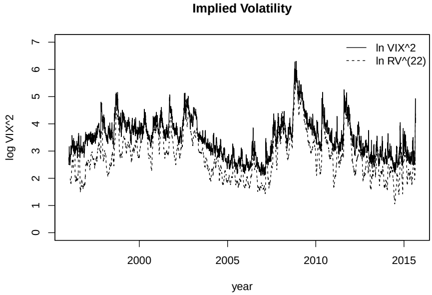
Separation of Continuous and Jump Components

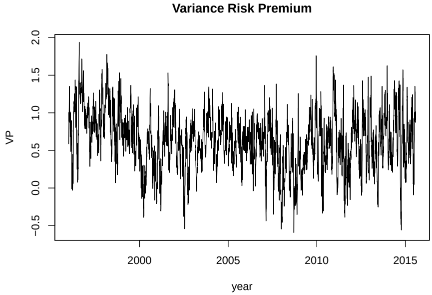
Table: HAR-RV vs HAR-RV-TCJ

									t_{MAC}			t_{EFB}			
	$\bar{z}/\hat{\sigma}_z$	MSE_1	MSE_2	$\hat{d}_{LW}$	$\hat{d}_{LPWN}$	t_{DM}	t_{HAC}	t_{FB}	0.7	0.75	0.8	0.2	0.4	0.6	0.8
$h=1$	0.122	0.409	0.375	**0.094***	0.127	**6.932**	**7.631**	**3.995**	**3.243**	**3.144**	**3.091**	**3.995**	**4.068**	**4.468**	**4.947**
												(2.610)	(3.154)	(3.693)	(4.228)
$h=5$	0.092	0.263	0.247	**0.072**	0.009	**3.666**	**3.790**	**2.789**	**3.620**	**3.853**	**4.277**	**2.789**	**3.981**	**5.093**	**5.848**
												(2.050)	(2.522)	(2.975)	(3.386)
$h=22$	0.045	0.292	0.285	**0.359***	**0.343***	0.776	0.912	0.666	0.140	0.152	0.171	0.666	0.925	1.064	1.164
						(1.645)	(1.645)	(2.092)		(1.645)		(4.701)	(5.551)	(6.413)	(7.281)

- For $h = 1$ we obtain $\hat{d}_{HAR-RV} = 0.096$, $\hat{d}_{HAR-RV-TCJ} = 0.070$ and $\hat{d}_{diff} = 0.094$.
- Tests uniformly confirm that HAR-RV-TCJ is superior.

Predictive Ability of the VIX





- Chernov [2007], variance risk premium: $VP_t = \ln VIX_t^2 - \ln RV_{t+22}^{22}$.
- Fractional cointegration: $\hat{d}_{\ln VIX} \approx \hat{d}_{\ln RV^{(22)}} \approx 0.8$ and $\hat{d}_{VP} \approx 0.2$, (cf. Nielsen [2007], Bollerslev et al. [2013],...).
- Inclusion of $\ln(VIX_t^2/12)$ in HAR-RV model: Becker et al. [2007], Becker et al. [2009] and Busch et al. [2011].

Predictive Ability of the VIX

Table: Predictive ability of the VIX for future RV and Jumps

	$\bar{z}/\hat{\sigma}_z$	$MSE1$	$MSE2$	$\hat{d}_{LW}$	$\hat{d}_{LPWN}$	t_{DM}	t_{HAC}	t_{FB}	t_{MAC} 0.7	t_{MAC} 0.75	t_{MAC} 0.8	t_{eFB} 0.2	t_{eFB} 0.4	t_{eFB} 0.6	t_{eFB} 0.8
HAR vs. HAR-VIX	0.135	0.292	0.269	**0.219***	**0.234***	**2.968**	**3.032**	**2.494**	0.929	1.038	1.188	2.494	2.754	2.985	2.849
												(3.404)	(4.064)	(4.750)	(5.388)
HAR-TCJ vs. HAR-TCJ-VIX	0.109	0.285	0.268	**0.175***	0.138	**2.421**	**2.455**	**2.097**	1.397	1.610	**1.892**	2.097	2.503	2.889	2.724
												(2.610)	(3.154)	(3.693)	(4.228)
HAR-TCJ-L vs. HAR-TCJ-L-VIX	0.082	0.282	0.269	**0.182***	0.163	**1.784**	**1.786**	1.819	0.889	1.016	1.192	1.819	2.153	2.430	2.317
						(1.645)	(1.645)	(2.092)		(1.645)		(3.404)	(4.064)	(4.750)	(5.388)

- Long memory between 0.13 and 0.24.
- Persistence in the loss differentials decreases for more complex HAR-RV-type models.
- Superior predictive ability of models including the VIX vanishes if robust tests are used.

Conclusion

- Conventional DM test is invalidated by long memory in loss differentials.
- DM statistic can readily be extended by memory robust long-run variance estimators.
- Extended fixed-*b* approach of McElroy and Politis [2012] with MQS kernel provides best size control.
- In terms of power the Bartlett kernel is superior.
- MAC suffers disproportionately under imprecise estimates of d. However, the power of the MAC is superior.
- Empirical findings highlight the importance of long-memory robust tests for forecast comparisons.