Modeling Tick-by-Tick Realized Correlations

Francesco Audrino[*,a,b], Fulvio Corsi[a,b]

[a] *University of St. Gallen*
[b] *University of Lugano*

Abstract

A tree-structured heterogeneous autoregressive (tree-HAR) process is proposed as a simple and parsimonious model for the estimation and prediction of tick-by-tick realized correlations. The model can account for different time and other relevant predictors' dependent regime shifts in the conditional mean dynamics of the realized correlation series. Testing the model on S&P 500 Futures and 30-year Treasury Bond Futures realized correlations, empirical evidence that the tree-HAR model reaches a good compromise between simplicity and flexibility is provided. The model yields accurate single- and multi-step out-of-sample forecasts. Such forecasts are also better than those obtained from other standard approaches, in particular when the final goal is multi-period forecasting.

Key words: High frequency data; Realized correlation; Stock–bond correlation; Tree-structured models; HAR; Regimes.

1. Introduction

The correlation of stock and bond returns plays a pivotal role for portfolio managers, risk analysts and financial researchers, being a key ingredient to portfolio diversification and risk management. It is well known that the stock-bond correlation changes over time. Characterizing these time-variations has relevant implications for understanding the economics of the joint stock-bond relation while an accurate forecast of stock-bond correlation may have important practical applications in asset allocation and risk management.

In the last twenty years, the increased performance of data base management systems allowed to create huge data bases consisting of all transactions observed in the market. Recently, the use of this kind of high frequency data has been advocated to improve the precision of asset volatility estimation, yielding to the so-called realized volatility (RV) approach proposed in a series of breakthrough papers by Andersen et al. (2001b, 2003), Barndorff-Nielsen and Shephard (2001, 2002a, 2002b, 2005), and Comte and Renault (1998). Regarding the realized volatility approach, the idea of employing high frequency data in the computation of covariances between two assets leads to the analogous concept of *realized covariance* (or covariation); for more details see Barndorff-Nielsen and Shephard (2004), Martens (2004), Hayashi and Yoshida (2005), Griffin and Oomen (2006), Palandri (2006), Sheppard (2006), and Voev and Lunde (2007). Recently, Corsi and Audrino (2008) proposed a modified tick-by-tick realized covariance

*Address for correspondence: University of St. Gallen, Institute of Mathematics and Statistics, Bodanstrasse 6, CH-9000 - St. Gallen, Switzerland. Tel: +41 71 224 2431. Fax: +41 71 224 2894.

Email addresses: `francesco.audrino@unisg.ch` (Francesco Audrino), `fulvio.corsi@lu.unisi.ch` (Fulvio Corsi)

estimator in cases where price arrival times are made imprecise by rounding in the reported time stamps, a typical situation for many financial data sets. *Realized correlations* are then constructed as quotients between realized covariances and products of realized standard deviations. This paper investigates only the fraction of daily realized correlations corresponding to the time when the markets are open (refer to (Andersen et al., 2007) for a more general discussion about how the night correlation, from the overnight returns, can be incorporated into our framework). A number of studies in the recent literature showed that using realized second moments the accuracy and performance of trading strategies and risk measures can be significantly improved. The idea that the whole information observed on the market is relevant for estimation and prediction of the unobserved asset returns correlations yielded to the use of the so-called tick-by-tick data.

Similar to the smooth transition heterogenous autoregressive technique proposed by McAleer and Medeiros (2008) which modeled the realized volatility of sixteen US stocks with good forecasting results, we propose a regime-dependent, tree-structured heterogeneous autoregressive (tree-HAR) model for the estimation and prediction of the tick-by-tick realized correlation series. The conditional mean dynamics of the realized correlation series follow local linear HAR processes and are subject to regime shifts depending on past values of certain relevant predictor variables, such as, for example, past returns, past realized volatilities or time. The local HAR processes are standard linear models where the explanatory variables are past realized correlations at three different horizons: daily, weekly, and monthly (for more details, see Corsi et al., 2006 and Corsi, 2009). This structure allows the model to take into account two important features exhibited by most real data realized correlation series: long memory and structural changes. Another feature of the tree-HAR model is that it belongs to the class of tree-structured threshold regime models, and can therefore be easily estimated and regimes can be interpreted in terms of relevant predictor variables; see for example, Audrino and Bühlmann (2001) and Audrino and Trojani (2006).

We test the accuracy of the tree-HAR model on the series of daily tick-by-tick realized correlations between S&P 500 and 30-year Treasury Bond futures for the time period 1990–2008. We collect empirical evidence that realized correlations constructed using tick-by-tick information show drastic regime shifts, supporting the evidence already found in other studies on classical correlations. The presence of structural breaks in the stock–bond correlations is already established in several works in the recent literature. Among these, Cappiello et al. (2006) provided statistical evidence of a structural break in stock–bond correlations due to the introduction of the Euro. Li (2002), Ilmanen (2003) and Christiansen and Ranaldo (2007) report that US stock–bond correlations went from positive to negative after 1997. Guidolin and Timmermann (2006) found empirical evidence that a four-state Markov regime-switching model is needed to capture the joint dynamics of US stock and bond returns. In their empirical study, Pastor and Stambaugh (2003) found that changes in stock–bond correlations are related to different levels of liquidity, and, therefore, allowing for structural breaks depending on a large set of predictor variables (for example, liquidity) may be relevant for improving forecasting accuracy.

We contribute to the literature on US stock–bond correlations by estimating local dynamics and incorporating structural breaks in a threshold-type model. The estimated tree-HAR model for daily US stock–bond realized correlations has four limiting regimes (endogenously estimated from the data) which can be interpreted as follows. First of all we find a structural break during the summer 2007. This break may be related to the US subprime financial crisis. The first three regimes refer to the time period before the summer 2007. The first regime is in reaction to US market crashes: in particular, the first regime is characterized by large negative past S&P 500 daily returns, the conditional mean dynamics of the realized correlations are highly persistent, and the volatility of realized correlations is large. The second and third regimes are both characterized by relatively positive (which, as usual in this context, means above the estimated threshold value) past S&P 500 daily returns, but for two different time periods. We identify a second structural break in time corresponding to the end of the year 1993. Since the common drivers of stock and bond prices should be the real interest rate and the expected inflation (see, for example, Ranaldo and Reynard, 2008), this structural break may be a consequence of the inversion in the 1993–94 period in the trend of the real interest rate and the target Fed funds rate (which affects market expected inflation because of the perceived Fed's superior information on future inflation) which after a long descending trend invert their tendency upwards. After the end of 1993 the persistence of the conditional mean dynamics and the volatility of the realized correlations increase significantly. Moreover, we also find that past individual realized volatilities are relevant predictors for future realized correlations.

We perform a series of out-of-sample tests for the superior predictive ability (SPA; see Hansen, 2005) of our model against a number of competitors using different goodness-of-fit statistics, to verify whether the greater flexibility allowed by the tree-HAR model (with a corresponding higher number of parameters to be estimated) has any value for forecasting. We empirically show that the tree-HAR model systematically outperforms the competitors, particularly when multi-period forecasts are considered.

The remainder of the paper is organized as follows: Section 2 proposes the tree-HAR process as a model for the estimation and forecast of realized correlations. Section 3 presents the empirical application to a bivariate series of S&P 500 and 30-year US Treasury Bond futures tick-by-tick data. Section 4 summarizes and concludes.

2. Modeling Realized Correlations

2.1. The Model

Empirical evidence on strong temporal dependence of realized correlations has been already shown in Andersen et al. (2001a, 2003), and Ferland and Lalancette (2006), among others. This evidence, together with our empirical results reported in Section 3, suggests that realized correlation series are best described by long-memory type of models.

Corsi (2009) and Corsi et al. (2006) recently proposed a class of "pseudo-long-memory" models called heterogeneous autoregressive (HAR) models, which, although not satisfying formally the long memory property, are able to successfully model the empirical long-memory behavior of financial variables in a simple and parsimonious way.

The basic idea was introduced in a study by Müller et al. (1997), where the long memory observed in the volatility was explained as the superimposition of only a few processes operating on different time scales. Corsi (2009) proposed a stochastic additive cascade of three different realized volatility components corresponding to the three main different time horizons present in the market: daily, weekly, and monthly. This stochastic volatility cascade leads to simple AR-type models in the realized volatility which also feature a consideration of realized volatilities defined over different time horizons (the HAR-RV models). Although the HAR models do not formally belong to the class of long-memory models, they are able to reproduce a memory decay which is almost indistinguishable from that observed in the empirical data.

The above-mentioned empirical evidence on the high degree of persistence of correlations suggests that the parsimonious HAR models could also be successfully applied to model the time series of realized correlations. Figure 1 shows the autocorrelogram of the empirical stock–bond correlations of our real data application of Section 3 matches well the one for correlations obtained simulating data from the HAR model with parameters similar to those estimated from real data.

A second important stylized fact which must be taken into account when building up a model for the realized correlation dynamics is the (possible) presence of structural breaks. Various studies in the recent

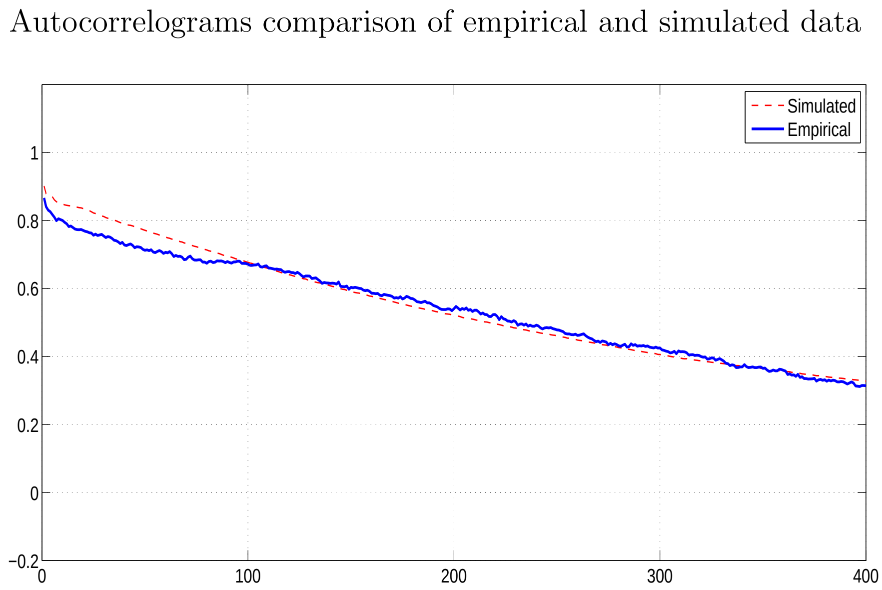

Figure 1: Comparison of autocorrelograms of correlations for the empirical US stock–bond data under investigation (solid) and for simulated data (dotted) from an HAR model with parameters estimated on the full US stock–bond real data sample.

literature on stock–bond correlations already report that stock–bond correlations went from positive to negative after 1997 (see, for example, Ilmanen, 2003). The reasons given to explain this pattern vary. One relates to market uncertainty and risk, introducing the "flight-to-quality" effect, which suggests the phenomenon of fleeing from stock to bond markets in times of worsening economic conditions (see for example, Ilmanen, 2003; or Connolly et al., 2005). Another explanation for the change of sign in stock–bond correlations relates to differences in inflation expectations or in the expectations of other macroeconomic variables (see for example, Li, 2002; or Christiansen and Ranaldo, 2007). Pastor and Stambaugh (2003) found that a kind of "flight-to-quality" effect appears in months with exceptionally low liquidity, that is months in which liquidity drops severely tend to be months in which stocks and fixed-income assets move in opposite directions. In two recent studies, Guidolin and Timmermann (2006) and Audrino and Trojani (2007) incorporated the possible regime shifts in the conditional dynamics of (realized) correlations using regime-dependent models. Guidolin and Timmermann (2006) analyzed the joint dynamics of US stock and bond returns using a Markov regime-switching model, and found empirical evidence of the presence of four different regimes.

Along these lines, we propose a tree-structured local HAR model for the dynamics of tick-by-tick realized correlations which is able to take into account the above-discussed stylized facts of realized correlation series: long-memory and structural breaks. Tree-structured models belong to the class of threshold regimes models, where regimes are characterized by some threshold for the relevant predictor variables. The class of tree-structured GARCH models was introduced by Audrino and Bühlmann (2001) in the financial volatility literature, and was generalized recently to capture simultaneous regime shifts in the first and second conditional dynamics of returns series, with good results for different forecasting applications (see for example, Audrino and Trojani, 2006). The constraint that conditional first and second moment dynamics must be subject to the same regimes are introduced to reduce computational costs and can be easily relaxed. Nevertheless, the resulting costs in the computational and model's complexity have not been found to be supported by a significant improvement in the accuracy of the forecasts.

Let $\{RC\}_{t\geq 1}$ be the daily tick-by-tick realized correlations computed from tick-by-tick realized covariance and tick-by-tick realized volatility measures, and let $\{\widetilde{RC}\}_{t\geq 1}$ be the daily Fisher-transformed (FT) series of the tick-by-tick realized correlations $\{RC\}_{t\geq 1}$. By considering Fisher-transformed correlations we do not have to impose any restriction on the parameters in the model to ensure the final estimates and forecasts to lie in the $[-1,1]$ interval. Note, however, that by performing the same tree-HAR analysis on the original correlations, we never got correlation estimates and predictions outside the $[-1,1]$ interval. The results of the analysis were qualitatively the same. In particular, the estimated regime structure was found to be exactly the same, with comparable local conditional first and second moment dynamics. The Fisher-transformed correlations are given by:

$$\widetilde{RC}_t = \frac{1}{2}\log\Big(\frac{1+RC_t}{1-RC_t}\Big),\ RC_t \in [-1,1].$$

We then model the series $\{\widetilde{RC}\}_{t\geq 1}$ as:

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

where $\{U_t\}_{t\geq 1}$ is a sequence of i.i.d. innovations following the distribution p_U with expected value 0 and variance 1, and $\mathbb{E}_t[\cdot]$ denotes (as usual) the conditional expectation given the information up to time t. The conditional dynamics of the FT correlations are given by:

$$\mathbb{E}_t[\widetilde{RC}_{t+1}] = \sum_{j=1}^{k}\big(a_j + b_j^{(d)}\widetilde{RC}_t + b_j^{(w)}\widetilde{RC}_t^{(w)} + b_j^{(m)}\widetilde{RC}_t^{(m)}\big)I_{[\mathbf{X}_t^{\text{pred}}\in\mathcal{R}_j]} \text{ and} \tag{2}$$

$$\sigma_{t+1}^2 = \sum_{j=1}^{k}\sigma_j^2 I_{[\mathbf{X}_t^{\text{pred}}\in\mathcal{R}_j]},\ \sigma_j^2>0, j=1,\ldots,k, \tag{3}$$

where $\theta = (a_j, b_j^{(d)}, b_j^{(w)}, b_j^{(m)}, \sigma_j^2 : j=1,\ldots,k)$ 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), and $\widetilde{RC}_t^{(w)}$ and $\widetilde{RC}_t^{(m)}$ are respectively the weekly and monthly FT-realized correlations, obtained as simple rolling averages of 5 respectively 22 daily FT-realized correlations. The choice of weekly and monthly averages is not restrictive. In fact, the model can be generalized to be of a AR(p)-type, where, as it has been already done in the literature (see, for example, Hillebrand and Medeiros, 2008), the number of past lags p is optimized using some selection criteria. Such studies found that the most relevant lags are exactly the lags 1,5, and 22. The regimes are characterized by partition cells $\mathcal{R}_j$ of the relevant predictor space G of $\mathbf{X}_t^{\text{pred}}$:

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

In our study, the relevant predictor variables in $\mathbf{X}_t^{\text{pred}}$ are past-lagged FT-realized correlations, and past-lagged realized volatilities and returns of the two instruments under investigation. All such predictor variables are considered at three different time horizons: daily, weekly, and monthly. We also consider time as an additional predictor variable.

To completely specify the conditional dynamics given in equations (2) and (3) of the FT-realized correlations, we determined the shape of the partition cells $\mathcal{R}_j$, which are admissible in the tree-HAR model. Similar to the standard classification and regression trees (CART) procedure (see Breiman et al., 1994), the only restriction we impose is that regimes must be characterized by (possibly high-dimensional) rectangular cells of the predictor space, with edges determined by thresholds on the predictor variables. Such partition cells are practically constructed using the idea of binary trees. Introducing this restriction has two major advantages: it allows a clear interpretation of the regimes in terms of relevant predictor variables, and also allows us to estimate the model using large-dimensional predictor spaces G.

As an illustration, in our empirical application on US stock–bond realized correlations presented in Section 3, the estimated partition cells for the time period before July 2007 are of the form:

$$
\begin{aligned}
\mathcal{R}_1 &= \{\mathbf{X}^{\text{pred}} : R_{\text{S\&P 500}} \leq -0.935\}, \\
\mathcal{R}_2 &= \{\mathbf{X}^{\text{pred}} : R_{\text{S\&P 500}} > -0.935 \text{ and } t \leq \text{ December 1993}\} \text{ and} \\
\mathcal{R}_3 &= \{\mathbf{X}^{\text{pred}} : R_{\text{S\&P 500}} > -0.935 \text{ and } t > \text{ December 1993}\},
\end{aligned}
$$

where $R_{\text{S\&P 500}}$ denotes the (annualized) daily returns of the US S&P 500 Index, and t denotes time. We find a first regime characterized by large losses of the US market index, and second and third regimes in reaction of positive and moderate losses of the US market, with an important structural break in time corresponding to December 1993. Section 3.2 contains a more structured discussion and interpretation of these results.

2.2. Estimation

The tree-HAR model introduced in equations (1) to (3) can be estimated using quasi-maximum likelihood (QML). Conditional on some reasonable starting values, the negative quasi-log-likelihood for model (1)-(3) is given by:

$$
-l\big(\theta; (\widetilde{RC}, \mathbf{X}^{\text{pred}})_1^n\big) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{t=1}^{n}\log\left(\sigma_t^2\right) + \frac{1}{2}\sum_{t=1}^{n}\frac{\left(\widetilde{RC}_t - \mathbb{E}_{t;\theta}[\widetilde{RC}_{t-1}]\right)^2}{\sigma_t^2}. \tag{4}
$$

Therefore, for any fixed sequence of partition cells, the tree-HAR model can be estimated by QML. The choice of the best partition cells (that is, splitting variables and threshold values) involves a model choice procedure for non-nested hypotheses. Similar to CART, the model selection of the splitting variables and threshold values can be performed using the idea of binary trees. Within any data-determined tree structure the best model is selected using the Bayesian-Schwartz information criterion (BIC). As pointed out by Hansen (1996), the use of model selection criteria to decide if the inclusion of another regime is relevant in threshold regression models such as the tree-HAR model has the following drawback: the location of the split cannot be estimated consistently when an irrelevant regime is added to the model. To overcome the problem, and also to ensure computational feasibility, we searched for threshold values over fixed grid points that are empirical quantiles of the different predictor variables. Alternatively, McAleer and Medeiros (2008) and Medeiros and Veiga (2009) recently proposed a sequence of tests to determine the number of regimes for a class of smooth transition models for the dynamics of financial (realized) volatility which circumvents the problem of identification in a way that controls the significance level of the tests in the sequence and computes an upper bound to the overall significance level. Such a strategy can be easily adapted to the case of fitting tree-HAR models. When choosing the splitting variables and threshold values using the above-mentioned sequence of tests, we obtained qualitatively the same regimes and, consequently, forecasting accuracy. For all details about the flexible procedure used to estimate the model, refer to Section 2.3 and Appendix A in Audrino and Trojani (2006). Proof of the consistency of the conditional mean and volatility estimates in the tree-HAR model under a possible model misspecification can be derived from Theorem 1 in Audrino and Bühlmann (2001).

3. Empirical application

3.1. Data and stylized facts

We considered a tick-by-tick bivariate returns series of the S&P 500 Futures and 30-year US Treasury Bond Futures for the period from January 1990 to March 2008, for a total of 4,451 daily observations. The data come from the Price-data.com database with time stamps rounded at the one minute frequency. The tick-by-tick realized covariances are computed with the First-Last estimator introduced by Corsi and Audrino (2008) as a generalization of the Hayashi and Yoshida (2005) realized covariance estimator to overcome the problem of working with imprecise arrival time due to rounding in the price time stamps. The First-Last covariance estimator can be seen as an average of two Hayashi and Yoshida-type of estimators. Both Hayashi and Yoshida are computed with return series constructed only with the first ticks in each interval of the grid induced by time stamps rounding for one asset, and the last ticks in each interval for the other asset.

Tick-by-tick realized volatility are computed employing the Multi-Scales Discrete Sine Transform realized volatility estimator proposed by Curci and Corsi (2003), which consists of a multi-frequency regression-based approach made robust by a discrete sine transform filter which optimally decorrelates the price signal from microstructure noise (we also constructed tick-by-tick realized volatilities using the two scales estimator of Zhang et al., 2005, obtaining similar results). We then have realized correlation measures where both the volatilities and the covariances are computed from tick-by-tick data.

The upper part of Figure 2 shows the time series of 4,451 daily tick-by-tick correlations we obtained for the time period between 1990 and 2008, together with their autocorrelogram. In the bottom part of Figure 2 Fisher-transformed (FT) realized correlations are plotted, again together with their autocorrelogram.

Figure 2 shows that (FT) realized correlations may be subject to regime shifts (in time or other relevant variables) and/or long-memory. It is difficult to disentangle the two features based on a descriptive analysis, because they may be in juxtaposition of one another. The results of the forecasting analysis performed in Section 3.3 will help us in determining the relevance of the two effects, i.e. structural breaks and long memory.

From a preliminary visual inspection of the correlation dynamics in Figure 2, we can recognize three important time changes of regime: the first occurring around the end of 1993 to the beginning of 1994, the second one around the end of 1997, and the last one in July 2007. As a confirmation of our visual inspection, the above mentioned four regimes are those we got when estimating the simple tree-AR(1) model on the FT-correlation series (we are going to use that model as a competing model in the forecasting analysis). The first break corresponds to a positive shock in the stock-bond correlations, whereas the second and third ones are negative shocks. The last two structural breaks in the stock-bond correlations' behavior with a consequent change in the sign of stock-bond correlations (from positive to negative) may be a consequence of a flight-to-quality effect induced by the Asian financial crisis and the recent subprime

S&P 500 – US Treasury Bond daily realized correlation from 1990 to 2008

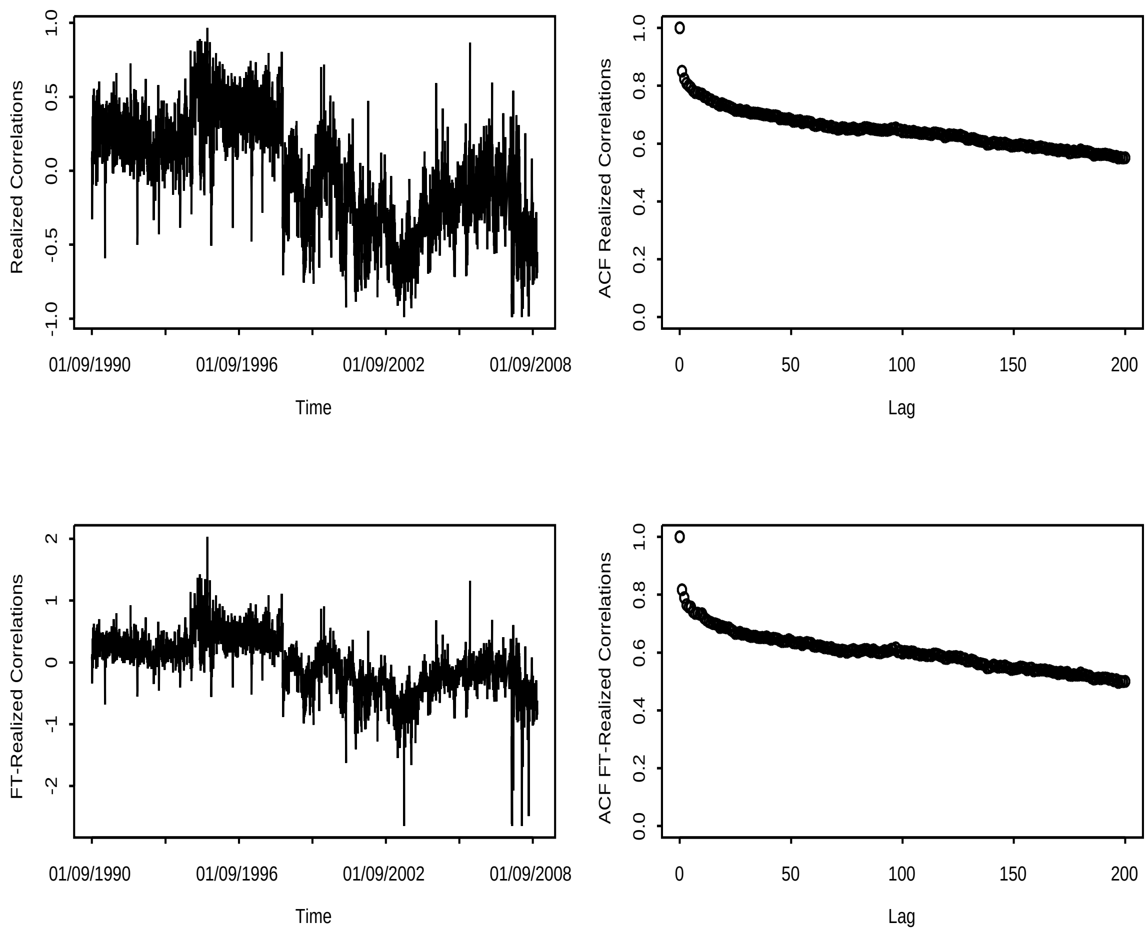

Figure 2: Time series of daily S&P 500 and 30-year US Treasury Bond realized correlations (upper-left panel) and Fisher-transformed (FT) realized correlations (lower-left panel) constructed using tick-by-tick data, together with their autocorrelograms (right panels). The time period under investigation is from January 1990 to March 2008.

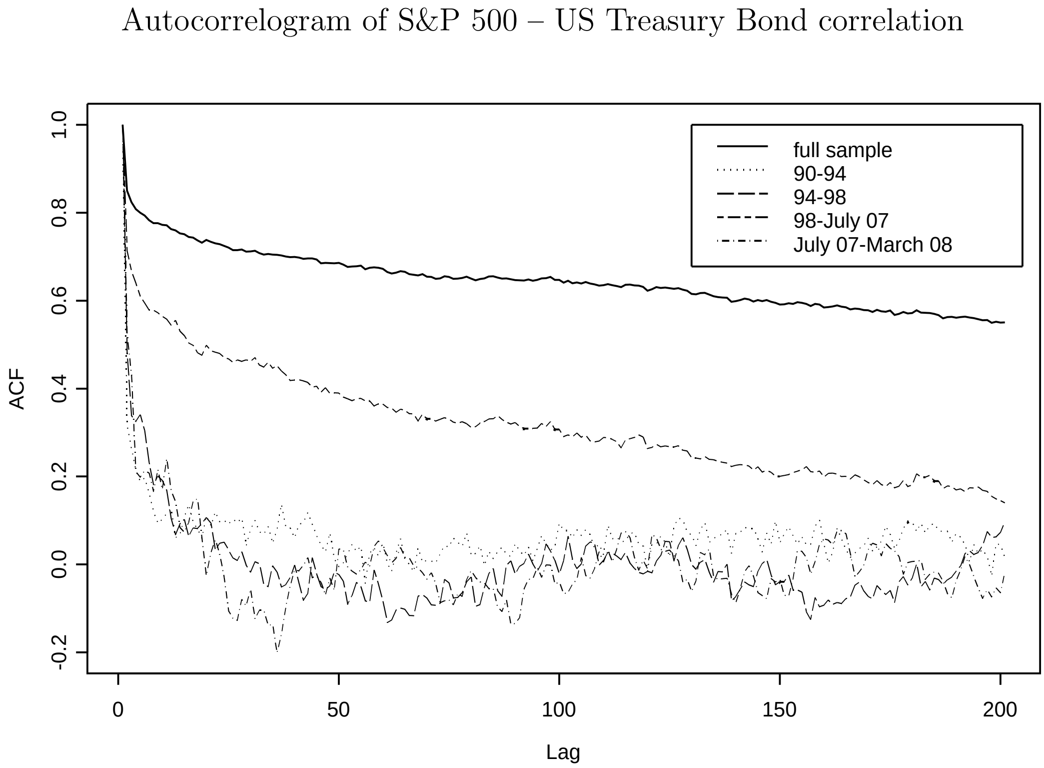

Figure 3: Autocorrelation functions of the S&P 500 and 30-year US Treasury Bond realized correlation for the full sample 1990–2008 (solid line) and the four sub-samples (time regimes): 90-94 (dotted line), 94-98 (long-dashed line), 98- July 07 (small-dashed line), and July 07 – March 08 (dot-dashed line).

financial crisis, respectively. In our opinion, the economic explanations for the first structural break may be twofold. First, the increase of US stock-bond correlations after the end of 1993 may be related to an increase in the investors' confidence in the US financial market during the long expansion cycle started in March 1991, also related to the European Monetary System crisis of 1992–1993. Second, this structural break may be a consequence of the inversion in the 1993–94 period in the trend of the real interest rate and the target Fed funds rate which after a long descending trend invert their tendency upwards.

More in detail, the correlation between the two series oscillates around a positive stable value of about 20 percent until 1994, around 40 percent from 1994 to 1997, while after the end of 1997, the correlation starts to exhibit a stronger dynamics and becomes predominantly negative. A last negative shock in the stock-bond correlations is observed in the 2007 summer.

When taking these structural changes into account, we get a different picture for the autocorrelations of the realized correlations. We computed the autocorrelation function in the four sub-periods: January 1990 to the end of 1993 (henceforth called the 90–94 period), from the beginning of 1994 to the end of 1997 (94–98 period), from the beginning of 1998 to July 2007 (98- July 07), and from July 2007 to the end of the sample period (July 07 - March 08). Their behavior is illustrated in Figure 3.

US S&P 500- T-Bond FT correlation estimated parameters

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

Table 1: Tree-HAR in-sample estimated parameters and regimes for the tick-by-tick Fisher-transformed (FT) realized correlation series of S&P 500 and US Treasury Bond returns. The sample period is from January 1990 to March 2008, for a total of 4,451 daily observations. $R_{\text{S\&P 500}}$ and t denote the past-lagged daily S&P 500 return and time, respectively. Model-based bootstrap standard errors computed using 1,000 replications are given in parentheses.

In the first period (90–94) the level of the autocorrelation is very low and quickly decaying. In the second period (94–98) the autocorrelation level and its persistence are even smaller. After the end of 1997 the memory of the process rises further. Finally, after the structural break in July 2007 the persistence significantly drops and the level of the autocorrelations becomes in most cases negative. This points to significant differences in the memory persistence of the stock–bond correlation in the subperiods.

Another interesting effect shown in Figure 3 is how time structural changes affect the global autocorrelation function computed on the full sample inducing an artificially high level in the autocorrelation coefficients. This phenomenon is discussed in detail by Diebold and Inoue (2001). Nonetheless, even without this structural break effect, the autocorrelation function of the realized correlation remains highly persistent, at least in the 98- July 07 period.

Following the recent literature on stock–bond correlations, we allow for structural breaks depending on a wide set of predictor variables, directly incorporated in a tree-HAR model specification. In addition to time, we considered as predictors past-lagged US S&P 500 and bond returns and realized volatilities, and past-lagged realized correlations at three different time horizons (daily, weekly, and monthly). The choice of the predictor space is clearly not restricted to the above mentioned variables and/or time horizons. One can think, for example, to use overnight returns squared, cross-products, or other relevant predictors in the modeling approach. In our case, we found that the mentioned variables at the daily horizon were the most informative ones to predict realized correlations. The results of this in-sample analysis are shown in the next section.

3.2. Estimation results

The analysis starts by estimating the tree-HAR model (1)-(3) on the whole data sample (from January 1990 to March 2008). Estimated coefficients, as well as the estimated regimes, are reported in Table 1. Classical model-based bootstrapped standard errors are given in parentheses (see Efron and Tibshirani, 1993).

Table 1 shows that almost all coefficients in the local dynamics of the conditional mean and variance of the Fisher-transformed (FT) realized correlations (with only one exception) are highly significant. About 11% of the observations in our sample belong to the first regime, which is characterized by relative large losses of the daily S&P 500 Index (i.e. larger than 1%), and therefore of the US stock market, for the time period preceding the US financial crisis started in the 2007 summer. In this regime the volatility of the FT-correlations is moderately high and the conditional mean dynamics are highly persistent. Moreover, the long-run mean is negative. This result supports and can be interpreted through the "flight-to-quality" effect already cited several times in the literature; see for example, Ilmanen (2003). In this regime characterized by bad stock market conditions, stocks and fixed-income assets tend to move in opposite directions, suggesting the phenomenon of fleeing from stock to bond markets.

The second and third regimes are characterized by "relatively good" stock market conditions in the time period preceding July 2007. Under such conditions, we identified a structural break in time at the end of 1993. Before that date, FT-realized correlations are moderately persistent, the long-run mean is positive and small, and volatility is also small. Such dynamics can be reasonably well associated with stable economic and market conditions. On the contrary, after the end of 1993 the conditional variance increases of approximately a factor of 2 (which corresponds to a 20% increase on the original, non-transformed correlations), conditional mean dynamics of FT-realized correlations become highly persistent, and the long-run mean drastically increases.

The last regime corresponds to the months following the 2007 summer in our data sample (i.e. 249 daily observations). This regime is characterized by extremely high volatility and a low persistence of FT-realized correlations. Similarly to the first regime, the long-run mean becomes negative. The same economic interpretations raised for regime 1 also apply to the final period in our sample.

The above described regime structure generates a large number of regime switches (from regime 1 and 2 before the structural break corresponding to the end of the year 1993, and from regime 1 and 3 after that break) that hover around the 10% of the data, independent if we count it before or after the break. The average duration of regime 1 before the end of 1993 is very short (around one day) and slightly increases after the break (around 1.2 days). This means that on average correlations follow the regime 1 dynamics for a very short period, and then switch to another regime. There is only one period of 5 consecutive days where the local dynamics of the FT-correlations stay in the regime 1. Average duration of regimes 2 and 3 are significantly higher; they are about 13 days and 8 days, respectively.

To end this section, we report the results of different goodness-of-fit statistics: the AIC and BIC criteria, and the mean absolute errors (MAE), the mean squared errors (MSE), and Ljung-Box (LB) tests at three different lags (5, 20, 50) for the residuals U_t in (1). For the purposes of comparison and to empirically verify whether one of the two effects (i.e. long-memory and structural breaks) is more important, we report the same statistics for different competitors introduced in the literature: the standard AR(1) model, the ARMA(1,1) model, the ARIMA(1,1,1) model introduced for non-stationary time series, a Tree-AR(1) model in the spirit of Audrino and Trojani (2006), the classical HAR model proposed by Corsi (2009), and a two-regimes Markov-switching HAR model (RS-HAR). The last comparison is particularly useful to investigate the possible advantages of using the tree-HAR model specification against a fair competitor that takes into account both long-memory and structural breaks. Results are summarized in Table 2

The superior ability of models taking into account both long-memory and structural breaks (i.e. the tree-HAR and the RS-HAR models) in estimating the dynamics of the FT-realized correlations is clearly shown by the values of the different goodness-of-fit statistics. For all performance measures considered, the tree-HAR or the RS-HAR models yield the best results. When neglecting to incorporate in the model long-memory or the possible presence of structural breaks, the estimates obtained for the conditional dynamics of the FT-realized correlations are highly inaccurate. Moreover, incorporating long-memory in the model seems to be more important than allowing for regime-shifts (see the comparison between the classical HAR model and the tree-AR(1) model).

In-sample performance 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

Table 2: In-sample goodness-of-fit results for the tree-HAR model, in comparison with the classical AR(1), ARMA(1,1), ARIMA(1,1,1) models, a tree-AR(1) model, the classical HAR model, and a two-regimes Markov-switching HAR model (RS-HAR). Data are FT-realized correlations between January 1990 and March 2008, for a total of 4,451 daily observations. The performance measure considered are the Akaike and Bayesian-Schwartz information criteria (AIC, BIC), and the mean absolute errors (MAE), the mean squared errors (MSE), and p-values of standard Ljung-Box tests at three different lags of the residuals.

3.3. Forecasting results

To better validate the usefulness of the tree-HAR model for the real data under investigation, we examine its forecasting ability, and compare this with the alternative methods introduced in the last section. We performed a series of out-of-sample tests to assess the forecasting power of the tree-HAR model for single-period and multi-period predictions of S&P - US Bond FT-realized correlations. Regarding goodness-of-fit statistics, we again considered the out-of-sample MAE and MSE of the residuals. In addition to these performance measures, we also report results for the out-of-sample log-likelihood in equation (4) in the single-period out-of-sample test, and the multiple R^2 obtained when regressing realizations against forecasts at the same time t (Mincer-Zarnowitz regression).

3.3.1. Single-period forecasts

To derive the daily forecasts we used a rolling strategy. The models are re-estimated every month (22 trading days) using all past data available in the sample. The initial in-sample period is from January 1990 to December 1999. Consequently, we obtained 1,979 out-of-sample daily forecasts (until March 2008). Results are summarized in Table 3. p-values of superior predictive ability (SPA) tests introduced by Hansen (2005) for the null-hypothesis that each model, taken as a benchmark, is not inferior to any of the alternatives are given in parentheses.

The tree-HAR model yields the best results for two out of the four goodness-of-fit statistics considered; with respect to the MSE statistic, the ARIMA(1,1,1) model is slightly better, and with respect to the log-likelihood the RS-HAR model yields the best result. Differences measured by the MAE and MSE statistics are in some cases very small and not statistically significant in the cases of the ARIMA(1,1,1), HAR and tree-HAR models. Only models that do not take into account long-memory (i.e. the AR(1), tree-AR(1), ARMA(1,1)) are clearly outperformed by the competitors. However, with respect to the out-of-sample likelihood, the tree-HAR and the RS-HAR models yield significant improvements in the accuracy of the FT-realized correlation forecasts over the competitors.

Summarizing, Table 3 suggests that when the focus of the analysis is daily forecasting, a simple HAR model, or even the ARIMA(1,1,1) model, yields very accurate results. Small improvements can be obtained when using more flexible models like the tree-HAR and the RS-HAR models. Improvements will be sometimes only marginal, depending on the performance measure and the competitor.

Single-period forecasting results

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

Table 3: Comparative results of one day forecasts of S&P 500 US bond FT-realized correlations obtained using the classical AR(1), ARMA(1,1), ARIMA(1,1,1) models, a tree-AR(1) model, the standard HAR model, a two-regimes Markov-switching HAR model, and the tree-HAR model. The forecasting time period is between January 2000 and March 2008, for a total of 1,979 daily observations. Out-of-sample forecasts are computed using a rolling strategy: the models are re-estimated every month (22 days) using the whole past information in the data sample. Performance is measured according to the out-of-sample log-likelihood (Loglik.), mean absolute errors (MAE) and mean squared errors (MSE) of the residuals, and multiple R^2 obtained when regressing realizations against forecasts at the same time t (Mincer-Zarnowitz regression). Values in parentheses are reported p-values of superior predictive ability (SPA) tests for the null-hypothesis that each model (taken as a benchmark model) is not inferior to any of the alternatives.

3.3.2. Multi-period forecasts

Practical asset allocation applications typically require correlation forecasts at longer time horizons. Therefore, we performed two out-of-sample tests at weekly (that is, five days) and monthly (that is, 22 days) horizons to assess the accuracy of the multi-period forecasts obtained using the different approaches. Such multi-period predictions are constructed using filtered historical simulation (FHS); see Barone-Adesi et al. (1998, 1999). FHS is a common, non-parametric version of the classical historical (Monte Carlo) simulation, where the new shocks are taken from the observed, empirical distribution of the fitted residuals (and not from a a-priori assumed theoretical distribution). Briefly, FHS works as follows. We generate 10,000 future scenarios at 5 (22) days horizons, bootstrapping the residuals estimated from the different models. Here, we considered the stationary bootstrap of Politis and Romano (1994) to account for the possible remaining autocorrelation in the residuals (see especially the Ljung-Box test results given in Table 2). Modification of the implementation to the block bootstrap of Künsch (1989) is straightforward. The forecast for the 5 (22) days ahead FT-realized correlation is then given by the median of the empirical distribution of the simulated future scenarios.

As in the previous out-of-sample experiment, we used the same rolling strategy and initial in-sample period. Results are summarized in Table 4. Again, p-values of SPA tests are reported in parentheses. The better forecasting power of the tree-HAR model with respect to all competitors for multi-period predictions is borne out by the results of the SPA tests. Especially for longer-time ahead forecasts (that is, 1 month), the predictions obtained using the tree-HAR model outperform those gleaned from the alternative approaches. Gains over the other competitors are, in most cases, statistically significant.

4. Conclusions

Combining realized covariances with realized volatilities, we obtained a realized correlation measure where both the volatilities and the covariances are computed from tick-by-tick data. We then propose a

Multi-period forecasting results: 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

Multi-period forecasting results: 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

Table 4: Comparative results of one week and one month forecasts of S&P 500 – US 30-years Treasury Bond FT-realized correlations obtained using the classical AR(1), ARMA(1,1), ARIMA(1,1,1) models, a tree-AR(1) model, the standard HAR model, a two-regimes Markov-switching HAR model, and the tree-HAR model. The forecasting time period is between January 2000 and March 2008. Out-of-sample forecasts are computed using a rolling strategy: the models are re-estimated every month (22 days) using the whole past information in the data sample. Performance is measured according to the mean absolute errors (MAE) and mean squared errors (MSE) of the residuals, and multiple R^2 obtained when regressing realizations against forecasts at the same time t (Mincer-Zarnowitz regression). Items in parentheses are reported p-values of superior predictive ability (SPA) tests for the null-hypothesis that each model (taken as a benchmark model) is not inferior to any of the alternatives.

tree-HAR model as a simple and parsimonious representation for the conditional dynamics of the (Fisher-transformed) realized correlations. The tree-HAR model is able to take into account two important stylized facts of realized correlations: strong temporal dependence (that is, long-memory) and structural breaks.

We estimated the tree-HAR model on the realized correlation series of S&P 500 and US bond returns, finding empirical evidence that the conditional dynamics are subject to regime changes depending on different values of past S&P 500 daily returns. We also identified two structural breaks in time, corresponding to the end of 1993 and the summer of 2007.

We then tested the prediction accuracy of the model using SPA tests for different goodness-of-fit statistics finding empirical evidence of its strong predictive power. The tree-HAR model significantly outperforms the competitors, particularly when the final goal is multi-period forecasting.

Our study can be generalized along different directions. Given the central role of stock-bond correlations in many practical financial problems, a natural extension of this study is to evaluate the tree-HAR predictions using some economically relevant loss functions, such as portfolio performance and trading strategies, risk-management (i.e. Value-at-Risk and Expected Shortfall), and option pricing. Although in this study we restricted our focus on the very simple US stock-bond bivariate application to compare our results to those already found in the literature, the same methodology can be generalized to other, more structured multivariate case studies. Achieving this goal will require the construction of valid positive definite tick-by-tick realized covariances and correlations matrix (which is, however, still an active area of ongoing research, see, for example, Barndorff-Nielsen et al., 2008) together with a proper generalization to the multivariate setting of the tree-structured thresholds model (as the one proposed in Audrino and Trojani, 2007).

Moreover, the model proposed in this study (i.e. the tree-HAR model) to estimate and forecast the dynamics of realized conditional correlations may be used as a starting model to develop further improvements. In particular, forecasts from the tree-HAR model can be combined with other model forecasts or can be improved by using machine learning methodologies (for instance, bootstrap aggregating). The tree-HAR model itself can be generalized to be able to take into account also smooth transitions across regimes. These and other investigations are left for future research.

Acknowledgments

Financial support through the Foundation of Research and Development of the University of Lugano is gratefully acknowledged. We thank Tim Bollerslev and two anonymous referees for helpful comments.

References

Andersen, T., Bollerslev, T., Diebold, F., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics* **61** (1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001b). The distribution of exchange rate volatility. *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71**(2), 579–625.

Andersen, T. G., Bollerslev, T., and Huang, X. (2007). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. CREATES Research Paper Series 14.

Audrino, F. and Bühlmann, P. (2001). Tree-structured GARCH models. *Journal of the Royal Statistical Society, Series B*, **63**, 727–744.

Audrino, F. and Trojani, F. (2006). Estimating and predicting multivariate volatility regimes in global stock markets. *Journal of Applied Econometrics* **21**, No. 3, 345–369.

Audrino, F. and Trojani, F. (2007). A general multivariate GARCH model with dynamic conditional correlations. VWA Discussion Paper Series 25, University of St. Gallen.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, **63**, 167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, **64**, 253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* **17**(5), 457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica* **72** (3), 885–925.

Barndorff-Nielsen, O. E. and Shephard, N. (2005). How accurate is the asymptotic approximation to the distribution of realized volatility?. In D. W. F. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg.* Cambridge University Press, pp. 306–331.

Barndorff-Nielsen, O.E. and Hansen, P.R. and Lunde, A. and Shephard, N. (2008). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Working Paper.

Barone-Adesi, G., Bourgoin, F., and Giannopoulos, K. (1998). Don't Look Back. *Risk* **11**, August, 100–104.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). VaR Without Correlations for Portfolio of Derivative Securities. *Journal of Futures Markets* **19** (April), 583–602.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont (CA).

Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* **4**, 537–572.

Christiansen, C. and Ranaldo, A. (2007). Realized bond-stock correlation: macroeconomic announcement effects. *Journal of Futures Markets* **27** (5), 439–469.

Comte, F. and Renault, E. (1998). Long memory in continuous time stochastic volatility models. *Mathematical Finance* **8**, 291–323.

Connolly, R., Stivers, C., and Sun, L. (2005). Stock market uncertainty and the stock-bond return relation. *Journal of Financial and Quantitative Analysis* **40**, 161–194.

Corsi, F. (2009). Simple approximate long memory models of realized volatility. *Journal of Financial Econometrics* **7**, 174–196.

Corsi, F. and Audrino, F. (2008). Realized covariance tick-by-tick in presence of rounded time stamps and general microstructure effects. Unpublished manuscript, University of St. Gallen.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2006). The volatility of realized volatility. Forthcoming in the *Special Issue of Econometric Reviews on Realized Volatility and Long Memory.*

Curci, G. and Corsi, F. (2003). A discrete sine transform approach for realized volatility measurement. NCCR FINRISK Working Paper No. 44.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics* **105**, 131-159.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London.

Ferland, R. and Lalancette, S. (2006). Dynamics of realized volatilities and correlations: An empirical study. *Journal of Banking and Finance* **30** (7), 2109–2130.

Griffin, J. E. and Oomen, R. C. A. (2006). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Unpublished manuscript.

Guidolin, M. and Timmermann, A. (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics* **21** (1), 1–22.

Hansen, B. E. (2006). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64**, 413–430.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics* **23**, 365–380.

Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11**(2), 359–379.

Hillebrand, E. and Medeiros. M.C. (2008). The Benefits of Bagging for Forecast Models of Realized Volatility. *Econometric Reviews*, forthcoming.

Ilmanen, A. (2003). Stock-Bond Correlations. *Journal of Fixed Income* **13**, 55–66.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**, 1217–1241.

Li, L. (2002). Macroeconomic factors and the correlation of stock and bond returns. Working paper, Yale International Center for Finance.

Martens, M. (2004). Estimating unbiased and precise realized covariances. Social Science Research Network Electronic Library.

McAleer, M. and Medeiros, M. C. (2008). A multiple regime smooth transition heterogenous autoregressive model for long memory and asymmetries. *Journal of Econometrics* **147**, 104–119.

Medeiros, M. C. and Veiga, A. (2009). Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model. *Econometric Theory* **25** (1), 117–161.

Müller, U., Dacorogna, M., Dave, R., Olsen, R., Pictet, O., and Weizsacker, J. (1997). Volatilities of different time resolutions – Analyzing the dynamics of market components. *Journal of Empirical Finance* **4**, 213–239.

Palandri, A. (2006). Consistent realized covariance for asynchronous observations contaminated by market microstructure noise. Unpublished manuscript.

Pastor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *The Journal of Political Economy* **111**, 642–685.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association* **89**, 1303–1313.

Ranaldo, A. and Reynard, S. (2008). Monetary Policy Effects on Long-Term Rates and Stock Prices. Unpublished manuscript.

Sheppard, K. (2006). Realized covariance and scrambling. Unpublished manuscript.

Voev, V. and Lunde, A. (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics* **5**, 68–104.

Zhang, L., Mykland, P.A., and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association* **100**, 1394–1411.