NTNU
Norwegian University of
Science and Technology

FX Markets and Google searches

- an investigation of Structural Breaks

Young Kim
Anton Emil Bergh Poulsson

Industrial Economics and Technology Management
Submission date: June 2016
Supervisor: Peter Molnar, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Problem description

Foreign exchange rates respond strongly to various events, whether these are market events, governmental actions or just geopolitical events. As a result, exchange rates might exhibit various regimes separated by structural breaks. We would therefore like to investigate a flexible model that endogenously detects these various regimes, and secondly, investigate if Google query searches contain information not already included in the model. This is an important topic regarding forecasts of foreign exchange dynamics, with practical value for investors, portfolio managers and governments.

Preface

This paper represents our final thesis as students in Master of Science in Industrial Economy and Technology Management at the Norwegian University of Science and Technology (NTNU). Our thesis is the product of 20 weeks of interesting, educational and, at times, tedious work. All the work presented henceforth was conducted by the authors themselves in the direction of empirical and quantitative finance.

We would like to accentuate the importance and inspiration the work of Francesco Audrino and Fulvio Corsi have had on our master thesis. Our model is mainly inspired by their tree-HAR model, and empirically tested on high-frequency foreign exchange data.

First of all, we would like to thank our supervisor, Peter Molnár, for his continuous guidance and expertise, which has been of invaluable help. We would also like to direct a thanks to family and friends who have helped us with guidance through our study. We thank you all for the support you have given us!

Trondheim, June 19, 2016

Anton Emil Bergh Poulsson

Young Kim

Contents

FX Markets and Google searches – an investigation of Structural Breaks

Young Kim, Anton E. B. Poulsson

Abstract

This paper studies the dynamics of volatilities and correlations calculated from high-frequency data for seven major currency pairs. We build our model upon the precise Heterogeneous Autoregresive (HAR) model of Corsi (2009). Since foreign exchange rate dynamics can change over time, we implement a flexible tree Heterogeneous Autoregressive (tree-HAR) model. The model is able to detect various regimes, and even the number of regimes is determined endogenously within the model. We find several regimes in each of the 7 volatility and 21 correlation time-series. Interestingly, these regimes are much more obvious for correlations. In addition, we find the tree-HAR model to react faster to structural breaks than the HAR model. Furthermore, we examine whether Google searches contain additional information not already included in modelling structural breaks. Despite not affecting regimes, Google search data provide as a helpful proxy measuring market uncertainty — the same uncertainty causing many of these structural breaks.

Keywords: FX, High-frequency Data, Google Trends, tree-HAR, Structural Breaks

Supervisor: Peter Molnár

Sammendrag

Hovedformålet med denne masterutredningen er å undersøke dynamikken til volatilitet og korrelasjon beregnet fra høyfrekvens data for syv valutapar. Modellen vår bygger på den nøyaktige Heterogene Autoregressive (HAR) modellen til Corsi (2009). Siden valutadynamikker endrer seg over tid velger vi å implementere en fleksibel tre-Heterogen Autoregressiv (tre-HAR) modell. Denne modellen klarer å oppdage flere regimer, hvor til og med antall regimer er bestemt endogent i modellen. Vi finner flere regimer for alle de 7 volatilitetene og 21 korrelasjonene. Interessant å merke seg er at disse regimene er mye tydeligere for korrelasjonstidsseriene. I tillegg finner vi tre-HAR modellen til å reagere raskere på strukturelle skift enn HAR-modellen. Videre undersøker vi om Google-søk inkluderer informasjon som ikke allerede er tatt i betraktning i modelleringen av de strukturelle skiftene. Selv om Google søk ikke har en signifikant påvirkningskraft på skiftene, fremstår det som en lovende estimator for å måle markedsuro — nettopp den samme uroen som forårsaker mange av de strukturelle skiftene.

All authors are from the Norwegian University of Science and Technology (NTNU). Correspondence should be sent to Peter Molnár, email address: peto.molnar@gmail.com

1. Introduction

Understanding the volatility and correlations of exchange rates[1] is an extremely important topic. It could assist governments in mitigating economic risks by better altering interest rates and trading policies according to the preferable economic course. Big corporations would be able to reduce hedging costs, while professional investors could possibly earn a profit from trading with uninformed investors.

In this paper we study the dynamics of foreign exchange (FX) rate volatilities and correlations calculated from high frequency data for seven major currency pairs. First, we investigate a flexible regression-tree model that endogenously estimates the number of regimes, as presented by Audrino and Corsi (2010). We document that various regimes are present in both volatilities and correlations, but most distinctly present for correlations. Second, we use daily Google search query data to investigate how they relate to FX volatilities and correlations.

Historically, volatility models were based on daily data. Daily volatility models either assume that volatility is randomly distributed (stochastic volatility models) or time-varying dependent on historical values (historical volatility models). However, usually these models cannot sufficiently capture all the observed properties of exchange rate volatility. For example, Andersen et al. (2001) study the distribution of exchange rate volatility between the Deutsche mark and Japanese yen, and document the existence of fat tails and long-memory behaviour.

On the other hand, realized volatility calculated from high-frequency data can be used to create volatility models that are simple to estimate, but at the same time very precise. One example of such a model is the Heterogeneous Autoregressive (HAR) model introduced by Corsi (2009). However, one challenge forecasting the FX market is that results might not be stable over time, due to regime shifts in monetary and fiscal policies (Bakaert and Hodrick, 1993). Such regimes need to be taken into account to get reliable estimates and forecasts. By creating models that quickly react to sudden structural breaks, financial market participants could get more precise forecasts of FX complex dynamics.

Studies indicate that FX rates tend to have clear structural breaks related to financial crises or shifts in monetary and fiscal policies (Bakaert and Hodrick, 1993, Engel and Hakkio, 1996). These structural breaks are highly related to uncertainty in the market, most commonly measured as market volatility. However, structural breaks are not always easily noticeable from solely studying the patterns of volatility. In fact, correlation time-series often reveal such breaks more clearly. Furthermore, studies regarding Internet searches find a significant relationship between increased investor attention and higher volatility and risk premiums (Andrei and Hasler, 2014, Kita and Wang, 2012, Vlastakis and Markellos, 2012). This gives reason to believe that Internet searches could prove to be a useful proxy for

[1] Foreign exchange rates are generally considered highly unpredictable (Fama, 1984, Meese and Rogoff, 1983, Taylor, 1987). This is not surprising, since any significant predictability would be quickly exploited by market participants seeking for profit opportunities. With this in mind, we find it more suitable to examine the predictability of foreign exchange rate volatility and correlations, instead of the exchange rates themselves.

market uncertainty, and that the relationship between Internet searches and volatilities and correlations might be varying for different regimes.

Bollen et al. (2000) find that even option prices, which are forward-looking, do not fully reflect regime-switching information. This means that present models, even those that include volatility implied from options prices, might not be able to model volatilities and correlations in a satisfying way. Hamilton and Susmel (1994) implement a regime-shifting model for financial time series. They uses a Markov chain to map the regimes. These regime-shifting models have been shown to improve forecasts of financial time series with structural breaks, or regimes. Engel and Hakkio (1996) investigates exchange rate mechanisms of the French franc, Italian lira, US dollar and Japanese yen against the Deutsche mark. They find that many of these exchange rates are characterised by long periods of stability interrupted by periods of extreme volatility. They use a model similar to Hamilton and Susmel (1994) to switch between regimes . Audrino and Bühlmann (2001) introduced a tree-structured model for regimes inspired by a classification and regression tree analysis. However, their model differentiates itself from regression trees by using likelihood values, and not residual sum of squares, as previous studies have done.

Other papers have used similar models to Audrino and Bühlmann (2001) and Hamilton and Susmel (1994) in their investigations of financial time series. Engel and Hamilton (1990) divide non-stationary foreign exchange time-series into two regimes to forecast exchange rates. Audrino and Trojani (2006) use a tree-AR-GARCH model to perform analysis of the global stock market. McAleer and Medeiros (2008) incorporates a HAR model in a smooth regime transition (HARST) model. They use several Dow Jones Industrial Average index stocks to forecast volatility. Audrino and Corsi (2010) develops a tree-HAR model to forecast realized correlation of S&P 500 and 30-year US Treasury Bond Futures. They find almost identical results using the HARST model.

The tree-structured Heterogeneous Autoregressive model for Realized Volatility (tree-HAR) by Audrino and Trojani (2006) has proven its ability to capture FX dynamics such as fat tails and long-memory behaviour. The main difference between the flexible tree-HAR model and standard models is that the tree-HAR model determines the number of regimes endogenously, whereas standard models implicitly assume just one regime, or, in the case of regime switching models, a given number of regimes (usually two regimes).

A relatively recent group of studies investigates the potential of using Internet search data in forecasting. Da et al. (2011) argue that Google searches can be used as a proxy for uninformed investors' effort to gain information. They also comment that professional traders often search for information through professional platforms, such as Reuters and Bloomberg. Since search for information is both time consuming and uncertain, it can be considered as an expense for investors (Bank et al., 2011, Drake et al., 2012). Similarly, Goddard et al. (2015) suggest that investor attention is a priced source of risk in FX markets. Due to these findings, investors will have to decide how much time to spend searching for information. This can cause asymmetries of information creating the opportunity for informed investors to earn a profit from trading with uninformed investors. Other interesting studies find that volatility and risk premiums increase with attention (Kita and Wang, 2012, Vlastakis and

Markellos, 2012). The intuition for this is that uncertain times make investors more risk averse.

Internet search data has become very popular in many fields, particularly for the purpose of nowcasting and forecasting[2]. Google search data is popular in financial forecasting too. Fink and Johann (2014) finds that the daily SVI is significantly positively related to daily turnover and volatility of a stock. Risteski and Davcev (2014) improves an E-GARCH model for the CAC40 index with SVI. Smith (2012) uses weekly Google Insight searches to measure investor sentiment and FX market volatility. He finds significant improvement by extending the GARCH(1,1) model with the SVI queries. Similarly, Goddard et al. (2015) examine investors' demand for information on seven currency pairs. They find that FX market volatility is affected by changes in SVI. Likewise, BitCoin is found to have a positive correlation with both queries from Google Trends and Wikipedia (Kristoufek, 2013). The author also investigates causality and detects a two-way influence between query searches and the BitCoin rate. However, Dimpfl and Jank (2015), supported by Goddard et al. (2015), find the SVI Granger-causing volatility, but not vice versa.

We contribute to the literature by investigating the dynamics of volatility and correlation for the seven most traded currencies against the US dollar. We therefore investigate 7 time-series for volatilities, and 21 time-series for correlations. The tree-HAR model allows us to identify distinct regimes for our sample, both for volatilities and correlations. In principle, these regimes could be linked to market conditions. This was done e.g. in Audrino and Corsi (2010) for the correlation between stocks and bonds. However, due to the large scope of our data set, we leave economic interpretation of how these regimes relate to particular market conditions and macroeconomic events to further research.

In the second part, we extend the HAR and tree-HAR model by adding Google search data as an exogenous variable. To our knowledge, we are the first paper to investigate whether Google query searches can improve FX volatility and correlation forecasts, and furthermore, the first study to investigate the role of Google searches within various regimes. Some papers have investigated the use of Google search data in forecasting FX volatility, but not with precise models based on high-frequency data. We find the SVI variable to be useful in forecasts of realized volatility and realized correlation. Moreover, the importance of the SVI variable varies across regimes of the tree-HAR model. Our results show that Google search queries have greater impact during more volatile regimes. Vlastakis and Markellos (2012) found a similar result for the 30 largest stocks traded in the US.

The rest of the paper is organised as follows. Section 2 presents the data, continued by model explanations in section 3. Section 4 and 5 present our results and concluding remarks.

[2] One of the first research areas that started to adopt Internet searches for practical purposes was syndromic surveillance, ie. early detection of disease outbreaks. Eysenbach (2006) finds Google searches to be a useful tool in predicting flu outbreaks in Canada. Other studies on syndromic surveillance find Internet search engines, such as Yahoo and Google, to have a significantly high correlation with outbreaks (Ginsberg et al., 2009, Polgreen et al., 2008). Google search data has also been used for nowcasting and forecasting in other interesting areas, such as housing prices (Kulkarni et al., 2009), tourism (Pan et al., 2012), unemployment rates (Anvik and Gjelstad, 2010, Choi and Varian, 2011), politics and social studies (Askita and Zimmermann, 2015, Reilly et al., 2012) and consumption (Vosen and Schmidt, 2011).

2. Data

2.1. FX data

Our original data consist of tick-by-tick foreign exchange rate volatility and correlation for seven currency pairs linked to the USD: AUD/USD, USD/CAD, USD/CHF, EUR/USD, GBP/USD, JPY/USD and USD/NZD. All the seven FX rates are sampled for the period from January 1, 2004 to October 31, 2015, using intervals of 30-minutes. We only include weekdays. We obtain foreign exchange rates data from the Swizz bank Ducascopy through Strategy Quant's program "Tick Downloader".

Next, we calculate daily realized volatility calculated as the standard deviation of the logarithmic intraday returns. Logarithmic intraday returns are given as $z_t = \log(x_t/x_{t-1})$, where x_t is the closing exchange rate on interval t. We calculate the daily realized volatility (RV) of the FX-rates in accordance with Andersen et al. (1999).

$$RV_t = \sqrt{\sum_{i=1}^{n} z_{t,i}^2}$$

where t is the index of the day on the interval i with n number of intraday return intervals. Weekly and monthly realized volatility is calculated using an equally weighted moving average (1), as in Corsi (2009):

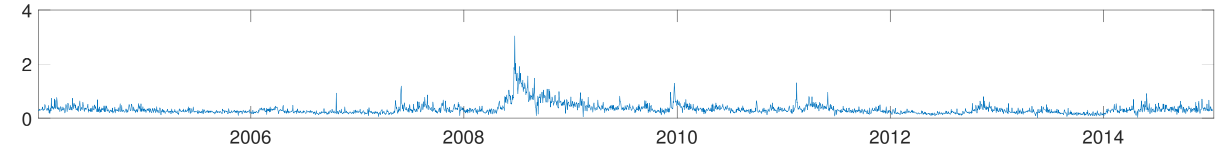
$$RV_t^p = \frac{1}{d}\sum_{i=1}^{d} RV_{t-i} \tag{1}$$

where d is the number of trading days in the given period p. For weekly realized volatility n = 5, while for monthly realized volatility n = 22. For simplicity, we will hereby denote daily, weekly and monthly realized volatility as RV^d, RV^w and RV^m. Daily realized covariance is calculated from tick-to-tick FX data. We use a simple covariance estimator as in Audrino and Corsi (2007). They show that the covariance estimator has no bias and a smaller dispersion than other estimators tested in their study. Daily realized correlation is then constructed from the realized volatility and covariance in the following way:

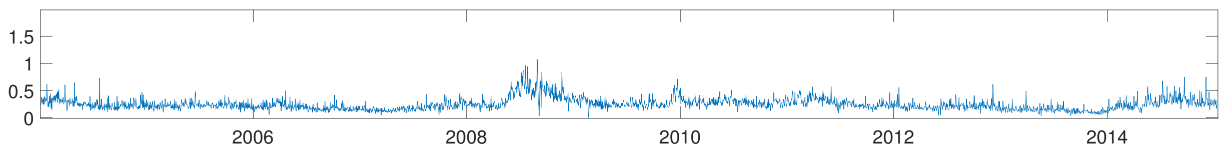
$$RCov_t = \sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} z_{i,s} z_{j,q} \tag{2}$$

$$RC_t = \frac{RCov_t}{\sqrt{\sum_{i=1}^{M_{i,t}} z_{i,s}^2 \sum_{j=1}^{M_{j,t}} z_{j,q}^2}} \tag{3}$$

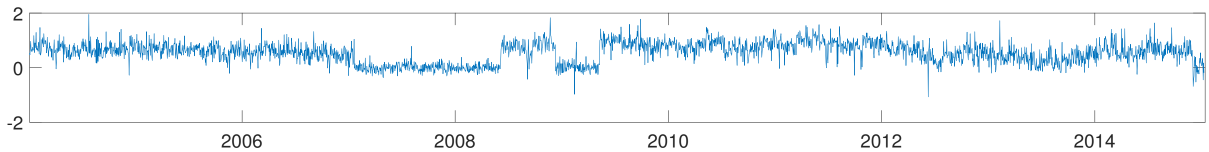
where M is the number of intraday samples for currency pair i and j on day t. Figure 1 shows the dynamics of volatilities and correlation for two currency pairs. It indicates that there are clear cut regimes seen for the correlation, but not for volatilities.



(a) Realized volatility of AUD/USD



(b) Realized volatility of EUR/USD



(c) Realized correlation of AUD/USD and EUR/USD

Figure 1. Time-series of the correlation of correlation of AUD/USD and EUR/USD and the volatility of the respective pairs.

2.2. Google Search Volume Index

We use Google search volume index (SVI) data downloaded from Google Trends. It is the most studied[3], easy accessible search volume data that has been used to measure investor attention. The earliest SVI data is available from January 1, 2004. The SVI provides a normalised ratio of Google query searches for a given search word. The SVI ranges from 0 to 100, where 100 is the maximum. Google Trends reports 0 if the number of searches for a period is under a certain threshold level. We remove these periods before the analysis. This causes some of the data sets to contain fewer observations, often related to a lower number of searches in the early stage of Google. In addition, the SVI shows seasonality with search volumes significantly smaller during weekends and around Christmas. We choose to remove data during these periods.

It is also possible to filter Google searches. However, we deem it unnecessary to use Google's finance filter as the queries themselves are unambiguous and mostly restricted to currency related searches, unlike for example the equivalent for company tickers and names.

[3]One downside of using the SVI is the varying popularity of Google as a search engine across the different countries we examine. This is most apparent for Japan, where Google has a 63.74% share of the search engine market compared to Europe and US's over 90%, as of October, 2015 (Applications, 2015). In addition, Google searches for tickers codes in Latin letters might fail to include searches in the local languages. This would only be a problem for the Japanese Yen. We leave this for further research.

Table 1: Cleaned SVI data.

This table present all the Google search queries and cleaned sample periods for the Google data downloaded from Google Trends. N is the number of observations within each sample period for the respective searches.

Ticker Codes	Search query	From date:	To date:	N
AUD	"AUD"	2005-05-24	2015-10-01	2639
CAD	"CAD"	2004-01-01	2015-10-01	2996
CHF	"CHF"	2004-07-28	2015-10-01	2847
EUR	"EUR"	2004-01-01	2015-10-01	2996
GBP	"GBP"	2004-07-01	2015-10-01	2866
JPY	"JPY"	2007-09-04	2015-10-01	2045
NZD	"NZD"	2007-03-05	2015-10-01	2181
USD	"USD"	2004-01-01	2015-10-01	2995

However, it should be mentioned that queries for the full currency names, such as "EURO" and "DOLLAR", have bigger search volumes, but are less likely to be caused by potential investors. This is elaborated in section 4.3. Table 1 lists all queries downloaded from Google Trends and their respective sample periods.

The raw data from Google trends is processed to make it comparable over time. Daily and weekly SVI data was downloaded in Python, processed in C# and analysed in Matlab. The first step we take is to make the daily data comparable over time. The daily SVI is only provided for two months at a time and respectively normalised for the separate periods. We use the weekly SVI, which is downloaded as a single time series for the entire search period, to combine the different daily data sets. We then normalise the SVI data in the following way:

$$z_i = \frac{x_i - \mathbf{E}(x_i)}{\sigma_i} \tag{4}$$

Where x_i is the logarithm of the SVI value on day i, and $\mathbf{E}(x_i)$ is the expected value for x on day i accounted for the linear trend in the data.[4] σ_i is the standard deviation over the last 252 days ending on day i. The use of the linear detrend function is motivated by the incremental trend in Google searches found from 2004 to 2015. The expected value is calculated with Matlab's detrend function of the past 252 days. The distributions of the normalised SVI and complete time-series are included in the appendix. Figure 2 shows the effect of the normalisation on the daily SVI for the EURO.

[4]For the sake of robustness, we also test a common fixed period normalisation and polynomial fit functions, with similar results.

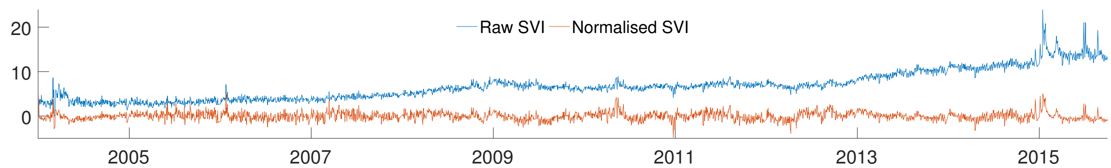


Figure 2. Time-series of absolute daily SVI and normalised daily SVI for the EURO.

2.3. *Descriptive statistics*

All tables of the descriptive statistics are reported in appendix C. The most obvious feature is that the realized volatility exhibit extremely high kurtosis[5]. Following Liu and Maheu (2008), Chiriac and Voev (2011) and Dimpfl and Jank (2015), we therefore perform a logarithmic transformation of the RV (log-RV). Log-RV has better statistical properties, and we therefore use log-RV in the subsequent analysis.

The normalisation of the SVI ensures that the SVI approximately follows a standard normal distribution, with a zero mean and standard deviation close to 1. Summary statistics distinguish CHF from other currencies. This is caused by few extreme observations. However, we keep all the observations in our analysis. The reason is that we want to see whether our method works even in the presence of extreme observations.

A Fisher-transformation is performed on the daily realized correlations, as reported in table C.3. It transforms the realized correlation so that it is not restricted to the interval [-1,1]. Moreover, the distribution of the Fisher-transformed correlations is closer to normal than the distribution of not-transformed correlations, which is convenient for modelling purposes. The formula for the Fisher-transformation is as following:

$$\tilde{RC}_t = \frac{1}{2} log\left(\frac{1 + RC_t}{1 - RC_t}\right) \tag{5}$$

Hereon, we will refer to the normalised SVI data as the SVI and Fisher-transformed realized correlation as $\tilde{RC}$.

Next, we test for stationarity and long-memory behaviour for realized volatilities and correlations. The unit-root test is based on the Dickey-Fuller test (Dickey and Fuller, 1979), and is performed with 10 augmented lags. The autocorrelation is run with up to a 100 day lag to test for long-memory behaviour. Complete results for all the time-series are presented in the appendix.

One feature of both realized volatility and correlation is that they clearly show a long-memory behaviour through a slow decay in autocorrelation. Even at a 100 day lag, the time-series exhibits long-memory behaviour. These findings would suggest the use of a long-memory model to modelling realized volatilities and correlations of FX rates. Other papers report of similar behaviour from foreign exchange dynamics. Andersen et al. (1999) reports

[5] The highest kurtosis (803.97) is observed for the RV of USD CHF currency pair. This could indicate that CHF has been exposed to extreme events causing abnormally high values, as for instance the January 2014 event when the Swiss Franc was depegged from the Euro.

Table 2: Summary of descriptive statistics

Panel A: Descriptive statistics of log realized volatility						
	Mean	St.dev	Kurtosis	Skewness	Max	Min
AUD/USD	-1.23	0.45	4.96	0.49	1.11	-3.85
CAD/USD	-1.49	0.43	6.43	-0.44	0.07	-5.11
CHF/USD	-1.54	0.43	4.31	0.30	0.23	-3.49
EUR/USD	-1.12	0.42	4.69	0.31	0.94	-3.81
GBP/USD	-1.46	0.43	5.03	-0.10	0.24	-4.86
JPY/USD	-1.39	0.42	6.14	0.05	2.00	-4.12
NZD/USD	-1.48	0.45	4.57	0.01	0.43	-4.10
Panel B: Descriptive statistics of normalised daily SVI						
	Mean	St.dev	Kurtosis	Skewness	Max	Min
USD/AUD	0.07	0.97	5.17	0.87	5.24	-2.79
USD/CAD	0.04	0.96	4.34	0.13	5.64	-4.49
USD/CHF	0.10	1.01	15.66	1.62	12.89	-3.80
USD/EUR	0.05	0.90	6.12	0.72	6.44	-4.42
USD/GBP	0.06	0.91	3.98	0.49	4.51	-3.11
USD/JPY	0.10	0.98	6.03	0.99	6.22	-2.96
USD/NZD	0.10	0.95	3.70	0.22	3.92	-3.63

of long-memory behaviour of the Deutsche mark/US dollar and Japanese yen/US dollar realized volatilities and correlations. This may suggest the presence of a unit root. However, we perform a Dickey-Fuller test with 10 augmented lags to find the null hypothesis of a unit root is rejected across all 10 lags, and currency pairs (table A.1 in the appendix). This allows us to model the time-series as stationary processes. The choice of model is explained in more detail in the next section.

3. The model

3.1. HAR model

The choice of model is motivated by the empirical features of our data sample, such as long-memory behaviour, fat-tails and structural breaks. Standard GARCH models and Stochastic Volatility models often fall short when forecasting time-series with the above-mentioned characteristics (Corsi, 2009). We therefore find the Heterogeneous Autoregressive model for Realized Volatility (HAR) by Corsi (2009) to be the most suitable model. Moreover, the HAR model has the convenient feature that it can be easily extended to include additional explanatory variables. Despite its simple structure, the HAR model shows good precision in modelling high-frequency volatility and correlation for foreign exchange rates.

Even though the model is not a long memory model, it has proven to be a "pseudo-long memory" model, outperforming other standard models. It takes into account both autocorrelations and empirical long memory behaviour. It should be mentioned that long-memory models, such as the FIGARCH and ARFIMA, are frequently used to model long memory volatility. However, the fractional difference operator used in these models cause a loss of observations which potentially could fail to capture the real structure of the data (Corsi, 2009). This is were the HAR model excels with is straightforward mathematics and clear economic interpretation.

The HAR model includes past values of the independent variable of different time intervals (daily, weekly and monthly). We use the same model structure for RV and $\tilde{RC}$, specified in the following way:

$$\mathbf{E_t}[\tilde{RV}_{t+1}] = \alpha_j + \beta_j^d \tilde{RV}_t^d + \beta_j^w \tilde{RV}_t^w + \beta_j^m \tilde{RV}_t^m + \epsilon_t \tag{6}$$

where RV^d (RC^d), RV^w (RC^w) and RV^m (RC^m) denote daily, weekly and monthly realized volatilities (correlations) calculated according to equation 1. The βs are the regression coefficients and ϵ_t is a white noise term with i.i.d.

3.2. Tree-HAR model

In order to capture structural breaks in the time-series, we implement a tree-structure to the HAR model, as inspired by Audrino and Corsi (2010). The tree-HAR model divides the time-series into regimes based on different predictor variables. The time series of the Fisher transformed realized correlations is modelled in the following way:

$$\tilde{RC}_{t+1} = \mathbf{E_t}[\tilde{RC}_{t+1}] + \sigma_{t+1} U_{t_1} \tag{7}$$

where U_{t+1} is a sequence of i.i.d. innovations following a Gaussian distribution. $\mathbf{E_t}$ is the conditional expectation given by:

$$\mathbf{E_t}[\tilde{RC}_{t+1}] = \sum_{j=1}^{k} \left(\alpha_j + \beta_j^d \tilde{RC}_t^d + \beta_j^w \tilde{RC}_t^w + \beta_j^m \tilde{RC}_t^m \right) I_{i,j} \tag{8}$$

$$\sigma_{t+1}^2 = \sum_{j=1}^{k} \sigma_j^2 I_{i,j} \tag{9}$$

where k is the number of regimes estimated endogenously from the data[6]. The conditional expectations is estimated from daily, weekly and monthly $\tilde{RC}$. The regression coefficients, βs, are calculated within each regime k. $I_{i,j}$ is an indicator function assigning day, i, to the regime, j. The regimes are given by disjunctive partition cells P_j that creates the whole in-sample forecast space, G:

[6] Tree-HAR offers the possibility to exploit exogenous variables to split the tree. Audrino and Corsi (2010) uses the $S\&P500$ as a possible separator.

$$G = \coprod_{j=1}^{k} P_j, \quad P_i \cap P_j = \emptyset (i \neq j) \tag{10}$$

We limit the minimum size of each partition P to 22.

3.3. The branching algorithm

We use the idea of a binary tree to branch our tree-HAR model. The branching is based on a negative quasi-maximum likelihood (QML) value calculated from the estimated parameters in equation 8 and 9. We maximise the QML for the tree for every iteration. The negative quasi-log likelihood function is given as following:

$$-l = \frac{n}{2} log(2\pi) + \frac{1}{2} \sum_{t=1}^{n} log(\sigma_t^2) + \frac{1}{2} \sum_{t=1}^{n} \frac{(\tilde{RC}_t - \mathbf{E_t}[\tilde{RC}_{t-1}])^2}{\sigma_t^2} \tag{11}$$

In addition, we restrict the numbers of regimes by implementing a Bayesian-Schwarz information criterion (BIC). The BIC uses the log likelihood values (l), number of estimated parameters (p), and sample size (n) to choose the best model. The BIC is given as following:

$$BIC = -2 \cdot (-l) + p \cdot log(n) \tag{12}$$

We proceed with an iterative branching algorithm to make a maximal span of the tree. The algorithm is as following:

1. The first step is to divide the sample in two. For each branch we estimate the conditional mean and sigma (equation 8 and 9) and compute QML (l). We then run an iterative process and create all possible subsets with two branches of the sample, given a step size.
2. Choose the branching option that minimises the QML-value. Calculate the BIC value.
3. Repeat this process by rebranching on one of the old branches until the BIC value deteriorates.

In most cases, an additional regime would improve the log-likelihood value of the model, but by implementing a BIC we penalise the branching for the increase in the number of free parameters to be estimated (see figure 3). The maximal tree, given by the algorithm, is then pruned by combining branches in a way that improves the BIC value. This is done by considering all possible combinations of regimes. The process stops when no further improvement of the BIC value is possible. The motivation is that regimes separated in time might share the same underlying parameters, which the original algorithm would fail to capture.

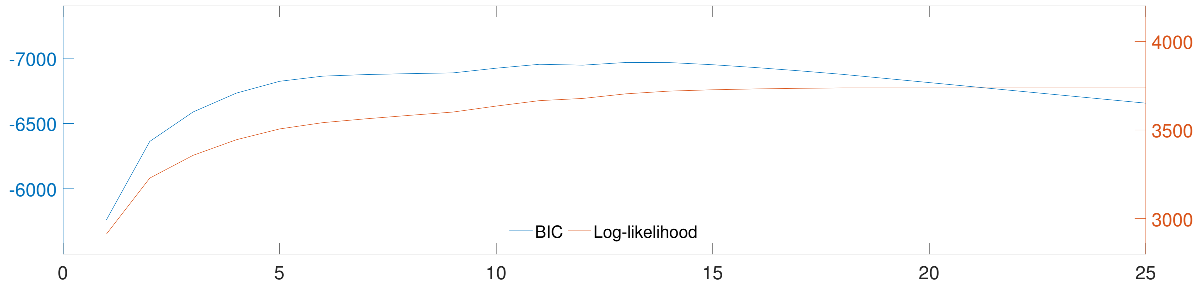


Figure 3. Plot of BIC and log-likelihood values from the tree-HAR model of the realized volatility of AUD/USD.

3.4. *Model extensions with SVI*

We extend the models by adding the daily SVI as an exogenous variable. We will hereby denote the respective models including SVI data as HAR-SVI and tree-HAR-SVI. The extended models are given as the following for the realized volatility:

$$RV_{t+1} = C + \beta^d RV_t^d + \beta^w \; RV_t^w + \beta^m RV_t^m + \beta^{SVI} SVI_t + \epsilon_t \tag{13}$$

$$\mathbf{E_t}[\tilde{RV}_{t+1}] = \sum_{j=1}^{k} \left(\alpha_j + \beta_j^d \tilde{RV}_t^d + \beta_j^w \tilde{RV}_t^w + \beta_j^m \tilde{RV}_t^m + \beta_j^{SVI} SVI_t \right) I_{i,j} \tag{14}$$

The same specification is used also for realized correlations. We combine the SVI data to include the respective currency searches. For instance, if we forecast the daily realized volatility of "GBP/USD" rate, we include the combined SVI variable[7] of the searches for "GBP" and "USD", as given in equation 15. For daily realized correlation, we use currency triplets consisting of three search queries. The SVI variable for the daily realized correlation between "GBP/USD" and "AUD/USD" includes a combination of all the three searches, "GBP", "USD" and "AUD". However, since both FX rates are tied to the US dollar, we weight the USD 50%.

$$SVI_t^{RV} = \frac{1}{2}(SVI_{C1,t} + SVI_{USD,t}) \tag{15}$$

$$SVI_t^{RC} = \frac{1}{2}(SVI_{C1,t} + SVI_{USD,t}) + \frac{1}{2}(SVI_{C2,t} + SVI_{USD,t}) \tag{16}$$

[7]Comment: we tested the models with separate SVI variables for each search query. We omit to present these results as they provided very similar results to the combined SVI models. In addition, the SVI coefficients was in general weighted similarly across the different search queries. This could indicate that for our data sample, the SVI variable was not able to differentiate which currency that was causing the movements. However, this is an interesting topic for further research.

4. Results

We investigate the forecasting abilities of the tree-HAR model for both realized volatility and correlation. The main results are presented in the following four subsections: in-sample results, out-of-sample results, structural breaks and Google search data. Due to the large scope of our study, the main results will only present figures and tables exemplifying the most interesting findings. Results for the whole sample are reported in the appendix.

4.1. In-sample

Tables 3 shows a summary of Mean Absolute Error (MAE), Mean Squared Error (MSE) and coefficient of determination (R^2) for the various models for both realized volatilities and correlations. We can clearly see the advantage of including cascades of different time horizons in modelling both realized volatility and correlation. The gap between the model fits are especially apparent from the AR(1) model to the HAR model.

The results are exclusively in favour for the tree-HAR model in all categories. The error terms (MAE and MSE) are lower for both RV and RC. There is also a substantial improvement in the R-squared value. This confirms that historical data provide as a good proxy for future volatility and correlation. Complete tables of the results are presented in appendix D.1. In order to illustrate the difference between the HAR and the tree-HAR model, we plot the fitted correlations together with the actual correlations for the currency pairs AUD/USD and EUR/USD. Figure 4 is zoomed in around regime shifts to highlight the tree-HAR's ability to quickly react to shifts in the correlation. The tree-HAR model exhibits excellent reactions in modelling the transition between regimes.

This improvement is expected considering the structure of the model. The tree-HAR model switches between HAR-models with different parameters for the different regimes. By differentiating periods, the tree-HAR model is able to better isolate regimes with different underlying dynamics. We will come back to this in the next subsection 4.4.

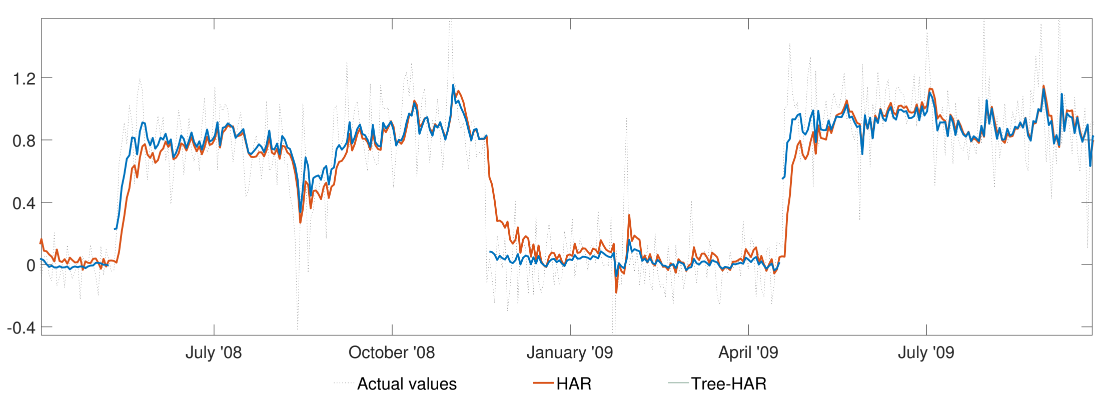


Figure 4. Time-series plot of in-sample estimates for the HAR model and tree-HAR model.

Table 3: In-sample results for realized volatility and correlation.

Note: This table reports MAE, MSE and R-squared values in-sample for realized volatility and correlation. Values are given in 10^{-2}.

Panel A: in-sample results for realized volatility.						
		AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD	MSE	0.571	0.465	0.436	0.423	0.399
	MAE	5.466	4.780	4.724	4.623	4.546
	R^2	64.97	71.65	73.41	74.16	75.58
EUR/USD	MSE	0.393	0.294	0.287	0.279	0.276
	MAE	4.660	3.996	3.973	3.894	3.899
	R^2	43.79	58.13	59.03	60.18	60.58
GBP/USD	MSE	0.327	0.239	0.234	0.229	0.217
	MAE	4.260	3.579	3.565	3.521	3.467
	R^2	58.15	69.45	70.14	70.71	72.29
NZD/USD	MSE	0.684	0.573	0.550	0.532	0.513
	MAE	6.204	5.506	5.452	5.378	5.285
	R^2	57.55	64.55	65.97	67.08	68.26
CAD/USD	MSE	0.393	0.294	0.288	0.283	0.276
	MAE	4.667	3.950	3.933	3.909	3.881
	R^2	51.00	63.49	64.16	64.84	65.72
CHF/USD	MSE	0.668	0.545	0.532	0.487	0.453
	MAE	5.382	4.650	4.662	4.509	4.541
	R^2	28.22	41.78	43.20	47.91	51.65
JPY/USD	MSE	0.656	0.560	0.525	0.548	0.470
	MAE	5.715	5.164	5.064	5.132	4.915
	R^2	39.40	48.43	51.66	49.60	56.70
Panel B: extract of in-sample results for realized correlation.						
		AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-CAD/USD	MSE	7.65	5.70	5.60	5.32	5.20
	MAE	21.67	18.23	18.29	17.74	17.50
	R^2	44.42	61.40	61.76	63.75	65.06
EUR/USD-CHF/USD	MSE	15.68	12.50	13.00	11.51	10.71
	MAE	29.96	26.29	26.70	25.28	24.72
	R^2	56.31	65.24	68.12	68.05	71.11
EUR/USD-GBP/USD	MSE	8.48	7.20	7.10	7.32	6.86
	MAE	22.79	20.92	20.86	20.76	20.49
	R^2	10.34	24.40	24.62	25.54	26.14
NZD/USD-JPY/USD	MSE	9.01	8.00	7.90	7.71	7.46
	MAE	23.57	21.71	21.74	21.56	21.07
	R^2	40.40	52.45	52.54	53.58	54.43

4.2. Out-of-sample

We perform out-of-sample forecasting using all past data for the AR(1) and HAR model. The tree-HAR model follows a different dynamic that allows it to combine regimes that are not continuous over time. This procedure is computationally more intensive compared to the HAR model. In order to mitigate the necessary number of computations, we only recalculate the tree every 22 days (approximately every month) as previously done by Audrino and Corsi (2010).

We apply a Model Confidence Set (MCS) test (Hansen et al., 2011) to compare the models' out-of-sample predictions. The MCS's null hypothesis of Equal Predictive Ability (EPA) is tested with a sequence of tests assigned by the user. The MCS procedure will report one, or a group of models, that give the best results, given a significance level. We use Absolute Error (AE) and Square-Error as loss functions for both realized volatilities and correlations.

Table 4 reports a summary of the out-of-sample results. We measure the fit of the out-of-sample results as MAE and MSE. The AR(1) model still shows poor performance compared to the HAR models. However, for the daily out-of-sample forecasts, the HAR model and tree-HAR model performs quite similarly. This is confirmed from the MCS test which cannot reject the null hypothesis of equal predictive abilities for both models. A possible explanation is that the tree-HAR model's out-of-sample prediction is reduced to a HAR prediction where the model restricts itself to predict based on values given in the relevant regimes. The standard HAR model would on the other hand benefit from all relevant values.

Table 5 displays a summary of the results for the MCS test on both realized volatility and correlation. The table documents the number of times each model is included in the superior set provided by the test. The test have greater difficulty in differentiating the models for the realized correlation than for the volatility. For the realized correlation, the MCS procedure shows a general preference for the HAR models. In turn, the HAR-SVI model surpasses the HAR model. The selection is more clear for the realized volatility where there is a clear preference for the HAR-SVI model. Interestingly, and expected, the AR model is not included in any of the superior sets. The results suggests that the SVI variable is indeed a beneficial supplement to the HAR model. The significance of the SVI is elaborated in subsection 4.3.

Table 4: Out-of-sample results.

Note: this table reports MSE and MAE for the out of sample results for realized volatility and correlation. All values are given in 10^{-2}.

Panel A: out-of-sample results for realized volatility.						
		AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD /USD	MSE	0.589	0.479	0.451	0.491	0.464
	MAE	5.486	4.882	4.794	4.932	4.835
EUR /USD	MSE	0.385	0.284	0.279	0.296	0.289
	MAE	4.716	3.990	3.963	4.049	4.026
GBP /USD	MSE	0.334	0.244	0.239	0.253	0.246
	MAE	4.251	3.596	3.586	3.684	3.663
NZD /USD	MSE	0.680	0.576	0.555	0.591	0.568
	MAE	6.222	5.572	5.529	5.653	5.589
CAD /USD	MSE	0.396	0.295	0.290	0.300	0.293
	MAE	4.683	3.948	3.938	3.997	3.977
CHF /USD	MSE	0.682	0.556	0.544	0.574	0.557
	MAE	5.433	4.671	4.656	4.698	4.716
JPY /USD	MSE	0.662	0.562	0.532	0.576	0.541
	MAE	5.918	5.220	5.136	5.363	5.180
Panel B: extract of the out-of-sample results for realized correlation.						
		AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-CAD/USD	MAE	21.88	18.21	18.10	18.30	18.26
	MSE	7.81	5.64	5.58	5.70	5.64
EUR/USD-GBP/USD	MAE	22.70	20.78	20.86	20.82	20.82
	MSE	8.36	7.04	7.06	7.06	7.04
EUR/USD-CHF/USD	MAE	30.66	26.37	26.38	26.66	26.75
	MSE	15.87	12.74	12.66	12.88	12.95
NZD/USD-JPY/USD	MAE	24.25	21.69	21.66	21.94	21.82
	MSE	9.59	7.86	7.79	8.09	7.96

Table 5: MCS

This table reports the frequency of which a model is included in the superior set provided using the Model Confidence Set (MCS) procedure with absolute error (AE) and square error (SE) as loss functions.

Panel A: MCS for realized volatility.				
	AE		SE	
	$\alpha = 0.15$	$\alpha = 0.25$	$\alpha = 0.15$	$\alpha = 0.25$
AR(1)	0	0	0	0
HAR	1	0	0	0
HAR-SVI	7	7	7	7
tree-HAR	0	0	0	0
tree-HAR-SVI	4	3	0	0
Panel B: MCS for realized correlation.				
	AE		SE	
	$\alpha = 0.15$	$\alpha = 0.25$	$\alpha = 0.15$	$\alpha = 0.25$
AR(1)	0	0	0	0
HAR	20	18	18	15
HAR-SVI	21	20	21	20
tree-HAR	16	8	10	8
tree-HAR-SVI	16	10	11	8

4.3. Google search data

In this part of our study we investigate the relationship between daily Google search query data and FX volatility and correlation. However, we find a significant improvement to the HAR models when we include the SVI variable. In the considered single-state HAR regressions, the SVI is nearly always significant at a 1% level. This indicates a relationship between Google search queries and movements in the FX rates. These results are in accordance with what Dimpfl and Jank (2015) found studying the equity market. From the in-sample and out-of-sample results, as presented earlier, we can see that the SVI model extensions improve both the HAR and tree-HAR model for the grand majority of the regressions.

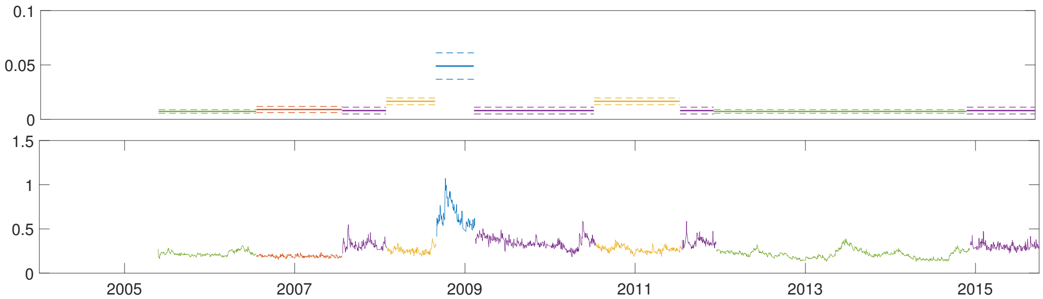


Figure 5. SVI coefficients for the realized volatility of AUD/USD in the different regimes.

However, an interesting aspect reveals itself from studying the regression coefficients of the tree-HAR regimes. Figure 5 displays the SVI coefficients for the different regimes, provided with confidence bands to include the significance of the coefficients. As seen here, the coefficients of the SVI variables are largest and most significant during volatile periods, like the quantitative easing of late 2008/early 2009. These results suggest that investors' search for information increases during uncertain times, making the impact of Google search query data on the model more significant. This is in accordance with previous literature reporting a strong relationship between market riskiness and investor attention (Da et al., 2011, Kita and Wang, 2012, Vlastakis and Markellos, 2012).

An interesting observation is that regimes are not affected by adding the SVI variable — they remain the same for the tree-HAR and tree-HAR-SVI model. Having said that, the MCS test shows overall a preference for the SVI model extensions. This indicates that Google search query data captures FX market movements to a certain extent. Although its significant presence in the models, the combined SVI extension's impact remains small with coefficients at the 1% significance level. Even though FX-rates are unlikely to be driven by private information (Kita and Wang, 2012), the difficulties of finding a structural model that can take profitable, risk free advantage of information lags in the market might indicate that publicly available information, and the cost of searching for information, is rapidly integrated into FX rates.

4.4. *Structural breaks*

An interesting aspect of the tree-HAR model is the division of regimes. As explained in section 3, the maximal tree is created by maximising the negative quasi log-likelihood function, only stopping when the BIC value deteriorates (see figure 3). Unfortunately, as a result of the quantity of the time series we are studying, we are not able to present in-depth analysis of the regimes. Nonetheless, we would like to present the regimes we have documented and some thoughts about the structural breaks.

In table 6 we can see how the BIC value is improved (minimised) by pruning regimes found in the maximal tree for the tree-HAR model. Note that the negative quasi log-likelihood value decreases by the pruning, while the BIC value increases. For most of the currencies, the model is able to combine similar regimes and reduce the overall number of regimes.

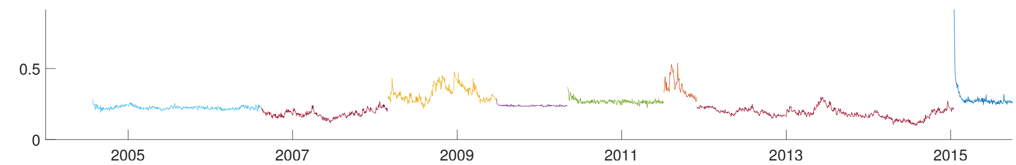
Table 6: Number of regimes for realized volatility and correlation.

Note: The LOG-value reported is the negative quasi-log likelihood value. Bayesian Information Criterion (BIC) is improved (minimised) by pruning the tree.

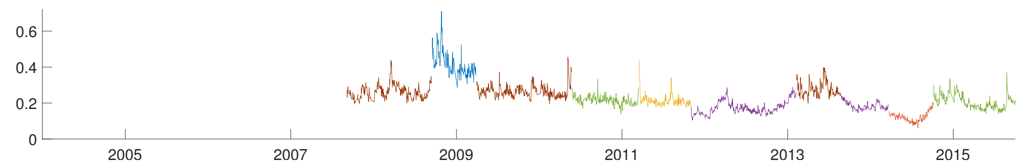
	Max tree			Pruned tree		
	Regimes	LOG	BIC	Regimes	LOG	BIC
AUDUSD	10	3638.3	-6882.8	5	3597.0	-6997.0
CADUSD	7	4633.2	-8986.2	6	4631.6	-9023.0
CHFUSD	8	3926.4	-7534.7	7	3925.5	-7572.7
EURUSD	7	4591.7	-8903.3	4	4572.9	-8985.8
GBPUSD	6	4712.1	-9185.4	4	4682.8	-9206.4
JPYUSD	10	2610.2	-4840.7	6	2597.7	-4966.8
NZDUSD	7	2655.1	-5041.1	4	2625.1	-5096.5
	Max tree			Pruned tree		
	Regimes	LOG	BIC	Regimes	LOG	BIC
AUD/USD-CAD/USD	6	190.77	-145.19	5	184.66	-172.36
EUR/USD-GBP/USD	3	-219.44	558.30	3	-219.44	558.30
EUR/USD-CHF/USD	16	-728.97	2094.26	9	-805.59	1969.11
NZD/USD-JPY/USD	5	-317.98	826.54	3	-330.81	775.97

To better visualise the regimes, we discern the particular regimes with colours in figure 6 and 7. For the AUD/USD and EUR/USD correlation series, the regimes seem to have an economical interpretation as the correlation distinctly shifts back and forth between regimes. The most distinct regimes are during financial unstable periods, such as the financial crisis in 2007-2009.

Another interesting aspect that arises from the tree-HAR model is the creation of different regimes. An interesting property with the regimes is that they are able to capture the effect of well known prior events. Common for all the currency pairs is a high volatility regime seen around the end of 2008 and start of 2009. Incidentally, this period corresponds to the start of the US Federal Reserve's quantitative easing in November 2008. Comparatively, the JPY/USD has a regime, starting in January 2013, that corresponds to reactions to a new Japanese monetary policy (of Japan, 2013).

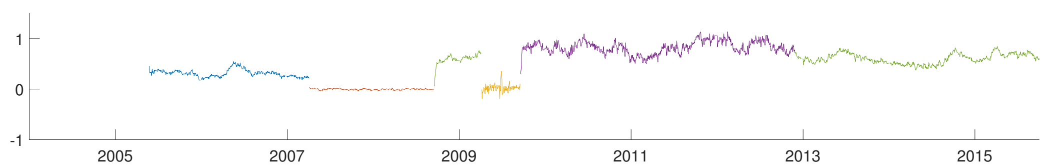


(a) CHF/USD

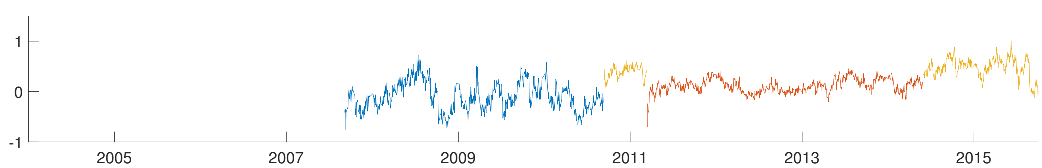


(b) JPY/USD

Figure 6. Regime plots of realized volatility.



(a) AUD/USD and CAD/USD

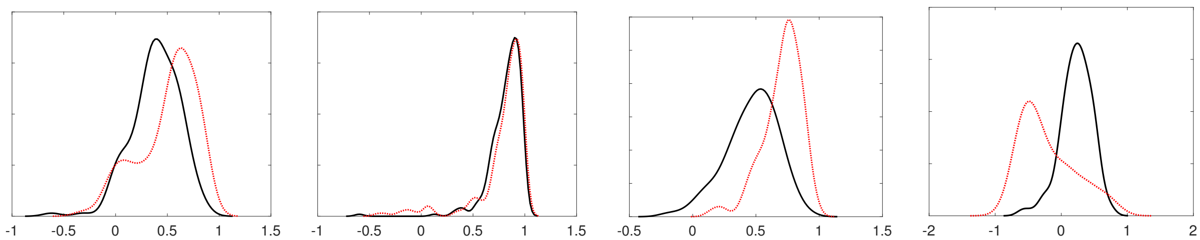


(b) NZD/USD and JPY/USD

Figure 7. Regime plots of realized correlation.

Another noteworthy pair is the CHF/USD. It experiences higher volatility on a couple occasions. The model neatly defines the two periods in their respective regimes. The first period in questions starts in September 2011 and corresponds to the decision to peg the CHF to the EUR in the wake of the CHF's strengthening against the EUR related to the European debt crisis (SNB, 2011). The next high volatility regime seen for the CHF starts with the Swiss National Bank's decision to discontinue the minimum exchange rate of the CHF in January 2015 (SNB, 2015).

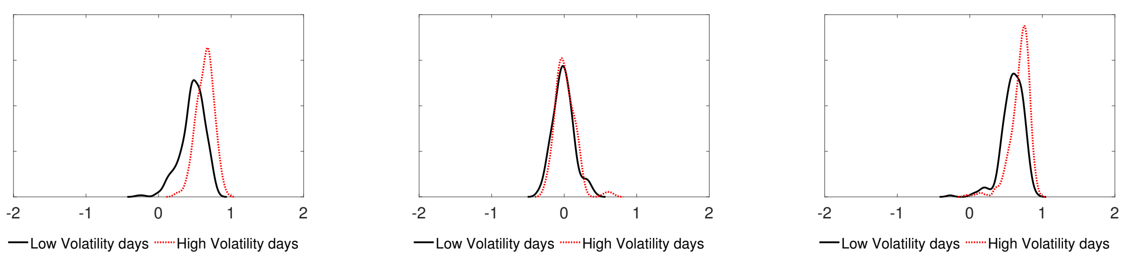
To further investigate FX market dynamics within regimes, we plot correlation during high and low volatility days to check for interesting patterns. High volatility days are defined as days where both base volatilities for the correlation are above the 90^{th} percentile, and low volatility days are days below the 10^{th} percentile. Earlier literature finds tendencies of higher correlation between assets on high volatility days, than on low volatility days (Andersen et al., 2003). Similarly, we find that for most currency triplets, this is the case. However, the pattern seem to be the opposite for correlations including JPY (figure 8(d)).



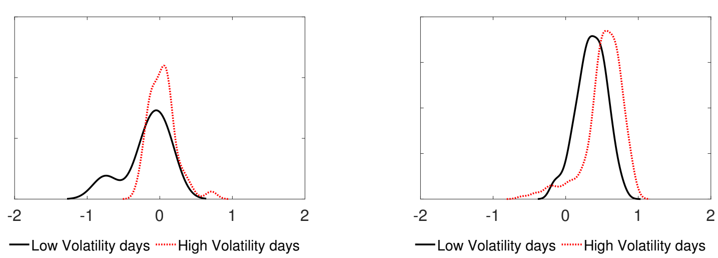
(a) AUD-CAD (b) EUR-CHF (c) EUR-GBP (d) NZD-JPY

Figure 8. Realized correlation on high/low volatility days. The correlation is denoted as the currencies tied to the US dollar. The plots are made from a Kernel distribution fit function.

In order to see what is driving the different correlation patterns for high and low volatile days, we also investigate the distributions of for each regime separately for the AUD/USD and CAD/USD. One interesting finding is that the difference between the correlation during the high/low volatility is lower within the regimes than for the whole sample period. This is a strong indication that each regime detected by the tree-HAR model is rather homogeneous, whereas there are large differences among regimes.



(a) Regime 1 (b) Regime 2 (c) Regime 3



(d) Regime 4 (e) Regime 5

Figure 9. Correlation during high and low volatility within each regime for AUD/USD and EUR/USD.

5. Concluding remarks

This paper investigates the use of a flexible tree-HAR model in forecasting foreign exchange (FX) market volatility and correlation. We empirically test the forecasting abilities of the model for seven of the most traded currencies (AUD, CAD, CHF, EUR, GBP, JPY and NZD) against the US dollar. Furthermore, we investigate the relationship between investor attention, here measured as daily Google query searches, and structural breaks in foreign exchange rates.

First, we find the tree-HAR model to possess excellent abilities to reproduce realized volatility and correlation in-sample. Our flexible tree-HAR model outperforms simpler, single-state autoregressive models, such as AR(1) and the HAR model. However, for the out-of-sample predictions, the Model Confidence Set (MCS) test slightly favours the HAR-model over the tree-HAR. For the extended models, we find that Google searches improve not only in-sample fit of the models, but also the out-of-sample forecasting performance. These results hold for the grand majority of the data samples.

We see a clear tendency that structural breaks for both volatility and correlation occur during financial turbulence. By extending the HAR and tree-HAR model with an exogenous SVI variable, we see indications of investors seeking more information during unstable times. The SVI coefficients for the tree-HAR model have significant SVI variables for the majority of the regressions during volatile times.

We believe that the tree-HAR model possesses a greater out-of-sample performance than we are able to present. In addition, it would be interesting to dwell further on the dynamics of the regimes. As commented, investors' search for attention seems to increase during unstable times – when information is most needed. Google searches seem to be an even better predictor for future FX dynamics during these times. We leave this for future research.

While this study does not offer a conclusive answer to what is causing structural breaks in FX markets, we quantitatively document the existence of such breaks and give indications of a relationship between financial unstability and structural breaks. This research raises important questions about the need of flexible models in order to better reproduce complex FX dynamics.

References

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. Realized volatility and correlation. *Paper*, 1999.

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, Vol.96: 42–55, 2001.

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, Vol.71:529–626, 2003.

Daniel Andrei and Michael Hasler. Investor attention and stock market volatility. *Review of Financial Studies*, 2014.

Christian Anvik and Kristoffer Gjelstad. Just google it. forecasting norwegian unemployment figures with web queries. *Master thesis*, 2010.

Net Applications. Desktop search engine market share, 2015. URL `https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0`.

Nikolaos Askita and Klaus F. Zimmermann. The internet as a data source for advancement in social sciences. *Working paper*, 2015.

Francesco Audrino and Peter Bühlmann. Tree structured generalized autoregressive conditional heteroscedastic models. *Journal of the Royal Statistical Society*, Series B 63:579–625, 2001.

Francesco Audrino and Fulvio Corsi. Realized correlation tick-by-tick. *paper*, 2007.

Francesco Audrino and Fulvio Corsi. Modeling tick-by-tick realized correlations. *Computational Statistics and Data Analysis*, Vol.54:2372–2382, 2010.

Francesco Audrino and Fabio Trojani. Estimating and predicting multivariate volatility thresholds in global stock markets. *Journal of Applied Econometrics*, Vol.21(3):345–369, 2006.

Geert Bakaert and Robert J. Hodrick. On biases in the measurement of foreign exchange risk premiums. *Journal of international money and finance*, Vol.12:115–138, 1993.

Matthias Bank, Martin Larch, and Georg Peter. Google search volume and its influence on liquidity and returns of german stocks. *Financial markets and portfolio management*, Vol.25(3):239–264, 2011.

Nicolas P.B. Bollen, Stephen F. Gray, and Robert E. Whaley. Regime switching in foreign exchange rates: Evidence from currency option prices. *Journal of Econometrics*, Vol.94: 239–276, 2000.

Roxana Chiriac and Valeri Voev. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947, 2011.

Hyunyoung Choi and Hal Varian. Predicting initial claims for unemployment benefits. *paper*, 2011.

Fulvio Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, page nbp001, 2009.

Zhi Da, Joseph Engelberg, and Pengjie Gao. In search of attention. *The journal of finance*, Vol. 66(5), 2011.

David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association*, Vol.79:427–431, 1979.

Thomas Dimpfl and Stephan Jank. Can internet search queries help to predict stock market volatility? *European Financial Management*, 2015.

Michael S. Drake, Darren T. Roulstone, and Jacob R. Thornock. Investor information demand: evidence from google searches around earnings announcement. *Journal of Accounting Research*, Vol. 50(4), 2012.

Charles Engel and Craig S. Hakkio. The distribution of exchange rates in the ems. *International Journal of Finance and Economics*, Vol.1:55–67, 1996.

Charles Engel and James D. Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *The American Economic Review*, Vol.80(4):689–713, 1990.

Gunther Eysenbach. Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. *AMIA 2006 Symposium*, pages 253–280, 2006.

Eugene F. Fama. Forward and spot exchange prices. *Journal of Monetary Economics*, Vol.14: 319–338, 1984.

Christopher Fink and Thomas Johann. May i have your attention, please: The market microstructure of investor attention. *Working Paper*, 2014.

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, Vol.457(19):1012–1015, 2009.

John Goddard, Arben Kita, and Qingwei Wang. Investor attention and fx market volatility. *Journal of International Financial Markets, Institutions and Money*, Vol.38:79–96, 2015.

James D. Hamilton and Raul Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, Vol.64:307–333, 1994.

Peter R. Hansen, Asger Lunde, and James M. Nason. The model confidence set. *Econometrica*, Vol.79(2):453–497, 2011.

Arben Kita and Qingwei Wang. Investor attention and fx market volatility. *Paper*, 2012.

Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: quantifying the relationship between phenomena of the internet era. *Scientific reports*, Vol.3, 2013.

Rajendra Kulkarni, Kingsley E. Haynes, R. R. Stough, and Jean H. P. Paelinck. Forecasting housing prices with google econometrics. *Research Paper*, 2009.

Chun Liu and John M Maheu. Are there structural breaks in realized volatility? *Journal of Financial Econometrics*, 6(3):326–360, 2008.

Michael McAleer and Marcelo C. Medeiros. A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics*, Vol.147:104–119, 2008.

Richard Meese and Kenneth Rogoff. The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? *National Bureau of Economic Research*, pages 67–112, 1983.

Bank of Japan. Joint statement of the government and the bank of japan on overcoming deflation and achieving sustainable economic growth, 2013. URL `https://www.boj.or.jp/en/announcements/release_2013/k130122c.pdf`.

Bing Pan, Chenguang (Doris) Wu, and Haiyan Song. Forecasting hotel room demand using search engine data. *Paper*, 2012.

Philip M. Polgreen, Yiling Chen, David M. Pennock, and Forrest D. Nelson. Using internet searches for influenza surveillance. *Clinical Infectious Diseases*, Vol.47:1443–1448, 2008.

Shauna Reilly, Sean Richey, and J. Benjamin Taylor. Using google search data for state politics research: An empirical validity test using roll-off data. *State Politics & Policy Quarterly*, Vol.12(2):146–159, 2012.

Dimche Risteski and Danco Davcev. Can we use daily internet search query data to improce predicting power of egarch models for financial time series volatility. In *International conference on computer science and information systems*, Dubai (UAE), Oct 2014.

Geoffrey Peter Smith. Google internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, Vol.9:103–110, 2012.

SNB. Swiss national bank sets minimum exchange rate at chf 1.20 per euro, 2011. URL `http://www.snb.ch/en/mmr/reference/pre_20110906/source/pre_20110906.en.pdf`.

SNB. Swiss national bank discontinues minimum exchange rate and lowers interest rate to –0.75 URL http://www.snb.ch/en/mmr/reference/pre_20150115/source/pre_20150115.en.pdf.

Stephen J. Taylor. Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*, Vol.3:159–170, 1987.

Nikolaos Vlastakis and Raphael N. Markellos. Information demand and stock market volatility. *Journal of Banking & Finance*, Vol.36:1808–1821, 2012.

Simeon Vosen and Torsten Schmidt. Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting*, Vol.30(6):565–578, 2011.

Appendices

Appendix A Dickey Fuller test

Table A.1: Dickey-Fuller Unit Root Test.

Note: This table reports the results from a Dickey-Fuller Unit Root Test with 10 augmentation lags. All daily realized volatility currency pairs reject the null hypothesis at a 0.01% significance level (critical value: -3.9669). As a result, we consider the time-series as stationary in our modelling.

	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10
AUD/USD	-22.56	-15.44	-12.24	-9.52	-7.46	-7.09	-6.80	-6.50	-5.92	-5.42	-5.23
EUR/USD	-26.48	-17.73	-13.54	-10.61	-8.44	-7.80	-7.19	-6.74	-6.23	-5.50	-5.25
GBP/USD	-24.28	-15.43	-11.85	-9.14	-7.18	-6.52	-5.96	-5.58	-5.23	-4.65	-4.55
NZD/USD	-25.22	-17.44	-13.51	-10.93	-8.53	-8.22	-8.05	-7.52	-6.96	-6.49	-6.11
CAD/USD	-26.00	-16.30	-12.50	-10.20	-7.70	-7.03	-6.40	-5.82	-5.55	-5.11	-4.77
CHF/USD	-29.54	-18.76	-14.25	-11.50	-9.17	-8.48	-7.71	-7.15	-6.61	-6.02	-5.72
JPY/USD	-28.68	-19.25	-15.10	-12.60	-10.66	-10.31	-9.44	-8.99	-8.38	-7.51	-7.05

Appendix B High-frequency data

Figure B.1 illustrates the difference in the correlation time-series from using daily and high-frequency FX rates. The high-frequency data absorbs much more information than the daily data , giving a clearer indication of structural breaks. Thus, we want to highlight the importance of using high-frequency data in order to get more precise observations and estimations of FX market movements.

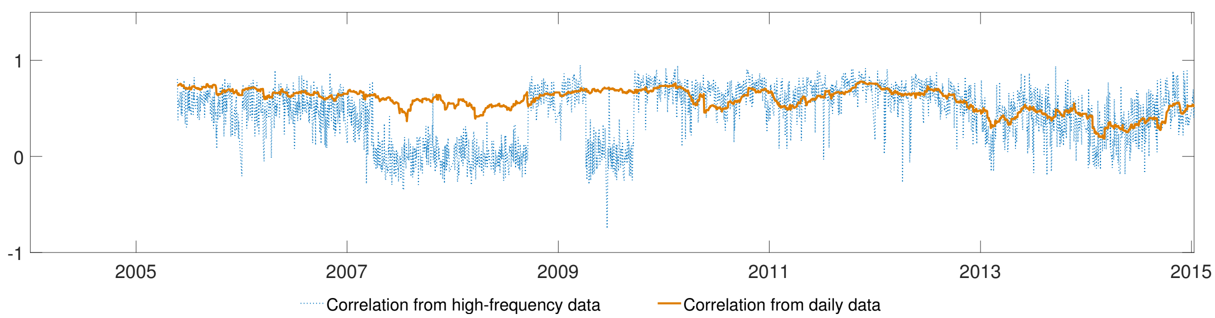


Figure B.1. Comparison of daily and high-frequency correlation for AUD/USD and EUR/USD.

Appendix C Extended stylised facts

Table C.1: Comparison of raw daily SVI and normalised daily SVI

Panel A: Descriptive statistics of raw daily SVI						
	Mean	St.dev	Kurtosis	Skewness	Max	Min
USD AUD	5.31	3.10	4.02	1.19	20.79	0.47
USD CAD	7.72	1.30	7.66	1.91	16.75	4.59
USD CHF	4.50	2.06	13.37	1.66	28.90	0.42
USD EUR	5.81	3.13	3.66	0.99	19.91	0.00
USD GBP	5.66	3.37	3.35	0.95	20.22	0.34
USD JPY	6.51	3.01	3.09	0.99	20.70	1.40
USD NZD	5.88	2.86	4.41	1.30	21.32	1.19
Panel B: Descriptive statistics of normalised daily SVI						
	Mean	St.dev	Kurtosis	Skewness	Max	Min
USD AUD	0.07	0.97	5.17	0.87	5.24	-2.79
USD CAD	0.04	0.96	4.34	0.13	5.64	-4.49
USD CHF	0.10	1.01	15.66	1.62	12.89	-3.80
USD EUR	0.05	0.90	6.12	0.72	6.44	-4.42
USD GBP	0.06	0.91	3.98	0.49	4.51	-3.11
USD JPY	0.10	0.98	6.03	0.99	6.22	-2.96
USD NZD	0.10	0.95	3.70	0.22	3.92	-3.63

Table C.2: Comparison of RV and log-RV

Panel A: Descriptive statistics of daily realized volatility						
	Mean	St.dev	Kurtosis	Skewness	Max	Min
AUD/USD	0.329	0.193	29.795	3.820	3.042	0.049
EUR/USD	0.247	0.113	8.271	1.698	1.071	0.006
GBP/USD	0.236	0.119	13.119	2.525	1.264	0.041
NZD/USD	0.359	0.180	19.282	2.934	2.564	0.093
CAD/USD	0.266	0.135	9.184	1.850	1.268	0.050
CHF/USD	0.275	0.181	803.970	21.506	7.370	0.041
JPY/USD	0.253	0.132	17.965	2.830	1.541	0.050
Panel B: Descriptive statistics of logarithmic scaled daily realized volatility						
	Mean	St.dev	Kurtosis	Skewness	Max	Min
AUD/USD	0.276	0.123	12.216	2.313	1.397	0.047
EUR/USD	0.217	0.086	5.775	1.226	0.728	0.006
GBP/USD	0.208	0.088	8.811	1.902	0.817	0.041
NZD/USD	0.299	0.117	9.069	1.874	1.271	0.089
CAD/USD	0.231	0.099	5.863	1.293	0.819	0.049
CHF/USD	0.238	0.095	61.337	4.046	2.125	0.040
JPY/USD	0.221	0.095	9.945	1.892	0.933	0.049

Table C.3: Realized correlation summary statistics

Note: This table reports the descriptive statistics of the realized correlation.

	Mean	St.dev	Kurtosis	Skewness	Max	Min
AUD/USD-EUR/USD	0.440	0.297	2.522	-0.667	0.950	-0.789
AUD/USD-GBP/USD	0.386	0.260	2.925	-0.685	0.921	-0.999
AUD/USD-NZD/USD	0.593	0.340	2.955	-1.155	0.973	-0.679
AUD/USD-CAD/USD	0.422	0.292	2.466	-0.556	0.934	-0.655
AUD/USD-CHF/USD	0.350	0.287	2.611	-0.454	0.916	-0.935
AUD/USD-JPY/USD	0.111	0.349	2.517	-0.155	0.940	-0.817
EUR/USD-GBP/USD	0.633	0.176	4.476	-1.044	0.982	-0.234
EUR/USD-NZD/USD	0.502	0.217	3.715	-0.813	0.983	-0.599
EUR/USD-CAD/USD	0.430	0.236	3.517	-0.662	0.947	-0.751
EUR/USD-CHF/USD	0.810	0.197	9.922	-2.303	0.999	-0.424
EUR/USD-JPY/USD	0.205	0.353	2.265	-0.268	0.962	-0.874
GBP/USD-NZD/USD	0.437	0.199	3.750	-0.560	0.984	-0.874
GBP/USD-CAD/USD	0.370	0.227	3.448	-0.587	0.928	-0.692
GBP/USD-CHF/USD	0.542	0.225	3.503	-0.818	0.974	-0.361
GBP/USD-JPY/USD	0.146	0.313	2.406	-0.180	0.928	-0.701
NZD/USD-CAD/USD	0.486	0.208	3.391	-0.664	0.959	-0.475
NZD/USD-CHF/USD	0.378	0.255	3.339	-0.622	0.966	-0.753
NZD/USD-JPY/USD	0.097	0.351	2.341	-0.165	0.962	-0.929
CAD/USD-CHF/USD	0.351	0.250	3.475	-0.527	0.943	-0.790
CAD/USD-JPY/USD	0.052	0.321	2.474	-0.052	0.914	-0.992
CHF/USD-JPY/USD	0.303	0.320	2.496	-0.372	0.939	-0.764

Table C.4: Fisher-transformed realized correlation summary statistics

Note: This table reports the descriptive statistics of the Fisher transformed realized correlation.

	Mean	St.dev	Kurtosis	Skewness	Max	Min
AUD/USD-EUR/USD	0.534	0.393	2.496	-0.132	1.831	-1.070
AUD/USD-GBP/USD	0.442	0.329	16.258	-1.146	1.593	-4.113
AUD/USD-NZD/USD	0.829	0.518	2.334	-0.549	2.143	-0.826
AUD/USD-CAD/USD	0.509	0.386	2.404	-0.009	1.687	-0.784
AUD/USD-CHF/USD	0.406	0.357	3.141	-0.014	1.562	-1.696
AUD/USD-JPY/USD	0.127	0.405	3.242	0.036	1.734	-1.148
EUR/USD-GBP/USD	0.802	0.305	3.603	0.143	2.348	-0.239
EUR/USD-NZD/USD	0.595	0.304	3.720	0.038	2.374	-0.691
EUR/USD-CAD/USD	0.497	0.309	3.523	0.021	1.806	-0.976
EUR/USD-CHF/USD	1.376	0.609	4.090	0.485	3.696	-0.452
EUR/USD-JPY/USD	0.242	0.420	2.885	0.101	1.972	-1.348
GBP/USD-NZD/USD	0.497	0.263	5.446	0.195	2.412	-1.352
GBP/USD-CAD/USD	0.415	0.278	3.544	-0.042	1.643	-0.851
GBP/USD-CHF/USD	0.665	0.339	3.244	0.129	2.161	-0.378
GBP/USD-JPY/USD	0.164	0.354	3.022	0.090	1.644	-0.870
NZD/USD-CAD/USD	0.568	0.286	3.177	0.072	1.932	-0.516
NZD/USD-CHF/USD	0.433	0.320	3.557	-0.013	2.030	-0.979
NZD/USD-JPY/USD	0.111	0.404	3.258	0.029	1.975	-1.651
CAD/USD-CHF/USD	0.397	0.310	3.879	0.028	1.763	-1.073
CAD/USD-JPY/USD	0.058	0.365	4.965	-0.175	1.549	-2.763
CHF/USD-JPY/USD	0.358	0.399	2.996	0.192	1.731	-1.006

Appendix D Full forecasting results

D.1 Full in-sample results

Table D.1: In-sample results for realized volatility

Note: This table reports MAE, MSE and R-squared values in-sample for realized volatility. Values are given in 10^{-2}.

		AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD	MSE	0.571	0.465	0.436	0.423	0.399
	MAE	5.466	4.780	4.724	4.623	4.546
	R^2	64.97	71.65	73.41	74.16	75.58
EUR/USD	MSE	0.393	0.294	0.287	0.279	0.276
	MAE	4.660	3.996	3.973	3.894	3.899
	R^2	43.79	58.13	59.03	60.18	60.58
GBP/USD	MSE	0.327	0.239	0.234	0.229	0.217
	MAE	4.260	3.579	3.565	3.521	3.467
	R^2	58.15	69.45	70.14	70.71	72.29
NZD/USD	MSE	0.684	0.573	0.550	0.532	0.513
	MAE	6.204	5.506	5.452	5.378	5.285
	R^2	57.55	64.55	65.97	67.08	68.26
CAD/USD	MSE	0.393	0.294	0.288	0.283	0.276
	MAE	4.667	3.950	3.933	3.909	3.881
	R^2	51.00	63.49	64.16	64.84	65.72
CHF/USD	MSE	0.668	0.545	0.532	0.487	0.453
	MAE	5.382	4.650	4.662	4.509	4.541
	R^2	28.22	41.78	43.20	47.91	51.65
JPY/USD	MSE	0.656	0.560	0.525	0.548	0.470
	MAE	5.715	5.164	5.064	5.132	4.915
	R^2	39.40	48.43	51.66	49.60	56.70

Table D.2: Number of regimes for realized volatility

Note: The LOG-value reported is the negative quasi-log likelihood value. Bayesian Information Criterion (BIC) is improved (minimised) by pruning the tree.

	Max tree			Pruned tree		
	Regimes	LOG	BIC	Regimes	LOG	BIC
AUDUSD	10	3638.3	-6882.8	5	3597.0	-6997.0
CADUSD	7	4633.2	-8986.2	6	4631.6	-9023.0
CHFUSD	8	3926.4	-7534.7	7	3925.5	-7572.7
EURUSD	7	4591.7	-8903.3	4	4572.9	-8985.8
GBPUSD	6	4712.1	-9185.4	4	4682.8	-9206.4
JPYUSD	10	2610.9	-4840.7	6	2597.7	-4966.8
NZDUSD	7	2655.1	-5041.2	4	2625.1	-5096.5

Table D.3: In-sample MAE comparison table for realized correlation

Note: This table reports the in-sample MAE values for the models compared. The data used was Fisher-transformed realized correlation. Values are given in 10^{-2}.

	AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-EUR/USD	22.33	18.75	18.72	18.10	18.13
AUD/USD-GBP/USD	20.71	17.27	17.24	16.75	16.84
AUD/USD-NZD/USD	23.98	20.37	20.22	19.44	19.04
AUD/USD-CAD/USD	21.67	18.23	18.29	17.74	17.50
AUD/USD-CHF/USD	22.14	19.02	19.04	18.23	18.39
AUD/USD-JPY/USD	23.06	21.02	20.97	20.44	20.63
EUR/USD-GBP/USD	22.79	20.92	20.86	20.76	20.49
EUR/USD-NZD/USD	20.64	18.97	18.97	18.79	18.58
EUR/USD-CAD/USD	22.15	19.81	19.83	19.61	19.55
EUR/USD-CHF/USD	29.96	26.29	26.70	25.28	24.72
EUR/USD-JPY/USD	26.59	23.88	23.77	23.04	23.05
GBP/USD-NZD/USD	19.31	18.07	18.07	17.91	17.76
GBP/USD-CAD/USD	20.44	18.68	18.65	18.57	18.36
GBP/USD-CHF/USD	24.38	21.65	21.76	21.60	21.33
GBP/USD-JPY/USD	24.54	22.03	22.08	21.69	21.71
NZD/USD-CAD/USD	20.82	18.61	18.68	18.55	18.16
NZD/USD-CHF/USD	20.95	18.92	18.93	18.76	18.63
NZD/USD-JPY/USD	23.57	21.71	21.74	21.56	21.07
CAD/USD-CHF/USD	22.09	20.07	20.08	19.76	19.69
CAD/USD-JPY/USD	22.80	20.91	20.93	20.55	20.64
CHF/USD-JPY/USD	24.88	22.11	22.01	21.64	21.50

Table D.4: In-sample MSE comparison table for realized correlation

Note: This table reports the MSE values in-sample for the models compared. The data used was Fisher-transformed realized correlation. Values are given in 10^{-2}.

	AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-EUR/USD	8.08	6.17	6.15	5.71	5.72
AUD/USD-GBP/USD	7.63	5.04	5.02	4.70	4.76
AUD/USD-NZD/USD	9.56	7.52	7.40	6.82	6.40
AUD/USD-CAD/USD	7.65	5.71	5.62	5.32	5.20
AUD/USD-CHF/USD	8.04	6.35	6.31	5.85	5.98
AUD/USD-JPY/USD	8.83	7.56	7.52	7.24	7.32
EUR/USD-GBP/USD	8.48	7.29	7.13	7.32	6.86
EUR/USD-NZD/USD	6.97	6.04	6.01	5.91	5.82
EUR/USD-CAD/USD	8.19	6.81	6.96	6.69	6.63
EUR/USD-CHF/USD	15.68	12.52	13.07	11.51	10.71
EUR/USD-JPY/USD	11.53	9.51	9.47	8.81	8.99
GBP/USD-NZD/USD	6.10	5.51	5.41	5.42	5.31
GBP/USD-CAD/USD	6.84	5.92	5.90	5.82	5.74
GBP/USD-CHF/USD	9.64	7.75	7.83	7.61	7.42
GBP/USD-JPY/USD	9.61	7.98	7.92	7.66	7.68
NZD/USD-CAD/USD	7.36	6.14	6.07	6.12	5.52
NZD/USD-CHF/USD	7.23	6.13	6.12	6.24	5.96
NZD/USD-JPY/USD	9.01	8.03	7.91	7.71	7.46
CAD/USD-CHF/USD	8.17	6.82	7.02	6.63	6.66
CAD/USD-JPY/USD	8.77	7.81	7.81	7.21	7.21
CHF/USD-JPY/USD	10.26	8.47	8.42	8.14	7.92

Table D.5: In-sample R^2 comparison table for realized correlation

Note: This table reports the R^2 values in-sample for the models compared. The data used was Fisher-transformed realized correlation. Values are given in 10^{-2}.

	AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-EUR/USD	44.53	60.63	60.77	62.30	62.99
AUD/USD-GBP/USD	26.99	49.77	50.00	53.23	52.79
AUD/USD-NZD/USD	56.41	72.59	72.79	75.04	76.23
AUD/USD-CAD/USD	44.42	61.40	61.76	63.75	65.06
AUD/USD-CHF/USD	35.34	50.50	50.65	54.12	53.08
AUD/USD-JPY/USD	42.27	52.38	52.44	54.14	55.59
EUR/USD-GBP/USD	10.34	24.40	24.62	25.54	26.14
EUR/USD-NZD/USD	17.77	35.92	36.39	37.22	37.26
EUR/USD-CAD/USD	14.99	29.14	32.84	30.56	31.06
EUR/USD-CHF/USD	56.31	65.24	68.12	68.05	71.11
EUR/USD-JPY/USD	38.75	46.34	46.37	49.63	48.98
GBP/USD-NZD/USD	8.46	24.15	24.83	24.99	23.75
GBP/USD-CAD/USD	11.62	24.77	27.13	25.60	25.75
GBP/USD-CHF/USD	16.47	33.36	32.34	34.13	35.57
GBP/USD-JPY/USD	35.69	38.69	38.64	40.71	39.07
NZD/USD-CAD/USD	20.96	31.70	32.37	32.97	33.71
NZD/USD-CHF/USD	20.06	40.21	40.24	41.57	41.77
NZD/USD-JPY/USD	40.40	52.45	52.54	53.58	54.43
CAD/USD-CHF/USD	14.71	28.53	32.01	30.96	31.48
CAD/USD-JPY/USD	28.37	40.96	40.92	43.62	44.26
CHF/USD-JPY/USD	31.90	46.27	46.30	48.31	48.80

Table D.6: Number of regimes for realized correlation

Note: The LOG-value reported is the negative quasi-log likelihood value. Bayesian Information Criterion (BIC) is improved (minimised) by pruning the tree.

	Max tree			Pruned tree		
	Regimes	LOG	BIC	Regimes	LOG	BIC
AUD/USD-EUR/USD	7	111.15	53.44	4	88.30	-19.05
AUD/USD-GBP/USD	7	309.79	-343.85	3	295.14	-393.34
AUD/USD-NZD/USD	5	-32.86	257.90	3	-57.08	229.46
AUD/USD-CAD/USD	6	190.77	-145.19	5	184.66	-172.36
AUD/USD-CHF/USD	9	61.95	230.61	5	25.80	145.36
AUD/USD-JPY/USD	8	-170.83	646.59	4	-210.25	572.96
EUR/USD-GBP/USD	3	-219.44	558.30	3	-219.44	558.30
EUR/USD-NZD/USD	5	35.80	120.58	2	-3.22	83.31
EUR/USD-CAD/USD	3	-189.68	499.43	3	-189.68	499.43
EUR/USD-CHF/USD	16	-728.97	2094.26	9	-805.59	1969.11
EUR/USD-JPY/USD	6	-475.79	1180.28	3	-501.36	1117.07
GBP/USD-NZD/USD	3	114.53	-113.75	2	103.63	-130.38
GBP/USD-CAD/USD	3	19.95	79.51	2	7.05	65.51
GBP/USD-CHF/USD	2	-378.99	837.51	2	-378.99	837.51
GBP/USD-JPY/USD	2	-345.67	767.57	2	-345.67	767.57
NZD/USD-CAD/USD	3	48.84	17.64	3	48.84	17.64
NZD/USD-CHF/USD	5	-30.42	253.02	3	-47.29	209.89
NZD/USD-JPY/USD	5	-317.98	826.54	3	-330.81	775.97
CAD/USD-CHF/USD	4	-213.38	585.85	3	-224.03	567.37
CAD/USD-JPY/USD	4	-290.27	733.00	4	-290.27	733.00
CHF/USD -JPY/USD	6	-353.65	936.00	3	-373.72	861.79

D.2 Full out-of-sample results

Table D.7: Out-of-sample results for realized volatility

Note: This table reports MAE and MSE values out-of-sample for realized volatility. Values are given in 10^{-2}.

		AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD /USD	MSE	0.589	0.479	0.451	0.491	0.464
	MAE	5.486	4.882	4.794	4.932	4.835
EUR /USD	MSE	0.385	0.284	0.279	0.296	0.289
	MAE	4.716	3.990	3.963	4.049	4.026
GBP /USD	MSE	0.334	0.244	0.239	0.253	0.246
	MAE	4.251	3.596	3.586	3.684	3.663
NZD /USD	MSE	0.680	0.576	0.555	0.591	0.568
	MAE	6.222	5.572	5.529	5.653	5.589
CAD /USD	MSE	0.396	0.295	0.290	0.300	0.293
	MAE	4.683	3.948	3.938	3.997	3.977
CHF /USD	MSE	0.682	0.556	0.544	0.574	0.557
	MAE	5.433	4.671	4.656	4.698	4.716
JPY /USD	MSE	0.662	0.562	0.532	0.576	0.541
	MAE	5.918	5.220	5.136	5.363	5.180

Table D.8: Out-of-sample MAE comparison table

Note: This table reports the MAE values out-of-sample for the models compared. The data used was the Fisher-transformed realized correlation. All values are given in 10^{-2}.

	AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-EUR/USD	22.305	18.747	18.720	19.031	18.884
AUD/USD-GBP/USD	20.691	17.680	17.674	17.697	17.907
AUD/USD-NZD/USD	23.529	20.164	20.165	20.102	20.253
AUD/USD-CAD/USD	21.875	18.210	18.102	18.302	18.261
AUD/USD-CHF/USD	22.237	19.071	19.095	19.587	19.714
AUD/USD-JPY/USD	23.780	21.576	21.522	21.926	21.934
EUR/USD-GBP/USD	22.695	20.782	20.857	20.824	20.820
EUR/USD-NZD/USD	20.852	18.800	18.814	18.798	18.873
EUR/USD-CAD/USD	22.154	19.886	19.826	19.944	19.792
EUR/USD-CHF/USD	30.664	26.367	26.380	26.662	26.745
EUR/USD-JPY/USD	26.458	23.790	23.733	23.989	23.936
GBP/USD-NZD/USD	19.579	17.949	17.935	18.254	18.116
GBP/USD-CAD/USD	20.396	18.560	18.508	18.636	18.534
GBP/USD-CHF/USD	24.406	21.596	21.645	21.698	21.626
GBP/USD-JPY/USD	23.887	21.867	21.805	21.887	21.897
NZD/USD-CAD/USD	20.535	18.440	18.359	18.889	18.544
NZD/USD-CHF/USD	21.172	18.854	18.867	19.160	19.114
NZD/USD-JPY/USD	24.248	21.687	21.655	21.943	21.821
CAD/USD-CHF/USD	22.163	19.988	20.000	20.080	20.485
CAD/USD-JPY/USD	23.485	21.099	21.114	21.352	21.482
CHF/USD-JPY/USD	25.310	22.363	22.355	22.650	22.671

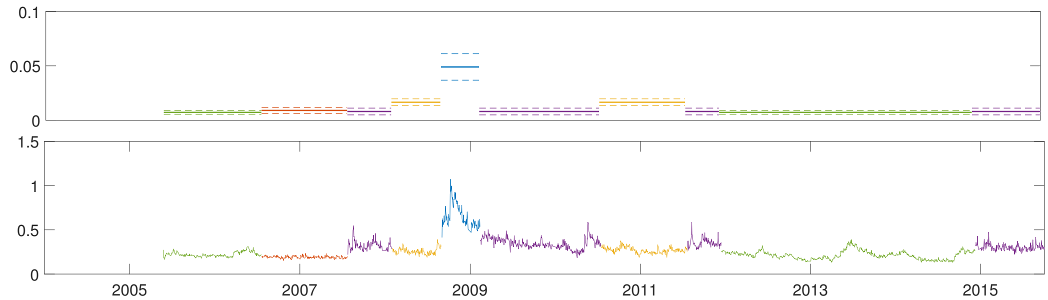
Table D.9: Out-of-sample MSE comparison table

Note: This table reports the MSE values out-of-sample for the models compared. The data used was the Fisher-transformed realized correlation. Values are given in 10^{-2}.

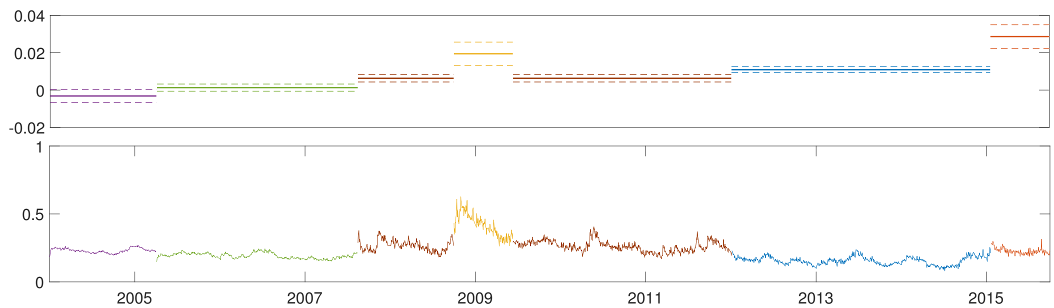
	AR(1)	HAR	HAR-SVI	Tree-HAR	Tree-HAR-SVI
AUD/USD-EUR/USD	8.162	6.167	6.146	6.381	6.306
AUD/USD-GBP/USD	7.982	6.096	6.068	6.145	6.222
AUD/USD-NZD/USD	9.552	7.297	7.286	7.417	7.479
AUD/USD-CAD/USD	7.807	5.643	5.583	5.697	5.636
AUD/USD-CHF/USD	8.183	6.385	6.396	6.740	6.777
AUD/USD-JPY/USD	9.560	7.962	7.919	8.291	8.242
EUR/USD-GBP/USD	8.355	7.037	7.060	7.062	7.039
EUR/USD-NZD/USD	7.133	5.908	5.907	5.954	5.963
EUR/USD-CAD/USD	8.100	6.819	6.778	6.887	6.779
EUR/USD-CHF/USD	15.873	12.735	12.662	12.877	12.947
EUR/USD-JPY/USD	11.494	9.610	9.570	9.741	9.756
GBP/USD-NZD/USD	6.271	5.379	5.383	5.583	5.503
GBP/USD-CAD/USD	6.809	5.831	5.804	5.892	5.836
GBP/USD-CHF/USD	9.616	7.638	7.613	7.744	7.632
GBP/USD-JPY/USD	9.137	7.865	7.832	7.869	7.861
NZD/USD-CAD/USD	6.931	5.742	5.675	5.962	5.812
NZD/USD-CHF/USD	7.445	6.051	6.057	6.266	6.177
NZD/USD-JPY/USD	9.592	7.863	7.790	8.091	7.963
CAD/USD-CHF/USD	8.138	6.863	6.856	6.939	7.194
CAD/USD-JPY/USD	9.392	7.967	7.986	8.127	8.229
CHF/USD-JPY/USD	10.705	8.645	8.640	8.765	8.802

Appendix E SVI coefficients

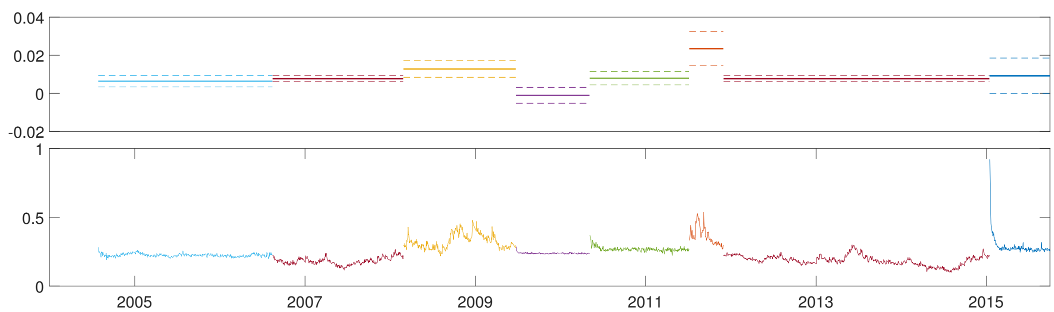
The following figures present the SVI-coefficient values from the regressions for the realized volatilities. The top panel reports the SVI-coefficients with a 95% confidence interval, whilst the bottom panel reports the regimes in the realized volatility time-series. Note that the SVI-coefficients are higher during volatile periods, and that the regimes are the same for the tree-HAR (reported in figure F.2) and tree-HAR-SVI regressions.



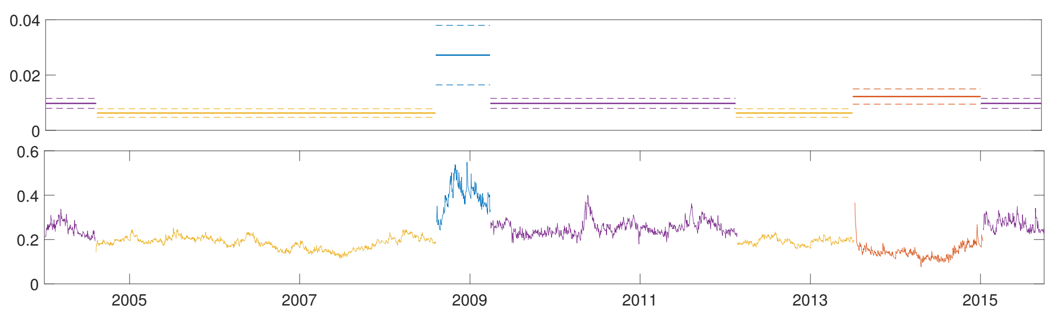
(a) AUD/USD



(b) CAD/USD

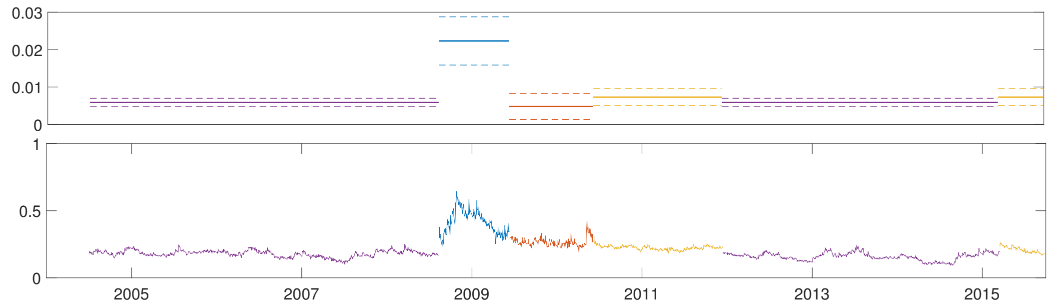


(c) CHF/USD

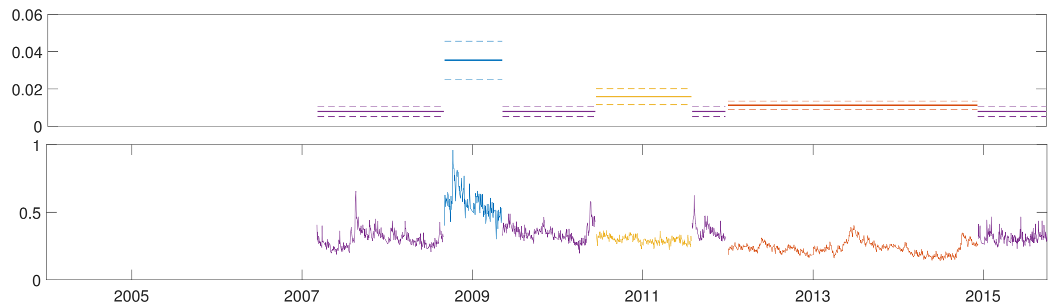


(d) EUR/USD

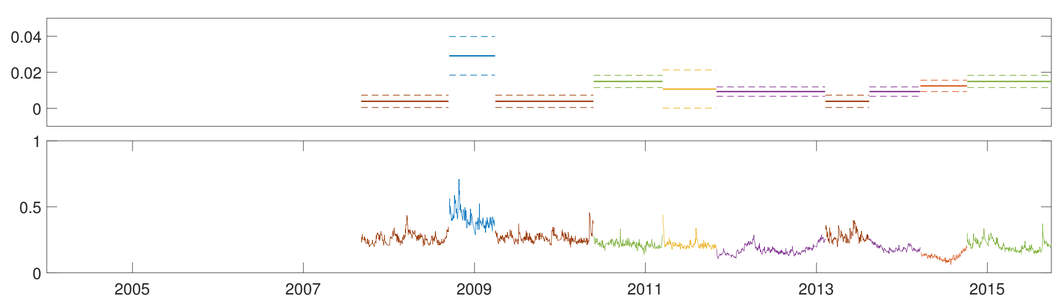
Figure E.1. SVI coefficients in different regimes for realized volatility.



(a) GBP/USD



(b) NZD/USD

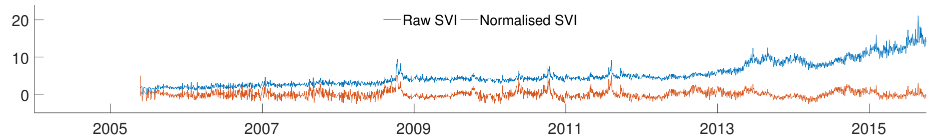


(c) JPY/USD

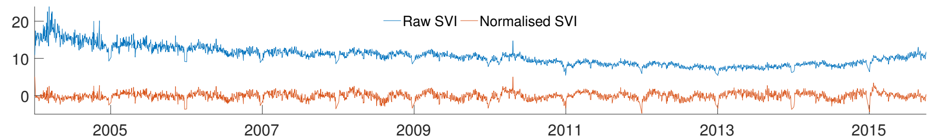
Figure E.2. SVI coefficients in different regimes for realized volatility.

Appendix F Time-series plots

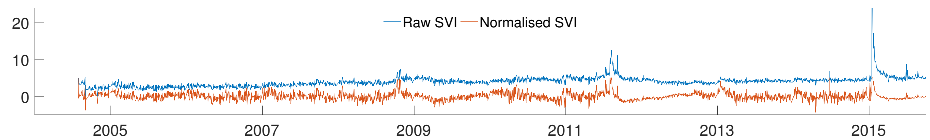
In this part of the appendix, we would like to present all the figures of the different time-series. The first figure presents the daily raw SVI data versus the normalised daily SVI data. The following figures after that present regimes found in the time-series of volatility and correlation. Each colour, within the respective time-series figures, illustrate the different regimes. We hope interesting patterns we have documented can be of inspiration and help for future studies.



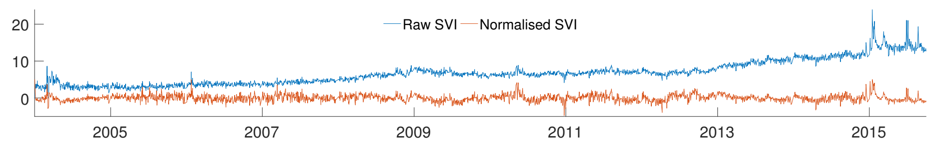
(a) AUD



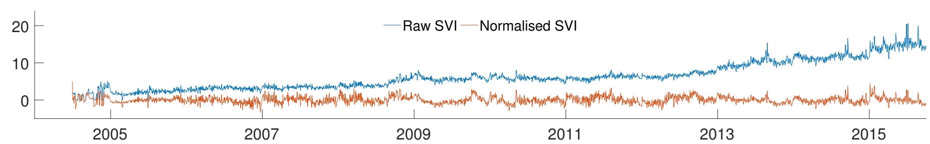
(b) CAD



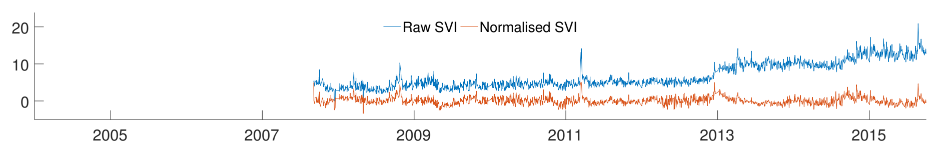
(c) CHF



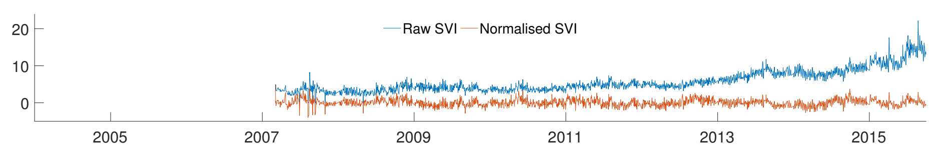
(d) EUR



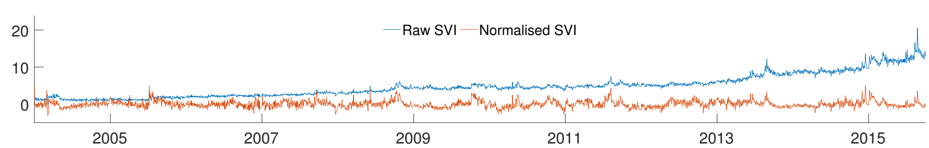
(e) GBP



(f) JPY

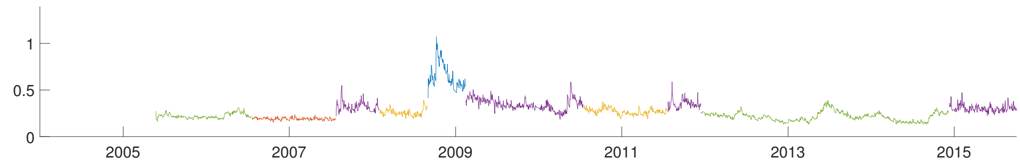


(g) NZD

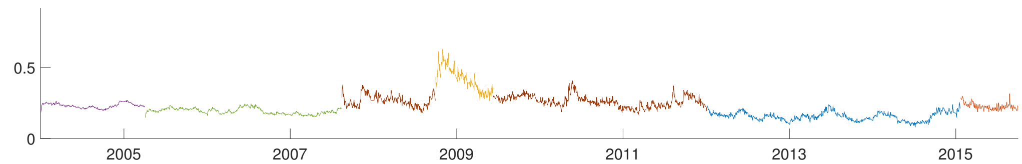


(h) USD

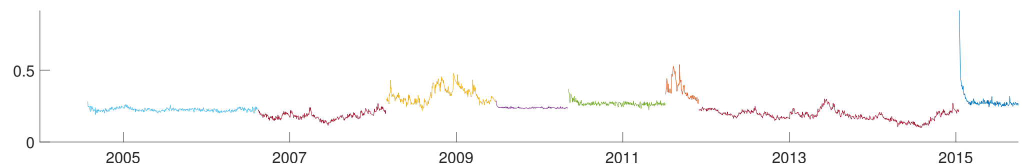
Figure F.1. Time-series of the cleand raw SVI and normalised SVI.



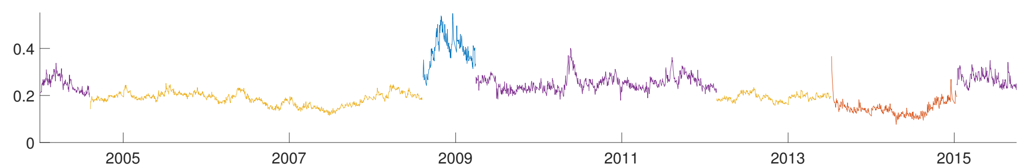
(a) AUD



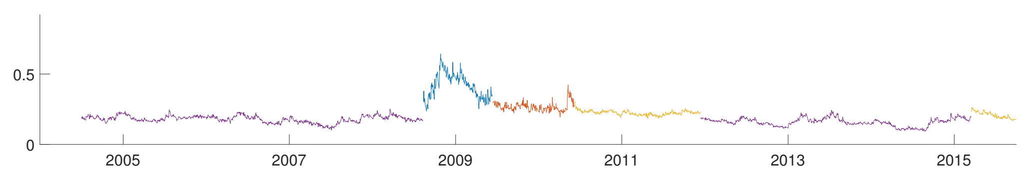
(b) CAD



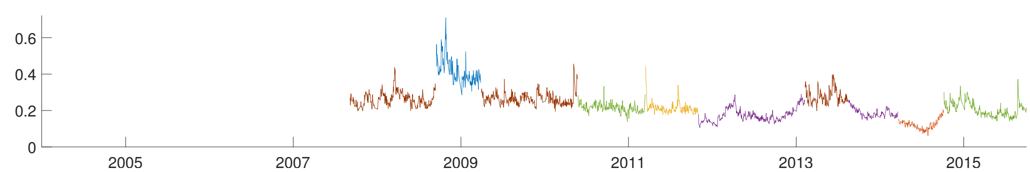
(c) CHF



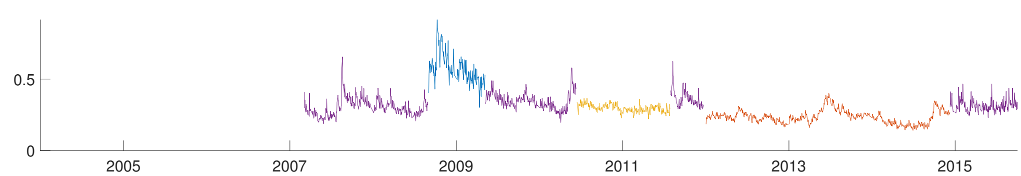
(d) EUR



(e) GBP

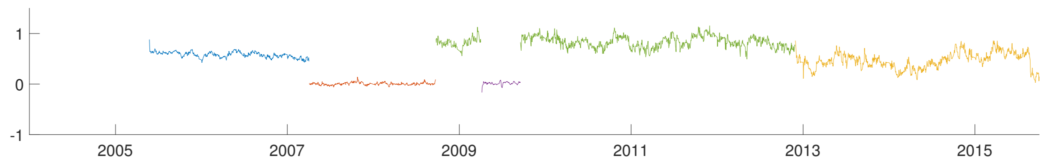


(f) JPY

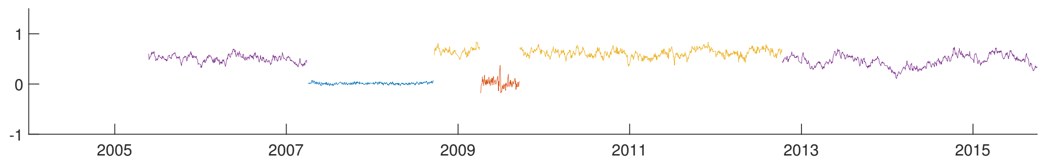


(g) NZD

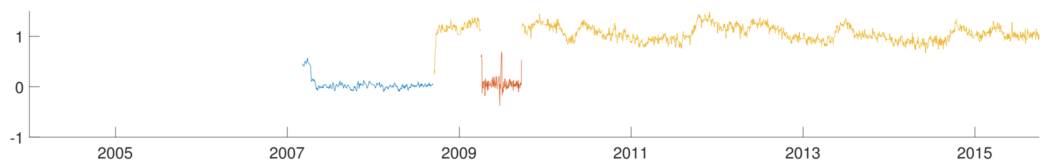
Figure F.2. Time-series of tree-HAR regimes for realized volatility.



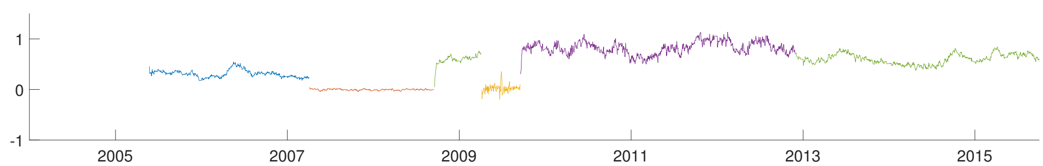
(a) AUD/USD and EUR/USD



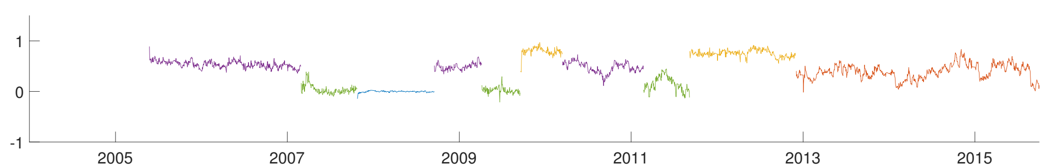
(b) AUD/USD and GBP/USD



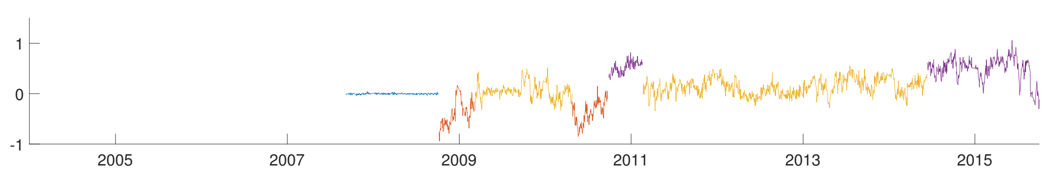
(c) AUD/USD and NZD/USD



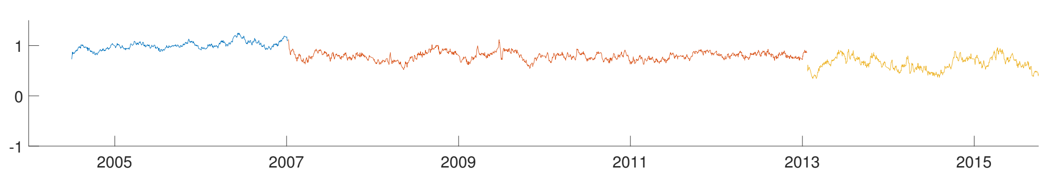
(d) AUD/USD and CAD/USD



(e) AUD/USD and CHF/USD

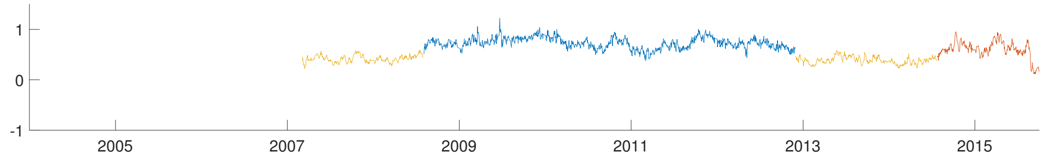


(f) AUD/USD and JPY/USD

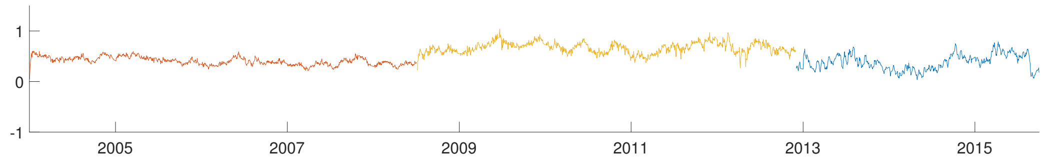


(g) EUR/USD and GBP/USD

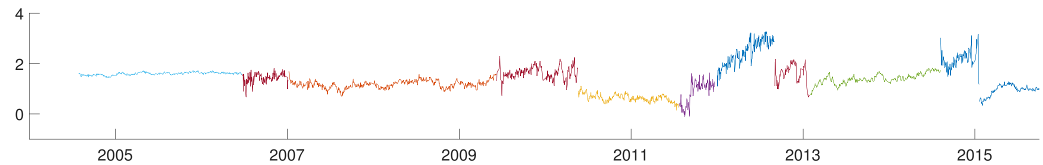
Figure F.3. correlation regime plots



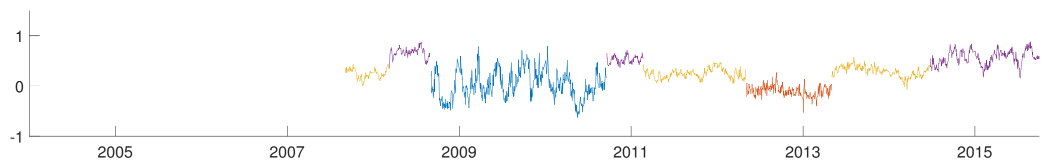
(a) EUR/USD and NZD/USD



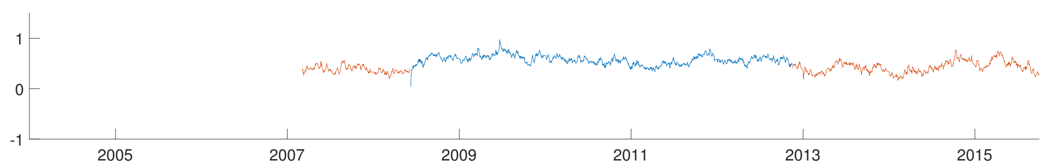
(b) EUR/USD and CAD/USD



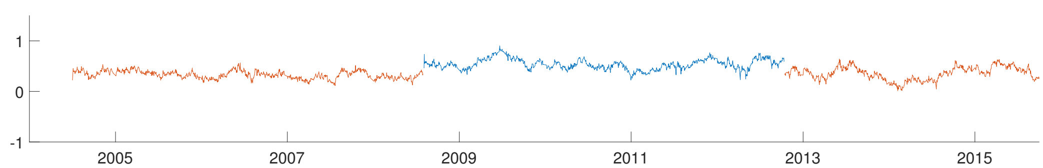
(c) EUR/USD and CHF/USD



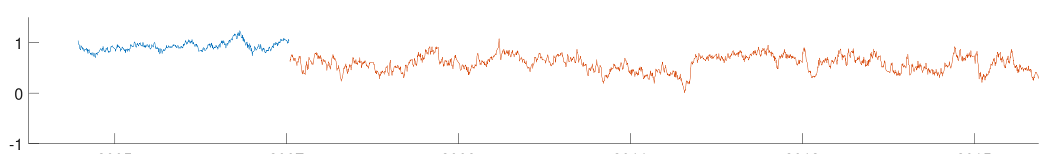
(d) EUR/USD and JPY/USD



(e) GBP/USD and NZD/USD

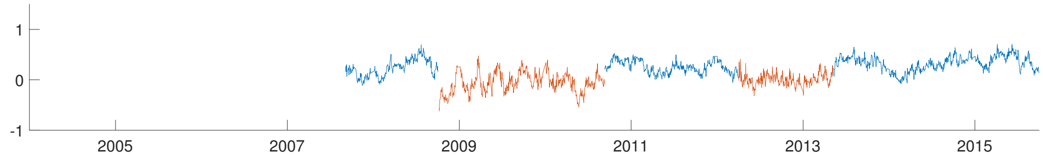


(f) GBP/USD and CAD/USD

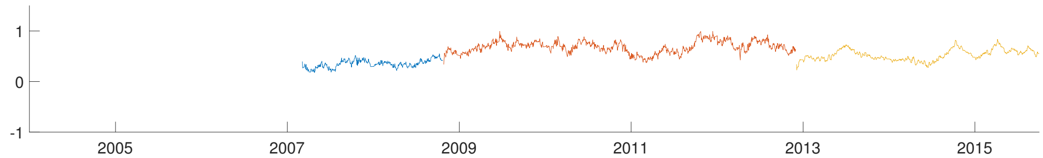


(g) GBP/USD and CHF/USD

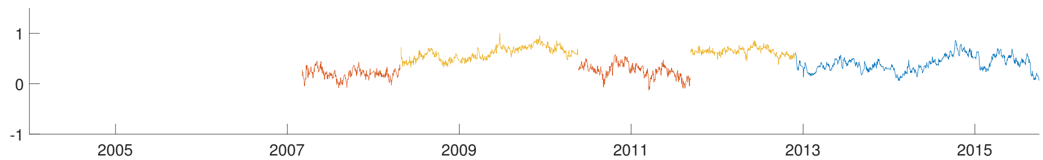
Figure F.4. correlation regime plots



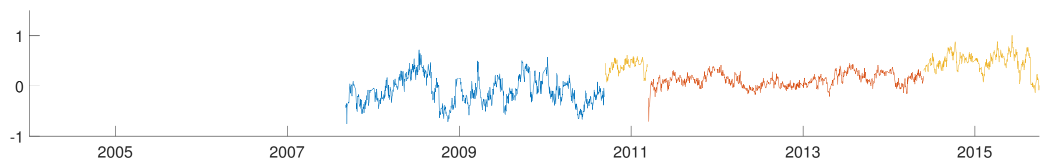
(a) GBP/USD and JPY/USD



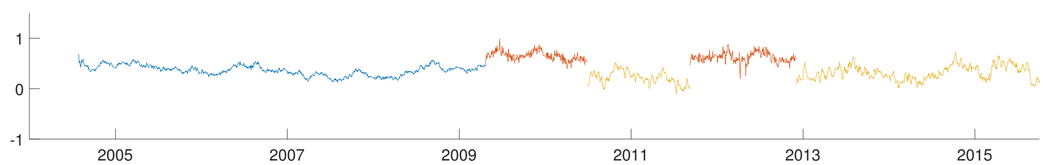
(b) NZD/USD and CAD/USD



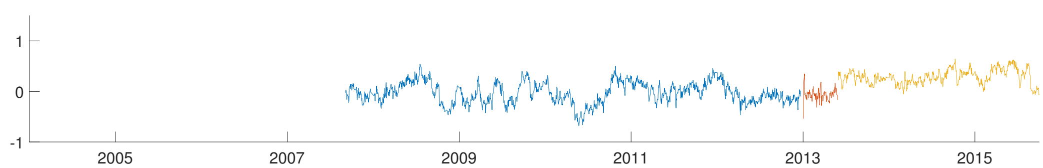
(c) NZD/USD and CHF/USD



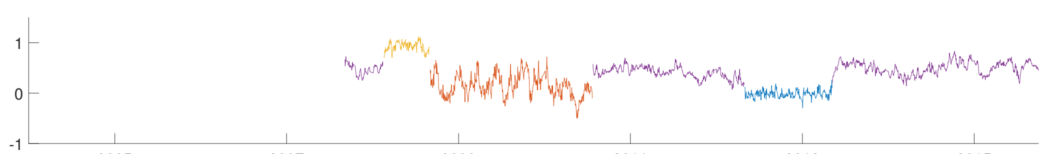
(d) NZD/USD and JPY/USD



(e) CAD/USD and CHF/USD



(f) CAD/USD and JPY/USD



(g) CHF/USD and JPY/USD

Figure F.5. correlation regime plots