INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

FORECASTING SOVEREIGN BOND REALIZED VOLATILITY USING TIME-VARYING COEFFICIENTS MODEL

Barbora Malinska

IES Working Paper 19/2021

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Malinska B. (2021): "Forecasting Sovereign Bond Realized Volatility Using Time-Varying Coefficients Model" IES Working Papers 19/2021. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Forecasting Sovereign Bond Realized Volatility Using Time-Varying Coefficients Model

Barbora Malinska[a]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic, Email: malinska.barbora@gmail.com

June 2021

Abstract:
This paper studies predictability of realized volatility of U.S. Treasury futures using high-frequency data for 2-year, 5-year, 10-year and 30-year tenors from 2006 to 2017. We extend heterogeneous autoregressive model by Corsi (2009) by higher-order realized moments and allow all model coefficients to be time-varying in order to explore dynamics in forecasting power of individual predictors across the term structure. We find realized kurtosis to be valuable predictor across the term structure with robust contribution also in out-of-sample analysis for the shorter tenors. Time-varying coefficient models are found to bring significant out-of-sample forecasting accuracy gain at the short end of the term structure. Further, we detect significant asymmetry in forecasting errors present for all the tenors as the constant-coefficient models were found to generate systemic under-predictions of future realized volatility.

JEL: C32, C53, G17
Keywords: Realized moments, Sovereign bonds, Volatility forecasting, High-frequency data, Time-varying coefficients

1. Introduction

Sovereign bonds are a traditional counterpart of stocks in investors' portfolios. Average daily trading volume of U.S. Treasury securities since 2009 has been double as compared to U.S. equities based on data published by Securities Industry and Financial Markets Association (SIFMA)[1] making the U.S. bond market one of the world's largest financial markets. Moreover, sovereign yields serve as a basis for pricing of other securities and derivatives as well as commercial loans which further strengthens the impact of the sovereign bond market on capital allocation in general. Forecasting bond return volatility is fundamental in terms of portfolio optimization, asset pricing and risk management but in contrast to equity markets empirical literature dealing with this topic is very scarce.

We believe that sovereign bonds represent very different investment proposition as compared to equities and even within the term structure there may be significant differences in context of volatility forecasting since short-term and long-term tenors are perceived as very different assets. The main contribution of this paper is threefold. First, we explore differences in predictability of volatility using multiple variations of HAR model across the entire term structure ranging from 2-year to 30-year maturity. Availability of high-frequency data allowed for significant improvement of volatility modeling and forecasting using methods relying on realized measures. In terms of forecasting, heterogeneous autoregressive model of realized volatility (HAR) by Corsi (2009) based on long memory feature of volatility has

[1] Data on average daily volumes available at https://www.sifma.org/resources/archive/research/statistics/

gained popularity thanks to its simplicity and sound forecasting performance. From the limited bond volatility literature, relevant studies include Andersen et al. (2007) or Busch et al. (2011) who, among other asset classes, inspected volatility of 30-year bonds. More compact view on bond market across various tenors was provided by Balduzzi et al. (2001) and on high-frequency level by Cieslak and Povala (2016) who performed a detailed analysis of realized volatility term structure. Recently, Özbekler et al. (2020) inspected HAR model including jumps and effect of monetary announcement on European sovereign bond market volatility.

Second, we examine informational content of additional explanatory variables and variation of their relevance across the term structure. In addition to variables which have demonstrated certain prediction power on bond markets in previous studies, we include higher-order realized moments, namely realized skewness and kurtosis. Recent surge of empirical literature inspecting value of higher-order moments on subsequent (mainly equity) returns emerged following the empirical work by Amaya et al. (2015) who found realized skewness to carry a significant power for predicting cross-section of equity returns. In context of volatility forecasting, higher-order realized moments are candidates for observable metrics of asymmetry or extremes resembling disaster events claimed to have substantial impact on price movements by number of theoretical studies (e.g. Kraus and Litzenberger (1976), Barro (2006)). Their value for volatility forecasting was reported by Mei et al. (2017) who found realized skewness to improve mid-term and long-term volatility forecasts of a Chinese equity index prices, Gkillas et al. (2019) who analyzed exchange-rate volatility or Demirer et al. (2019) who found realized skewness to be relevant for future gold price volatility.

Third, in addition to detecting differences in bond market volatility forecasting for various tenors we aim to explore whether (and how) the relationships evolve over time. We advance the complex analysis across entire term structure outlined above to fully dynamic environment using time-varying coefficients methodology (TVC) as recently introduced by Chen et al. (2018) who allowed all HAR model coefficients to be time-varying and found the TVC model to outperform the constant-coefficient benchmark especially in crisis period with unduly high volatility and in longer forecasting horizons. Other works examining the HAR model in dynamic environment on stock markets include Bollerslev et al. (2016) or Bekierman and Manner (2018) who advanced the constant-coefficient HAR model to a model time-varying in the autoregressive coefficient, Wu and Hou (2019) who replaced the autoregressive parameter by positive and negative semi-variance or Zhu et al. (2020) who included also a jump component. Our ambition is to demonstrate how the time-varying volatility forecasting models perform out-of-sample as compared to the static counterparts, how the potential accuracy gain varies for individual tenors and whether the models generate any systemic under- or over-predictions.

The remainder of this papers is organized as follows. Section 2 explains applied methodology, section 3 provides with detailed dataset description, section 4 summarizes in-sample estimation results comparing various model specifications both in static and dynamic forms, section 5 evaluates out-of-sample forecasting performance of all models and section 6 concludes.

2. Methodology

2.1. Realized moments

Thanks to availability of high-frequency data, we can observe stronger presence of model-free data-driven volatility measurements to the detriment of parametric models relying on strict assumptions in order to to capture the latency of volatility including GARCH (Bollerslev, 1986) or stochastic volatility (SV) models (Taylor, 1986), or alternatively, option-based implied volatility measures. As Andersen and Bollerslev (1998) or Andersen et al. (2003) show, realized volatility measures based on intra-day data bring significant reduction in noise and improve stability of the results as compared to the measures relying on daily return observations. We use realized variance measure as introduced by Andersen and Bollerslev (1998):

$$RV_t = \sum_{i=1}^{N} r_{t,i}^2 \tag{1}$$

where $r_{t,i}$ generally represents the $i-$th log-return on trading day t, N is the number of equispaced returns on the respective trading day. Intraday return is defined as $r_{t,i} = (p_{t,i} - p_{t,i-1}) \times 100$ where $p_{t,i}$ refers to logarithm of a intra-day price on day t. We note terms *(realized) variance* and *(realized) volatility* may be used interchangeably in the further text.

In additon to variance, higher moments embed information about occurrence of extreme events or asymmetry and can be associated with concepts such as tail risk (Bollerslev et al., 2015) or disaster risk (Kozhan et al., 2013). Thanks to availability of high-frequency data, we use realized skewness measure which was first introduced by Neuberger (2012). as constructed by

Amaya et al. (2015):

$$RSK_t = \frac{\sqrt{N}\sum_{i=1}^{N}(r_{t,i})^3}{RV_t^{3/2}} \tag{2}$$

where $r_{t,i}$ generally represents the $i-$th log-return on trading day t, N is the number of equispaced returns on the respective trading day and RV_t refers to realized variance as outlined above.

Similarly, in construction of realized kurtosis we follow Amaya et al. (2015):

$$RKU_t = \frac{N\sum_{i=1}^{N}(r_{t,i})^4}{RV_t^2} \tag{3}$$

where $r_{t,i}$ generally represents the $i-$th log-return on trading day t, N is the number of equispaced returns on the respective trading day and RV_t refers to realized variance as outlined above.

2.2. Extended HAR model

The basic building block of the analysis is the well-known HAR model by Corsi (2009). We define the initial model which serves also as a benchmark model to our further extensions introduced below as follows:

$$\log RV_t = \beta_0 + \beta_1 \log RV_{t-1} + \beta_2 \log RV_{w,t-1} + \beta_3 \log RV_{m,t-1} + \epsilon_t \tag{4}$$

where $\log RV_{w,t}$ represents $\log RV_{t:t-5}$ and $\log RV_{m,t}$ represents $\log RV_{t:t-21}$ defined as $\log RV_{t:t-j} = \frac{1}{j}\sum_{k=0}^{j} \log RV_{t-k}$ and RV refers to realized variance as defined above. The autoregressive elements is also referred to as $\log RV_d$.

We extend the basic model by two groups of additional explanatory variables. Similarly to Mei et al. (2017) or Gkillas et al. (2019) following up on the recent surge of literature focusing on realized moments, we include first lags of realized skewness (differenced) and kurtosis and construct HAR-M (moments) model:

$$\begin{aligned} \log RV_t = \beta_0 + \beta_1 \log RV_{t-1} + \beta_2 \log RV_{w,t-1} + \\ +\beta_3 \log RV_{m,t-1} + \gamma_1 \Delta RSK_{t-1} + \gamma_2 RKU_{t-1} + \epsilon_t \end{aligned} \tag{5}$$

Second, we include additional control variables which shall from its nature be useful in explaining future volatility on bond markets such as Federal Funds rate as proxy for a risk-free rate, VIX volatility index or economic policy uncertainty index. For now, we label vector of these variables as C_{t-1} and provide with detailed description in Section 3. HAR-C (controls) model is expressed as:

$$\begin{aligned} \log RV_t = \beta_0 + \beta_1 \log RV_{d,t-1} + \beta_2 \log RV_{w,t-1} + \\ +\beta_3 \log RV_{m,t-1} + C_{t-1}\delta + \epsilon_t \end{aligned} \tag{6}$$

Final extension is called HAR-CM model and represents a combination of HAR-C and HAR-M models in order to reveal additional explanatory power of realized moments which has not been captured by the control variables:

$$\begin{aligned} \log RV_t = \beta_0 + \beta_1 \log RV_{d,t-1} + \beta_2 \log RV_{w,t-1} + \\ +\beta_3 \log RV_{m,t-1} + \gamma_1 \Delta RSK_{t-1} + \gamma_2 RKU_{t-1} + C_{t-1}\delta + \epsilon_t + \epsilon_t \end{aligned} \tag{7}$$

2.3. Time-varying coefficients model

One of the ambitions of this paper is to explore whether assumption of constant coefficients in the models described above is appropriate or whether allowing dynamics of these coefficients can improve in-sample and out-of-sample performance as compared to the benchmark static HAR model.

Having an extensive data set covering 12 years in hand, we have a solid base for exploration of the dynamics between the variables. Since fully non-parametric models which do not require any assumptions on the relationships between the variables do not often provide with a straightforward inference, semi-parametric models where the coefficients change over time in a specified manner allow for more flexibility than parametric linear model without the drawbacks of the fully non-parametric methods.

Before turning to concrete time-varying model specifications, we would like to highlight main features of TVC models generally expressed as:

$$y_t = x_t^\intercal \beta(z_t) + u_t, t = 1, ..., T \tag{8}$$

where y_t is the dependent variable, x_t is the vector of independent variables, and u_t is the disturbance term with $E(u_t|x_t) = 0$ and $E(u_t^2|x_t) = \sigma^2$. As compared to the constant-coefficient linear model, $\beta(z_t)$ represent the regression coefficients being the unknown functions of time (linear constant method) or of a random variable changing with time (local linear method). Pioneering work of the former was Robinson (1989) and Hastie and Tibshirani (1993) for the latter who analysed the time-varying parameter linear models using stationary variables[2]. Recently, the fully time-varying coefficient methodology was applied by Chen et al. (2018) who forecasted volatility

[2]We refer to Casas and Fernandez-Casal (2019) for a comprehensive literature review on time-varying coefficients methodology.

of S&P500 returns.

Assuming $\beta(\cdot)$ to be twice differentiable, Taylor-rule-approximation around z is expressed by $\beta(z_t) = \beta(z) + \beta(z)^{(1)}(z_t - z)$. Combining OLS and local linear kernel method (Fan and Gijbels, 1996) solves the following minimization problem:

$$(\hat{\beta}(z_t), \hat{\beta}^{(1)}(z_t)) = \arg\min_{\theta_0,\theta_1} \sum_{t=1}^{T} [y_t - x_t^\intercal \theta_0 - (z_t - z) x_t^\intercal \theta_1]^2 K_b(z_t - z) \qquad (9)$$

where the kernel $K_b(z_t - z) = \frac{1}{b} K(\frac{z_t - z}{b})$ weights the local regressions within a chosen bandwith b.We refer to Chen et al. (2018) for detailed explanation of the theoretical framework of the local linear TVC method as also applied in our paper.

We use the above presented TVC with local linear kernel methodology for estimation of models presented in Section 2.2 allowing all coefficients to be time-varying. The respective models are labeled as TVHAR, TVHAR-M, TVHAR-C and TVHAR-CM, respectively.

3. Data

For construction of realized measures, we use 1-minute U.S. Treasury futures data (active contracts) from Tick Data database for each tenor traded at CME Globex platform under tickers TU (2-year tenor), FV (5-year tenor), TY (10-year tenor) and US (30-year tenor).

Apart from crucial benefit for a data-driven methodology of having clean and reliable high-frequency data from a renowned database, there are other advantages of analyzing front contract futures data instead of cash market. First, U.S. Treasury futures market, with futures prices are tightly linked to underlying bond prices (and yields), has been gaining relative importance to the detriment of the cash market.[3] Moreover, also due to lower transaction costs, futures market was detected to be dominant to cash market in reaction to news and price discovery process (Andersen et al., 2007).

Our sample period covers futures price data from January 3, 2006 to December 5, 2017. We exclude weekends, public holidays and other days with no trading activity (i.e. days with a single unique price) resulting in 3,066 days included in the sample. On each day we take into account observations between 5:20 a.m. to 4:00 p.m. CT including also usual times of publication of annoucements of Federal Reserve System closely followed by fixed income investors.[4]

In line with consensus in the literature, we sample the datapoints according to 5-minute interval to achieve the optimal trade-off between sufficiently high frequency and lowest possible bias due to microstructure noise (Hansen and Lunde, 2006). Final dataset containing 395,514 price observations for each tenor is used to construct measures of realized volatility, realized skewness and realized kurtosis as defined in Section 2.1.

Figure 1 shows mean monthly volatility (corresponding the the $RV_{t:t-21}$ components of the HAR model) for our four tenors included in our analysis. Unsurprisingly, we find the period of global financial crisis to be the most

[3]See "The New Treasury Market Paradigm", CME Group, June 2016, available at https://www.cmegroup.com/education/files/new-treasury-market-paradigm.pdf.

[4]Thanks to electronic platform CME Globex, trading activity occurs also outside CME trading hours (07:20 - 14:00 CT). Therefore, we extend the trading day by additional 4 hours to capture significant trading activity.

turbulent period in our sample. Importantly, development of volatility shows similar patters across all maturities. However, we have a clear evidence of volatility term structure to be upward-sloping. As discussed in Malinska (2020), in addition to classical duration-related topics, increased volatility for longer tenors could be explained by different drivers of investor sentiment on short and long ends of the term structure. Since short-maturity securities tend to follow monetary policy news (which shall be relatively stable or predictable), longer maturities embed rather general sentiment of economic development. This is also our motivation to include variables capturing these dynamics and mood on the market into our volatility forecasting models.

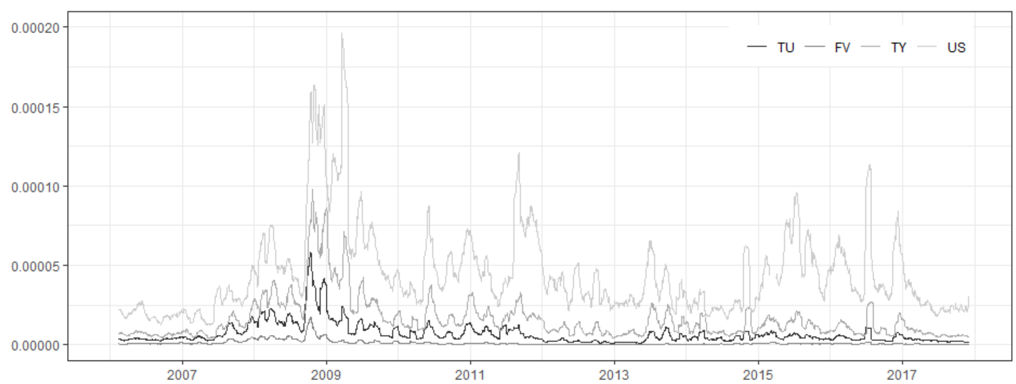


Figure 1: Mean monthly realized variance. Note: Monthly realized variance calculated as average from preceding 22 days.

Figure 2 contains plots of realized variance, skewness and kurtosis for the 10-year tenor as an example. Over the sample period, mean values of realized variance, skewness and kurtosis were 1.6e-05, -0.057 and 15.77, respectively whereas median values were 9.4e-06, -0.018 and 8.29, respectively. We remind that we use realized variance in logarithm and realized skewness in first-difference in our models. Detailed descriptive statistics of variables as applied in further estimation for all tenors is summarized in Table 1.

(a) Realized variance (b) Realized skewness (c) Realized kurtosis

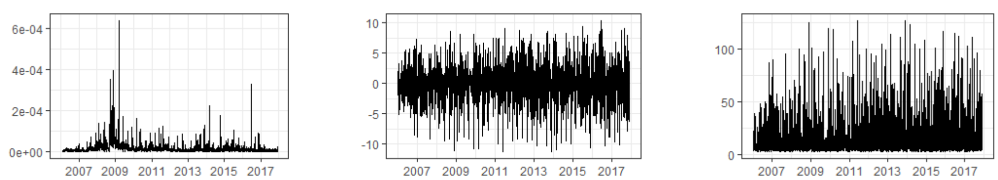


Figure 2: Realized moments of 10-year futures returns (TY)

We include several control variables capturing mood and expectations on the market which we believe capture information likely to be taken by bond investors into account or which were found to be valauble in volatility forecasting in existing literature. As a basis, we include a 3-month constant maturity Treasury bill rate[5] widely perceived as a risk-free rate of the U.S. money market embedding also expectations on monetary policy change. Risk-free rate is applied in first-difference in our models. Further, as compared to Chao (2016) who include 15 economic variables when forecasting bond volatility in monthly frequency, we strive to find few index variables

[5]Board of Governors of the Federal Reserve System (US), 3-Month Treasury Constant Maturity Rate [DGS3MO], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DGS3MO, April 11, 2021.

providing with a complex description of market situation, investor sentiment and future expectations of those in order to limit dimensionality of the model having a rather computationally-demanding procedure of TVC estimation ahead.

(a) Risk-free rate (b) EPU (c) VIX

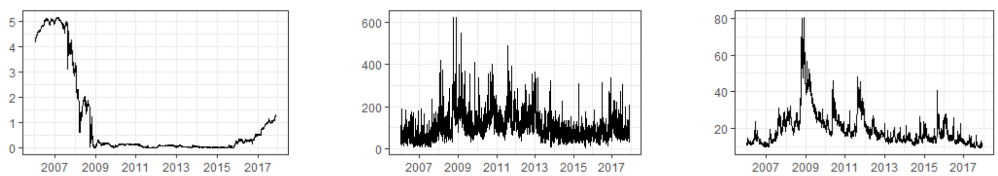


Figure 3: Bond market control variables

We also include Economic Policy Uncertainty index (EPU) for U.S. market introduced by Baker et al. (2016) which was used in multiple works dedicated to volatility forecasting (e.g. Gkillas et al. (2019)). The index is constructed based on newspaper coverage related to economic policy uncertainty, expiration schedule of U.S. tax provisions and discrepancies among economic forecasts.[6]

Since investors often optimize their portfolios balancing exposure to stocks and bonds, we believe that an indicator giving information on expectations on future stock market volatility might have a valuable content also for sovereign bond futures volatility forecasting. Therefore, we include differenced CBOE

[6]More details on index construction as well as actual data series available at https://www.policyuncertainty.com/.

Volatility Index (VIX) derived from near-term options prices of S&P 500 index which is closely followed by market participants as a key indicator of "fear" on the stock market globally.

Table 1 sets forth a statistical summary of the variables described above in functional forms as applied in the actual models. Importantly, test statistics of the augmented Dickey-Fuller test indicate rejection of null hypothesis of unit root presence in all variables. Further we report Q statistics of Ljung-Box test of serial correlation of up to 22 lags indicating persistence in all the series.

Table 2 sets forth a correlation analysis of all explanatory variables (weekly and monthly means of $\log RV$ are omitted). Correlation of logarithm of realized variance with differenced realized skewness is very close to zero across all maturities. Realized skewness shows positive slightly positive correlation with $\log RV$ for the shortest maturity, while the remaining tenors report low but negative correlation. Correlation of variance with changed in risk-free rate is negative and relatively stable for all tenors. Correlation of economic policy uncertainty index and variance is decent and significantly stronger for the long-term maturities (being 10-year and 30-year tenors) as compared to short- and mid-term maturities. On the contrary, changes of VIX index do not seem much correlated with log realized variance for any maturity.

Further, we demonstrate that looking even at basic components of the HAR model, there is a reasonable suspicion that relative importance of daily, weekly or monthly means for next day's realized variance probably vary over time. Figure 4 provides with plots of 250-day rolling correlations between

log RV	TU	FV	TY	US
Mean	-14.308	-12.444	-11.487	-10.327
Median	-14.539	-12.532	-11.586	-10.373
Std. dev	0.933	0.959	0.849	0.743
Min.	-16.914	-15.554	-14.155	-13.035
Max.	-9.769	-7.857	-7.354	-4.705
ADF test	-3.997***	-4.944***	-5.168***	-5.243***
Ljung-Box (22)	2.804e04***	1.785e04***	1.454e04***	1.197e04***

ΔRSK	TU	FV	TY	US
Mean	-0.002	-0.002	-0.002	-0.001
Median	-0.037	-0.036	0.006	0.067
Std. dev	3.040	3.942	3.804	3.361
Min.	-14.122	-15.974	-15.874	-14.047
Max.	14.244	16.623	17.821	16.933
ADF test	-24.308***	-24.699***	-25.069***	-25.494***
Ljung-Box (22)	7.894e02***	7.835e02***	7.964e02***	7.964e02***

RKU	TU	FV	TY	US
Mean	11.421	16.624	15.772	13.515
Median	5.206	8.645	8.289	7.114
Std. dev	16.315	19.364	18.480	16.177
Min.	1.985	2.436	2.407	2.390
Max.	127.121	128.041	127.346	127.836
ADF test	-10.511***	-13.443***	-13.714***	-12.863***
Ljung-Box (22)	1.580e02***	6.768e01***	5.855e01***	1.001e02***

Controls	ΔRF	EPU	ΔVIX	
Mean	-0.001	104.190	0.000	
Median	0.000	87.420	-0.055	
Std. dev	0.052	67.449	1.816	
Min.	-0.810	3.320	-17.360	
Max.	0.760	626.030	16.540	
ADF test	-15.057***	-5.507***	-15.800***	
Ljung-Box (22)	4.130e02***	1.633e04***	1.407e02***	

*** indicates p-value below 1%.

Table 1: Descriptive statistics

		$\log RV$	ΔRSK	RKU	ΔRF	EPU	ΔVIX
TU	$\log RV$	1.00	0.01	0.12	-0.11	0.08	0.05
	ΔRSK	0.01	1.00	-0.12	-0.08	-0.01	0.15
	RKU	0.12	-0.12	1.00	-0.01	-0.02	0.00
	ΔRF	-0.11	-0.08	-0.01	1.00	-0.03	-0.13
	EPU	0.08	-0.01	-0.02	-0.03	1.00	-0.03
	ΔVIX	0.05	0.15	0.00	-0.13	-0.03	1.00
FV	$\log RV$	1.00	0.00	-0.04	-0.10	0.15	0.05
	ΔRSK	0.00	1.00	-0.06	-0.07	-0.01	0.18
	RKU	-0.04	-0.06	1.00	0.01	0.02	0.01
TY	$\log RV$	1.00	0.01	-0.04	-0.10	0.39	0.05
	ΔRSK	0.01	1.00	-0.08	-0.07	0.00	0.16
	RKU	-0.04	-0.08	1.00	0.00	-0.01	0.01
US	$\log RV$	1.00	0.01	-0.02	-0.08	0.29	0.05
	ΔRSK	0.01	1.00	-0.07	-0.06	-0.01	0.17
	RKU	-0.02	-0.07	1.00	0.00	0.03	0.01

Table 2: Correlation matrix of explanatory variables (mutual correlations of control variables displayed in the TU section only)

$\log RV$ and $\log RV_{t-1}$, $\log RV_{t-1:t-6}$ and $\log RV_{t-1:t-22}$ for the 10-year tenor.

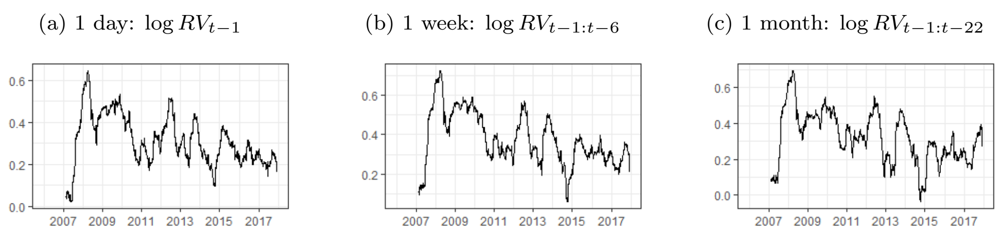


Figure 4: 250-day rolling correlation of $logRV$ with preceding day's, week's and month's variance (in logarithm) for 10-year tenor

We see that correlation of today's and yesterday's variance peaked during financial crisis (0.6) and since then there is apparent a decreasing trend with 2017 values around 0.2 with local peaks in the periods of U.S. debt crises or other turbulent periods (such as Fed's announcement to reduce pace of quantitative easing in 2013) with similar trend and pattern also in case of weekly and monthly means. Therefore, we believe it is worth to explore time-variation of the coefficients within HAR model and its extensions since a constant relationship may not be a realistic assumption.

4. Results

4.1. In-sample estimation results

First, we summarize key findings of constant-coefficient HAR model and its extensions across the term structure (variables were standardized prior estimation). As expected, in Table 3 we see significant coefficients of all HAR regressors but the effect of a change of individual elements by one standard deviation of the HAR model increases (or declines) across the term structure. The same applies for goodness-of-fit where we see clear downward trend with increasing maturity.

	TU	FV	TY	US
(Intercept)	0.002	0.002	0.005	0.006
	(0.011)	(0.013)	(0.013)	(0.014)
$\log RV_d$	0.092 **	0.113 ***	0.133 ***	0.144 ***
	(0.028)	(0.027)	(0.026)	(0.027)
$\log RV_w$	0.347 ***	0.305 ***	0.277 ***	0.248 ***
	(0.067)	(0.042)	(0.037)	(0.034)
$\log RV_m$	0.383 ***	0.330 ***	0.309 ***	0.299 ***
	(0.058)	(0.036)	(0.034)	(0.034)
R^2	63.5%	50.5%	45.6%	41.3%

The number in paranthesis is heteroskedasticity and autocorrelation robust Newey-West standard error. ***, ** and * denotes siginficance at 0.1%, 1% and 5%, respectively.

Table 3: Estimation results of constant-coefficient HAR model

Table 4 sets forth estimation result of HAR-M model, i.e. HAR model extended for realized moments. Here we observe similar outcome for all tenors since realized skewness is not significant in contrast with realized kurtosis having negative and significant effect across all maturities with coeffcient showing an inverted U-shape. It is worth highlighting that inclusion (and significance) of realized kurtosis had impact on coefficients of lagged log RV as relative importance of HAR components turn to be more balanced[7]. Most significant changes in $\hat{\beta}_1$ are visible for the 5-year and 10-year tenors for which also R^2 differential as compared to HAR model is the highest. It seems that capturing of intra-day fluctuations by second and fourth moments is of value since due to clustering behaviour extremes occurring on one day are likely to be repeated tomorrow. As a robustness check of information carried by realized kurtosis, we have included also a jump component. We examined presence of jumps using approach by Barndorff-Nielsen and Shephard (2006). The variable refers to $BNSJ_t = (RV_t - BP_t) \times I_t$ where BP is bi-power variation and I is equal to 1 if null hypothesis of no jumps was rejected at 1% significance level and zero otherwise. We found the jump-related variable to be significant only in absence of realized kurtosis in the model.

Further, we include multiple control variables and present estimation results of HAR-CM model in Table 5. As expected, controls having a significant impact across the entire term structure are risk-free rate and VIX index. The effect of the former is similar for all maturities, whereas of the latter the ef-

[7] We estimated the HAR-M model also including weekly and monthly means of realized kurtosis with negligible impact on coefficients of log RV_w or log RV_m. Results are available upon request.

	TU	FV	TY	US
(Intercept)	0.002	0.003	0.005	0.007
	(0.011)	(0.013)	(0.013)	(0.014)
$\log RV_d$	0.285 ***	0.277 ***	0.282 ***	0.243 ***
	(0.039)	(0.033)	(0.030)	(0.029)
$\log RV_w$	0.267 ***	0.237 ***	0.215 ***	0.208 ***
	(0.056)	(0.042)	(0.037)	(0.034)
$\log RV_m$	0.330 ***	0.274 ***	0.261 ***	0.272 ***
	(0.054)	(0.036)	(0.033)	(0.031)
ΔRSK	0.012	0.002	0.007	0.001
	(0.010)	(0.013)	(0.013)	(0.014)
RKU	-0.149 ***	-0.167 ***	-0.166 ***	-0.131 ***
	(0.018)	(0.018)	(0.019)	(0.017)
R^2	64.6%	52.2%	47.4%	42.5%

The number in paranthesis is heteroskedasticity and autocorrelation robust Newey-West standard error. ***, ** and * denotes siginficance at 0.1%, 1% and 5%, respectively.

Table 4: Estimation results of constant-coefficient HAR-M model

fect strengthens with time to maturity. The remaining control capturing economic policy uncertainty turns out to be significant for longest maturities with the largest coefficient in case of the 30-year tenor where we see also the largest contribution of inclusion of controls to the goodness-of-fit whereas for other maturities the improvement is much more limited as compared to the HAR-M model.

Next, we estimate the models using time-varying coefficients methodology in order to inspect dynamics of the relationships. Plots of the individual HAR-CM model coefficients for all tenors are presented below together with 90% confidence intervals and constant-coefficient OLS estimates.

On Figure 5 we see that for all maturities there are relatively long periods

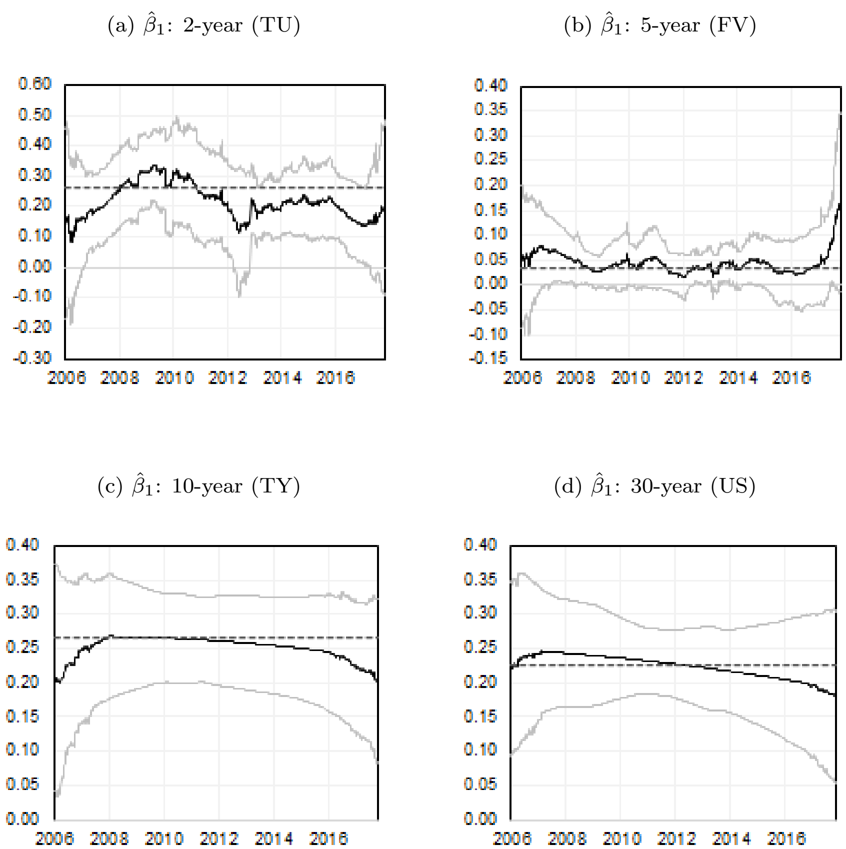
	TU	FV	TY	US
(Intercept)	0.002	0.003	0.005	0.006
	(0.011)	(0.013)	(0.013)	(0.014)
log RV_d	0.266 ***	0.262 ***	0.266 ***	0.226 ***
	(0.029)	(0.026)	(0.025)	(0.024)
log RV_w	0.273 ***	0.245 ***	0.221 ***	0.212 ***
	(0.041)	(0.037)	(0.036)	(0.035)
log RV_m	0.337 ***	0.274 ***	0.254 ***	0.260 ***
	(0.035)	(0.033)	(0.031)	(0.030)
ΔRSK	0.003	-0.008	-0.003	-0.011
	(0.011)	(0.013)	(0.013)	(0.014)
RKU	-0.143 ***	-0.162 ***	-0.162 ***	-0.128 ***
	(0.015)	(0.016)	(0.016)	(0.016)
ΔRF	-0.042 ***	-0.039 **	-0.045 ***	-0.041 **
	(0.011)	(0.013)	(0.013)	(0.014)
EPU	-0.010	0.008	0.033 *	0.052 ***
	(0.011)	(0.013)	(0.014)	(0.015)
ΔVIX	0.034 **	0.042 **	0.045 ***	0.057 ***
	(0.011)	(0.013)	(0.013)	(0.014)
R^2	64.9%	52.5%	47.9%	43.2%

The number in paranthesis is heteroskedasticity and autocorrelation robust Newey-West standard error. ***, ** and * denotes siginficance at 0.1%, 1% and 5%, respectively.

Table 5: Estimation results of constant-coefficient HAR-CM model

(exceeding 1 year) where the OLS coefficient is not covered within the confidence interval of the time-varying estimates. In these periods, TVC approach estimates $\hat{\beta}_1$ to decrease be below the OLS estimate towards zero.

Coefficient estimates on the weekly variance component plotted on Figure 6 shows OLS estimates to be included in the TVC confidence bands throughout the period, however we find the confidence intervals to cover zero for the last 2-3 years of the period. On the other hand, since 2014 the OLS $\hat{\beta}_3$ turns out to be well out of the confidence bands.



(a) $\hat{\beta}_1$: 2-year (TU)

(b) $\hat{\beta}_1$: 5-year (FV)

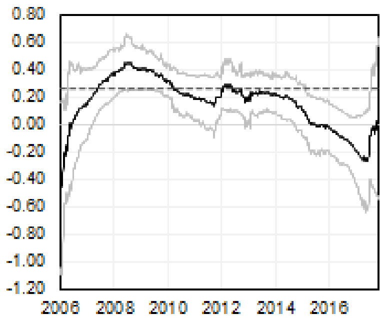
(c) $\hat{\beta}_1$: 10-year (TY)

(d) $\hat{\beta}_1$: 30-year (US)

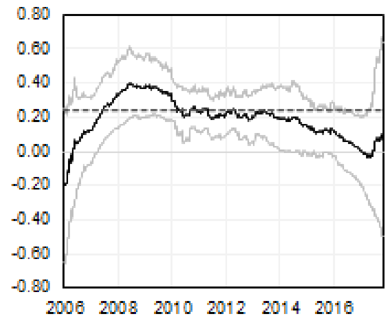
Figure 5: Time-varying estimates of the coefficient functions $\beta_1(\cdot)$ (black solid line) of TVHAR-CM model ($\log RV_d$ variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

In case of $\log RV_m$ presented on Figure 7, we see OLS estimate to overstate the β_3 in time of global financial crisis of 2007-2009 for the 2-year maturity asset. Other maturities report U-shaped development of the coefficient with OLS estimate included within the 90% confidence intervals.

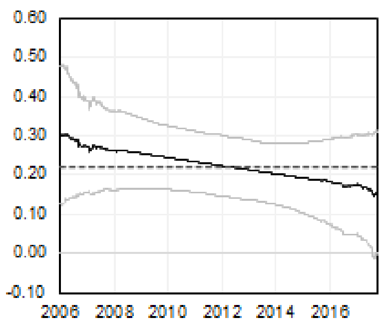
(a) $\hat{\beta}_2$: 2-year (TU)



(b) $\hat{\beta}_2$: 5-year (FV)



(c) $\hat{\beta}_2$: 10-year (TY)



(d) $\hat{\beta}_2$: 30-year (US)

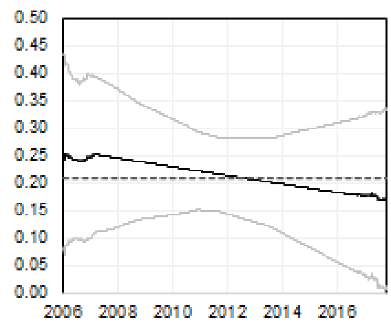
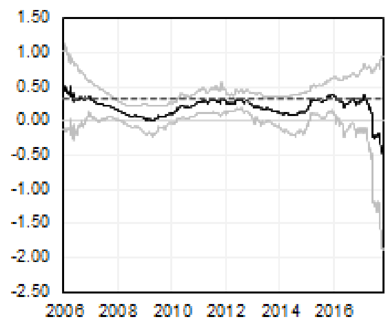


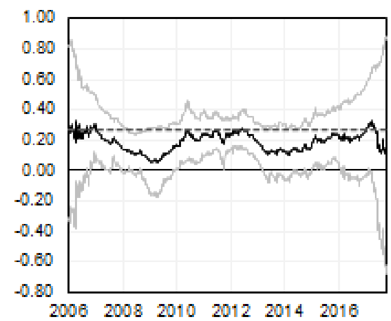
Figure 6: Time-varying estimates of the coefficient functions $\beta_2(\cdot)$ (black solid line) of TVHAR-CM model ($\log RV_w$ variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

Standard OLS estimation results reported no significance of realized skewness (differenced) on next days volatility for all inspected tenors. This is confirmed by TVC analysis with coeffient functions summarized in Figure 8 except for the shortest 2-year tenor where the 90% confidence band shifts well above zero starting end of 2015. On the other hand, for the longer-term tenors, there is evidence of limited (but significant) negative effect around

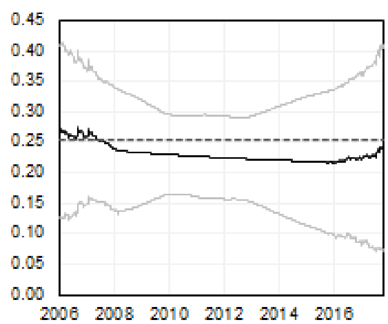
(a) $\hat{\beta}_3$: 2-year (TU)



(b) $\hat{\beta}_3$: 5-year (FV)



(c) $\hat{\beta}_3$: 10-year (TY)



(d) $\hat{\beta}_3$: 30-year (US)

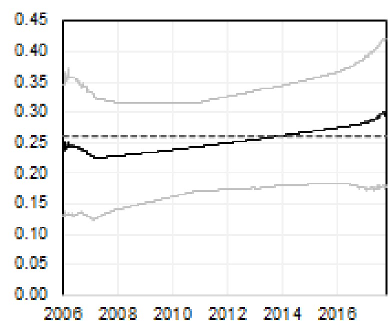


Figure 7: Time-varying estimates of the coefficient functions $\beta_3(\cdot)$ (black solid line) of TVHAR-CM model ($\log RV_m$ variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

2007-2008 but it is fair to highlight that the upper limit of the confidence bands are very close to zero.

Constants-coefficient estimate on realized kurtosis was detected as negative and significant for all the maturities. However, for all tenors we see the OLS estimate to lie out of the confidence bands for at least last 2 years (see

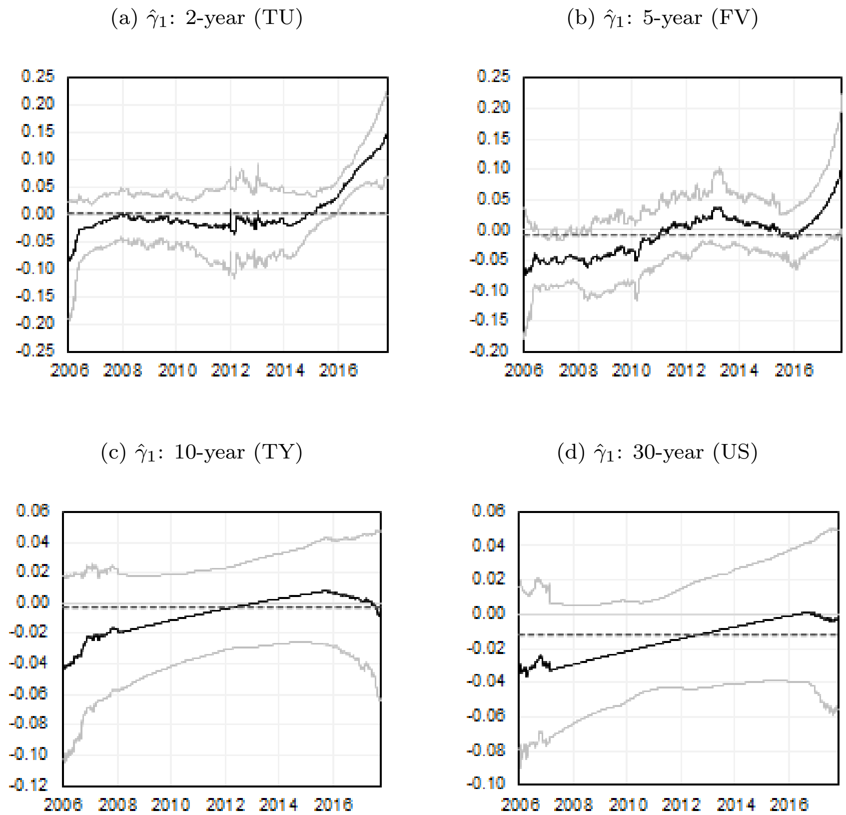
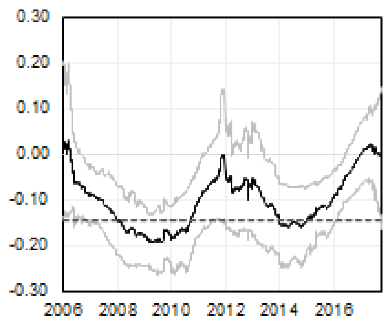


Figure 8: Time-varying estimates of the coefficient functions $\gamma_1(\cdot)$ (black solid line) of TVHAR-CM model (ΔRSK variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

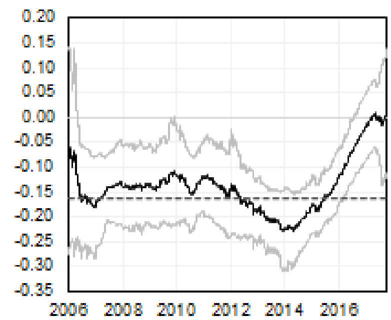
Figure 9). Importantly, values of $\hat{\gamma}_2$ including confidence bands remain well below zero for the vast majority of the sample period for all maturities.

As presented on Figure 10, development of $\hat{\delta}_1$ functions confirms significance of changes in risk-free rate detected by the static estimation only in the period 2007-2009 and since then the width of the confidence bands includes also zero across all maturities.

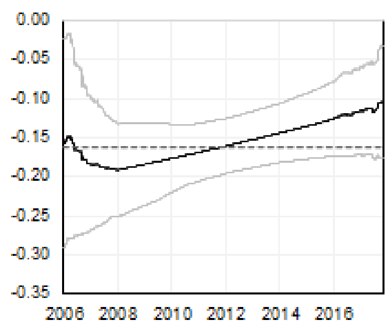
(a) $\hat{\gamma}_2$: 2-year (TU)



(b) $\hat{\gamma}_2$: 5-year (FV)



(c) $\hat{\gamma}_2$: 10-year (TY)



(d) $\hat{\gamma}_2$: 30-year (US)

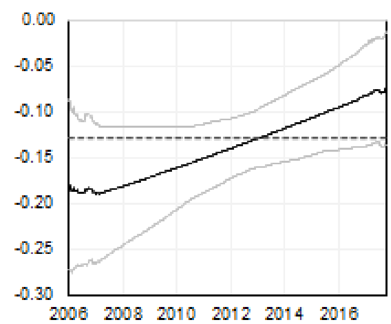


Figure 9: Time-varying estimates of the coefficient functions $\gamma_2(\cdot)$ (black solid line) of TVHAR-CM model (RKU variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

Effect of economic policy uncertainty was detected as significant for the 10-year and 30-year tenors only using standard OLS. Result of time-varying analysis summarized on Figure 11 provides a sligthly different view. For the shortest tenor (and to the large extent also for the 5-year tenor), we find several approximately 1-year windows where confidence bands of EPU effect is well above zero (around 2008 and 2017) and on the other hand periods with

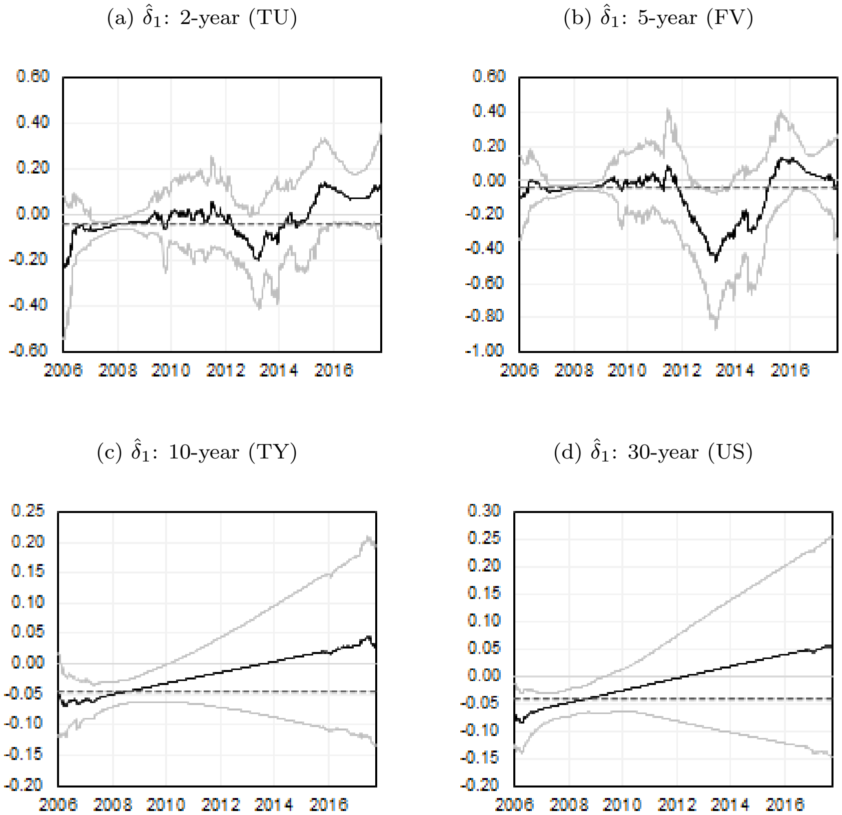
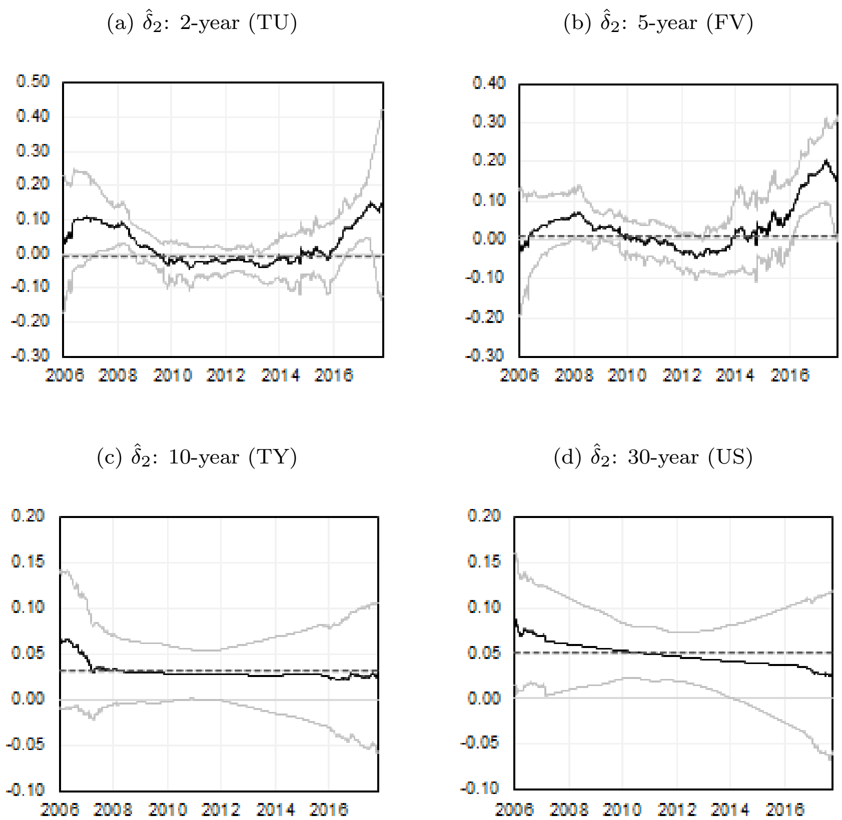


Figure 10: Time-varying estimates of the coefficient functions $\delta_1(\cdot)$ (black solid line) of TVHAR-CM model (ΔRF variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

90% probability of δ_2 to be negative (around 2013). Periods where economic policy uncertainty has positive effect on $\log RV$ are valid for the 10-year and 30-year tenors as well, but we do not find any evidence of potentially negative effect in case of these tenors.

The remaining control variable capturing changed in expectations of stock market volatility (ΔVIX) shows relatively stable and similar pattern for



(a) $\hat{\delta}_2$: 2-year (TU)

(b) $\hat{\delta}_2$: 5-year (FV)

(c) $\hat{\delta}_2$: 10-year (TY)

(d) $\hat{\delta}_2$: 30-year (US)

Figure 11: Time-varying estimates of the coefficient functions $\delta_2(\cdot)$ (black solid line) of TVHAR-CM model (EPU variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

all tenors much in line with the OLS estimates being included within the confidence bands of the time-varying estimates (see Figure 12).

We have examined the in-sample fit Model Confidence Set procedure (Hansen et al., 2011) (as presented in the following section in more detail) based mean square error and mean absolute error and we have confirmed the TVHAR-CM model to provide the best in-sample fit among other model

(a) $\hat{\delta}_3$: 2-year (TU) (b) $\hat{\delta}_3$: 5-year (FV)

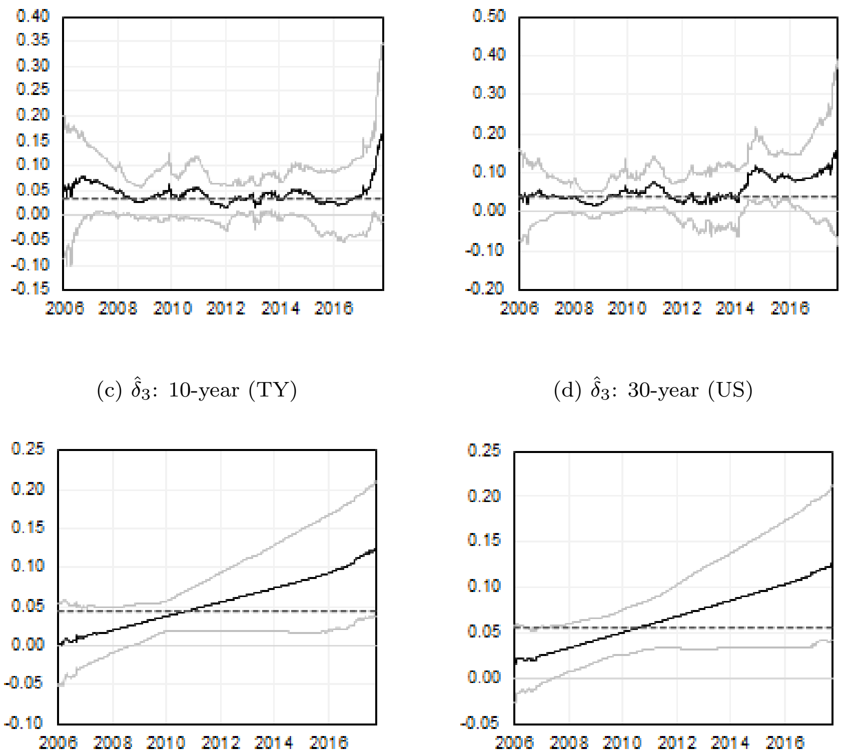


Figure 12: Time-varying estimates of the coefficient functions $\delta_3(\cdot)$ (black solid line) of TVHAR-CM model (ΔVIX variable) with 90% bootstrap confidence interval (grey solid line) and respective constant-coefficient OLS estimate (grey dashed line)

specifications.[8]

[8]Since we focus primarily on the out-of-sample performance, results of MCS procedure for the in-sample analysis are untabulated and are available upon request. TVHAR-CM model was the only model included in the 20% set of superior models for all the tenors.

4.2. Out-of-sample forecasting performance

Whereas we have found the TVHAR-CM model a clear outperformer in-sample, out-of-sample volatility forecasting is more important for market practitioners and since HAR model by (Corsi, 2009) is very popular thanks to its strong out-of-sample performance, beating HAR is a true challenge for remaining seven model specifications.

We inspect the forecasting performance at various horizons. First, we generate 1-day ($logRV_{t+1}$), 5-day ($logRV_{t+6}$) and 21-day ($logRV_{t+22}$) ahead forecasts of volatility on the respective day. Second we also inspect ability to forecast average volatility in coming week ($logRV_{w,t+1}$) and month ($logRV_{m,t+1}$).

To generate forecasts at various horizons based on constant-coefficient models, we start with first 2,300 days leaving the remaining 766 days ($\approx$ 25% of data) as out-of-sample evaluation period. Rolling-estimation window moves by 1 day (having fixed length of 2,300 days) until the end of out-of-sample period. Similarly, for 1-step-ahead forecast based on time-varying models we use first 2,300 observations using optimal bandwidth set in all sample estimation for respective model specification and tenor[9] to obtain local linear estimates of the respective coefficient functions. We note that for the purpose of the out-of-sample forecasting procedure, we set the minimum bandwidth to 0.21 corresponding to the widely used normal reference rule ($b = 1.06\sigma n^{-1/5}$) used as a rule-of-thumb optimum bandwidth value.

[9] In ideal case the optimal bandwidth should be optimized for each estimation round. However, as claimed also by Chen et al. (2018), there is only limited accuracy gain at a cost of significantly longer and more demanding computation process.

In order to evaluate in-sample fit of the competing models for individual maturities, we consider multiple loss functions. First, we employ commonly used mean absolute error (MAE) and mean squared error (MSE) defined as:

$$MAE = \frac{1}{T}\sum_{i=1}^{T} |\hat{RV}_t - RV_t| \tag{10}$$

$$MSE = \frac{1}{T}\sum_{i=1}^{T} (\hat{RV}_t - RV_t)^2 \tag{11}$$

We take into account also additional loss functions in order to control for a potential systemic asymmetry in forecasting errors of individual models. We follow Nomikos and Pouliasis (2011) and construct two mean mixed error functions each penalizing either over- or under-predictions:

$$MME(O) = \frac{1}{T}\left(\sum_{i\in U} |\hat{RV}_t - RV_t| + \sum_{i\in O} \sqrt{|\hat{RV}_t - RV_t|}\right) \tag{12}$$

$$MME(U) = \frac{1}{T}\left(\sum_{i\in U} \sqrt{|\hat{RV}_t - RV_t|} + \sum_{i\in O} |\hat{RV}_t - RV_t|\right) \tag{13}$$

where O and U are sets of days of over-predictions and under-predictions, respectively. All error metrics for each model are reported in relative terms to the forecasting errors generated by constant-coefficient HAR model serving as a benchmark in our analysis.

In order to pick the best performing models from the total of 8 model specifications (namely HAR, HAR-M, HAR-C and HAR-CM and their time-varying counterparts) for each maturity, we apply a Model Confidence Set (MCS) methodology introduced by Hansen et al. (2011). As compared to Diebold-Mariano test (Diebold and Mariano, 2002) or CPA test (Giacomini and White, 2006) often used in related literature for comparison of two competing models, MCS is designed to select a superior subset of tested models with statistically equal performance. Let M_0 denote set of compared models. MCS procedure consists of sequential testing of zero hypothesis of equal performance of all model combinations $H_{0,M} = E|L_{i,t} - L_{j,t}| = 0$, where L is a chosen loss function and $i, j \in M$. Sequential elimination of worst-performing models (p-value $\leq \alpha$) results in model confidence set of surviving models $\hat{M}^*_{1-\alpha}$ containing the best-performing model at $1 - \alpha$ confidence. For our out-of-sample performance evaluation we consider 80% confidence level with an average block length of 3 days[10].

When evaluating the out-of-sample forecasting performance we proceed as follows. First, we focus on each forecasting horizon separately to reveal differences across the term structure. Second, we look at individual tenors and find the most versatile models irrespective of the forecasting horizon. Third, we take a closer look at potential asymmetry of the forecasting errors to reveal potential systemic over- or under-predictions of the models. Finally, we check whether there is a single model performing reasonably well across horizons and maturities significantly outperforming the HAR model popular among practitioners for its versatility and solid forecasting performance.

Table 6 shows one-day ahead loss relative to HAR model for each tenor and loss function. We see that except for the shortest tenor, HAR model is always included in the model confidence set but with the only exception never ranks among top three best performing models based on any loss func-

[10]We use the default setting of the MCS package for R software. Change of block length does not have any material impact on our results.

tion. Even if the MCS procedure has not eliminated many models, we detect the first signals of systemic forecasting performance differences for the short and long maturities which become more evident for the longer forecasting horizons. First, we find that inclusion of control variables (risk-free, EPU index or VIX index) significantly under-perform the others both in constant-coefficient and time-varying specification for short tenors. On the other hand, inclusion of realized moments is significantly improving the forecasting accuracy especially in the time-varying version (TVHAR-M ranks among top 3 for all the loss functions). On the other hand, we find the control variables to contribute significantly to forecasting performance for the longest tenor. Generally we find the time-varying models to be inferior to the constant coefficient models for 30-year tenor since except for TVHAR none of the models ranks among top 3 for any loss function.

Model confidence set for forecast of mean volatility in the next 5 days shows similar results as the next-day forecasts described above with few interesting differences. First, we find the HAR model to significantly under-perform also in case of the 5-year tenor in addition to the shortest one. Interestingly, in contrast to the 1-day ahead case, we find the time-varying models to be included in the model confidence set for all loss functions also in case of the 30-year tenor. Looking at the top 3 best performing models, we detect clear under-performance of the constant-coefficient models in general both for 2-year and 5-year tenors. On the other hand, TVHAR-M has ranked top 3 for both tenors based on any loss function. Reverse conclusions hold for the 10-year and 30-year tenors where the time-varying models rank among three best based on a single loss function at most and constant-coefficient

		Daily: $logRV_{t+1}$			
Loss	Model	**TU**	**FV**	**TY**	**US**
MAE	HAR	1.000	**1.000**	**1.000**	**1.000**
	HAR-M	<u>**0.980**</u>	<u>**0.993**</u>	**1.000**	**1.000**
	HAR-C	0.996	**0.994**	<u>**0.990**</u>*	<u>**0.991**</u>*
	HAR-CM	**0.981**	<u>**0.989**</u>*	<u>**0.993**</u>	<u>**0.992**</u>
	TVHAR	**0.980**	<u>**0.994**</u>	<u>**0.995**</u>	<u>**0.998**</u>
	TVHAR-M	<u>**0.971**</u>*	**0.999**	**1.002**	1.025
	TVHAR-C	**0.983**	1.056	**1.001**	**1.007**
	TVHAR-CM	<u>**0.976**</u>	**0.999**	**1.000**	1.014
MSE	HAR	1.000	**1.000**	**1.000**	<u>**1.000**</u>
	HAR-M	**0.977**	<u>**0.981**</u>	<u>**0.990**</u>	**1.004**
	HAR-C	0.998	**0.995**	**0.991**	<u>**0.992**</u>*
	HAR-CM	**0.979**	<u>**0.977**</u>*	<u>**0.981**</u>*	<u>**0.997**</u>
	TVHAR	<u>**0.971**</u>	**0.987**	**0.991**	**1.009**
	TVHAR-M	<u>**0.961**</u>*	<u>**0.983**</u>	<u>**0.991**</u>	1.042
	TVHAR-C	**0.984**	1.108	**1.019**	1.027
	TVHAR-CM	<u>**0.973**</u>	**0.993**	**0.999**	1.034
MME(O)	HAR	**1.000**	**1.000**	**1.000**	**1.000**
	HAR-M	<u>**0.984**</u>	<u>**0.999**</u>	**1.004**	<u>**0.997**</u>
	HAR-C	**0.994**	<u>**0.995**</u>	<u>**0.992**</u>*	<u>**0.990**</u>*
	HAR-CM	<u>**0.983**</u>*	<u>**0.995**</u>*	<u>**0.999**</u>	<u>**0.991**</u>
	TVHAR	**0.994**	**0.999**	<u>**0.997**</u>	**0.998**
	TVHAR-M	<u>**0.990**</u>	**1.006**	**1.005**	1.016
	TVHAR-C	**0.995**	1.048	**1.001**	**1.000**
	TVHAR-CM	**0.993**	**1.006**	**1.005**	**1.002**
MME(U)	HAR	1.000	**1.000**	**1.000**	**1.000**
	HAR-M	**0.981**	<u>**0.995**</u>	**0.999**	**0.998**
	HAR-C	1.000	**0.997**	<u>**0.992**</u>*	<u>**0.994**</u>
	HAR-CM	**0.983**	<u>**0.993**</u>*	<u>**0.994**</u>	<u>**0.991**</u>*
	TVHAR	<u>**0.981**</u>	<u>**0.996**</u>	<u>**0.996**</u>	<u>**0.995**</u>
	TVHAR-M	<u>**0.970**</u>*	**0.998**	**1.003**	1.020
	TVHAR-C	**0.984**	1.036	**0.997**	**1.005**
	TVHAR-CM	<u>**0.974**</u>	**0.998**	**0.999**	1.013

The table shows MAE, MSE, MME(O) and MME(U) loss functions relative to the loss of HAR model. Values in bold indicate the model to be included in the 80% model confidence set for the respective tenor. Underlined values indicate the model to rank among top 3 best performing models and * denotes the best model according to T_{max} statistics.

Table 6: Out-of-sample evaluation: 1-day volatility forecasts ($logRV_{t+1}$)

		Weekly: $logRV_{w,t+1}$			
Loss	Model	**TU**	**FV**	**TY**	**US**
MAE	HAR	1.000	1.000	**1.000**	**1.000**
	HAR-M	**0.975**	**0.989**	**0.992**	**0.997**
	HAR-C	0.997	**0.990**	**0.987**	**0.985**
	HAR-CM	**0.971**	**0.981**	**0.978** *	**0.982** *
	TVHAR	**0.944**	**0.976**	**1.008**	**1.006**
	TVHAR-M	**0.943** *	**0.966**	**1.013**	**1.026**
	TVHAR-C	**0.948**	**0.970**	**0.992**	**1.005**
	TVHAR-CM	**0.943**	**0.959** *	**0.996**	**1.018**
MSE	HAR	1.000	1.000	**1.000**	**1.000**
	HAR-M	**0.955**	**0.970**	**0.987**	**1.002**
	HAR-C	0.982	**0.982**	**0.971**	**0.979** *
	HAR-CM	**0.939**	**0.953**	**0.957** *	**0.982**
	TVHAR	**0.869**	**0.938**	**0.999**	**1.074**
	TVHAR-M	**0.871** *	**0.920** *	**1.025**	**1.099**
	TVHAR-C	**0.878**	**0.949**	**1.006**	**1.084**
	TVHAR-CM	**0.872**	**0.932**	**1.029**	**1.102**
MME(O)	HAR	**1.000**	**1.000**	**1.000**	**1.000**
	HAR-M	**0.986**	**0.994**	**0.991**	**1.002**
	HAR-C	**0.995**	**0.993**	**0.992**	**0.990** *
	HAR-CM	**0.978** *	**0.987**	**0.982** *	**0.992**
	TVHAR	**0.990**	**0.989**	**1.018**	**1.004**
	TVHAR-M	**0.990**	**0.984**	**1.016**	**1.020**
	TVHAR-C	**1.002**	**0.990**	**1.012**	**1.011**
	TVHAR-CM	**0.997**	**0.977** *	**1.010**	**1.024**
MME(U)	HAR	1.000	**1.000**	**1.000**	**1.000**
	HAR-M	0.978	**0.995**	**0.998**	**0.995**
	HAR-C	1.006	**0.992**	**0.993**	**0.987**
	HAR-CM	0.986	**0.992**	**0.988**	**0.981**
	TVHAR	**0.938**	**0.982**	**0.995**	**0.984**
	TVHAR-M	**0.936**	**0.978**	**1.001**	**1.001**
	TVHAR-C	**0.935**	**0.970**	**0.977** *	**0.978** *
	TVHAR-CM	**0.932** *	**0.967** *	**0.981**	**0.985**

The table shows MAE, MSE, MME(O) and MME(U) loss functions relative to the loss of HAR model. Values in bold indicate the model to be included in the 80% model confidence set for the respective tenor. Underlined values indicate the model to rank among top 3 best performing models and * denotes the best model according to T_{max} statistics.

Table 7: Out-of-sample evaluation: 1-week volatility forecasts ($logRV_{w,t+1}$)

HAR-CM shows the most robust performance across the loss functions.

More differentiated forecasting performance in case of mean volatility in the next month (or 21 days) translates to more restricted model confidence sets (see Table 8). Except for the 2-year tenor, all time-varying model specifications are included in the model confidence set. For the shortest tenor, HAR-M model shows the most robust performance ranking among top three for 3/4 of the loss functions. For the longer maturities, static models never rank among top three models with very few exceptions. TVHAR and TVHAR-M models are superior for the 5-year tenor, whereas TVHAR and TVHAR-C always rank among the top in case of 10-year tenor.

When we look at forecasts of daily volatility in 5-days (Table 9), we again find contrasting poles of 2-year and 30-year tenors. In case of the former, we find the constant-coefficient models to be oftern excluded from the MCS whereas the time-varying models are always included and except for the TVHAR-CM model rank among top three models for each of the loss functions. On the contrary, none of the time-varying models is included in the model confidence set in case of the 30-year tenor. For the 5-year tenor, TVHAR and TVHAR-M show the most consistent performance and HAR-CM model always ranks among top three for both longer maturities. Benchmark HAR model never ranks among the top models for any loss functions and maturity.

For the 21-day horizon, we find only TVHAR model to be included in model confidence sets for all tenors based on any loss function. Together with TVHAR-M model it is a consistent out-performer ranking among top three for any loss function so is the constant-coefficient HAR-C model for

		Monthly: $logRV_{m,t+1}$			
Loss	Model	**TU**	**FV**	**TY**	**US**
MAE	HAR	**<u>1.000</u>**	**1.000**	1.000	**1.000**
	HAR-M	**<u>0.973</u>***	**0.983**	0.988	**0.996**
	HAR-C	**1.011**	1.014	0.994	**0.993**
	HAR-CM	**0.977**	**0.999**	0.982	**0.986**
	TVHAR	**1.030**	**<u>0.932</u>**	**<u>0.861</u>***	**<u>0.934</u>***
	TVHAR-M	**1.037**	**<u>0.918</u>***	**<u>0.874</u>**	**<u>0.946</u>**
	TVHAR-C	**1.056**	**0.964**	**<u>0.863</u>**	**<u>0.947</u>**
	TVHAR-CM	**1.059**	**<u>0.951</u>**	**0.879**	**0.964**
MSE	HAR	**<u>1.000</u>**	**1.000**	1.000	**1.000**
	HAR-M	**<u>0.931</u>***	**0.963**	**0.975**	**0.997**
	HAR-C	**1.015**	**1.011**	**0.988**	**0.999**
	HAR-CM	**<u>0.940</u>**	**0.971**	**0.962**	**1.000**
	TVHAR	**1.029**	**<u>0.917</u>**	**<u>0.790</u>***	**<u>0.897</u>***
	TVHAR-M	**1.052**	**<u>0.890</u>***	**<u>0.823</u>**	**<u>0.924</u>**
	TVHAR-C	**1.050**	**0.984**	**<u>0.836</u>**	**<u>0.955</u>**
	TVHAR-CM	**1.061**	**<u>0.956</u>**	**0.871**	**0.982**
MME(O)	HAR	**1.000**	**1.000**	**1.000**	**1.000**
	HAR-M	**<u>1.000</u>**	**<u>0.989</u>**	**0.991**	**<u>0.997</u>**
	HAR-C	**<u>0.980</u>**	**1.007**	**0.996**	**<u>0.995</u>**
	HAR-CM	**<u>0.969</u>**	**0.998**	**0.986**	**<u>0.987</u>**
	TVHAR	1.140	**<u>0.987</u>**	**<u>0.939</u>**	**1.008**
	TVHAR-M	1.142	**<u>0.970</u>***	**<u>0.946</u>**	**1.010**
	TVHAR-C	1.185	**1.028**	**<u>0.938</u>**	**1.021**
	TVHAR-CM	1.186	**1.011**	**0.948**	**1.030**
MME(U)	HAR	1.000	1.000	1.000	1.000
	HAR-M	**0.967**	0.986	0.994	1.001
	HAR-C	1.033	1.016	0.996	**0.994**
	HAR-CM	0.998	1.003	0.989	**0.990**
	TVHAR	**<u>0.912</u>***	**<u>0.914</u>**	**<u>0.859</u>**	**<u>0.904</u>***
	TVHAR-M	**<u>0.917</u>**	**<u>0.909</u>***	**0.868**	**<u>0.917</u>**
	TVHAR-C	**<u>0.923</u>**	**0.922**	**<u>0.856</u>***	**<u>0.907</u>**
	TVHAR-CM	**0.926**	**<u>0.916</u>**	**<u>0.867</u>**	**0.924**

The table shows MAE, MSE, MME(O) and MME(U) loss functions relative to the loss of HAR model. Values in bold indicate the model to be included in the 80% model confidence set for the respective tenor. Underlined values indicate the model to rank among top 3 best performing models and * denotes the best model according to T_{max} statistics.

Table 8: Out-of-sample evaluation: 1-month volatility forecasts ($logRV_{m,t+1}$)

		5 days ahead: $logRV_{t+5}$			
Loss	Model	**TU**	**FV**	**TY**	**US**
MAE	HAR	1.000	**1.000**	**1.000**	1.000
	HAR-M	0.995	**0.994**	**0.995**	**0.993**
	HAR-C	1.004	**0.999**	**0.995**	**0.994**
	HAR-CM	0.998	**0.993**	**0.990***	**0.988***
	TVHAR	**0.933***	**0.969**	**0.978**	1.019
	TVHAR-M	**0.939**	**0.966***	**0.978**	1.016
	TVHAR-C	**0.941**	**0.976**	**0.984**	1.044
	TVHAR-CM	**0.946**	**0.972**	**0.973**	1.025
MSE	HAR	**1.000**	**1.000**	**1.000**	**1.000**
	HAR-M	**1.003**	**0.999**	**1.002**	**0.993***
	HAR-C	**0.994**	**0.998**	**0.996***	**1.000**
	HAR-CM	**0.997**	**0.997**	**0.997**	**0.994**
	TVHAR	**0.971***	**0.985**	**0.998**	1.069
	TVHAR-M	**0.981**	**0.986***	**1.001**	1.065
	TVHAR-C	**0.971**	**0.995**	**1.006**	1.116
	TVHAR-CM	**0.980**	**0.995**	**1.001**	1.071
MME(O)	HAR	**1.000**	**1.000**	**1.000**	1.000
	HAR-M	**0.998**	**0.998**	**0.994**	**0.996**
	HAR-C	**1.002**	**0.998**	**0.995**	**0.994**
	HAR-CM	**0.997**	**0.997**	**0.989**	**0.991***
	TVHAR	**0.976***	**0.996**	**0.993**	1.037
	TVHAR-M	**0.980**	**0.994***	**0.993**	1.025
	TVHAR-C	**0.985**	**0.999**	**0.995**	1.057
	TVHAR-CM	**0.988**	**0.995**	**0.986***	1.036
MME(U)	HAR	1.000	**1.000**	**1.000**	1.000
	HAR-M	0.994	**0.993**	**0.993**	**0.995**
	HAR-C	1.006	**1.000**	**0.996**	**0.994**
	HAR-CM	0.999	**0.994**	**0.991***	**0.991***
	TVHAR	**0.954***	**0.993**	**1.006**	1.027
	TVHAR-M	**0.957**	**0.989***	**1.004**	1.030
	TVHAR-C	**0.957**	**0.998**	1.011	1.049
	TVHAR-CM	**0.961**	**0.995**	**1.002**	1.036

The table shows MAE, MSE, MME(O) and MME(U) loss functions relative to the loss of HAR model. Values in bold indicate the model to be included in the 80% model confidence set for the respective tenor. Underlined values indicate the model to rank among top 3 best performing models and * denotes the best model according to T_{max} statistics.

Table 9: Out-of-sample evaluation: 5-day ahead volatility forecasts ($logRV_{t+5}$)

		21 days ahead: $logRV_{t+21}$			
Loss	Model	**TU**	**FV**	**TY**	**US**
MAE	HAR	**1.000**	1.000	**1.000**	**1.000**
	HAR-M	**0.991**	**0.988**	**0.993**	**0.999**
	HAR-C	**1.015**	1.004	**0.999**	**0.996**
	HAR-CM	**1.005**	**0.991**	**0.992**	**0.995***
	TVHAR	**0.985**	**0.963**	**0.949**	**1.009**
	TVHAR-M	**0.992**	**0.963***	**0.949***	**1.034**
	TVHAR-C	**0.984***	**0.975**	**0.957**	1.056
	TVHAR-CM	**0.989**	**0.975**	**1.010**	1.084
MSE	HAR	**1.000**	**1.000**	1.000	**1.000**
	HAR-M	**0.994***	**0.987**	0.992	**1.000***
	HAR-C	**1.004**	**1.002**	0.998	**1.002**
	HAR-CM	**0.996**	**0.989**	0.990	**1.002**
	TVHAR	**1.053**	**0.984**	**0.942***	**1.021**
	TVHAR-M	1.064	**0.984***	**0.943**	**1.056**
	TVHAR-C	**1.046**	**0.993**	**0.950**	1.107
	TVHAR-CM	**1.054**	**0.992**	1.080	1.144
MME(O)	HAR	**1.000**	1.000	1.000	**1.000**
	HAR-M	**0.994***	0.990	0.996	**1.000**
	HAR-C	**1.006**	1.005	0.999	**0.996***
	HAR-CM	**1.000**	0.994	0.994	**0.998**
	TVHAR	**1.015**	**0.961**	**0.955**	**1.026**
	TVHAR-M	**1.019**	**0.960***	**0.954***	1.046
	TVHAR-C	**1.020**	0.975	0.965	1.067
	TVHAR-CM	**1.019**	0.976	1.013	1.089
MME(U)	HAR	1.000	1.000	1.000	**1.000**
	HAR-M	0.992	**0.990**	0.995	**0.998**
	HAR-C	1.015	1.001	1.000	**0.993**
	HAR-CM	1.006	**0.990**	**0.994**	**0.992***
	TVHAR	**0.960**	**0.971**	**0.966**	**0.994**
	TVHAR-M	**0.963**	**0.972***	**0.966***	**1.009**
	TVHAR-C	**0.955***	**0.980**	**0.970**	**1.023**
	TVHAR-CM	**0.958**	**0.980**	**0.999**	1.043

The table shows MAE, MSE, MME(O) and MME(U) loss functions relative to the loss of HAR model. Values in bold indicate the model to be included in the 80% model confidence set for the respective tenor. Underlined values indicate the model to rank among top 3 best performing models and * denotes the best model according to T_{max} statistics.

Table 10: Out-of-sample evaluation: 21-day ahead volatility forecasts ($logRV_{t+21}$)

the 30-year tenor. In case of the 2-year tenor there is no model ranking in the top three for more than two out of four loss functions.

Figure 13 summarises how frequently each of the models was included in the model confidence set or ranked among the three best performing models according to the T_{max} statistic as described in Hansen et al. (2011) revealing the most versatile models for each tenor irrespective of the forecasting horizon. We find time-varying models to outperform the constant-coefficient models for the 2-year, 5-year and 10-year maturity whereas constant-coefficient models especially those including the control variables show to be useful tools for the longest tenor. Despite HAR model has been included in the respective model confidence sets for 50%-80% cases, it ranked among top 3 only in 0-20% cases. In terms of model performance ranking, TVHAR and TVHAR-M are found as most versatile for the two shortest tenors, TVHAR and HAR-CM for the 5-year and the two static models containing the control variables for the longest maturity.

(a) Model confidence set

Model	TU	FV	TY	US
HAR	50%	70%	70%	80%
HAR-M	80%	90%	75%	95%
HAR-C	50%	75%	75%	100%
HAR-CM	75%	90%	80%	100%
TVHAR	95%	100%	100%	80%
TVHAR-M	90%	100%	100%	55%
TVHAR-C	95%	75%	90%	60%
TVHAR-CM	95%	95%	90%	45%

(b) Top 3 ranking

Model	TU	FV	TY	US
HAR	20%	0%	0%	20%
HAR-M	40%	35%	30%	50%
HAR-C	5%	5%	45%	75%
HAR-CM	35%	30%	60%	80%
TVHAR	55%	75%	65%	35%
TVHAR-M	65%	85%	40%	15%
TVHAR-C	40%	10%	40%	20%
TVHAR-CM	40%	55%	15%	0%

Figure 13: Frequency of inclusion to MCS and top-3 ranking (based on T_{max} statistic) across all forecasting horizons and loss functions

Finally, we analyze any potential systemic over- or under-predictions of the individual models. For the 1-day forecasting horizon, MME(U) and MME(O) do not show any permanent asymmetry of the forecasts and report only slight differences in rankings. The only discrepancy worth highlighting is that MME(O) detect HAR-M and HAR-CM models with the lowest errors whereas the MME(U) loss reports the lowest errors for their time-varying counterparts suggesting that inclusion of moments for the 2-year tenor volatility forecasting is the beneficial in any case and the dynamics in coefficients is nice-to-have if we care more about under-predictions.

For the weekly volatility predictions, we find a systemic under-predictions of the constant-coefficient models across the maturities. Looking at the opposite poles of the term structure, we observe the lowest MME(O) errors for the static models including moments (i.e. HAR-M and HAR-CM) for the 2-year and those including controls (i.e. HAR-C and HAR-CM) of the 30-year tenor. In both cases, MME(U) reports the comparable or lower errors for their respective time-varying counterparts. This shift is even stronger for the monthly volatility predictions where we see the strongest evidence of systemic under-prediction of the constant-coefficient models across the entire term structure irrespective of a concrete model specification.

In case of the 5-day horizon, we find similar outcome as for the 1-day horizon with no major shifts in model performance evaluation. However, in case of 21-day ahead forecasts, we again observe significant under-predictions for the 2-year tenor.

5. Conclusion

Volatility forecasting on bond markets is of indisputable interest to investors and researchers due to immense trading volumes of sovereign fixed-income securities and very different investment proposition to stocks not allowing for an automatic extension of evidence found on the stock market (heavily studied by researchers) also to the bond market. In this paper we study predictability of U.S. sovereign bond futures realized volatility over period 2006-2017 for 2-year, 5-year, 10-year and 30-year tenor. We extend heterogeneous autoregressive realized volatility model by Corsi (2009) serving as benchmark by higher-order realized moments, namely realized skewness and kurtosis while controlling for risk-free rate, economic policy uncertainty and expected equity market volatility captured by VIX index.

There are multiple contributions of this paper both for researchers and practitioners. We provide the first consistent study of volatility forecasting across the entire term structure confirming that different investment propositions of short and long tenors are also reflected in predictability of future volatility. Most importantly, we discover two main patterns in the out-of-sample analysis using Model Confidence Set evaluation procedure by Hansen et al. (2011) robust across considered loss functions. First, inclusion of higher moments (namely realized kurtosis) is beneficial but the effect is diminishing with increasing time to maturity. On the other hand, for the longer tenors, we find bond market control variables to contribute significantly. Second, allowing variation of the coefficients in time is of significant value for the shorter end of the term structure while extended constant-coefficient models (especially for the bond-market controls) are performing well for the longest tenors. Generally, we found the forecasting performance of the inspected models to decrease with time to maturity. We also inspect any systemic asymmetry using mean mixed errors introduced by Nomikos and Pouliasis (2011) penalizing over- or under-predictions. We found a strong tendency of the static models to generate under-predictions in our out-of-sample period.

References

Amaya, D., Christoffersen, P., Jacobs, K., Vasquez, A., 2015. Does realized skewness predict the cross-section of equity returns? Journal of Financial Economics 118, 135–167.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review , 885–905.

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Vega, C., 2007. Real-time price discovery in global stock, bond and foreign exchange markets. Journal of International Economics 73, 251–277.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. The quarterly journal of economics 131, 1593–1636.

Balduzzi, P., Elton, E.J., Green, T.C., 2001. Economic news and bond prices: Evidence from the us treasury market. Journal of financial and Quantitative analysis , 523–543.

Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. Journal of financial Econometrics 4, 1–30.

Barro, R.J., 2006. Rare disasters and asset markets in the twentieth century. The Quarterly Journal of Economics 121, 823–866.

Bekierman, J., Manner, H., 2018. Forecasting realized variance measures using time-varying coefficient models. International Journal of Forecasting 34, 276–287.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics 31, 307–327.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1–18.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. Journal of Financial Economics 118, 113–134.

Busch, T., Christensen, B.J., Nielsen, M.Ø., 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. Journal of Econometrics 160, 48–57.

Casas, I., Fernandez-Casal, R., 2019. tvreg: Time-varying coe fficient linear regression for single and multi-equations in r. Available at SSRN .

Chao, S.W., 2016. Do economic variables improve bond return volatility forecasts? International Review of Economics & Finance 46, 10–26.

Chen, X.B., Gao, J., Li, D., Silvapulle, P., 2018. Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. Journal of Business & Economic Statistics 36, 88–100.

Cieslak, A., Povala, P., 2016. Information in the term structure of yield curve volatility. The Journal of Finance 71, 1393–1436.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Demirer, R., Gkillas, K., Gupta, R., Pierdzioch, C., 2019. Time-varying risk aversion and realized gold volatility. The North American Journal of Economics and Finance 50, 101048.

Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. Journal of Business & economic statistics 20, 134–144.

Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications: monographs on statistics and applied probability 66. volume 66. CRC Press.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Gkillas, K., Gupta, R., Pierdzioch, C., 2019. Forecasting (downside and upside) realized exchange-rate volatility: Is there a role for realized skewness and kurtosis? Physica A: Statistical Mechanics and its Applications 532, 121867.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business & Economic Statistics 24, 127–161.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. Journal of the Royal Statistical Society: Series B (Methodological) 55, 757–779.

Kozhan, R., Neuberger, A., Schneider, P., 2013. The skew risk premium in the equity index market. The Review of Financial Studies 26, 2174–2203.

Kraus, A., Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. The Journal of finance 31, 1085–1100.

Malinska, B., 2020. Time-varying pricing of risk in sovereign bond futures returns. IES Working Papers .

Mei, D., Liu, J., Ma, F., Chen, W., 2017. Forecasting stock market volatility: Do realized skewness and kurtosis help? Physica A: Statistical Mechanics and its Applications 481, 153–159.

Neuberger, A., 2012. Realized skewness. The Review of Financial Studies 25, 3423–3455.

Nomikos, N.K., Pouliasis, P.K., 2011. Forecasting petroleum futures markets volatility: The role of regimes and market conditions. Energy Economics 33, 321–337.

Özbekler, A.G., Kontonikas, A., Triantafyllou, A., 2020. Volatility forecasting in european government bond markets .

Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters, in: Statistical Analysis and Forecasting of Economic Structural Change. Springer, pp. 253–264.

Taylor, S., 1986. Modelling Financial Time Series. Wiley.

Wu, X., Hou, X., 2019. Forecasting realized variance using asymmetric har model with time-varying coefficients. Finance Research Letters 30, 89–95.

Zhu, Q., You, M., Wu, S., 2020. Forecasting volatility with time-varying coefficient regressions. Discrete Dynamics in Nature and Society 2020.

	TU	FV	TY	US
(Intercept)	0.002	0.002	0.004	0.006
	(0.011)	(0.013)	(0.013)	(0.014)
$\log RV_d$	0.080 ***	0.104 ***	0.120 ***	0.130 ***
	(0.021)	(0.021)	(0.021)	(0.021)
$\log RV_w$	0.348 ***	0.310 ***	0.282 ***	0.251 ***
	(0.040)	(0.041)	(0.036)	(0.035)
$\log RV_m$	0.389 ***	0.329 ***	0.302 ***	0.288 ***
	(0.036)	(0.036)	(0.031)	(0.031)
ΔRF	-0.050 ***	-0.047 ***	-0.052 ***	-0.045 **
	(0.011)	(0.014)	(0.013)	(0.014)
EPU	0.004	0.005	0.029	0.051 **
	(0.011)	(0.013)	(0.014)	(0.015)
ΔVIX	0.038 ***	0.044 **	0.049 **	0.059 ***
	(0.011)	(0.013)	(0.014)	(0.014)
R^2	63.9%	50.9%	46.2%	42.0%

The number in paranthesis is heteroskedasticity and autocorrelation robust Newey-West standard error. ***, ** and * denotes siginficance at 0.1%, 1% and 5%, respectively.

Table .11: Estimation results of constant-coefficient HAR-C model

IES Working Paper Series

2021

1. Mahir Suleymanov: *Foreign Direct Investment in Emerging Markets: Evidence from Russia since the 2000s*
2. Lenka Nechvátalová: *Multi-Horizon Equity Returns Predictability via Machine Learning*
3. Milan Scasny, Matej Opatrny: *Elasticity of Marginal Utility of Consumption: The Equal-Sacrifice Approach Applied for the Czech Republic*
4. Javier Garcia-Bernardo, Petr Jansky and Vojtech Misak: *Common Agricultural Policy Beneficiaries: Evidence of Inequality from a New Data Set*
5. Petr Jakubik, Saida Teleu: *Suspension of Insurers' Dividends as a Response to the Covid-19 Crisis: Evidence from Equity Market*
6. Boris Fisera, Menbere Workie Tiruneh, David Hojdan: *Currency Depreciations in Emerging Economies: A Blessing or a Curse for External Debt Management?*
7. Vojtech Molnar: *Price Level Targeting with Imperfect Rationality: A Heuristic Approach*
8. Alex Cobham, Tommaso Faccio, Javier Garcia-Bernardo, Petr Jansky, Jeffery Kadet, Sol Picciotto: *A Practical Proposal to End Corporate Tax Abuse: METR, a Minimum Effective Tax Rate for Multinationals*
9. Evžen Kočenda, Ichiro Iwasaki: *Bank Survival Around the World: A Meta-Analytic Review*
10. Michal Kuchta: *Scenario Generation for IFRS9 Purposes using a Bayesian MS-VAR Model*
11. Jozef Barunik, Josef Kurka: *Frequency-Dependent Higher Moment Risks*
12. Petr Pleticha: *Who Benefits from Global Value Chain Participation? Does Functional Specialization Matter?*
13. Alex Cobham, Petr Jansky, Chris Jones, Yama Temouri: *An Evaluation of the Effects of the European Commission's Proposals for the Common Consolidated Corporate Tax Base*
14. Javier Garcia-Bernardo, Petr Jansky: *Profit Shifting of Multinational Corporations Worldwide*
15. Tomas Domonkos, Boris Fisera and Maria Siranova: *Income Inequality as Long-term Conditioning Factor of Monetary Transmission to Bank Interest Rates in EA Countries*
16. Petr Jansky, Natalia Li: *Improving the Corruption Perceptions Index: Additional Data Sources and Their Effects*
17. Dominika Ehrenbergerova, Josef Bajzik, Tomas Havranek: *When Does Monetary Policy Sway House Prices? A Meta-Analysis*
18. Daniel Kolář: *Inequality in Pre-Industrial Urban Bohemia: The City of Budweis*
19. Barbora Malinska: *Forecasting Sovereign Bond Realized Volatility Using Time-Varying Coefficients Model*

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz