Modelling and forecasting realised volatility in German-Austrian continuous intraday electricity auction prices

Ciarreta, A.[a], Muniain, P.[b] and Zarraga, A.[c]

[a] Department of Economic Analysis II, BiRE, BETS and BRiDGE, University of the Basque Country, UPV/EHU. Avda. Lehendakari Aguirre, 83. 48015 Bilbao. Spain. E-mail: aitor.ciarreta@ehu.eus.

[b] Department of Economic Analysis II, University of the Basque Country, UPV/EHU. Avda. Lehendakari Aguirre, 83. 48015 Bilbao. Spain. E-mail: peru.muniain@ehu.eus.

[c] (*Corresponding author*) Department of Applied Economics III, BiRE and BETS, University of the Basque Country, UPV/EHU. Avda. Lehendakari Aguirre, 83. 48015 Bilbao. Spain. E-mail: ainhoa.zarraga@ehu.eus.

Abstract

This paper uses high-frequency continuous intraday electricity auction price data from the EPEX market to estimate and forecast realised volatility. Three different jump tests are used to break down the variation into jump and continuous components using quadratic variation theory. Several heterogeneous autoregressive models are then estimated for the logarithmic and standard deviation transformations. GARCH structures are included in the error terms of the models when evidence of conditional heteroscedasticity is found. Model selection is based on various out-of-sample criteria. Results show that decomposition of realised volatility is important for forecasting and that the decision whether to include GARCH-type innovations might depend on the transformation selected. Finally, results are sensitive to the jump test used in the case of the standard deviation transformation.

Keywords: Realised volatility, jumps, GARCH, volatility forecasting

JEL classification: C14, C53, Q47

1 Introduction

With the liberalisation of European electricity markets, a need for transparency in electricity price formation has arisen. Market participants rely on these prices as a signpost for making optimal consumption and production decisions. A better understanding of their dynamics is therefore crucial. Electricity is a non-storable good, which makes prices very volatile. This is a significant difference with respect to other traded commodities. Moreover, characteristics such as mean reversion, seasonality and stationarity lead to frequent price spikes. The need to continuously balance demand and supply complicates market architecture and has led to the development of market solutions such as intra-day markets, futures and other derivatives. Hence, price dynamics need to be modelled and forecasted to enable participants to hedge against the risk of unforeseen contingencies. As Geman and Roncoroni (2006) point out, electricity price risk forces the energy industry to develop more accurate forecasting models. In particular, the analysis of volatility is crucial in ensuring market efficiency and derivative pricing.

With the availability of high-frequency, highly volatile data, quadratic variation theory has proved to be a useful approach for analysing the daily realised volatility (RV) of prices (Barndorff-Nielsen and Shephard, BNS, 2004, 2006 and Andersen *et al.*, ABD, 2007). It is employed to identify significant price jumps and decompose total variation into its jump and non-jump components non-parametrically, as introduced by BNS (2004) and modified by ABD (2007) so as to prevent the process of jump detection from being downward biased. Applications to electricity markets can be found in Chan *et al.* (2008), Ullrich (2012), Haugom *et al.* (2011) and Haugom and Ullrich (2012), among others. Additionally, other jump-robust test statistics have been developed using different types of estimators for continuous variation in price processes. Some examples are Andersen *et al.* (2012) and Corsi *et al.* (2010). To our knowledge they have not been applied to electricity markets. The continuous nature of the price process required means that the analysis is better suited to continuous intra-day markets such as the German-Austrian one.

RV has been treated econometrically in several ways, starting with the heterogeneous autoregressive process (HAR) introduced by Müller *et al.* (1990) for financial time series (subsequently improved by Corsi, 2004) and the so-called HAR-RV model, in which RV is parameterised as a function of its own lags averaged over different time horizons. Based on the HAR-RV model, ABD (2007) introduce the jump component of variation and develop the so-called HAR-RV-CJ forecasting model. Chan *et al.* (2008) modify the model to fit electricity price time series more closely. In general, benefits in forecasting are obtained when total variation is decomposed into jump and non-jump components.

GARCH-type models have also been employed to model the second moment of electricity prices, but they are thought to be less suited than RV models to capturing the intraday volatility of high-frequency data. In fact, Andersen *et al.* (2003) suggest that simple models of RV explain electricity price dynamics better than GARCH and related stochastic volatility models in out-of-sample forecasting. Chan *et al.* (2008) compare the HAR-CV-JV model with EGARCH in five power markets in Australia and find mixed results. Frömmel *et al.* (2014) compare the EGARCH model to Realised GARCH-type models in the estimation of price volatility in the EPEX market and find better one-step-ahead predictions using the latter. Corsi *et al.* (2008) add GARCH innovations to the HAR model to account for volatility clusters and non-Gaussian distribution of RV using financial data.

The aim of this paper is to estimate different econometric models for the RV obtained from the prices of continuous intraday 15-minute blocks on the German and Austrian electricity market and select the best one in terms of volatility forecasting performance. This is why this market is used to make almost-real-time adjustments. Although the volume traded is only 1-2% of total electricity trading, its importance is growing because it enables selling and purchasing agents to optimise generation and consumption plans, respectively. Therefore, there is growing interest in understanding the dynamics of markets of this type. Following ABD (2007), two non linear transformations of RV are considered in the analysis: the logarithmic, whose distribution is closer to the Gaussian, and the square root, which is the standard deviation. The HAR-RV and HAR-CV-JV models by Chan *et al.* (2008) are estimated, in which the total RV is decomposed using three robust-to-jumps tests. Additionally to the BNS test, widely employed in electricity markets, the Andersen *et al.* (2012) and Corsi *et al.* (2010) jump tests are also used. The characteristics of the innovations in the models are also studied to determine whether GARCH-type structures are appropriate, as per Corsi *et al.* (2008) for financial markets. In this paper, both simple GARCH and more elaborate EGARCH structures are considered. The models are compared in terms of out-of-sample forecast performance according to several criteria, thus enabling conclusions to be drawn in regard, for instance,to the importance of including jumps and the sensitivity of the results to the jump test used.

The rest of the paper is organised as follows: Section **??** explains how quadratic variation theory is used to decompose RV into its continuous and jump components. Section **??** describes the EPEX continuous auction and the data used in the analysis. Section **??** presents the econometric models used and the estimation results. Section **??** presents several out-of-sample forecast criteria used to compare the predictive power of each model estimated. Finally, Section **??** summarises and concludes.

2 Realised volatility and jump detection

RV is taken as a measure for the unobserved volatility of a high-frequency time series. It is estimated based on quadratic variation theory, which also enables total price variation to be decomposed non-parametrically into its continuous and jump components for sufficiently equally spaced observations.[1]

Assume that prices are set M times each day t and that there are T days in the sample. The RV for day t can be estimated as:

[1] A detailed explanation can be found in BNS (2004, 2006) and ABD (2007).

$$RV_t = \sum_{j=1}^{M} r_j^2, \tag{1}$$

where r_j is the intraday price difference for time j.[2]

Integrated volatility (IV), which captures the continuous, predictable component of RV, can be estimated using different methods. Although each estimator has its own properties, there is no single method which is preferred to the rest. In this paper three jump-robust type estimators are considered. First, BNS (2004, 2006), based on realised bipower variation, second Andersen *et al.* (2012), based on the minimum or median RV and finally Corsi *et al.* (2010), based on threshold bipower variation. BNS (2004, 2006) estimate the IV using bipower variation (BV), given by:

$$BV_t = 1.57 \frac{M}{(M-1)} \sum_{j=2}^{M} |r_j||r_{j-1}|$$

Huang and Tauchen (2005) suggest the following statistic, based on the difference between RV and the jump-robust measure of variance, BV, to detect significant jumps:

$$\sqrt{M} \frac{(RV_t - BV_t)/RV_t}{\sqrt{0.61 \max[1, TQ_t/BV_t^2]}} \tag{2}$$

The denominator represents the effect of integrated quarticity, which can be estimated using tripower quarticity, $TQ_t = 1.74 \left(\frac{M^2}{M-2}\right) \sum_{j=3}^{M} (|r_j||r_{j-1}||r_{j-2}|)^{4/3}$.

In finite samples the BV is upward biased in the presence of jumps and consequently the jump component is underestimated, but on the other hand the BV is affected by zero returns, thus reducing its value and consequently detecting more jumps.

Corsi *et al.* (2010), CPR henceforth, propose a consistent and nearly unbiased estimator of IV based on threshold bipower variation:

$$CTBPV_t = 1.57 \sum_{j=2}^{M} Z_1(r_j, \vartheta_j) Z_1(r_{j-1}, \vartheta_{j-1}),$$

where the function Z_1 is defined as:

[2]Negative prices in the sample make it unfeasible to calculate returns as log price differences. Therefore, we use price differences as in Chan *et al.* (2008). Moreover, electricity is a non-storable good so price differences should not be considered as returns in a traditional sense, but we refer to them as such for the sake of notational convenience. Given the continuous nature of the electricity market, the first value of r for each day is calculated as the difference between the first price of that day and the last one of the previous day.

$$Z_1(r_j, \vartheta_j) = \begin{cases} |r_j| & \text{if } r_j^2 \leq \vartheta_j \\ \\ 1.094\vartheta_j^{1/2} & \text{if } r_j^2 > \vartheta_j \end{cases}$$

which depends on the returns and the value of the threshold $\vartheta_j = c_\vartheta^2 \hat{V}_j$. As Corsi *et al.* (2010) suggest, we consider $c_\vartheta = 3$ and use a non-parametric filter to obtain the local volatility estimate $\hat{V}_j$ based on an iterative process and the use of a Gausssian kernel. The authors show that the choice of the threshold does not affect the robustness of the IV estimator, and its impact on estimation is marginal.

The corresponding jump test statistic is:

$$\sqrt{M}\frac{(RV_t - CTBPV_t)/RV_t}{\sqrt{0.61max\left[1, \frac{CTTriPV_t}{CTBPV_t^2}\right]}}, \tag{3}$$

where $CTTriPV_t = 1.74M\sum_{j=3}^{M}\prod_{k=1}^{3} Z_{4/3}(r_{j-k+1}, \vartheta_{j-k+1})$ and $Z_{4/3}$ is defined in the original paper. The test statistic is more powerful than those based on multipower variation, but it is also affected by zero returns.

Finally, Andersen *et al.* (2012) propose the MinRV and MedRV jump-robust consistent estimators of IV using the nearest neighbour truncation. These estimators provide better jump-robustness than BV in finite samples, but they have a larger asymptotic variance than the CTBPV. Moreover, the MinRV is exposed to zero returns and is less efficient than the MedRV estimator, which is why we only consider the latter. This is based on the median on blocks of three consecutive returns:

$$MedRV_t = 1.42\left(\frac{M}{M-2}\right)\sum_{j=2}^{M-1} med(|r_{j-1}|, |r_j|, |r_{j+1}|)^2$$

and can be used to construct the jump test statistic:

$$\sqrt{M}\frac{(RV_t - MedRV_t)/RV_t}{\sqrt{0.96max\left[1, \frac{MedRQ_t}{MedRV_t^2}\right]}}, \tag{4}$$

where $MedRQ_t = 0.92\frac{M^2}{M-2}\sum_{j=2}^{M-1} med(|r_{j-1}|, |r_j|, |r_{j+1}|)^4$ is the estimator of the integrated quarticity.

Andersen *et al.* (2012) argue that increasing the block length of the returns in the estimation of IV means losing efficiency as occurs in the case of higher order multipower variation measures. Therefore, we conduct all the jump tests for the IV estimators based on adjacent returns.

It can be proved that the jump test statistics obtained with the three methods (equations (??), (??) and (??)) converge to a normal distribution when $M \to \infty$. Hence, if the statistic exceeds the critical value $\Phi_{1-\alpha}$, day t is classified as a jump day. For a chosen significance level α, the jump component of volatility at day t is obtained as:

$$JV_t = I_{Z_t > \Phi_{1-\alpha}} (RV_t - \widehat{IV}_t),$$

where Z_t is the jump test statistic and $\widehat{IV}_t$ the corresponding estimator of IV obtained with each method, and $I_{Z_t > \Phi_{1-\alpha}}$ is 1 if $Z_t > \Phi_{1-\alpha}$, and 0 otherwise.

Once total RV and jump variation (JV) are estimated, the continuous component of the total variation, CV, is given by the difference between the two, $CV_t = RV_t - JV_t$, and measures the variation without taking jumps into account.

3 Data

EPEX is the European Power Exchange platform for day-ahead and intraday trading of electricity on the French, German, Austrian and Swiss markets. These four countries account for approximately a third of Europe's electricity consumption. Although the volume traded on the intraday market is not as big as on the day-ahead market, it is designed to fit generation to real time demand, so there is increasing interest on the part of the different market agents in forecasting close-to-delivery electricity demand. Within the intraday market, we focus on the 15-minute contract continuous auction for Germany and Austria, which is the closest-to-real-time market mechanism used to balance demand and supply.

In the intraday continuous market, 15-minute periods can be traded until 45 minutes before delivery begins. Starting at 4pm on the current day, all 15-minute periods in the following day can be traded. The prices range from €-9999.99 to €9999.99/MWh and the minimum increment is €0.01/MWh. The sample period runs from November 18, 2012 to April 30, 2016 (EPEX, 2016). Before that time, intraday continuous auctions for 15-minute contracts were not sufficiently developed at either pole and very few quarters with observations are available. Thus, overall we have 1260 days and 120,956 observations.

Table ?? reports the descriptive statistics for the total price series and those depending on the season of the year and the day of the week. The mean price is €33.44/MWh with a standard deviation of €20.52/MWh, but several seasonal effects are observed. Mean prices are higher in autumn and summer, probably due to the lower availability of renewable power generation (especially hydroelectric), which means that fossil fuel plants must be used to meet demand. Day of the week seasonality is observed because lower economic activity at weekends results in lower prices. Volatility decreases for summer and Saturdays. It is also demonstrated that the price series distribution of probability is skewed and presents excess kurtosis. According to the Jarque-Bera test, the price distribution of probability is not normal. Finally, seasonal behaviour is observed in the quarter-hours, since the mean price decreases with the quarter-hours for the first and last hours of the day. The behaviour of mean prices in the central hours of the day is more irregular but in general prices are higher in peak hours.[3]

Table 1: Descriptive statistics for prices per season and day of the week

	Wtr.	Spr.	Sum.	Aut.	Mn.	Tu.	Wd.	Th.	Fr.	St.	Sn.	Total
Minimum	-135.69	-117.06	-73.06	-211.84	-170.34	-211.84	-188.91	-67.98	-84.08	-117.06	-135.69	-211.84
Maximum	212.44	240.99	236.35	250.00	203.97	211.32	174.18	250.00	240.99	136.41	140.16	250.00
Mean	32.88	30.86	34.26	35.83	34.65	37.58	38.03	36.34	36.08	28.89	22.53	33.44
Median	32.00	29.67	32.90	35.53	33.53	35.73	36.01	34.12	34.27	28.80	23.58	32.38
St. Dev.	20.12	19.63	18.54	22.96	20.76	21.35	21.06	19.66	19.05	17.44	19.21	20.52
Skewness	0.26^a	0.72^a	0.90^a	-0.66^a	0.18^a	-0.19^a	-0.23^a	0.96^a	1.39^a	-0.35^a	-0.75^a	0.16^a
Kurtosis (Ex.)	3.52^a	7.35^a	4.87^a	7.99^a	3.81^a	8.86^a	9.65^a	4.67^a	10.04^a	3.86^a	3.37^a	6.21^a
Jarque-Bera	18311.00^a	67912.62^a	29763.33^a	83820.35^a	10569.94^a	56600.86^a	67265.73^a	18355.67^a	78200.61^a	11079.78^a	9780.75^a	194852.25^a

Wtr., Spr., Sum. and Aut. stand for winter, spring, summer and autumn, respectively. Mn., Tu., Wd., Th., Fr., St. and Sn. are the days of the week starting from Monday. St. Dev. and Kurtosis (Ex.) denote standard deviation and excess kurtosis, respectively. a indicates rejection of the corresponding null hypothesis (no skewness, no excess kurtosis and normal distribution) at the 1% level.

Following Ullrich (2012), the medians of the returns are subtracted from the returns for each month of the year, day of the week and quarter-hour of the day to take into account the seasonal effects observed. The resulting adjusted returns are:

$$r_j^* = r_j - \hat{r}_{m,d,q},$$

where $\hat{r}_{m,d,q}$ is the median of the month (m), day of the week (d) and quarter-hour (q). The returns are replaced by adjusted returns, r^*, in all the calculations.

In empirical applications two non-linear transformations of RV are usually considered. The first is the square root of the volatility, $\sqrt{RV}$, which is the realised standard deviation and can be of interest in many applications, such as financial ones, and the second is the logarithmic realised volatility, log RV, which is also used because its distribution is closer to the normal.[4] The concavity of the non-linear forms means that high values of the time series decrease more than low values, so that the time series becomes smoother and the jumps are therefore less pronounced. The logarithmic function is more concave than the square root function. Hence, the smoothest time series is the one with the logarithmic form. In this paper both non-linear transformations of RV are considered to analyse the robustness of the results to the type of transformation.

Table **??** presents the descriptive statistics of the RV, defined in equation (**??**), and the two transformations defined above. As expected, the standard deviation decreases as more concave non-linear forms are chosen. The three time series distribution of probability shows positive skewness and leptokurtosis and none of them is normal, as confirmed by the Jarque-Bera statistic.

Table **??** reports the descriptive statistics for each component of the total RV, for each transformation and for each jump test. Jump detection is tested at the 0.1% significance

[3] To save space, descriptive statistics for the quarter-hours are not shown. They are however available from the authors upon request.

[4] Ullrich (2012) and Chan *et al.* (2008) consider the standard deviation, while Corsi *et al.* (2008), Haugom and Ullrich (2012) and ABD (2007) use both transformations.

Table 2: Descriptive statistics of RV

	RV	$\sqrt{RV}$	$\log RV$
Minimum	1496.00	38.68	7.31
Maximum	297913.72	545.81	12.60
Mean	13626.28	107.18	9.20
Median	9264.61	96.25	9.13
St. Dev.	16065.86	46.27	0.73
Skewness	7.01^a	2.38^a	0.58^a
Kurtosis (Ex.)	90.51^a	10.90^a	0.67^a
Jarque-Bera	440356.40^a	7420.86^a	94.14^a

St. Dev. and kurtosis (Ex.) denote standard deviation and excess kurtosis, respectively. a indicates rejection of the corresponding null hypothesis (no skewness, no excess kurtosis and normal distribution) at the 1% level.

level, that is, $\alpha = 0.001$. The difference in the number of days classified as jump-days with each test is noteworthy. 484 of the 1260 days (38.41%) are classified as jump-days under the CPR test, but there are only 6 jump-days (0.48%) under the Med. In a moderate position, the BNS test detects jumps in 214 days (16.98%). Consequently, the largest standard deviation of the JV component corresponds to the CPR test followed by the BNS and the Med. As expected, the mean CV under any of the jump tests is lower than the mean RV.

Table 3: Descriptive statistics of CV and JV

	CV			JV		
	BNS	CPR	Med	BNS	CPR	Med
Minimum	1411.30	1397.26	1496.00	794.78	574.81	1046.39
Maximum	297913.72	297913.72	297913.72	62408.37	88202.70	4622.52
Mean	12590.42	10566.52	13611.46	6098.97	7965.49	3112.90
Median	8612.94	7766.97	9205.58	3699.52	4413.30	3650.04
St. Dev.	14791.38	12026.14	16070.91	7724.16	10546.16	1531.10
Skewness	7.71^a	12.11^a	7.01^a	4.30^a	3.87^a	-0.58
Kurtosis (Ex.)	116.85^a	262.55^a	90.41^a	23.34^a	19.64^a	-1.96
Jarque-Bera	729295.69^a	3649823.21^a	439477.03^a	5513.96^a	8986.13^a	1.30
Jump days (%)				16.98	38.41	0.48

St. Dev. and kurtosis (Ex.) denote standard deviation and excess kurtosis, respectively. a indicates rejection of the corresponding null hypothesis (no skewness, no excess kurtosis and normal distribution) at the 1% level.

As mentioned in Section **??**, both BNS and CPR tests are affected by zero returns leading to a larger number of jumps detected.[5] However, the BNS test is upward biased in the presence of jumps underestimating the jump component, whereas the CPR combines the estimation of the BV and a threshold and hence the number of jumps is greater than when using BNS. The Med test is less influenced by zero returns but it has the

[5]There are 2.89% of zero returns in the sample.

disadvantage of lower efficiency than CPR in the estimation of IV. Given the difference in the results of the jump tests we conduct the rest of the analysis considering all three to check the robustness of the results to the choice of the test.

4 Realised volatility models and estimation results

To start with, we estimate the HAR-RV model, proposed by Corsi (2004), for both the square root and logarithmic transformations:

$$\sqrt{RV_t} = \beta_0 + \beta_1\sqrt{RV_{t-1}} + \beta_2\sqrt{RV_{w,t-1}} + \beta_3\sqrt{RV_{m,t-1}} + a_t,$$
$$\log RV_t = \beta_0 + \beta_1 \log RV_{t-1} + \beta_2 \log RV_{w,t-1} + \beta_3 \log RV_{m,t-1} + a_t,$$

where average RV over the previous week, $RV_{w,t} = \frac{1}{7}\sum_{l=1}^{7} RV_{t-l}$, and over the previous month, $RV_{m,t} = \frac{1}{30}\sum_{l=1}^{30} RV_{t-l}$, are considered.[6]

Once RV is separated into jump and continuous components using the different jump tests, the HAR-CV-JV model, introduced by Chan *et al.* (2008), is also considered for both transformations:

$$\begin{aligned}\sqrt{RV_t} = {} & \lambda_0 + \lambda_1\sqrt{CV_{t-1}} + \lambda_2\sqrt{CV_{w,t-1}} + \lambda_3\sqrt{CV_{m,t-1}} \\ & + \theta_1\sqrt{JV_{t-1}} + \theta_2\sqrt{JV_{w,t-1}} + \theta_3\sqrt{JV_{m,t-1}} + a_t,\end{aligned}$$

$$\begin{aligned}\log RV_t = {} & \lambda_0 + \lambda_1 \log CV_{t-1} + \lambda_2 \log CV_{w,t-1} + \lambda_3 \log CV_{m,t-1} \\ & + \theta_1 \log JV_{t-1} + \theta_2 \log JV_{w,t-1} + \theta_3 \log JV_{m,t-1} + a_t,\end{aligned}$$

where $CV_{w,t}$ and $JV_{w,t}$ are the average CV and JV, respectively, over the previous week and $CV_{m,t-1}$ and $JV_{m,t-1}$ are the average CV and JV, respectively, over the previous month[7].

The HAR-RV and HAR-CV-JV models are estimated using OLS, where heteroscedasticity and autocorrelation-corrected consistent standard errors are considered. Table **??** shows the estimation results of the HAR-RV model. All the coefficients are positive and significant at the 5% level, showing a strong degree of volatility persistence for both non-linear transformations. The estimated coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are quite similar for both

[6]The correlogram of RV suggests a high autocorrelation up to 30 lags and a considerable reduction for more lags.

[7]Regarding non-jump days, when the logarithmic transformation is implemented $\log 0$ is replaced by 0 to maintain the same number of observations.

transformations and it is observed that the estimated effect of $RV_{m,t-1}$ on RV is larger than that of $RV_{w,t-1}$. This is due to the presence of volatility clusters in periods longer than a week.

Table 4: HAR-RV model estimation results

HAR-RV			
		$\sqrt{RV}$	$\log RV$
Intercept	$\hat{\beta}_0$	14.90^a	1.15^a
RV_{t-1}	$\hat{\beta}_1$	0.44^a	0.45^a
$RV_{w,t-1}$	$\hat{\beta}_2$	0.13^b	0.16^a
$RV_{m,t-1}$	$\hat{\beta}_3$	0.27^a	0.26^a
adj-R^2		42.03%	47.27%

a and b denote significance at the 1% and 5% levels, respectively. adj-R^2 is the adjusted R^2.

Table ?? reports the HAR-CV-JV model estimation results for the three jump tests considered (BNS, CPR and Med). The most similar estimated coefficients in terms of significance and magnitude correspond to the cases of the BNS and Med jump tests. In general, there is high volatility persistence in the CV component, and it is slightly higher in magnitude in the logarithmic form due to the greater concavity of the transformation. Again, the estimated effect of the previous month CV on total RV is greater than that of the previous week in most cases analysed. With regard to the jump component, persistence is much lower and the jumps which are most important in explaining RV come, in general, from the previous day. This could be due to the fact that jumps are by nature short-lived, rare occurrences.

Table 5: HAR-CV-JV model estimation results

HAR-CV-JV							
		$\sqrt{RV}$			$\log RV$		
		BNS	CPR	Med	BNS	CPR	Med
Intercept	$\hat{\lambda}_0$	18.77^a	24.26^a	14.93^a	1.45^a	1.89^a	1.11^a
CV_{t-1}	$\hat{\lambda}_1$	0.43^a	0.42^a	0.44^a	0.44^a	0.43^a	0.45^a
$CV_{w,t-1}$	$\hat{\lambda}_2$	0.12^b	0.07	0.13^b	0.13^a	0.12^b	0.16^a
$CV_{m,t-1}$	$\hat{\lambda}_3$	0.22^a	0.12	0.27^a	0.25^a	0.06^b	0.27^a
JV_{t-1}	$\hat{\theta}_1$	0.21^a	0.20^a	0.47^a	0.04^a	0.03^a	0.06^a
$JV_{w,t-1}$	$\hat{\theta}_2$	0.02	0.06	-0.21	$4.1x10^{-3}$	0.02^b	0.01
$JV_{m,t-1}$	$\hat{\theta}_3$	0.18^c	0.11^b	0.07	0.01^c	0.09^a	0.01
adj-R^2		42.13%	42.52%	41.98%	47.45%	47.27%	47.26%

a, b and c denote significance at the 1%, 5% and 10% levels, respectively. adj-R^2 is the adjusted R^2.

The goodness of fit depends on the jump test. For the square root form the decomposition of total RV is preferred when the jumps are detected using BNS and CPR tests, therefore the HAR-CV-JV explains a larger proportion of RV variation. For the logarithmic transformation the explanatory power of the HAR-CV-JV model is greater when using the BNS jump test. There is no improvement in the goodness of fit using the Med test for any transformation. This could be explained by the fact that the number of jump-days detected with the Med is so small that there is no gain in dividing total RV.

The HAR-type models assume identically and independently distributed Gaussian innovations. However, as Corsi *et al.* (2008) point out, in empirical applications the residuals of these models usually show volatility clustering and their probability distributions are characterised by skewness and excess kurtosis. If this is the case there is a loss of efficiency when estimating the models and consequently a decrease in the forecast accuracy, which could be solved by modelling the conditional variance of realised volatility with GARCH specifications.

The error terms in both HAR models, a_t, are analysed in order to determine whether GARCH structures are appropriate for our data. Based on the Ljung-Box statistic, they are justified for the standard deviation transformation. By contrast, the logarithmic transformation of RV provides an unconditional distribution closer to the Gaussian and GARCH models are not justified for modelling the volatility of RV.[8] Out of the various GARCH models analysed in the literature, two different models are specified in this paper for the error term $a_t = \sigma_t \epsilon_t$, where ϵ_t are *iid* random standard normal distributed variables. The first is the GARCH(1,1) model, which in general accounts for the volatility clusters and the main features of the distribution of the innovations. Unlike Corsi *et al.* (2008), GARCH(1,1) structure is considered here for the errors of both the HAR-RV and the HAR-CV-JV models. The second is Nelson's (1991) EGARCH(1,1) model, where asymmetric effects are considered. This structure is widely used in the relevant literature because the incorporation of the leverage effect reduces the skewness in the distribution of the errors and the model fits the characteristics of the data. In particular, Chan *et al.* (2008) model daily volatility of price changes using EGARCH(1,1), but it is not included in the error terms of the HAR model.

The GARCH(1,1) is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 \sigma_{t-1}^2,$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 < 1$ ensure that the variance is positive and the process is stationary.

The asymmetric EGARCH(1,1) is given by:

$$\log \sigma_t^2 = \delta_0 + \delta_1 \frac{|a_{t-1}|}{\sigma_{t-1}} + \delta_2 \log \sigma_{t-1}^2 + \delta_3 \frac{a_{t-1}}{\sigma_{t-1}},$$

where the coefficients are not required to be positive, since even if $\log \sigma_t^2 < 0$, the volatility will always be positive. The asymmetric response of the volatility due to shocks of different

[8] The Ljung-Box test shows evidence of uncorrelated a_t and a_t^2 series for the logarithmic transformation, while there is evidence of uncorrelated errors but correlated squared errors for the standard deviation transformation. Results are available upon request.

signs is captured by δ_3. A positive (negative) coefficient implies an inverse (direct) leverage effect.

The resulting models are called HAR-GARCH-RV, HAR-EGARCH-RV, HAR-GARCH-CV-JV and HAR-EGARCH-CV-JV and are estimated by maximum likelihood.

Tables **??** and **??** show the estimation results for the models considering total RV and decomposition of RV for each jump test considered. The estimated coefficients of GARCH(1,1) ensure that the variance is positive and the time series is stationary. The estimated coefficients of the HAR-EGARCH models show evidence of an inverse leverage effect, meaning that positive shocks have more effect on RV than negative shocks. In general, the magnitude and significance of the coefficients estimated are similar between the models with GARCH and EGARCH innovations although some differences appear in the estimated effect of the JV component depending on the jump test. In terms of log-likelihood the HAR-EGARCH is preferred for both total RV and its decomposition. This means that the leverage effect is important in modelling the standard deviation of RV.

Table 6: HAR-GARCH-RV and HAR-EGARCH-RV models estimation results ($\sqrt{RV}$)

HAR-GARCH-RV			HAR-EGARCH-RV		
Intercept	$\hat{\beta}_0$	14.42^a	Intercept	$\hat{\beta}_0$	15.91^a
RV_{t-1}	$\hat{\beta}_1$	0.45^a	RV_{t-1}	$\hat{\beta}_1$	0.48^a
$RV_{w,t-1}$	$\hat{\beta}_2$	0.19^a	$RV_{w,t-1}$	$\hat{\beta}_2$	0.16^a
$RV_{m,t-1}$	$\hat{\beta}_3$	0.20^a	$RV_{m,t-1}$	$\hat{\beta}_3$	0.19^a
Intercept	$\hat{\alpha}_0$	41.67^a	Intercept	$\hat{\delta}_0$	0.21^a
a^2_{t-1}	$\hat{\alpha}_1$	0.09^a	$\lvert a_{t-1}\rvert/\sigma_{t-1}$	$\hat{\delta}_1$	0.07^a
σ^2_{t-1}	$\hat{\alpha}_2$	0.88^a	$\log \sigma^2_{t-1}$	$\hat{\delta}_2$	0.96^a
			a_{t-1}/σ_{t-1}	$\hat{\delta}_3$	0.14^a
Log-likelihood		-5961.84	Log-likelihood		-5936.62

a denotes significance at the 1% level.

Table 7: HAR-GARCH-CV-JV and HAR-EGARCH-CV-JV models estimation results ($\sqrt{RV}$)

HAR-GARCH-CV-JV					HAR-EGARCH-CV-JV				
		BNS	CPR	Med			BNS	CPR	Med
Intercept	$\hat{\lambda}_0$	19.24^a	22.91^a	14.22^a	Intercept	$\hat{\lambda}_0$	18.13^a	19.10^a	14.55^a
CV_{t-1}	$\hat{\lambda}_1$	0.43^a	0.41^a	0.44^a	CV_{t-1}	$\hat{\lambda}_1$	0.47^a	0.48^a	0.48^a
$CV_{w,t-1}$	$\hat{\lambda}_2$	0.16^a	0.13^b	0.20^a	$CV_{w,t-1}$	$\hat{\lambda}_2$	0.13^a	0.16^a	0.15^a
$CV_{m,t-1}$	$\hat{\lambda}_3$	0.16^a	0.09	0.20^a	$CV_{m,t-1}$	$\hat{\lambda}_3$	0.19^a	0.08	0.21^a
JV_{t-1}	$\hat{\theta}_1$	0.20^a	0.21^a	0.47^a	JV_{t-1}	$\hat{\theta}_1$	0.22^a	0.19^a	0.44^b
$JV_{w,t-1}$	$\hat{\theta}_2$	0.18^a	0.14^b	-0.26	$JV_{w,t-1}$	$\hat{\theta}_2$	0.07	0.03	-0.26
$JV_{m,t-1}$	$\hat{\theta}_3$	0.07	0.19^b	0.17	$JV_{m,t-1}$	$\hat{\theta}_3$	0.08	0.22^a	0.38^b
Intercept	$\hat{\alpha}_0$	61.54^a	54.08^a	44.08^a	Intercept	$\hat{\delta}_0$	0.21^a	0.19^a	0.19^b
a^2_{t-1}	$\hat{\alpha}_1$	0.13^a	0.10^a	0.09^a	$\|a_{t-1}\|/\sigma_{t-1}$	$\hat{\delta}_1$	0.08^a	0.06^a	0.05^a
σ^2_{t-1}	$\hat{\alpha}_2$	0.83^a	0.85^a	0.87^a	$\log\sigma^2_{t-1}$	$\hat{\delta}_2$	0.96^a	0.97^a	0.97^a
					a_{t-1}/σ_{t-1}	$\hat{\delta}_3$	0.13^a	0.13^a	0.14^a
Log-likelihood		-5954.98	-5956.23	-5959.46	Log-likelihood		-5934.41	-5934.07	-5933.14

a and b denote significance at the 1% and 5% levels, respectively.

The standardised error terms and their squares follow a white noise process in all four models.[9] Therefore, the GARCH(1,1) and EGARCH(1,1) correctly fit the data working with the standardised deviation form of RV.

5 Forecast

The predictive power of the different models is measured using various out-of-sample criteria. To that end, observations from November 18, 2012 to December 31, 2015, covering 1109 days, are considered as in-sample data, while observations in 2016, covering 121 days, are considered as out-of-sample data. The forecasts for 2016 are obtained from a recursive estimation of the models starting with the first 1109 observations and expanding the estimation period by adding one new observation each time. Comparisons between observed and predicted values are made using MAE (mean absolute error), RMSE (root mean square error) and MAPE (mean absolute percentage error) criteria[10]:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|RV_t - \widehat{RV_t}|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(RV_t - \widehat{RV_t})^2}$$

[9]The Ljung-Box test is applied to the standardised error terms and their squares. Results are available upon request.

[10]The best model in terms of prediction is the one with the lowest value of any criterion implying that the forecast error is the smallest.

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \frac{|RV_t - \widehat{RV_t}|}{RV_t},$$

where N is the number of forecast observations and $\widehat{RV_t}$ is the predicted value of RV using the models described in Section **??**.

Table **??** reports the results of the MAE, RMSE and MAPE criteria for both transformations of RV and each jump test. For both transformations, models that decompose RV into JV and CV components clearly provide better forecasts than those using total RV. For the standard deviation form, the selection of the best model depends on the jump test. Specifically, the HAR-GARCH-CV-JV model consistently outperforms the other models in terms of forecasting ability for all three criteria under the BNS jump test, followed by the HAR-CV-JV model. By contrast, GARCH structures do not improve forecasts under CPR and Med tests, which select the HAR-CV-JV model. Contrary to the results of the in-sample forecast, there is no gain in adding EGARCH innovations in terms of out-of-sample forecast performance. In the logarithmic transformation, the use of GARCH structures is not justified, so forecasting ability is assessed in the HAR-RV and the HAR-CV-JV models. All three criteria show higher predictive power for the HAR-CV-JV model; specifically, the best results are obtained using BNS jump detection.

Table 8: MAE, RMSE and MAPE criteria.

$\sqrt{RV}$	MAE	RMSE	MAPE
HAR-RV	15.0767	19.5739	0.1784
HAR-CV-JV (BNS)	14.7714	19.2089	0.1752
HAR-CV-JV (CPR)	14.8110	19.4492	0.1756
HAR-CV-JV (Med)	15.0610	19.3712	0.1785
HAR-GARCH-RV	15.0827	19.5883	0.1780
HAR-GARCH-CV-JV (BNS)	**14.6218**	**19.0844**	**0.1733**
HAR-GARCH-CV-JV (CPR)	14.8349	19.5145	0.1757
HAR-GARCH-CV-JV (Med)	15.1113	19.4293	0.1785
HAR-EGARCH-RV	15.2375	19.8298	0.1814
HAR-EGARCH-CV-JV (BNS)	15.0213	19.5349	0.1789
HAR-EGARCH-CV-JV (CPR)	14.9806	19.4672	0.1761
HAR-EGARCH-CV-JV (Med)	15.2884	19.6601	0.1820
$\log RV$	MAE	RMSE	MAPE
HAR-RV	0.3382	0.4189	0.0381
HAR-CV-JV (BNS)	**0.3330**	**0.4127**	**0.0375**
HAR-CV-JV (CPR)	0.3351	0.4179	0.0377
HAR-CV-JV (Med)	0.3370	0.4159	0.0380

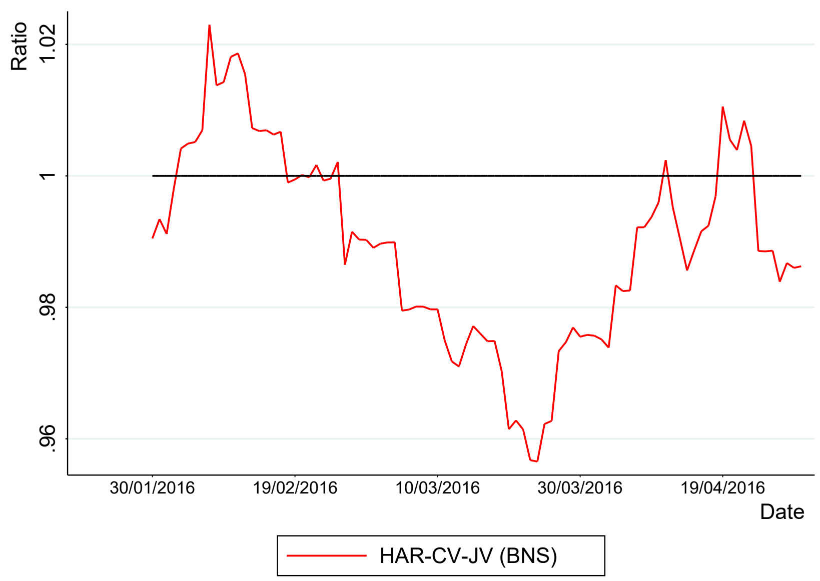
Numbers in bold show the lowest value for each criterion and RV transformation.

As shown, in both transformations the models obtained using the BNS jump test provide more accurate forecasts than the corresponding models using the CPR and Med tests. This could be due to the fact that the CPR test is too greatly affected by zero returns and detects too many jumps, whereas the difference between MedRV and RV is very small, so too few jumps are detected with the Med test.

The criteria used are calculated for the forecast errors of the whole out-of-sample period, but it is also possible to assess the behaviour of errors during that period by computing a rolling ratio of any of the criteria as the ratio of the criterion value for each model and that of the benchmark (HAR-RV) for a moving window of 30 forecasts. A ratio greater than one means that the predictive power of the HAR-RV model is better than that of the other model with which it is compared.

Figures **??** and **??** show the rolling RMSE ratio for the logarithmic and square root transformations of RV, respectively, for the BNS jump test[11]. The former shows that, according to the RMSE criterion, the HAR-CV-JV model in its logarithmic form outperforms the HAR-RV model throughout the forecast period, except in a few observations at the beginning and end of the period. In the case of the standard deviation form, the HAR-CV-JV model is preferred to the HAR-RV model throughout the forecast period. The predictive power of the HAR-GARCH-CV-JV model is higher than that of the benchmark in most of the period, especially at the beginning and end. This result is in line with the selection of the HAR-GACH-CV-JV model according to the forecast criteria (see Table **??**). It is important to underline that the forecasting accuracy of the HAR-GARCH-RV and the HAR-EGARCH-RV models is lower than that of the HAR-RV model in the whole sample. This fact indicates that forecasts are not improved by including GARCH structures when total RV is considered.

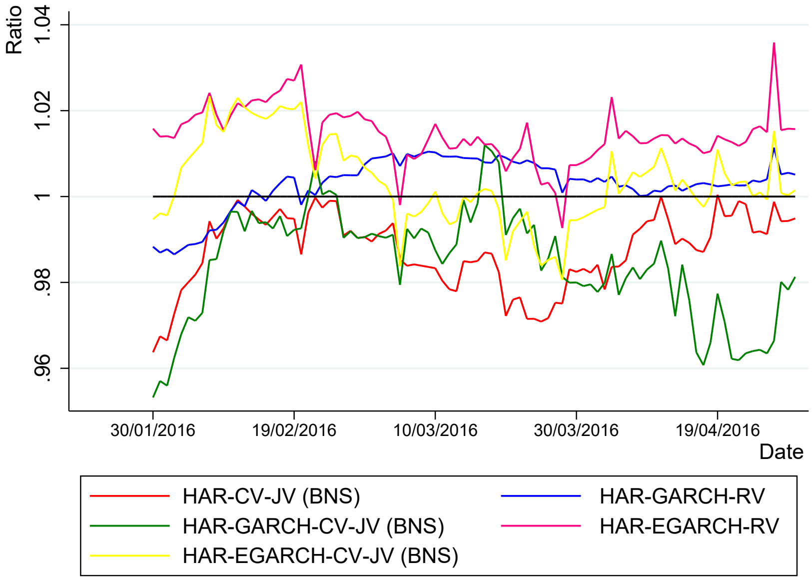
Figure 1: Rolling RMSE ratio (log RV)



Haugom *et al.* (2011) also find that the forecast accuracy is improved when jumps are

[11]Rolling ratios when using the CPR and the Med jump tests have also been calculated. The former provides a pattern similar to that found when using the BNS test while the latter moves around the value of one due to the poor performance in the jump detection. There are no significant differences when rolling ratios for MAE and MAPE criteria are considered. Results are available upon request.

Figure 2: Rolling RMSE ratio ($\sqrt{RV}$)



considered for the Nord Pool forward electricity market and Chan *et al.* (2008) show a slight improvement in volatility forecasts for the NEMMCO decomposing the total RV.

The approach of Andersen *et al.* (2003) is also applied to compare the forecasting performance of different pairs of models. This is an out-of-sample approach based on the regression of the observed RV on the estimated value of RV for the two models to be compared. In the case of the logarithmic transformation of RV there are only two models to be compared: HAR-RV and HAR-CV-JV. For the standard deviation form six different models for each jump test are compared in pairs. However, no conclusive results are obtained[12].

6 Conclusions

The formation of electricity prices has become more complex with the liberalisation of the market. Balancing markets have proved useful in managing real demand and have become more sophisticated as agents have learnt how to trade optimally in this new environment. Thus, there is growing interest in forecasting volatility in electricity prices in these close-to-real-time delivery markets.

In this paper we take realised volatility as a measure of the unobserved volatility process of high-frequency data such as the continuous intraday auction for 15-minute contracts in Germany and Austria. The variation in electricity prices is decomposed into continuous and jump components using three different jump tests. This approach enables us to analyse and compare the forecasting performances of two different models: one for total realised volatility and the other with the aforementioned decomposition. Furthermore,

[12]Results are available from the authors upon request.

GARCH structures are included in the error terms of the models whenever appropriate. The estimation and forecasting power are analysed for the square root and logarithmic transformations of realised volatility.

Out-of-sample forecast criteria support the HAR-CV-JV model for the logarithmic transformation under all three jump tests and for the standardised RV under the CPR and Med jump tests, while the HAR-GARCH-CV-JV model improves the accuracy of volatility forecasting under the BNS jump test and the standard deviation form. A comparison of the different jump tests revels that models using the BNS provide the best results in terms of forecasting ability. These results highlight several conclusions: first, under both transformations the decomposition of total realised volatility is important for forecasting purposes. Second, the choice of the transformation is also important because more complex models including GARCH innovations might be selected for the standardised RV, while the distribution of the innovations in the logarithmic transformation is closer to the normal and simpler models would suffice. Moreover, although EGARCH innovations outperform GARCH innovations in terms of explanatory power in estimation, they are not selected in terms of forecast accuracy. Three, the results are sensitive to which jump test is used at least in the standardised form of RV.

The forecast results obtained are of interest to market participants in the German-Austrian market as they can fit their spot positions close to delivery. Within EPEX, the conclusions can be expected to hold in the Swiss new intraday 15 minute continuous market with similar rules.

Acknowledgements

The authors would like to thank two anonymous reviewers for valuable comments and suggestions that helped to improve the paper. Financial support under research grant ECO2015-64467-R (MINECO/FEDER), from Dpto. de Educación, Universidades e Investigación del Gobierno Vasco under research grant IT-783-13, and from Dpto. de Educación, Política Lingüística y Cultura del Gobierno Vasco through Beca Predoctoral de Formación de Personal Investigador no Doctor is acknowledged.

References

Andersen TG, Bollerslev T, Diebold FX. 2007. Roughing it up: Including jump components in the measurement, modelling and forecasting of return volatility. *Review of Economics and Statistics* **89**: 701-720.

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2003. Modelling and forecasting realized volatility. *Econometrica* **71**: 579-625.

Andersen TG, Dobrev, D, Schaumburg E. 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* **169**: 75-93.

Barndorff-Nielsen O, Shephard N. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* **2**: 1-37.

Barndorff-Nielsen O, Shephard N. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* **4**: 1-60.

Chan KF, Gray P, van Campen B. 2008. A new approach to characterizing and forecasting electricity volatility. *International Journal of Forecasting* **24**: 728-743.

Corsi F. 2004. A simple long memory of realized volatility, Working paper. Switzerland University of Lugano.

Corsi F, Mittnik S, Pigorsch C, Pigorsch U. 2008. The Volatility of Realized Volatility. *Econometric Reviews* **27**: 46-78.

Corsi, F, Pirino, D, Reno, R. 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* **159**: 276-288.

Geman H, Roncoroni A. 2006. Understanding the fine structure of electricity prices. *Journal of Business* **79**: 1225-1261.

EPEX, 2016. European Power Exchange. https://www.epexspot.com/en/market-data/intradaycontinuous/intraday-table/-/DE.

Frömmel M, Han X, Kratochvil S. 2014. Modeling the daily electricity price volatility with realized measures. *Energy Economics* **44**: 492-502.

Haugom E, Ullrich CJ. 2012. Forecasting spot price volatility using the short-term forward curve. *Energy Economics* **34**: 1826-1833.

Haugom E, Westgaard S, Solibakke PB, Lien G. 2011. Realized volatility and the influence of market measures on predictability: Analysis of Nord Pool forward electricity data. *Energy Economics* **33**: 1206-1215.

Huang X, Tauchen G. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* **16**: 99-117.

Müller UA, Dacorogna MM, Olsen RB, Puctet OV, Schwarz M, Morgenegg C. 1990. Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling

Law, and Intraday Analysis. *Journal of Banking and Finance* **14**: 1189-1208.

Nelson DB. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**: 347-370.

Ullrich CJ. 2012. Realized volatility and price spikes in electricity markets: The importance of observation frequency. *Energy Economics* **34**: 1809-1818.