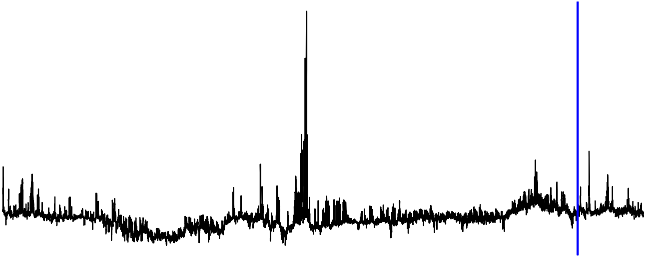
Bayesian Pooling Approach in Forecasting Electricity Prices

Maciej Kostrzewski, Jadwiga Kostrzewska



Cracow University of Economics

(The research is supported by the National Science Centre Poland
through grant 2016/23/B/HS4/03018)

Overview

Motivation

- This research is focused on forecasting hourly day-ahead electricity prices.
- Electricity markets are different from many other commodity and financial markets due to the difficulty in storing large quantities of electricity under the need of keeping stability of the power system (balance between production and consumption).
- The Bayesian approach - promising previous experience:
  - Prices of **gas** forward contracts TTF Month 1 (Netherlands TTF Hub)
  - $\mathbf{CO}_2$ emission allowance futures contracts prices (ICE ECX contracts)
  - Hourly day-ahead locational marginal **electricity** prices for the Jersey Central Power and Light Company of the Pennsylvania–New Jersey–Maryland Interconnection (U.S.)
- Jumps or Stochastic Volatility? Do we have to choose?
- Individual models, the pooling approach or generalised models?
- The criterion: "A model is as good as its predictions"

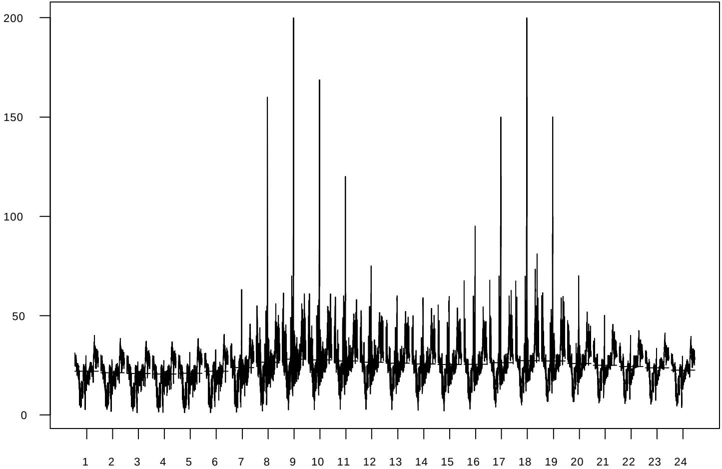
Nord Pool - Data Presentation

- Nord Pool is a leading power market and the largest power market in Europe.
- 380 companies from 20 countries are active at Nord Pool.
- The system is balanced using three different markets: Day-ahead Market (Elspot, volume 98%), Intraday Market (Elbas, volume 1%), Balancing Market (volume 1%).
- The system price (EUR/MWh) is an unconstrained market clearing reference price. It is calculated without any congestion restrictions by setting capacities to infinity. All orders from Nordic (Denmark, Finland, Norway, Sweden) and Baltic regions (Lithuania, Latvia, Estonia) are included in system price calculation.
- Most standard financial contracts traded in the Nordic region use the system price as a reference price.

(Source: NordPoolGroup.com)

The system price at each of 24 hours

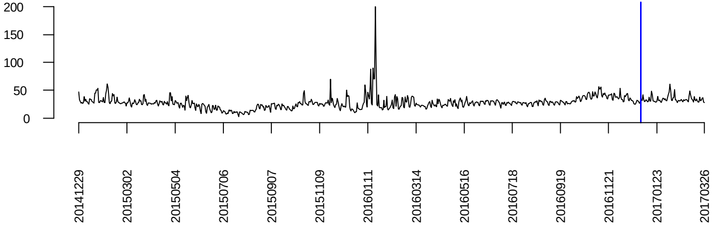
The data set: December 29, 2014 - March 26, 2017

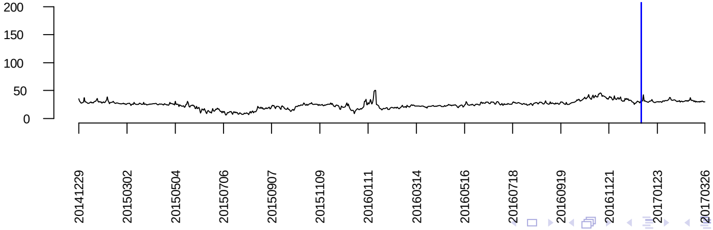


Nord Pool - Data Presentation

- The hourly day-ahead electricity prices
- In-sample data time range: December 29, 2014 - January 1, 2017
  - a number of observations 735 (105 weeks)
- Out–of-sample data time range: January 2, 2017 - March 26, 2017
  - a number of observations 84 (12 weeks)

System price at #9 and #21





Models

- Electricity supply and demand depend on weather conditions: temperature, wind speed, precipitation, solar radiation. Demand is subject to daily, weekly or yearly seasonality due to e.g. an intensity of business activities (working hours, peak hours, weekdays, holidays and near holidays) etc. These features have a significant impact on the market and price behaviour. They might result in extreme spot price volatility and may in consequence bring the existence of sharp price movements, called jumps (or spikes).

- The hallmark of electricity prices is time-variable volatility and jumps.

Models

$\{y_t\}$ - log-prices (EUR/MWh) $t = 1, ..., n$,

$$\left\{\varepsilon_t^{(1)}\right\}, \left\{\varepsilon_t^{(2)}\right\} \sim iidN(0,1),\ \{J_t\} \sim iid,$$

$$\begin{aligned} p_{J_t}(x) = \quad & p_D \frac{1}{\eta_D} \exp\left(\frac{1}{\eta_D} x\right) \mathbb{I}_{(-\infty,0)}(x) + p_0 \delta_{(0)}(x) + \\ & + p_U \frac{1}{\eta_U} \exp\left(-\frac{1}{\eta_U} x\right) \mathbb{I}_{(0,\infty)}(x) \end{aligned}$$

- The **DEJDX** model (Kostrzewski (2015))

$$y_{t+1} = \quad y_t + \psi_1 X_{1,t+1} + ... + \psi_k X_{k,t+1} + \sigma \varepsilon_{t+1} + J_{t+1}$$

- The **SVX** model:

$$\begin{aligned} y_{t+1} = \quad & y_t + \psi_1 X_{1,t+1} + ... + \psi_k X_{k,t+1} + \sqrt{\exp(h_t)} \varepsilon_{t+1}^{(1)} \\ h_{t+1} = \quad & h_t + \kappa_h (\theta_h - h_t) + \sigma_h \left(\rho \varepsilon_{t+1}^{(1)} + \sqrt{1-\rho^2} \varepsilon_{t+1}^{(2)}\right) \end{aligned}$$

- The **SVDEJX** model (Kostrzewski (2016)):

$$\begin{aligned} y_{t+1} = \quad & y_t + \psi_1 X_{1,t+1} + ... + \psi_k X_{k,t+1} + \sqrt{\exp(h_t)} \varepsilon_{t+1}^{(1)} + J_{t+1} \\ h_{t+1} = \quad & h_t + \kappa_h (\theta_h - h_t) + \sigma_h \left(\rho \varepsilon_{t+1}^{(1)} + \sqrt{1-\rho^2} \varepsilon_{t+1}^{(2)}\right) \end{aligned}$$

Models

- **We considered following X variables:**

X_1- **the logarithm of electricity consumption forecast**
X_2- **the minimum of the previous day's hourly log-prices**
X_3- **weekday (a dummy variable)**

- **We considered following models**:
    - The **DEJDX, SVX, SVDEJX** models
    - The SVCJ model (Eraker, Johannes, Polson (2003)) with exogeneous variables

Finally, we apply the DEJDX, SVX, SVDEJX and the pooled model
The models with exogeneous variables, jumps or stochastic volatility

Bayesian Approach

$y = (y_1, ..., y_n)$ the observed data
θ the vector of unknown parameters

$p(\theta)$ the prior density

$p(y|\theta)$ the sampling density

$p(y, \theta) = p(y|\theta)\, p(\theta)$ the Bayesian model

$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$ the posterior density

Why Bayesian ... ?

1. Latent variables
allow formal inference about:
a) jump occurrence times and their sizes
b) stochastic volatility

2. Predictive densities
handle the uncertainty about:
a) unknown parameters
b) stochastic structures

3. Pooling method
handles the uncertainty about:
a) unknown parameters
b) stochastic structures
c) a choice of models

SVDEJX - Bayesian Approach

The vector of unknown parameters:

$$\underbrace{(\kappa_h, \theta_h, \sigma_h, \rho, \eta_D, \eta_U, p_D, p_0, p_U, \psi_1, ..., \psi_k,}_{=\theta}$$

$$\underbrace{h_{t_1}, ..., h_{t_n}}_{=h}, \underbrace{q_{t_1}, ..., q_{t_n}}_{=q}, \underbrace{\xi^D_{t_1}, \xi^U_{t_1} ..., \xi^D_{t_n}, \xi^U_{t_n})}_{=\xi}$$

Parametrisation: $(\sigma_h, \rho) \to (\phi_h, \omega_h)$; $\phi_h = \sigma_h \rho$, $\omega_h = \sigma_h^2 \left(1 - \rho^2\right)$

The prior structure:

$\kappa_h \sim N(1, 6)\, I_{(0,2)}$, $\theta_h \sim N(0, 10)$,

$\omega_h \sim IG\left(3, \frac{1}{20}\right)$, $\phi_h \,|\omega_h \sim N\left(0, \frac{1}{2}\omega_h\right)$,

$p(\eta_D) \sim IG(1.86, 0.43)$, $p(\eta_U) \sim IG(1.86, 0.43)$,

$\psi_i \sim N(0, 1)$, $(p_D, p_0, p_U) \sim Dirichlet\,(1, 1, 1)$

Latent variables: $p\left(q_{t_i} \,|\theta\right) = \begin{cases} p_D & \text{on } q_{t_i} = -1 \\ p_0 & \text{on } q_{t_i} = 0 \\ p_U & \text{on } q_{t_i} = 1 \end{cases}$,

$\xi^D_{t_i} \,|\theta \sim Exp(\eta_D)$, $\xi^U_{t_i} \,|\theta \sim Exp(\eta_U)$

Values of jumps: $J_{t_i} = -\xi^D_{t_i} \cdot \mathbb{I}\left(q_{t_i} = -1\right) + \xi^U_{t_i} \cdot \mathbb{I}\left(q_{t_i} = 1\right)$

The Bayesian model: $p(y, \theta, h, q, \xi) = p(y\,|\theta, h, q, \xi)\, p(\theta, h, q, \xi)$

Predictive Bayes Factors (Geweke, Amisano (2010))

- The **predictive density** for y_t and a model M:

$p\left(y_t | y_{t-1}, ..., y_1, M\right) = \int_{\Theta} p\left(y_t | y_{t-1}, ..., y_1, \theta, M\right) p\left(\theta | y_{t-1}, ..., y_1, M\right) d\theta$

- The one-step-ahead **predictive likelihood**

$PL_M(t) = p\left(y_t | y_{t-1}, ..., y_1, M\right)$, where $y_t, y_{t-1}, ..., y_1$ are known

- The **predictive Bayes factor** for two competing models M_1 and M_2 :

$\frac{PL_{M_1}(t)}{PL_{M_2}(t)}$ (in favor of M_1 over M_2 for observation t).

- Let us regard $p\left(\theta | y_S, ..., y_1, M\right)$ as the prior distribution for θ and $y_S, ..., y_1$ as the training sample, then the **log predictive Bayes factor**:
  $\ln\left(\frac{p(y_T, ..., y_1 | y_S, ..., y_1, M_1)}{p(y_T, ..., y_1 | y_S, ..., y_1, M_2)}\right) = \sum_{t=S+1}^{T} \ln\left(\frac{PL_{M_1}(t)}{PL_{M_2}(t)}\right)$
- The **cumulative log predictive Bayes factor**
  $r \longmapsto \ln\left(\frac{p(y_{S+r}, ..., y_1 | y_S, ..., y_1, M_1)}{p(y_{S+r}, ..., y_1 | y_S, ..., y_1, M_2)}\right)$.
- **Weights** for the pooling method: $P\left(M_1 | y_{S+1}; y_S, ..., y_1\right)$,
  $P\left(M_2 | y_{S+1}; y_S, ..., y_1\right)$

Empirical Results

The MCMC methods

Combining the Gibbs sampler and the Metropolis-Hastings algorithm facilitates the Bayesian inference under the models. These techniques are numerical methods of multidimensional integration, which belong to the Markov Chain Monte Carlo (MCMC) methods.

- The results are presented for the DEJDX, SVX, the pooling of DEJDX and SVX model (called the pooled model) and SVDEJX model

Prediction Intervals

- **Bayes_Q** (B_Q): Quantiles of predictive distribution e.g $(Q_{0.25}, Q_{0.75})$
- **Bayes_HPD** (B_HPD): The Highest Predictive Density Interval

Empirical Results

- The hourly day-ahead electricity prices
- In-sample data time range: December 29, 2014 - January 1, 2017
  - a number of observations 735 (105 weeks)

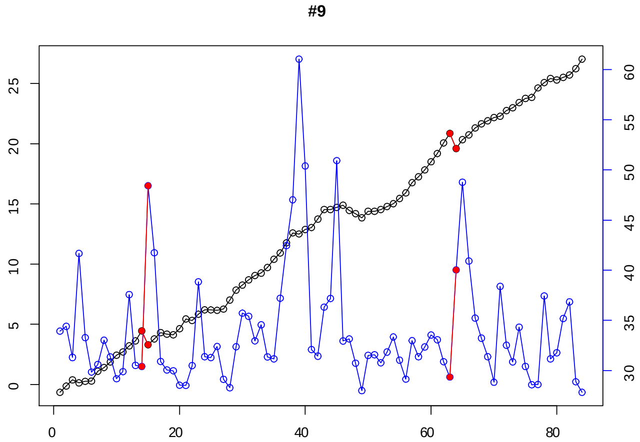
One day-ahead predictions of electricity spot prices:

- Out–of-sample data time range: January 2, 2017 - March 26, 2017
  - a number of observations 84 (12 weeks)

The first day-ahead forecast is calculated under the model estimated over the in-sample data set. The estimation is based on 200,000 MCMC draws, preceded by 1,000,000 burn-in cycles. Subsequent re-estimations and forecasts are based on a data window expanded by a consecutive observation. The re-estimations of the model are based on 10,000 MCMC draws, preceded by 20,000 burn-in cycles.

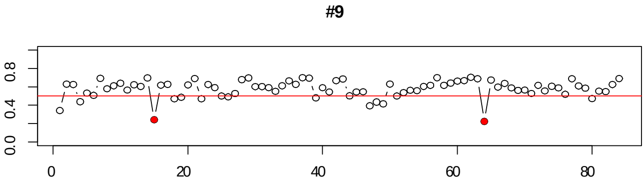
The cumulative Bayes factors - SVX over DEJDX

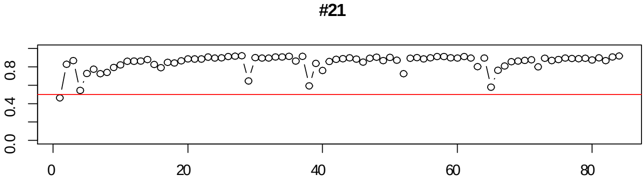
Out–of-sample data time range: January 2, 2017 - March 26, 2017 (84days)



The cumulative log predictive Bayes factors (black), the prices (navy)

The Weights - SVX over DEJDX





- Jumps or Stochastic Volatility?

For most hours the predictive Bayes factors favour the SVX models.

Unconditional coverage of the 50% and 90% day-ahead PIs by the actual spot price

PI(%)	B_Q			B_HPD		
	DEJDX	SVX	Pooling	DEJDX	SVX	Pooling
Unconditional Coverage						
50	17.51	53.37	48.66	17.71	52.38	47.07
90	89.73	91.52	89.98	89.83	91.52	89.63
Mean (standard deviation) of the PI width						
50	1.09	3.4	2.78	1.05	3.36	2.74
	(1.04)	(2.39)	(1.42)	(1.02)	(2.32)	(1.37)
90	12.73	10.05	8.54	12.19	9.91	8.38
	(6.15)	(6.85)	(4.46)	(5.79)	(6.55)	(4.24)
Median (inter-quantile range) of the PI width						
50	0.71	2.74	2.46	0.68	2.72	2.45
	(0.55)	(2.17)	(1.64)	(0.52)	(2.15)	(1.62)
90	13	8.33	7.58	12.52	8.27	7.51
	(6.09)	(6.44)	(5.18)	(5.6)	(6.33)	(5.02)

The Christoffersen tests (Christoffersen, 1998)

- The unconditional coverage test and conditional coverage test (the joint test of unconditional coverage and independence)
- $\alpha = 0.05$
- The tests are conducted separately for each of the 24 hours.

The number of times (hours) the test doesn't reject the null hypothesis:

PI(%)	B_Q			B_HPD		
	DEJDX	SVX	Pooling	DEJDX	SVX	Pooling
Unconditional Coverage						
50	0	20	**23**	0	21	**22**
90	13	20	**21**	15	20	**23**
Conditional Coverage						
50	0	10	**16**	0	11	**15**
90	4	7	**12**	4	8	**12**

The Christoffersen tests (Christoffersen, 1998)

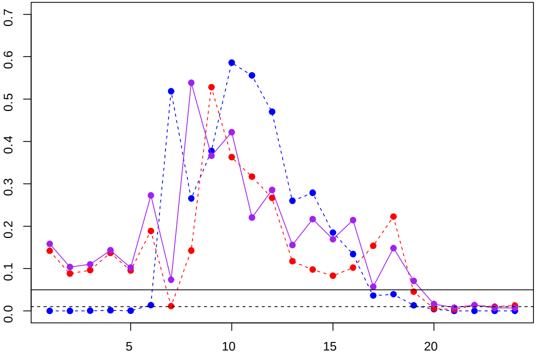
- The unconditional coverage test and conditional coverage test (the joint test of unconditional coverage and independence)
- $\alpha = 0.01$
- The tests are conducted separately for each of the 24 hours.

The number of times (hours) the test doesn't reject the null hypothesis:

PI(%)	B_Q			B_HPD		
	DEJDX	SVX	Pooling	DEJDX	SVX	Pooling
Unconditional Coverage						
50	0	**24**	**24**	1	**24**	**24**
90	15	**23**	**23**	16	23	**24**
Conditional Coverage						
50	0	18	**21**	0	**20**	19
90	8	**15**	14	8	15	**16**

The Berkowitz Test (Berkowitz (2001))

- The test for the goodness-of-fit of the predictive distributions



The p-values for the DEJDX (navy), SVX (red), Pooled (purple) models. The dashed and solid horizontal black lines represent 1% and 5% significance levels, respectively.

The pinball loss function

(Maciejowska, Nowotarski 2016, Nowotarski, Weron 2018)

- The pinball loss function is a special case of an asymmetric piecewise linear loss function

$$Pinball_{h,t}(q) = \begin{cases} (1-q)\left(Q_{h,t}(q) - P_{h,t}\right), & P_{h,t} < Q_{h,t}(q) \\ q\left(P_{h,t} - Q_{h,t}(q)\right), & P_{h,t} \geq Q_{h,t}(q) \end{cases}$$

$P_{h,t}$ is an electricity spot price, h is an hour, t is a day index ($t = 1, ..., 84$)
$Q_{h,t}(q)$ is the q-th quantile of the predictive distribution ($q = 0.01, ..., 0.99$)

- The pinball loss is a measure of fit for a predictive quantile.
- The final score for a day t is an average of the pinball loss across 24 hours and 99 quantiles.
- The lower score indicates the better probabilistic forecast.
- The number of days with the lowest value of the final score:

DEJDX	SVX	Pooling
17	25	**42**
—	27	**57**

- Individual models or the pooling approach? The pooling approach!

The pooling approach or generalised model?

The SVDEJX model is a generalised structure of DEJDX and SVX.

$$\begin{aligned} y_{t+1} = & \quad y_t + \psi_1 X_{1,t+1} + ... + \psi_k X_{k,t+1} + \sqrt{\exp\left(h_t\right)} \varepsilon_{t+1}^{(1)} + J_{t+1}, \\ h_{t+1} = & \quad h_t + \kappa_h \left(\theta_h - h_t\right) + \sigma_h \left(\rho \varepsilon_{t+1}^{(1)} + \sqrt{1-\rho^2} \varepsilon_{t+1}^{(2)}\right), \end{aligned}$$

- The criterion: “A model is as good as its predictions”

The Christoffersen tests (Christoffersen, 1998)

- The unconditional coverage test and conditional coverage test (the joint test of unconditional coverage and independence)
- $\alpha = 0.05$
- The tests are conducted separately for each of the 24 hour.

The number of times (hours) the test doesn't reject the null hypothesis:

PI(%)	B_Q		B_HPD	
	SVDEJX	Pooling	SVDEJX	Pooling
Unconditional Coverage				
50	21	**23**	**22**	**22**
90	18	**21**	18	**23**
Conditional Coverage				
50	13	**16**	14	**15**
90	7	**12**	7	**12**

The Christoffersen tests (Christoffersen, 1998)

- The unconditional coverage test and conditional coverage test (the joint test of unconditional coverage and independence)
- $\alpha = 0.01$
- The tests are conducted separately for each of the 24 hour.

The number of times (hours) the test doesn't reject the null hypothesis:

PI(%)	B_Q		B_HPD	
	SVDEJX	Pooling	SVDEJX	Pooling
Unconditional Coverage				
50	24	**24**	**24**	**24**
90	21	**23**	23	**24**
Conditional Coverage				
50	18	**21**	19	**19**
90	14	**14**	14	**16**

Hour	The Berkowitz Test	
	Pooling	SVDEJX
1	1	1
2	1	0
3	1	0
4	1	0
5	1	0
6	1	1
7	1	0
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0

- The test of goodness-of-fit of the predictive distributions
- Zero if the test rejects the null hypothesis
- One if the test doesn't reject
- $\alpha = 0.05$
- The more green the better

Hour	The Berkowitz Test	
	Pooling	SVDEJX
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	0
21	0	0
22	1	1
23	0	0
24	0	0

- The test of goodness-of-fit of the predictive distributions
- Zero if the test rejects the null hypothesis
- One if the test doesn't reject
- $\alpha = 0.01$
- The more green the better

The pinball loss function

- The lower score indicates the better probabilistic forecast.
- The number of days with the lowest value of the final score:

DEJDX	SVX	Pooling	SVDEJX
17	25	**42**	—
—	27	**57**	—
—	—	**54**	30

Point Forecast

Hour	MAPE (%)	
	Pooling_Medians	SVDEJX_Medians
1	4.8	**4.78**
2	5.17	**5.11**
3	5.25	**5.21**
4	5.17	**5.12**
5	**4.68**	**4.68**
6	4.47	**4.46**
7	**5.11**	5.3
8	**9.16**	9.35
9	11.5	**11.1**
10	10.24	**9.95**
11	9.1	**8.87**
12	7.53	**7.39**
13	6.87	**6.76**
14	6.57	**6.49**
15	6.48	**6.46**
16	**6.14**	**6.14**
17	**7.42**	7.43
18	9.38	**9.34**
19	**8.17**	8.22
20	**5.28**	5.31
21	**3.93**	3.96
22	**3.65**	**3.65**
23	3.58	**3.55**
24	4.26	**4.25**

- Point forecasts equal to medians of predictive distributions
- The mean absolute percentage error
- The lower MAPE the better point forecasts

Conclusions and Remarks

1. Jumps or stochastic volatility? Stochastic volatility
2. We do not have to choose between jumps and SV. We can use the pooling approach or SVDEJX model.
3. The pooling approach gives better results than individual models.
4. The pooling approach or generalised model?
    1. the probabilistic forecasts $\Longrightarrow$ the pooling approach
    2. the point forecasts $\Longrightarrow$ The SVDEJX model
5. The quantiles prediction intervals (Bayes_Q) and the HPD prediction intervals (Bayes_HPD) give similar results.

Thank you very much for your attention!

kostrzem@uek.krakow.pl

1. Christoffersen, P. F. (1998), Evaluating Interval Forecasts, International Economic Review, 39(4), 841–862.
2. Eraker B, Johannes M., Polson N. (2003), The Impact of Jumps in Volatility and Returns, The Journal of Finance, Vol. LVIII, No. 3, pp. 1269-1300
3. Kostrzewski M. (2015), Bayesian DEJD Model and Detection of Asymmetry in Jump Sizes, Central European Journal of Economic Modelling and Econometrics, 7, 43–70.
4. Kostrzewski M. (2016), Bayesian SVLEDEJ Model for Detecting Jumps in Logarithmic Growth Rates of One Month Forward Gas Contract Prices, Central European Journal of Economic Modelling and Econometrics, 8(3).
5. Geweke J., Amisano G. (2010), Comparing and Evaluating Bayesian Predictive distributions of Asset Returns, International Journal of Forecasting, Vol. 26, Issue 2, pp. 216-230
6. Szerszen, P. (2009), Bayesian Analysis of Stochastic Volatility Models with Levy Jumps: Application to Risk Analysis, Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C, (40), 1–52.
7. Maciejowska, Nowotarski (2016), A hybrid model for GEFCom2014 probabilistic electricity price forecasting, International Journal of Forecasting, Vol. 32, Issue 3, pp.1051-1056
8. Nowotarski J., Weron R. (2018), Recent advances in electricity price forecasting: A review of probabilistic forecasting, Renewable and Sustainable Energy Reviews