Faculty of Computer Science
Master of Data Science

Moscow
2024

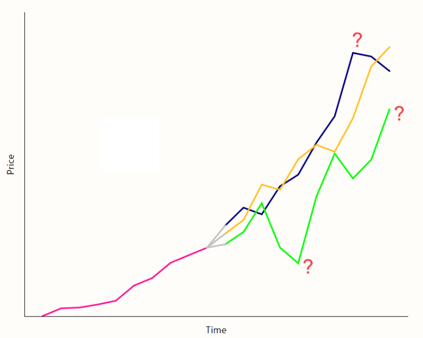
Electricity Spot Prices Forecasting Using Stochastic Volatility Models

Student: Batyrov Andrei Renatovich
Supervisor: Kasianova Ksenia Alekseevna

Table of Contents

Electricity prices forecasting is important

- Electricity is essential commodity, no doubt
- Electricity demand depends on individual and business activities, climate, weather and other factors
- Electricity is non-storable – subject to sudden and large price change (shocks)



Research goal – generate probabilistic forecasts of day-ahead electricity prices in a spot marker for European and Siberian price zones for Peak and Off-peak hours, employing stochastic volatility models

Approaches to Forecasting

Various approaches

- Multi-agent
- Fundamental
- Reduced-form
- **Statistical**
- Computational intelligence
- Hybrid solutions

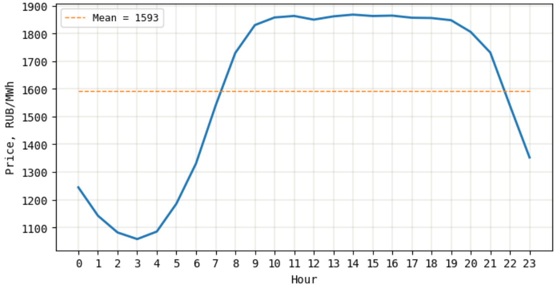
Statistical approaches

- Similar-day and exponential smoothing
- Regression (AR, ARMA, etc.)
- **Heteroscedastic (non-constant volatility)**
  1. (Generalized) Autoregressive Conditional Heteroscedasticity ((G)ARCH)
  2. **Stochastic Volatility (SV)**

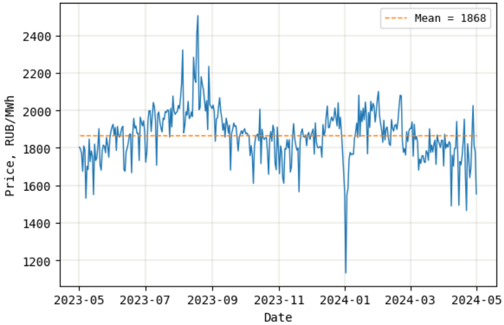
> *...stochastic volatility models almost always outperform their GARCH counterparts, suggesting that stochastic volatility models might provide a better alternative to the more conventional GARCH models. [1]*

Consumer prices

Modeling time frame: 01.05.2023 – 30.04.2024, one year back



Mean hourly profile for European price zone



Trace plot for peak hour #14

Data: Daily Indices and Volumes of The Day-ahead Market [2]

Stationarity

Augmented Dickey-Fuller test

$\mathcal{H}_0$: Unit root (Non-stationary)
$\mathcal{H}_1$: Stationary

Hour	Statistic	p-value	Result at $\alpha = 0.05$
Peak #14	-2.56	0.10	Non-stationary
Off-peak #3	-2.71	0.07	Non-stationary

Consumer price is a non-stationary process – variance of price (volatility) can be modeled as a stochastic process.

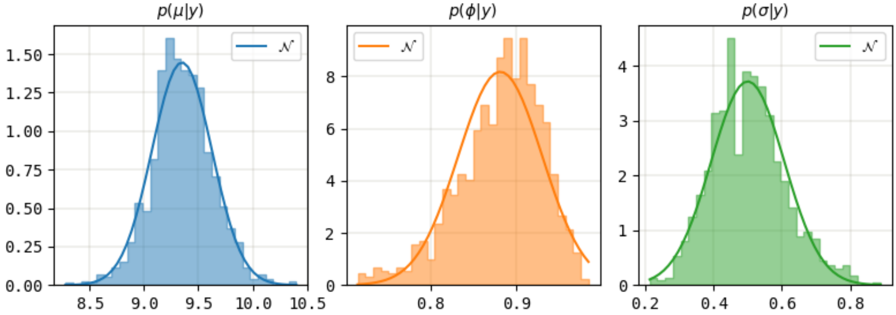
Consumer price at time t

$$y_t \sim \mathcal{N}\left(\bar{y}, e^{h_t/2}\right). \tag{1}$$

- $\bar{y}$, mean price
- h_t, latent log volatility at time t; $h_t = \mu + \phi(h_{t-1} - \mu) + \delta_t \sigma$; $h_1 \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{1-\phi^2}}\right)$; $\delta_t \sim \mathcal{N}(0, 1)$
- μ, mean log volatility
- ϕ, persistence of volatility
- σ, white noise shock scale

Parameters estimation results

- Stan – proprietary probabilistic language with Bayesian inference and prediction

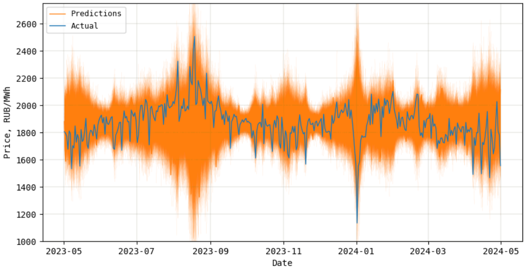


Posterior distributions of estimated parameters

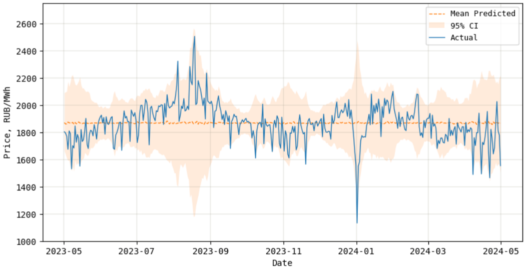
SV Baseline Model

Electricity Spot Prices Forecasting Using Stochastic Volatility Models

In-sample predictions

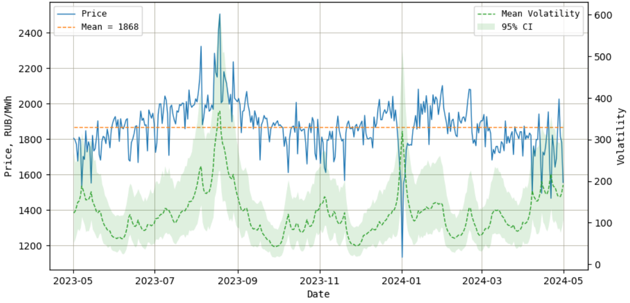


Actual price and 1,000 predictions

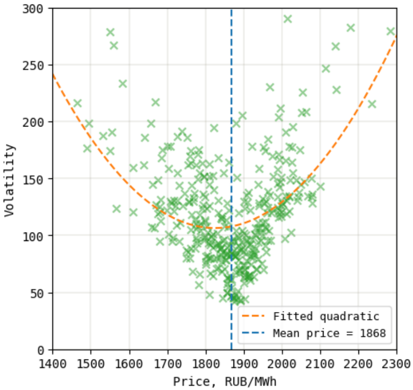


95% CI for 1,000 predictions

Volatility



Consumer prices and learned volatility

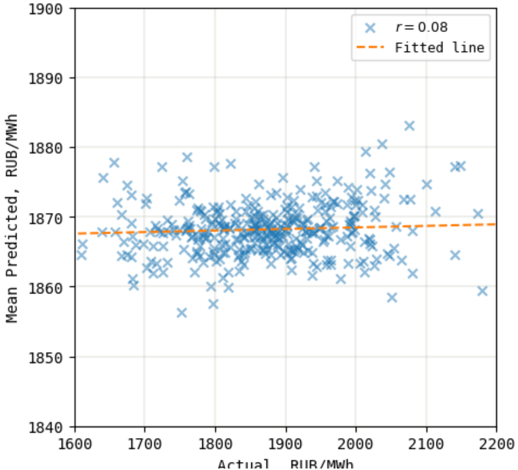


Volatility vs Price

SV Baseline Model

Electricity Spot Prices Forecasting Using Stochastic Volatility Models

Goodness-of-fit



MAE and RMSE metrics for predicted price

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|. \qquad (2)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}. \qquad (3)$$

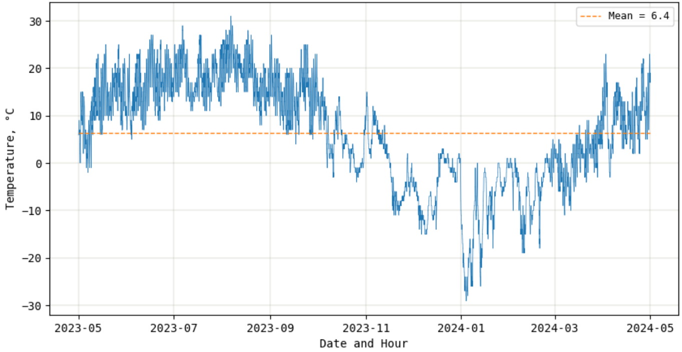
Metric	SV Baseline
MAE	99.18
RMSE	137.04

What about other factors?

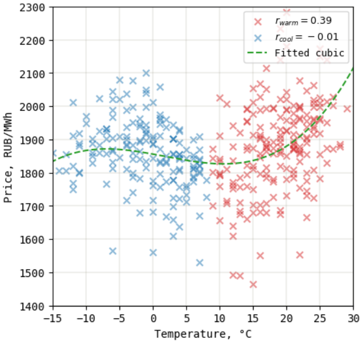
Hypotheses:

- High and low air temperatures drive demand and thus price
- Hourly demand profiles differ during the week
- Autoregression – use yesterday's information

Air temperature



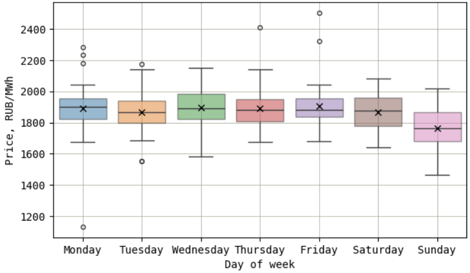
Air temperature in Moscow



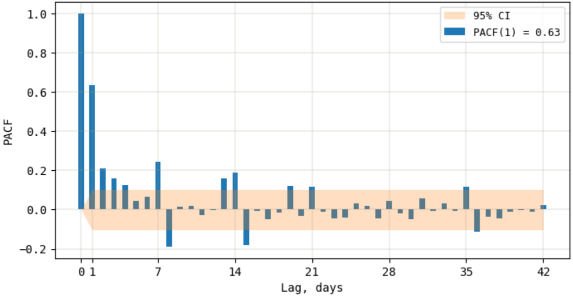
Price vs Temperature

Data: Reliable Prognosis [3]

Day of the week & Autoregression



Price vs day of the week



PACF

Consumer price at time t

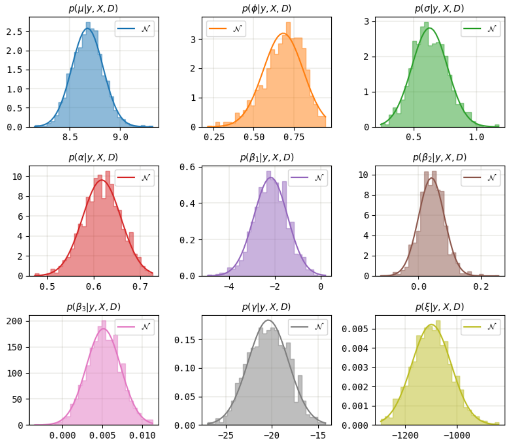
$$y_t \sim \mathcal{N}\left(\bar{y} + \alpha y_{t-1} + \beta_3 X_{t-1}^3 + \beta_2 X_{t-1}^2 + \beta_1 X_{t-1} + \gamma D_t + \xi, e^{h_t/2}\right). \qquad (4)$$

X_{t-1} – hourly air temperature at time $t-1$, to prevent target leakage, lagged one day behind $(t-1)$; D_t – day of the week at time t.

6 new model parameters for 3 introduced exogenous regressors:

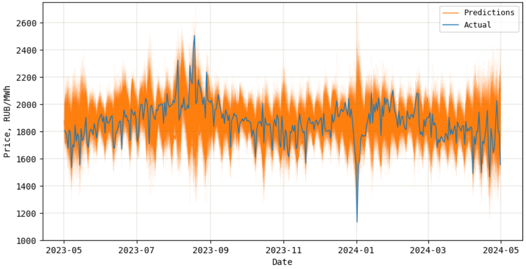
- α, autoregressive component
- $\beta_{i=1\ldots3}$, air temperature regressor
- γ, day of the week regressor
- ξ, constant term (intercept) for all exogenous regressors

Parameters estimation results

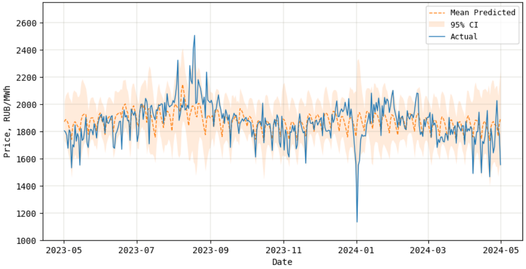


Posterior distributions of estimated parameters

In-sample predictions



Actual price and 1,000 predictions

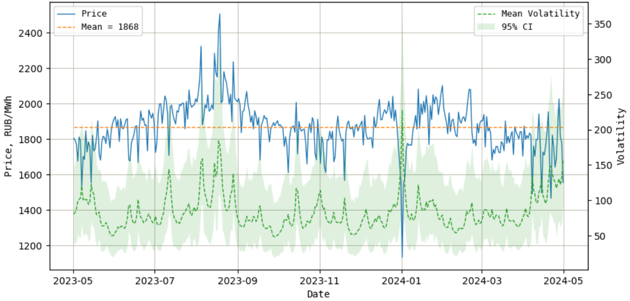


95% CI for 1,000 predictions

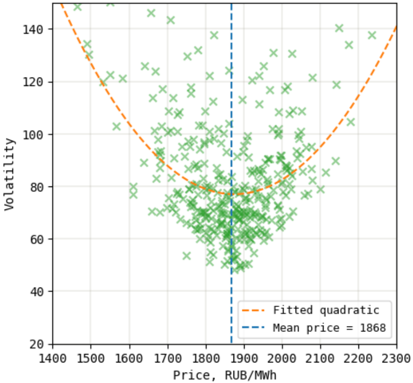
SV X Model

Electricity Spot Prices Forecasting Using Stochastic Volatility Models

Volatility

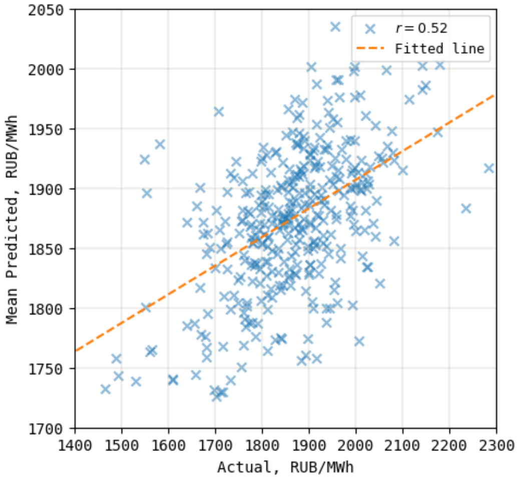


Consumer prices and learned volatility



Volatility vs Price

Goodness-of-fit



MAE and RMSE metrics for predicted price

Metric	SV Baseline	SV X
MAE	99.18	**85.23** (-14.06%)
RMSE	137.04	**117.91** (-13.96%)

SV X model is a better fit

Model combinations

Combination #	Model family	Hour	Price zone
1	SV Baseline	Peak	1 (European)
2		Peak	2 (Siberian)
3		Off-peak	1
4		Off-peak	2
5	SV X	Peak	1
6		Peak	2
7		Off-peak	1
8		Off-peak	2

- Time frame: 23.06.2014 – 30.04.2024 (3600 days), approx. ten years back
- Number of date sliding windows $N_w = \frac{3600 - 30 \times 12}{30 \times 3} = 36$

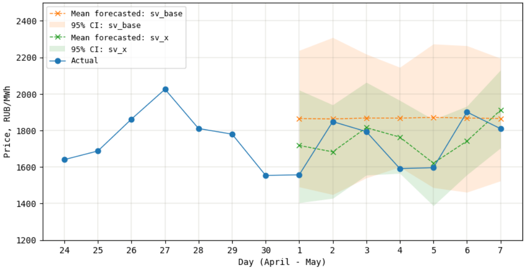
Summary of metrics for all model combinations

Metric	Hour	Price zone	SV Baseline	SV X
MAE	Peak	1	105.39	99.45
	Peak	2	119.14	115.56
	Off-peak	1	104.42	93.92
	Off-peak	2	134.86	118.18
		Average	115.95	**106.78** (-7.91%)
RMSE	Peak	1	130.53	123.59
	Peak	2	146.45	142.10
	Off-peak	1	137.50	129.59
	Off-peak	2	168.31	150.73
		Average	145.69	**136.50** (-6.31%)

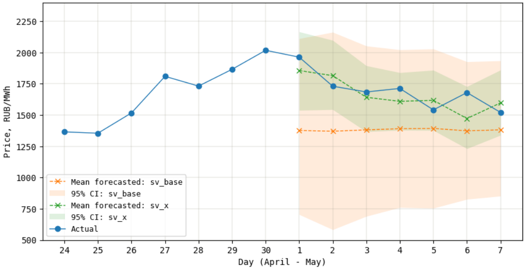
Setup

- 8 model combinations, same as for cross-validation
- First train time frame: 01.05.2023 – 30.04.2024, same as for modeling
- Forecast time frame: 01.05.2024 – 07.05.2024, 1 week ahead

Result plots: Peak hour

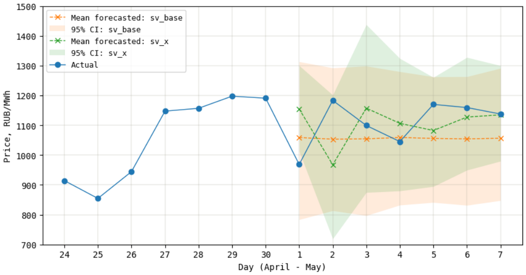


Price zone 1

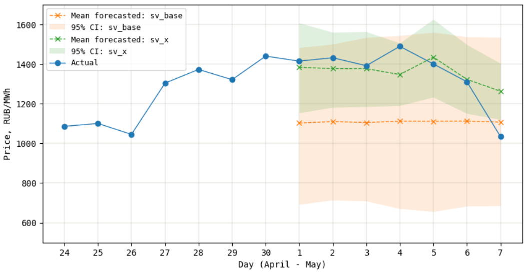


Price zone 2

Result plots: Off-peak hour



Price zone 1



Price zone 2

Summary of metrics for all model combinations

Metric	Hour	Price zone	SV Baseline	SV X
MAE	Peak	1	148.06	114.95
	Peak	2	309.28	99.56
	Off-peak	1	82.84	91.94
	Off-peak	2	265.35	73.97
		Average	201.38	**95.11** (-52.77%)
RMSE	Peak	1	191.49	130.60
	Peak	2	338.86	110.59
	Off-peak	1	91.02	117.76
	Off-peak	2	281.16	105.44
		Average	225.63	**116.10** (-48.54%)

Main results

Build and tested 8 model combinations

Cross-validation best performer

	European	Siberian
Peak	SV X	SV X
Off-peak	SV X	SV X

Forecasting best performer

	European	Siberian
Peak	SV X	SV X
Off-peak	SV Baseline	SV X

- Results confirm the applicability and robustness of the enhanced SV X model
- This model may be used in financial derivative instruments for hedging the risk associated with electricity trading

Conclusion

Future work

- Build models tailored for each hour for both price zones – 48/96 models
- Introduce other climate and weather factors driving the electricity demand: humidity, precipitation, solar irradiance, wind speed, etc.
- Develop custom library to increase computation speed thus allowing to build more complex models and validate them more rigorously and promptly in production-like scenarios [4], [5]

"Prediction is difficult – particularly when it involves the future." – Mark Twain

Thank you!

References

[1] Joshua C.C. Chan, Angelia L. Grant, "Modeling Energy Price Dynamics: GARCH versus Stochastic Volatility," *Energy Economics (54)*, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0140988315003539

[2] Administrator of Trading System (ATS). Daily indices and volumes of the day-ahead market. [Online]. Available: https://www.atsenergo.ru/results/rsv/index

[3] Raspisaniye Pogodi Ltd. Reliable Prognosis. [Online]. Available: https://rp5.ru/

[4] Andrei Batyrov. (2024) Master's Thesis Materials. GitHub Repository. National Research University Higher School of Economics. Faculty of Computer Science. [Online]. Available: https://github.com/andrewha/mds2022/tree/main/Thesis

[5] ——, "Sometimes you really want to re-invent the wheel: EasyML," *Medium.com*, 2024. [Online]. Available: https://medium.com/@handrewkha/sometimes-you-really-want-to-re-invent-the-wheel-easyml-8fa3ab99aed5