1. Motivation

Since the 1990s, a liberalization of the electric market has emerged in European Union's countries in order to improve efficiency as well as reduce electricity prices. Following the accession into the European community, Hungary started to privatize its energy companies in the second half of the 1990s (Tamás Szőke et al., 2021). The Hungarian electricity market was liberalized in 2003. One of the critical stylized facts on the electric market is that electricity prices will be determined based on the contracts on regular markets. Moreover, the relation between supply and demand plays an essential role in pricing electricity (Massimiliano Serati et al., 2008).

Forecasting electricity spot prices is of central interest to most researchers and practitioners in the electricity field. The study of electricity spot price is not only crucial for profitability analysis and power planning in the long run but also useful for pricing derivative contracts in the medium run. There is a wide range of models that are applied to predict electricity prices. In this project, I extend the "Expert" model by adding relevant dependent variables in order to get an outperforming model.

2. Exploratory Data Analysis:

2.1 Correlation structure of electricity prices

2.1.1 ACF and PACF

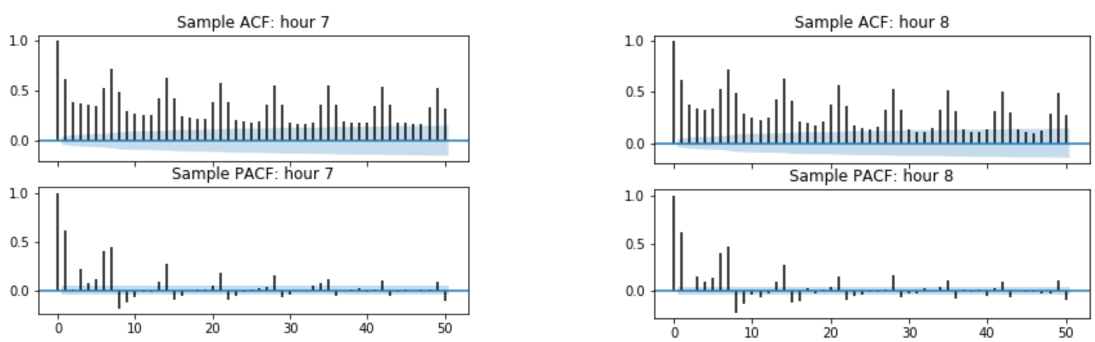


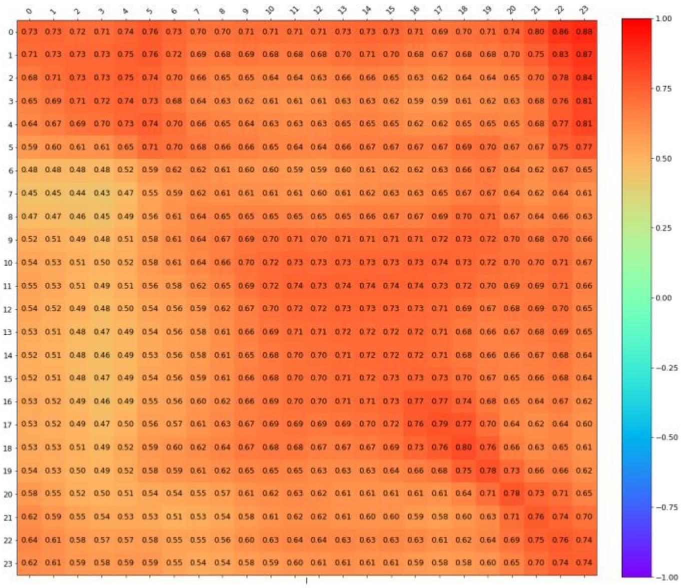
Figure 1: ACF and PACF for Hungary's electricity prices

In addition to autoregressive lags 1, 2 and 7 in Expert model, k = 14 and k = 21 autoregressive lags have significant difference from 0 PACF, especially for 7th and 8th hour. However, the PACF interpretation would be valid when the process of generating data is stationary. In this study, the electricity price of Hungary is likely to be not stationary because of the high dependency on weather conditions, which could be explained by seasonal factors. Nevertheless, it is worth including the larger lags in my modified models. In the end, the evaluation method will help me to decide whether these two lags should stay in the final model.

2.1.2 Cross-period dependencies

It can be seen from Figure 2.a that the diagonal values are significantly large for all hours (s = 0 to s = 23). It is in line with Expert model where the price $Y_{d,s}$ depends on the previous day's price $Y_{d-1,s}$. Furthermore, the price at the last hour of the previous day correlates dramatically with the current day's prices, especially for the first five hours (from 0 am to 5 am).

In Figure 2.b, the autocorrelations are computed conditioning on $Y_{d-1,23}$ and $Y_{d-1,s}$. It is obvious that all autocorrelations tend to be captured. Hence, I will include the $Y_{d-1,23}$ into my modified models.



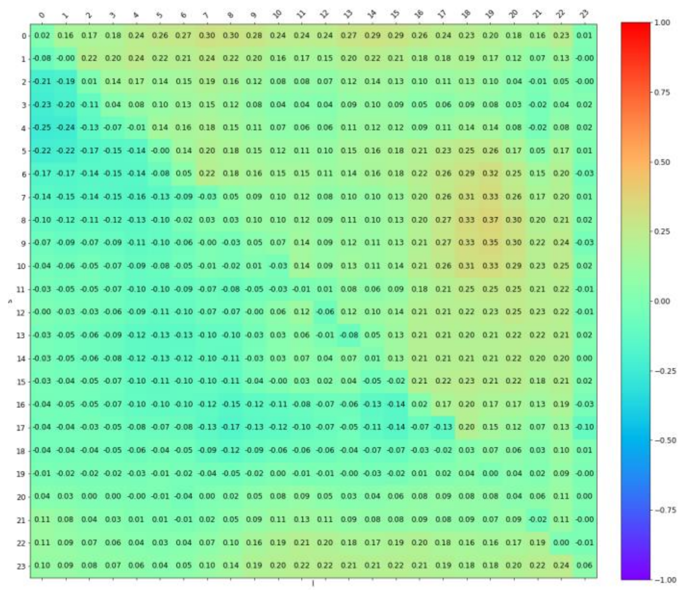


Figure 2.a: Cross-period autocorrelations $Cor(Y_{d,s},Y_{d-1,l})$

Figure 2.b: Cross-period autocorrelations $Cor(Y_{d,s},Y_{d-1,l} \mid Y_{d-1,s}, Y_{d-1,23})$

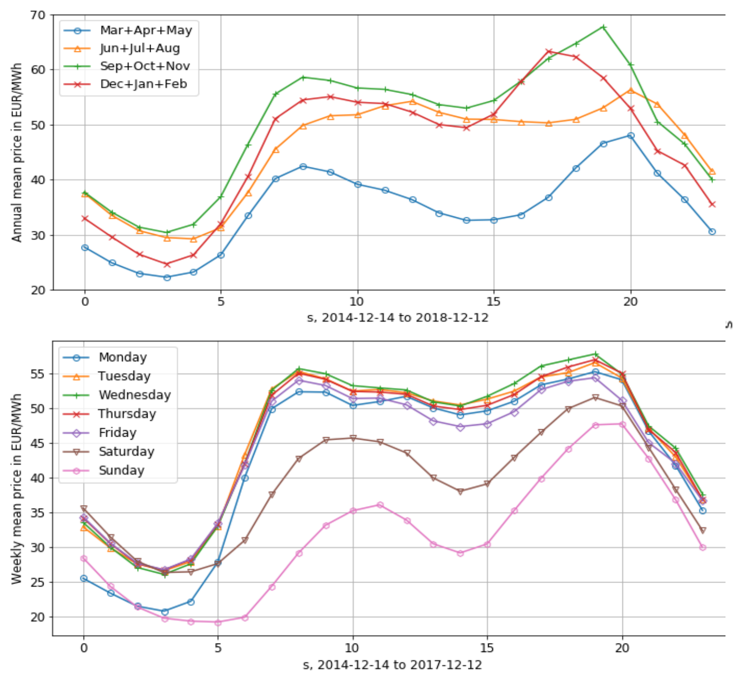
2.2 Seasonal structures

It can be observed from Figure 3 that electricity prices of Hungary vary overtime. There are big differences between the prices on daytime and nighttime. The gap is also different for weekend and working days. Hence, I will add a category variable that is number of hour from 0 to 23 in order to capture the hour effect. Overall, the electricity is much cheaper on weekend, is which usually considered as weekend effect. There is another popular effect in electricity marker, so-called transition effects. Electricity prices during night/morning hours of Monday are lower than one for other working days. Additionally, electricity prices during afternoon/evening hours of Friday are lower than one for other working days. Dummy variables represented for Monday, Saturday and Sunday are included in expert model in order to capture these effects. Moreover, electricity prices vary across seasons. In detail, it is highest in Autumn and lowest in Spring. Hence, it is promising to include seasonal dummies to improve the performance of models.

In Figure 4, the autocorrelations are calculated conditioning on last hour $Y_{d-1,23}$, $Y_{d-1,s}$, $Y_{d-2,s}$, $Y_{d-7,s}$, and dummy variables for Monday, Saturday and Sunday. It is clear that Tuesday and Friday have large correlations with electricity prices. Therefore, I will include two more dummy variables for Tuesday and Friday in my modified models.

2.2 External regressors:

It can be seen from Figure 5 that day-ahead load and electricity prices are positive correlated. In script, the correlation coefficients of load, emission allowance prices, gas prices, oil prices, and coal prices and electricity prices are presented in detailed. For most of hours, the correlation coefficients are significant, especially for day-ahead load and gas price (around 0.6 and 0.35 respectively). Hence, I will keep these regressors in my modified models.



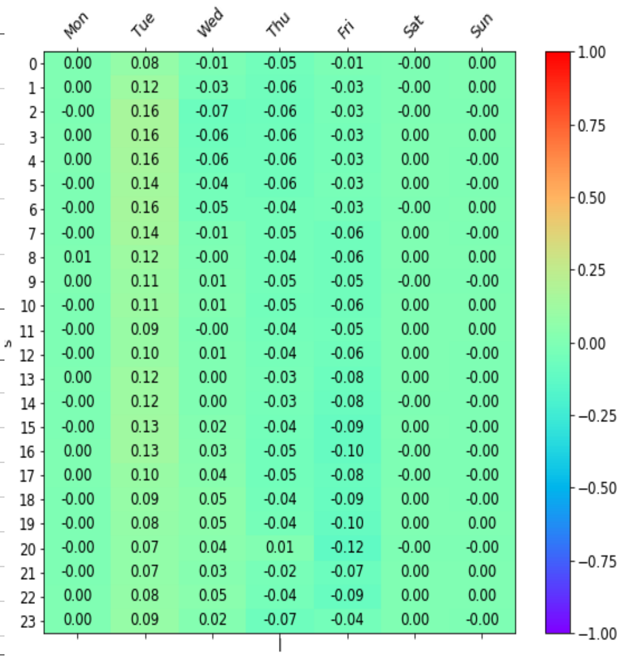


Figure 3: Hungary electricity prices over time

Figure 4: Conditional correlation Cor($Y_{d,s}$,DoW_d^l|X_d) with X_d as regressors of expert model

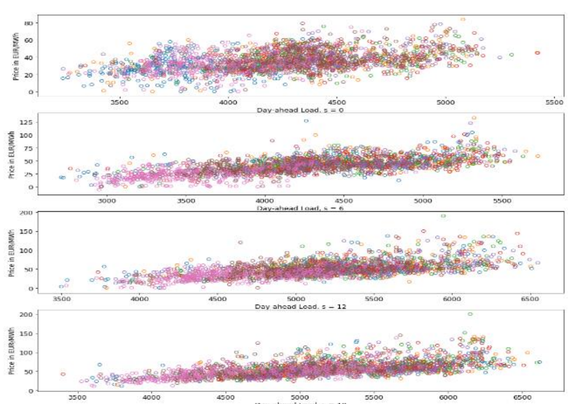


Figure 5: Correlation of day-ahead load and electricity price for s = 0, 6, 12, 18

3. Models and evaluation criteria

As discussed in Section 2, I use more variables in order to improve the accuracy of predictions. In general, I consider Naïve, Expert and Expert.REDADV model as benchmark models. However, there is no $X_{d,s}^{DA\text{-}RES}$ in Expert.REDADV due to the missing data in Hungarian dataset. In total, I try 9 models and compare their performances with those benchmark models, and these 9 implementing models are as follows:

– M1: Expert model with two more autoregressive lags 14 and 21.
– M2: Expert model with two more weekday dummies for Tuesday and Friday.
– M3: Expert model with last hour prices of previous day, $Y_{d-1, 23}$.
– M4: Expert model with seasonal dummies (Winter, Spring, and Summer).
– M5: Expert model with category variables for the hour (0 – 23) X_s, X_s^2, X_s^3, X_s^4, X_s^5 and X_s^6.
– M6: Expert model with all additional variables mentioned-above.
– M7: Multivariable-mixed expert model with all additional variables mentioned-above.
– M8: Expert.REDADV model with all additional variables mentioned-above.
– M9: Expert.REDADV model with all additional variables mentioned-above, applying advanced estimation method (LASSO together with cross-validation).

In addition, I also apply the idea of forecast combination to see if the performance of these models would be improved. I use two simple approaches of forecast combination:

– Naïve combination (N-C): the weights will be distributed equally for all 9 models.
– Long-short combination (LS-C): the weights will be distributed equally for top best models.

In order to evaluate the performances of these models as well as to make comparison with the benchmark models, I will use MAE to get rid of any possible significant effects of outliers.

4. Results:

4.1 Performance of models

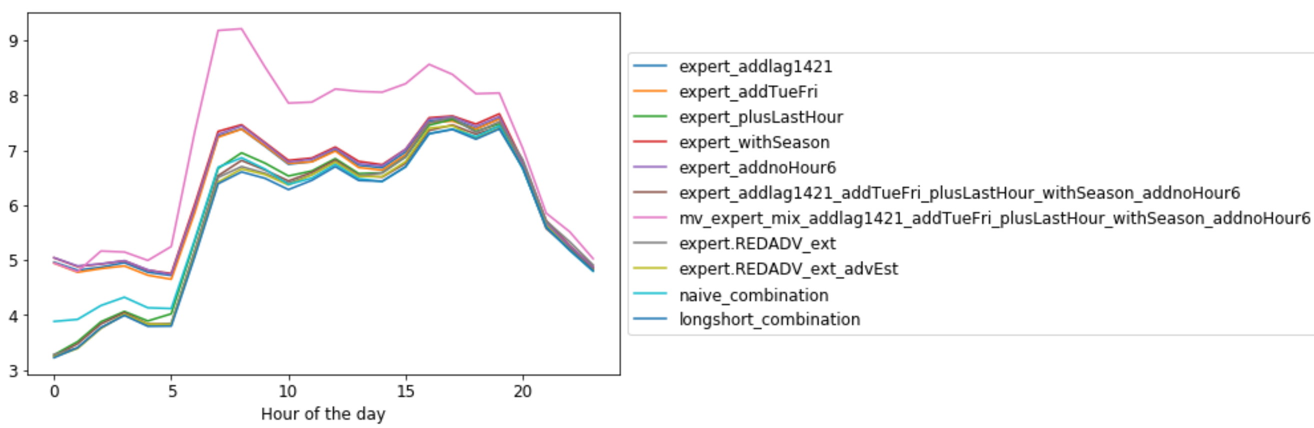


Figure 6: MAE of models over hours

It can be observed from Figure 6 that the performances of these models are better during night/morning hours, and getting worse for the rest of a day. The performance of extended Expert.REDADV model using LASSO estimation and cross-validation is better the one using OLS, which is determined by Diebold-Mariano (DM) test (p-value = $6.6 \cdot 10^{-5}$). In general, the expert model with hour category variables (M5), expert model with all additional variables (M6), extended Expert.REDADV models (OLS and LASSO estimator) (M8 and M9) are the four best models. In long-short combination, I weight four best models equally to come up with the final forecast for electricity prices. However, it is unclear from graph 6 which is the best model; hence, it requires a statistical test. I use DM test in order to draw a concrete conclusion on the best model. It is obvious in Table 1 that LS-C outperforms others.

Table 1: p-value of DM test between long-short combination and other models

	M1	*M2*	*M3*	*M4*	*M5*	*M6*	*M7*	*M8*	*M9*	*N-C*
LS-C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.0004	0.000

4.2 Comparison with benchmark models

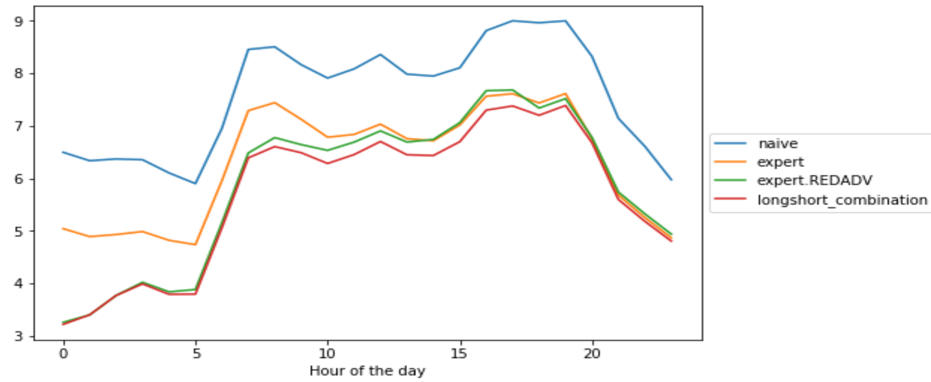


Figure 7: MAE of long-short combination and benchmark models

Long-short combination outperforms all naïve and expert models during hours of day (Figure 7). Comparing to Expert.REDADV model, MAE of long-short combination is lower most of the time. However, it is unclear if long-short combination is better than Expert.REDADV during morning time by looking at the graph only. Table 2 provides statistical proof that long-short combination outperforms all benchmarks.

Table 2: p-value of DM test between long-short combination and benchmark models

	Naïve Model	*Expert Model*	*Expert.REDADV*
LS-C	0.000	0.000	0.000

5. Conclusion:

The Expert.REDADV model performs exceptionally well with the Hungarian electricity dataset. I can improve the accuracy of Expert.REDADV by adding more relevant regressors that are investigated through exploratory data analysis. The application of advanced estimation LASSO not only helps to get rid of over-fitting problem, but also improve the performance. Finally, simple forecasting combination contribute to increase the accuracy of predicting models.

According to Viktor Miklos Kiss et al. (2016), Hungary does not produce enough electricity but depends heavily on imported electricity to cover demand. The main import partners are Slovakia, Austria, and Ukraine. Hence, it is worth considering day-ahead loads in Slovakia, Austria, and Ukraine while modeling forecasting models to predict the electricity prices of Hungary.

References:

Kiss, Viktor Miklos, Zsolt Hetesi, and Tibor Kiss. "Issues and solutions relating to Hungary's electricity system." *Energy* 116 (2016): 329-340.

Serati, Massimiliano, Matteo Manera, and Michele Plotegher. "Modeling electricity prices: from the state of the art to a draft of a new proposal." No. 833-2016-55461. 2008.

Szőke, Tamás, Olivér Hortay, and Richárd Farkas. "Price regulation and supplier margins in the Hungarian electricity markets." *Energy Economics* 94 (2021): 105098.