Forecasting Spot Prices of Commodities Using Futures Prices

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Thi Nhat Le Pham

First Supervisor: Prof. Dr. Andreas Thümmel | Second Supervisor: Mr. Declan O'Connor
Darmstadt University of Applied Sciences | Department of Natural Science and Mathematics & Computer Science | Course of Studies Data Science (M. Sc.)

Motivation

Commodities are standardized goods for which price plays a central role in purchasing decisions. Due to the low degree of differentiation among suppliers, pricing policy becomes particularly important—especially in industrial goods markets, where price negotiations are common and heavily influenced by the negotiating skills of the involved parties [1]. In this context, reliable price forecasts are gaining importance: those who can better anticipate future spot prices gain advantages in negotiations as well as in operational decisions such as production planning, inventory management, and investment timing [2].

Price forecasts serve as a strategic tool in risk management and help optimize returns and planning certainty. Futures prices, in particular, are considered a promising approach to forecasting future spot prices, as they reflect aggregated market expectations [3]. Therefore, the aim of this thesis is to systematically analyze the forecasting potential of futures prices and, in doing so, improve the decision-making basis for market participants in the commodity sector.

The methodological focus lies on comparing different model classes within the field of time series analysis. Four approaches are considered: the Random Walk and ARIMA model as classical statistical methods, the Kalman filter as a state space model, and XGBoost as a modern algorithm from the field of machine learning.

Methodology

- **Original Data:** Skimmed Milk Powder (SMP) and butter data are provided by QuaRisMa GmbH. Spot prices are available at a monthly frequency, while futures prices are recorded daily.
- **Data Preprocession:** To ensure comparability, daily futures prices were aggregated to monthly end-of-month values, as these offer a clear, standardized reference point and better reflect recent market expectations—especially for less liquid products like butter and SMP. Additionally, a structural aggregation was performed: futures contracts were grouped into 1-M, 3-M, and 6-M categories based on time to maturity. This approach reduces complexity while preserving relevant short- and medium-term market information.
- **Data Characteristics:** To assess the time series used in the study, statistical tests were conducted for stationarity (ADF test), heteroskedasticity (ARCH test), structural breaks (CUSUM test), and seasonality (F-test).
- **Models:**

1. **Random Walk**: Serves as a simple benchmark model, assuming that future prices equal the most recent observed spot price.
2. **ARIMA**: A classical statistical time series model that relies solely on historical spot price data to capture autoregressive and moving average structures.
3. **Kalman Filter**: A state-space model that incorporates futures prices as an exogenous variable, enabling dynamic updating and handling of latent variables.
4. **XGBoost**: A modern machine learning algorithm that also integrates futures prices as exogenous inputs to enhance predictive performance through nonlinear modeling.

- **Evaluation:** By combining fixed and rolling window approaches for training and testing, this study ensures a comprehensive assessment of model robustness and forecasting accuracy across six time horizons.
  The evaluation metrics MAE and MASE were selected based on their compatibility with the statistical properties of time series data, as emphasized by Hewamalage et al. (2023) [4], thereby further enhancing the reliability of the model comparison.

Results

- **Data Characteristics:**

The formal statistical tests provide the following key insights about the data:

- None of the time series are stationary, as indicated by **ADF** p-values greater than 0.05.
- There is strong evidence of heteroskedasticity, with **ARCH** test p-values well below 0.05.
- All time series exhibit structural breaks, as the **CUSUM** test p-values are below 0.05.
- No significant monthly seasonality is detected, as the **F-test** p-values are approximately 1.0.

These results support the characterization of the time series as non-stationary, heteroskedastic, structurally unstable, and non-seasonal.

- **Forecasting:**

To evaluate forecasting accuracy, the error metrics **MAE** and **MASE** were calculated using a Rolling Window and averaged across all windows for each model and both datasets.

The **ARIMA model** consistently performs worse or no better than the naive model across all forecast horizons and datasets—with a single exception (Butter, horizon 1).

Horizon	SMP		Butter	
	MAE	MASE	MAE	MASE
1	112	1.06	162	0.631
2	184	1.73	298	1.17
3	214	2.02	452	1.77
4	212	2.00	606	2.37
5	209	1.97	779	3.05
6	230	2.17	977	3.84

Table 1. Model evaluation (MAE and MASE) for ARIMA across SMP and Butter datasets, aggregated using arithmetic mean

It is therefore unsuitable for reliable forecasting.

In contrast, **Kalman Filter** and **XGBoost** were evaluated more extensively, using `Fut_plus_1M`, `Fut_plus_3M`, and `Fut_plus_6M` as exogenous variables. For this reason, the results cannot be displayed in table form here, but the key findings are summarized below.

- With `Fut_plus_1M`: both models show significantly improved results for SMP:
  XGBoost offers stable performance across all horizons.
  Kalman filter performs particularly well for short-term forecasts (horizons 1–2).
  For butter, **Kalman filter** with `Fut_plus_1M` performs close to the naive model, while **XGBoost** remains consistently reliable.
- With `Fut_plus_3M`:
  XGBoost remains stable for SMP but deteriorates significantly for Butter, especially at longer horizons.
  Kalman filter is usable for short horizons, but generally underperforms at longer ones.
- With `Fut_plus_6M`:
  XGBoost performs worse than the naive model across both datasets.
  Kalman filter shows average performance, without clear or consistent superiority.

Discussion

Overall, the **ARIMA** model proves unsuitable for reliable forecasting of complex time series exhibiting non-stationarity, heteroskedasticity, and structural instability, due to its reliance on stationarity and stable data structure assumptions.

In contrast, the **Kalman filter** and **XGBoost** demonstrate greater robustness:

- The **Kalman filter** employs a state-space approach that adapts to changing system dynamics, making it well-suited for evolving time series structures.
- **XGBoost** effectively captures nonlinear relationships and complex interactions, handling noise and feature dependencies flexibly.

Incorporating futures prices as exogenous variables significantly improves forecasting quality for both models, as futures embed important market expectations. The **Kalman filter** shows particular strength and sensitivity at longer forecast horizons, while **XGBoost** adapts flexibly to uncertainties but may be more prone to overfitting.

The **Kalman filter** benefits from additional information and exogenous inputs but is limited by assumptions such as linear state transitions and Gaussian process noise. **XGBoost**, though more difficult to interpret and often requiring extensive tuning, excels in modeling nonlinear effects.

Both models, especially when using the exogenous variable `Fut_plus_1M`, substantially outperform naive and simpler benchmarks, highlighting that combining advanced modeling techniques with relevant exogenous information greatly enhances commodity spot price forecasting accuracy.

Conclusion and Future Work

This study investigated the extent to which futures prices improve spot price forecasting. Results show that including futures as exogenous variables significantly enhances accuracy, especially for short one-month horizons. Both the **Kalman filter** and **XGBoost** handle complex time series features well, providing robust forecasts.

The **Kalman filter** suits linear, sequential processes but is more sensitive at longer horizons, while **XGBoost** models nonlinearities flexibly but needs careful tuning and is harder to interpret.

Future research should explore additional commodities, higher-frequency data, and advanced **Kalman filter** variants (e.g., **Extended** or **Unscented Kalman filter**) to relax model assumptions. Improved **XGBoost** parameter optimization and extensions like **ARIMAX** or **SARIMAX** could also be beneficial. Finally, further investigation of exogenous variables and aggregation methods may enhance forecasting and understanding of futures–spot price dynamics.

References

[1] Markus Moeth and Uta Herbst. Preisverhandlungen auf commodity-märkten. In Margit Enke & Anja Geigenmüller, editor, *Commodity Marketing*, pages 151–170. Gabler, 2011.

[2] Faster Capital. Preisprognosen so nutzen sie preisprognosen um vorauszuplanen und risiken zu verwalten, 2025. Accessed on May 01, 2025.

[3] Ranran Li. Forecasting energy spot prices: A multiscale clustering recognition approach. *Resources Policy*, 2023.

[4] Hansika Hewamalage, Klaus Ackermann, and Christoph Bergmeir. Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37(3):788–832, 2023.