Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis International Bachelor in Econometrics & Operation Research

Complex Dynamics of Commodity Market Returns: A Kernel Principal Component Approach

Warnier de Wailly Jules (602905)

Supervisor:	Daan Opschoor
Second assessor:	Pierluigi Vallarino
Date final version:	30th June 2024

Abstract

Commodities play a pivotal role in global economies, influencing both developed and developing nations. This research delves into the ability to forecast the dynamics of price movements within commodity sectors, uncovering interconnections between the returns. Such insights are essential for optimizing resource allocation, potentially guiding decisions along the entire commodity value chain, from producers upstream to traders in the midstream and ultimately to consumers downstream. This research explores how Principal Component Analysis (PCA) aids in identifying connections between commodities, leading to more accurate estimations. It further extends the PCAs with kernel capturing non-linear movement and reduces forecasts errors. The study evaluates model performance across three distinct stages: return, volatility, and Value at Risk (VaR) forecasts. It provides a comprehensive analysis of all applications of the models. Ultimately, the research demonstrates the presence of complexity in the short run with reductions in forecasts errors for complex models. These methods perform especially well during period of high volatility where the inclusion of a Kernel has mixed effects, further improving accuracy over a simple PCA under certain conditions. It also shows cluster of commodities in reaction to complexity.

1 Introduction

The Commodity Global Value Chain is the core of today's economy. Primary commodities serve as crucial inputs in all production processes (Fally & Sayre, 2018). This chain not only ensures the sustenance of the global economy by delivering essential soft commodities like food to areas of need but also propels economic advancement through fuelling innovation. All this brings high interest from researchers seeking to understand its implications more comprehensively.

In addition to its pivotal role for producers and consumers, traders occupy an important position in the midstream. They face the challenge of balancing supply and demand, with primary commodities accounting for approximately 16% of world trade, as highlighted in the research by Fally and Sayre (2018). Understanding the difficulties of this chain is paramount for creating value and ensuring its success. Furthermore, Ge and Tang (2020) discovered that commodity returns can robustly predict GDP growth in the subsequent quarter. They underscored the significance of comprehending commodity prices. Their study concluded that commodity prices can serve as leading indicators of economic growth. An increase in commodity prices and basis values result in a stronger future economy. These findings underscore the critical importance of understanding the dynamics of commodity prices in driving economic growth.

Now, this research delved into the interconnections among various commodities. Building on Palaskas and Varangis (1991) which supported the hypothesis of co-integrated movement in price. Now, this research employs Principal Component Analysis (PCA) to identify common factors and reduce noise. We may assume that the price of fuel may impact the prices of soft commodities due to the use of fuel for transporting goods from production to consumption, but this study seek to discover and quantify impacts. It employs PCA, and implements Kernel tricks, to aim to uncover non linear relationships. This tricks introduces non-linear links in a PCA (Hofmann, Schölkopf & Smola, 2008). It has already proved improvement in the field of macroeconomics (Kutateladze, 2022). This is supported by some work documenting the superiority of nonlinear models in the context of time series (Teräsvirta, Tjøstheim & Granger, 1994). Hereon, this research seeks to implement and reduce noise for more precision in return and financial instrument forecasts.

Samuelson (2016) lays down the foundations, proving with a general stochastic model for price change that price differences are random if not completely independent of the previous differences. This has brought focus to long-term forecasting of commodities. Hereon, the benchmark of our analysis is structured in similar fashion reflecting simplicity of the returns. In addition, Tolmasky and Hindanov (2002) focused on its use for risk management, particularly addressing the long-term high volatilities of raw materials that require analysis of price movements crucial for correct contracts. In the case of petroleum spreads, Girma and Paulson (1998) explored statistical relationships and seasonalities with the purpose of testing trading rules. According to them, the relationship concerning inter-commodity spreads can be complex and difficult to understand, and there is a lack of concrete theoretical foundation that can explain the behavior of these spreads. This is the focus of our research.

Ye, Zyren and Shore (2005) found that the short-run demand elasticity of commodities is much lower than the short-run inventory elasticity in the United States. Short-term volatility is not as pronounced as in other industries. Indeed, consumption can be quite well anticipated in the short term. Additionally, they observed that the production elasticity is virtually zero. Their reasoning suggests that in this raw sector of the industry, short-term actions are often impractical. However, it's worth noting that their research is constrained by their assumption of 'normal market circumstances'. This assumption has been challenged, particularly in recent times, such as during conflicts in the Middle East, where volatility expectations for petrol surged in response to the news, contradicting their assumption. Nevertheless, the key takeaway is that prices deviations are sourced beyond supply and demand. Now, this research delves into other drivers within the domain of commodities. Furthermore, a sub sample analysis is conducted, it reveals the model's varying reactions and adaptability to both 'normal and non-normal market circumstances' over time.

Utilizing more complex models, the research investigates whether identifying interrelationships between commodities reduce errors in volatility and return forecasts. Extending on Tolmasky and Hindanov (2002) research, this research also aims to increase precision in Value at Risk (VaR) forecasting. VaR is a widely used risk management tool in finance that quantifies the potential loss in value of an investment portfolio over a defined period for a given confidence interval (Duffie & Pan, 1997). It provides an estimate of the maximum loss that might be expected, thus helping investors and financial institutions manage and control the level of risk they are taking.

Ultimately, this research demonstrates the benefits of incorporating PCA for forecasting returns and related financial instruments. It also highlights the improvements from using a kernel trick to address non-linear relationships. Additionally, it emphasizes the potential for more precise short-term forecasts. However, it shows that long-term forecasts are more challenging for complex models to outperform the historical average. Finally, some commodity cluster react differently to the addition of complexity.

Moving forward, the research delves deeper into specific methodologies and empirical analyses to address the question. Section 2 summarize the data employed for the research. Then, Section 3 explains the theoretical frameworks of the advanced statistical techniques used. Finally, the results are shared in Section 4 followed by the conclusion and future research directions in Section 5.

2 Data

The data used in this research are commodity prices since 1990. This information is publicly available on the International Monetary Fund (IMF) website, facilitating replication. The dataset is updated on a monthly basis by the IMF. Its strength lies in its broad coverage and diverse sources. Soft commodities such as food and beverage prices are obtained from the United States Department of Agriculture and from the International Coffee/Cacao/Tea Organization respectively. Metals data are sourced from a combination of the London Metal Exchange and the ICE Benchmark Administration. Energy prices are acquired from Refinitiv and Argus.

For ease of comparison and comprehension, the data is reported in indexes. The Primary Commodity Price Index (PCPI) have for base year 2016. Furthermore, the IMF collection methods are helping the analysis. Indeed, all indexes are computed from Unites States Dollars (USD) prices helping in scaling, key for the success of PCA.

However, there are some limitations. The analysis focuses on worldwide prices, acknowledging that commodity prices can vary significantly due to factors beyond supply and demand, such as quality and origin. Despite these variations, the increasing importance of trade, as highlighted by Fally and Sayre (2018), emphasizes the significance of examining cost changes and the benefits of commodity trading. Their research also underscores the low price elasticity of demand due to difficulty in finding substitutes. Low price elasticity of supply, and high dispersion of natural resources across countries, suggest a convergence of worldwide prices. Therefore, studying world prices holds importance in the literature.

Here is a resume of the collected data used in the research and their meaning:

- $\mathcal{I}_{\textbf{food}}$: Food Price Index includes Cereal, Vegetable Oils, Meat, Seafood, Sugar, Apple (non-citrus fruit), Bananas, Chana (legumes), Fishmeal, Groundnuts, Milk (dairy), Tomato (veg) (recorded in USD/mt)

- $\mathcal{I}_{\textbf{beverage}}$: Beverage Price Index includes Coffee, Tea, and Cocoa (recorded in cts/kg)

- $\mathcal{I}_{\textbf{metals}}$: Base Metals Price Index includes Aluminum, Cobalt, Copper, Iron Ore, Lead, Molybdenum, Nickel, Tin, Uranium and Zinc (recorded in USD/mt)

- $\mathcal{I}_{\textbf{pre.metals}}$: Precious Metals Price Index includes Gold, Silver, Palladium and Platinum (recorded in USD/troy ounce)

- $\mathcal{I}_{\textbf{gas}}$: Natural Gas Price Index includes European, Japanese, and American Natural Gas (recorded in USD/MMBtu)

- $\mathcal{I}_{\textbf{coal}}$: Coal Price Index includes Australian and South African Coal (recorded in USD/t)

- $\mathcal{I}_{\mathbf{oil}}$: Crude Oil (petroleum), Price index is the average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh (recorded in USD/bbl)

This research focuses on a grouped dataset to facilitate the initial application of the methods and to reduce computational demands. An additional analysis, detailed in the appendix, applies the same method to Oil indexes using the complete dataset provided by the IMF, demonstrating the potential for further enhancements.

Figure 1 presents a plot of the time series, showing the overall increasing price indexes over time. Two spikes are noticeable in the plot, corresponding to two notable time spans: the financial crisis of 2008 and the COVID-19 pandemic in the early 2020s. Both periods experienced high inflation driven by negative impacts on the supply chain and the fragility of the financial system. However, while both crises significantly impacted the economy, the 2020 crisis has unique features due to the worldwide pandemic and the aggressive fiscal and monetary responses.

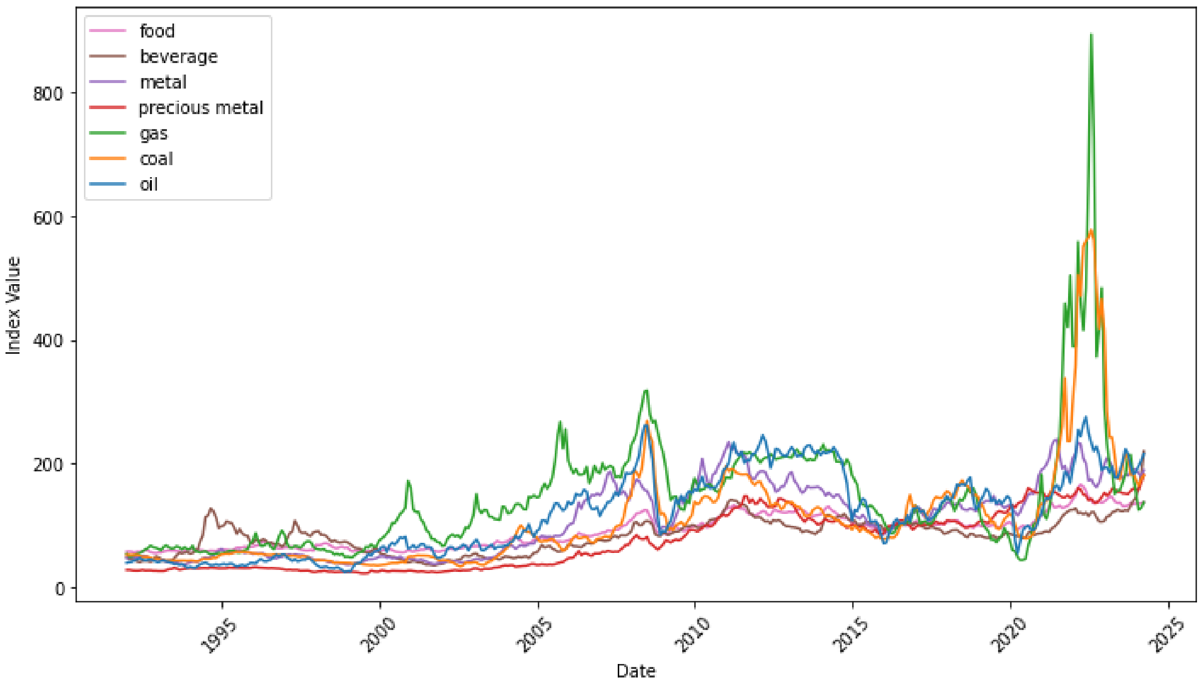


Figure 1: Time Series of Commodities Indexes

Co-movements are confirmed by Figure 1, Table 1 displays the correlation between the series. The overall high numbers support co-movement hypothesis (Palaskas & Varangis, 1991).

Table 1: Correlation Matrix of the Indexes

	$\mathcal{I}_{\textbf{food}}$	$\mathcal{I}_{\textbf{beverage}}$	$\mathcal{I}_{\textbf{metals}}$	$\mathcal{I}_{\textbf{pre.metals}}$	$\mathcal{I}_{\textbf{gas}}$	$\mathcal{I}_{\textbf{coal}}$	$\mathcal{I}_{\textbf{oil}}$
$\mathcal{I}_{\textbf{food}}$	1	0.78	0.93	0.94	0.68	0.80	0.91
$\mathcal{I}_{\textbf{beverage}}$	0.78	1	0.74	0.76	0.43	0.59	0.67
$\mathcal{I}_{\textbf{metals}}$	0.93	0.74	1	0.89	0.62	0.71	0.90
$\mathcal{I}_{\textbf{pre.metals}}$	0.94	0.76	0.89	1	0.52	0.69	0.80
$\mathcal{I}_{\textbf{gas}}$	0.68	0.43	0.62	0.52	1	0.86	0.73
$\mathcal{I}_{\textbf{coal}}$	0.80	0.59	0.71	0.69	0.86	1	0.73
$\mathcal{I}_{\textbf{oil}}$	0.91	0.67	0.90	0.80	0.73	0.73	1

Table 2 shows a summary statistic of the indexes. It indicate that energy commodities have a higher average index and significantly higher standard deviations. This is also apparent in their wide range (max-min). Secondly, the food sector appears much less extreme with relatively low indexes and the lowest standard deviations. Finally, the metal sector is the most diverse. Precious metals have lower indexes and standard deviations compared to base metals.

Table 2: Summary Statistic of the Indexes

	Food	**Beverage**	**Metal**	**Pre. Metal**	**Gas**	**Coal**	**Oil**
Mean	90,6	83,9	109,0	75,0	145,7	109,5	117,4
Std	27,1	27,6	58,4	46,5	99,5	88,8	66,3
Min	55,1	35,2	36,4	22,5	43,9	33,6	24,1
Max	165,7	219,5	238,8	181,4	893,1	577,6	275,4

3 Methodology

This section details the research methodology. It starts with an overview of the data preprocessing steps, followed by a description of the models employed to extract returns forecasts used for forecasting volatility to further compute the Value at Risk (VaR). Subsequently, it explains the method used to compare the divergence between different approaches.

3.1 Data Manipulation

This research focuses on commodity prices, Samuelson (2016) established properties of returns. To achieve stationary and correct distributions in the data, log-returns are considered. Let the log-return at time t be $\log(P_t) - \log(P_{t-1})$ where P_t is the price of the commodity at time t. However, it's important to note that indexes are collected rather than direct prices. Nonetheless, since the base is common within a time series, applying the same function to indices yields the correct return and thus does not bias interpretation.

$$\begin{aligned}\log(\mathcal{I}_t) - \log(\mathcal{I}_{t-1}) &= \log\left(\frac{P_t}{P_{2016}}\right) - \log\left(\frac{P_{t-1}}{P_{2016}}\right) \\ &= \log(P_t) - \log(P_{2016}) - \log(P_{t-1}) + \log(P_{2016}) \\ &= \log(P_t) - \log(P_{t-1}) \\ &= r_t\end{aligned}$$

As both stationary and correct distribution are required assumptions for further models, it has to be verified. In this research, the Augmented Dickey Fuller (ADF) (Dickey & Fuller, 1979) (Dickey & Fuller, 1981) test is used. It tests the null hypothesis (H_0) that the time series has a unit root. The time series can be considered statistically stationary if H_0 is rejected. It is achieved by estimating the following equation by Ordinary Least Squares (OLS):

$$\Delta r_t = \alpha r_{t-1} + \sum_{i=1}^{p} \beta_i \Delta r_{t-i} + \epsilon_t$$

where p is the number of lag differences included. The longest statistically significant lag is selected, using the BIC. More explanation on the BIC are given further down the research. Then, by testing for significance of the coefficient α from its t-statistic and comparing it to the Dickey Fuller critical value (Cheung & Lai, 1995) the H_0 may be rejected.

To test for normality, the research employs a Jarque-Bera test (Jarque & Bera, 1980). It is a statistical test that assesses whether sample data have the 3rd and 4th moment, respectively the Skewness (S) and Kurtosis (K), matching a normal distribution. The null hypothesis (H_0) tests that the data are normally distributed. The Jarque-Bera test statistic (JB) is defined as:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \sim \chi^2_2$$

Under the null hypothesis of normality, the JB statistic follows a chi-squared distribution with 2 degrees of freedom.

To test for a Student-t distribution in the returns, a Kolmogorov-Smirnov test (Massey Jr, 1951) is employed. It compares actual returns to the Student-t fitted distributions. It measures the maximum difference between the empirical cumulative distribution of the returns and the hypothetical cumulative distribution of the Student-t. The empirical distribution is derived from the characteristics of the returns. The test evaluates the goodness of fit.

3.2 Modeling of the Returns

The following subsection explains how, at each time t, forecasts returns for horizons of 1, 6, and 12 months are computed. This is accomplished using a rolling window of 10 years minus the forecast horizon.

3.2.1 Benchmark

Initially, a benchmark is established. The non-random component (estimated return) is assigned to be equal to the mean of previous returns, denoted by μ. This can also be obtained by conducting an Ordinary Least Squares (OLS) regression of the time series on a constant term. Such a procedure is similar to the initial phase of a two-step Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) estimation with a constant. Breaking down the process into two steps facilitates comparison with subsequent models.

3.2.2 Principal Component

The next model is based on simple Principal Component Analysis (PCA). PCA is a statistical technique used for dimensionality reduction in data analysis. It identifies patterns in data, finds the directions of maximum variance (principal components), and projects the data onto these components. It reduces the number of dimensions and preserves most of the information. Simple PCA focuses on linear relationships.

Additionally, for an initial exploration of non linearity, PC2 and SPC are used. On the one hand, PC2 is identical to the previous model with the addition of the squares of factor estimates in the forecasting equation. On the other hand, SPC (Bai & Ng, 2017), applies the standard PCA algorithm to the original set of variable augmented by its square.

Finally, kernel tricks are implemented in the remaining models to tackle non linearity. It allows for the expansion of dimensions of the factors in terms of a Kernel function. This function maps the original data non linearly into a high-dimensional space using subset function: $\varphi(\cdot) : \mathcal{R} \to \mathcal{F}$. Here, $\mathcal{F}$ denotes the feature space and $\mathcal{R}$ is the set of returns. Let the Kernel function $k(r_i, r_j)$ equal $\varphi(r_j)'\varphi(r_i)$. This reflects common factors between data points. The result is a Kernel function where: $k(.,.) : \mathcal{R} \times \mathcal{R} \to \mathbb{R}$. The kernel matrix K is created such that $\{\mathcal{K}\}_{ij} = \varphi(r_j)'\varphi(r_i)$. One key note is that $\varphi(\cdot)$ does not have to be known, as for a valid Kernel, there always exist a corresponding $\varphi(\cdot)$. For a Kernel function to be valid, it musts be positive-definite, meaning $\int\int f(r_i)k(r_i, r_j)f(r_j)\, dr_i\, dr_j \geq 0$ for any square-integrable function f. In this research the following valid kernel are employed (Kutateladze, 2022):

1. Sigmoid Kernel:
$$k(r_i, r_j) = \tanh(\gamma(r_i'r_j) + 1)$$
2. Radial Basis Function (RBF) Kernel:
$$k(r_i, r_j) = e^{-\gamma\|r_i - r_j\|^2}$$
3. Quadratic polynomial (poly(2)) Kernel:
$$k(r_i, r_j) = (r_i'r_j + 1)^2$$

Where the parameters γ are determined over a grid of values by time series cross-validation. More details on that are provided below.

Furthermore, to model the series, an Autoregressive Diffusion Index (ARDI) model is employed. The AR part allows for lags of the dependant to be include, it is a key aspect of time series. The DI part allows for addition of factor and their laggs in the equation. It is computed mathmaticaly as

$$r_{t+h} = \beta_0^h + \sum_{p=1}^{P_t^h} \beta_{Y,p}^h r_{t-p+1} + \sum_{m=1}^{M_t^h} \beta_{F,m}^{h\,\prime} F_{t-m+1} + \epsilon_{t+h} \tag{1}$$

Where P_t^h, M_t^h, K_t^h are the number of lags of the target variable, the number of lags of factors, and the number of factors, respectively. Such that, the factor F_{t-m+1} and the loading $\beta_{F,m}^h$ are $K_t^h \times 1$ vectors. Finally, r_{t+h} are commodity returns.

The modeling of the (K)PCA on the commodity return dataset has been applied in such a way: First, depending on the forecast horizons the indexes are scaled to apply direct forecast; $\Delta_{hor} Log(\mathcal{I})$. Simultaneously, the factors are created from the simple returns and transformed with respect to the correct PCA method.

Second, the correct lags are identified. It is achieved by looping over each possible combination of lags, updating the dataset accordingly, and fitting the dataset with OLS according to Equation 1. Hereon, the Bayesian Information Criterion (BIC) is computed. It evaluates the trade-off between goodness of fit and model simplicity. It is computed as $-2ln(L) + k * ln(n)$. The first part values better fits, it is computed from the likelihood function (L). This reflects the likelihood of certain parameters given observed outcomes is the same as the probability of those outcomes given the parameters. That is, an increase in the error variance decreases the likelihood function. The second part is the penalty given to complex models, computed from the number of parameters (k) and the number of observations (n). The combination of lags that minimizes the BIC is chosen.

Optionally, if different hyperparameters are available (e.g., for Kernels with Sigmoid and RBF functions), Gamma is determined over a grid of values using time series cross-validation. Specifically, the latest five available observations are consecutively predicted, and the hyperparameter that minimizes the average error is selected. The hyperparameter grid for γ is defined as $\gamma \in \{10^z \mid z \in [-6, -3]$ with a step size of $0.5\}$. These small values are used to scale. For the Sigmoid Kernel, γ scales the dot product between observations. For the RBF Kernel, γ scales the distance between observations.

Finally, the optimal model is used to make h-step ahead forecasts. By retrieving the coefficients and updating the dataset according to the method, horizon and lags, the model is fitted and the forecast is generated.

3.3 Volatility Modeling - GARCH

In this subsection, the foundation of the GARCH method for modeling and forecasting volatility are elucidated (Bollerslev, 1986). The conditional distribution of returns r_t is defined as $r_t = f(\mathbf{x}_t) + \varepsilon_t$, where $\mathbf{x}_t$ represents factors and $f(.)$ denotes the loading function. For the benchmark, the $f(.)$ is the mean and for other models it is the ARDI model, Equation 1. The error term ε follows: $E(\varepsilon_t | \mathcal{I}_{t-1}) = 0$ and $E(\varepsilon_t^2 | \mathcal{I}_{t-1}) = \sigma_t^2$, where $\mathcal{I}_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$. This formulation allows for an alternative expression of the returns: $r_t = f(\mathbf{x}_t) + z_t \sigma_t$, where $z_t \sim$ iid $\mathcal{D}(0, 1)$.

The differences between the models arise from the variation in $f(\mathbf{x}_t)$ and differences in the parameterization of the distribution $\mathcal{D}$. Additionally, potential differences may arise from the return distribution as previously explained, where skewed-Normal and Student's t-distributions are considered.

Now, the GARCH(1,1) model (Bollerslev, 1986) describes volatility (σ^2) as:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

The rationale behind this lies in volatility clustering, indicating that σ_t^2 is closely linked to σ_{t-1}^2 and the squared previous unexpected return $(r_{t-1} - \mu)^2 = \varepsilon_{t-1}^2$. To ensure $\sigma_t^2 \geq 0$ for all t, the loading parameters are required to follow $\omega > 0$, $\alpha > 0$, and $\beta \geq 0$.

3.4 Value at Risk

Hereon, with an expected volatility, Values at Risk (VaR) are computed. VaR are the minimum return that could occur over a given time period with a specified probability. It essentially is a quantile of the distribution of returns. For any $0 < q < 1$, the VaR at $100*(1-q)\%$ for a specific period is the return that is expected to be exceeded with probability $1 - q$. It is computed as:

$$\text{VaR}_t(1-q, h) = \hat{r}_{t+h} + z_q\sigma_{t+h},$$

where $\hat{r_{t+1}}$ is the expected return, $z_q\sigma_{t+h}$ the quantile value and h the horizon. In this research, $q \in \{0.01, 0.05, 0.10\}$ is used. It represent respectively a confidence level of 99%, 95% and 90%.

3.5 Comparative Evaluation

This research compares model's performances at three distinct steps. This approach helps determine if incorporating models and interconnections into return forecasting instruments helps capture deviations.

First, the capacity of predicting the returns are evaluated by the Mean Squared Prediction Error (MSPE). That is the average of the squared differences between the returns forecast and the actual returns, it is given by:

$$\text{MSPE} = \frac{1}{T}\sum_{t=t_0}^{T}(r_t - \hat{r}_t)^2$$

where T is the final date of the available range, r_t is the actual return at time t, $\hat{r_t}$ is the predicted return for time t.

To gain insights into the different models' adaptability over various periods, the cumulative sum of squared forecast error differentials similar to Pettenuzzo and Timmermann (2017) is analysed. Mathematically, at time t, it is represented as:

$$\text{CSSED}_t = \sum_{\tau=0}^{t}(e_{\tau,1}^2 - e_{\tau,2}^2)$$

where: $e_{\tau,i}^2$ is the squared forecast error of model i at time τ. This approach allows to assess how model's forecast accuracy evolves over time compared to another model.

Secondly, the MSPE of the volatility forecasts are computed using the squared daily return as proxy for the true conditional variance (Brownlees, Engle & Kelly, 2011):

$$\text{MSPE} = \frac{1}{T}\sum_{t=1}^{T}(\sigma_t^2 - r_t^2)^2$$

where T is the final date of the available range, σ_t^2 is the volatility forecast at time t and r_t^2 is the return squared at time t.

In addition, to account for the proxy, a robust methods explained in Patton (2011) is used. This involves computing the QLIKE loss function for the forecasted volatility and the proxy (squared returns). To apply this function, certain assumptions are required: the returns must follow either a normal distribution or a Student's t-distribution with a degree of freedom higher than 2. Then, let $u_{i,t} = L(\hat{\sigma}_t^2, h)$ be the output of the loss function for model i at time t, where $L(\hat{\sigma}_t^2, h) = \log(h) + \frac{\hat{\sigma}_t^2}{h}$, withe forecast and the proxy as inputs. Then, the Diebold–Mariano–West for equal predictive accuracy of the benchmark and the different models is applied to assess precision of the models compare to the benchmark.

Finally, the conditional coverage of the VaR forecasts is tested. To assess the accuracy of the VaR forecasts as in Kupiec et al. (1995), the indicator function is defined as:

$$\mathcal{I}_{t+1} = \begin{cases} 1 & \text{if } r_t < \text{VaR}_t(1-q, 1) \\ 0 & \text{if } r_t \geq \text{VaR}_t(1-q, 1) \end{cases}$$

such that $\mathcal{I}_{t+1}$ indicates 'violations' of the VaR. Correctly specified VaR should have correct unconditional coverage. It allows to test for the following null hypothesis (H_0): $P(\mathcal{I}_{t+1} = 1) = E(\mathcal{I}_{t+1}) = q$. This null hypothesis is tested using a Likelihood Ratio test (Kupiec et al., 1995). Given independence, the likelihood function for interval forecasts with coverage probability $p = P[I_{t+1} = 1]$ is given by:

$$\begin{aligned} L(p; I_T, I_{T-1}, \ldots, I_1) &= P[I_T = i_T, I_{T-1} = i_{T-1}, \ldots, I_1 = i_1] \\ &= P[I_T = i_T]P[I_{T-1} = i_{T-1}] \cdots P[I_1 = i_1] \\ &= (1-p)^{T_0} p^{T_1}, \end{aligned}$$

where $T_1 = \sum_{t=1}^{T} i_t$, $T_0 = T - T_1$.

Furthermore, the Likelihood Ratio (LR) test compares the likelihood under the null $p = q$ with the likelihood under the alternative $p = \pi$, where π is to be estimated with maximum likelihood. Under the null hypothesis of correct unconditional coverage $p = q$: $L(q; I_T, I_{T-1}, \ldots, I_1) = (1-q)^{T_0} q^{T_1}$. Under the alternative hypothesis, $p = \pi$ for some $\pi \neq q$: $L(\pi; I_T, I_{T-1}, \ldots, I_1) = (1-\pi)^{T_0} \pi^{T_1}$. The maximum likelihood estimate of π is equal to:

$$\hat{\pi} = \hat{P}[I_{t+1} = 1] = \frac{T_1}{T_0 + T_1}.$$

The likelihood ratio test of correct unconditional coverage is computed as:

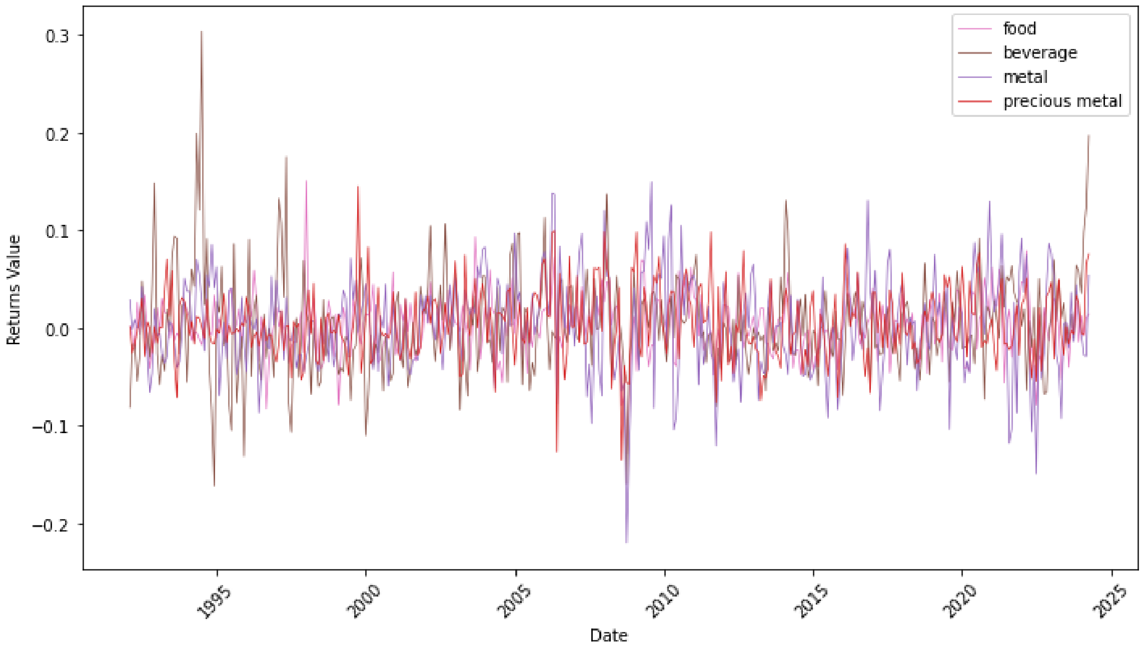
$$LR_{uc} = -2\log\left(\frac{L(q; I_T, I_{T-1}, \dots, I_1)}{L(\hat{\pi}; I_T, I_{T-1}, \dots, I_1)}\right) \sim \chi^2(1).$$

Then the test statistic is compared to the critical values of the Chi-Square distribution with 1 degree of freedom at a 5% confidence level.

4 Results

4.1 Data Manipulation

Figure 2 shows the commodities returns over time. Visually the expected results from the transformations appears achieved. There is no more long increase in the series like in Figure 1, but only sudden spikes fluctuating around the mean. This supports possible stationary. Furthermore, the Figure shows alternating periods of large and small movements in prices. This known as volatility clustering, it supports the use of GARCH model in volatility forecasting. Furthermore, similar patterns appears between commodities further supporting co-movement.



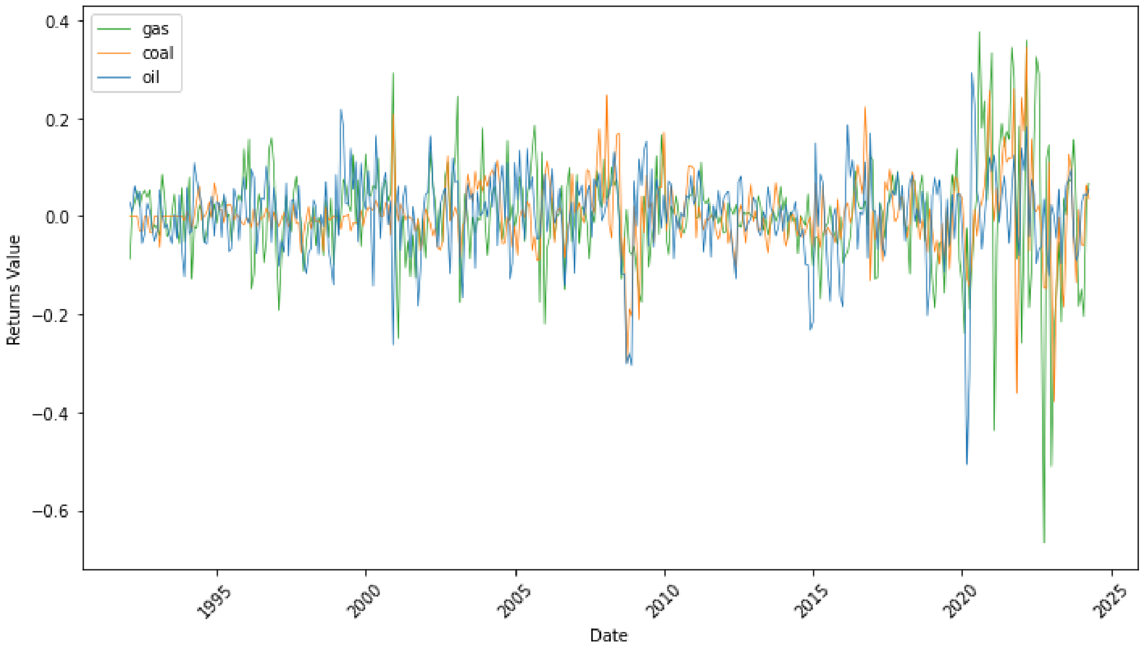


Figure 2: Time Series of Commodity Returns

Additionally, Figure 3 displays density plots of the returns. The observations indicate that the returns mostly appear to adhere to a more controllable distribution with a mean around 0. Food seems to be slightly skewed. However, their standard deviation exhibits fluctuations, further justifying the use of GARCH model for each series.

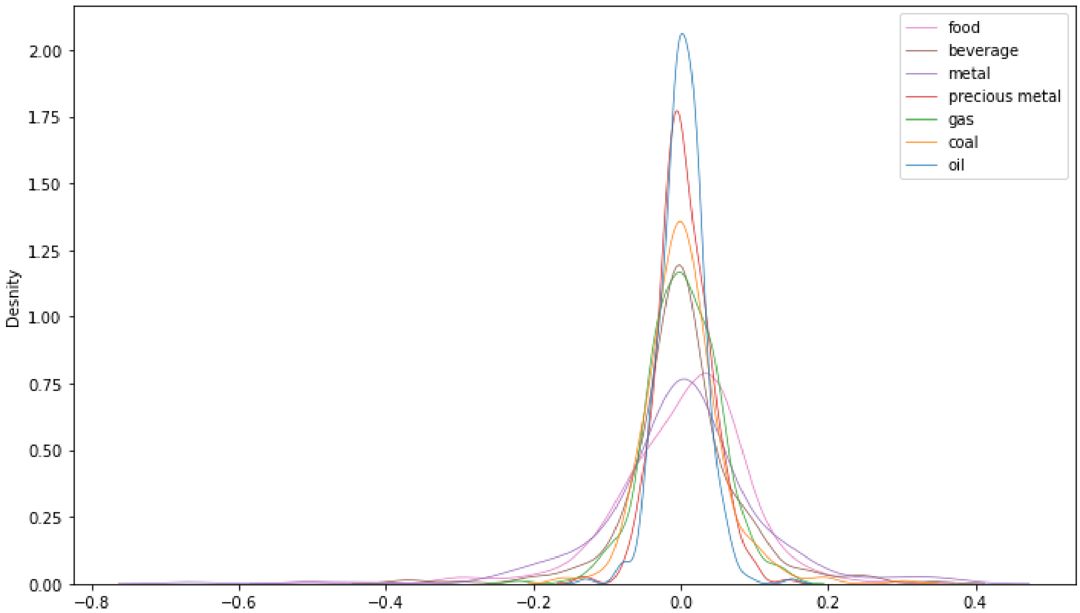


Figure 3: Density Plot of Commodity Returns

Table 3 shows the summary statistic of the returns. It confirms visual assumptions where means fluctuate around zeros but are overall slightly positive. Furthermore, standard deviations and ranges diverge between the returns. On the one hand, the Jarque Bera test returns small p-values, it implies the data does not follow normal distributions. This supports the presence of fat tails. On the other hand, KS test for the fit to a t-distribution shows high p-values. This indicates that there is no significant difference between the different returns and the t-distribution at a 5% significance level. In other words, the data can be considered to follow the t-distribution, hence this is specified in the modelling of volatility using GARCH. Furthermore, the degrees of freedom are relatively high supporting fat tails. Finally, all p-values of the ADF test are lower than .05. That is strong evidence against the null hypothesis of the test. The time series does not have a unit root, hence are likely stationary.

Table 3: Summary Statistics of Commodity Returns

	Food	**Beverage**	**Metal**	**Pre. Metal**	**Gas**	**Coal**	**Oil**
Mean	0,002	0,004	0,004	0,005	0,002	0,003	0,004
Std	0,029	0,050	0,049	0,035	0,106	0,074	0,085
Min	-0,129	-0,162	-0,220	-0,135	-0,667	-0,379	-0,507
Max	0,151	0,303	0,149	0,145	0,376	0,345	0,293
p values:							
JB	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Student-t	0,857	0,958	0,718	0,451	0,960	0,197	0,430
(df)	(7)	(4)	(9)	(9)	(2)	(2)	(4)
ADF	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

4.2 Modeling

Table 4 presents the relative MSPEs of different return forecasts. The benchmark model serves as the base for each horizon and series.

Table 4: Relative MSPEs of Returns Forecasts

Each value represents the ratio of overall out-of-sample MSPE of the returns from the different models compare to the benchmark. Values in green, lower than one, imply lower forecasts errors from the model. Values in red imply the opposite. The forecasts are h-step ahead and are computed from a rolling window of length 120 months minus the horizon.

	Horizons (Months)	**PCA**	**SPC**	**PC2**	**KPCA poly**	**KPCA sigmoid**	**KPCA rbf**
Food	**1**	0.769	0.873	0.786	0.802	0.769	0.768
	6	1.020	1.083	1.035	1.166	1.020	1.019
	12	1.208	1.623	1.595	1.468	1.207	1.208
Beverage	**1**	0.908	0.930	0.920	0.918	0.908	0.909
	6	1.176	1.267	1.297	1.373	1.175	1.176
	12	1.534	1.521	1.723	1.774	1.536	1.534
Metal	**1**	0.749	0.829	0.756	0.775	0.757	0.755
	6	1.199	1.328	1.264	1.612	1.199	1.203
	12	1.439	1.529	1.748	1.810	1.444	1.441
Pre. Metal	**1**	0.925	1.020	0.978	0.992	0.917	0.920
	6	1.226	1.287	1.239	1.261	1.218	1.225
	12	1.521	1.579	1.464	1.497	1.518	1.519
Gas	**1**	1.036	1.025	1.101	1.195	1.035	1.035
	6	1.126	1.143	1.321	1.391	1.127	1.127
	12	1.010	2,305	1.578	1.782	1.009	1.010
Coal	**1**	0.780	0.811	0.822	0.773	0.783	0.781
	6	0.873	0.926	0.932	0.889	0.874	0.873
	12	1.062	1509	1.247	1.174	1.063	1.063
Oil	**1**	0.790	0.896	0.823	0.863	0.789	0.790
	6	0.889	1.088	2,609	5,001	0.913	0.899
	12	1.159	1.302	1.338	2,052	1.156	1.156

Long-run return forecasts proved more challenging to improve upon the benchmark. Most relative MSPEs exceeds 1 as the forecast horizon reaches half a year. This suggests that in the long run, the most accurate forecast is the average return of previous periods, aligning with findings in inflation return literature (Stock & Watson, 2007). In the long run, simplicity is key to estimating returns, as the Random Walk or Auto-Regressive models are often difficult to outperform.

In the short term, however, more complex models shows improvements over simpler models.

PCA-based models reduces MSPE in 6 out of 7 cases, and KPCA-based models further reduce MSPE in 5 out of 6 cases. The use of KPCArbf forecasts reduces MSPE by about 24% for food and metals, for 1-month horizon compared to the benchmark. Nevertheless, for metals, a simple PCA model also achieves a notable increase in precision without requiring Kernel methods. Additionally, SPC and PC2 models rarely outperformed all other models, contributing minimally to accuracy.

The analysis suggests that adding PCA provides the most significant improvement, while Kernel methods do not notably enhance overall forecast accuracy further. Although Kernel methods sometimes increase overall accuracy, the overall improvement is not substantial enough to justify the additional time and computational costs for now.

Regarding commodities, certain clusters emerged. Complex model for Gas, and precious metals struggle to outperform the benchmark. Gas returns are the only ones where no model surpassed the benchmark. Nick and Thoenes (2014) concluded that in the short-run natural gas prices are affected by temperature and supply shocks. However, those characteristic are not perceived by current model implying the difficulty to improve forecasts. Precious metals such as gold are key for exchange rate and thus are strictly monitored making it hard to predict. In this way, Hassani, Silva, Gupta and Segnon (2015) emphasized the difficulty to outperform a Random Walk model. However, the model still manage to slightly reduce the forecast error. Finally, for the remaining commodities more complex model helped increase accuracy over the benchmark.

Now, Figure 4 and 5 present the Cumulative Sum of Squared Error differentials of the benchmark against the PCA method for 1 month and 12 month ahead forecast respectively. The Figure for 6 months ahead forecasts is stored in appendix.

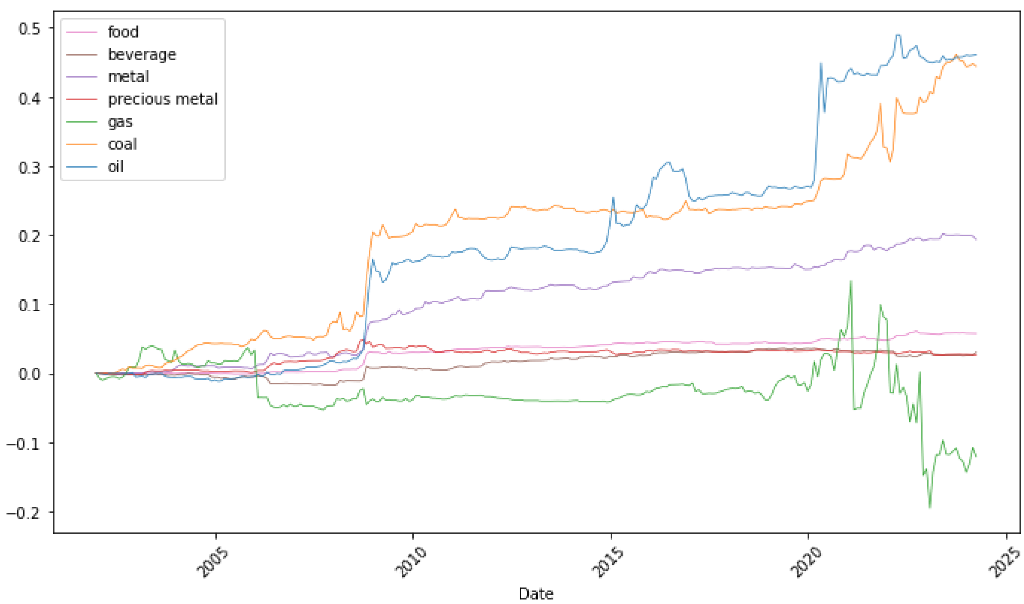


Figure 4: Cumulative Sum of Squared Error of the Benchmark vs PCA on 1-Month Ahead Forecasts

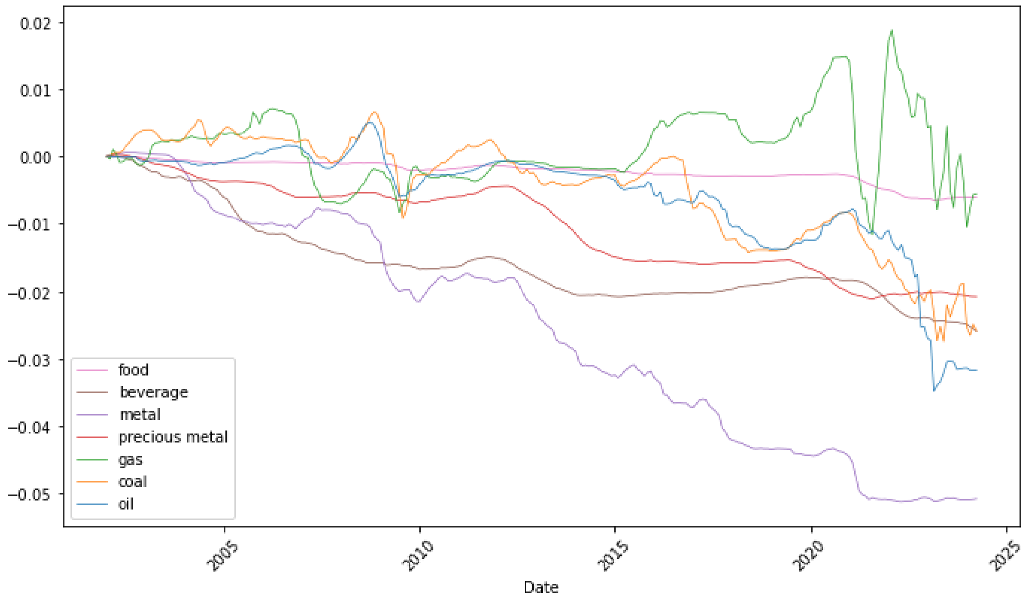


Figure 5: CCSE of the Benchmark vs PCA model on 1-Month Ahead Forecasts

First, PCA performs poorly for gas return forecasts as it mainly remains negative. It experiences a sudden spike in the beginning of the 2020s but is shortly followed by a drop, confirming the non-predictive power of the model. However, for all other commodities, the PCA model helps in reducing errors. Especially for oil, coal, and metals, it shows an overall increasing trend, implying better forecasts over all periods. Upon closer inspection, there are two spikes during the crises in 2008 and 2021. This indicates that the PCA model helps capture movements and provides less error in a highly volatile environment.

Figure 5 showcases the power of the benchmark for long-run forecasts. This graph also reflects drops in late 2009 and late 2022, indicating that the PCA model's reaction to financial crises negatively impacts long-run forecasts. From both figures, it can be concluded that the focus on complexity is important mainly in the short run.

To further examine the addition of a kernel, Figure 6 shows the CSSE of the PCA model against the KPCA model using the RBF kernel function.

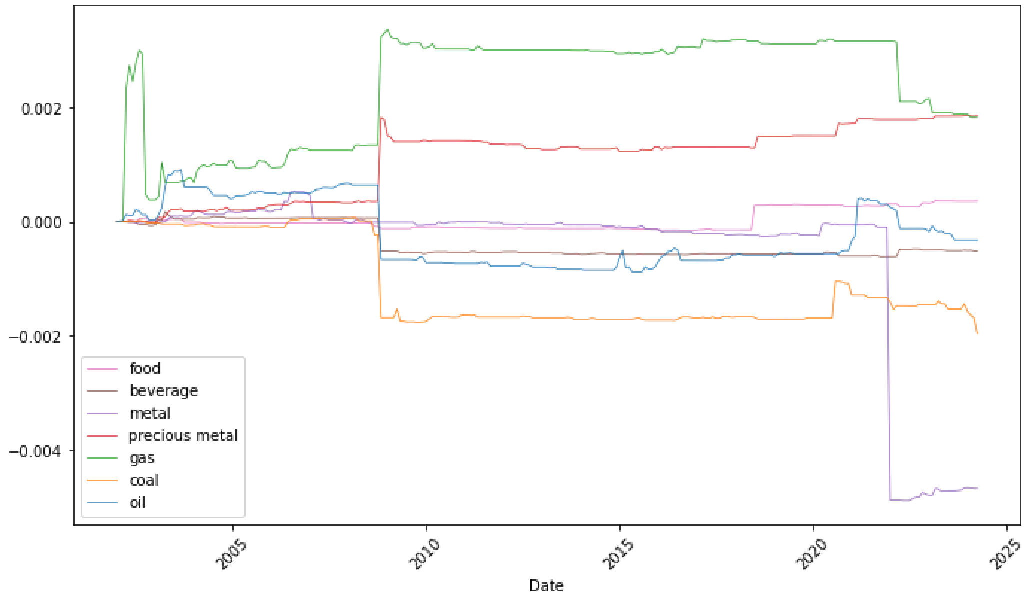


Figure 6: CSSE of the PCA vs KPCA rbf on 1-Month Ahead Forecasts

Figure 6 shows a mixed reaction to the addition of the kernel. First, the total range of difference is relatively small ($\simeq 0.008$), implying minimal differences in forecasts error between the models. However, even if the slope appears flat over time, divergences appears around 2008 and 2022. Those divergences suggest different scales and adaptations to the movement by the 2 models. The kernel improves the accuracy for gas and precious metals but decreases it for the remaining commodities. This is notable since these two commodities were previously classified as the hardest to improve upon. These results suggest that the forecasting of these primitive commodities could be enhanced beyond simple PCA with nonlinear modeling.

4.3 Volatility Modeling - Garch

Table 6 in the Appendix presents the relative MSPEs for the volatility forecasts. Although the PCA and kPCA models appear to offer greater precision overall, no clear patterns emerge that help identify the drivers of decreasing MSPEs.

Table 3 validates the assumptions necessary for conducting the DM test on volatility, showing that all returns follow a t-distribution with degrees of freedom greater than or equal to 2. Table 7 in Appendix, presents the p-values resulting from the test. All values are negative but have an absolute value smaller than 1.96. It indicates that none of the results are statistically significant at the 5% confidence level. Despite this, every t-statistic is negative, suggesting a slight improvement in predictive accuracy over the benchmark.

In summary, the improvement in volatility forecasts for the models compared to the benchmark is mixed. Using MSPE as a factor does not provide much insight. However, accounting for the proxy with a loss function shows some improvement towards complex models with negative t-statistics. While these results are not statistically significant, they indicate the models are heading in the right direction.

4.4 Value at Risks

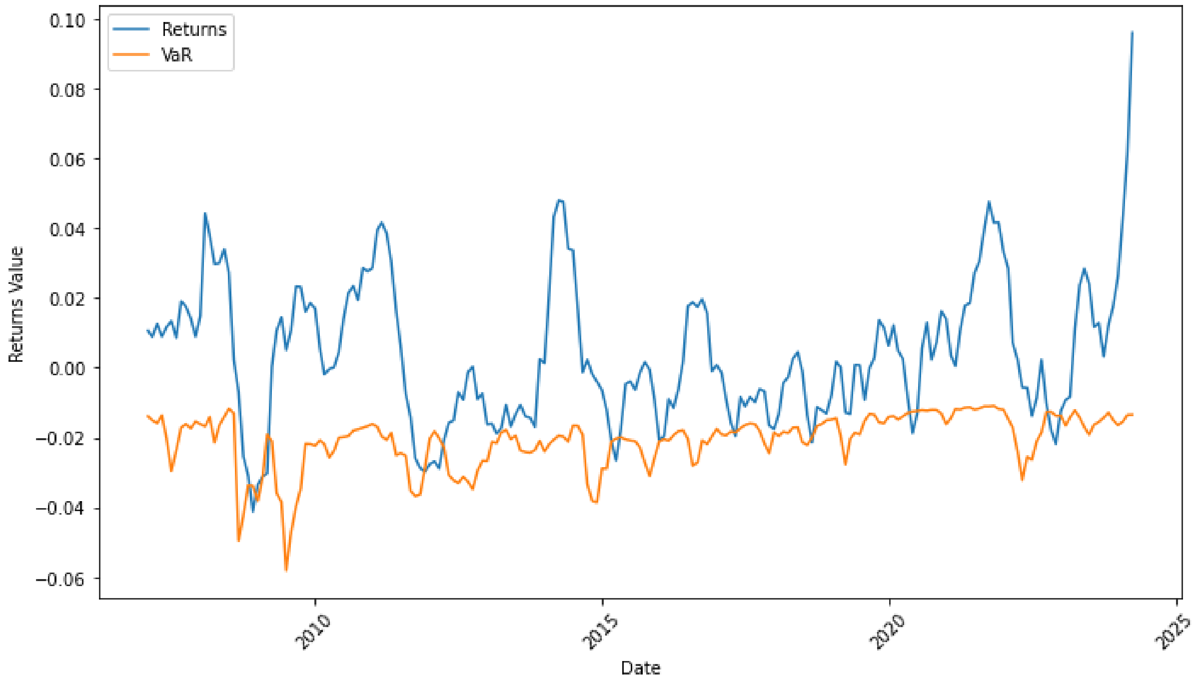


Figure 7: 1 Month VaR at 90% Confidence and Returns for Beverage Returns Using PCA

Figure 7 illustrates the 1-month VaR at 90% confidence for beverage returns, computed using PCA. The graph demonstrates how the VaR responds to spikes in returns. It is observed that each violation of the VaR occurs immediately after a sudden upward spike followed by a sharp drop. It is interesting to test for correct coverage of the VaR forecasts. However, for future application, this is notable and to be careful of while operating as this indicates that abrupt changes in returns, especially when there is a rapid increase followed by a decline, are key drivers of risk exposure in the beverage sector. Such behavior could imply that the market is particularly reactive to sudden positive shocks, which are often followed by corrections. This pattern highlights the importance of closely monitoring market conditions and the potential for quick adjustments in risk management strategies. The results underscore the necessity for dynamic risk models that can account for rapid fluctuations and their subsequent impacts on VaR, ensuring more robust risk mitigation in highly volatile environments.

Table 5 shows the results of the LR test for correct coverage at 90% confidence. The test results for 95% and 99% confidence levels are stored in the Appendix. They have qualitatively similar insights.

Table 5: P-Values of Correct Coverage Test for VaR at 90%

Each value represents the p-values of the LR test for correct coverage at a 90% confidence. Values colored in green implies significant statistical proof for correct coverage. More details on the test are stored in Methodology section.

	Horizons (Months)	Bench.	PCA	SPC	PC2	KPCA poly	KPCA sigmoid	KPCA rbf
Food	**1**	0.068	0.945	0.333	0.6	0.236	0.945	0.945
	6	0.765	0.871	0.945	0.6	0.69	0.871	0.871
	12	0.161	0.259	0.333	0.008	0.455	0.259	0.259
Beverage	**1**	0.871	0.523	0.871	0.871	0.523	0.378	0.523
	6	0.057	0.378	0.378	0.945	0.057	0.523	0.523
	12	0.765	0.69	0.765	0.945	0.765	0.69	0.69
Metal	**1**	0.333	0.333	0.042	0.161	0.236	0.333	0.333
	6	0.6	0.106	0.042	0.014	0.042	0.106	0.068
	12	0.333	0.333	0.042	0	0.008	0.333	0.333
Pre. Metal	**1**	0.259	0.871	0.69	0.259	0.523	0.69	0.871
	6	0.6	0.101	0.871	0.945	0.69	0.057	0.057
	12	0.101	0.106	0.871	0.455	0.945	0.106	0.106
Gas	**1**	0.106	0.333	0.945	0.945	0.6	0.333	0.333
	6	0	0.004	0	0.002	0	0.001	0.002
	12	0	0	0	0	0	0	0
Coal	**1**	0.455	0.6	0.6	0.6	0.6	0.945	0.6
	6	0.236	0.765	0.106	0.333	0.765	0.765	0.6
	12	0.03	0.101	0.455	0.6	0.69	0.167	0.101
Oil	**1**	0.161	0.945	0.236	0.523	0.871	0.945	0.945
	6	0.106	0.455	0.333	0.945	0.455	0.333	0.333
	12	0.001	0.004	0	0.004	0	0.002	0.004

First, considering gas, previous conclusions indicated it is challenging to model and forecast. Even though all models can produce correct VaR coverage at 1 month and 90% confidence, these models, including the benchmark, do not seem to achieve correct coverage at higher confidence levels or further horizon. For oil, similar patterns can be observed. Even if short to mid-term VaR forecasts are plausible at 90% and 95% confidence levels, correct coverage does not seem achievable. Difficulties of predicting correct volatility is reflecting in these commodities VaR. However, in contrast to previous clustering, precious metals VaR appears to achieve correct coverage. This may follow from precious metals being used in banks and thus monitored and not only due to match in supply and demand like Oil or Gas.

Nevertheless, for the remaining commodities, coverage have overall significance. Similar to return forecast MSPEs, beverage, and food VaR show correct coverage. Once again PCA models, along with KPCArbf and KPCAsigmoid, return significant VaR coverage most consistently.

5 Conclusion

The analysis reveals several significant findings concerning the efficacy of Kernel Principal Component modeling in estimating returns across various commodities. The study highlights the notable significance of complex models in short-term forecasts, while in the long run, benchmark models utilizing historical averages proved challenging to surpass. Given the considerable volatility inherent in commodity trading, influenced by a multitude of factors, further research might benefit from focusing on even shorter time horizons. Moreover, incorporating macroeconomic variables could enhance the precision of short-term reaction analysis to significant shocks. Links to influencing factors are more pronounced in short term reactions.

Additionally, Principal Component demonstrates notable enhancement in prediction accuracy of the commodity returns compared to the historical average benchmark. Nevertheless, incorporating complex models using a Kernel trick did not substantially augment the overall capabilities beyond those of PCA. Given the computational demands and resource requirements of such complex models, the simplicity and efficiency of PCA may be preferable. Nonetheless, kPCA helps scale reactions during financial crisis and may be preferred in such times.

In addition, the research sheds light on the movement of various commodities within clusters. Gaz returns proved difficult to improve, literature explains the need of external factors, further supporting the need of additional variable for further research. Then, precious metals, even with difficulties to outperformed in literature, complex models helps slightly decrease the forecast error in the short run. For soft commodities, Beverage and Food return forecasts can be improved over the benchmark by more than 10% and 23% respectively. Finally, Metal, Coil and Oil showed improvements, even for mid term forecasts.

Then, the models show mixed improvement in volatility forecasts compared to the benchmark. MSPE as a factor is not very insightful, but incorporating a proxy with a loss function shows some improvement in complex models with negative t-statistics. Although not statistically significant, these results suggest the models are progressing positively. However, this still allowed for correct VaR coverage for most commodities at a 90% confidence level across all models. Gas and oil stand apart with less precise coverage for long-run forecasts and require more care when handled.

References

Bai, J. & Ng, S. (2017). Principal components and regularized estimation of factor models. *arXiv preprint arXiv:1708.08137*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Brownlees, C. T., Engle, R. F. & Kelly, B. T. (2011). A practical guide to volatility forecasting through calm and storm. *Available at SSRN 1502915*.

Cheung, Y.-W. & Lai, K. S. (1995). Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, *13*(3), 277–280.

Dickey, D. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431.

Dickey, D. & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*(4), 1057-72.

Duffie, D. & Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, *4*(3), 7–49.

Fally, T. & Sayre, J. (2018). *Commodity trade matters* (Tech. Rep.). National Bureau of Economic Research.

Ge, Y. & Tang, K. (2020). Commodity prices and gdp growth. *International Review of Financial Analysis*, *71*, 101512.

Girma, P. B. & Paulson, A. S. (1998). Seasonality in petroleum futures spreads. *Journal of Futures Markets (1986-1998)*, *18*(5), 581.

Hassani, H., Silva, E. S., Gupta, R. & Segnon, M. K. (2015). Forecasting the price of gold. *Applied Economics*, *47*(39), 4141–4152.

Hofmann, T., Schölkopf, B. & Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*.

Jarque, C. M. & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, *6*(3), 255–259.

Kupiec, P. H. et al. (1995). *Techniques for verifying the accuracy of risk measurement models* (Vol. 95) (No. 24). Division of Research and Statistics, Division of Monetary Affairs, Federal ....

Kutateladze, V. (2022). The kernel trick for nonlinear factor modeling. *International Journal of Forecasting*, *38*(1), 165–177.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68–78.

Nick, S. & Thoenes, S. (2014). What drives natural gas prices?—a structural var approach. *Energy Economics*, *45*, 517–527.

Palaskas, T. B. & Varangis, P. N. (1991). *Is there excess co-movement of primary commodity prices: a co-integration test* (Vol. 758). World Bank Publications.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246-256. (Realized Volatility)

Pettenuzzo, D. & Timmermann, A. (2017). Forecasting macroeconomic variables under model instability. *Journal of business & economic statistics*, *35*(2), 183–201.

Samuelson, P. A. (2016). Proof that properly anticipated prices fluctuate randomly. In *The world scientific handbook of futures markets* (pp. 25–38). World Scientific.

Stock, J. H. & Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, *39*, 3–33.

Teräsvirta, T., Tjøstheim, D. & Granger, C. W. (1994). Aspects of modelling nonlinear time series. *Handbook of econometrics, ch48*, *4*, 2917–2957.

Tolmasky, C. & Hindanov, D. (2002). Principal components analysis for correlated curves and seasonal commodities: The case of the petroleum market. *Journal of Futures Markets*, *22*(11), 1019–1035.

Ye, M., Zyren, J. & Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, *21*(3), 491-501.

A Bonus Analysis

This section highlights significant improvements that support the previous claims while addressing a broader scope. Specifically, this short analysis focuses on one-month-ahead forecasts, as earlier research concluded that complexity appears in the short run, with one month being the shortest timespan available. The difference from previous analysis arises from the inputs used for modeling the PCAs, utilizing the entire dataset provided by the IMF. This significantly increased computation time and power required but the scope is reduced to only Oil as it showed possibility of improvement and due to its importance in the economy.

The first observation is that the Mean Squared Prediction Errors (MSPEs) decrease significantly. The PCA model reduces MSPE by more than 60% and the KPCA model using the rbf function reduces MSPE by more than 70% compared to the benchmark MSPE. (relative MSPE to the benchmark are 0.386, and 0.279 respectively) Figure 8 and 9 illustrate the CSSE and helps clarify the differences.

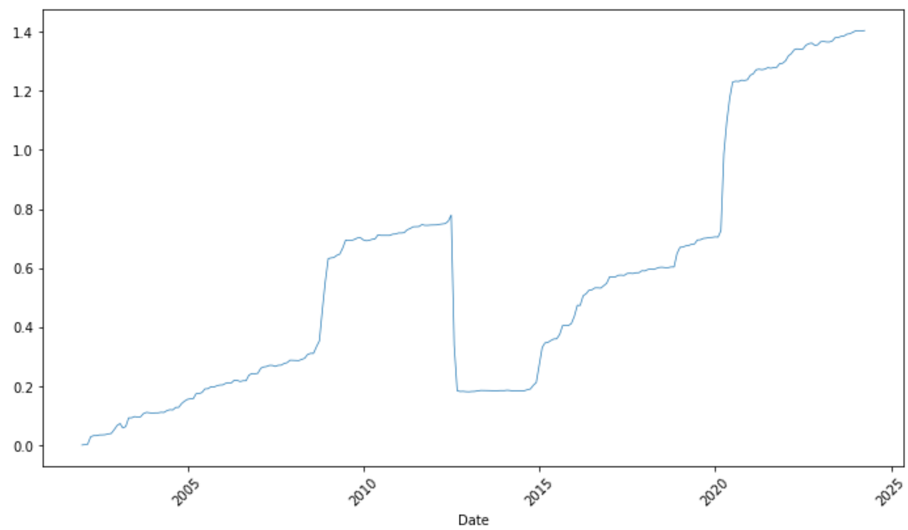


Figure 8: CSSE of the Benchmark vs PCA on 1m Ahead Forecasts

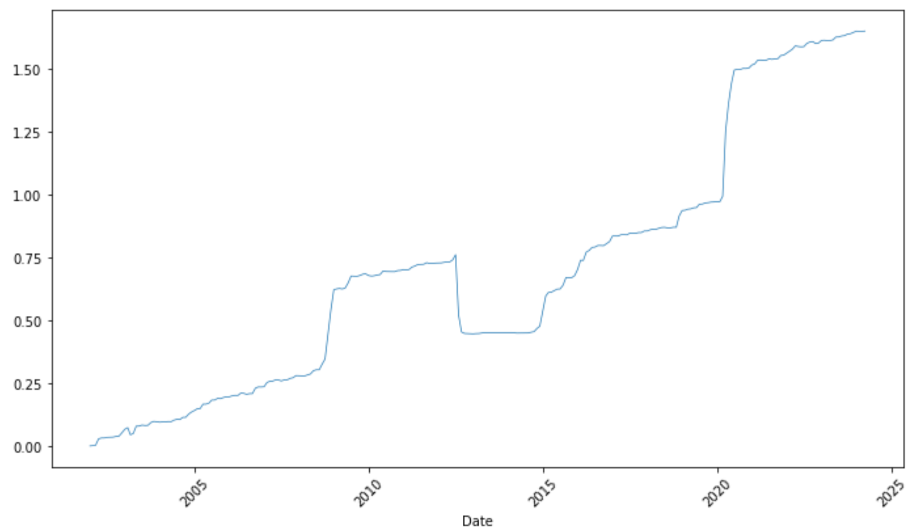


Figure 9: CSSE of the Benchmark vs KPCA rbf on 1m Ahead Forecasts

The figures indicate overall increasing trends, demonstrating the success of the models in reducing forecast errors. The slope is notably steeper than previous analysis highlighting the significant improvement. Notably, there are steep increases during financial crises, highlighting the models' high capacities in highly volatile environments. However, compared to previous analyses, both models appear to be less efficient in 2012.

Figure 10 compares the performance of the different PCA models, with notable differences emerging in 2012. This suggests that the more complex model (kPCA) provides more precise forecasts by reacting less aggressively. Both models captures factors that deviate the forecast from the mean, although the benchmark scales it, allowing it to outperform in this specific 2012 situation.

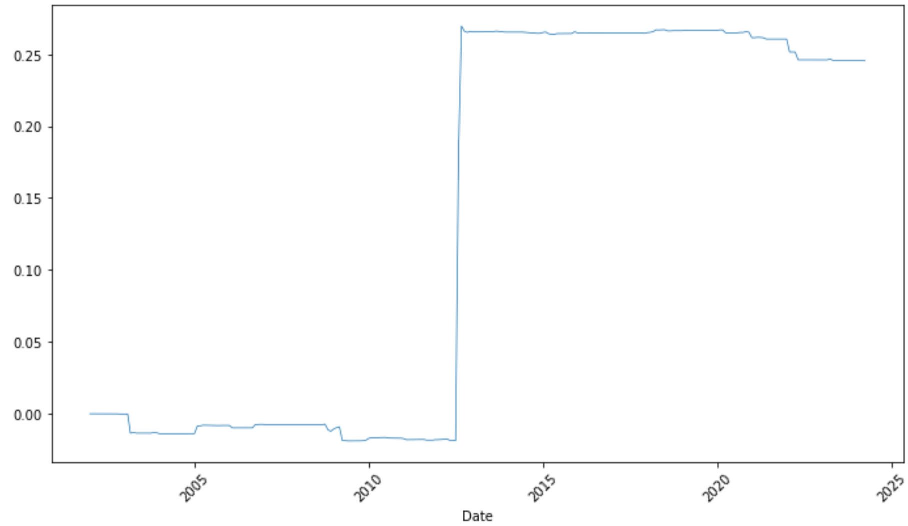


Figure 10: CSSE of the PCA vs KPCA rbf on 1-Month Ahead Forecasts

B Extra Tables & Figures

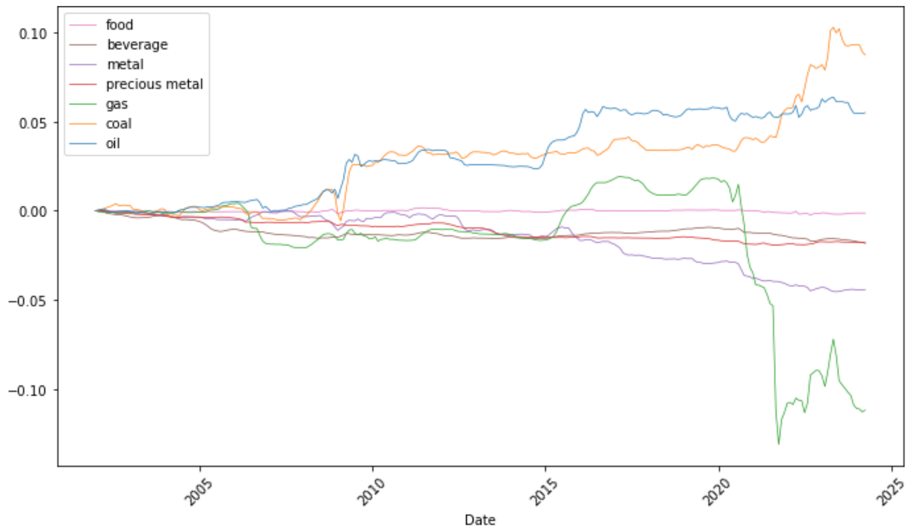


Figure 11: CSSE of the Benchmark vs PCA on 6-Months Ahead Forecasts

Table 6: Relative MSPEs of Volatility Forecasts

Each value represents the ratio of overall out-of-sample MSPE of volatility from the different models compare to the benchmark. Values in green, lower than one, imply lower forecasts errors from the model. Values in red imply the opposite. The forecasts are h-step ahead and are computed from a rolling window of length 120 months minus the horizon.

	Horizons (Months)	**PCA**	**SPC**	**PC2**	**KPCA poly**	**KPCA sigmoid**	**KPCA rbf**
Food	**1**	0.490	1.125	0.760	0.969	0.488	0.488
	6	1.217	2.272	0.950	3.894	1.211	1.156
	12	0.881	1.898	1.915	1.695	0.915	0.881
Beverage	**1**	0.717	1.333	0.886	1.881	0.704	0.711
	6	0.713	1.925	3.287	5.700	0.696	0.710
	12	2.545	2.345	5.000	21.273	2.545	2.491
Metal	**1**	0.422	1.687	0.960	3.410	0.421	0.419
	6	0.508	3.549	0.828	17.611	0.503	0.502
	12	0.688	2.847	2.591	7.179	0.760	0.706
Pre. Metal	**1**	1.499	6.247	4.629	3.557	1.483	1.326
	6	0.675	1.619	1.738	1.553	0.632	0.675
	12	1.358	3.716	2.418	2.627	1.343	1.328
Gas	**1**	1.529	2.247	4.371	4.482	1.388	1.531
	6	1.569	1.960	4.054	3.075	1.552	1.562
	12	0.681	2,305	3.024	5.893	0.714	0.672
Coal	**1**	1.067	0.823	1.172	1.252	1.086	1.049
	6	0.587	0.753	0.794	0.815	0.499	0.518
	12	1.059	1509	1.050	1.428	1.081	1.069
Oil	**1**	0.385	1.382	1.525	2.786	0.385	0.384
	6	1.435	3.310	2,609	5,001	1.426	1.422
	12	1.007	1.372	5.152	2,052	0.970	0.983

Table 7: Diebold–Mariano t-statistics of equal predictive accuracy

Each value represents the t-statistic of the DM test equal predictive accuracy. A t-statistic greater than 1.96 in absolute value indicates rejection of the null hypothesis of equal predictive accuracy at the 0.05 significance level. A negative t-statistic suggests that the model's forecast resulted in lower average loss compared to the benchmark forecast.

	Horizons (months)	**PCA**	**SPC**	**PC2**	**KPCA poly**	**KPCA sigmoid**	**KPCA rbf**
	1	-0.353	-0.590	-0.345	-0.483	-0.350	-0.350
food	**6**	-0.376	-0.560	-0.441	-0.674	-0.375	-0.368
	12	-0.114	-0.566	-0.705	-0.368	-0.122	-0.116
	1	-0.155	-0.392	-0.201	-0.401	-0.158	-0.156
beverage	**6**	-0.043	-0.407	-0.309	-0.742	-0.032	-0.038
	12	-0.242	-0.244	-0.562	-0.629	-0.239	-0.236
	1	-0.466	-0.616	-0.562	-0.773	-0.467	-0.465
metal	**6**	-0.518	-1.228	-0.784	-1.478	-0.526	-0.522
	12	-0.718	-1.160	-1.605	-1.604	-0.738	-0.737
	1	-0.455	-0.577	-0.498	-0.559	-0.440	-0.446
pre. metal	**6**	-0.233	-0.338	-0.559	-0.484	-0.228	-0.234
	12	-0.264	-0.161	-0.454	-0.177	-0.250	-0.258
	1	-0.424	-0.642	-1.178	-1.141	-0.423	-0.412
gas	**6**	-0.683	-0.719	-0.906	-1.247	-0.700	-0.632
	12	-0.300	-1.681	-1.097	-1.519	-0.321	-0.295
	1	-0.664	-0.506	-0.523	-0.608	-0.662	-0.662
coal	**6**	-0.167	-0.315	-0.181	-0.344	-0.171	-0.170
	12	-0.619	-0.984	-0.742	-0.734	-0.621	-0.622
	1	-0.685	-1.081	-0.969	-1.074	-0.686	-0.686
oil	**6**	-0.641	-1.228	-1.525	-1.678	-0.618	-0.571
	12	-0.598	-1.123	-1.020	-1.503	-0.577	-0.591

Table 8: P-Values of Correct Coverage Test for VaR at 95%

Each value represents the p-values of the LR test for correct coverage at a 95% confidence. Values colored in green implies significant statistical proof for correct coverage. More details on the test are stored in Methodology section.

	Horizons (months)	Bench.	PCA	SPC	PC2	KPCA poly	KPCA sigmoid	KPCA rbf
Food	**1**	0.013	0.607	0.163	0.837	0.416	0.607	0.607
	6	0.094	0.268	0.094	0.026	0.026	0.268	0.268
	12	0.013	0.607	0.094	0.003	0.163	0.607	0.607
Beverage	**1**	0.094	0.837	0.837	0.911	0.436	0.837	0.837
	6	0.133	0.133	0.436	0.66	0.006	0.133	0.133
	12	0.258	0.911	0.416	0.911	0.416	0.911	0.911
Metal	**1**	0.163	0.026	0.001	0.013	0.013	0.013	0.013
	6	0.268	0.094	0.026	0.026	0.163	0.094	0.051
	12	0.051	0.268	0.003	0	0.001	0.268	0.268
Pre. Metal	**1**	0.436	0.436	0.258	0.059	0.059	0.436	0.436
	6	0.911	0.837	0.607	0.416	0.607	0.837	0.837
	12	0.133	0.013	0.268	0.051	0.094	0.006	0.013
Gas	**1**	0.006	0.026	0.026	0.163	0.094	0.026	0.026
	6	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0
Coal	**1**	0.163	0.66	0.094	0.607	0.416	0.66	0.66
	6	0.013	0.094	0.607	0.013	0.094	0.094	0.094
	12	0.133	0.133	0.911	0.607	0.133	0.258	0.133
Oil	**1**	0.006	0.163	0.026	0.268	0.416	0.163	0.268
	6	0	0.006	0.001	0.051	0.013	0.003	0.006
	12	0	0	0	0	0	0	0

Table 9: P-Values of Correct Coverage Test for VaR at 99%

Each value represents the p-values of the LR test for correct coverage at a 99% confidence. Values colored in green implies significant statistical proof for correct coverage. More details on the test are stored in Methodology section.

	Horizons (Months)	Bench.	PCA	SPC	PC2	KPCA poly	KPCA sigmoid	KPCA rbf
Food	**1**	0.543	0.232	0.083	0.543	0.232	0.232	0.232
	6	0.002	0.026	0	0.002	0.007	0.026	0.026
	12	0	0.083	0	0.002	0.002	0.083	0.083
Beverage	**1**	0.543	0.961	0.961	0.232	0.961	0.543	0.961
	6	0.041	0.041	0.961	0.543	0.961	0.041	0.041
	12	0.543	0.083	0.083	0.543	0.026	0.026	0.083
Metal	**1**	0.026	0.083	0.232	0.543	0.083	0.083	0.083
	6	0.002	0.083	0.002	0.026	0.083	0.083	0.232
	12	0.002	0.232	0.007	0.007	0	0.232	0.232
Pre. Metal	**1**	0.406	0.406	0.041	0.406	0.406	0.961	0.406
	6	0.232	0.083	0.026	0.002	0.007	0.083	0.083
	12	0.406	0.007	0.026	0	0.002	0.007	0.007
Gas	**1**	0.002	0	0	0.002	0.007	0	0
	6	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0
Coal	**1**	0.007	0.026	0.083	0.007	0.026	0.026	0.026
	6	0.002	0.007	0.002	0.026	0.002	0.026	0.026
	12	0.041	0.961	0.961	0.083	0.406	0.961	0.961
Oil	**1**	0	0.007	0.026	0.026	0.083	0.026	0.026
	6	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0

At the 1% confidence level, the coverage is generally poor. This might be attributed to the sample size of about 200, where deviations from two positive indicator functions may vary significantly, even if the coverage is technically correct. More data would be required to accurately identify its statistical significance.

C Explanation of the Accompanying ZIP File

C.1 Inputs

The inputs used for the research are stored in the Excel file titled 'IMF Commodities since 1990'. The original data extracted from the IMF is located on the 'External Sheet'. The sheet named 'Python' contains the dataset used for analysis in Python, which includes only the indices analyzed in the research. The 'Methodology' sheet contains all important information regarding the dataset. All of these files can be accessed on the IMF website, with the link provided in the 'Methodology' sheet.

C.2 Code

All code is run on an Intel i5 computer with 16.0 GB (15.3 GB usable) of RAM. To implement the code, the software environments Spyder is used with the programming language Python(3.11). The code successfully completed in less than an hour ($\sim 55min$).

The code is stored in two files. The first file, 'Main', contains the primary code that extracts the inputs, displays summary statistics, and formats the data for application of the method. This main code calls functions from the second file, 'Models'. The 'Models' file contains the code that applies the models to the returns as described in the methodology section of the research. Finally, the main file utilizes these models to complete the remaining methodology and generates outputs.

C.3 Outputs

The numerical outputs are saved in the Excel file 'ANALYSIS', and the plots are stored in the folder 'Plots'. All of these results are documented in the research paper.

D Replication of Original Thesis Paper

As this research is part of my Bachelor Thesis, for completeness, in this section I share the original output showing the improvement of the MSPE of macroeconomic variable from the use of non-linear PCA. In this case the PCA model is set as the benchmark.

Table 10: Relative MSPE of the Used Models on Macroeconomic Variables (PCA is the benchmark model)

	Horizons (Months)	SPC	PC2	kPCA poly	kPCA sigmoid	kPCA RBF
RPInc.	**1**	1,033	1,013	1,077	1,001	1,000
	3	1,319	1,174	2,297	0,989	0,997
	6	1,531	1,271	3,160	0,965	0,998
	9	1,858	1,086	2,535	0,883	0,932
	12	1,375	1,250	1,970	0,889	0,937
	18	1,165	38,160	2,206	0,838	0,927
	24	1,198	19,446	1,328	0,869	0,989
Civil. Employ.	**1**	1,026	0,987	1,431	1,002	1,000
	3	1,476	1,136	1,627	0,983	0,976
	6	1,638	1,218	2,737	0,924	0,945
	9	2,255	1,281	2,044	0,921	0,932
	12	1,916	1,478	1,798	0,908	0,938
	18	1,315	20,826	1,627	0,917	1,003
	24	1,362	386,273	1,670	0,914	0,970
Housing starts: Privately Owned	**1**	1,010	1,098	1,068	0,995	0,995
	3	1,553	1,189	1,347	0,974	0,959
	6	2,353	1,762	2,867	0,979	0,999
	9	1,918	1,841	1,639	0,958	0,922
	12	1,460	1,393	2,209	0,958	0,941
	18	1,219	67,184	2,202	0,945	0,894
	24	1,795	23,410	2,862	0,977	0,941
RPCons.	**1**	1,079	1,046	1,172	1,004	0,992
	3	1,310	1,158	2,099	0,990	0,989
	6	1,641	1,231	1,898	0,955	0,961
	9	2,034	1,318	1,896	0,931	0,936
	12	1,560	1,179	2,568	0,907	0,960
	18	1,136	11,539	2,820	0,913	0,962
	24	1,340	19,689	1,778	0,948	0,980
M1 Money Stock	**1**	1,066	1,345	1,966	1,007	0,989
	3	1,596	1,464	3,999	0,963	0,999
	6	1,371	1,133	2,665	0,927	0,966
	9	1,544	1,199	2,626	0,918	0,957
	12	1,710	2,520	2,526	0,931	0,976
	18	1,484	27,087	2,106	0,898	0,952
	24	1,317	157,372	1,601	0,830	0,983

Table 11: (Continued) Relative MSPE of the Used Models on Macroeconomic Variables (PCA is the benchmark model)

	Horizons (Months)	**SPC**	**PC2**	**kPCA poly**	**kPCA sigmoid**	**kPCA RBF**
Effect. Fed. Funds Rate	**1**	0,926	1,228	1,317	0,992	1,039
	3	1,482	1,347	2,258	0,925	0,978
	6	3,126	1,398	3,885	0,836	0,954
	9	2,674	2,336	2,824	0,850	0,912
	12	2,417	1,220	1,330	0,865	0,962
	18	1,771	9,057	2,841	0,809	0,933
	24	2,223	23,787	1,674	0,795	0,891
CPI	**1**	1,044	1,193	1,629	0,990	0,997
	3	2,034	1,660	3,014	1,019	1,060
	6	2,272	1,205	3,915	0,942	1,010
	9	1,361	1,932	1,955	0,966	1,014
	12	1,323	1,548	1,475	0,954	0,980
	18	1,352	12,581	1,319	0,952	0,950
	24	1,317	34,103	1,233	0,954	0,956
S.P 500	**1**	1,123	1,113	1,619	0,989	1,000
	3	1,333	1,279	1,748	0,968	0,988
	6	2,031	1,429	2,533	0,934	0,970
	9	2,088	1,681	2,350	0,871	0,968
	12	2,088	1,981	2,038	0,863	0,945
	18	1,585	41,795	1,420	0,813	0,832
	24	2,044	43,136	1,340	0,875	0,944