Conversation on Predicting TTEF and ENI.MI Returns

Introduction

This document explores the prediction of TTEF returns using oil futures and the development of features to capture the mean reversion of TTEF using only oil data. Additionally, it discusses the application of these concepts to ENI.MI returns.

Capturing Mean Reversion in TTEF Using Oil Futures

To capture mean reversion in TTEF (Toronto-Traded Energy Fund) using only oil futures data, the goal is to detect when TTEF deviates significantly from its expected value based on oil prices, assuming a long-term relationship between the two.

Feature: Z-score of the Spread Between TTEF and Oil Futures

1. **Step 1: Assume a Linear Relationship**

   Model TTEF as linearly dependent on oil futures price:

   $$TTEF_t = \alpha + \beta \cdot Oil_t + \epsilon_t$$

2. **Step 2: Estimate the Spread (Residual)**

   Fit a simple linear regression between historical TTEF and oil futures prices. Compute the residual (spread):

   $$Spread_t = TTEF_t - (\hat{\alpha} + \hat{\beta} \cdot Oil_t)$$

   This spread is a candidate for a mean-reverting signal.

3. **Step 3: Compute the Z-score of the Spread**

$$Z_t = \frac{Spread_t - \mu_{spread}}{\sigma_{spread}}$$

Where: μ_{spread} and σ_{spread} are the rolling mean and standard deviation of the spread (e.g., 30-day window).

Use as a Feature

Use the Z-score of the spread as a feature in the TTEF return model. A high absolute Z-score suggests potential mean reversion (e.g., if TTEF is much higher or lower than predicted by oil, it may revert).

Alternatives & Enhancements

- Kalman Filter Regression (for time-varying alpha/beta)
- Cointegration Test (Johansen/Engle-Granger) to confirm long-term relationship
- Oil returns vs. TTEF returns spread
- Construct a rolling regression of returns instead of prices.
- Residual ADF test to confirm mean reversion of spread

Predicting Mean-Reverting Returns of TTEF

The return of TTEF is already mean-reverting, and the goal is to predict this behavior using only oil futures data.

Goal Recap

- Target: Return of TTEF, which is known to be mean-reverting.
- Constraint: Use only oil futures data to generate features.
- Objective: Create a feature from oil futures that aligns with or helps predict this mean-reverting behavior in TTEF.

Best Feature Option: Oil Return Deviation from Equilibrium Level

Since TTEF is influenced by oil and mean-reverts, build a proxy for the "equilibrium" oil price level that TTEF tends to revert to.

Feature: Oil Return Z-Score (Short vs. Long Window)

This captures mean-reversion pressure in oil, which may mirror the behavior in TTEF due to their correlation.

1. Define returns:
$$r_t^{oil} = \log(oil_t) - \log(oil_{t-1})$$
2. Compute short-term and long-term moving averages:
$$\mu_{short}(r^{oil}), \mu_{long}(r^{oil}), \sigma_{long}(r^{oil})$$
3. Z-score of short-term return relative to long-term:
$$Z_t^{oil} = \frac{\mu_{short}(r^{oil}) - \mu_{long}(r^{oil})}{\sigma_{long}(r^{oil})}$$

This Z-score indicates how "stretched" oil is and therefore how much mean-reversion pressure might be acting on TTEF returns.

Summary of the Feature

Feature Name	Description
oil_return_zscore	Z-score of short-term oil return vs long-term average return

Alternative Ideas (also oil-only)

- Oil Momentum Reversal Indicator
- Price-to-MA Ratio of Oil
- Oil Term Structure Slope (if futures curve available)

Volatility-Adjusted Oil Misalignment Score (VOMS)

A custom indicator that captures how extreme oil price moves are, adjusted for recent volatility, and normalized to avoid overreacting to peaks.

VOMS Formula

Let:

$$r_t^{oil} : \text{daily log return of oil}$$
$$\mu_N^{oil} : \text{long-term average oil return (e.g. 20–30 days)}$$
$$\sigma_N^{oil} : \text{standard deviation of oil returns over same window}$$

Then define:

$$VOMS_t = \tanh\left(\frac{r_t^{oil} - \mu_N^{oil}}{\sigma_N^{oil}}\right)$$

The tanh function compresses extreme values, ensuring robustness under volatility spikes. The score ranges from -1 to +1, indicating overbought/oversold conditions in oil, with mean-reversion implications for ENI.

Why This Works for ENI.MI

- ENI is tied to energy/oil fundamentals and tends to mean-revert around oil-driven fair value.
- When oil return is far from recent average but not explosively volatile, ENI may respond by reverting its own return toward a fair value.
- The VOMS score indicates how misaligned the current oil return is with its history in a controlled, volatility-adjusted way.

Summary of the Feature

Feature Name	Description
VOMS	Volatility-Adjusted Oil Misalignment Score

Alternative Ideas

- Skewness-Based Oil Return Features
- Multi-Horizon Normalized Return Strength (Trend Strength)
- MIDAS-Level Reversion Zones
- Volume-Aware Volatility-Reversion Alpha
- Bollinger Band + RSI Filter

Using PCA in Multi-Target Regression

Using PCA (Principal Component Analysis) in a multi-target regression setup like predicting ENI and TTEF returns can exploit shared structure between the two assets.

Why Use PCA in Multi-Target Regression?

- ENI and TTEF likely share exposure to oil prices, broader energy sector trends, and European equity risk premia.
- Compress their joint behavior into orthogonal components.
- Predict those components (which may be more stationary and predictable).
- Reconstruct the original targets from the predicted components.

Step-by-Step PCA for Multi-Target Regression

1. Apply PCA to your targets.
2. Train LightGBM to predict the PCA components.
3. Predict and transform back.

Benefits of Using PCA

- Dimensionality reduction
- Captures co-movement
- Smoother targets
- Noise filtering

Optional Enhancements

- Drop the 2nd PCA component if it adds mostly noise.
- Use PCA on residuals instead of raw targets.
- Apply PCA across time as features.

Caveats

- PCA is linear: works best when ENI and TTEF are linearly related.
- You must invert-transform predictions, which adds slight numerical error.
- PCA components don't have economic meaning (latent), so interpretability is lower.

[12pt]article amsmath amsfonts graphicx hyperref enumitem geometry array booktabs
margin=1in
Mean-Reversion Feature Engineering Using Oil Futures to Predict ENI and TTEF Returns

Objective

We aim to predict the return of TTEF (and potentially ENI.MI) using only oil futures data. TTEF returns are known to be mean-reverting, and the goal is to find oil-derived features that reflect this behavior.

Feature Engineering to Capture Mean Reversion

Z-Score of the Spread Between TTEF and Oil Futures

Step 1: Assume a Linear Relationship

$$\text{TTEF}_t = \alpha + \beta \cdot \text{Oil}_t + \varepsilon_t$$

Step 2: Estimate the Spread (Residual)

$$\text{Spread}_t = \text{TTEF}_t - (\hat{\alpha} + \hat{\beta} \cdot \text{Oil}_t)$$

Step 3: Compute the Z-Score of the Spread

$$Z_t = \frac{\text{Spread}_t - \mu_{\text{spread}}}{\sigma_{\text{spread}}}$$

where μ_{spread} and σ_{spread} are the rolling mean and standard deviation of the spread.

Use as a Feature

The Z-score of the spread can be used as a feature in the TTEF return model. A high absolute Z-score suggests potential mean reversion.

Alternatives & Enhancements

- Kalman Filter Regression (for time-varying alpha/beta)
- Cointegration Test (Johansen/Engle-Granger) to confirm long-term relationship
- Oil returns vs. TTEF returns spread
- Construct a rolling regression of returns instead of prices.
- Residual ADF test to confirm mean reversion of spread

Predicting Mean-Reverting Returns of TTEF

Goal Recap

- **Target:** Return of TTEF, which is known to be mean-reverting.
- **Constraint:** Use only oil futures data to generate features.
- **Objective:** Create a feature from oil futures that aligns with or helps predict this mean-reverting behavior in TTEF.

Best Feature Option: Oil Return Deviation from Equilibrium Level

Feature: Oil Return Z-Score (Short vs. Long Window)

1. Define returns:
$$r_t^{\text{oil}} = \log(\text{oil}_t) - \log(\text{oil}_{t-1})$$
2. Compute short-term and long-term moving averages:
$$\mu_{\text{short}}(r^{\text{oil}}), \quad \mu_{\text{long}}(r^{\text{oil}}), \quad \sigma_{\text{long}}(r^{\text{oil}})$$
3. Z-score of short-term return relative to long-term:
$$Z_t^{\text{oil}} = \frac{\mu_{\text{short}}(r^{\text{oil}}) - \mu_{\text{long}}(r^{\text{oil}})}{\sigma_{\text{long}}(r^{\text{oil}})}$$

Summary of the Feature

Feature Name	Description
`oil_return_zscore`	Z-score of short-term oil return vs long-term average return

Alternative Ideas

- Oil Momentum Reversal Indicator
- Price-to-MA Ratio of Oil
- Oil Term Structure Slope (if futures curve available)

Volatility-Adjusted Oil Misalignment Score (VOMS)

VOMS Formula:
$$\text{VOMS}_t = \tanh\left(\frac{r_t^{\text{oil}} - \mu_N^{\text{oil}}}{\sigma_N^{\text{oil}}}\right)$$

Why This Works for ENI.MI

- ENI is tied to energy/oil fundamentals and tends to mean-revert around oil-driven fair value.
- The VOMS score indicates how misaligned the current oil return is with its history in a controlled, volatility-adjusted way.

Summary of the Feature

Feature Name	Description
`VOMS`	Volatility-Adjusted Oil Misalignment Score

Using PCA in Multi-Target Regression

Why Use PCA in Multi-Target Regression?

- ENI and TTEF likely share exposure to oil prices, broader energy sector trends, and European equity risk premia.
- Compress their joint behavior into orthogonal components.
- Predict those components (which may be more stationary and predictable).
- Reconstruct the original targets from the predicted components.

Step-by-Step PCA for Multi-Target Regression

1. Apply PCA to your targets.
2. Train LightGBM to predict the PCA components.
3. Predict and transform back.

Benefits of Using PCA

- Dimensionality reduction
- Captures co-movement
- Smoother targets
- Noise filtering

Optional Enhancements

- Drop the 2nd PCA component if it adds mostly noise.
- Use PCA on residuals instead of raw targets.
- Apply PCA across time as features.

Caveats

- PCA is linear: works best when ENI and TTEF are linearly related.
- You must invert-transform predictions, which adds slight numerical error.
- PCA components don't have economic meaning (latent), so interpretability is lower.

[12pt]article amsmath amsfonts graphicx hyperref enumitem geometry array booktabs

margin=1in

Mean-Reversion Feature Engineering Using Oil Futures to Predict ENI and TTEF Returns

Objective

We aim to predict the return of TTEF (and potentially ENI.MI) using only oil futures data. TTEF returns are known to be mean-reverting, and the goal is to find oil-derived features that reflect this behavior.

Feature Engineering to Capture Mean Reversion

Z-Score of the Spread Between TTEF and Oil Futures

Step 1: Assume a Linear Relationship

$$\text{TTEF}_t = \alpha + \beta \cdot \text{Oil}_t + \varepsilon_t$$

Step 2: Estimate the Spread (Residual)

$$\text{Spread}_t = \text{TTEF}_t - (\hat{\alpha} + \hat{\beta} \cdot \text{Oil}_t)$$

Step 3: Compute the Z-Score of the Spread

$$Z_t = \frac{\text{Spread}_t - \mu_{\text{spread}}}{\sigma_{\text{spread}}}$$

where μ_{spread} and σ_{spread} are the rolling mean and standard deviation of the spread.

Use as a Feature

The Z-score of the spread can be used as a feature in the TTEF return model. A high absolute Z-score suggests potential mean reversion.

Alternatives & Enhancements

- Kalman Filter Regression (for time-varying alpha/beta)
- Cointegration Test (Johansen/Engle-Granger) to confirm long-term relationship
- Oil returns vs. TTEF returns spread
- Construct a rolling regression of returns instead of prices.
- Residual ADF test to confirm mean reversion of spread

Predicting Mean-Reverting Returns of TTEF

Goal Recap

- **Target:** Return of TTEF, which is known to be mean-reverting.
- **Constraint:** Use only oil futures data to generate features.
- **Objective:** Create a feature from oil futures that aligns with or helps predict this mean-reverting behavior in TTEF.

Best Feature Option: Oil Return Deviation from Equilibrium Level

Feature: Oil Return Z-Score (Short vs. Long Window)

1. Define returns:
$$r_t^{\text{oil}} = \log(\text{oil}_t) - \log(\text{oil}_{t-1})$$

2. Compute short-term and long-term moving averages:
$$\mu_{\text{short}}(r^{\text{oil}}), \quad \mu_{\text{long}}(r^{\text{oil}}), \quad \sigma_{\text{long}}(r^{\text{oil}})$$

3. Z-score of short-term return relative to long-term:
$$Z_t^{\text{oil}} = \frac{\mu_{\text{short}}(r^{\text{oil}}) - \mu_{\text{long}}(r^{\text{oil}})}{\sigma_{\text{long}}(r^{\text{oil}})}$$

Summary of the Feature

Feature Name	Description
`oil_return_zscore`	Z-score of short-term oil return vs long-term average return

Alternative Ideas

- Oil Momentum Reversal Indicator
- Price-to-MA Ratio of Oil
- Oil Term Structure Slope (if futures curve available)

Volatility-Adjusted Oil Misalignment Score (VOMS)

VOMS Formula:
$$\text{VOMS}_t = \tanh\left(\frac{r_t^{\text{oil}} - \mu_N^{\text{oil}}}{\sigma_N^{\text{oil}}}\right)$$

Why This Works for ENI.MI

- ENI is tied to energy/oil fundamentals and tends to mean-revert around oil-driven fair value.
- The VOMS score indicates how misaligned the current oil return is with its history in a controlled, volatility-adjusted way.

Summary of the Feature

Feature Name	Description
`VOMS`	Volatility-Adjusted Oil Misalignment Score

Using PCA in Multi-Target Regression

Why Use PCA in Multi-Target Regression?

- ENI and TTEF likely share exposure to oil prices, broader energy sector trends, and European equity risk premia.
- Compress their joint behavior into orthogonal components.
- Predict those components (which may be more stationary and predictable).
- Reconstruct the original targets from the predicted components.

Step-by-Step PCA for Multi-Target Regression

1. Apply PCA to your targets.
2. Train LightGBM to predict the PCA components.
3. Predict and transform back.

Benefits of Using PCA

- Dimensionality reduction
- Captures co-movement
- Smoother targets
- Noise filtering

Optional Enhancements

- Drop the 2nd PCA component if it adds mostly noise.
- Use PCA on residuals instead of raw targets.
- Apply PCA across time as features.

Caveats

- PCA is linear: works best when ENI and TTEF are linearly related.
- You must invert-transform predictions, which adds slight numerical error.
- PCA components don't have economic meaning (latent), so interpretability is lower.

Research Articles on Mean Reversion Features in Commodities Market

Academic & Peer-Reviewed Research

- **Mean-Reverting Statistical Arbitrage Strategies in Crude Oil Markets**: This study constructs mispricing portfolios across crude benchmarks (e.g., WTI, Brent, Dubai), tests cointegration, and implements strategies exploiting mean-reversion in the mispricing residuals and their predictability.
- **Momentum and Mean-Reversion in Commodity Spot and Futures Markets (Chaves & Viswanathan, 2016)**: Shows momentum trading works better in futures; mean-reversion strategies are more effective in spot markets. Also decomposes basis behavior to better understand relevancy for spreads/features.
- **Nonlinear Mean Reversion in Oil and Stock Markets (Jawadi & Bellalah, 2011)**: Demonstrates non-linear, asymmetric mean-reversion dynamics between oil and equity markets; such frameworks can produce features that respond differently to large vs small deviations.
- **Monte Carlo Analysis of Mean Reversion in Commodity Futures Prices (Irwin et al., 1996)**: Challenges asymptotic regression results on mean reversion in agricultural futures, cautioning about sample biases—important when designing statistical features on short horizons.
- **Time-Changed Ornstein–Uhlenbeck Processes for Commodity Prices (Li & Linetsky, 2012)**: Presents a stochastic model combining OU mean-reversion, Lévy jumps, and stochastic volatility—useful to derive features like instantaneous reversion speed or jump intensity.
- **Calibration & Filtering for Multi-Factor Commodity Models (Peters, Briers, Shevchenko et al., 2011)**: Explores latent factor models with mean-reverting trends plus seasonality, estimated via particle-filtered Kalman techniques—features like filtered deviations or reversion speed become measurable.

- **Deep Learning Regime-Switching for Energy Commodity Prices (2022)**: Combines deep neural nets with mean-reverting regimes to model commodity price dynamics—suggesting features like regime probability or network-inferred residuals could serve as robust mean-reversion indicators.

Relevant Mechanistic Models & Index Design

- **DBLCI Mean Reversion Index (Deutsche Bank)**: A fully rule-based commodity index that dynamically shifts weights when prices are ±5% off their 5-year moving average—capitalizing on mean-reversion across multiple commodity exposures.

- **Theory of Storage / Schwartz Model (1997)**: Describes mean-reversion of convenience yield and spot prices in commodity markets—this underpins features based on term-structure deviations or implied convenience yield gaps. https://quantpedia.com/strategies/momentum-effect-in-commodities?utm$_source = chatgpt.com$