README — Commodity Futures: Aligned Monthly Data and Forecasts

When using this dataset, please cite: Farag, M., S. Snudden, and G. Upton (2024) *Can Futures Prices Predict the Real Price of Primary Commodities?* LCERPA Working Paper 2024-3. https://www.lcerpa.org/files/LCERPA_2024_3.pdf

Purpose and Structure

This package provides two complementary sets of monthly futures data for 17 primary commodities. Both sets are intended for general research use. A separate replication package contains the exact subset of data and all code used in the paper.

1. **Futures Forecasts**
   These files contain the futures-based forecasts under five curve-construction assumptions. The data are organized so that each column provides a forecast sequence from a specific information set (end-of-month m–yyyy) and construction variant. This structure aligns with the out-of-sample forecast evaluation in the paper.

2. **Futures Data**
   These files contain the raw monthly alignment of commodity futures contracts by delivery horizon. Contracts are listed directly by contract number (e.g., CL1…CL130 for WTI), without interpolation for missing maturities. This version of the data is organized by contract and delivery date, not as forecast paths. It is useful for researchers who want to work directly with the underlying nominal futures prices before applying curve-construction or interpolation methods.

Commodities covered (17)

- *Energy*: WTI crude oil, Henry Hub natural gas, Heating Oil, Gasoline, Ethanol
- *Metals*: Copper, Aluminum, Nickel, Zinc, Tin, Lead, Gold, Silver, Platinum
- *Agriculture*: Wheat, Corn, Soybeans

How the data are provided

- There are 17 Excel workbooks (one per commodity), each titled "<commodity>futuresdata.xlsx".
- Both the Futures Forecasts folder and the Futures Data folder contain these 17 workbooks.
- Each workbook in both folders has the same five tabs, corresponding to the curve-construction assumptions listed below.

The difference lies in the content:

- In the Futures Forecasts folder, each tab contains forecast paths (f_fut#_m_yyyy) generated from the end-of-month futures curve under the corresponding assumption.
- In the Futures Data folder, each tab contains the raw contract-level futures prices aligned by delivery horizon (e.g., CL1…CL130 for WTI), before interpolation or forecast construction.

Tabs and mapping to paper tables (curve assumptions)

- Last_day_month — Table A.6, column 3 ("No adjustment").
- EoM_ADJ — Table 4 and Table A.6, column 2 ("Baseline" adjustment).
- 5day_Average — Table B.2 ("Five-Day End-of-Month Futures Average").
- Monthly_Average — Table 5 ("Monthly average of daily settlements").

- ADJ_BC12_LD — Table A.6, last column ("Adjustment (4)": applies average curvature from the first 12 contracts to the front contract).

Futures Forecasts Folder

Within each sheet, the columns are:

- time — Month reference date (YYYY-MM-01).
- x — Monthly spot average for the commodity (column name matches the commodity; e.g., wti, copper, wheat).
- x_lastday — End-of-month (EOM) spot; the closing price on the last trading day of the month.
- f_fut#_m_yyyy — Futures-based forecast series (see interpretation below).

Interpretation of the futures forecast columns (f_fut#_m_yyyy)

- "#" is a number that differentiates the curve-construction variant used within the tab's assumption (1…6).
- "m" is the calendar month (1…12) defining the information set used to construct the forecasts in that column.
- "yyyy" is the year defining the information set used to construct the forecasts in that column.
- Each f_fut#_m_yyyy column is the forecast path generated from the futures curve observed at the end of month m of year yyyy. The column provides the entire sequence of forecasts for horizons t+1 to T from that information set.
- Alignment: the first available observation in the column is the first out-of-sample forecast date. For example, f_fut5_1_1990 uses data available through the end of January 1990; its first forecast month is 1990-02-01 (the one-step-ahead forecast from that information set).
- Longer-horizon forecasts use observed futures prices where available; if a required futures settlement is not observed (e.g., due to contract availability), a no-change assumption on the observed futures contract is applied as detailed in the paper's Appendix.

Futures Data folder

This folder mirrors the Futures Forecasts folder in structure and naming. Each file has the same five tabs, but the contents differ: instead of forecast paths, the tabs report raw contract-level settlements. Within each tab, columns include:

- month — Numeric month (1–12).
- year — Four-digit year.
- data — Reference column for spot or settlement value (naming consistent across commodities).
- CL1…CLn — Futures contracts indexed by delivery horizon (e.g., CL1…CL130 for WTI).

These data are nominal and aligned to contract delivery dates. No interpolation has been applied for missing contracts, so gaps appear when contracts are unavailable. This version is intended for researchers who need to work with the raw panel of contracts rather than with constructed forecast paths.

Ticker/source metadata and sample periods

Tickers, markets, sources, and sample periods are listed in Table 3 of the paper ("Bloomberg Tickers for Commodity Futures, Sample Periods, and Contract Details"). Please consult that table when reproducing or extending the dataset and cite the original sources accordingly.

Usage notes for forecasting

- In the *Futures Forecasts* folder, choose the tab whose curve-construction assumption fits your research design or matches the specification you wish to emulate.
- Use the f_fut#_m_yyyy columns to select the information set (m, yyyy) and curve-variant (#) you require.
- The x and x_lastday spot columns are provided for reference and benchmarking; they are not forecasts.

For exact reproduction of the paper's results, use the separate replication package.

Terms of use

This dataset is available for academic use and not for resale. It is recommended that the dataset be distributed under a non-commercial Creative Commons license (e.g., CC-BY-NC). When using the data, please cite the working paper above and the original data sources listed in the paper.

Contact

Stephen Snudden
Wilfrid Laurier University
https://stephensnudden.com/data-code/
ssnudden@wlu.ca