ChatGPT

1. S&P 500 Index Daily Returns – *Source:* **Kaggle (Henry Han) and Yahoo Finance. This dataset provides daily historical prices of the S&P 500 index from 1927 to 2020 (over 23,500 trading-day observations)** [1] **. It can be downloaded as a CSV (e.g.,** `SPX.csv` **on Kaggle) containing dates and daily Open-High-Low-Close values. Time Interval: 1927–2020 (daily). Format: CSV. Description: A financial market time series representing U.S. stock market performance. Asymmetry: The distribution of daily returns is non-Gaussian with heavy tails and negative skew. For example, including the 1987 crash, sample skewness was measured around –2.39 (dropping to –0.26 if the crash day is excluded)** [2] **. This indicates a pronounced left-tail (large negative moves), so ARIMA residuals deviate from normality.** *Download link:* **Kaggle S&P 500 Historical Data** [1] **(no registration needed to access via the Kaggle API or Kaggle's dataset page).**

2. Crude Oil WTI Spot Prices – *Source:* **U.S. Energy Information Administration (EIA) via FRED. Daily West Texas Intermediate (WTI) crude oil prices at Cushing, OK are available from January 1986 to present (updated daily)** [3] **. The dataset (FRED series "DCOILWTICO") has 10,000+ daily observations and can be downloaded in CSV or Excel format from FRED/EIA** [4] [5] **. Time Interval: 1986–Present (daily). Format: CSV, XLS (via EIA/FRED API). Description: Economic commodity price series reflecting oil market fluctuations. Asymmetry: The distribution of daily returns is non-Gaussian and exhibits significant skewness. Empirical studies show crude oil returns (WTI and Brent) are negatively skewed – large price drops ("oil shocks") occur more than large upward moves** [6] **. This heavy left-tail implies ARIMA model innovations would be skewed (non-normal).** *Download link:* **FRED (series DCOILWTICO) – e.g. "Download Data" on the FRED page yields a CSV** [3] **.**

3. Nile River Annual Flow Volumes – *Source:* **Time Series Data Library (Hyndman) / R dataset. This classic dataset contains annual averages of the Nile River's flow volume measured at Aswan for years 1871–1970 (100 observations)** [7] **. Time Interval: 1871–1970 (yearly). Format: CSV (available via the R `datasets` package or the TSDL repository), also in Excel in academic sources** [7] **. Description: Natural hydrological time series (river discharge in 10^8 m³) with a known changepoint around 1898 due to dam construction** [7] **. Asymmetry: The series is non-Gaussian. A level shift (post-1898) and occasional extreme flood years indicate deviation from normality (skewed and heteroskedastic behavior). The presence of a structural break and outliers means ARIMA residuals would not be normally distributed (innovations show skew/heavy tails).** *Download link:* **[Rdatasets – Nile.csv](#) (public domain)** [7] **.**

4. Canadian Lynx Trappings (Annual Counts) – *Source:* **R `datasets` package (built-in) originally from Hudson's Bay Company records. This dataset records the annual number of lynx trapped in the Mackenzie River district of Canada from 1821–1934 (114 observations)** [8] **. Time Interval: 1821–1934 (yearly). Format: CSV (e.g., via [Rdatasets](#)), also included in R. Description: A famous ecological time series with cyclic boom-bust dynamics in lynx population. Asymmetry: The data are highly right-skewed (several years of explosive population booms). In fact, "the lynx data show very strong right-skewness," so a log-transformation is typically applied before ARIMA modeling** [9] **. This indicates non-Gaussian innovations – the raw residuals of an ARIMA fit would be far from normal without transforming the data.** *Download link:* **Included in R (`lynx` dataset)** [8] **; also available as CSV in the Rdatasets repository.**

5. England & Wales Precipitation (EWP) Series – *Source:* UK Met Office Hadley Centre. This is a monthly meteorological time series of area-averaged precipitation (rainfall+snowfall) over England and Wales from January 1766 to present (over 250 years; ~3,100 monthly observations) [10] **. Time Interval: 1766–2023 (monthly, with an annual aggregation available)** [11] **. Format: Plain text (CSV-like) available for download from the Met Office website** [11] **. Description: Long-term climate record used to study rainfall trends and variability. Asymmetry: Precipitation amounts have a heavily skewed distribution – most months are moderate, but a few months have extreme rainfall totals** [12] **. Annual totals range widely (e.g. ~600 mm in very dry years up to ~1200+ mm in wet years), yielding a right-skewed distribution of totals. This non-Gaussian character (many low-rainfall months and few very wet outliers) means ARIMA residuals would not be normally distributed** [12] **. *Download link:* Met Office HadUKP portal (e.g.** `HadEWP_monthly_totals.txt` **)** [11] **, freely accessible.**

Each of the above datasets is openly available and sufficiently long (>100 data points). Their documented skewness or heavy-tailed behavior suggests **non-Gaussian innovations**, meaning that while ARIMA models can be fitted, one should expect skewed or leptokurtic residuals rather than ideal Gaussian white noise [2] [6] . The sources provided include direct download links or repositories for obtaining the data in CSV/Excel format, along with evidence of asymmetry (e.g. skewness coefficients or notes in the literature).

---

1 S&P 500 Historical Data - Kaggle
https://www.kaggle.com/datasets/henryhan117/sp-500-historical-data

2 Skewness | Man Group
https://www.man.com/insights/skewness

3 4 Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (DCOILWTICO) | FRED | St. Louis Fed
https://fred.stlouisfed.org/series/DCOILWTICO

5 Cushing, OK WTI Spot Price FOB (Dollars per Barrel)
https://www.eia.gov/dnav/pet/hist/rwtcd.htm

6 bayes-cid.com
https://www.bayes-cid.com/pdf/issues/2024-winter/publications/Pages-30_35-CID-Winter-2024-Carnero-etal.pdf

7 8 Help for package datasets
https://mirrors.ibiblio.org/pub/mirrors/CRAN/doc/manuals/r-patched/packages/datasets/refman/datasets.html

9 Microsoft Word - ATSA-Scriptum-SS2015.docx
https://stat.ethz.ch/education/semesters/ss2015/atsa/ATSA_Scriptum_v1_SS15.pdf

10 Met Office Hadley Centre HadUKP Data Download
https://www.metoffice.gov.uk/hadobs/hadukp/data/download.html

11 www.metoffice.gov.uk
https://www.metoffice.gov.uk/hadobs/hadukp/data/monthly/HadEWP_monthly_totals.txt

12 Changes in Observed Daily Precipitation over Global Land Areas …
https://journals.ametsoc.org/view/journals/clim/34/1/jcliD190965.xml