

FRED-MD Macroeconomic Dataset: Complete Analysis Guide

1. Dataset Overview

FRED-MD (Federal Reserve Economic Data - Monthly Database) is a comprehensive macroeconomic dataset maintained by the Federal Reserve Bank of St. Louis.

Dataset Characteristics

- **Time Period:** January 1959 to August 2025 (800+ monthly observations)
- **Variables:** 127 macroeconomic indicators
- **Coverage:** 67 years of US economic data
- **Data Source:** Federal Reserve Economic Data (FRED)
- **Update Frequency:** Monthly

2. Key Economic Indicators Definitions

Core Economic Activity

- **RPI:** Real Personal Income - inflation-adjusted personal income
- **INDPRO:** Industrial Production Index - measures manufacturing output
- **PAYEMS:** Nonfarm Payrolls - total employment excluding agriculture
- **DPCERA3M086SBEA:** Real Personal Consumption Expenditures - consumer spending
- **CMRMTSPLx:** Real Manufacturing and Trade Industries Sales
- **RETAILx:** Retail Sales Index - consumer spending indicator

Labor Market Indicators

- **UNRATE:** Unemployment Rate - percentage of labor force unemployed
- **CLF16OV:** Civilian Labor Force (16+ years)
- **CE16OV:** Civilian Employment (16+ years)
- **UEMPMEAN:** Average Duration of Unemployment (weeks)
- **CLAIMSx:** Initial Unemployment Claims - weekly job losses indicator
- **AWOTMAN:** Average Weekly Overtime Hours (Manufacturing)
- **AWHMAN:** Average Weekly Hours (Manufacturing)

Housing Market

- **HOUST**: Housing Starts - new residential construction projects
- **PERMIT**: Building Permits - leading indicator for housing activity
- **HOUSTNE/MW/S/W**: Regional housing starts (Northeast, Midwest, South, West)

Financial Markets

- **S&P 500**: S&P 500 Stock Price Index
- **S&P div yield**: S&P 500 Dividend Yield
- **S&P PE ratio**: S&P 500 Price-to-Earnings Ratio
- **FEDFUNDS**: Federal Funds Rate - key monetary policy tool
- **TB3MS**: 3-Month Treasury Bill Rate
- **GS10**: 10-Year Treasury Constant Maturity Rate
- **AAA**: Moody's Seasoned AAA Corporate Bond Yield
- **BAA**: Moody's Seasoned BAA Corporate Bond Yield

Monetary Aggregates

- **M1SL**: M1 Money Supply - most liquid money measures
- **M2SL**: M2 Money Supply - broader money supply measure
- **M2REAL**: Real M2 Money Supply (inflation-adjusted)
- **TOTRESNS**: Total Reserves of Depository Institutions
- **NONBORRES**: Non-borrowed Reserves
- **BUSLOANS**: Commercial and Industrial Loans

Price Indices (Inflation Measures)

- **CPIAUCSL**: Consumer Price Index (All Urban Consumers)
- **PCEPI**: Personal Consumption Expenditures Price Index
- **PPICMM**: Producer Price Index (Commodities)
- **WPSFD49207**: Producer Price Index (Finished Goods)
- **OILPRICEx**: Crude Oil Prices

International Trade & Exchange

- **TWEXAFEGSMTHx**: Trade Weighted Exchange Index (Advanced Foreign Economies)
- **EXSZUSx**: Switzerland/US Exchange Rate
- **EXJPUSx**: Japan/US Exchange Rate
- **EXUSUKx**: US/UK Exchange Rate

- **EXCAUSx**: Canada/US Exchange Rate

Business Confidence & Sentiment

- **UMCSENTx**: University of Michigan Consumer Sentiment Index
- **VIXCLSx**: CBOE Volatility Index (VIX) - market fear gauge

3. Principal Component Analysis Results

Component Summary

- **Total Components**: 125 (after data cleaning)
- **PC1 Explains**: 59.18% of total variance
- **PC1-5 Combined**: 88.47% of total variance
- **Components for 90% Variance**: 6 components
- **Components for 95% Variance**: 10 components

Principal Component Interpretations

PC1 (59.18% variance): General Economic Growth & Price Level

Top Features: CPI indices, PCE, retail sales, wages

Economic Meaning:

- Dominant "size factor" capturing long-term economic growth
- Reflects secular inflation trends and nominal value increases
- All major price indices load heavily on PC1
- Higher PC1 = higher overall economic activity and price levels

PC2 (10.91% variance): Interest Rate & Monetary Policy Cycle

Top Features: Fed funds rate, Treasury yields (3M-30Y), corporate bond yields

Economic Meaning:

- Captures Federal Reserve monetary policy stance
- Shows coordination across entire yield curve
- Reflects credit market conditions and borrowing costs
- High PC2 = high interest rate environment (tight monetary policy)

PC3 (8.43% variance): Labor Market & Housing Cycle

Top Features: Unemployment rate (+), Housing starts (-), Building permits (-)

Economic Meaning:

- Employment-housing cycle relationship
- Housing is procyclical (moves opposite to unemployment)
- Captures business cycle dynamics
- High PC3 = high unemployment + weak housing market (recession indicator)

PC4 (6.47% variance): Term Structure & Credit Spreads

Top Features: Yield spreads (5Y-10Y term premiums), Credit spreads (AAA-BAA)

Economic Meaning:

- Yield curve shape and slope
- Credit risk premiums and flight-to-quality episodes
- Term structure of interest rates
- Market stress and risk appetite indicators

PC5 (3.49% variance): Labor Productivity & Trade

Top Features: Manufacturing hours, Trade-weighted dollar (-), Labor productivity

Economic Meaning:

- Labor market intensity vs. automation trends
- International competitiveness via exchange rates
- Manufacturing sector dynamics
- Work intensity and productivity changes

4. Transform Codes Explanation

Transform codes (first row in dataset) indicate data preprocessing:

- **Code 1:** No transformation (levels)
- **Code 2:** First difference (Δx)
- **Code 3:** Second difference ($\Delta\Delta x$)
- **Code 4:** Natural logarithm ($\ln(x)$)
- **Code 5:** First difference of log ($\Delta \ln(x)$) = **percentage change/growth rate**
- **Code 6:** Second difference of log ($\Delta\Delta \ln(x)$)

Most variables use Code 5, meaning data represents **month-over-month percentage changes**, not levels.

5. Economic Interpretation & Real-World Applications

Historical Event Analysis

2008 Financial Crisis Impact

- **PC1** (Economic Activity): Remained relatively stable (7.51-9.18)
- **PC2** (Interest Rates): Sharp decline (-4.56 to 1.29) - Fed easing
- **PC3** (Unemployment/Housing): Major spike (0.96-9.30) - labor/housing stress

COVID-19 Pandemic (2020-2021)

- **PC1**: Sharp increase (9.90-14.89) - massive fiscal/monetary stimulus
- **PC2**: Negative territory (-3.00 to 0.03) - zero interest rate policy
- **PC3**: High volatility (-6.40 to 9.46) - unemployment spike then recovery

Recent Period (2023-2025)

- **PC1**: Elevated levels (13.89-15.62) - persistent inflation
- **PC2**: Positive (0.26-1.37) - Fed tightening cycle
- **PC3**: Negative (-5.19 to -3.09) - tight labor market, strong housing

Practical Applications

1. **Economic Forecasting**: Use PC scores to predict turning points
2. **Policy Analysis**: Track monetary policy transmission via PC2
3. **Risk Management**: Monitor PC3 for recession early warning
4. **Portfolio Management**: Use PCs for factor-based investment strategies
5. **Business Cycle Dating**: Combine PCs to identify expansion/contraction phases

6. Analysis Results Summary

Correlation Insights

- **Strong Positive**: Employment measures with industrial production ($r=0.99$)
- **Strong Negative**: Unemployment with stock market performance ($r=-0.30$)
- **Policy Coordination**: Fed funds rate highly correlated with all Treasury yields
- **Inflation Linkage**: CPI strongly correlated with money supply growth (M2)

Variance Analysis Key Findings

- **Dimensionality Reduction:** 95% of variance captured by just 10 components (vs. 125 original variables)
- **Dominant Factor:** PC1 alone explains nearly 60% of all economic variation
- **Efficient Representation:** First 4 PCs capture 85% of information
- **Practical Threshold:** 6 components sufficient for most economic modeling (90% variance)

Statistical Properties

- **Data Quality:** High completeness after cleaning (800 observations, 125 variables)
- **Time Coverage:** Comprehensive 67-year span captures multiple business cycles
- **Cross-Correlation:** Strong interdependence among economic indicators confirms factor structure
- **Stationarity:** Transform codes ensure most series are stationary for analysis

7. Technical Implementation Notes

Data Preprocessing Steps

1. Remove transform codes row
2. Convert date column and set as index
3. Drop variables with >40% missing values
4. Forward-fill then backward-fill remaining missing values
5. Standardize all variables (mean=0, std=1) before PCA

PCA Implementation

```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA

scaler = StandardScaler()
X_scaled = scaler.fit_transform(data)
pca = PCA()
components = pca.fit_transform(X_scaled)
```

Key Output Files

- `pca_loadings.csv`: Component loadings matrix
- `pca_scores_timeseries.csv`: Principal component time series
- `pca_variance_explained.csv`: Variance analysis results
- `correlation_matrix.csv`: Key indicator correlations

This document provides a comprehensive guide to understanding and analyzing the FRED-MD macroeconomic dataset using Principal Component Analysis. For additional technical details or specific variable definitions, consult the Federal Reserve Bank of St. Louis FRED database documentation.