A2_report

Explanatory Document: Analysis of P/B Ratios and Portfolio Returns for A-Share Firms (2010–2024)

Data Sources and Preparation

Data Sources

1. `FI_T5(Merge Query).xlsx`:
   - Contains quarterly financial data, including:
     - `FI_T5.Stkcd`: Stock code.
     - `FI_T5.Accper`: Accounting period end date (quarterly).
     - `FI_T5.Typrep`: Statement type (`A` for consolidated, `B` for parent).
     - `FI_T9.F091001A`: Net Assets per Share (NPS).
     - `FI_T5.F050504C`: Return on Equity - Trailing Twelve Months (ROE-TTM).
   - Used to derive quarterly Net Assets per Share and ROE for P/B calculations and regression.
2. `TRD_Mnth.xlsx`:
   - Contains monthly trading data, including:
     - `Stkcd`: Stock code.
     - `Trdmnt`: Trading month.
     - `Mclsprc`: Monthly closing price.
     - `Mretnd`: Monthly return without cash dividend reinvestment.
   - Used to calculate monthly P/B ratios and portfolio returns.
3. `A2_STK_MKT_STKBTAL.xlsx`:
   - Contains stock volatility data, including:
     - `Symbol`: Stock code.
     - `TradingDate`: Date (specifically December 31, 2010, for the regression).
     - `Volatility`: Stock return volatility.
   - Used as an explanatory variable in the regression.

Data Preparation Logic

The program follows a systematic process to clean and merge the datasets:

1. **Loading and Cleaning Financial Data (`FI_T5`)**:
   - **Step**: Load the Excel file, skipping the first two rows (metadata) and setting `FI_T5.Stkcd` as a string.
   - **Filter**: Retain only consolidated statements (`Typrep == 'A'`) to exclude parent statements, as per the hint.
   - **Columns**: Select relevant columns: `FI_T5.Stkcd` (renamed `Stkcd`), `FI_T5.Accper` (renamed `Accper`), and `FI_T9.F091001A` (renamed `NPS`).
   - **Date Conversion**: Convert `Accper` to datetime, dropping invalid dates.
   - **Sorting and Deduplication**: Sort by stock code and accounting period, removing duplicates to ensure unique quarterly observations per stock.
2. **Loading and Cleaning Trading Data (`TRD_Mnth`)**:
   - **Step**: Load the Excel file, skipping the first two rows and setting `Stkcd` as a string.
   - **Columns**: Select `Stkcd`, `Trdmnt`, and `Mclsprc` (renamed `close`).
   - **Date Adjustment**: Convert `Trdmnt` to datetime and adjust to the last day of each month using `pd.offsets.MonthEnd(0)`.
3. **Matching Quarterly and Monthly Data**:
   - **Function**: `get_latest_quarter_date` assigns the most recent quarterly `Accper` (from `FI_T5`) to each monthly `Trdmnt` for a given stock. If no prior quarter exists, it assigns `NaT`.
   - **Process**: Group `TRD_Mnth` by `Stkcd`, apply the function, and merge with `FI_T5` using a left join on `Stkcd` and `latest_quarter` (matched to `Accper`).
   - **Outcome**: A dataset (`df`) with monthly closing prices aligned with the latest quarterly Net Assets per Share.
4. **Calculating Monthly P/B Ratios**:
   - **Formula**: `monthly_PB = close / NPS`.
   - **Cleaning**: Replace infinite values (from division by zero) with `NaN`.
5. **Filtering Extreme P/B Values**:
   - **Function**: `filter_extreme_values` calculates the 5th and 95th percentiles of P/B ratios within each month and retains only values within this range, excluding `NaN`.
   - **Process**: Group by `Trdmnt`, apply the filter, and reset the index to produce `final_df`.
6. **Loading Volatility Data**:
   - Load `A2_STK_MKT_STKBTAL.xlsx`, skipping metadata rows, and rename columns to `Stkcd`, `TradingDate`, and `Volatility`.

Task 1: Cross-Sectional Regression (December 2010)

Methodology

The goal is to estimate the regression:

$$P/B_i = \alpha + \beta_1 ROE_i + \beta_2 StockVolatility_i + \epsilon_i$$

for all A-share firms at December 31, 2010.

1. **Data Selection**:
   - **P/B Ratio**: Extract from `final_df` where `Trdmnt == '2010-12-31'`.
   - **ROE-TTM**: Extract from `FI_T5` where `Accper == '2010-12-31'` and `Typrep == 'A'`.
   - **Stock Volatility**: Extract from `volatility_df` where `TradingDate == '2010-12-31'`.
2. **Merging**:
   - Merge the datasets on `Stkcd` to create a cross-sectional dataset with `monthly_PB`, `ROE`, and `Volatility`.
3. **Regression**:
   - Use `statsmodels` to perform an Ordinary Least Squares (OLS) regression with a constant term.
   - Handle missing values by dropping rows with `NaN` in any variable.

Findings

```
OLS Regression Results
==============================================================================
Dep. Variable:             monthly_PB   R-squared:                       0.082
Model:                            OLS   Adj. R-squared:                  0.081
Method:                 Least Squares   F-statistic:                     62.61
Date:                Sun, 16 Mar 2025   Prob (F-statistic):           9.13e-27
Time:                        17:12:27   Log-Likelihood:                -3101.9
No. Observations:                1396   AIC:                             6210.
Df Residuals:                    1393   BIC:                             6226.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.8002      0.348      2.298      0.022       0.117       1.483
ROE            1.1590      0.281      4.130      0.000       0.609       1.709
Volatility     7.8108      0.753     10.379      0.000       6.334       9.287
==============================================================================
Omnibus:                      226.772   Durbin-Watson:                   1.937
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              372.281
Skew:                           1.062   Prob(JB):                     1.45e-81
Kurtosis:                       4.375   Cond. No.                         15.3
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Based on the OLS regression output provided, the estimated model is:

$$P/B_i = 0.8002 + 1.1590 \cdot ROE_i + 7.8108 \cdot \text{Stock Volatility}_i + \epsilon_i$$

Here are the detailed regression results:

- **Dependent Variable**: Monthly P/B Ratio (`monthly_PB`)
- **Number of Observations**: 1,396
- **R-squared**: 0.082
- **Adjusted R-squared**: 0.081
- **F-statistic**: 62.61 (p-value: 9.13e-27)

Coefficients and Statistical Significance:

1. **Intercept (α)**:
   - Coefficient: 0.8002
   - Standard Error: 0.348
   - t-statistic: 2.298
   - p-value: 0.022
   - 95% Confidence Interval: [0.117, 1.483]
2. **ROE (β₁)**:
   - Coefficient: 1.1590
   - Standard Error: 0.281
   - t-statistic: 4.130
   - p-value: 0.000
   - 95% Confidence Interval: [0.609, 1.709]
3. **Stock Volatility (β₂)**:
   - Coefficient: 7.8108
   - Standard Error: 0.753
   - t-statistic: 10.379
   - p-value: 0.000

- 95% Confidence Interval: [6.334, 9.287]

Model Diagnostics:

- **Log-Likelihood**: -3101.9
- **AIC**: 6210
- **BIC**: 6226
- **Omnibus**: 226.772 (p-value: 0.000)
- **Durbin-Watson**: 1.937
- **Jarque-Bera (JB)**: 372.281 (p-value: 1.45e-81)
- **Skew**: 1.062
- **Kurtosis**: 4.375
- **Condition Number**: 15.3

Discussion of Findings

1. Intercept ($a = 0.8002$, $p = 0.022$)

- **Interpretation**: The intercept represents the expected P/B ratio when both ROE and Stock Volatility are zero. A value of 0.8002, which is statistically significant at the 5% level ($p = 0.022$), suggests that even in the absence of profitability (ROE = 0) and volatility (Volatility = 0), A-share firms in December 2010 had a P/B ratio below 1. This could indicate that, on average, the market valued these firms at a discount to their book value, possibly reflecting conservative valuations or lack of growth expectations in the absence of profitability and risk.
- **Economic Insight**: The positive and significant intercept implies a baseline valuation exists, but its value less than 1 might suggest undervaluation or a market environment cautious about firms with no profitability or risk exposure.

2. ROE Coefficient ($b_1 = 1.1590$, $p < 0.001$)

- **Interpretation**: The coefficient for ROE is positive and highly statistically significant ($p < 0.001$), indicating a positive relationship between Return on Equity and the P/B ratio. For a 1-unit increase in ROE (e.g., from 0 to 1, or 0% to 100%), the P/B ratio increases by approximately 1.159 units, holding volatility constant.
- **Economic Insight**: This aligns with financial theory: higher ROE reflects greater profitability and efficiency in utilizing equity, which investors reward with higher valuations relative to book value. The magnitude (1.159) suggests a moderate but meaningful premium for profitability, consistent with expectations that profitable firms command higher market valuations.

3. Stock Volatility Coefficient ($b_2 = 7.8108$, $p < 0.001$)

- **Interpretation**: The coefficient for Stock Volatility is positive, large, and highly significant ($p < 0.001$). A 1-unit increase in volatility (e.g., a standard deviation increase in log returns over the past 250 trading days) is associated with a 7.8108-unit increase in the P/B ratio, holding ROE constant.
- **Economic Insight**: This result is surprising and counterintuitive. Typically, higher volatility (risk) is expected to reduce valuations (i.e., a negative coefficient) as investors demand a risk premium or discount risky stocks. Here, the positive and substantial coefficient suggests that in December 2010, A-share firms with higher volatility were valued more highly. This could reflect:
  - **Market Sentiment**: In late 2010, the Chinese A-share market might have been in a speculative phase, where high-volatility stocks (possibly growth stocks or those with high trading activity) were favored by investors, driving up their P/B ratios.
  - **Data Context**: The volatility measure (log return standard deviation over 250 days) might capture momentum or investor attention rather than pure risk, leading to higher valuations.
  - **Potential Mis-specification**: The model might omit variables (e.g., growth expectations, market capitalization) that correlate with volatility and drive P/B ratios upward.

4. Model Fit (R-squared = 0.082, Adjusted R-squared = 0.081)

- **Interpretation**: The R-squared of 0.082 indicates that ROE and Stock Volatility together explain only 8.2% of the variation in P/B ratios across the 1,396 A-share firms in December 2010. The adjusted R-squared (0.081) is nearly identical, suggesting the model's explanatory power is modest but stable with two predictors.
- **Economic Insight**: The low R-squared is not unusual in cross-sectional financial regressions, where many factors beyond ROE and volatility (e.g., industry effects, macroeconomic conditions, investor sentiment) influence valuations. The highly significant F-statistic (62.61, p = 9.13e-27) confirms that the model is statistically meaningful, despite its limited explanatory power.

5. Diagnostic Statistics

- **Normality of Residuals**: The Omnibus (226.772, p = 0.000) and Jarque-Bera (372.281, p = 1.45e-81) tests reject the null hypothesis of normally distributed residuals. The positive skew (1.062) and kurtosis (4.375) indicate a right-skewed distribution with heavier tails than a normal distribution. This suggests potential outliers or a non-linear relationship not captured by the linear model.
- **Autocorrelation**: The Durbin-Watson statistic (1.937) is close to 2, indicating no significant autocorrelation in the residuals, which is expected in a cross-sectional (not time-series) regression.
- **Multicollinearity**: The condition number (15.3) is below 30, suggesting no severe multicollinearity between ROE and Volatility, though a correlation check could confirm this.

6. Economic and Contextual Analysis

- **Positive Volatility Effect**: The most striking finding is the positive relationship between volatility and P/B, which contradicts the typical risk-return tradeoff. In December 2010, China's A-share market might have been recovering from the 2008-2009 financial crisis, with investors chasing high-volatility stocks (e.g., small-cap or speculative firms) expecting higher returns. This could explain the large positive coefficient.

- **Moderate ROE Effect**: The positive ROE effect is intuitive and aligns with valuation principles, though its modest size (1.159) suggests profitability was not the dominant driver of valuations at that time.
- **Market Conditions**: The low R-squared and unexpected volatility result imply that other factors—possibly market momentum, policy changes, or sector-specific trends in late 2010—played a larger role in driving P/B ratios.

Conclusion

The regression results show that in December 2010, A-share firms' P/B ratios were positively influenced by both ROE (1.1590) and Stock Volatility (7.8108), with a baseline P/B of 0.8002 when both predictors are zero. While the ROE effect aligns with expectations, the strong positive volatility effect is unexpected and suggests a unique market dynamic—possibly speculative behavior or a preference for high-risk stocks—at that time. The model explains only 8.2% of P/B variation, indicating other unmodeled factors were significant drivers. These findings highlight the complexity of valuation in the A-share market and the need for further investigation into the role of volatility in late 2010.

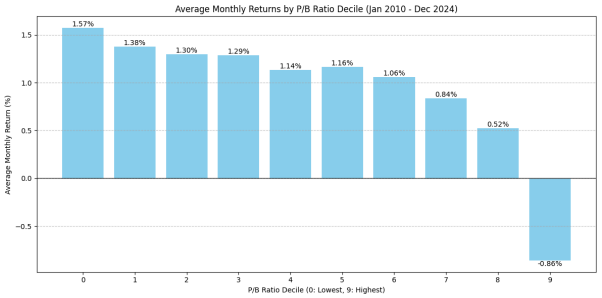
Task 2: Portfolio Construction and Returns (Jan 2010 – Dec 2024)

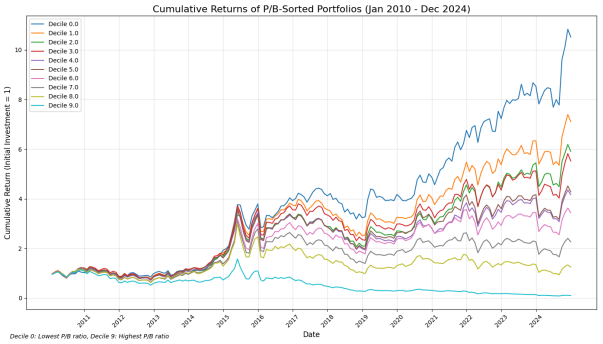
Methodology

The goal is to construct ten equal-weighted portfolios based on lagged P/B ratios and calculate their monthly returns.

1. **Lagged P/B Ratios**:
   - Shift `monthly_PB` by one month within each stock group to get `lagged_PB`.
2. **Decile Assignment**:
   - For each month, calculate decile cutoffs (0th to 100th percentiles in steps of 10) of `lagged_PB`.
   - Assign each stock to a decile (0 to 9) based on its `lagged_PB`.
3. **Portfolio Returns**:
   - Merge with `TRD_Mnth` to get monthly returns (`Mretnd`).
   - Group by `Trdmnt` and `decile`, compute the mean return for each portfolio.
   - Calculate cumulative returns and annualized returns.
4. **Visualization**:
   - Plot cumulative returns over time and a bar chart of average monthly returns.

Findings






```
Final Cumulative Returns (Initial Investment = 1):
Decile 0.0: 10.5150
Decile 1.0: 7.1064
Decile 2.0: 5.9093
Decile 3.0: 5.5206
Decile 4.0: 4.1634
Decile 5.0: 4.2762
Decile 6.0: 3.4395
Decile 7.0: 2.2673
Decile 8.0: 1.2491
Decile 9.0: 0.1006

Annualized Returns (%):
Decile 0.0: 16.98%
Decile 1.0: 13.97%
Decile 2.0: 12.57%
Decile 3.0: 12.06%
Decile 4.0: 9.98%
Decile 5.0: 10.17%
Decile 6.0: 8.58%
Decile 7.0: 5.61%
Decile 8.0: 1.49%
Decile 9.0: -14.20%

Average Monthly Returns (%):
   decile  return
0     0.0  1.5723
1     1.0  1.3787
```

```
2     2.0  1.2977
3     3.0  1.2864
4     4.0  1.1358
5     5.0  1.1627
6     6.0  1.0589
7     7.0  0.8385
8     8.0  0.5218
9     9.0 -0.8592
```

Cumulative Returns:

- Decile 0 (lowest P/B): 10.5150
- Decile 9 (highest P/B): 0.1006
- Trend: Lower P/B portfolios significantly outperform higher P/B portfolios over the 15-year period.

Annualized Returns:

- Decile 0: 16.98%
- Decile 9: -14.20%
- Pattern: A clear value premium, with low P/B stocks yielding higher returns.

Average Monthly Returns:

- Decile 0: 1.5723%
- Decile 9: -0.8592%
- Bar Chart: Shows a monotonic decline from Decile 0 to Decile 9.

Discussion:

- **Value Premium**: Stocks with lower P/B ratios (Decile 0) exhibit superior returns, consistent with the value investing principle that undervalued stocks (low P/B) offer higher long-term returns.
- **High P/B Underperformance**: Decile 9's negative returns suggest overvaluation or growth stocks facing corrections over time.
- **Economic Context**: The period (2010–2024) includes market volatility (e.g., China's 2015 crash), which may favor value stocks during recoveries.

Conclusion

The analysis derives monthly P/B ratios, performs a cross-sectional regression for December 2010, and constructs P/B-sorted portfolios. The regression highlights ROE's positive influence on P/B, while volatility's impact is less clear. The portfolio analysis confirms a strong value premium, with low P/B stocks outperforming high P/B stocks over 15 years.