# Analyzing Changing "Investor Exuberance": The Determinants of S&P Composite Index Total Return CAPE Changes

C. N. V. Krishnan, Xiyao Tan, and Jiemin Yang *

**Abstract**

We analyze the determinants of changes in S&P Composite Index Total Return Cyclically Adjusted Price-to-Earnings ratio (TR CAPE), to better understand changing "investor exuberance". We use three different methods - linear regression using PCA, Lasso, and Ridge regression techniques, as well as ElasticNet method – and a large number of explanatory variables, to explain changing investor exuberance. Across all methods, we find that monthly changes in Michigan sentiment index is significantly associated with monthly changes in TR CAPE. When we cross check the results using annual changes (rather than monthly changes), across all methods, annual changes in Michigan sentiment index and changes in core inflation are significantly associated with annual changes in TR CAPE.

**Keywords**: TR CAPE, Cyclically Adjusted Price-to-Earnings ratio, investor exuberance, PCA, Principal Components Analysis, Lasso regression, Ridge regression, determinants, stock market variables, economy-wide variables, fixed income variables, commodity variables.

**JEL Code**: G10

## I. INTRODUCTION

Well-known valuation indicators such as the CAPE (Cyclically Adjusted Price-to-Earnings) ratio have been studied for their predictive power in financial markets. For example, Siegel (2016) discusses how CAPE can be used to predict future stock returns over long horizons, providing insights into market valuation and investment strategy. Mauboussin and Callahan (2014) examine the relationship between CAPE and economic cycles, highlighting its utility in signaling economic downturns and recoveries. Market sentiment analysis also incorporates CAPE to gauge investor behavior. Baker and Wurgler (2007), for example, discuss how sentiment-driven market fluctuations may be linked to understanding CAPE trends, providing insights into investor behavior and market dynamics.

Our objective in this paper is to explain the contemporaneous relation between changes in various key variables and changing investor exuberance, captured by changes in S&P Composite Index Total Return Cyclically Adjusted Price-to-Earnings ratio (TR CAPE) (Shiller, 2005). In other words, we use an explanatory model, not a predictive model (see Shmueli, 2010). Any change in any explanatory variable should affect price contemporaneously (the numerator in P/E ratio) as it changes investor perception (Fama, 1970). A high TR CAPE ratio be a sign of exuberance or speculative bubbles. Conversely, a low TR CAPE ratio can indicate pessimism about stock performances.

* C. N. V. Krishnan, cnk2@case.edu, Case Western Reserve University; Xiyao Tan, xxt187@case.edu, Case Western Reserve University; Jiemin Yang, jxy1144@case.edu, Case Western Reserve University.

There are many possible explanatory variables, including economy wide variables, market variables, fixed income variables, and commodity variables. To analyze this, we use different methods and compare results. We compare the results of linear regression using Principal Component Analysis (PCA), Lasso Regression, and Ridge Regression. Integrating PCA, Lasso, and Ridge regression techniques can enhance financial research by leveraging the strengths of each method. Hastie, Tibshirani, and Wainwright (2015), for example, provide a comprehensive guide on combining these methods to improve model robustness and predictive accuracy, illustrating the benefits of an integrated approach in financial econometrics. The significance of selecting the appropriate methodology must take into account the fact that different techniques can yield varying results in terms of variable significance, as highlighted by Zou and Hastie (2005). This enables us to compare and contrast the results.

Principal Component Analysis (PCA) has been extensively used to reduce dimensionality in large datasets by identifying key factors that drive movements. Jolliffe (2002) elaborates on the mathematical foundations and applications of PCA, emphasizing its role in reducing model complexity while retaining essential information. Baker and Wurgler (2006) explore the impact of investor sentiment on stock returns using PCA regression, demonstrating the importance of accounting for psychological factors in financial modeling. Stock and Watson (2002) applied PCA analysis for macroeconomic forecasting, to identifying the key factors from many possible predictors. Connor and Korajczyk (1986) demonstrate how PCA analysis improves understanding of market dynamics by isolating significant variables from noise. Litterman and Scheinkman (1991) applied PCA to the term structure of interest rates, showing its utility in identifying principal components that explain the variance in bond returns.

Lasso regression, introduced by Tibshirani (1996), performs variable selection and regularization by imposing an L1 penalty on regression coefficients (a penalty on the sum of the absolute values of the regression coefficients), effectively shrinking some coefficients to zero. This means that the Lasso method can identify and keep only the most significant variables in the model, while excluding others that are less important. This method is particularly useful in financial modeling, where multicollinearity is a common issue. Studies by Fan and Li (2001) have demonstrated the effectiveness of Lasso in improving financial econometric models by selecting relevant predictors and reducing the influence of less significant variables.

Ridge regression addresses multicollinearity by adding an L2 penalty to the regression equation, stabilizing coefficient estimates without reducing them to zero. The L2 penalty, also known as the Ridge penalty, is on the sum of the squared values of the regression coefficients. Unlike the L1 penalty used in Lasso regression, the L2 penalty does not set any coefficients to zero. Instead, it shrinks all coefficients toward zero. Hoerl and Kennard (1970) initially proposed Ridge regression, which has since been applied in various financial contexts. This technique has been shown to enhance the stability and predictive accuracy of economic indicators and financial forecasting models (Elliott & Timmermann, 2008). Hoerl and Kennard (1970) demonstrate the utility of Ridge regression in dealing with multicollinearity and improving the accuracy of economic predictions. Stock and Watson (2002) applied Ridge regression in macroeconomic forecasting, illustrating its effectiveness in handling high-dimensional economic data.

The comparative analysis of PCA, Lasso, and Ridge regression techniques has been a focal point in recent financial research. Hastie, Tibshirani, and Friedman (2009) provide a detailed comparison, emphasizing the unique advantages and drawbacks of each method. James, Witten, Hastie, and Tibshirani (2013) provide insights into the applications of Lasso and Ridge regression in modern financial econometrics, highlighting their effectiveness in selecting significant predictors and improving model performance.

We also use ElasticNet, introduced by Zou and Hastie (2005), as a combination of Lasso and Ridge regressions. While Lasso tends to select a few key variables and forces the rest to zero, Ridge regression includes all variables but shrinks coefficients uniformly. ElasticNet creates a balance by blending the penalties of both methods, allowing for variable selection like Lasso while retaining the regularization properties of Ridge. This makes it particularly effective when dealing with correlated predictors, which is often the case in financial datasets and the case here as well. By using ElasticNet, researchers can achieve a more robust and interpretable model that integrates the strengths of Lasso's feature selection and Ridge's handling of multicollinearity, and represents a comprehensive methodology that captures both predictive accuracy and variable significance in financial econometrics.

We use all these methods in this paper. PCA based linear regression results show that monthly change in Michigan sentiment index and one commodity (Zinc) price change are significantly associated with monthly change in Index TR CAPE. Lasso regression (using 3 different methodologies) shows that monthly change in Michigan sentiment index and change in the 5-year Treasury yield are significantly associated with monthly change in Index TR CAPE, consistently across the 3 methods. Ridge and ElasticNet regressions show that monthly change in Michigan sentiment index, change in 5-year Treasury yield and change in 10-Year High-Quality-Market Corporate-Bond Spot Rate are significantly associated with monthly change in TR CAPE, across the 2 methods.

Therefore, across all methods, monthly changes in Michigan sentiment index is significantly associated with monthly changes in TR CAPE. Consumer sentiment is seen as a leading indicator of economic activity. Carroll, Fuhrer, and Wilcox (1994) and Ludvigson (2004), for example, argue that consumer sentiment can significantly influence stock market valuations. When consumers feel confident about the economy, they may be more likely to invest in the stock market, driving up stock valuations and increasing the market-index TR CAPE ratio (Brown and Cliff, 2004; and Fisher and Statman, 2000). Conversely, a decline in sentiment might indicate a future drop in stock prices (Baker and Wurgler, 2007). The Michigan Sentiment Index (based on survey of consumers) is a widely recognized measure of consumer confidence, which can have a significant impact on economic expectations and investor behavior, and hence on TR CAPE (also see, for example, Lemmon and Portniaguina, 2006).

When we cross check the results using annual changes in Index TR CAPE (rather than monthly changes), we find, using linear regression using PCA, that the annual change in Michigan sentiment index, a commodity (Zinc) price change, and change in core inflation are significantly associated with the annual change in TR CAPE. From LASSO regression (using three different methodologies), we find that the annual change in GDP index, change in core inflation, change in money supply, change in Michigan sentiment index, and change in 5-year Treasury yield are consistently associated with the annual change in TR CAPE across all three methods. From Ridge and ElasticNet regressions, we find that the annual change in core inflation, change in Michigan sentiment index, change in 5-year Treasury yield, change in Moody's BAA rate, Dow Jones return, Nikkei 225 (Japan) and Hang Seng Index (Hong Kong) are associated with annual change in TR CAPE, across the 2 methods.

Therefore, across all methods, annual changes in Michigan sentiment index and changes in core inflation are significantly associated with annual changes in TR CAPE.

We have already discussed the importance of consumer confidence above. Core inflation, which excludes food and energy prices, tends to be more stable as it reflects the underlying trend of inflation without the noise from volatile components. This makes it a more stable indicator, providing a clearer view of long-term inflation trends without the noise of short-term price changes. Higher core inflation can lead to increased interest rates, which raises the discount rate, reducing the present value of future cash flows and, in turn, lowering stock market valuations, including the TR CAPE ratio. Campbell and Shiller (1998) demonstrated that inflationary pressures, especially core inflation, directly affect the discount rates applied to future earnings, thereby potentially influencing valuation metrics like the TR CAPE ratio. Fama and Schwert (1977) also argue that core inflation impacts the discount rates used to value future earnings, thereby impacting valuation metrics.

## II. DATA AND VARIABLES

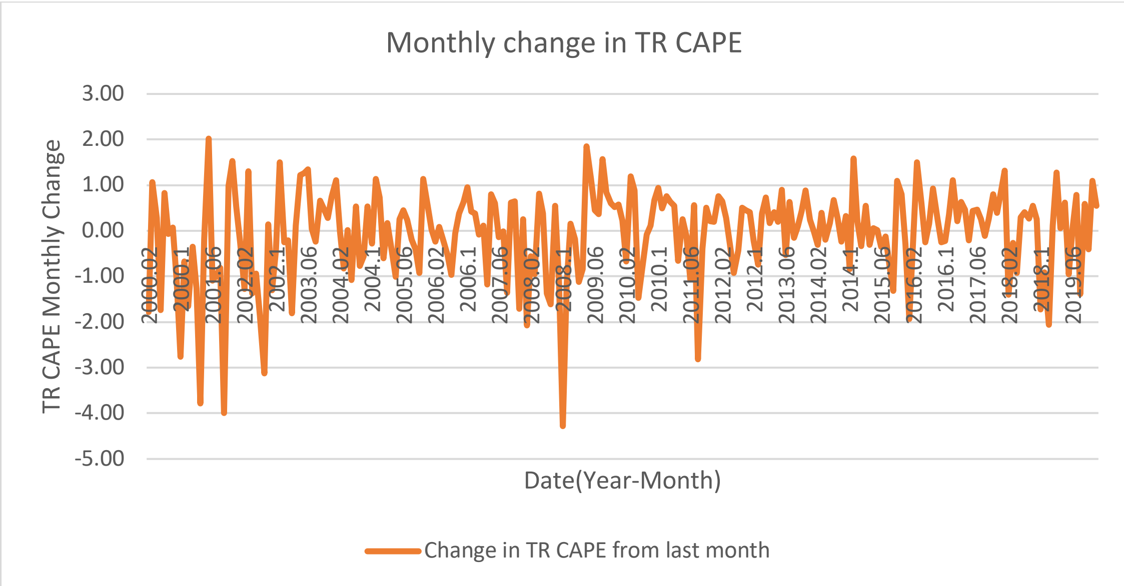
The variable we examine is the monthly Cyclically Adjusted S&P Composite Index Total Return Price Earnings Ratio (TR CAPE) from February 2000 to December 2019 taken from Robert Shiller's website (this data, updated, was used in "Irrational Exuberance" by Robert Shiller, Princeton University Press, 2000, 2005, 2015). Cyclically Adjusted Price Earnings Ratio" (CAPE), also known as P/E-10, and is a valuation measure that divides the current price of a stock or index by the average of ten years of earnings, adjusted for inflation. This ratio is used to assess whether a stock or market is over- or undervalued by comparing the current CAPE ratio to historical averages. TR CAPE (or TR P/E10) is similar to CAPE ratio but adjusted to include total return, which incorporates dividends and capital gains. It divides the current price of a stock or index by the average of ten years of earnings, adjusted for inflation and total returns. This ratio provides a more comprehensive view of the market's valuation by considering the full return profile. Therefore, we use this to specifically examine the determinants of the monthly change TR CAPE, to examine what determines changes in investor exuberance from month to month. As an additional check, we examine yearly changes in TR CAPE. We collect 42 explanatory variables from Bloomberg, FRED database of Federal Reserve Bank of St Louis and WRDS (Wharton Research Database System) as possible determinants to changes in Index TR CAPE.

We first examine monthly changes in the S&P Index TR CAPE. From Figure 1, the average monthly change in Index TR CAPE over 239 months in the database is approximately -0.060, with a standard deviation of 0.976 and a median of 0.150. The negative mean suggests a general downward trend in TR CAPE from the previous month. The standard deviation is moderate, indicating that the changes in TR CAPE can vary from month to month. The positive median, slightly higher than the average, suggests that while the average change is negative, the distribution of changes is such that the central tendency (median) remains positive, indicating that there are more positive changes than the negative mean might imply.

The first set of explanatory variables we consider that can influence changes in Index TR CAPE are the economy-wide explanatory variables. We consider monthly changes as in (This month - Last month) for US GDP index, US CPI index, Core inflation, Federal Funds Rate, Unemployment rate, Industrial production, M2 money supply, Personal income index, Producer price index, and Michigan sentiment index. Some of the key ones are US GDP index, Core inflation, M2 money supply, and Michigan sentiment index. Fama (1990) highlighted the critical role of GDP growth as a driver of stock market valuations. An increase in GDP is typically indicative of economic expansion, leading to higher corporate earnings and, consequently, higher stock prices. Higher core inflation generally leads to increased interest rates, which raises the discount rate, reducing the present value of future cash flows and, in turn, lowering stock market valuations. Friedman and Schwartz (1963) explored the relationship between monetary aggregates, such as the M2 money supply, and stock market valuations. Their findings suggest that an increase in the M2 money supply, often a result of expansionary monetary policy, enhances liquidity in financial markets, thereby pushing up asset prices as more money chases the same number of assets. As explained earlier, the Michigan Sentiment Index (based on survey of consumers) is a widely recognized measure of consumer confidence, which can have a significant impact on economic expectations and investor behavior. Figure 2 shows the plots of a few key economic indicators, to get a sense of the variabilities, while Table 1 shows the descriptive statistics.

**Figure 1: Plot of Monthly change in TR CAPE**

*This figure shows the plot of monthly changes in S&P Composite Index Total Return CAPE from February 2000 to December 2019, the data for which taken from Robert Shiller's website.*




From Figure 2, the change in the US GDP index shows significant variability over the period, particularly during the 2008 financial crisis. The GDP index reflects the overall economic output, and sharp movements can be observed during periods of economic expansion or contraction. The most pronounced fluctuations occur around the periods of economic recessions and recoveries, such as in 2008-2009, where GDP changes are larger and more erratic. The change in core inflation is less variable compared to the other variables. Core inflation, which excludes food and energy prices, tends to be more stable as it reflects the underlying trend of inflation without the noise from volatile components. The relative stability in core inflation suggests that underlying inflationary pressures remained moderate throughout the period, even during economic downturns. The change in M2 money supply is one of the more variable indicators, particularly showing large spikes around periods of economic intervention, such as during the financial crisis and subsequent quantitative easing programs. The Federal

Reserve's actions to increase liquidity in the financial system are evident in the large positive changes in M2. This variability reflects the monetary policy actions aimed at stabilizing the economy during periods of financial stress. The Michigan Sentiment Index, which measures consumer confidence, also exhibits notable variability, especially during economic downturns like the 2008 financial crisis. Changes in consumer sentiment are typically more volatile as they reflect consumer reactions to economic news, labor market conditions, and overall economic health. The sentiment index tends to react quickly to economic conditions, making it a likely candidate for a leading indicator of economic activity. Thus, the variability in these economic indicators highlights different aspects of the economic environment.

**Figure 2: Monthly Changes in Key Economy Wide Explanatory Variables**

*This figure shows the plots (in different colors) of monthly changes in 4 key economic indicators from February 2000 to December 2019, taken from different sources: Bloomberg, FRED, and WRDS.*

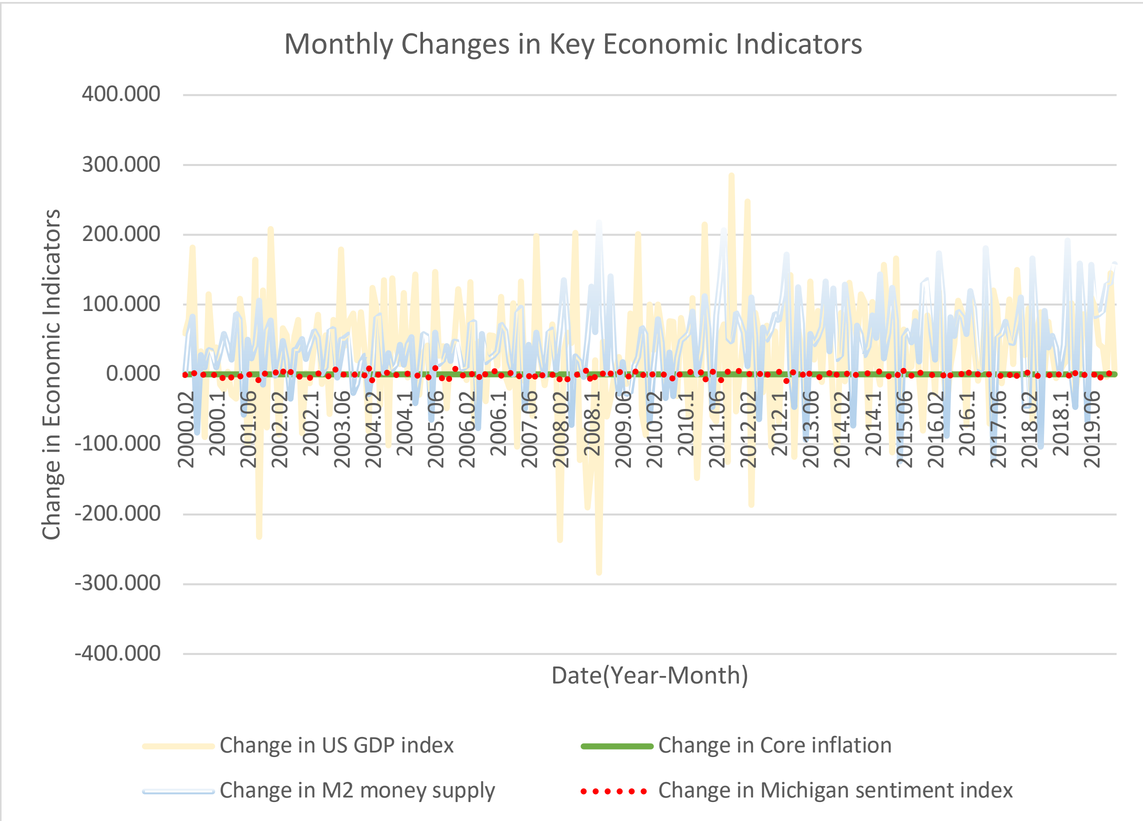



Table 1 presents the descriptive statistics of monthly changes in various key explanatory variables. The means of most variables show small deviations from zero, except for the change in the M2 money supply, personal income index, and US GDP index, which have positive means, indicating a general trend of increase over the observed period. The standard deviations vary, with the M2 money supply and personal income index showing the highest variability, reflecting significant fluctuations in these indicators. The change in the Federal Funds Rate and unemployment rate have medians close to zero, suggesting general stability in these measures during the observed period. As mentioned above, the Michigan Sentiment Index also exhibits notable variability.

**Table 1: Descriptive Statistics of Monthly changes in key Economic Variables**

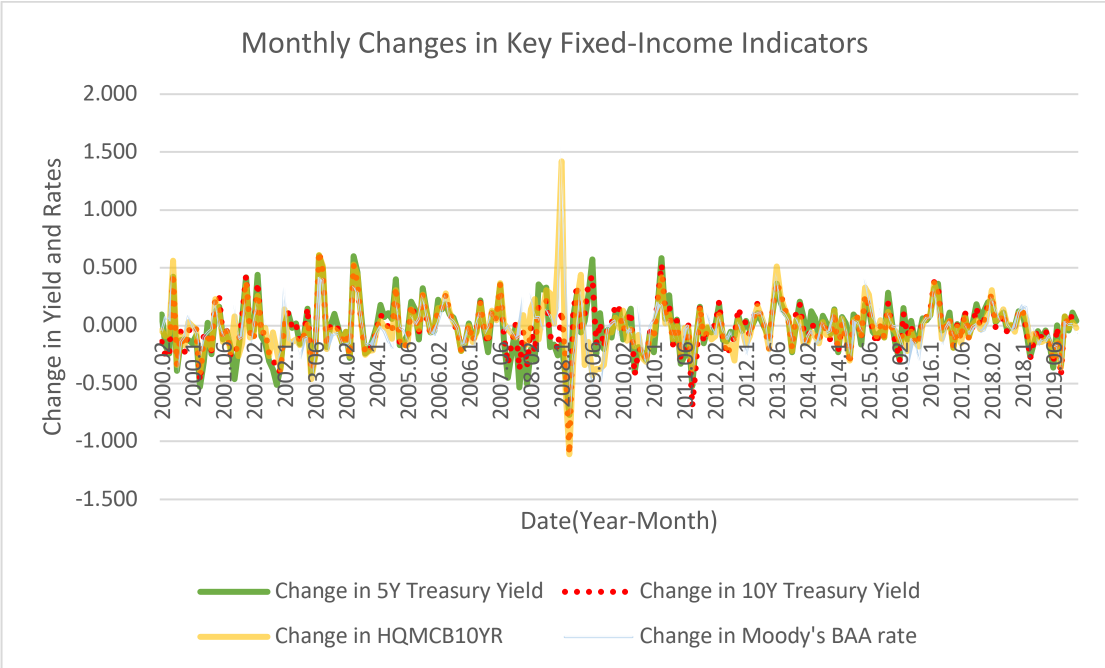
*This table shows the mean, median, standard deviation (SD) of monthly changes, as well as the number of months of observations (N), of all the economy-wide explanatory variables we collect. The sources of data are Bloomberg, WRDS and FRED databases*

| | N | Mean | SD | Median |
|---|---|---|---|---|
| Change in US GDP index | 239 | 30.071 | 84.260 | 35.645 |
| Change in US CPI index | 239 | 0.374 | 0.613 | 0.406 |
| Change in Core inflation | 239 | 0.363 | 0.168 | 0.393 |
| Change in Federal Funds Rate | 239 | -0.016 | 0.166 | 0.002 |
| Change in Unemployment rate | 239 | -0.002 | 0.160 | 0.000 |
| Change in Industrial production | 239 | 0.044 | 0.623 | 0.082 |
| Change in M2 money supply | 239 | 44.864 | 57.705 | 48.840 |
| Change in Personal income index | 239 | 38.518 | 85.332 | 41.200 |
| Change in Producer price index | 239 | 0.296 | 2.047 | 0.400 |
| Change in Michigan sentiment index | 239 | -0.053 | 4.069 | -0.200 |

The next set of explanatory variables we consider that can influence changes in Index TR CAPE are the fixed-income explanatory variables. We consider monthly changes, computed as (This month - Last month) for 1M Treasury Yield, 6M Treasury Yield, 1Y Treasury Yield, 5Y Treasury Yield, 10Y Treasury Yield, 30Y Treasury Yield, HQMCB10YR (10-Year High Quality Market Corporate Bond Spot Rate), Moody's AAA rate, Moody's BAA rate, and TED spread. Some of the key ones are 5-year (5Y) Treasury Yield, 10-year (10Y) Treasury Yield, HQMCB10YR, and Moody's BAA rate. Campbell and Viceira (2002) found that changes in intermediate-term Treasury yields, such as the 5-year Treasury yield, have a direct impact on equity valuations. As Treasury yields rise, the discount rate applied to future earnings increases, leading to lower present values of expected cash flows and a reduction in the TR CAPE ratio. Conversely, lower yields can support higher equity valuations by reducing the discount rate. Shiller (2005) analyzed the effect of long-term Treasury yields on stock market valuations. He concluded that the 10-year Treasury yield, often used as a benchmark for long-term interest rates, influences investor expectations about future inflation and economic growth. Higher 10-year yields may lead to lower TR CAPE ratios as they signal higher discount rates and potential economic slowing. Fama and French (1989) explored the relationship between high-quality corporate bond yields and equity valuations. They found that changes in these yields reflect the credit risk perceptions of high-quality corporate bonds, which can signal broader economic conditions. Rising corporate bond yields often correlate with increased risk aversion among investors, leading to lower equity valuations. Chen, Roll, and Ross (1986) investigated the impact of credit spreads, (e.g., Moody's BAA rate), on stock market valuations. They demonstrated that wider BAA spreads indicate higher credit risk and economic uncertainty, leading investors to demand higher returns on equities. This can depress stock prices and can reduce the TR CAPE ratio. Conversely, narrowing spreads suggest improving economic conditions, which can boost equity valuations. Figure 3 shows the plots of a few key fixed income indicators, to get a sense of the variabilities, while Table 2 shows the descriptive statistics.

**Figure 3: Monthly changes in Fixed Income Variables.**

*This figure shows the plots (in different colors) of monthly changes in 4 key fixed-income indicators from February 2000 to December 2019, taken from different sources: Bloomberg, FRED, and WRDS.*




From Figure 3, the 5Y Treasury Yield exhibits moderate variability over the period, influenced by short- to medium-term economic outlooks and monetary policy decisions. It tends to react strongly to changes in Federal Reserve policy and shifts in investor sentiment regarding inflation and economic growth. The 10 year Treasury Yield also shows moderate variability but is generally more stable than the 5Y Treasury Yield. This is because the 10-year yield represents long-term interest rates and is less sensitive to short-term economic fluctuations. However, it does respond to long-term economic expectations, inflation forecasts, and major policy announcements. Notable fluctuations are observed during periods of economic stress, such as the financial crisis of 2008. HQMCB10YR, which reflects high-quality corporate bond yields over a 10-year period, shows significant variability, particularly during the 2008 financial crisis. The spikes in this yield reflect periods of increased risk aversion among investors, where demand for safe assets like government bonds increases, and corporate bond yields rise due to perceived higher credit risk. This yield is sensitive to changes in economic conditions that affect corporate creditworthiness. Moody's BAA rate, which tracks yields on lower investment-grade corporate bonds, exhibits the highest variability among these variables, especially during the 2008 crisis. The pronounced spikes in the BAA rate reflect periods of heightened credit risk, where investors demand higher yields for taking on additional risk associated with lower-rated bonds. This rate is highly sensitive to economic downturns and periods of financial instability. Thus, this plot illustrates how different fixed income variables react to economic conditions. Treasury yields (5Y and 10Y) generally reflect broader economic expectations and are influenced by monetary policy, while corporate bond yields (HQMCB10YR and Moody's BAA rate) are more sensitive to credit risk and investor sentiment.

**Table 2: Descriptive Statistics of Fixed Income Variables**

*This table shows the mean, median, standard deviation (SD) of monthly changes, as well as the number of months of observations (N), of all the fixed-income explanatory variables we collect. The sources of data are Bloomberg, WRDS and FRED databases.*

| | N | Mean | SD | Median |
|---|---|---|---|---|
| Change in 1M Treasury Yield | 221 | 0.007 | 0.303 | 0.003 |
| Change in 6M Treasury Yield | 239 | -0.017 | 0.171 | -0.001 |
| Change in 1Y Treasury Yield | 239 | -0.019 | 0.171 | -0.002 |
| Change in 5Y Treasury Yield | 239 | -0.02 | 0.217 | -0.028 |
| Change in 10Y Treasury Yield | 239 | -0.02 | 0.21 | -0.023 |
| Change in 30Y Treasury Yield | 239 | -0.018 | 0.187 | -0.03 |
| Change in HQMCB10YR | 239 | -0.021 | 0.225 | -0.04 |
| Change in Moody's AAA rate | 239 | -0.02 | 0.174 | -0.026 |
| Change in Moody's BAA rate | 239 | -0.019 | 0.205 | -0.022 |
| Change in TED spread | 239 | -0.001 | 0.193 | -0.001 |

Table 2 presents the descriptive statistics for various fixed income variables. The means of most variables are slightly negative, indicating a slight downward trend in these fixed income rates over the observed period. The standard deviations vary, with the highest observed in the changes in 1 month Treasury Yield and HQMCB10YR, reflecting higher volatility in these indicators. The medians also display a mixture of positive and negative values, with the majority being negative, suggesting a general downward central tendency among the fixed income variables, though with some slight upward tendencies in certain cases, such as the 1M Treasury Yield.

The stock market explanatory variables that we consider are monthly returns computed as (Percentage change) = ((This month - Last month)/Last month)×100 for Dow Jones Index, S&P 500 Index, Russell 5000 Index, Value-weighted Center for Research in Security prices (VWCRSP) index, Equally weighted CRSP (EWCRSP) index, FTSE 100 (UK) Index, DAX (Germany) Index, CAC 40 (France) Index, Nikkei 225 (Japan) Index, Hang Seng (Hong Kong) Index, S&P/ASX 200 (Australia) index, and TSX Composite Index (Canada) Index. Some of the key ones are Dow Jones return (a blue-chip index that includes 30 large, well-established U.S. companies), S&P 500 return, Nikkei 225 (Japan) and Hang Seng Index (Hong Kong). Asness (2000) finds that positive returns in the Dow Jones Industrial Average often reflect strong economic fundamentals and investor optimism, leading to higher stock valuations. This upward pressure on valuations can result in an increased TR CAPE ratio as stock prices rise relative to long-term earnings. Shiller (1981) argues that strong performance in the S&P 500, which represents a broad measure of the U.S. stock market, is typically associated with increased investor confidence and expectations of future earnings growth. This can drive the TR CAPE ratio higher, as investors are willing to pay more for stocks in anticipation of continued growth. Froot and Ramadorai (2008) find that the performance of the Nikkei 225 can influence investor sentiment globally, especially in markets with close economic ties to Japan. Strong returns in the Nikkei 225 can boost confidence in Asian markets and contribute to higher TR CAPE ratios globally. Bekaert and Harvey (1997) find that fluctuations in the Hang Seng Index, which reflects the economic and financial conditions in Hong Kong and China, can impact global investor sentiment. Positive returns in the Hang Seng Index can lead to increased valuations in other markets, pushing up the TR CAPE ratio as investors become more optimistic about global economic prospects. Figure 4 shows the plots of a few key stock market returns, to get a sense of the variabilities, while Table 3 shows the descriptive statistics.

**Figure 4: Monthly returns in Stock Market Variables**

*This figure shows the plots (in different colors) of monthly changes in 4 key stock market returns from February 2000 to December 2019, taken from different sources: Bloomberg, FRED, and WRDS.*

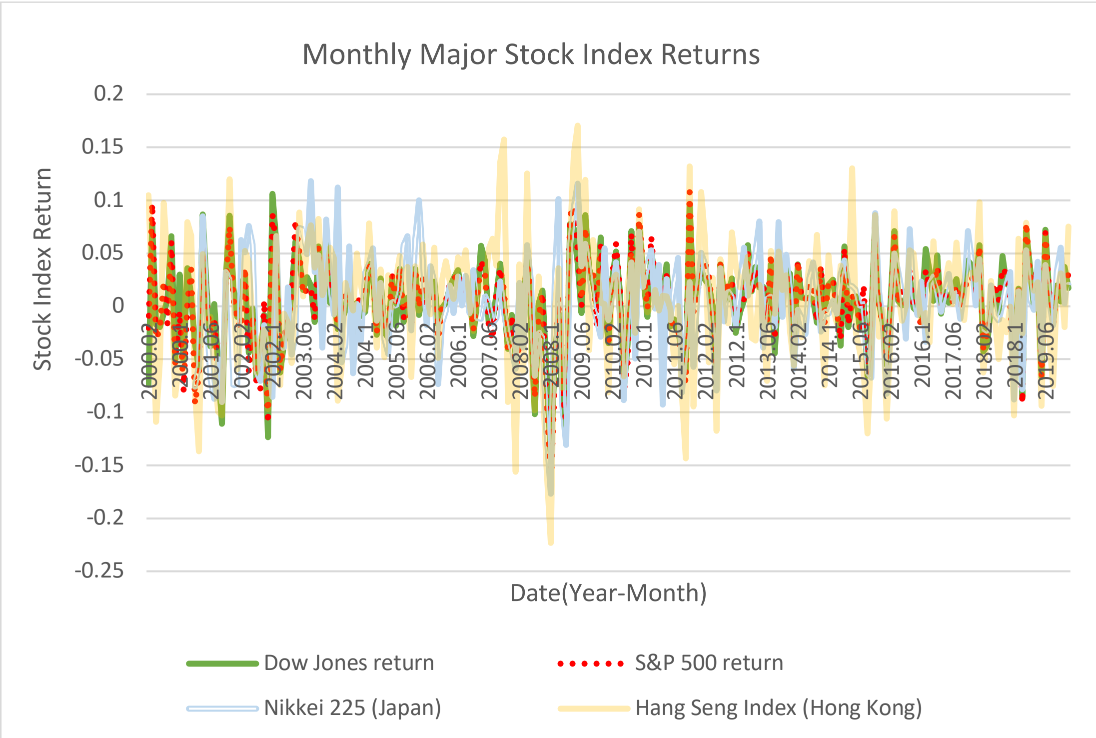



Figure 4 shows that the Dow Jones return is relatively stable compared to the other indices but still exhibits significant fluctuations, particularly during periods of economic stress such as the 2008 financial crisis. It tends to reflect broader economic conditions in the U.S. The S&P 500 return is like the Dow Jones in terms of variability but slightly more volatile due to its broader composition of 500 companies across various sectors. The S&P 500 is often seen as a better representation of the overall U.S. stock market, and its returns show a clear reaction to major economic events, such as the 2008 financial crisis, where significant dips are observed. The Nikkei 225 return exhibits moderate variability, reflecting the performance of the Japanese stock market. The index is sensitive to both domestic and international economic conditions, as Japan's economy is highly integrated into global trade and finance. Notable volatility can be seen during global economic downturns, such as in 2008, where the Nikkei experienced significant swings. The Hang Seng Index return is the most volatile among the four indices. This high variability is due to the Hang Seng Index's sensitivity to regional and global economic conditions, particularly those affecting China and the broader Asia-Pacific region. The index shows large swings during periods of financial instability, such as the 2008 financial crisis and subsequent global market events. The Hang Seng's high volatility reflects the dynamic nature of the Hong Kong market, which is influenced by both local factors and broader geopolitical and economic developments. The variability in these stock market indices highlights their sensitivity to different economic environments.

**Table 3: Descriptive Statistics of Stock Market Variables**

*This table shows the mean, median, standard deviation (SD) of monthly returns, as well as the number of months of observations (N), of all the stock market explanatory variables we collect. The sources of data are Bloomberg, WRDS and FRED databases.*

| | N | Mean | SD | Median |
|---|---|---|---|---|
| Dow Jones return | 239 | 0.005 | 0.04 | 0.008 |
| S&P 500 return | 239 | 0.004 | 0.042 | 0.01 |
| Russell 5000 return | 186 | 0.007 | 0.041 | 0.012 |
| VWCRSP index return | 239 | 0.004 | 0.042 | 0.01 |
| EWCRSP index return | 239 | -0.052 | 0.921 | 0.01 |
| FTSE 100 (UK) return | 239 | 0.001 | 0.046 | 0.003 |
| DAX (Germany) return | 239 | 0.006 | 0.067 | 0.007 |
| CAC 40 (France) return | 239 | 0.003 | 0.059 | 0.005 |
| Nikkei 225 (Japan) return | 227 | 0.004 | 0.048 | 0.009 |
| Hang Seng Index (Hong Kong) return | 239 | 0.004 | 0.06 | 0.01 |
| S&P/ASX 200 (Australia) return | 239 | 0.005 | 0.059 | 0.01 |
| TSX Composite Index (Canada) return | 239 | 0.005 | 0.056 | 0.008 |

The table shows that the means of the stock market variables are slightly positive, indicating small average monthly returns across the observed indices. The standard deviations are relatively low, with the highest observed for the EWCRSP index.[†] The medians are all positive, suggesting a general upward tendency in returns. However, the EWCRSP index shows a negative mean, one reason for which is the vulnerability of small-cap stocks to economic downturns (Fama & French, 1989). The equal-weighting method amplifies the impact of these losses since smaller firms are given the same weight as larger ones, further driving down this index's overall performance (Bekaert & Harvey, 1997).

The commodity prices explanatory variables that we consider are changes in prices computed as (This month price - Last month price) for sugar, coffee, soybean,

† The high SD of EWCRSP index is attributable to several factors. First, small-cap stocks, which play an equal role in equal-weighted indices like EWCRSP, tend to be more sensitive to changes in investor sentiment, leading to greater volatility compared to large-cap stocks (Baker & Wurgler, 2006). Additionally, small cap firms typically face lower liquidity, which amplifies price swings and contributes to higher variability in returns (Fama & French, 1989). Finally, small cap firms tend to experience higher overall volatility due to increased market risks, further explaining the larger standard deviation observed in the EWCRSP index (Bekaert & Harvey, 1997).

tobacco, crude oil, natural gas, gold, zinc, and rice. Some of the key ones are soybean price change, crude oil price change, gold price change and zinc price change. Bessler and Kling (1989) find that changes in soybean prices can influence the earnings of companies in the agricultural sector, as well as companies dependent on agricultural products. Additionally, soybean prices can impact inflation expectations, which in turn influence equity valuations. Hamilton (2009) find that fluctuations in oil prices significantly affect corporate earnings across various sectors, especially energy-intensive industries. High oil prices can increase production costs, reduce profit margins, and ultimately lower stock valuations, leading to a decrease in TR CAPE. Baur and McDermott (2010) find that rising gold prices often signal investor concerns about inflation or financial instability, which can lead to lower stock valuations and a decrease in TR CAPE as investors shift their portfolios toward safer assets. Erten and Ocampo (2013) find that changes in zinc prices can reflect broader economic conditions, particularly in industries such as construction and manufacturing. An increase in zinc prices can indicate higher demand in these sectors, potentially leading to higher earnings and stock prices, which could increase TR CAPE. Figure 5 shows the plots of a few key commodity price changes, to get a sense of the variabilities, while Table 4 shows the descriptive statistics.

**Figure 5: Monthly Changes in Commodity Prices Variables**

*This figure shows the plots (in different colors) of monthly changes in 4 key commodity price changes from February 2000 to December 2019, taken from different sources: Bloomberg, FRED, and WRDS.*

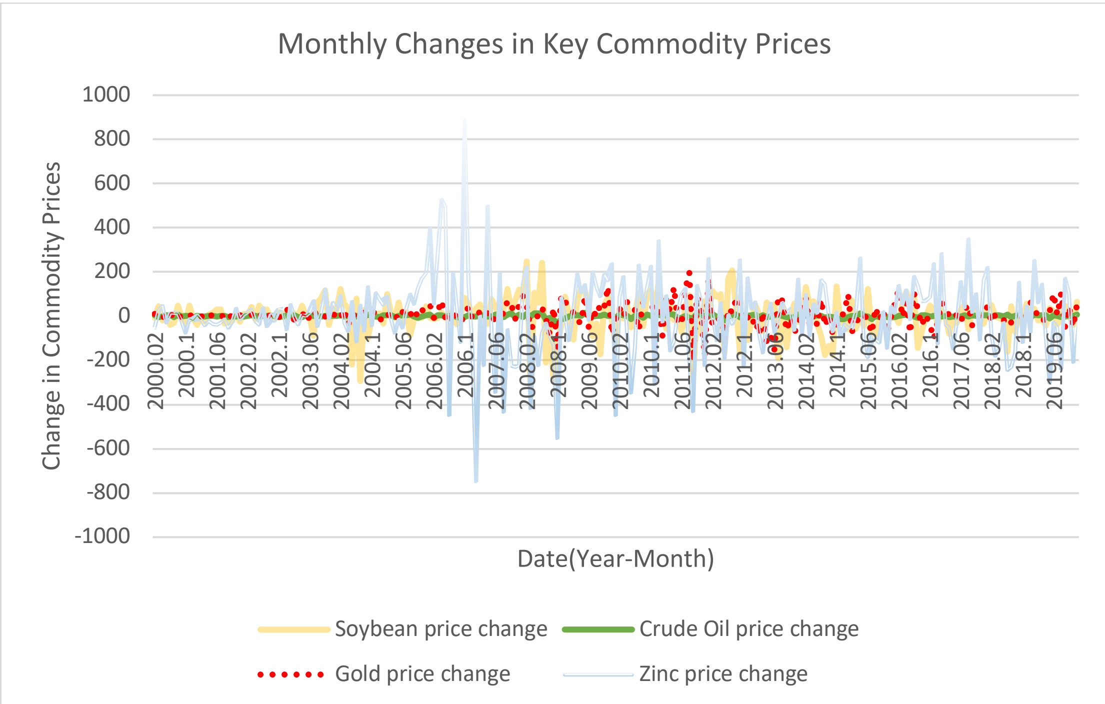



Figure 5 shows that the Soybean price changes show moderate variability. Soybean prices are influenced by factors such as weather conditions, crop yields, global demand, and trade policies. Crude oil price changes are more variable, with significant fluctuations throughout the period, especially during the mid-2000s and the 2008 financial crisis. Oil prices are highly sensitive to geopolitical events, changes in supply and demand dynamics, and OPEC's production decisions. Gold price change demonstrates relatively moderate variability, with notable spikes during times of economic uncertainty, such as the 2008 financial crisis. Gold is often considered a safe-haven asset, so its price tends to rise during periods of economic instability or when inflation expectations increase. Zinc price change is the most variable among the four commodities, with significant spikes and drops, particularly noticeable around 2006-2008. Zinc prices are closely tied to industrial demand, especially in construction and manufacturing. The high variability reflects the cyclicality of these industries and the impact of global economic conditions on demand for base metals like zinc, showing zinc's strong correlation with industrial demand.

**Table 4: Descriptive Statistics of Commodity Price Changes**

*This table shows the mean, median, standard deviation (SD) of monthly changes, as well as the number of months of observations (N), of all the stock market explanatory variables we collect. The sources of data are Bloomberg, WRDS and FRED databases.*

| | N | Mean | SD | Median |
|---|---|---|---|---|
| Sugar price change | 239 | 0.033 | 1.564 | 0.010 |
| Wheat price change | 239 | 1.266 | 51.627 | 1.250 |
| Coffee price change | 239 | 0.078 | 12.078 | -0.900 |
| Soybean price change | 239 | 1.820 | 81.130 | 5.250 |
| Tobacco price change | 239 | 3.207 | 8.599 | 1.800 |
| Crude Oil price change | 239 | 0.140 | 5.978 | 0.670 |
| Natural Gas price change | 239 | -0.002 | 0.903 | 0.015 |
| Gold price change | 239 | 5.162 | 52.085 | 4.550 |
| Zinc price change | 239 | 4.732 | 174.889 | 6.000 |
| Rice price change | 239 | 0.031 | 0.926 | 0.040 |

Table 4 presents the descriptive statistics for various commodity price changes over the observed period. The mean values for most commodities are positive, indicating that, on average, there were slight increases in these commodity prices. However, the standard deviations vary significantly across the different commodities, with some like zinc and soybean showing high variability, reflecting greater price fluctuations. Soybean, gold, and zinc price changes have notably high positive medians, suggesting a general upward trend in their prices. On the other hand, coffee and natural gas exhibit negative and near-zero median values, indicating a tendency for price stability or slight declines.

## III. RESULTS

### *Linear Regression using PCA*

We use a standard method of determining the principal components (PCs) that capture the bulk of the variability of the underlying. First, we compute the z-scores of monthly changes in all the explanatory variables, to remove dimensionality. When variables are grouped, PCA focuses on explaining the variance within each group, meaning the principal components reflect the dominant patterns within those specific groups. However, when all variables are taken together, the principal components prioritize the overall variance in the dataset, leading to shifts in the principal component structure and revealing relationships that may not be representative of each specific group (Kritzman, Li, Page, & Rigobon, 2011). In other words, variables that may not capture the variation of each group of explanatory variables can become significant contributors in explaining the overall variance (Adrian, Moench, & Shin, 2010). Therefore, following standard practice, we have determined the PCs by the groups – economy-wide, fixed income, stock market and commodity variables.

We then calculate the Explained Variance Ratio for each principal component. We examine the cumulative variance and select the number of principal components required to achieve 75% cumulative interpretation variance (see Jolliffe, 2002). For example, the cumulative variance from PC1 to PC6 is 0.780598 for economy-wide variables, which exceeds 75%, so we choose the first 6 principal components. The same approach applies to other groups.

**Table 5: Variance Explained by each Principal Component**

*This table shows the variation of the underlying variables explained by each PC (Principal component) and the cumulative variance explained, for economy-wide variables (Panel A), fixed-income variables (Panel B), stock market variables (Panel C), and commodity variables (Panel D). In bold are shown the PC, up to which explain at least 75% of the variability of the underlying variables.*

**Panel A:** Monthly changes in Economy-wide Variables PC Explained Variance

| PC | Explained Variance | Cumulative Variance |
|---|---|---|
| PC1 | 0.218533 | 0.218533 |
| PC2 | 0.152287 | 0.37082 |
| PC3 | 0.118664 | 0.489484 |
| PC4 | 0.108884 | 0.598368 |
| PC5 | 0.098247 | 0.696615 |
| **PC6** | **0.085739** | **0.782354** |
| PC7 | 0.075755 | 0.858109 |
| PC8 | 0.066458 | 0.924568 |
| PC9 | 0.05685 | 0.981418 |
| PC10 | 0.018582 | 1 |

**Panel B:** Monthly changes in Fixed Income Variables PC Explained Variance

| PC | Explained Variance | Cumulative Variance |
|---|---|---|
| PC1 | 0.512412 | 0.512412 |
| **PC2** | **0.240501** | **0.752913** |
| PC3 | 0.098469 | 0.851382 |
| PC4 | 0.073137 | 0.924519 |
| PC5 | 0.038981 | 0.9635 |
| PC6 | 0.018625 | 0.982125 |
| PC7 | 0.007523 | 0.989648 |
| PC8 | 0.006996 | 0.996644 |
| PC9 | 0.002348 | 0.998992 |
| PC10 | 0.001008 | 1 |

**Panel C:** Monthly Stock Market Returns PC Explained Variance

| PC | Explained Variance | Cumulative Variance |
|---|---|---|
| **PC1** | **0.818558** | **0.818558** |
| PC2 | 0.059178 | 0.877736 |
| PC3 | 0.035813 | 0.913549 |
| PC4 | 0.029085 | 0.942634 |
| PC5 | 0.021458 | 0.964092 |
| PC6 | 0.014734 | 0.978826 |
| PC7 | 0.009985 | 0.988812 |
| PC8 | 0.00504 | 0.993851 |
| PC9 | 0.004048 | 0.9979 |
| PC10 | 0.001934 | 0.999834 |
| PC11 | 0.000155 | 0.999989 |
| PC12 | 0.000011 | 1 |

**Panel D:** Monthly changes in Commodity Prices PC Explained Variance

| PC | Explained Variance | Cumulative Variance |
|---|---|---|
| PC1 | 0.265532 | 0.265532 |
| PC2 | 0.118514 | 0.384046 |
| PC3 | 0.114201 | 0.498247 |
| PC4 | 0.102295 | 0.600542 |
| PC5 | 0.092274 | 0.692816 |
| **PC6** | **0.078743** | **0.771559** |
| PC7 | 0.069366 | 0.840926 |
| PC8 | 0.064938 | 0.905863 |
| PC9 | 0.051243 | 0.957106 |
| PC10 | 0.042894 | 1 |

As Table 5 shows, 6 PCs are needed to explain at least 75% of variability in monthly changes in economy-wide variables and monthly changes in commodity prices, only 2 for monthly changes in fixed income variables (which arguably move together), only 1 for monthly stock market returns. Using the important PC's from Table 5 (that explain at least 75% of the variability), we regress monthly change in TR CAPE on these important PC's to find the significant ones.‡ We find monthly changes in economy-wide PC2 and PC3, monthly changes in Commodity Prices PC1, and PC6 are the only significant ones, as shown in Table 6 below.

To find the loadings of the original variables for each significant PC (from Table 6), the significant PC loading information is first extracted from PCA results. We then calculate the loading of each significant PC on the original variable in its group, and select variables with an absolute value greater than 0.5, indicating that these variables

‡ If we use only the first two PCs from each group, for example, we find fewer PC's were significant in explaining changes in TR CAPE, which leads to the result (not consistent with the other methods) that none of the original explanatory variables associated with these principal components were found to be significant for explaining changes in TR CAPE. Thus, using all PCs that capture the bulk of the variability in each group of explanatory variables is important.

have a significant contribution to the principal component (see Yang, Florescu, and Islam, 2020).[§]

**Table 6: Important PCs for changes in monthly TR CAPE**

*This table shows the regression coefficients and the associated t statistics (in parenthesis) of only the significant PCs (at the 1% level) from out of all the explanatory variables that are significantly associated with monthly changes in all explanatory variables.*

| Significant PCs | Monthly change in TR CAPE |
|---|---|
| Monthly changes in economy-wide variables PC2 | -0.1320<br>(-2.639) |
| Monthly changes in economy-wide variables PC3 | -0.2574<br>(-4.426) |
| Monthly changes in commodity prices PC1 | -0.1421<br>(-3.801) |
| Monthly changes in commodity prices PC6 | -0.2498<br>(-3.494) |

**Table 7: Loadings of Original Variables on PCs**

*The two panels of this table show the loading of the significant PCs on the original explanatory variables. In bold are the ones with "significant" loadings.*

**Panel A:** Loadings of Original Variables on economy-wide variables PC2 and PC3

| Original explanatory variable | PC2 | PC3 |
|---|---|---|
| Change in US GDP index | -0.427662 | 0.354378 |
| Change in US CPI index | 0.416444 | 0.041977 |
| Change in Core inflation | 0.321924 | -0.057046 |
| Change in Federal Funds Rate | -0.258503 | -0.279961 |
| Change in Unemployment rate | 0.363341 | 0.190894 |
| Change in Industrial production | -0.459959 | 0.287033 |
| Change in M2 money supply | 0.005516 | 0.302857 |
| Change in Personal income index | -0.061184 | -0.014348 |
| Change in Producer price index | 0.307141 | 0.0161 |
| **Change in Michigan sentiment index** | **-0.177574** | **-0.761581** |

[§] Sticking to a 0.5 cutoff point, as highlighted in the literature by Yang, Florescu, and Islam (2020), is consistent with prior studies that advocate for selecting variables with significant loadings (i.e., above 0.5) to reduce model complexity while retaining essential variability. Using a standardized cutoff like 0.5 ensures comparability with other studies and reduces overfitting, as supported by Jolliffe (2002) in the context of PCA.

**Panel B:** Loadings of Original Variables on Commodity Prices PC1 and PC6

| Original explanatory variable | PC1 | PC6 |
|---|---|---|
| Sugar price change | -0.216772 | 0.185001 |
| Wheat price change | -0.426898 | 0.017523 |
| Coffee price change | -0.407641 | 0.382389 |
| Soybean price change | -0.435825 | -0.029871 |
| Tobacco price change | -0.037542 | 0.003803 |
| Crude Oil price change | -0.31942 | -0.153091 |
| Natural Gas price change | -0.165806 | 0.360197 |
| Gold price change | -0.320895 | 0.281196 |
| **Zinc price change** | **-0.290882** | **-0.71722** |
| Rice price change | -0.310061 | -0.26774 |

Finally, we regress monthly change in TR CAPE on these important original variables.

**Table 8: Significant Original Variables for Monthly Change in TR CAPE**

*This table shows the regression coefficients and the associated t statistics (in parenthesis) of significant original variables (from Table 7) on monthly Change in TR CAPE. In bold are the significant original variables (at 1% level).*

| Original explanatory variable | Monthly change in TR CAPE |
|---|---|
| Change in Michigan sentiment index | **0.3058** |
| | **(5.770)** |
| Zinc price change | **0.2268** |
| | **(4.279)** |

From Table 8, we find that both consumer sentiment and commodity prices can influence investor expectations about future economic conditions. We find that monthly change in Michigan sentiment index and Zinc price change are significantly associated with monthly changes in TR CAPE. As discussed earlier, a rise (drop) in the Michigan Sentiment Index can lead to higher (lower) stock prices and an increase (decrease) in the TR CAPE ratio. Changes in zinc prices can be indicative of global economic activity. A rise in zinc prices may suggest increased industrial demand, which could be associated with economic growth and higher corporate earnings, potentially driving up stock prices and the TR CAPE ratio (Fattouh, 2011).

As a check, we examine the determinants of annual change in TR CAPE next. We just report only the final results in Table 9 (that corresponds to the results in Table 8 for monthly changes in TR CAPE).

**Table 9: Significant Original Variables for Annual Changes in TR CAPE**

*This table shows the regression coefficients and the associated t statistics (in parenthesis) of significant original variables (on annual changes in TR CAPE. In bold are the significant original variables (at 1% level).*

| Original explanatory variable | Annual change in TR CAPE |
|---|---|
| Change in Michigan sentiment index | **2.1875** |
| | **(8.008)** |
| Zinc price change | **1.3037** |
| | **(4.960)** |
| Sugar price change | 0.0110 |
| | (0.046) |
| Change in M2 money supply | 0.4457 |
| | (1.732) |
| Change in Core inflation | **-0.8024** |
| | **(-3.077)** |

Consistent with the results for the monthly changes in TR CAPE, annual change in Michigan sentiment index, and Zinc prices are significantly associated annual change sin TR CAPE. The additional explanatory variable is change in core inflation. Higher core inflation can lead to increased interest rates, which can lower stock market valuations, and affect TR CAPE.

**Lasso Regression**

LASSO (Least Absolute Shrinkage and Selection Operator) performs variable selection and regularization by imposing L1 penalties on the regression coefficients. LASSO regression reduces some regression coefficients to zero, effectively selecting the variables that are most important to the response variable. We change the penalty function and only report the non-zero coefficient variables in different penalty function scenarios for comparative analysis with Lasso. Recent advancements have integrated non-convex penalties like smoothly clipped absolute deviation (SCAD) (Fan and Li, 2012), and minimax concave penalty (MCP) (Breheny and Huang, 2015) to enhance variable selection consistency and reduce bias. Zou (2006) introduces the adaptive Lasso method, which improves feature selection by applying different penalties to different coefficients, thereby addressing challenges posed by high-dimensional data. This method enhances the Lasso's ability to select significant features by adapting the penalty weights based on the importance of the features, often informed by prior information such as correlation or other criteria. Therefore, using adaptive Lasso, SCAD and MCP, we get the following results.

From Table 10, we find that monthly changes in Michigan sentiment index and the 5-year Treasury Yield are associated with monthly changes in TR CAPE, consistently across the 3 methods. We have already discussed the importance of consumer sentiment. Interest rates, particularly the 5-year Treasury yield, can be another critical factor in stock valuations. This rate is often used as a benchmark for discounting future earnings in stock valuation models. Asness (2000) discusses how changes in interest rates affect the equity risk premium. Campbell and Shiller (1988) also explored the relationship between bond yields and equity prices, noting that rising bond yields can make bonds more attractive relative to stocks.

**Table 10: Non-Zero Coefficients for monthly Change in TR CAPE**

*This table shows the non-zero coefficients using three different LASSO regression met hods – adaptive LASSO, SCA and MCP. In bold are the coefficients that are non-zero across all 3 methods.*

| | Adaptive Lasso | SCAD | MCP |
|---|---|---|---|
| Change in Core inflation | | -0.000756 | |
| Change in Michigan sentiment index | **0.0728** | **0.111027** | **0.076238** |
| Change in 5Y Treasury Yield | **0.140772** | **0.292441** | **0.266117** |
| Change in HQMCB10YR | | -0.208182 | -0.20448 |
| Change in Moody's BAA rate | -0.069698 | | |
| Dow Jones return | | 0.386654 | 0.386594 |
| Russell 5000 return | 0.357357 | | |
| Nikkei 225 (Japan) | 0.093038 | 0.064351 | |
| TSX Composite Index (Canada) | 0.015752 | | |
| Tobacco price change | | -0.036195 | |
| Change in TED spread | -0.009251 | | |
| FTSE 100 (UK) | 0.014113 | | |

We check results with annual changes in TR CAPE. From Table 11, we find that, in addition to the explanatory variables found important for monthly changes in TR CAPE – change in Michigan sentiment index and the 5-year Treasury Yield - annual change in US GDP index, change in Core inflation and M2 money supply are also significantly associated with annual change in TR CAPE, consistently across the 3 methods. Barro (1990) demonstrated that the growth rate in GDP is a significant determinant of stock market valuations via increased corporate earnings. Christiano, Motto, and Rostagno (2003) explored the relationship between monetary aggregates like M2 and stock market valuations via liquidity.

## Ridge and ElasticNet Regressions

The optimization of Ridge regression using techniques such as matrix differential calculus, as initially explored by Hoerl and Kennard (1970), highlighting the potential of this method to improve model performance, particularly in addressing multicollinearity issues in regression analysis.In Ridge regression, to determine which variable is the primary determinant, we can look at the absolute value of the coefficient. Gelman and Hill (2007) suggest that coefficients with absolute values greater than 0.1 (when predictors are standardized) may be considered large enough to be practically significant. We can use a combination of statistical significance (p-values) and practical significance (magnitude of standardized coefficients) to define a threshold for "large enough" coefficients.

**Table 11: Non-Zero Coefficients for annual changes in TR CAPE.**

*This table shows the non-zero coefficients using three different LASSO regression met hods – adaptive LASSO, SCA and MCP. In bold are the coefficients that are non-zero across all 3 methods.*

| Variables | **Adaptive Lasso** | **SCAD** | **MCP** |
|---|---|---|---|
| **Change in US GDP index** | **0.062228** | **0.859216** | **0.886479** |
| Change in US CPI index | -0.062286 | | |
| **Change in Core inflation** | **-0.50784** | **-0.603869** | **-0.597813** |
| **Change in M2 money supply** | **0.022026** | **0.292416** | **0.309301** |
| **Change in Michigan sentiment index** | **0.286443** | **0.773844** | **0.910871** |
| **Change in 5Y Treasury Yield** | **0.570288** | **0.758615** | **1.547698** |
| Change in 10Y Treasury Yield | 0.212297 | 0.613633 | |
| Change in Moody's BAA rate | -0.521703 | -0.551752 | |
| Dow Jones return | 2.081543 | | |
| Nikkei 225 (Japan) | 0.109781 | | |
| Hang Seng Index (Hong Kong) | 0.161715 | 0.163172 | |
| Coffee price change | 0.032653 | 0.054182 | |
| Gold price change | -0.080118 | | |
| EWCRSP index return | | 0.073658 | |
| DAX (Germany) | | 0.605722 | 1.147329 |
| Change in Unemployment rate | | -0.197706 | |
| Change in Producer price index | | 0.256226 | 0.364755 |
| Change in 1M Treasury Yield | | -0.306875 | -0.452185 |
| Change in 30Y Treasury Yield | | 0.18989 | |
| Change in HQMCB10YR | | -0.463175 | -0.739327 |

ElasticNet is chosen for its ability to combine the advantages of LASSO and Ridge regression, providing a robust approach to handling high-dimensional data and multicollinearity. It offers a balanced solution by performing variable selection and ensuring model stability, leading to improved predictive performance. The integration of both L1 and L2 penalties allows ElasticNet to address the limitations of each method individually, making it a preferred choice certain financial data analyses. ElasticNet can perform automatic variable selection, setting some coefficients to zero and thus simplifying the model (Zou and Hastie, 2005). The L2 penalty in Ridge regression helps ensure model stability and reduces sensitivity to small changes in the data, effectively addressing the limitations of purely L1 penalization in LASSO, particularly in situations involving multicollinearity (Hoerl and Kennard, 1970).

From Table 12, we find that monthly change in Michigan sentiment index, and 5Y Treasury Yield (consistent with Lasso regression methods) and the change in HQMCB10YR are associate with monthly change in TR CAPE, across the 2 methods. HQMCB10YR, which represents the yield on high-quality corporate bonds, impacts corporate financing costs and the relative attractiveness of stocks versus bonds. When the yield on these bonds increases, it can make bonds more attractive relative to equities, potentially leading to lower stock prices (see, for example, Adrian, Moench, and Shin (2010)).

**Table 12: Significant Coefficients for monthly change in TR CAPE**

*This table shows the coefficients using 2 different regression methods –Ridge and ElasticNet. In bold are the coefficients that are non-zero across both methods.*

| Variables | RIDGE | ElasticNet |
|---|---|---|
| Change in Michigan sentiment index | **0.128087** | **0.115114** |
| Change in 5Y Treasury Yield | **0.152981** | **0.120158** |
| Change in 10Y Treasury Yield | 0.146129 | |
| Change in 30Y Treasury Yield | 0.122688 | |
| Change in HQMCB10YR | **-0.191198** | **-0.102890** |
| Change in Moody's AAA rate | -0.138462 | |
| Dow Jones return | 0.164454 | |
| DAX (Germany) | 0.169977 | |
| CAC 40 (France) | -0.203325 | |

We check results with annual changes in TR CAPE.

**Table 13: Significant Coefficients for annual change in TR CAPE**

*This table shows the coefficients using 2 different regression methods –Ridge and ElasticNet. In bold are the coefficients that are non-zero across both methods.*

| Variables | RIDGE | ElasticNet |
|---|---|---|
| Change in US GDP index | 0.176222 | |
| Change in US CPI index | -0.154916 | |
| **Change in Core inflation** | **-0.306288** | **-0.531418** |
| Change in Unemployment rate | -0.114378 | |
| Change in Industrial production | 0.107554 | |
| Change in M2 money supply | 0.134871 | |
| **Change in Michigan sentiment index** | **0.320523** | **0.298494** |
| **Change in 5Y Treasury Yield** | **0.228477** | **0.628361** |
| Change in 10Y Treasury Yield | 0.220049 | |
| Change in 30Y Treasury Yield | 0.164942 | |
| Change in HQMCB10YR | -0.116616 | |
| **Change in Moody's BAA rate** | **-0.278571** | **-0.498345** |
| **Dow Jones return** | **0.386095** | **2.089984** |
| S&P 500 return | 0.278705 | |
| Russell 5000 return | 0.272973 | |
| VWCRSP index return | 0.280822 | |
| EWCRSP index return | 0.22005 | |
| DAX (Germany) | 0.157595 | |
| **Nikkei 225 (Japan)** | **0.207744** | **0.127994** |
| **Hang Seng Index (Hong Kong)** | **0.214298** | **0.176161** |
| Sugar price change | -0.101073 | |
| Coffee price change | 0.132685 | |

From Table 13, we find that, in addition to the explanatory variables found important for monthly changes in TR CAPE – change in Michigan sentiment index and the 5-year Treasury Yield - we find that annual change in Core inflation, Moody's BAA rate, Dow Jones return, Nikkei 225 (Japan) and Hang Seng Index (Hong Kong) are all also associated with annual change in TR CAPE, across the 2 methods.

## IV. Conclusion

Recent advances in data science and machine learning have further enhanced the use of regression techniques in financial studies. Gu, Kelly, and Xiu (2020), for example, have explored empirical asset pricing using machine learning, demonstrating significant improvements in predictive accuracy and robustness. These advancements illustrate the ongoing evolution of regression techniques, driven by the need to handle increasingly complex financial datasets and improve the reliability of financial models. We do the same in this paper. We analyze the determinants of changes in S&P Composite Index Total Return Cyclically Adjusted Price-to-Earnings ratio (TR CAPE), to better understand changing "investor exuberance". We use three different methods - linear regression using PCA, Lasso, and Ridge regression techniques – and a large number of explanatory variables, to compare and contract significant determinants.

The results differ with the method used. But, across all the methods we use, monthly changes in Michigan sentiment index is significantly associated with monthly changes in TR CAPE. Cross-checking these results using annual changes in TR CAPE and annual changes in explanatory variables, across all methods, we find that annual changes in Michigan sentiment index and in core inflation are significantly associated with annual changes in TR CAPE.

Overall, changes in the Michigan Sentiment Index appears to have significant association with changes in investor exuberance. This is a measure of consumer sentiment, when high, typically reflects optimism about future economic growth, leading to increased consumer spending and higher corporate earnings expectations. This positive outlook can boost investor confidence, driving up stock prices and, consequently, increasing the TR CAPE ratio as markets anticipate stronger future earnings. In contrast, the financial crisis of 2008, for example, led to a sharp decline in the Michigan Sentiment Index as consumer confidence plummeted due to fears of a prolonged recession.

## References

Adrian, T., Moench, E., & Shin, H. S. 2010. Financial intermediation, asset prices, and macroeconomic dynamics. Federal Reserve Bank of New York Staff Reports, 422.

Alexander, C. 2001. *Market models: A guide to financial data analysis*.

Ang, A., & Piazzesi, M. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50(4), 745-787.

Asness, C. S. 2000. Stocks versus bonds: Explaining the equity risk premium. *Financial Analysts Journal*, 56(2), 96-113.

Baker, M., & Wurgler, J. 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645-1680.

Baker, M., & Wurgler, J. 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129-151.

Barro, R. J. 1990. The stock market and investment. *Review of Financial Studies*, 3(1), 115-131.

Baur, D. G., & McDermott, T. K. 2010. Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8), 1886-1898.
Bekaert, G., & Harvey, C. R. 1997. Emerging equity market volatility. *Journal of Financial Economics*, 43(1), 29-77.
Bessler, D. A., & Kling, J. L. 1989. Agricultural prices and the broader economy: Evidence from the United States. *American Journal of Agricultural Economics*, 71(2), 543-552.
Breheny, P., & Huang, J. 2015. Generalized regression estimators with concave penalties and a comparison to lasso type estimators. *METRON*.
Brown, G. W., & Cliff, M. T. 2004. Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1), 1-27.
Campbell, J. Y., & Shiller, R. J. 1988. Stock prices, earnings, and expected dividends. *Journal of Finance*, 43(3), 661-676.
Campbell, J. Y., & Ammer, J. 1993. What moves the stock and bond markets? A variance decomposition for long-term asset returns. *Journal of Finance*, 48(1), 3-37.
Campbell, J. Y., & Shiller, R. J. 1998. Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, 24(2), 11-26.
Campbell, J. Y., & Viceira, L. M. 2002. Strategic asset allocation: Portfolio choice for long-term investors. *Oxford University Press*.
Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. 1994. Does consumer sentiment forecast household spending? If so, why? *American Economic Review*, 84(5), 1397-1408.
Chen, N., Roll, R., & Ross, S. A. 1986. Economic forces and the stock market. *Journal of Business*, 59(3), 383-403.
Christiano, L. J., Motto, R., & Rostagno, M. 2003. The Great Depression and the Frieman-Schwartz hypothesis. *Journal of Money, Credit, and Banking*, 35(6), 1119-1197.
Connor, G., & Korajczyk, R. A. 1986. Performance measurement with the arbitrage-pricing theory: A new framework for analysis. *Journal of Financial Economics, 15*(0), 373-394.
DeMiguel, V., Garlappi, L., and Uppal, R. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5), 1915-1953.
Elliott, G., & Timmermann, A. 2008. Economic forecasting. *Journal of Economic Literature, 46*(0), 3-56.
Erten, B., & Ocampo, J. A. 2013. Super cycles of commodity prices since the mid-nineteenth century. World Development, 44, 14-30.
Fama, E. F.1970. Efficient Capital Markets: A Review of Theory and Empirical Work, *The Journal of Finance*, 25 (2), 383-417.
Fama, E. F., & Schwert, G. W. 1977. Asset returns and inflation. *Journal of Financial Economics*, 5(2), 115-146.
Fama, E. F., & French, K. R. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1), 23-49.
Fama, E. F. 1990. Stock returns, expected returns, and real activity. *Journal of Finance*, 45(4), 1089-1108.
Fan, J., & Li, R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96*(0), 1348-1360.
Fan, J., & Li, R. 2012. Variable selection in linear mixed effects models. *Annals of Statistics, 40*(4), 2043-2064.

Fattouh, B. 2011. An anatomy of the crude oil pricing system. *Oxford Institute for Energy Studies.*

Fisher, K. L., & Statman, M. 2000. Investor sentiment and stock returns. *Financial Analysts Journal*, 56(2), 16-23.

Friedman, M., & Schwartz, A. J. 1963. A monetary history of the United States, 1867-1960. *Princeton University Press*.

Froot, K. A., & Ramadorai, T. 2008. Institutional portfolio flows and international investments. *Review of Financial Studies*, 21(2), 937-971.

Gelman, A., & Hill, J. 2007. Data analysis using regression and multilevel/hierarchical models. *Cambridge University Press*.

Gu, S., Kelly, B., and Xiu, D. 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.

Hamilton, J. D. 2009. Understanding crude oil prices. *Energy Journal*, 30(2), 179-206.

Hamilton, J. D. 2009. Causes and consequences of the oil shock of 2007-08. *Brookings Papers on Economic Activity*, 40(1), 215-261.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The elements of statistical learning: Data mining,inference, and prediction* (2nd ed.). Springer.

Hastie, T., Tibshirani, R., and Wainwright, M. 2015. *Statistical learning with sparsity: The lasso and generalizations.* CRC Press.

Hoerl, A. E., & Kennard, R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(0), 55-67.

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. *An introduction to statistical learning: With applications in R*.

Jolliffe, I. T. 2002. *Principal component analysis* (2nd ed.). Springer.

Kritzman, M., Li, Y., Page, S., and Rigobon, R. 2011. Principal components as a meaure of systemic risk. *The Journal of Portfolio Management*, 37(4), 112-126.

Lemmon, M., & Portniaguina, E. 2006. Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, 19(4), 1499-1529.

Litterman, R., & Scheinkman, J. 1991. Common factors affecting bond returns. *Journal of Fixed Income, 1*(0), 54-61.

Ludvigson, S. C. 2004. Consumer confidence and consumer spending. *Journal of Economic Perspectives*, 18(2), 29-50.

Mauboussin, M. J., & Callahan, D. 2014. *Measuring the moat: Assessing the magnitude and sustainability of value creation.* Credit Suisse.

Ng, A. Y. 2004. Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning*, 78.

Ng, S. 2014. Boosting regressions. *Journal of Econometrics, 178*(0), 351-362.

Schmeling, M. 2009. Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, 16(3), 394-408.

Shiller, R. J. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71(3), 421-436.

Shiller, R. J. 2005. Irrational Exuberance (2nd ed.). *Princeton University Press*.

Shmueli, G. 2010, To Explain or to Predict?, *Statistical Science* 2010 25(3), 289.

Siegel, J. J. 2016. *Stocks for the long run: The definitive guide to financial market returns & long-term investment strategies* (5th ed.). McGraw-Hill Education.

Stock, J. H., & Watson, M. W. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, 97*(0), 1167-1179.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(0), 267-288.

Wang, X., Zhu, Z., and Zhang, H. H. 2020. Spatial heterogeneity automatic detection and estimation. *Computational Statistics & Data Analysis*, 180, 107667.

Yang, S., Florescu, I., and Islam, M. T. 2020. Principal component analysis and factor analysis for feature selection in credit rating. *Journal of Risk and Financial Management*, 13(6), 117.

Zou, H., & Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(0), 301-320.

Zou, H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418-1429.