The Gettysburg Economic Review

Volume 11 | Article 3

2019

Are Price-Earnings Ratios Mean Reverting? An Empirical Study

Kevin Klassen
Gettysburg College
Class of 2019

Are Price-Earnings Ratios Mean Reverting? An Empirical Study

Abstract
Mean reversion in stock prices is a highly studied area in the financial literature with controversial findings. While some economists have found evidence of mean reverting processes in stock prices, many argue in favor of the Efficient Market Hypothesis which states stock prices are random walk processes. This paper seeks to add to the literature on mean reversion but testing for evidence in price-earnings ratios rather than stock prices. The study employs a robust regression model controlling for company-specific and general market factors that influence price-earnings ratio deviations. After correcting for heteroskedasticity, serial correlation, and unit root processes, the results indicate mean reverting behavior does exist in US equities from 2008-2017 and mean reversion in price-earnings ratios may occur more quickly than mean reversion of stock prices. The outcome of this paper also implies some level of endogeneity in the Three-Factor-Model proposed by Fama and French (1992).

Keywords
mean reversion, stock market, Efficient Market Hypothesis, regression model, Three-Factor-Model

Are Price-Earnings Ratios Mean Reverting? An Empirical Study

Kevin Klassen
Gettysburg College
klaske01@gettysburg.edu
April 2019

Abstract

Mean reversion in stock prices is a highly studied area in the financial literature with controversial findings. While some economists have found evidence of mean reverting processes in stock prices, many argue in favor of the Efficient Market Hypothesis which states stock prices are random walk processes. This paper seeks to add to the literature on mean reversion but testing for evidence in price-earnings ratios rather than stock prices. The study employs a robust regression model controlling for company-specific and general market factors that influence price-earnings ratio deviations. After correcting for heteroskedasticity, serial correlation, and unit root processes, the results indicate mean reverting behavior does exist in US equities from 2008-2017 and mean reversion in price-earnings ratios may occur more quickly than mean reversion of stock prices. The outcome of this paper also implies some level of endogeneity in the Three-Factor-Model proposed by Fama and French (1992).

Acknowledgements:

I would like to thank Professor Cadigan, Professor Murphy, Professor Cushing-Daniels, Professor Weise, Professor O'Brien, Polina Rozhkova, and Luca Menicali for helpful suggestions in my topic selection, data organization, empirical analysis and paper structure.

I. Introduction

Whether stock prices and ratios can be described as random walk or mean reverting processes is highly controversial within the financial literature. Mean reversion refers to a tendency of asset prices or ratios to return to a trend path. This paper sets out to examine whether the price-earnings (P/E) ratios of US companies have transitory components and thus exhibit mean reverting behavior. Fama and French (1988) and Poterba and Summers (1987) are among the first to provide direct empirical evidence that mean reversion occurs in US stock prices over long horizons. At the same time, other economists are critical of their results. Richardson and Stock (1989) and Richardson (1993) report that correcting for small sample bias may reverse the results found by Fama and French and Poterba and Summers mentioned above. Moreover, Kim et al. (1991) argue that mean reversion is a pre-World War II phenomenan and current stock prices exhibit mean averting behavior.

The question of whether stock price-earnings contain transitory components poised in this paper is important for financial practice and theory. For example, consider technical analysis of stock price movements. If stock price-earnings ratios contain large transitory components, then observing a stock with a P/E ratio statistically far from its mean may establish a trend that could be traded technically. The notion of stock price trends is harshly rejected by many economists who argue in favor of the Efficient Market Hypothesis (EMH), which states that share prices reflect all information about a security, including information derived from fundamental and technical analysis. Therefore, it is theoretically impossible to consistently produce risk-adjusted excess returns, or alpha, and only inside information can result in outsized risk-adjusted returns.

This paper can also be used to evaluate the claims made by Keynes in his book *The General Theory of Employment, Interest and Money* (1936) where he states, "all sorts of considerations enter into market valuation which are in no way relevant to the prospective yield." Poterba and Summers (1987) state that "if divergences between the market and fundamental value of a stock exist, but at beyond some limit are eliminated by speculative forces, then stock prices exhibit mean reversion." Thus, if Keynes' claim is true and the psychology of speculators can cause the market valuation of stocks to diverge from their fundamental values, evidence of mean reversion in P/E ratios should exist.

Lastly, the results of this paper could have interesting implications on the Three-Factor-Model proposed by Fama and French (1992). To expand on the traditional Capital Asset Pricing Model (CAPM), Fama and French suggest stock returns are explained by size and valuation factors in addition to market risk. The valuation factor they employ is related to book-to-market value of a stock, which is highly correlated to the price-earnings ratio. If P/E ratios are mean reverting processes, there may be endogeneity in their valuation factor that is not properly accounted for. A more in-depth discussion of these implications is located in the Theory and Methodology section.

This study fits in an extensively researched section of the financial literature but seeks to test for mean reverting behavior in stock price-earnings ratios rather than stock prices and utilizes a slightly different methodology than those used by economists such as Fama and French (1988). I utilize quarterly stock and sector data gathered from Bloomberg. The sample period ranges from 2008 to 2017. The outcome variable of interest is the distance of the current P/E ratio from its trailing five-year average and the explanatory variable of interest is its lagged value. This is a similar model used to test for mean reversion in stock prices by Balvers et al

(2000), but I introduce several more controls to achieve more accurate estimators. Moreover, much of the previous literature employs variance ratio tests and standard unit root tests for mean reversion. However, econometric studies by Campbell and Perron (1991), Cochran's (1991) and DeJong et al. (1992) indicate that standard unit roots tests have very low power against local stationary alternatives in small samples. Further, Zhen (2010) argue that panel data can be used to generate more accurate unit root estimation. In this paper, I employ a linear regression model using panel data from S&P 500 companies to test for mean reverting processes in price-earnings ratios.

While most of the previous literature examine stock price mean reversion, these results can be misleading. Stock price movements occur for a wide variety of reasons, many of which are either difficult or impossible to isolate. So, it will be difficult to isolate a reversion coefficient due to potential endogeneity from many unobserved variables. The price-earnings ratio of a company has well-grounded determinants, including expected growth, consistency of dividends, company size, and extent of analyst converge, to name a few. Including these variables as controls in a regression will allow me to get a more accurate and unbiased estimation of the presence of mean reverting behavior.

The estimators of interest used in this paper are likely subject to several statistical biases due to the nature of the data. Issues that I found to be present through the use of rigorous econometric testing are heteroskedasticity, serial correlation and unit roots. The paper addresses them by employing heteroskedastic-robust standard errors while differencing and detrending each variable. There is also likely to be survivorship bias and small sample bias present in this analysis. I address the former by using both time-series panels and pooled panels but fail to address the latter due to limited time and resources. Regression results from both datasets support the mean reverting hypothesis, showing evidence of mean reverting processes in US company price-earnings ratios from 2008-2017.

The remainder of the paper is organized as follows. Section II will discuss the previous literature on stock mean reversion and the relationship between price-earnings ratios and stock returns. Section III will describe the theory behind my model and define the methodology used to achieve unbiased estimators of my coefficients. Further, Section IV will review the data used to address the research question. Finally, Section V will examine the results of the regression output and will be followed by a comprehensive conclusion for this paper.

II. Literature Review

Most of the existing literature relating to this topic simply employs unit root and stationarity tests to detect mean reversion in stock prices. There is also controversy over the existence of mean reverting behavior in financial assets. Many economists argue in favor of the EMH which, as mentioned in the introduction, asserts that all asset prices follow a random walk and thus, do not exhibit mean reverting behavior. Some economists have found evidence of mean reverting behavior through the use of variance ratio tests, but others believe correcting for biases negates their findings.

This paper focuses on mean reverting behavior in price-earnings ratios, which has not been widely studied in the financial literature. However, P/E ratios have been researched extensively on their relation to excess returns in equity markets and as a determinant of equity prices. Basu (1977) conducts an empirical study to test whether P/E ratios are related to investment performance in common stocks. He does so by creating five diversified portfolios, each with different portfolio P/E ratios. The results indicate lower P/E stocks are underpriced relative to the market and tend to experience the highest unexplained excess returns. A study by

Gill et al. (2012) finds that price-earnings ratios explain a significant portion of the variation in equity share prices in the United States.

There have been ample studies into stock price mean reversion with conflicting results. A study by Poterba and Summers (1987) aims to test whether transitory components account for a large fraction of the variance in common stock returns using variance ratio tests. Using data on firms form the United States and 17 other countries over the period 1926 – 1985, the authors find positive autocorrelation in stock returns over short horizons and negative autocorrelation over long horizons. They also report that transitory components in stock prices account for more than half of the variance in monthly returns. They conclude mean reversion does occur in stock prices and it is likely due to slowly-decaying "price fads" that cause stock prices to deviate from fundamental value. Fama and French (1988) provide further evidence of transitory stock price components in a study focusing on the relationship between dividend yield and stock returns. They find that the power of dividend yields to forecast stock returns increases with the return horizon and concluded this is likely due to time-varying expected returns generating temporary components of prices. Another study by Fama and French (1988) investigates the permanent and transitory components of stock prices during 1926 – 1985. They consider a time series dataset and employ variance ratio tests similar to those used by Poterba and Summers (1987). Their findings indicate a slowly mean-reverting component of stock prices tends to induce negative autocorrelation in returns.

More recent studies also indicate mean reverting behavior in stock prices. Mukherji (2011) uses a powerful nonparametric block bootstrap method and fresh data to examine the unresolved issue of mean reversion in stock returns. The results show that both large and small company stocks experienced significant mean reversion in returns for periods of 1 through 5 years during 1926–1966. In 1967–2007, there was significant mean reversion in 5 year returns of large company stocks, and 1, 4, and 5 year returns of small company stocks. The findings indicate that, although mean reversion in stock returns has weakened in recent decades, it persists, particularly for small company stocks. Another study uses panel data from national stock market indices of 18 countries from 1969 to 1996 (Balvers et al. 2000). They find strong evidence of mean reversion in relative stock index prices and a significantly positive speed of mean reversion with a half-life of three to three and a half years. According to their findings, investment strategies that fully exploit mean reversion across national indexes outperform buy-and-hold strategies.

Other publications reject the possibility of mean reverting behavior and argue in favor of the Efficient Market Hypothesis. Kim et al. (1991) compare stock return data before and after World War II. Using randomization tests to calculate significance levels under the null hypothesis that returns are distributed independently of their ordering in time, they find that mean reverting behavior is an entirely pre-war phenomena and current stock prices exhibit mean averting behavior. They interpret these results as evidence of a fundamental change in the stock return process and conjecture that it may be due to the resolution of the uncertainties of the 1930s and 1940s. A paper by Zhu (2010) asserts that conventional unit-root tests have weak power against stationary alternatives. His study uses unit-root tests in panel data to re-examine the time-series properties of the stock prices as unit root tests on panel data appear to have increased power of unit root tests. The results cannot reject the random-walk hypothesis for G-7 country stock-price indices. Richardson and Stock (1989) and Richardson (1993) develop an asymptotic distribution theory for statistics involving multiyear returns and correct for the small sample bias that they believe was present in previous mean reversion analyses. Their alternative theory provides substantially better approximations to the relevant finite-sample distributions used in conventional financial theory. It also leads to empirical inferences much less at odds with the hypothesis of no mean reversion and they claim their results may negate those found by Poterba and Summers (1987) and Fama and French (1988).

III. Theory and Methodology

A typical formulation of a stochastic process for an asset displaying mean reversion to a simple moving average, in this case in the asset's P/E ratio, is as follows:

$$\left|ln\left(\frac{P/E_t^i}{\overline{P/E_t^i}}\right)\right| = \pi^i + \lambda^i\left\{\left|ln\left(\frac{P/E_{t-1}^i}{\overline{P/E_{t-1}^i}}\right)\right|\right\} + \varepsilon_t^i, \quad (1)$$

where P/E_t^i is the price-earnings ratio of company i at time t, $\overline{P/E_t^i}$ is the trend price-earnings ratio of company i at time t, π^i is a constant, and ε_t^i is a stationary shock term with an unconditional mean of zero. The parameter λ^i measures the impact of increasing the distance of the previous P/E ratio from trend P/E ratio in the previous period on the distance of the current P/E ratio from trend P/E ratio in the current period. To accept the alternative hypothesis that mean reverting behavior exists in P/E ratios, λ^i must be statistically significant and $0 < \lambda^i < 1$. If $0 < \lambda^i < 1$, deviations in P/E ratio from the trend are reversed as *t* increases which, by definition, is mean reversion.

However, there is likely to be serial correlation as the model proposed is an autoregressive process of order one. Therefore, first differencing will be applied to equation (1) to yield the following:

$$\left|ln\left(\frac{P/E_t^i}{\overline{P/E_t^i}}\right)\right| - \left|ln\left(\frac{P/E_{t-1}^i}{\overline{P/E_{t-1}^i}}\right)\right| = \tau^i + \delta^i\left\{\left|ln\left(\frac{P/E_{t-1}^i}{\overline{P/E_{t-1}^i}}\right)\right| - \left|ln\left(\frac{P/E_{t-2}^i}{\overline{P/E_{t-2}^i}}\right)\right|\right\} + \Delta\varepsilon_t^i, \quad (2)$$

which can be written more simply as:

$$\Delta\left|ln\left(\frac{P/E_t^i}{\overline{P/E_t^i}}\right)\right| = \tau^i + \delta^i\left\{\Delta\left|ln\left(\frac{P/E_{t-1}^i}{\overline{P/E_{t-1}^i}}\right)\right|\right\} + \Delta\varepsilon_t^i. \tag{3}$$

The interpretation of δ^i is slightly different than in equation (1). In this case, the δ^i measures the speed of reversion between t – 1 and t. $\Delta\left|ln\left(\frac{P/E_{t-1}^i}{\overline{P/E_{t-1}^i}}\right)\right|$ represents the change in the distance of the observed P/E ratio for company i from the trend P/E ratio for company i from t – 2 to t – 1. If $\Delta\left|ln\left(\frac{P/E_{t-1}^i}{\overline{P/E_{t-1}^i}}\right)\right|$ has a positive value, that means the P/E ratio for company i diverged from the trend P/E ratio during the period between t – 2 to t – 1. Therefore, to accept the alternative hypothesis that mean reverting behavior exists in P/E ratios, δ^i must yield a statistically significant and negative result. An intuitive interpretation of this model is given that the P/E ratio of company i diverged from its trend value over the previous period, the P/E ratio should converge towards its trend value over the current period if mean reverting behavior exists.

To estimate equations (1) and (3), I will employ the following econometric models:

$$abs\_diff_{it} = \beta_0 + \beta_1 abs\_diff_{it-1} + \beta_i \boldsymbol{X}_{it} + e_{it}, \tag{4}$$

$$\Delta abs\_diff_{it} = \theta_0 + \theta_1 \Delta abs\_diff_{it-1} + \theta_i \Delta\boldsymbol{X}_{it} + \Delta e_{it}, \tag{5}$$

where "abs_diff_{it}" is the distance of company i's current P/E ratio in the current quarter from its 20-period simple moving average as a percentage. In the context of this paper, the 20 period simple moving average is equal to the trailing 5-year average P/E ratio for company i at time t. "abs_diff_{it-1}" is the one period lagged value of "abs_diff_{it}". "$\boldsymbol{X}_{it}$" is a vector representing the set of control variables used, which includes *pe_ratio*, *pe_ratio2*, *sector_delta*, *eps_growth*, *lvolume*, and *lmarket_cap*. Full descriptions of those variables can be found on Table A in the Appendix and are discussed further in the Data Section. The variable *pe_ratio* is used to control for companies that trade at unusually high P/E ratios; *pe_ratio2* is the squared value of *pe_ratio,* which is used to control for the decrease in marginal effect of increasing *pe_ratio* by one when the value of *pe_ratio* gets very large; *sector_delta* is used to control for business cycle changes where certain sectors tend to trade at higher or lower P/E ratios; *eps_growth* is a key determinant of P/E ratio as suggested by financial theory since companies that having accelerating earnings growth can often sustain expanding P/E ratios for extended periods of time; *lvolume* is used as a proxy for shock factors that may cause P/E ratios to deviate from their trend value since trading volume tends to increase when investor sentiment is highly positive or highly negative; *lmarket_cap* is used to control for the size of the company since larger companies tend to have more analysts covering their stock, so there is more information available for investors to consider when making an investment decision. Therefore, I would expect larger companies to trade closer to their trend values. The coefficient of interest for equations (4) and (5) are β_1 and θ_1, respectively, and the sign and significance of these coefficients will indicate whether mean reverting behavior is exhibited.

Due to the nature of the data, there are several econometric issues that will need to be addressed to get unbiased estimators. The four most important issues that this paper addresses are heteroskedasticity, survivorship bias, serial correlation, and unit roots. Heteroskedasticity is likely to exist in financial time series data, as indicated by the prior literature. Survivorship bias is likely present in the time series panel data used in this paper and is discussed further in the Data section of this paper.

Similar to many other financial time series datasets, the one used in this paper is likely to be serial correlated. One of the most important predictors of company *i*'s P/E ratio this quarter is the P/E ratio from the previous quarter. Specifically,

$$P/E_{it} = \xi + \phi P/E_{it-1} + \varepsilon_{it},$$

$$(6)$$

where ϕ is a statistically significant coefficient different from zero. This is likely an issue for most of the variables employed in my study. In addition, the similar issue of unit roots is likely to arise where $\phi = 1$ or is close to 1. Unit root processes occur when the stochastic process that determines the variable of interest is non-stationary and often appears in financial time series datasets.

The presence of these econometric issues is tested for and discussed in further detail in the Results section of this paper. The econometric technique that allows me to correct for heteroskedasticity is using heteroskedastic-robust standard errors. Survivorship bias is minimized through the use of a second pooled dataset but is not completely eliminated. To control for serial correlation and unit roots, differencing and detrending are applied to the model.

While it was not the original intent of the present study, the results yielded could have interesting implications for the three-factor model proposed by Fama and French (1992). Consider the Capital Asset Pricing Model (CAPM) shown in equation (7), which is used to determine a theoretically appropriate required rate of return of an asset to make decisions about adding assets to a well-diversified portfolio:

$$r_i = r_f + \beta_1(r_m - r_f) + \alpha_i + e_i. \tag{7}$$

Fama and French expand on the CAPM by including two more factors believed to explain the variation in required return of an asset. Market risk is still the primary determinant, but also included is a company size factor (*SMB*) and a company value factor (*HML*):

$$r_i = r_f + \beta_1(r_m - r_f) + \beta_2(SMB) + \beta_3(HML) + \alpha_i + e_i \tag{8}$$

Their results indicate that small companies tend to outperform large companies, and companies with high book-to-market ratios tend to outperform companies with low book-to-market ratios. The book-to-market ratio of a company is defined as the inverse of the price-to-book ratio, so higher book-to-market ratios indicate higher values. They found similar results from other value ratios; companies with lower price-earnings ratios tend to outperform companies with higher price-earnings ratios. The value factor they propose is of interest since the mean-reverting behavior of this factor is the focus of my study. If P/E ratios exhibit mean reverting behavior, there may be a mean reverting function that goes unmodeled in their three-factor specification. Thus, the value factor they propose may be endogenous, which would lead to bias in the estimation of β_3 in equation (8).

IV. Data

The datasets used in this paper were pulled from a Bloomberg Terminal and every variable is measured quarterly from the first quarter of 2008 to the last quarter of 2017. One dataset is organized in time-series panels where the same companies are followed from 2008Q1 to 2017Q4. The 50 companies used were randomly selected from the set of companies that have remained in the S&P500 from 2008Q1 to 2017Q4. This is likely to cause survivorship bias because companies that remain in the S&P500 for 10 years are likely high-quality companies. To address this issue, I use a second dataset where 50 companies are randomly selected for every observed year. The companies selected in the second dataset must have remained in the S&P500 for the entire year observed. This is meant to minimize survivorship bias and provide a robustness check for the results yielded by the time series panels. While the time series dataset only follows 50 companies, the pooled dataset follows 307 total companies and allows for the companies observed to have released shares to the public after 2008. The same variables were gathered from both datasets.

The outcome variable of interest is the absolute value of the percentage difference between the current price-earnings ratio and the trend price-earnings ratio. The main explanatory variable is lagged value of the outcome variable. The variables for the value of the P/E ratio and the P/E ratio squared as controls for companies that trade at unconventional P/E ratios. I include the squared term since as a company's P/E ratio gets very large, increasing the value by 1 will not have as large an effect on the percentage difference from its trend. I use the trailing 20-period simple moving average as a proxy for the trend value, which is used to calculate the outcome and explanatory variables. I use the change in P/E ratio of each sector as for the business cycle, as mentioned in the Theory and Methodology section. Finally, the earnings-per-share growth, log of market cap and log of per-period volume act as controls for the determinants of a company's P/E ratio.

V. Results

The empirical results from the different models tested both indicate mean reverting behavior in P/E ratios. However, as discussed in the Theory and Methodology section, the interpretations are slightly different. The simple AR(1) OLS regression is displayed in Tables I. The estimated coefficient on *abs_diff_1* is positive, between 0 and 1, and statistically significant at the 99% confidence level, which indicates mean reverting behavior is present in P/E ratios between 2008 and 2017. Increasing the distance of a company's current P/E from the company's historical P/E in the previous period by 1% is estimated to increase the difference in the current period by 0.7381%. The estimated coefficients from the simple OLS regression on the remaining variables are consistent with my expectations.

As mentioned earlier, to control for possible survivorship bias and as a robustness check, I run the same regression using a pooled panel dataset. The results of the simple OLS model are confirmed by the pooled regression, estimating very similar coefficients with the only one change in significance levels coming from *sector_delta*. The coefficient on *sector_delta* also changed signs, but the pooled regression was inconclusive with a t-statistic of 1.25.

As with many financial datasets, the datasets I use likely suffer from heteroskedasticity in the error terms. To test for the presence of heteroskedasticity, I run Breusch-Pagan's test for heteroskedasticity and the results for both datasets can be seen on Table E. As expected, both the time series panels and pooled panels suffer from heteroskedasticity. Therefore, I will continue my analysis using heteroskedastic-robust standard errors.

There is likely some time-constant, firm or sector specific unobserved factors lying in the error term that may contribute to the distance of a firm's P/E ratio from its historical average in both the time series panel and pooled panel regressions. Specifically,

$$abs\_diff_{it} = \beta_0 + \beta_1 abs\_diff_{it-1} + \beta_i \boldsymbol{X}_{it} + a_i + u_{it}, \quad (9)$$

where a_i represents the unobserved, time-constant factors and $e_{it} = a_i + u_{it}$. There are two methods primarily used to address this issue, demeaning and first differencing. First differencing can also be used to correct for serial correlation while demeaning cannot correct this issue by itself. I discussed the theoretical possibility of serial correlation within the datasets used in the Theory and Methodology section. After running the simple OLS, I find further evidence of serial correlation since the R^2 values from both time-series panels and pooled panels seems to be quite high relative to the R^2 values achieved by other papers in the financial literature. To confirm the presence of serial correlation in my dataset, I estimate the impact of the lagged values of each variable on the current period's value along with the impact of the lagged residuals gathered from a regression of each variable on the time trend. The results are displayed in Table F and Table G, respectively, and indicate that every variable suffers from serial correlation. One difference to note between the time-series and pooled datasets was the serial correlation was of significantly less magnitude in the pooled regression. This is likely because different companies were used every year and I would not expect variable x of company i in year t to be serially correlated to variable x of company j in year t+1. Nonetheless, there is evidence of serial correlation in both datasets, so I opt to apply differencing to each variable rather than demeaning. I also detrend each variable as detrending can also be used to address this type of bias.

To detrend, I regress each variable on the time trend t and gathered the residuals. More specifically, to detrend variable x_{it}, I estimate the model:

$$x_{it} = \alpha_0 + \alpha_1 t + e_{it}, \tag{10}$$

where $e_{it} = \ddot{x}_{it}$, which is the portion of the variation in x_{it} not explained by the time trend. Every variable is replaced with its detrended counterpart, so the model can now be written as:

$$abs\ddot{\_}diff_{it} = \beta_0 + \beta_1 abs\ddot{\_}diff_{it-1} + \beta_i \ddot{\boldsymbol{X}}_{it} + e_{it}, \tag{11}$$

where the accent over each variable indicates that is has been detrended. Next, I apply first differencing to each variable for i firms and proceed to estimate the following regression:

$$\Delta abs\ddot{\_}diff_{it} = \beta_0 + \beta_1 \Delta abs\ddot{\_}diff_{it-1} + \beta_i \Delta \ddot{\boldsymbol{X}}_{it} + \Delta u_{it.} \tag{12}$$

Notice that the time-constant unobserved factors a_i drops out, so I am left with serially uncorrelated and exogenous variables.

Furthermore, an issue that often arises when dealing with financial time series data is non-stationarity and random walks. These issues are known to be present if a variable follows a unit root process. More specifically, variable x_{it} follows a random walk with a drift, a special type of unit root process, when

$$x_{it} = \mu_0 + \rho_i x_{it-1} + e_{it}, \tag{13}$$

where ρ_i is not statistically different from one. Moreover, since I am estimating an AR(1) model, it is also important to see that the absolute value of ρ_i for every x_{it} is less than one to ensure I have a weakly dependent, stable AR(1) process. To test for the presence of unit roots, I run a modified form of the Augmented Dickey-Fuller test for panel data. I employ a Levin-Lin-Chu unit test, which involved fitting an augmented Dickey-Fuller regression for each panel (Balvers et al, 2000). One critical assumption that must hold for this test to yield accurate results is a common autoregressive parameter for all panels. This means that the test does not allow for the possibility that some panels contain unit roots while others do not. The results can be seen in Table H and indicate that only two variables do not display unit root processes in the time-series panels, and only one variable does not display a unit root process in the pooled. Conveniently, if a variable has a unit root process, the first difference of the variable is stationary. Differencing was already applied to correct for serial correlation, so I can continue my analysis using the model specified in equation (12).

The results of the first differenced and detrended model confirm the findings of the simple OLS model.

Variable	Simple OLS *Absolute Difference*	Detrended and Differenced (Robust SE) *Absolute Difference*
First Lag of Absolute Difference	0.7381*** (0.1333)	-0.1270*** (0.0339)
P/E Ratio	0.0037*** (0.0004)	0.0143*** (0.0014)
P/E Ratio^2	-3.00e-06*** (4.87e-07)	-0.00001*** (1.70e-06)
Sector P/E Ratio Delta	0.0001** (0.00005)	0.00005* (0.00003)
EPS Growth Trailing 1 Year	-9.99e-07 (6.63e-06)	2.54e-08 (1.67e-06)
Log of Volume	0.0191*** (0.0050)	0.0595*** (0.0195)
Log of Market Cap	-0.0133*** (0.0050)	-0.3025*** (0.0511)
Constant	-0.2997	4.74e-06
R^2	0.7206	0.4319
Observations	2,000	1,950

Note: (***) denotes statistical significance at the 99% confidence level, (**) denotes statistical significance at 95% confidence level, (*) denotes statistical significance at 90% confidence level

As discussed in the Theory and Methodology section, the interpretation of this regression is slightly different from the simple OLS model. The sign on the coefficient for *abs_diff_1* flipped from positive to negative while remaining statistically significant at the 99% confidence level, as expected if mean reverting behavior exists. It is estimated that a 1% increase in the distance of a company's current P/E from the company's historical P/E over the previous period is estimated to decrease the difference over the current period by 0.1270%. Therefore, the difference between the current P/E ratio and the trend will approach 0 over time, which is consistent with mean reverting behavior. Again, this model estimates how a change in the distance of the current P/E ratio from the trend over the period $t-2$ through $t-1$ affects the change in the distance over the period $t-1$ through t. It is not surprising to see that the level of R^2 dropped rather significantly between the simple and robust regressions from 0.7206 to 0.4319. This is likely due to detrending the second equation, since the time trend probably accounted for a large portion of the R^2 in the simple model.

Similar results were yielded from the pooled regression.

Variable	**Simple OLS** *abs_diff*	**Detrended and Differenced (Robust SE)** *abs_diff*
First Lag of Absolute Difference (*abs_diff_1*)	0.7391*** (0.0157)	-0.2748*** (0.0644)
P/E Ratio (*pe_ratio*)	0.0015*** (0.0002)	0.0092*** (0.0032)
P/E Ratio^2 (*pe_ratio2*)	-1.42e-06*** (3.48e-07)	-0.00002*** (7.18e-06)
Sector P/E Ratio Delta (*sector_delta)*	-0.0001 (0.00008)	-0.00005 (0.00008)
EPS Growth Trailing 1 Year (*eps_growth*)	-3.38e-08 (4.33e-06)	1.13e-06 (2.52e-06)
Log of Volume (*lvolume*)	0.0271*** (0.0043)	0.0346 (0.0223)
Log of Market Cap (*lmarket_cap*)	-0.0254*** (0.0044)	-0.3025*** (0.0739)
Constant	-0.4157	0.0006
R^2	0.7206	0.4319
Observations	2,000	1,550

Note: (***) denotes statistical significance at the 99% confidence level, (**) denotes statistical significance at 95% confidence level, (*) denotes statistical significance at 90% confidence level

As in the time series panel regression, the coefficient for *abs_diff_1* is negative while remaining statistically significant at the 99% confidence level. The magnitude of the effect is actually larger in the pooled regression, as a 1% increase in the distance of a company's current P/E from the company's historical P/E in the previous period decreases the difference in the current period by 0.2748%, as opposed to 0.1270% in the time series. This may be the most interesting result because I would have expected the effect to be lower in the pooled regression due to the lower variance of the data. About 25% of the observations needed to be dropped in the first differenced model since the companies observed changes every year and it would not make sense to difference the data from two different companies. Further, previous research suggests that mean reversion takes several years (Balvers et al, 2000 and Poterba and Summers, 1987). If that is the case, it would be unlikely to observe mean reversion within the year-long sample period collected for each company. Thus, it is possible that the results yielded from this paper contradict those of previous studies that support mean reversion. It may also be that P/E mean reversion happens faster than price mean reversion, which was the primary topic of study in previous works. Additionally, it may be that companies that released shares to the public after 2008 display more mean reverting tendencies. Only companies that were in the S&P500 from 2008 to 2017 were used in the first set of regressions, while new companies were selected every year for the second.

Another difference between the simple and robust regressions to note is the increased magnitude of the coefficient on *lmarket_cap*. After correcting for the biases mentioned above, the magnitude coefficient increases 25 and 15-fold for the time series and pooled regressions, respectively. A 1% increase in the market cap of a company is estimated to decrease the distance of the company's current P/E from the company's historical P/E in the current period by 0.3025% in both the time series and pooled regressions. This result is consistent with financial theory since larger companies are often highly covered by investment analysts, so there are higher quantities of analysis on the company, so its price should act more efficiently and not deviate as far from fundamental value.

A final detail to point out is the constant terms from both regressions. In the simple OLS regression for both time series panels and pooled panels, the constant term was quite far from 0. Holding all the employed variables constant, I would expect the difference between a company's current P/E and its historical P/E to be relatively close to 0, assuming markets are mostly efficient. The constant term yielded by the robust regression for both the time series panels and pooled panels were consistent with this hypothesis, further indicating that the biases discussed earlier were corrected for in the robust regression.

One issue that was not addressed but could be corrected for with further research is small sample bias. While the results of the differenced and detrended model imply that P/E ratios exhibit mean reverting behavior, small sample bias may be affecting the coefficients since only 10 years of data on 50 companies were collected. Richardson and Stock (1989) and Richardson (1993) report that correcting for small-sample bias problems may reverse the Fama and French (1988a) and Poterba and Summers (1988) results. Both Fama and French's and Poterba and Summers' results provided the foundation for price mean-reversion investment strategies when they were published. If correcting for small sample bias reverses their results, it is possible the same can happen to my results. However, I will point out that both studies employed variance ratio tests for mean reversion, which were not used in this paper.

VI. Conclusion

In this paper, an attempt was made to empirically determine whether price-earnings ratios exhibit mean reverting behavior. The research conducted falls in the section of economic literature on the Efficient Market Hypothesis; specifically, it aims to test the alternative hypothesis of mean reverting processes in price-earnings ratios of a stock against that of a random walk process. Previous literature on the topic rely on variance ratio tests and conventional unit root tests using time series data to detect mean reversion. However, some economists have found these tests to have little power against the stationary alternative and panel data can be used to increase their power. This paper contributes a robust linear model using panel data from US equites to achieve the a more accurate test for mean reversion. Further, the paper directly addresses and corrects for heteroskedasticity, serial correlation and unit roots which attempting to minimize survivorship bias.

The results provide evidence of mean reverting processes in the price-earnings ratios of US equities and appear to be robust to presence of survivorship bias. Previous works that found evidence of stock price mean reversion state that mean reversion typically takes between three to three and a half years to occur (Balvers et al, 2000). The output of this empirical study, however, suggest price-earnings ratio mean reversion may occur much faster. Moreover, endogeneity in the Fama and French Three-Factor-Model may be an important consequence of this paper, but further research should aim to test this hypothesis directly. In addition, future studies should attempt to correct for small-sample bias and increase the sample period to acquire more consistent and unbiased estimators.

Appendix

Table A – Variable Descriptions

Variable Name	Description
pe_ratio	Company's P/E ratio in the current period
pe_ratio2	$(pe\_ratio)^2$
hist_pe	Company's trailing 5 year average P/E ratio
ldiff_hist	ln(*pe_ratio*) – ln(*hist_pe*). Shows how far company's current P/E ratio is away from its 5 year average P/E ratio as a percentage.
abs_diff	Absolute value of *ldiff_hist*
abs_diff_1	1 period lag of *abs_diff*
sector_delta	Change in average sector PE ratio from the last period to the current period
eps_growth	Trailing 1 year earnings-per-share growth
lmarket_cap	Log of the company's market capitalization in the current period
lvolume	Log of the number of company shares traded in the current period

Table B – Summary Statistics from Time Series Panel Data

Variable	Observations	Mean	Std. Dev.	Min.	Max.
abs_diff	2,000	0.385	0.443	0.00014	3.540
pe_ratio	2,000	22.996	53.134	1.73	781.60
pe_ratio2	2,000	3,350.62	38,984.87	3.01	610,904.8
hist_pe	2,000	19.748	17.431	6.13	172.07
sector_delta	2,000	.1922	115.01	-1,526.18	1,524.38
eps_growth	2,000	9.686	805.19	-31,900	5,100
volume	2,000	4.19e+08	5.78e+08	4,124,512	6.40e+09
lvolume	2,000	19.306	1.028	15.232	22.579
market_cap	2,000	51.539	79.737	0.608	729.29
lmarket_cap	2,000	3.225451	1.145	-0.498	6.592

Table C – Summary Statistics from Pooled Panel Data

Variable	Observations	Mean	Std. Dev.	Min.	Max.
abs_diff	2,000	0.2216	0.3066	0.00018	2.657
pe_ratio	2,000	24.46	44.227	0.942	759.44
pe_ratio2	2,000	2,553.4	26,750.64	0.888	576,749.1
hist_pe	2,000	23.69	49.705	2.983	1016.435
sector_delta	2,000	0.503	53.821	-598.87	815.71
eps_growth	2,000	9.178	996.76	-39,770.69	9,300
volume	2,000	4.10e+08	1.26e+09	20,260	2.99e+10
lvolume	2,000	19.086	1.116	9.916	24.294
market_cap	2,000	34.575	59.858	0.028	729.29
lmarket_cap	2,000	2.837	1.141	-3.309	6.592

Table D – Ramsey RESET Test from Time Series Panels

Models	F-Statistic	p-value
Without P/E Ratio^2	114.13	0.0000***
With P/E Ratio^2	2.08	0.1028
R^2	0.7146	0.7198

Note: (***) denotes statistical significance at the 99% confidence level

Table E – Breusch-Pagan Test for Heteroskedasticity

Time Series Panel		**Pooled Panel**	
chi2(1)	*p-value*	*chi2(1)*	*p-value*
1931.21	0.0000***	2413.97	0.0000***

Note: (***) denotes statistical significance at the 99% confidence level

Table F – Serial Correlation (First Lag)

Variable	Time Series Panel: Coefficient on first lag	Time Series Panel: Adjusted R^2	Pooled Panel: Coefficient on first lag	Pooled Panel: Adjusted R^2
abs_diff	0.8300***	0.6864	0.7952***	0.5899
pe_ratio	0.8862***	0.7813	0.2885***	0.0791
pe_ratio2	0.8828***	0.7791	0.3403***	0.1142
sector_delta	-0.5221***	0.2577	-0.0756***	0.0044
eps_growth	0.0964***	0.0088	0.1793***	0.0048
lvolume	0.9739***	0.9412	0.0719***	0.0044
lmarket_cap	0.9918***	0.9812	0.1573***	0.0237

Note: (***) denotes statistical significance at the 99% confidence level

Table G – Serial Correlation (Lagged Residuals)

Variable	Time Series Panel: Coefficient on lagged residual from time trend regression	Time Series Panel: Adjusted R^2	Pooled Panel: Coefficient on lagged residual from time trend regression	Pooled Panel: Adjusted R^2
abs_diff	0.8291***	0.6847	0.1529***	0.0208
pe_ratio	0.8845***	0.7778	0.2741***	0.0710
pe_ratio2	0.8822***	0.7774	0.3355***	0.1111
sector_delta	-0.5225***	0.2581	-0.0757***	0.0044
eps_growth	0.0940***	0.0083	0.1786***	0.0048
lvolume	0.9703***	0.9380	0.0583**	0.0026
lmarket_cap	0.9881***	0.9801	0.0832***	0.0061

Note: (***) denotes statistical significance at the 99% confidence level, (**) denotes statistical significance at 95% confidence level

Table H – Levin-Lin-Chu Test for Panel Unit Roots

Variable	Time Series Panel (*p-value*)	Pooled Panel (*p-value*)
abs_diff	0.1769	0.7422
pe_ratio	0.5973	0.9997
pe_ratio2	0.0009***	0.9353
sector_delta	0.0001***	0.0000***
eps_growth	0.9984	0.9993
lvolume	0.9025	0.9932
lmarket_cap	0.8749	0.9671

Note: (***) denotes statistical significance at the 99% confidence level

Table I – Regression Output from Time Series Panels

Variable	Simple OLS *abs_diff*	Detrended and Differenced (Robust SE) *abs_diff*
First Lag of Absolute Difference (*abs_diff_1*)	0.7381*** (0.1333)	-0.1270*** (0.0339)
P/E Ratio (*pe_ratio*)	0.0037*** (0.0004)	0.0143*** (0.0014)
P/E Ratio^2 (*pe_ratio2*)	-3.00e-06*** (4.87e-07)	-0.00001*** (1.70e-06)
Sector P/E Ratio Delta (*sector_delta)*	0.0001** (0.00005)	0.00005* (0.00003)
EPS Growth Trailing 1 Year (*eps_growth)*	-9.99e-07 (6.63e-06)	2.54e-08 (1.67e-06)
Log of Volume (*lvolume)*	0.0191*** (0.0050)	0.0595*** (0.0195)
Log of Market Cap (*lmarket_cap)*	-0.0133*** (0.0050)	-0.3025*** (0.0511)
Constant	-0.2997	4.74e-06
R^2	0.7206	0.4319
Observations	2,000	1,950

Note: (***) denotes statistical significance at the 99% confidence level, (**) denotes statistical significance at 95% confidence level, (*) denotes statistical significance at 90% confidence level

Table J – Regression Output from Pooled Panels

Variable	Simple OLS *abs_diff*	Detrended and Differenced (Robust SE) *abs_diff*
First Lag of Absolute Difference (*abs_diff_1*)	0.7391*** (0.0157)	-0.2748*** (0.0644)
P/E Ratio (*pe_ratio*)	0.0015*** (0.0002)	0.0092*** (0.0032)
P/E Ratio^2 (*pe_ratio2*)	-1.42e-06*** (3.48e-07)	-0.00002*** (7.18e-06)
Sector P/E Ratio Delta (*sector_delta)*	-0.0001 (0.00008)	-0.00005 (0.00008)
EPS Growth Trailing 1 Year (*eps_growth*)	-3.38e-08 (4.33e-06)	1.13e-06 (2.52e-06)
Log of Volume (*lvolume*)	0.0271*** (0.0043)	0.0346 (0.0223)
Log of Market Cap (*lmarket_cap*)	-0.0254*** (0.0044)	-0.3025*** (0.0739)
Constant	-0.4157	0.0006

R^2	0.7206	0.4319
Observations	2,000	1,550

Note: (***) denotes statistical significance at the 99% confidence level, (**) denotes statistical significance at 95% confidence level, (*) denotes statistical significance at 90% confidence level

References

Balvers, Ronald, et al. "Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies." The Journal of Finance, vol. 55, no. 2, 2000, pp. 745–772.

Basu, S. "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis." The Journal of Finance, vol. 32, no. 3, 1977, p. 663.

Cunado, J., et al. "Mean Reversion in Stock Market Prices: New Evidence Based on Bull and Bear Markets." Research in International Business and Finance, vol. 24, no. 2, 2010, pp. 113–122.

Fama, Eugene F., and Kenneth R. French. "Permanent and Temporary Components of Stock Prices." Journal of Political Economy, vol. 96, no. 2, 1988, pp. 246–273.

Fama, Eugene F., and Kenneth R. French. "Dividend Yields and Expected Stock Returns." Journal of Financial Economics, vol. 22, no. 1, 1988, pp. 3–25.

Fama. Eugene F. and Kenneth R. French. 1992a. The cross-section of expected stock returns. Journal of Finance 47. 427 – 465.

Kim, Myung Jig, Nelson, Charles, and Starz, Richard. "Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence." The Review of Economic Studies, vol. 58, no. 3, 1991, p. 515.

Mukherji, Sandip. “Are Stock Returns Still Mean-Reverting?” Review of Financial Economics, vol. 20, no. 1, 2011, pp. 22–27.

Poterba, James, and Lawrence Summers. “Mean Reversion in Stock Prices: Evidence and Implications.” Massachusetts Institute of Technology, 1987.

Richardson, Matthew. “Temporary Components of Stock Prices: A Skeptic's View.” Journal of Business & Economic Statistics, vol. 11, no. 2, 1993, p. 199.

Richardson, Matthew, and James H. Stock. “Drawing Inferences from Statistics Based on Multiyear Asset Returns.” Journal of Financial Economics, vol. 25, no. 2, 1989, pp. 323–348.

Taylor, Mark P., and David A. Peel. “Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals.” Journal of International Money and Finance, vol. 19, no. 1, 2000, pp. 33–53.

Zhu, Zhen. “The Random Walk of Stock Prices: Evidence from a Panel of G-7 Countries.” Applied Economics Letters, vol. 5, no. 7, 1998, pp. 411–413.