ACADEMY of ECONOMICS and FINANCE JOURNAL

AEF

VOLUME 14 2023

editors

NICHOLAS MANGEE

MICHAEL TOMA

ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 14 2023

Aims and Scope of the *Academy of Economics and Finance Journal*:

The *Academy of Economics and Finance Journal (AEFJ)* is the refereed publication of selected research presented at the annual meetings of the Academy of Economics and Finance. All economics and finance research presented at the meetings is automatically eligible for submission to and consideration by the journal. The journal and the AEF thrive at the intersection of economics and finance, but special topics unique to each field are suitable for presentation and later publication. The *AEFJ* is published annually. The submission of work to the journal will imply that it contains original work, unpublished elsewhere, that is not under review, nor a later candidate for review, at another journal. The submitted paper will become the property of the journal once accepted, and not until the paper has been rejected by the journal will it be available to the authors for submission to another journal.

Volume 14:

Volume 14 of the *AEFJ* is sponsored by the Academy of Economics and Finance and Georgia Southern University's departments of Economics and Finance.

ACADEMY of
ECONOMICS
and FINANCE
JOURNAL

Volume 14 **Table of Contents** 2023

The Effect of US Treasury Auctions on Returns: The 1990s Experience

James J. Forest, SUNY New Paltz
Scott P. Mackey, Roger Williams University

Abstract

This study fills a decade-long gap in Treasury auction research, focusing on the impact of auction surprises on returns. Using a comprehensive dataset, bid-to-cover ratios, noncompetitive bids, and volumes are examined, while controlling for macro announcements. Findings reveal a positive relationship between auction surprises and returns, but with inconsistencies in the significance of auction statistics across maturities, a departure from previous studies. Auctions that were removed from the auction cycle behaved atypically. Lower returns were found for the 1-year bill when auction size changes. This research contributes to the literature by highlighting the influence of structural changes during this falling-deficit regime.

JEL Codes: E44, E52
Keywords: US Treasury auctions, GARCH, Federal Reserve, macroeconomic announcements

Introduction

The US Treasury market, as the broadest and most liquid financial market in the world, plays a critical role in the global economy. Its operations, particularly Treasury auctions, have far-reaching implications on financial markets. However, the existing literature has largely overlooked the impact of these auctions during periods of falling deficits and reduced auction volumes. This study aims to fill this significant gap in the literature by examining the effect of Treasury auction announcements on interest rates during the 1990s, a period characterized by economic expansion, declining federal deficits, and significant technological advances.

This decade stands in stark contrast to the preceding and following periods. The early 1980s saw the Federal Reserve drive up interest rates to rein in inflation amidst elevated levels of spending and reduced tax rates. The 21st Century, on the other hand, has been marked by increased government borrowing to fund various crises and wars. Spurred by technological advances which brought increased business productivity, the 1990s was a time of economic expansion and declining federal deficits. It was also the period of Federal Reserve Chair Alan Greenspan where Fed and Treasury policies were heavily informed by macroeconomic announcements, as shown in Forest and Mackey (2023). During this period, Greenspan (2007) suggests that both entities worked cooperatively and engineered a robust expansion while keeping inflation well contained.

By focusing on the 1990s, a period of falling deficits, this study offers a unique perspective on the impact of Treasury auctions on returns, contributing to a more comprehensive understanding of the dynamics of the US Treasury market. The fewer auctions, fewer maturities, and more time between auctions made it difficult to anticipate the level of demand from bidders. Missing from the existing literature is a conditional model of return behavior relative to auction demand and size during this era. The findings of this research have significant implications for traders, researchers, and policy makers, offering valuable insights into the forces that drive the financial market.

For traders, the results show the existence of a highly significant risk premium on T-bills when auction volumes were changed. This stands in contrast to existing studies that suggest rates tend to be lower on auction days. For researchers, results suggest that modeling macroeconomic announcements is essential in reducing omitted-variable bias. For policymakers, it suggests that money was left on the table, and that rates would have been lower if not for less predictable borrowing levels.

The research findings reveal intriguing dynamics in the US Treasury market. A positive relationship between auction surprises and returns is found, indicating that unexpected changes in auction outcomes can influence market performance. However, the statistical significance of these auction statistics varied across different maturities, suggesting a complex interplay between auction outcomes and market responses. Interestingly, the findings diverge from previous studies that assessed the impact on TIPS and futures markets. In the analysis, lower returns were observed for the 1-year bill when auction size changes, highlighting the sensitivity of short-term securities to auction dynamics.

The paper begins with a review of Treasury market concepts and a survey of the relevant literature, followed by a detailed discussion of the methodology used in the study. The subsequent sections present the findings of the research, a discussion of these findings in the context of the existing literature, and the conclusions drawn from the study.

Treasury Market Background

Within this market, an active over-the-counter secondary market exists with most trading volume occurring between a group of primary dealers, which numbered approximately 40 in the 1990s, but have since shrunk to just 24.[i] By 1997, an average of $125 billion worth of US Treasury securities traded daily in a market that functions virtually around the clock – about 1.5% of year-end gross domestic product (GDP). Currently, this amount has grown to $715 billion – or 2.7% of GDP.

In addition to its tremendous size and depth, the US Treasury market plays a vital role in the financial system by establishing benchmark risk-free rates for each maturity. Derivative products exist on these issues and variable-rate instruments reset based on Treasury yields. Treasury securities and inflation adjusted Treasuries, known as TIPS, are also extremely important in giving economists and market participants a real-time means for calculating breakeven inflation rates. Additionally, the bills, notes and bonds traded are widely accepted as “risk-free” assets as the US Government has never defaulted on its debt – a legacy dating back to Treasury Secretary Alexander Hamilton’s post-Revolutionary War debt repayment policy. The leading role of this market within the financial system illustrates the importance of understanding potential sources of disruption. Further, the central nature of this market to the global financial system suggests volatility in this market can potentially be transmitted to other sectors of the financial market and the world economy.

Literature Review

This section provides a brief review of the literature regarding auction and other announcement effects in the US Treasury and related markets. The Treasury market offers an opportunity to examine how an increase in the supply and demand for government securities affects the prevailing interest rate – i.e., the government’s cost of borrowing. Two notable papers, Schirm, Sheehan and Ferri (1989) and Wachtel and Young (1987), focus on the effect of debt and deficit announcements on interest rates. The former finds issuance of irregular cash management bills were disruptive to the market, while the latter found deficit announcements to impact markets negatively. Wachtel and Young (1990) find a small but significant response to post-auction demand results but no response to pre-auction announcements of auction volume, indicating that markets had priced in the additional supply between deficit announcements and the actual auctions. This is consistent with the efficient market hypothesis and shows that traders adjusted expectations of the greater borrowing needs ahead of issuance.

Bahamin, Cebula et al. (2012) look at bid dispersion in auctions from 1998 to 2010, finding it positively related to bid-to-cover ratio but negatively associated with both the percentage of accepted competitive bids and of noncompetitive bids. Lou, Yan and Zhang (2013) explore the pre- and post-auction price behavior over a 28-year period. They demonstrate a general increase in secondary market yields prior to Treasury auctions, followed by a subsequent decline. They estimate that this phenomenon results in a 9 to 18 basis point issuance cost to the Treasury. Smales (2021) investigates the behavior of futures markets for Treasury securities and finds that mutual funds, not dealers, drive the auction cycle yield premium.

An important aspect of this study is controlling for other regularly scheduled sources of variation that are known to market participants *a priori,* specifically macroeconomic announcements. For example, a number of studies have examined their effect on rates, including: Cornell (1983), Jones, Lamont and Lumsdaine (1998), and Fleming and Remolona (1999). Additionally, Kuttner (2002) examines the effect of FOMC policy changes on interest rates while Bernanke and Kuttner (2005) studied the Fed policy effect on equity markets. These papers, in general, document the existence of an announcement day effect arising from surprises in the release of macroeconomic and monetary policy information.

Other studies examining announcement effects on capital markets include, Engle and Ng (1993); Cook and Hahn (1989); Christie-David, Chaudhry and Lindley (2003); Bollerslev, Cai and Song (2000); Balduzzi, Elton and Green (2001); and Urich and Wachtel (1984). But few studies control for macroeconomic announcements while modeling auction demand expectations. Amin and Tédongap (2023), control for auction demand while modeling the TIPS auction cycle but fail to control for macroeconomic announcements. Only Wachtel and Young (1990), Smales (2021), and *this study* model *both* auction demand and macroeconomic announcements. It is important to model both because there may be a propensity for announcements to distort results when not factored into the model – i.e., omitted variables bias. This study controls for 11 widely reported macroeconomic announcements and surprises in the Federal Funds Rate to avoid this common pitfall.

A natural point of comparison exists between the Federal Reserve and the Treasury. Just as the central bank is expected to conduct open market policy without disrupting the market, the US Treasury is charged with financing its budgetary needs while not disturbing financial markets. Considering the incredible size of government borrowings, this is a significant task. According to Nandi (1997), the US government issued approximately $2 trillion in securities during 1995 alone – this was more than 25% of that year’s total US GDP. As of June of 2023, Treasury has already issued $9.9 trillion – 37% of GDP based on the most recent SIFMA data.

Methodology and Data

This section introduces the methodology employed and describes the data that was used in conducting this research. It commences with a description of the auction statistics under consideration, provides descriptive statistics and compares them to more recent auctions, and provides examples of how these demand statistics were reported in the media of that era. This shows the relatively frequent volume changes in the bill sector and infrequency of auctions in the bond sector. Expectations are modeled based on ARIMAX methods and these formulations are evaluated.

Auction Statistics of Interest

Traders often assess the level of auction demand by analyzing statistics that are made available by the US Treasury following each auction. Such information is released shortly after the auction closes through the wire services and can be found the following day in the Wall Street Journal. This release includes statistics such as the auction yield, bid-to-cover ratio (aka coverage ratio), and noncompetitive bids (aka noncomps). The latter two measures offer market participants insight into the level of demand during the auction process and tend to be the most widely reported and followed of the statistical release.

Table 1. Auction Supply Descriptives
Auction Increases vs. Decreases

	Decreased	Increased	Unchanged	Total
1-Year Bill	35	40	7	82
(%)	42.68%	48.78%	8.54%	100%
5-Year Note	9	17	74	100
(%)	9.00%	17.00%	74.00%	100%
30-Year Bond	4	9	17	30
(%)	13.33%	30.00%	56.67%	100%

Note: Sample for Notes and Bond: 1/2/1990 to 12/31/1999, Sample for 1-Year Bill: 9/24/1993 to 2/31/1999.

This information is relevant to traders in the secondary market, especially in cases when a surprise in auction demand is conveyed. During this era, the financial press often related post-auction performance to signals provided in the auction results. For example, on November 5, 1998, Gregory Zuckerman of the Wall Street Journal reported the following:

> *"The tone in the market was badly hurt by an auction of $12 billion of 10-year that proved 'just terrible' in the words of a trader. The bid-to-cover ratio, or ratio of bids to available securities, was just 1.52, well below the average of 2.3 from the past dozen auctions and the lowest in 20 years, according to Goldman Sachs." (Wall Street Journal, 1998)*

The author clearly suggested that market participants benchmark auction statistics based on the trend they have observed for recent auctions at a given maturity. Likewise, market analysts often view noncompetitive bidding as an indication of demand for the new issue. Journalist Sonoko Setaishi of the Wall Street Journal also quoted a bond trader's post-auction assessment:

> *"'Strong 'noncomps' offset the bid-to-cover ratio,'"*

This is another example of how practitioners adapt to information from the auction bidding process. It is indicative of how the market also uses noncompetitive bidding as a measure of auction demand. Further, it shows that the surprise in one of the post-auction statistics (in this case: bid-to-cover ratio) can potentially be offset by another statistic (noncompetitive bids), or vice versa – *suggesting that both should be modeled*. But how exactly are these statistics defined?

The bid-to-cover ratio is defined as total auction bids divided by the accepted bids. This is the most popular auction demand statistic by the financial wire services and convention suggests that higher ratios indicate stronger demand. Noncompetitive bids are typically made by individual investors or small banks as opposed to the primary dealers that actively compete in the auctions. An elevated level of noncompetitive bids indicates strength in underlying retail demand, which suggests that dealers will have an easier time re-selling the supply purchased at the auction.

To convey basic information about the auction process, descriptive statistics for auction results are provided in Table 2, below. It shows that the average bid-to-cover ratio decreases moving from the bill sector, where auctions are 3.20 times "oversubscribed" on average, to the 30-year bond, averaging only 2.27. The 5-year note shows a coverage ratio of 2.64. Furthermore, the standard deviation of this statistic also decreases with term to maturity.

It is important to compare these results with those from the post-2000 sample in Smales (2021), who also showed that the ratio decreases further out on the yield curve. However, both auction size and bid-to-cover ratios were noticeably different during the 1990s period. For example, the ratio for the 5-year note decreased by 4.4% but has grown by 6.2% for the 30-year bond since the sample. With respect to auction size, the 5-year note average auction size grew 141.4%, while the 30-year bond size grew only 18.9%. Thus, the increased supply in the note has altered demand for that maturity, while the increased relative scarcity of the bond appears to have increased its demand.

Unlike the case of macroeconomic announcements, where sources like Bloomberg and Investing.com publish surveys of market consensus, *no source* provides market expectations estimates for auction demand statistics. Market participants must rely on past auctions as a benchmark for auction demand or else for alternative metrics for projecting auction outcomes. Using time-series forecasts of post-auction statistics, however, one can quantify auction expectations based on the information available to traders prior to the auction and thereby evaluate the impact of a surprise auction outcome on interest-rate levels and volatility.

Table 2. Descriptive Statistics – Auction Results

	Issue	Mean	Max	Min	Std Dev	Obs
Auction Size	30-Year	10.43	12.00	8.25	0.85	30
($ Billion)	5-Year	11.10	16.00	3.00	1.73	100
	1-Year Bill	14.82	19.44	10.00	3.58	82
Bid-to-Cover	30-Year Bond	2.27	2.82	1.48	0.35	30
	5-Year Note	2.64	3.76	1.74	0.44	100
	1-Year Bill	3.20	6.44	2.08	0.80	82
Noncomps	30-Year Bond	320.07	937.00	47.00	168.81	30
($ Million)	5-Year Note	569.40	1,172.00	169.00	216.07	100
	1-Year Bill	897.18	1,643.90	347.00	231.36	82

Note: Noncomps indicate noncompetitive bids. Note and Bond sample: 1/2/1990 to 12/31/1999, 1-Year Bill sample: 9/23/1993 012/31/1999.

Time series models for the auction variables were created using standard ARIMAX methods with exogenous regressors. Note, the goal here is not to create a model that captures the most variation in the dependent variable. In fact, overfitting would undermine the forthcoming analysis. Rather, a reasonable proxy for the market expectation for the auction result is desired. The structure of these models is summarized in Table 3. The final parsimonious model structure was determined based on Akaike information criteria.

Table 3. Time Series Models for Auction Statistics

	Model	Regs	Adj.		Prob	
Bid-to-Cover	*ARIMAX(p,d,q,x)*	*X*	*R^2*	*F-stat*	*(F-stat.)*	*MAPE*
1-Year Bill	(2,0,0,1)	A	0.461	18.267	0.000	13.7
5-Year Note	(1,0,1,1)	A	0.251	12.044	0.000	12.1
30-Year Bond	(0,1,1,0)	N/A	0.352	8.884	0.001	13.0
Noncomps						
1-Year Bill	(1,0,0,0)	N/A	0.679	86.782	0.000	10.9
5-Year Note	(1,0,0,1)	A	0.620	54.911	0.000	20.7
30-Year Bond	(1,1,0,0)	N/A	0.127	3.100	0.061	34.3

Note: ARIMAX variables: p = order of Autoregressive term, d = number of differencing to achieve time series stationarity, q = order of Moving Average term, X = exogenous regressors. X: A = Auction Volume, N/A = No exogenous regressors. DW = Durbin Watson statistic for autocorrelation in residuals. MAPE = Mean Absolute Percentage Error. Note and Bond Sample: 1/2/1990 to 12/31/1999. T-Bill Sample: 9/23/1993 to 12/31/1999.

The models have autoregressive order p, order of integration d, moving-average order q, and x exogenous predictors. Exogenous predictors include the previously announced auction volume. Most models have a single AR parameter, but both the 30-year bond coverage ratio and noncomps have an I(1) structure. This nonstationarity in demand matters when forming expectations for future auctions, as it is evidence that the distribution of auction demand changed over this period. The consistent percentage error in the coverage ratio compared to an increasing MAPE with maturity for noncomps is noted.

Combining the forecasts for auction variables with forecasts of macroeconomic variables disentangles the effects of contemporaneous announcement effects and thereby obtains a clearer picture of the pressure that the auctions exert on the market. As a result, the conditional impact of auctions on the market is better assessed to gauge the relative importance of auction announcements relative to macroeconomic announcements and discern which auction statistics hold the most weight with market participants.

Survey data are from Standard & Poor's MMS and have been widely used in the existing literature as the basis for estimating standardized surprises in macroeconomic data. The on-the-run (OTR) and off-the-run (FTR) US Treasury return data are extracted from the CRSP Daily Treasury database. Treasury auction results were compiled from the Treasury Direct website and checked against Bloomberg and the Wall Street Journal. Three maturities across the yield curve are examined to enable additional analysis from the perspective of three distinct segments – bills, notes, and bonds. This study proceeds with an analysis of the results for the mean equations for OTR and FTR returns, then continues with additional Wald tests associated with the variance equations. The empirical analysis concludes with an evaluation of the importance of controlling for macroeconomic announcements and other controls when modeling the auction cycle.

GARCH-X Model Specification

This section presents results from GARCH-X(1,1) models of Treasury returns on auction statistics, macroeconomic announcements, Fed policy and dummy variables for quiet days – those days when there are no auctions, Fed announcements, nor any of the eleven macroeconomic announcements. The model is presented below, followed by a table explaining variables and coefficients. The model takes the following form:

$$r_t = \mu + \sum_{i=1}^{4} \theta_i x_{i,t} + \sum_{j=1}^{11} \lambda_j z_{j,t} + \phi \mathcal{FFS}_t + \gamma QD_t + \epsilon_t \tag{1}$$
$$h_t = \omega + a\varepsilon_{t-1}^2 + \beta h_{t-1} + \delta_1 AUCDUM + \partial_2 FOMC + \partial_3 QD_t$$

Variables:

r_t	= Daily total return on US Treasury Security at time t
μ	= Constant (intercept term)
$x_{i,t}$	= 4 auction variables (Bid-to-Cover Std. Surprise, Noncomps Std. Surprise, Decrease Dummy, Increase Dummy)
$z_{j,t}$	= 11 macroeconomic standardized surprise variables (Listed in regression results)
$\mathcal{FFS}_t$	= Surprise in Federal Funds Rate based on Kuttner (JME, 2001), not standardized
QD_t	= Binary equal to 1 on "quiet days" (no macro announcements nor maturity-specific auction), 0 otherwise
ϵ_t	~ $IID(0, \sigma)$ error term following a Student's t-distribution with τ degrees of freedom estimated
h_t	= Conditional variance
ε_{t-1}^2	= ARCH term (lagged squared error term)
h_{t-1}	= GARCH term (lagged conditional variance)
$AUCDUM$	= Maturity-specific dummy variable
$FOMC$	= Binary equal to 1 on FOMC meetings or conference call days, 0 otherwise

Here, r_t is one-day total return on the Treasury security at time t. Four auction variables, $x_{i,t}$, are standardized surprise variables for the bid-to-cover ratio, noncompetitive bids, and two (1,0) dummy variable series indicating announcements of increased or decreased volume, respectively. $z_{i,t}$ is the standardized surprise in economic indicator i at time t, $\mathcal{FFS}_t$ is the surprise in the Federal Funds Rate in basis points and ε_t is the residual at time t. Standardized surprises in macroeconomic indicators are calculated by subtracting the expected value of the economic variable from the as-reported result from the official release and dividing by the sample standard deviation. Standardization allows us to easily assess the return associated with a one standard deviation surprise in an auction or macroeconomic variable.

Empirical Results

The results from GARCH estimation of OTR securities are provided in Table 4 and show positive mean equation coefficients on the bid-to-cover ratio for all three maturities; The 5-year note coefficient, however, is the only one that achieves acceptable statistical significance. The magnitude of the bid-to-cover surprise on the 5-year note is greater, in absolute terms, than that of the unemployment rate, core-PPI, durable goods orders and retail sales. The signs of the coefficients are consistent with the prior expectation that a larger-than-expected coverage ratio indicates strong demand. However, the benchmark 30-year bond is insignificant with a p-value of 0.12. Given the relative infrequency of bond auctions, the findings warrant *cautious* attention because the statistical power of the parameter estimates should be considered – e.g., there were only 30 auctions at the longest maturity.

Importantly, the lack of significance stands in contrast to the Treasury futures market findings of Smales (2021), which suggest small but highly significant bid-to-cover effects after the turn of the century. Coverage ratio *surprises* are not controlled for but controls are used for the *level* and for macro surprises. Importantly, there were cases of nonstationarity in the ratio, suggesting that the distribution changed over time. Amin and Tédongap (2023) evaluate TIPS cash market findings with additional bidding data for primary dealers versus direct and indirect bidders. They find coefficients are small and insignificant for the former, but large and highly significant for the latter. The data are from 2005-2019. They show a large and significant relationship between the ratio for both direct and indirect bidders across three maturities. This consistency stands in stark contrast to the results of this study.

Table 4. GARCH(1,1) Regressions on Daily On-the-Run Returns
Treasury Rates and Auction Results 1990 - 1999
GARCH(1,1) estimates based on Student's t-distribution

	1-Year Bill		5-Year Note		30-Year Bond	
Auction Announcement	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
θ_1 Bid-to-Cover Surprise	0.003	0.864	0.542	0.000**	0.615	0.118
θ_2 Noncomps Surprise	-0.001	0.966	0.106	0.474	0.742	0.075*
θ_3 Decrease Dummy	-0.063	0.006**	0.141	0.646	-0.792	0.497
θ_4 Increase Dummy	-0.044	0.036**	0.013	0.969	0.183	0.800
Economic Indicators	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
λ_1 Capacity	-0.075	0.000**	-0.296	0.007**	-0.457	0.010**
λ_2 Confidence	-0.051	0.001**	-0.543	0.000**	-0.760	0.000**
λ_3 CPI (Core)	-0.096	0.000**	-0.614	0.000**	-0.842	0.000**
λ_4 Durable Goods	-0.044	0.055*	-0.439	0.000**	-0.749	0.000**
λ_5 ECI	-0.079	0.000**	-1.127	0.000**	-1.717	0.000**
λ_6 Hourly Earnings	-0.083	0.000**	-0.679	0.000**	-0.835	0.000**
λ_7 NAPM	-0.083	0.000**	-0.753	0.000**	-1.037	0.000**
λ_8 Nonfarm Payrolls	-0.135	0.000**	-0.727	0.000**	-1.155	0.000**
λ_9 PPI (Core)	-0.010	0.498	-0.385	0.001**	-0.735	0.000**
λ_{10} Retail Sales	-0.057	0.002**	-0.297	0.007**	-0.440	0.012**
λ_{11} Unemployment	0.056	0.001**	0.173	0.087*	0.114	0.426
ϕ Fed Funds Rate	-0.020	0.000**	-0.059	0.000**	-0.081	0.000**
γ Quiet Day	0.021	0.004**	-0.104	0.055*	-0.237	0.005**
μ Constant	0.067	0.000**	0.171	0.000**	0.242	0.000**
ω C	0.017	0.000**	0.097	0.013**	0.048	0.414
α RESID(-1)^2	0.167	0.000**	0.036	0.000**	0.029	0.000**
β GARCH(-1)	0.401	0.000**	0.939	0.000**	0.956	0.000**
δ_1 AUCDUM	-0.007	0.201	-0.254	0.164	0.751	0.173
δ_2 Fed Meeting	0.011	0.269	-0.069	0.667	-0.318	0.290
$\delta 3$ Quiet Day	-0.009	0.000**	-0.074	0.223	0.073	0.510
τ T-Dist DOF	4.095	0.000**	6.507	0.000**	8.159	0.000**
Durbin Watson	2.32		1.86		1.94	
Adjusted R-squared	0.08		0.08		0.07	
Log likelihood	717		-4362.9		-5464.4	

Note: Capacity = Capacity Utilization, Confidence = Consumer Confidence, CPI(Core) = Core Consumer Price Index, Durable Goods = Durable Goods orders, ECI = Employment Cost Index, Hourly Earnings = average Hourly Earnings, NAPM = National Association of Purchasing Managers diffusion index, Nonfarm Payrolls = Nonfarm Payrolls, PPI (Core) = Core Producer Price Index, Retail Sales = Retail Sales, Unemployment = Unemployment rate. Sample for Notes and Bond: 1/2/1990 to 12/31/1999, Sample for 1-Year Bill: 9/24/1993 to 12/31/1999. ** Indicates significant at 5%, * Indicates significant at 10% level.

With respect to surprises in noncompetitive bidding, the 30-year bond coefficient has the expected sign and is statistically significant at the 10% level. The magnitude of the coefficient is greater than important macroeconomic indicators: capacity utilization, retail sales and unemployment. It is also a larger effect than that of the coverage ratio. Noncomps have fallen out of favor in the research since the 1980s, but results suggest that they may deserve additional attention.

Three key economic indicators appeared to be much more important to the market: the employment cost index, nonfarm payrolls, and the diffusion index produced by the National Association of Purchasing Managers (now known as the ISM Manufacturing Index). These three reports were known favorites of then-Federal Reserve Chair Alan Greenspan who was a macroeconomic forecaster prior to his tenure at the Fed. Additionally, surprises in Fed Funds Rate policy had a highly significant negative effect on returns across all three maturities in this era when the Fed held their interest rate decisions close to the vest. Smales (2021) shows that the Fed no longer shocks this market, as surprises in Fed Funds Rate changes only affect the 1-year bill significantly in a 21st Century sample. This is a function of modern transparency in the Fed's forward guidance policy since Greenspan's tenure.

Turning to the coefficients on auction volume increases and decreases, first compare results to two earlier studies. Wachtel and Young (1990) found that neither auction volume levels nor surprises in auction volume had a significant effect on Treasury rates during the early 1980s. In a similar study, Wachtel and Young (1987) find that government deficit announcements significantly affect rates. This is consistent with concerns with increased government spending and lower marginal tax rates in the 1980s that sent the deficit on a steady upward trajectory. But while their earlier study showed a general sensitivity to higher-than-expected deficits, the effect appeared to be fully priced into the market by the time the Treasury announced how much they would borrow at each maturity.

This result is confirmed for the note and bond, but 1-year bill returns were significantly lower when auction size changed. The magnitude of the coefficient is larger than nonfarm payroll; note that there were frequent size changes for that issue, as only 8.5% of 82 auctions were unchanged in volume. This is likely due to the increasing scarcity of Treasury securities during the era and its effect on auction expectations. This result suggests that volume data should be incorporated into empirical auction studies. Further, it shows that maturities that are eliminated from the auction cycle can behave differently, such structural changes have implications for model specification (See, for example, Hendry, Pagan and Sargan (1984).

Table 5. GARCH(1,1) Regressions on Daily 1st Off-the-Run Returns
Treasury Rates and Auction Results 1990 - 1999
GARCH(1,1) estimates based on Student's t-distribution

	1-Year Bill		5-Year Note		30-Year Bond	
Auction Announcement	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
θ_1 Bid-to-Cover Surprise	0.009	0.588	0.300	0.000**	0.483	0.082*
θ_2 Noncomps Surprise	0.007	0.686	0.057	0.539	0.595	0.047**
θ_3 Decrease Dummy	-0.047	0.022**	0.106	0.569	-0.923	0.305
θ_4 Increase Dummy	-0.044	0.054*	0.144	0.565	-0.071	0.904
Economic Indicators	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
λ_1 Capacity	-0.067	0.000**	-0.161	0.019**	-0.417	0.028**
λ_2 Confidence	-0.040	0.001**	-0.361	0.000**	-0.742	0.000**
λ_3 CPI (Core)	-0.080	0.000**	-0.434	0.000**	-0.846	0.000**
λ_4 Durable Goods	-0.032	0.083*	-0.299	0.000**	-0.787	0.001**
λ_5 ECI	-0.112	0.000**	-0.574	0.000**	-1.389	0.000**
λ_6 Hourly Earnings	-0.080	0.000**	-0.448	0.000**	-0.876	0.000**
λ_7 NAPM	-0.062	0.000**	-0.504	0.000**	-1.026	0.000**
λ_8 Nonfarm Payrolls	-0.114	0.000**	-0.614	0.000**	-1.198	0.000**
λ_9 PPI (Core)	-0.014	0.333	-0.217	0.002**	-0.651	0.000**
λ_{10} Retail Sales	-0.055	0.001**	-0.181	0.005**	-0.415	0.020**
λ_{11} Unemployment	0.032	0.021**	0.153	0.015*	0.390	0.013**
ϕ Fed Funds Rate	-0.018	0.000**	-0.046	0.000**	-0.035	0.003**
γ Quiet Day	0.017	0.008**	-0.065	0.043**	-0.185	0.028**
μ Constant	0.071	0.000**	0.140	0.000**	0.180	0.001**
ω C	0.001	0.039**	0.045	0.003**	3.993	0.001**
α RESID(-1)^2	0.055	0.000**	0.042	0.000**	0.025	0.054*
β GARCH(-1)	0.919	0.000**	0.932	0.000**	0.588	0.000**
δ_1 AUCDUM	0.000	0.943	-0.081	0.227	-2.561	0.159
δ_2 Fed Meeting	0.000	0.916	-0.047	0.422	-5.502	0.000**
$\delta 3$ Quiet Day	-0.001	0.200	-0.046	0.040**	-1.428	0.009**
τ T-Dist DOF	5.023	0.000**	6.507	0.000**	3.259	0.000**
Durbin Watson	2.07		1.87		1.96	
Adjusted R-squared	0.10		0.11		0.07	
Log likelihood	920.7		-3085.5		-5503.6	

Note: Capacity = Capacity Utilization, Confidence = Consumer Confidence, CPI (Core) = Core Consumer Price Index, Durable Goods = Durable Goods orders, ECI = Employment Cost Index, Hourly Earnings = average Hourly Earnings, NAPM = National Association of Purchasing Managers diffusion index, Nonfarm Payrolls = Nonfarm Payrolls, PPI (Core) = Core Producer Price Index, Retail Sales = Retail Sales, Unemployment = Unemployment rate. Sample for Notes and Bond: 1/2/1990 to 12/31/1999, Sample for 1-Year Bill: 9/24/1993 to 12/31/1999. ** Indicates significant at 5%, * Indicates significant at 10% level.

Table 5 provides results for the FTR Treasury securities, which have not been investigated in related studies. Interestingly, results suggest that auction effects are not exclusive to the OTR issues. While most US Treasury market trading volume is in the OTR securities, FTR security return regressions produce similar but noticeably smaller coefficients for the note and bond auction demand statistics. Interestingly, the improved statistical significance is seen for the bid-to-cover ratio on the FTR 30-year bond, with a p-value of 0.082. The bid-to-cover coefficient for the FTR 5-year note is nearly half that of its OTR counterpart, but still highly significant. This decrease in magnitude is not particularly surprising as auction demand is expected to matter most to dealers who maintain inventories of the in-demand OTR securities and adjust these inventories over the auction cycle. As bonds go further off the run, they tend to be held increasingly by buy-and-hold investors. The FTR macro coefficients are measurably less negative for 10 of the 11 macroeconomic indicators. The largest reduction in sensitivity was seen with the employment cost index.

The results are again mixed but show that both auction and macroeconomic announcement surprises could affect the limited market for more-seasoned US Treasury issues, but that those securities were less sensitive to shocks than their OTR counterparts. Thus, it would take an extremely severe auction surprise to disrupt the off-the-run segment of the market. Inconsistency in 1-year bill regressions stand out compared to that of the 5- and 30-year securities. Note that the data series starts in late 1993. Further, the bill auctions were cut in size from about $20 billion per auction to just $10 billion at the end of the decade and were phased out during the early 2000s. Treasury resumed issuance in 2009.

Announcement Effects on Volatility

This section examines results from the variance equation in the GARCH-X models, focusing on the effect that auctions exert on return volatility. Variance effects can also be compared to existing literature, as they were also examined with respect to Treasury auction announcements in the futures markets by Smales (2021). The equation in this study bears similarity to that of Jones, Lamont and Lumsdaine (1998), in that dummy variables enter into the variance equation specification.

Table 6. Tests of Coefficient Equality

	1-Year Bill			5-Year Note			30-Year Bond		
Panel A. On-the-Run GARCH Equations									
Test 1. Null Hypothesis: C(22)=C(23)	Value	df	Prob.	Value	df	Prob.	Value	df	Prob.
F-statistic	**2.67**	**(1,1544)**	**0.10**	0.61	(1,2477)	0.44	**2.71**	**(1,2477)**	**0.10**
Panel B. 1st Off-the-Run GARCH Equations									
Test 1. Null Hypothesis: C(22)=C(23)	Value	df	Prob.	Value	df	Prob.	Value	df	Prob.
F-statistic	0.02	(1,1525)	0.90	0.16	(1,2477)	0.69	2.40	(1,2477)	0.12

Note: Note and Bond Sample: 1/2/1990 to 12/31/1999. Bill Sample: 9/23/1993 to 12/31/1999. C(22) = Coefficient on Auction Day, C(23) = Coefficient on FOMC Call of Meeting Day, C(24) = Coefficient on Quiet Day. **Bold** indicates significance at 10%.

From prior Tables 4 and 5, the variance equation coefficients on the auction dummy variables are all insignificant. This indicates that auctions did not amplify volatility in the US Treasury market in the 1990s, as most coefficients were negative. Only the 30-year bond suggested increased volatility, with a sizeable parameter estimate of 0.751, but with an insignificant p-value of 0.17. Comparatively, Smales (2021) estimated a much-smaller parameter for the futures contract at 0.054, with a p-value of 0.08. Thus, the 21st Century futures market estimate is significant but small in terms of magnitude (See Table 4 in Smales (2021).

Another common feature between the two studies is the auction dummy corresponding to the 5-year note. Smales (2021) estimates a sensitivity of 0.054 and a p-value of 0.02, while this study estimated a value of -0.081 with an insignificant p-value of 0.227. In other words, volatility on 5-year notes appeared lower on auction days during the 1990s but was positive (and significant) for futures on this maturity in a more recent sample. This implies that auctions matter for futures but, in the environment of the 1990s cash market, volatility was not increased by the frequent 5-year auctions.

This research also shows that quiet days tended to have a negative effect on volatility, as expected. The OTR results are only significant for the 1-year bill, where the coefficient is negative but highly significant and relatively small. FTR quiet day volatility effects are significant for 5- and 30-year securities. Note that both the 1-year bill and 30-year bond were eliminated from the auction cycle for a period in the early 2000s because of federal budget surpluses during the early part of the decade.

Table 6 presents results of Wald tests of coefficient equality between auction day and Fed policy announcement day dummy variables in the variance equations – thereby providing a natural contrast between fiscal versus monetary perspectives. Panel A presents the OTR results, while panel B shows FTR results. The latter case fails to reject equality of the coefficients. But, importantly, OTR tests rejected equality for the 1-year and 30-year issues and are significant at 10%. At these opposing ends of the term structure, the Treasury and Fed impacted Treasury returns asymmetrically. The two coefficients had opposite signs in Table 4, with a positive (negative) coefficient on auctions (Fed policy) at the long end of the curve and a negative (positive) effect on the short end.

Discussion

Until now, the literature on auction announcements has suggested that auction demand, as measured by the bid-to-cover ratio, is a statistically significant factor in auction day returns across the yield curve. However, event studies in the Treasury auction demand literature are all for periods of rising budget deficits and increased borrowing. This study fills a large gap in the empirical literature by evaluating an economic era of falling deficits and decreased borrowing for three different maturities and for both on- and off-the-run securities. Results are compared with prior findings before and after the sample.

Overall, the OTR coverage ratio results are mixed. Some consistency exists with Wachtel and Young (1990), Smales (2021) and Amin and Tédongap (2023) in that bid-to-cover increases have a positive effect on Treasury returns, but acceptable significance is lacking for the 30-year bond and the 1-year bill and is highly insignificant. This is a departure from earlier studies which suggest consistently high significance across maturities. Sharp decreases in offering in both maturities appear to have complicated the expectations forming process for these maturities. For the larger and more frequent 1-year bill, these were most meaningful, as the structural change in borrowings induced a risk premium on days when auction volume was changed. This suggests that Treasury incurred additional costs associated with an uncertainty premium. Thus, Treasury officials should carefully consider changes in the auction cycle as they can result in potentially undesirable results.

The lack of significance in the case of noncompetitive bids on bills and notes may be an indicator that this measure fell out of favor with market participants since the 1980s when Wachtel and Young (1990) performed their study. The measure is investigated by neither Smales (2021) nor Amin and Tédongap (2023), on more recent samples. Additionally, the 1990s was a declining deficit period, as opposed to the late skyrocketing deficit decade of the 1980s, that was marked with increased cold war military spending. As a result, noncompetitive bidding data may have been less of a factor. A modern analysis of this measure may be worthy of attention, given the trajectory of the federal budget.

Conclusion

This study has provided an analysis of the impact of US Treasury auction announcements on interest rates during the 1990s. The findings reveal a complex interplay between auction outcomes and market responses, with a positive relationship between auction surprises and returns. However, the statistical significance of these auction statistics varied measurably across different maturities, indicating the less consistent influence of Treasury auctions on 1990s US Treasury market returns.

Interestingly, the findings diverge from recent studies that assessed the impact on TIPS and futures markets. In the analysis, lower returns were observed for the 1-year bill when auction size changes, highlighting the sensitivity of short-term securities to auction dynamics. These findings underscore the nuanced influence of Treasury auctions on market returns and provide a fresh perspective on the functioning of the US Treasury market.

Furthermore, results suggest that the US Treasury's financing operations were conducted in a manner that exerted no more pressure on the market than most macroeconomic announcements. This is a significant finding, as it provides a clearer picture of the pressure that the auctions exert on the market, gauges the relative importance of auction announcements relative to macroeconomic announcements, and discerns which auction statistics held the most weight with market participants.

However, the relative size of coefficients on auction demand surprises versus macroeconomic announcements should not be interpreted as suggesting that adverse auction outcomes do not matter. Results suggest that a large negative surprise at auction that occurs contemporaneously with negative macroeconomic surprises would likely be troublesome for traders, particularly at the 30-year maturity. During the 1990s, bond auctions were sometimes held on days when CPI and employment report data were released.

This study fills a significant gap in the literature by examining a unique period of falling deficits, less-frequent auctions, and reduced auction volumes. The insights gained from this research have implications for traders, policy makers, and researchers, offering valuable insights into the forces that drive the Treasury market. Future research could further explore the dynamics of the US Treasury market while including periods of economic instability and significant policy changes. Studies should, when possible, include 1990s era auctions to account for alternative borrowing regimes. Further, by conditioning on important macroeconomic and monetary policy announcements, researchers will be able to assess the relative importance of announcements on the market.

Acknowledgements

The authors would like to thank Ben Branch, Steven Dennis, John Harding, Hossein Kazemi, Kenneth Kuttner, Nelson Lacey, Saira Latif, Michael Melton, Robert Mooradian, Sanjay Nawalkha, Ralph Norton, and Kim Rupert for helpful comments and suggestions. Seminar participants at the UMass – Amherst; Roger Williams; Central Connecticut State; UConn and attendees of FMA and Academy of Economics and Finance conferences are thanked for useful suggestions; and to Nam Tran and Svetlana Doronkina for research assistance. Any remaining errors are the sole responsibility of the authors. The authors thank Kenneth Kuttner for making his Federal Reserve expectations data available.

References

Amin S, Tédongap R (2023) The changing landscape of treasury auctions. Journal of Banking and Finance 148: 106714 https://doi.org/10.1016/j.jbankfin.2022.106714

Bahamin P, Cebula RJ, et al. (2012) The demand for treasury securities at auction. Academy of Economics and Finance Journal 3:23-32 https://economics-finance.org/aefj/issues/AEFJ_volume3.pdf

Balduzzi P, Elton E, Green TC (2001) Economic news and bond prices: Evidence from the U.S. Treasury market. Journal of Financial and Quantitative Analysis 36(4):523-543 https://doi.org/10.2307/2676223

Bernanke BS, Kuttner KN (2005) What explains the stock market's reaction to federal reserve policy? Journal of Finance 60(5): 1221-1257

Bollerslev T, Cai J, Song F (2000) Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the U.S. Treasury bond market. Journal of Empirical Finance 7:37-55

Christie-David R, Chaudhry M, Lindley JT (2003) The effects of unanticipated macroeconomic news on debt markets. Journal of Financial Research XXVI(3):319-339

Cook T, Hahn T (1989) The effect of changes in the federal funds rate target on market interest rates in the 1970s. Journal of Monetary Economics 24:331-351

Cornell B (1983) The money supply announcements puzzle: Review and interpretations. American Economic Review 73:644-657

Engle RF, Ng VK (1993) Measuring and testing the impact of news on volatility. Journal of Finance Volume 48(5):1749-1778

Fleming M, Remolona E (1999) Price formation and liquidity in the U.S. Treasury market: The response to public information. Journal of Finance LIV(5):1901-1915 https://www.jstor.org/stable/222509

Forest JJ, Mackey S (2023) The effect of macroeconomic announcements on credit markets: An autometric general-to-specific analysis of the Greenspan era

Greenspan A (2007) The age of turbulence, Penguin Books

Hendry DF, Pagan A, Sargan JD (1984) Dynamic specification. Z. Griliches and M. D. Intriligator

Jones CM, Lamont O, Lumsdaine RL (1998) Macroeconomic news and bond market volatility. Journal of Financial Economics 47(3):315-337 https://doi.org/10.1016/s0304-405x(97)00047-0

Kuttner KN (2002) Monetary policy surprises and interest rates. Federal Reserve Bank of New York

Lou D, Yan H, Zhang J (2013) Anticipated and repeated shocks in liquid markets. Review of Financial Studies 26(8):1891-1912 10.1093/rfs/hht034

Nandi S (1997) Treasury auctions: What do the recent models and results tell us? Economic Review - Federal Reserve Bank of Atlanta 82(4):4-15

Schirm DC, Sheehan RG, Ferri MG (1989) Financial market responses to treasury debt announcements. Journal of Money, Credit and Banking 21(3):394-400

Smales LA (2021) Macroeconomic news and treasury futures return volatility: Do treasury auctions matter? Global Finance Journal 48: 100537 https://doi.org/10.1016/j.gfj.2020.100537

Urich T, Wachtel P (1984) The effects of inflation and money supply announcements on interest rates. Journal of Finance 39(4):1177-1188

Wachtel P, Young J (1987) Deficit announcements in interest rates. American Economic Review 77(5):1007-1012

Wachtel P, Young J (1990) The impact of treasury auction announcements on interest rates. Quarterly Review of Economics and Business 30(3):63-72

Wall Street Journal (1998) Prices of Treasury Bonds are Sent Tumbling by Stocks' Strength, Fed Worries, Weak Auction" (Zuckerman G) Nov. 5, 1998

Identifying Weak-Form Market Inefficiencies using the Hurst Exponent

R. Stafford Johnson, Xavier University

Abstract

In this paper, the Hurst exponent is calculated for the S&P 500 for 12 annual periods from 2010 to 2022 to identify price patterns along with the corresponding period returns from technical strategies. The ex-post simulations show abnormal returns from some moving-average and moving-average band strategies during strong trending periods as detected by the Hurst coefficient, but not in periods when the Hurst exponent indicates random patterns. The simulations provide evidence of weak-form market inefficiency, as well as an argument for the use of the Hurst coefficient as a metric for identifying non-random, trending periods where some technical strategies are profitable.

JEL: G14, G17
Keywords: Hurst exponent; Technical strategies, Market efficiency

Introduction

Many weak-form tests of the efficient market hypothesis have examined whether abnormal returns can be earned from technical trading strategies. The statistical tests used in a number of studies include serial correlation, run, and filter-rule tests. The findings reached from these weakly efficient market tests are mixed. Some studies find that the returns on securities are serially uncorrelated and that the runs observed on stocks are not significantly different than the runs obtained from a random number generator. Other studies suggest that certain trading rules based on moving averages and moving-average band strategies yield above average profits (Fama and Blume, 1966; Sweeney, 1988; Brock, Labonishok, and LaBaron, 1991; and Bessembinder and Chan, 1998). A number of these technical strategies are based on momentum or mean-reversion. In momentum-based strategies, investors try to capitalize on the continuance of the existing market trend, while mean-reversion strategies are based on detecting whether stock returns and volatility revert to their long-term average over time. The Hurst exponent is a time series measure named after the British hydrologist Howard Hurst (1880–1978). The exponent was originally used in hydraulic engineering to study the volatility patterns of rain observed over a long period of time. In finance, the Hurst exponent can be used to identify price patterns hidden within seemingly random stock price trends.

This paper examines the weak-form efficient market hypothesis by examining (1) whether trading rules based on moving-average and moving-average-band strategies earned risk-adjusted returns significantly greater than a naive buy-and-hold strategy in periods when price trends are present as measured by the size of the Hurst exponent, and (2) whether a naive buy-and-hold strategy outperforms moving-average and moving-average-band strategies in periods when stock price trends are random as measured by the Hurst exponent. In the next section, the Hurst exponent is defined. This is followed with an analysis of the relationship between the Hurst exponent, price trends, and strategies based on moving-averages and moving-average bands. In the last section, a summary is presented of the findings from back tests conducted to determine trending and random periods for the S&P 500 index using the Hurst coefficient and the associated return performances of buy-and-hold, moving-average, and moving-average-band strategies. The ex-post simulations show abnormal returns from some technical moving-average and moving-average band strategies during strong trending periods as detected by the Hurst coefficient, but not in periods when the Hurst exponent indicates random patterns. These trends, in turn, provide evidence of weak-form market inefficiency, as well as an argument for the use of the Hurst coefficient as a metric for identifying non-random, trending periods where some technical strategies are profitable.

Hurst Exponent

The Hurst exponent is a nonparametric statistic. The exponent is calculated by using a rescaled range (R/S) analysis in which the data is transformed into a number of segments and then examined by looking at the logarithmic range and scale of each segment relative to the number of segments. Mathematically the Hurst exponent compares the diffusion of a time series to that of a geometric Brownian motion. For descriptions of how it is computed, see May (1999), Kaabar (2019), and Singh, Divakar, and Garg (2018).

In general, a time series can be persistent with a tendency to continue its up or down pattern, anti-persistent in which it has a higher tendency to reverse its current pattern, or random. In general, the Hurst exponent detects long-term memory that is a bias in the time series referred to as fractional Brownian motion. The bias relates to the autocorrelations of the time series and the rate at which these decrease as the lag between pairs of values increases. If the Hurst exponent, H, decreases towards zero, the price series may be more mean-reverting and if it increases more towards one, the price series may be more trending. If the series are random and cannot be forecasted (zero autocorrelation), then it follows a random walk and H is near 0.5:

- Geometric random walk (H = 0.5)
- Mean-reverting series (H < 0.5)
- Trending Series (H > 0.5)

The Hurst exponent can be calculated on a Bloomberg terminal for most time series. It can also be calculated from the online PRACMA package (Practical Numerical Math Functions). Bloomberg's Hurst calculation is found on its GPO KAOS screen and is based on the work of Christopher May (1999) who applied the Hurst exponent to nonlinear price patterns. The KAOS screen calculates H for lookback periods ranging from 12 to 250 periods. As noted, if a price trend is random, the Hurst coefficient continuously has a value close to 0.5. If not, then there is pattern to the stock price movement. For the period from 6/18/1990 to 11/8/2023, the Hurst exponent's daily value for the S&P 500 averaged 0.6439, suggesting patterns to overall market trends and some periods in which technical strategies would be profitable. For example, the Hurst exponent's average daily value for the two-year period from 1/28/18 to 1/31/20 was H(100) = 0.6725 for a 100-day lookback and H(50) = 0.6629 for a 50-day lookback. These coefficients suggest technical strategies would generate returns better than a naive buy-and-hold strategy during that period. In contrast, for the one-year period from 1/31/13 to 1/31/14 the Hurst exponent's average daily value for the S&P 500 averaged H(100) = 0.5615 for a 100-day lookback and H(50) = 0.4955 for a 50-day lookback. These coefficients suggest technical strategies would not generate returns any better than a naive buy-and-hold strategy during that period (see Exhibit 2).

Exhibit 1. Hurst Exponent, 6/18/1990-11/8/2023: 100-day lookback period; Average = 0.6439

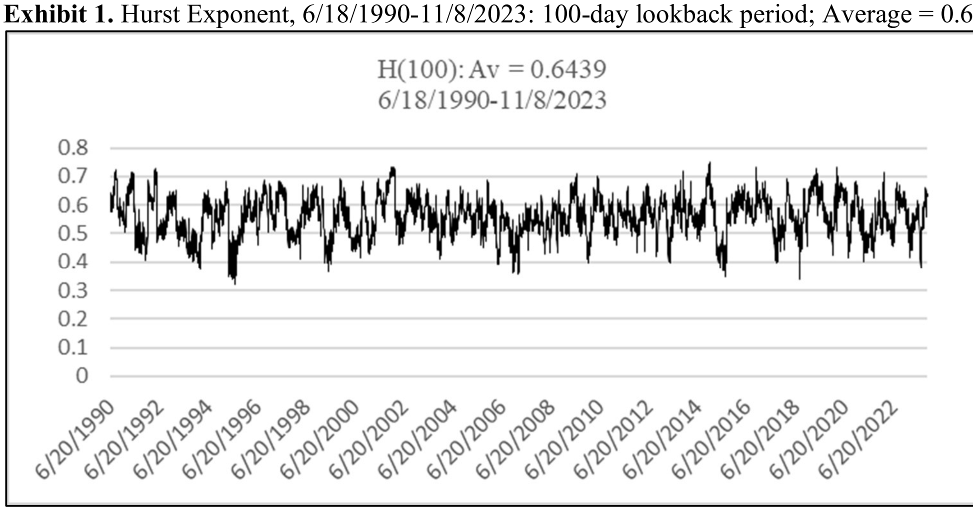
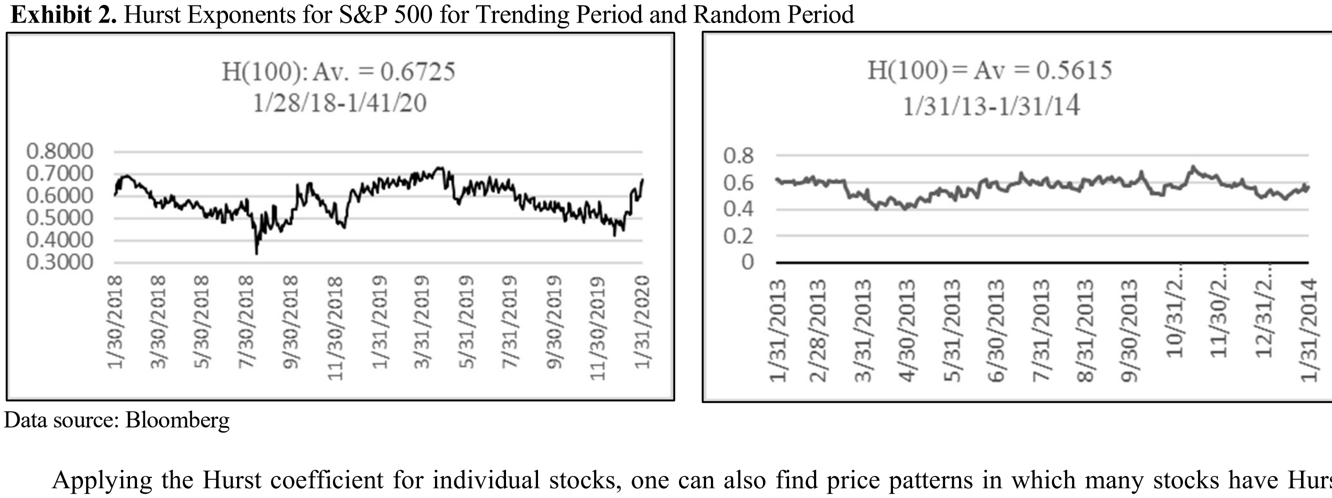


Exhibit 2. Hurst Exponents for S&P 500 for Trending Period and Random Period



Data source: Bloomberg

Applying the Hurst coefficient for individual stocks, one can also find price patterns in which many stocks have Hurst exponent values higher or lower than 0.5. On 5/31/2023, for example, 18 of the 30 Dow stocks had Hurst coefficients for a 50-day lookback period that deviated plus or minus by at least 20% from 0.5: H > 0.6 or H < 0.4 (see Exhibit 3).

Exhibit 3. Dow Stocks, Hurst exponent, 5/31/2023, 50-day lookback period

Company	Hurst	Company	Hurst
PROCTER & GAMBLE CO/THE	0.72	WALGREENS BOOTS ALLIANCE INC	0.68
NIKE INC -CL B	0.66	3M CO	0.50
MERCK & CO. INC.	0.70	CHEVRON CORP	0.65
COCA-COLA CO/THE	0.76	UNITEDHEALTH GROUP INC	0.63
AMGEN INC	0.74	HONEYWELL INTERNATIONAL INC	0.43
JP MORGAN CHASE & CO	0.42	VERIZON COMMUNICATIONS INC	0.62
CISCO SYSTEMS INC	0.70	AMERICAN EXPRESS CO	0.48
SALESFORCE INC	0.62	INTL BUSINESS MACHINES CORP	0.81
WALMART INC	0.65	CATERPILLAR INC	0.46
WALT DISNEY CO/THE	0.52	MICROSOFT CORP	0.50
JOHNSON & JOHNSON	0.69	HOME DEPOT INC	0.41
TRAVELERS COS INC/THE	0.48	BOEING CO/THE	0.50
INTEL CORP	0.59	VISA INC-CLASS A SHARES	0.53
APPLE INC	0.39	MCDONALD'S CORP	0.75
GOLDMAN SACHS GROUP INC	0.64	DOW INC	0.68

Data source: Bloomberg

Technical strategies and weak-form efficient market tests

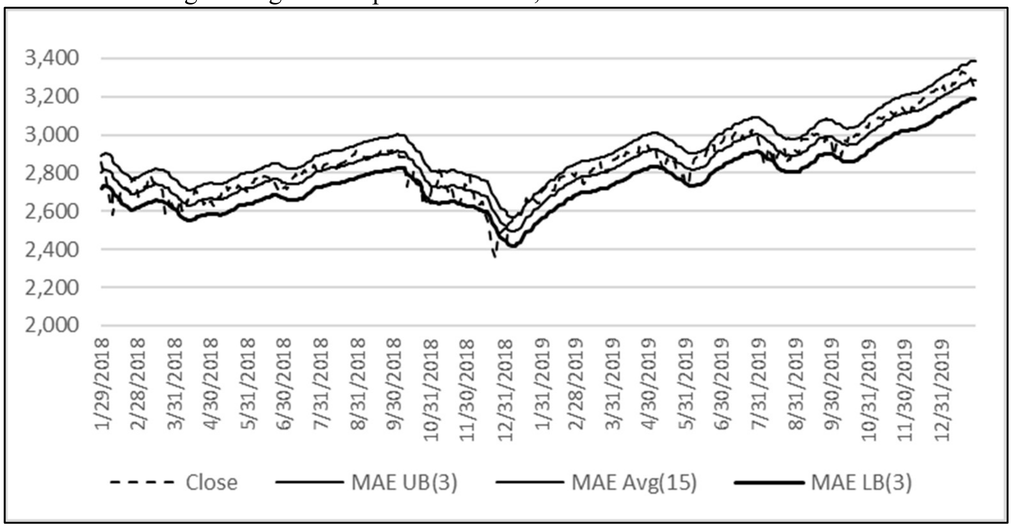
Many technical strategies are based on stock price movements relative to their moving averages. Moving averages are used to determine the overall price trend. If the overall trend is decreasing, then the moving average line is above the price line, and when the overall price trend is increasing, the moving average line is below the price line. Technicians identify price trend reversals whenever a price line breaks its moving average line. Specifically, a signal to a technician of a reversal of a declining trend would be when prices start increasing such that the price line breaks through the moving average line from below. A technician might see this as a strong signal if it breaks the moving average line from below on heavy volume. A signal of a reversal of a rising trend would be when prices start decreasing such that the price line breaks through the moving average line from above. Again, if the breakthrough is accompanied with heavy volume, then the reverse signal may be considered a strong indicator. There are several types of moving averages: simple, weighted, triangular, and exponential. A weighted-moving average weights newer observations more than older. For example, a 30-day weighted moving average would weigh the most recent date by 30, the day before by 29, and so. A triangular moving average weighs the data in the middle dates more. When used as signals for price trend changes, a weighted moving average generates the first signal, followed by the simple moving average, and then the triangular average. Other types of moving averages are exponential averages and variable moving averages that incorporate volatility. Moving averages also vary by periodicity (e.g., 52-week average and a 30-day average).

Stocks also tend to trade in channels or bands defined by upper and lower resistance and support lines. Bull, bear, and sideway trends are often described by resistance and support levels. A resistance level is the ceiling above which the price is not expected to rise. When a price rises to its resistance level, an increase in selling and an excess supply is expected, causing a price reversal. A support level is a floor beneath which the price is not expected to fall. When the price falls to its support level, an increase in buying and an excess demand is expected, causing the price to reverse. Price trends are characterized by prices rising until they meet their resistance level and falling back until they meet their support level. The resistance and support line of a channel can be thought of as points where the stock is either overbought or oversold. If a stock price breaks a support or resistance level, it is a signal that something very significant is occurring and a new trend is developing. When a strong buying surge pushes the stock's price past its resistance level, it is considered a breakout, with the stock expected to rise to a new higher resistance level and a new support level that is often the old resistance level. In contrast, when a strong selling surge pushes the stock's price below its support level, there is a breakout with the stock expected to fall until it reaches a new support level and new resistance level. Technicians, in turn, try to define trends in terms of resistance and support levels to help them identify breakouts. Often, technicians look at volume information to confirm a breakout.

In practice, many technicians create support and resistance bands around moving averages. Moving average bands, also called moving average envelopes, can be constructed from different moving average intervals, from weighted and triangular moving averages, and with different percentage increases and decreases from the averages. Exhibit 4 shows a moving-average envelope graph for the Hurst-determined trending period identified previously from 1/28/2018 to 1/28/20, with the moving average calculated on 15-day intervals and with the upward and lower bounds generated from plus and minus three standard deviations from the mean. Trading points for a moving-average-envelope (MAE) reversal strategy consist of going long when the closing price hits the lower band of the MAE and short when the close hits the upper band. For this Hurst-identified trending period, the MAE strategy generated a profit of $55.45 million from a $100 million starting investment for a rate of return of 55.45%.

The profits were based on six long positions and five short positions. The position's Sharpe ratio was 2.06. In contrast, a naive buy-and-hold strategy generated a return for the period of only 15.66% with a Sharpe ratio of 0.60.

Exhibit 4. Moving-Average Envelope for S&P 500, 1/28/2018 to 1/28/2020



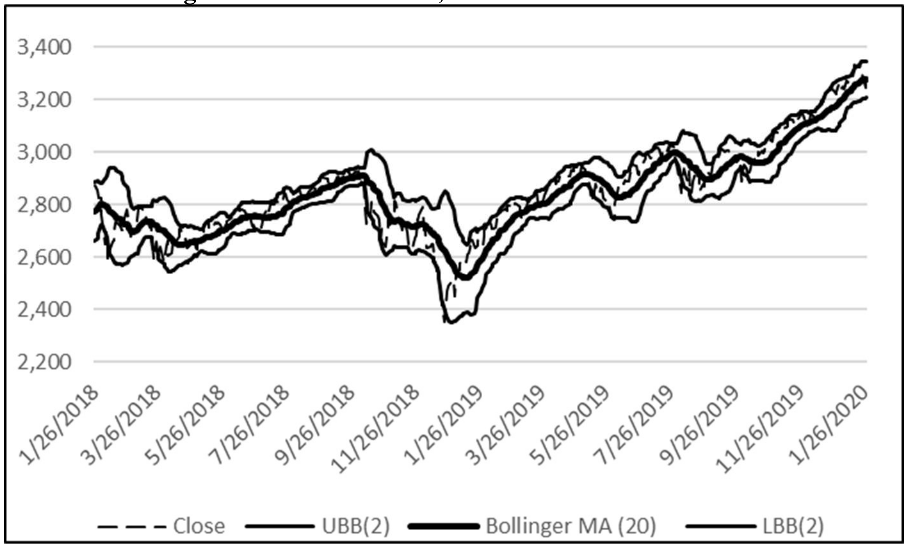
Trade #	Position	Entry Date	Entry Price	Exit Date	Exit Price	Size	Trade Profit	Cumulative Profit
1	Long	02/06/18	2,615	02/27/18	2,780	38,244	6,335,883	6,335,883
2	Short	02/27/18	2,780	03/23/18	2,647	38,244	5,114,753	11,450,636
3	Long	03/23/18	2,647	11/08/18	2,806	42,109	6,723,544	18,174,180
4	Short	11/08/18	2,806	11/21/18	2,658	42,109	6,259,082	24,433,262
5	Long	11/21/18	2,658	12/04/18	2,782	46,819	5,837,861	30,271,123
6	Short	12/04/18	2,782	12/17/18	2,591	46,819	8,974,266	39,245,389
7	Long	12/17/18	2,591	01/09/19	2,580	53,747	-577,780	38,667,609
8	Short	01/09/19	2,580	05/14/19	2,820	53,747	-12,905,730	25,761,879
9	Long	05/14/19	2,820	06/21/19	2,953	44,594	5,912,718	31,674,597
10	Short	06/21/19	2,953	08/06/19	2,861	44,594	4,081,689	35,756,286
11	Long	08/06/19	2,861	01/28/20	3,276	47,447	19,693,352	55,449,638
Naive	Long	01/30/18	2,833	01/28/20	3,276	35,301	15,655,994	15,655,994

Data source: Bloomberg

A popular band constructed with moving averages is the Bollinger band. The band is formed by creating lines that are two standard normal deviations above and below a 20-day or 30-day moving average. The usefulness of the band is that the price of a security, in turn, should remain within the bands 95% of the time provided the underlying variability does not change significantly. Exhibit 5 shows a graph of a Bollinger band generated from Bloomberg for the trending period from 1/28/2018 to 1/28/20, with the moving average calculated on 20-day intervals and with the upward and lower bounds generated from plus and minus two standard deviations from the mean. The trading points for the Bollinger strategy consist of going long when the closing price hits the lower band of the MAE and short when the close hits the upper band. For this period, the Bollinger strategy generated a profit of $34.435 million from a $100 million investment for a rate of return of 34.4%. The profits were based on six long positions and six short positions. The position's Sharpe ratio was 2.06. As previously noted, the naive buy-and-hold strategy generated a return for the period of only 15.66% with a Sharpe ratio of 0.60.

The profitability observed from the moving average envelope and Bollinger strategies (Exhibits 4 and 5) are consistent with the sizes of the Hurst exponents for this two-year period from 1/28/18 to 1/31/20: H(100) = 0.6725 and H(50) = 0.6629 (Exhibit 2). As a metric, the Hurst coefficients in this case suggest trending patterns in which technical strategies would have been profitable, generating abnormal returns that outperformed a naive buy and hold strategy.

Exhibit 5. Bollinger Bands for S&P 500, 1/28/2018 to 1/28/2020



Trade #	Position	Entry Date	Entry Price	Exit Date	Exit Price	Size	Trade profit	Cum profit
1	Long	02/06/18	2,615	05/11/18	2,723	38,244	4,127,292	4,127,292
2	Short	05/11/18	2,723	06/28/18	2,699	38,244	918,238	5,045,531
3	Long	06/28/18	2,699	08/08/18	2,857	38,924	6,153,884	11,199,415
4	Short	08/08/18	2,857	10/11/18	2,777	38,924	3,110,806	14,310,221
5	Long	10/11/18	2,777	03/19/19	2,841	41,165	2,630,032	16,940,253
6	Short	03/19/19	2,841	05/14/19	2,820	41,165	849,646	17,789,899
7	Long	05/14/19	2,820	07/05/19	2,984	41,767	6,855,218	24,645,117
8	Short	07/05/19	2,984	08/02/19	2,944	41,767	1,685,298	26,330,415
9	Long	08/02/19	2,944	09/06/19	2,980	42,912	1,563,284	27,893,699
10	Short	09/06/19	2,980	10/02/19	2,925	42,912	2,383,762	30,277,461
11	Long	10/02/19	2,925	11/29/19	3,147	44,542	9,906,141	40,183,602
12	Short	11/29/19	3,147	01/28/20	3,276	44,542	-5,748,591	34,435,011
Naive	Long	01/30/18	2,833	01/28/20	3,276	35,301	15,655,994	15,655,994

Data source: Bloomberg

In contrast to the Hurst-identified trending period, the returns from a naive buy-and-hold and the Bollinger strategy were almost identical (see Exhibit 6) for the one-year Hurst-identified random period from 1/31/13 to 1/31/14 (H(100) = 0.5615 and H(50) = 0.4955; Exhibit 2): Bollinger: Return = 18.06%; Naive: Return = 18.99%. Moreover, the naive strategy significantly outperformed the moving-average enveloped strategy, which had a return of -0.87% (see Exhibit 7).

Evidence of profitability from technical moving-average and moving-average band strategies during Hurst-identified trending periods

Ex-post simulation tests were conducted by the author to determine:

1. If trading rules based on moving-average and moving-average-band strategies earned risk-adjusted returns significantly different than a naive buy-and-hold strategy in periods when price trends were present as measure by the Hurst exponent being greater than 0.6 or less than 0.4.
2. If a naive buy-and-hold strategy outperformed trading rules based on moving-average and moving-average-band strategies in periods when stock price trends were random as measured by a Hurst exponent near 0.5.

The simulations examined daily closing price trends for the S&P 500 index and used Bloomberg's Hurst Coefficient (GPO KAOS) and back-testing (BTST) screens to determine trending patterns and returns for different periods. Specifically:

1. A 12-year period from 1/18/10 to1/18/22 was divided into 12 annual periods: 1/18/10-11/18/11, 1/18/11-1/18/12, …, 1/18/21-1/18/22.
2. In each period, the Hurst exponent was calculated using Bloomberg for 25-day, 50-day, and 100-day lookback periods: H(25), H(50), and H(100).
3. Based on the sizes of the exponents, three trends were identified:

- Strong Trend when at least two of the three exponents deviated plus or minus 20% from H = 0.5 (H > 0.6 or H < 0.4)
- Partial Trend when at least one of the three exponents deviated plus or minus 20% from H = 0.5 (H > 0.6 or H < 0.4)
- Random when the exponent was between 0.4 and 0.6 for all three lookbacks: 0.4 < H < 0.6

4. For each period, returns from back testing were calculated for a buy-and-hold strategy of going long on the first day and closing on the last day and for the following moving -average and moving-average-band strategies:
    - A Simple Moving Average Strategy of going long when the close crossed a 30-day moving average line from below and short when the close crossed the moving average line from above.
    - Bollinger Band Strategy of going long when the close hit the lower band and short when the close hit the upper band. The Bollinger Bands were set two standard deviations from a 20-day moving average line.
    - Moving-Average Envelope strategy of going long when the close hit the lower band and short when the close hit the upper band. The default bands were set to three standard deviations from a 15-day moving average.
    - Exponential Moving Average Strategy of going long when the close crossed the exponential moving average line from below and short when the close crossed from above. The Exponential Moving Average line was based on the previous 50 days and was calculated by applying a percentage of the current closing price to the previous day's moving average value.
    - Weighted Moving Average Strategy of going long when the close crossed the Weighted Moving line from below and short when the close crossed the line from above. The Weighted Moving Average line was based on the previous 50 days and placed more weight on recent data and less on past.
    - The Triangular Moving Average Strategy of going long when the close crossed the triangular Moving Average line from below and short when the close crossed the line from above. The Triangular Moving Average line was based on the previous 50 days and placed the majority of weight on the middle portion of the price series.
5. The following statistics were calculated on the trade returns for each strategy:
    - Number of long trades
    - Number of short trades
    - Number of total trades
    - Cumulative dollar returns from a $100m investment
    - Cumulative return as percent of a $100m investment
    - Sharpe ratio: risk premium (rate of return – risk-free return) per unit of risk (standard deviation)
6. Strategies were then ranked by their Sharpe Ratios.
7. The Sharpe ratio rankings of the technical moving-average and moving-average band strategies were compared to naive buy-and-hold strategies in each period to determine if technical strategies were consistent in outperforming naive strategies in periods of strong or partial trending periods, and whether the naive strategy was consistent in outperforming the technical strategies in random periods.

Exhibit 6. Bollinger trades from 1/31/13 to 1/31/14 when H(100) = 0.5615 and H(50) = 0.4955

Trade #	Position	Entry Date	Entry Price	Exit Date	Exit Price	Size	Trade Profit	Cum. Profit
1	Short	02/20/13	1,531	02/2613	1,488	65,319	2,814,596	2,814,596
2	Long	02/2613	1,488	03/06/23	1,540	69,102	3,589,158	6,403,754
3	Short	03/06/23	1,540	06/06/23	1,609	69,102	-4,802,589	1,601,165
4	Long	06/06/23	1,609	07/12/13	1,675	63,134	4,164,950	5,766,115
5	Short	07/12/13	1,675	08/16/23	1,661	63,134	886,401	6,652,516
6	Long	08/16/23	1,661	09/12/13	1,689	64,201	1,796,986	8,449,502
7	Short	09/12/13	1,689	10/09/13	1,657	64,201	2,068,556	10,518,058
8	Long	10/09/13	1,657	10/18/13	1,737	66,698	5,317,832	15,835,890
9	Short	10/18/13	1,737	12/13/13	1,778	66,698	-2,751,959	13,083,930
10	Long	12/13/13	1,778	12/24/13	1,828	63,602	3,182,644	16,266,574
11	Short	12/24/13	1,828	01/27/14	1,791	63,602	2,352,638	18,619,212
12	Long	01/27/14	1,791	01/27/14	1,783	66,229	-558,973	18,060,240
Naïve	Long	02/01/13	1,498	01/31/14	1,783	66,750	18,989,040	18,989,040

Exhibit 7. Moving-average envelope trades from 1/31/13 to 1/31/14 when H(100) = 0.5615 and H(50) = 0.4955

Trade #	Position	Entry Date	Entry Price	Exit Date	Exit Price	Size	Trade Profit	Cum. Profit
1	Short	02/22/13	1,502	02/25/13	1,516	66,559	-877,248	-877,248
2	Long	02/25/13	1,516	02/26/13	1,488	65,401	-1,814,878	-2,692,125
3	Short	02/26/13	1,488	02/28/13	1,516	65,401	-1,840,384	-4,532,510
4	Long	02/28/13	1,516	04/04/13	1,554	62,973	2,374,082	-2,158,427
5	Short	04/04/13	1,554	04/05/13	1,560	62,973	-396,100	-2,554,528
6	Long	04/05/13	1,560	04/08/13	1,553	62,465	-419,765	-2,974,292
7	Short	04/08/13	1,553	04/09/13	1,563	62,465	-615,280	-3,589,573
8	Long	04/09/13	1,563	04/16/13	1,552	61,678	-663,039	-4,252,611
9	Short	04/16/13	1,552	04/17/13	1,575	61,678	-1,369,868	-5,622,480
10	Long	04/17/13	1,575	04/18/13	1,552	59,938	-1,351,003	-6,973,482
11	Short	04/18/13	1,552	04/24/13	1,579	59,938	-1,603,342	-8,576,824
12	Long	04/24/13	1,579	06/03/13	1,632	57,907	3,065,018	-5,511,806
13	Short	06/03/13	1,632	06/19/13	1,652	57,907	-1,165,089	-6,676,895
14	Long	06/19/13	1,652	06/20/13	1,625	56,496	-1,537,256	-8,214,151
15	Short	06/20/13	1,625	07/08/13	1,634	56,496	-541,232	-8,755,383
16	Long	07/08/13	1,634	08/12/13	1,688	55,834	3,024,528	-5,730,855
17	Short	08/12/13	1,688	08/14/13	1,694	55,834	-307,645	-6,038,500
18	Long	08/14/13	1,694	08/15/13	1,680	55,471	-791,571	-6,830,071
19	Short	08/15/13	1,680	09/10/13	1,674	55,471	293,442	-6,536,630
20	Long	09/10/13	1,674	10/01/13	1,682	55,821	451,592	-6,085,038
21	Short	10/01/13	1,682	10/02/13	1,692	55,821	-529,741	-6,614,779
22	Long	10/02/13	1,692	10/04/13	1,679	55,195	-723,606	-7,338,386
23	Short	10/04/13	1,679	10/11/13	1,691	55,195	-678,899	-8,017,284
24	Long	10/11/13	1,691	12/06/13	1,788	54,392	5,290,710	-2,726,574
25	Short	12/06/13	1,788	12/09/13	1,806	54,392	-970,897	-3,697,472
26	Long	12/09/13	1,806	12/12/13	1,782	53,317	-1,306,267	-5,003,738
27	Short	12/12/13	1,782	12/19/13	1,809	53,317	-1,455,021	-6,458,759
28	Long	12/19/13	1,809	01/14/14	1,821	51,708	639,111	-5,819,648
29	Short	01/14/14	1,821	01/15/14	1,841	51,708	-990,725	-6,810,373
30	Long	01/15/14	1,841	01/24/14	1,827	50,632	-686,570	-7,496,943
31	Short	01/24/14	1,827	01/31/14	1,783	50,632	2,246,542	-5,250,401
Naïve	Long	02/01/13	1,498	01/31/14	1,783	66,750	18,989,040	18,989,040

Data source: Bloomberg

Findings

Exhibit 8 summarizes the Hurst coefficient values and Sharpe ratios from the simulations (details of the trade simulations and cumulative returns are in the appendix). As shown in the exhibit, seven of the 12 periods were identified as strong price trends with at least two of the three Hurst exponents with H > 0.6 or H < 0.4. In six of those seven periods with strong trending periods, a number of the technical strategies outranked the naive strategy. The one exception was the 2016-2017 period when the buy-and-hold strategy outperformed the Bollinger strategy (Sharpe (Buy & Hold) = 2.244, Sharpe (Bollinger) = 1.545, Sharpe (MA Envelope) = 0.874). In one of the 12 periods, a partial trend was identified in which only one of the three exponents had H > 0.6 or H < 0.4. In that period, a technical strategy also dominated. Finally, four of the 12 periods were identified as random with 0.4 < H < 0.6. In each of those periods the naive buy-and hold strategy dominated all of the technical strategies.

These ex-post simulations thus provide evidence of weak-form market inefficiency, as well as an argument for the use of the Hurst coefficient as a metric for identifying non-random, trending periods where some technical strategies are profitable. Specifically, the findings show:

1. Some trading rules based on moving-average and moving-average-band strategies earned risk-adjusted returns greater than a naive buy-and-hold strategy in periods when price trends were present as measured by the Hurst exponent being greater than 0.6 or less than 0.4.
2. Naive buy-and-hold strategies outperformed trading rules based on moving-average and moving-average-band strategies in periods when stock price trends were random as measured by a Hurst exponent near 0.5.

Conclusion

One of the most influential theories to emerge out of the finance literature over the last 60 years is the efficient market hypothesis (EMH). Introduced by Burton Malkiel in the 1960s, the EMH precipitated a considerable amount of controversy between proponents of the EMH and practitioners who employed fundamental and technical analysis. EMH proponents argued that if the market consisted of a sufficient number of fundamentalists, then their actions would force the market price of a security to its equilibrium value. Similarly, EMH proponents argued that if the market consisted of enough technicians, then their actions would eliminate the possibility of earning any abnormal return from identifying trends in security prices. Since its introduction, the EMH has spurred an extensive amount of empirical research. The research can be divided along the lines of the weak-form, semi-strong-form, and strong-form tests. The weak-form tests of the EMH try to examine whether there are price trends in which investors can earn abnormal returns from trading strategies based on the trends. This paper contributes to that literature by identifying periods of market inefficiencies characterized as strong trending in which some technical trading rules outperform a naïve buy-and-hold strategy. Empirical research on the EMH has also provided fundamental and technical analysts with new methods and statistical tools for selecting securities and evaluating security price trends. This paper contributes to that literature by introducing the Hurst coefficient as a metric for identifying if a period is strong trending, partial trending, or random.

Exhibit 8. Hurst exponents and Sharpe rankings from technical moving-average and moving-average band strategies: 2010-2022

Period	Hurst Coefficient	Trending	Sharpe Rank (index)	Consistency
1/18/10-1/18/11	H(25) = 0.1965, H(50) = 0.6185, H(100) = 0.5092	Strong	Simple MA = 1.543, Exponential MA = 1.543, Buy & Hold = 0.997	Yes
1/18/11-1/18/12	H(25) = 0.7456, H(50) = 0.6682, H(100) = 0.5189	Strong	Moving Average Envelope = 1.540, Bollinger = 1.149, Buy & Hold = 0.196	Yes
1/18/12-1/18/13	H(25) = 05872, H(50) = 0.5934, H(100) = 0.5982	Random	Buy & Hold = 1.386, MA Envelope = 0.246, Simple MA = 0.158	Yes
1/18/13-1/18/14	H(25) = 05533, H(50) = 0.4437, H(100) = 0.5293	Random	Buy & Hold = 2.803, Bollinger = 1.824, Exponential MA = 1.127	Yes
1/18/14-1/18/15	H(25) = 06521, H(50) = 0.5595, H(100) = 0.6425	Strong	MA Envelope= 1.788, Bollinger = 1.784, Buy & Hold= 1.064	Yes
1/18/15-1/18/16	H(25) = 06133, H(50) = 0.5840, H(100) = 0.6497	Strong	MA Envelope= 0.297, Buy & Hold= -0.430, Bollinger = -0.547	Yes
1/18/16-1/18/17	H(25) = 0.7250, H(50) = 0.7073, H(100) = 0.6169	Strong	Buy & Hold = 2.244, Bollinger = 1.545, MA Envelope = 0.874	No
1/18/17-1/18/18	H(25) = 0.5301, H(50) = 0.4812, H(100) = 0.4569	Random	Sharpe Index: Buy & Hold = 4.151, Exponential MA = 1.55, Simple MA = 1.235	Yes
1/18/18-1/18/19	H(25) = 0.8210, H(50) = 0.7045, H(100) = 0.7053	Strong	MA Envelope= 2.450, Bollinger = 0.681, Weighted MA = -0.133; Buy & Hold = -0.310	Yes
1/18/19-1/18/20	H(25) = 0.5423, H(50) = 0.5024, H(100) = 0.4623	Random	Buy & Hold = 2.648, MA Envelope = 1.812, Simple MA = 1.356	Yes
1/18/20-1/18/21	H(25) = 0.4475, H(50) = 0.6375, H(100) = 0.6257	Strong	Weighted MA = 1.307, Triangular MA = 1.017, Exponential MA = 0.914, Buy & Hold = 0.721	Yes
1/18/21-1/18/22	H(25) = 0.6406, H(50) = 0.4650, H(100) = 0.5251	Partial	Bollinger = 2.566, Buy & Hold = 2.011, Exponential MA = 0.423	Yes

Trending: Strong Trend when at least two of the three exponents deviated plus or minus 20% from H =0.5 (H > 0.6 or H < 0.4); Partial Trend when at least one of the three exponents deviated plus or minus 20% from H =0.5 (H > 0.6 or H < 0.4); Random when the exponent was between 0.4 and 0.6 for all three lookbacks: 0.4 < H < 0.6

Sharpe Ratio: Risk premium (rate of return – risk-free return) per unit of risk (standard deviation)

Consistency: Yes, when Sharpe ratio rankings for some of the technical moving-average and moving-average band strategies were greater than a buy-and-hold strategy for a Hurst-identified trending periods; Yes, when Sharpe ratio for buy-and-hold strategy exceeded technical moving-average and moving-average band strategies for a Hurst-identified random periods.

References

Bessembinder HK, Chan K (1998) Market efficiency and the returns to technical analysis. Financial Management 27(2):5−17

Brock WJ, Lakonishok J, LeBaron B (1992) Simple technical trading rules and the stochastic properties of stock returns. The Journal of Finance 47:1731-1764

Fama E, Blume M (1966) Filter rules and stock market trading. Journal of Business 39:226–241

Hurst HE (1951) The long-term storage capacity of reservoirs. Transactions of the Amer. Soc. of Civil Engineers 116:770-799

Johnson S (2014) Equity Markets and Portfolio Analysis. Wiley, New York: Wiley: 537-622

Kaabar S (2019) The Hurst exponent in trading: https://www.linkedin.com/pulse/hurst-exponent-trading-sofien-kaabar
Malkiel BG (2003) The efficient market hypothesis and its critics. The Journal of Economic Perspectives 17:59–82
Mandelbrot BB (1997) The variation of certain speculative prices. New York, Springer: 371-418
Mandelbrot BB, Wallis JR (1969) Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. Water Resources Research 5:967–988
May C (1999) Nonlinear Pricing: Theory and Applications. Wiley, New York:120-121
PRACMA (nd) Practical Numerical Math Functions: https://www.rdocumentation.org/packages/pracma/versions/1.9.9
Singh V, Divakar V, Garg A (2018) The Hurst exponent: https://blog.quantinsti.com/hurst-exponent/
Sweeney RJ (1988) Some new filter rule tests: Methods and results. Journal of Financial and Quantitative Analysis 23:285−300

Appendix

Exhibit A. Hurst exponents, trades, Sharpe ratios, rankings from technical moving-average and moving-average band strategies

1/18/10-1/18/22: Partial Trending; Sharpe Rankings:
H(25) = 0.6391, H(50) = 0.4606, H(100) = 0.5220
Sharpe Index: Buy & Hold = 0.974, Bollinger = .770,
MA Envelope = 0.246

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$300,466,746	0.974
Bollinger Bands	31	31	62	$202,299,184	0.770
Simple MA	169	170	339	-$37,681,579	-0.233
Exponential MA	114	115	229	-$6,003,377	0.027
Weighted MA	134	135	269	-$29,362,381	-0.155
Triangular MA	100	101	201	-$31,994,146	-0.180
MA Envelopes	23	22	45	$32,170,332	0.246

1/18/10-1/18/11: Strong Trending; Sharpe Rankings: Consistent
H(25) = 0.1965, H(50) = 0.6185, H(100) = 0.5092
Sharpe Index: Simple MA = 1.543, Exponential MA = 1.543,
Buy & Hold = 0.997

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$12,811,415	0.997
Bollinger Bands	2	2	4	$3,987,867	0.377
Simple MA	9	9	18	$19,612,521	1.562
Exponential MA	6	6	12	$19,142,047	1.543
Weighted MA	11	11	22	$3,225,781	0.330
Triangular MA	7	7	14	$7,086,543	0.621
MA Envelopes	3	3	6	$3,536,766	0.343

1/18/11-1/18/12: Strong Trending; Sharpe Rankings: Consistent
H(25) = 0.7456, H(50) = 0.6682, H(100) = 0.5189
Sharpe Index: Moving Average Envelope = 1.540,
Bollinger = 1.149, Buy & Hold = 0.196

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$1,044,393	0.196
Bollinger Bands	3	3	6	$18,170,260	1.149
Simple MA	17	17	34	-$23,084,924	-1.137
Exponential MA	10	10	20	-$1,884,981	0.005
Weighted MA	11	11	22	-$10,022,811	-0.455
Triangular MA	9	9	18	-$7,806,257	-0.330
MA Envelopes	4	4	8	$24,546,588	1.540

1/18/12-1/18/13: Random; Sharpe Rankings: Consistent
H(25) = 05872, H(50) = 0.5934, H(100) = 0.5982
Sharpe Index: Buy & Hold = 1.386, MA Envelope = 0.246,
Simple MA = 0.158

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$13,600,864	1.386
Bollinger Bands	2	3	5	-$5,549,344	-0.440
Simple MA	13	13	26	$913,014	0.158
Exponential MA	12	12	24	-$12,684,706	-1.215
Weighted MA	11	11	22	-$2,240,493	-0.165
Triangular MA	8	8	16	-$3,764,205	-0.327
MA Envelopes	2	1	3	$6,247,952	0.697

1/18/13-1/18/14: Random; Sharpe Rankings: Consistent
H(25) = 05533, H(50) = 0.4437, H(100) = 0.5293
Sharpe Index: Buy & Hold = 2.803, Bollinger = 1.824,
Exponential MA = 1.127

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$23,736,292	2.803
Bollinger Bands	5	6	11	$15,587,305	1.824
Simple MA	15	15	30	-$6,902,524	-0.732
Exponential MA	5	5	10	$6,658,397	1.127
Weighted MA	12	12	24	$902,077	0.163
Triangular MA	5	5	10	-$1,237,175	-0.105
MA Envelopes	0	1	1	-$9,756,060	-1.484

1/18/14-1/18/15: Strong Trending; Sharpe Rankings: Consistent
H(25) = 06521, H(50) = 0.5595, H(100) = 0.6425
Sharpe Index: MA Envelope= 1.788, Bollinger = 1.784,
Buy & Hold= 1.064

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$9,470,854	1.064
Bollinger Bands	3	2	5	$15,880,496	1.784
Simple MA	16	17	33	-$7,224,576	-0.685
Exponential MA	8	9	17	-$2,105,427	-0.156
Weighted MA	10	11	21	-$10,384,488	-1.013
Triangular MA	8	9	17	-$9,219,205	-0.890
MA Envelopes	2	1	3	$15,405,566	1.788

1/18/15-1/18/16: Strong Trending; Sharpe Rankings: Consistent
H(25) = 06133, H(50) = 0.5840, H(100) = 0.6497
Sharpe Index: MA Envelope= 0.297, Buy & Hold= -0.430, Bollinger = -0.547

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	-$6,923,070	-0.430
Bollinger Bands	1	0	1	-$8,038,355	-0.547
Simple MA	24	24	48	-$21,882,011	-1.626
Exponential MA	19	19	38	-$15,776,131	-1.160
Weighted MA	24	24	48	-$22,745,914	-1.701
Triangular MA	20	20	40	-$17,317,322	-1.291
MA Envelopes	2	1	3	$2,347,292	0.297

1/18/16-1/18/17: Strong Trending; Sharpe Rankings: Inconsistent
H(25) = 0.7250, H(50) = 0.7073, H(100) = 0.6169
Sharpe Index: Buy & Hold = 2.244, Bollinger = 1.545, MA Envelope = 0.874

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$21,090,947	2.244
Bollinger Bands	1	2	3	$11,494,538	1.545
Simple MA	19	18	37	-$12,149,841	-1.273
Exponential MA	15	14	29	-$10,944,990	-1.218
Weighted MA	11	10	21	-$3,646,704	-0.374
Triangular MA	10	9	19	-$8,522,609	-0.924
MA Envelopes	2	2	4	$7,707,956	0.874

1/18/17-1/18/18: Random; Sharpe Rankings: Consistent
H(25) = 0.5301, H(50) = 0.4812, H(100) = 0.4569
Sharpe Index: Buy & Hold = 4.151, Exponential MA = 1.55, Simple MA = 1.235

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$23,158,138	4.151
Bollinger Bands	3	4	7	-$10,199,580	-1.648
Simple MA	11	11	22	$7,310,906	1.235
Exponential MA	7	7	14	$9,536,181	1.555
Weighted MA	8	8	16	$4,195,167	0.560
Triangular MA	7	7	14	$6,392,523	0.869
MA Envelopes	0	0	0	$0	0.000

1/18/18-1/18/19: Strong Tending; Sharpe Rankings: Consistent
H(25) = 0.8210, H(50) = 0.7045, H(100) = 0.7053
Sharpe Index: MA Envelope= 2.450, Bollinger = 0.681, Weighted MA = -0.133

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	-$4,705,967	-0.310
Bollinger Bands	3	2	5	$9,940,145	0.681
Simple MA	14	14	28	-$4,709,730	-0.328
Exponential MA	11	12	23	-$17,595,941	-1.184
Weighted MA	9	9	18	-$1,892,474	-0.133
Triangular MA	6	7	13	-$8,466,592	-0.575
MA Envelopes	4	4	8	$33,792,218	2.450

1/18/19-1/18/20: Random; Sharpe Rankings: Consistent
H(25) = 0.5423, H(50) = 0.5024, H(100) = 0.4623
Sharpe Index: Buy & Hold = 2.648, MA Envelope = 1.812, Simple MA = 1.356

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$25,272,874	2.648
Bollinger Bands	3	4	7	$12,927,335	0.937
Simple MA	8	8	16	$17,244,404	1.356
Exponential MA	7	7	14	$3,050,918	0.143
Weighted MA	4	4	8	$8,886,442	0.545
Triangular MA	4	4	8	-$2,769,187	-0.246
MA Envelopes	2	1	3	$25,620,375	1.812

1/18/20-1/18/21: Strong Trending; Sharpe Rankings: Consistent
H(25) = 0.4475, H(50) = 0.6375, H(100) = 0.6257
Sharpe Index: Weighted MA = 1.307, Triangular MA = 1.017, Exponential MA = 0.914

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$13,159,609	0.721
Bollinger Bands	2	2	4	$2,925,696	0.320
Simple MA	9	9	18	$12,819,522	0.754
Exponential MA	6	6	12	$16,563,606	0.914
Weighted MA	9	9	18	$23,822,662	1.307
Triangular MA	5	5	10	$17,842,899	1.017
MA Envelopes	4	4	8	-$26,187,792	-0.679

1/18/21-1/18/22: Partial Trending; Sharpe Rankings: Consistent
H(25) = 0.6406, H(50) = 0.4650, H(100) = 0.5251
Sharpe Index: Bollinger = 2.566, Buy & Hold = 2.011, Exponential MA = 0.423

Strategy	Long	Short	Total	$ Return	Sharpe Ratio
Buy & Hold	1	0	1	$20,285,577	2.011
Bollinger Bands	4	5	9	$24,365,638	2.566
Simple MA	14	15	29	-$11,365,616	-0.995
Exponential MA	7	8	15	$3,492,972	0.423
Weighted MA	14	15	29	-$15,372,260	-1.354
Triangular MA	10	11	21	-$5,018,877	-0.432
MA Envelopes	1	1	2	-$8,828,577	-0.780

Long: Number of long trades
Short: Number of short trades
Total: Total number of trades
$ Return: Cumulative return from $100m investment
Sharpe Ratio: Risk premium (Rate of Return - Risk-Free Return) per unit of risk (Standard Deviation)

Data Source: Bloomberg

Actual Versus Expected Performance, Discrimination, and the Retention and Dismissal of Major League Baseball Managers

Rodney Paul and Alison Gilmore, Syracuse University
Andrew Weinbach, Coastal Carolina University
Mark Wilson, St. Bonaventure University

Abstract

A simple financial market instrument from the sports wagering marketplace, season win totals, are used as a proxy for expectations of team success in Major League Baseball. This win total is compared to actual wins and this figure is used to help evaluate performance compared to expectations to help in modeling the retention and dismissal of managers. Actual versus expected performance is an important determinant in models of CEO turnover and can directly be applied to managers in Major League Baseball. In this context, treatment discrimination is investigated with respect to race and former player status.

JEL Codes: L83, Z2, Z22
Keywords: Sports, Baseball, Manager, Discrimination, Race

Introduction

Discrimination against workers is a potential issue for all levels of occupations across the economy. Whether that discrimination takes place before groups are ever hired or if they are subject to different standards once hired, identifying and devising strategies to protect and help these groups can lead to a more equitable work environment for everyone.

Coaches and managers in the sports world are not immune to the possibilities of discrimination in the workplace. Allegations of different treatment based on race, gender, or other factors have surfaced in the world of sport and some studies have found evidence of its existence over time. In this study, the role of potential discrimination is examined in two forms for Major League Baseball managers, who is the equivalent of the head coach in other sports such as football and basketball. First, race is examined for any significant role in the decision to retain or dismiss a manager, either after or during the season. Second, being a former Major League Baseball player is assessed if it leads to a separate set of rules for these managers, specifically if they are treated differently than managers who were not former players in the top league in baseball.

To test for discrimination, the importance of performance is compared to expectations in the retention and dismissal of managers. The supposition, based on previous studies noted in the literature review below and general observation, is that managers, like corporate CEOs, are typically not only judged by actual performance, but by their performance compared to what was expected of them over a certain period. If a manager does well compared to their peers in terms of actual performance, whether that has to do with revenues and profits in most industries or, in the case of baseball managers, in terms of wins, this is often not enough information to garner whether the manager was successful unless the baseline standard of expectations is known. For instance, a Major League Baseball manager could be extraordinarily successful or a disappointment with a "five hundred" (0.5) win percentage. The season is likely deemed a success if expectations were below that level or a disappointment if expectations were above that level.

One of the advantages of studying sports for a variety of business and societal issues is that data in many areas has been more plentiful and available to researchers. The case of evaluating manager performance compared to expectations is no different. While in many industries, information on expectations of performance may be difficult to ascertain, in sports it is often straightforward, and better-yet, market driven. The betting market for major sports leagues offers preseason wagers on season win totals. A season win total is an over/under wager on the number of wins a team will have in each season. An over bet is successful if the team wins more than the posted number, while an under bet is successful if the team wins fewer games than the posted figure. These prices (season win totals) are formed in a simple financial market and are available for wagering by the public before the season begins. These prices serve as expectations for the upcoming season for the team and its manager and is a market-based equivalent of analysts' forecasts used as a proxy for expectations for companies.

Season win totals for Major League Baseball and other sports are archived on the website sportsoddshistory.com. Season totals presented on the website are those posted before the start of the regular season. These totals can be compared to the actual performance of managers by simply comparing the win percentage implied by the season total to the actual win percentage of the team and its manager. The assumptions are, if the actual win percentage exceeds the expected win percentage, based on the betting market season total, then the manager outperformed compared to expectations, while if the actual win percentage was less than the expected win percentage, the manager disappointed. It is hypothesized this is the major driving force behind why managers are retained or dismissed and set forth to evaluate this through a logit model. With the role of expectations in place, it is then straightforward to test for discrimination by race and former player status through the inclusion of dummy variables representing minority status and if the manager was a former player.

This paper is structured as follows. The following section if a literature review of studies of CEO and manager retention and dismissal, including past studies of discrimination. The third section discusses the data and the computation of the key variable of interest as it relates to performance compared to expectations for Major League Baseball managers. The fourth section presents the logit model and its results, while the last section discusses the results and concludes the paper.

Literature Review

The key area of focus of this paper is the role, if any, of discrimination in the marketplace for Major League Baseball managers. Research has previously been undertaken as it relates to the role of discrimination in the hiring and firing of managers and similar position, both outside and inside of sports. When dealing with discrimination in hiring practices, Heilman and Caleo (2018) make the important distinction between access and treatment discrimination. The key difference is when these specific types of discrimination could take place. In the time before hiring, access discrimination could be prevalent. Access discrimination occurs when groups in society are restricted in their opportunities to be hired. In this type of discrimination, it is not based on actual or expected performance in the role, but due to discrimination based on factors such as race and gender. After individuals have been hired, a different type of discrimination is possible. This discrimination, known as treatment discrimination, could lead to people in certain identifiable groups not being treated equally, as they could be dismissed from their position or passed over for promotion due to race or gender. In a study of the NFL, Foreman and Turick (2021) found that Black coaches in the NFL are less likely to be promoted from position coaches to central coaching positions. Cunningham (2019) found similar promotion obstacles related to race and gender in the general workplace outside of sports. Unequal treatment of workers based on race and gender has also been used to show and note the persistence of earning gaps between specific groups of workers (Wicker et al., 2021).

Analyzing the performance of female jockeys in UK and Irish horseracing, Brown and Yang (2015) found women are slightly underestimated (outperforming expectations based on betting odds) in the sample, but underestimated to a greater extent in jump racing, where participation by women is particularly low. In another study related to the investigation of discrimination, Gomez-Gonzalez et al. (2018) found that Black head coaches were 8.1% more likely to be dismissed than white coaches in a sample of 20 years of NBA game and betting market data.

Madden (2004) studied discrimination against minority head coaches in the NFL. Minority head coaches were found to be dismissed earlier than white coaches with similar performance characteristics. This qualifies as a form of treatment discrimination as described in previously mentioned literature. A later study by Madden and Ruther (2011) did not find evidence of treatment discrimination, which they attributed to the introduction of the Rooney Rule. The Rooney Rule in the NFL is a policy that requires teams to interview minority candidates for head coaching and senior football operation positions. It aims to promote diversity and equal opportunities within the league by ensuring that minority candidates are considered for these key roles.

In their 2011 study, discrimination against minority coaches was not found to be statistically significant in their sample. Salaga and Juravich (2020) also did not find discrimination against minority coaches using the ex-post measure of expectations based on individual game results against-the-spread and other factors. Paul et al. (2022) found that when accounting for performance compared to expectations by using betting market futures as a measure of expectations, race was not shown to be a significant determinant of dismissal of coaches, as the difference between actual and expected wins was the main driver of retention and dismissal of coaches in professional football.

Manager performance and expectations has been researched both inside and outside the sports industry. A variety of studies focused on theoretical models of turnover of Chief Executive Officers (Frederickson et al., 1988; Franck, et. al., 2010; and Holmes, 2010). Frederickson et al. (1988) designed a model of turnover based on expectations and performance. Their study focused on the relative power of the current CEO, which individuals oversaw dismissal or retention, and the availability of other viable CEO candidates. On the other hand, Franck et al. (2010) focused on the risk and its role in CEO turnover. Holmes (2010) used a Bayesian learning model of estimating the true abilities of CEOs and devised a cost-benefit approach to retention and dismissal. Frick et al. (2010) focused on turnover models using standard principal-agent theory. For all these studies noted, whether CEOs were retained or dismissed was heavily dependent on performance compared to expectations.

A common way in the literature to obtain ex-ante expectations of CEO performance is through analyst forecasts. Analyst forecasts serve as a proxy for anticipated performance in the marketplace. Farrell and Whidbee (2003) note the many advantages and disadvantages of the analyst forecast approach to expectations. Brickley (2003) found that performance expectations based on industry analyst forecasts appear to be a significant determinant of CEO turnover in corporate settings.

In relation to sports, it can be argued that managers and coaches are like corporate CEOs. Therefore, their retention or dismissal can be modeled in the same way including the use of performance compared to expectations as a determinant. Wangrow, Schepker, and Barker (2018) found NBA head coach dismissals were linked to expectations, specifically to underperformance. They used a variety of factors to measure expectations including last year's performance, playoff metrics, and attendance metrics. Humphreys et al. (2016) used point spreads from the betting market as a direct measure of expectations. The authors found college football coaches who outperform expectations were more likely to be retained, while those that disappointed compared to expectations were more likely to be fired. In a comparable manner, Salaga and Juravich (2020) used the same technique in the NFL, using ex-post record against-the-spread and found a significant effect of the variable on retention and dismissal in professional football.

The use of betting odds as a proxy for expectations has been used in European soccer to evaluate coaching performance. Papers by Pierer, Nuesch, and Franck (2014), Buraimo, Bryson, and Simmons (2017), and Elaad, Jelnov, and Kantor (2018) incorporated betting odds to evaluate coaching performance in countries such as England, Germany, Italy, and Spain. Similar findings were discovered across these studies in that performance relative to expectations as measured by betting market odds played a significant role in the likelihood of coach dismissal. Barros, Frick, and Passos (2009) find that coaches with higher payrolls tend to be dismissed earlier in German soccer. In Dutch football, Van Ours and van Tuijl (2016) found teams are more likely to dismiss more experienced coaches, other things equal.

The role of team behavior and prior playing experience of coaches has also been investigated in the sports literature. Betting data from Sportsoddshistory.com was used to study behavior of teams in Major League Baseball (Roach, 2020). Findings revealed teams with higher win totals and teams that underperform expectations tend to increase spending on payroll to a greater degree. Del Corral, Maroto, & Gallardo (2017) found some evidence that former professional players appeared to be more efficient as coaches in the Spanish Top League of Basketball.

Data

The key data for this study is from the futures market for Major League Baseball teams. Commonly referred to as "Season Totals" or "Win Totals," the futures created in the betting market are an over/under wager on the total number of wins a team will have in the 162-game season of baseball. These figures are available for wager before the season begins for each team in the league. Data on managers and MLB team performance were obtained from www.baseball-reference.com. Minority coaches were taken from the Wikipedia entry on African-American managers in Major League baseball history (https://en.wikipedia.org/wiki/Category:African-American_baseball_managers), a history of Hispanic managers in Major League Baseball from a news story on the Pennsylvania Patriot-News (https://www.pennlive.com/sports/2016/04/the_latin-born_managers_in_maj.html), and a news story from the Japanball website on the history of Asian-heritage managers in Major League Baseball history (https://japanball.com/past-asian-americans-in-baseball-mlb-players-coaches-and-executives/). The website www.baseball-reference.com was used to determine which coaches were former Major League Baseball players.

The Season Totals wager is set at a specific number of wins. An over wager is a winner if the team wins more than the posted total. An under wager is successful if the team wins fewer than the posted total. When the actual number of wins equals the posted total, all bets are returned, and the wager is considered a push. Data on this type of wager is available on the website www.sportsoddshistory.com, which archives these betting market prices for many years across different sports. The data used in this study include the 1990-2021 Major League Baseball seasons.

Every manager who starts a season is included in the sample. Coaches who are hired mid-season are not included in the sample as updated season totals data for that specific period are not available. A manager does not need to complete the season to be included in the sample, however, as the actual win percentage at the time of dismissal is compared to the expected win percentage (computed as the season win total divided by 162).

When a team changes managers, it is assumed those separated from a team are dismissed for underperformance. While it could be the case that a coach is terminated for inappropriate behavior off the field, (Foreman, Soebbing, and Seifried, 2021), dismissed because better coaching options exist for the team (Foreman and Soebbing, 2015), dismissed in response to one particularly disappointing close loss (Lefgren at al., 2019), or leave voluntarily, there is no clear way to distinguish among those causes with this data. In most cases, it would be challenging to distinguish the true nature of the separation, even with all publicly available information, as some managers may voluntarily leave their team based on a perceived probability of being dismissed based on the outcome of games not yet played. This is a potential weakness for this dataset for which no clear remedy is available. However, the connection between performance and retention/dismissal has been well established in the literature, as noted previously, despite potential shortcomings of not having complete information for every managerial change.

The futures market for season win totals in Major League Baseball serves as a proxy of expected performance for teams and their managers. Season win totals, formed as a price in a simple financial market, is assumed to include all available information about the team, its opponents, the manager, and other factors which are likely to determine success or failure. Expectations are likely to be important to ownership and the front office of a Major League Baseball team when determining if they should retain or dismiss a manager. It is assumed the betting market prices are a good proxy of this figure in the minds of ownership and front office personnel of Major League Baseball teams. A simple regression model with actual wins as the dependent variable and the season win total as independent variable yields a coefficient of 0.92. Therefore, it is assumed that actual performance only captures part of how managers are evaluated, with performance compared to expectations ultimately being a much more crucial factor when considering their retention or dismissal.

As a straightforward example, assume that a Major League Baseball team finishes the season with a “five hundred” record, in other words, they finish the season 81-81. Whether this season is a success, or a failure depends on how good a given team was expected to be. If expectations were low, such as those represented by a season win total of 71, this season is likely to be considered successful and the manager is more likely to retain their position. On the other hand, if expectations were for a 91-win season (using the betting market as a proxy), the team and manager have likely disappointed compared to expectations, increasing the probability the manager will be dismissed. As can be seen from this example, the betting futures market provides market-based insights which are important to owners and the team front office when considering managerial changes for the next season (or within-season).

Model

The logit model used in this study is as follows:

$$(\text{Manager Dismissal})_{i,t} = \alpha_0 + \beta_1(\text{Age}_{i,t}) + \beta_2(\text{First Year Manager Dummy}_{i,t}) + \beta_3(\text{Years of Experience}_{i,t}) + \beta_4(\text{Years of Experience}_{i,t})^2 + \beta_5(\text{Other Managerial Experience}_{i,t}) + \beta_6(\text{Other Managerial Experience}_{i,t})^2 + \beta_7(\text{Win Percentage}_{i,t}) + \beta_8(\text{Difference Between Actual and Expected Win Percentage}_{i,t}) + \beta_9(\text{Minority Dummy}_{i,t}) + \beta_{i,t}(\text{Former MLB Player Dummy}_{i,t}) + \varepsilon_{,t} \quad (1)$$

The dependent variable is a dummy variable for if the manager was dismissed or retained. Dismissed takes a value of one, while retained takes a value of zero, for this variable. With the dependent variable being binary in nature, a logit model was run to investigate the role of performance compared to expectations, race, and former player status on manager dismissal and retention. Probit models were also run, with comparable results.

The independent variables included in the regression model include manager age in years, a dummy variable for the first year managing a team, years with the organization (tenure) and its square, other managerial experience (tenure) in years and its square, team win percentage for the season, the difference between actual and expected win percentage, and dummy variables for minority managers and/or managers who were former MLB players.

Age is included as older managers may not be as effective in communicating with young players, so it could have an impact on the decision of a team to dismiss. Recent studies of coach dismissal, such as Gomez-Gonzalez et al. (2018), included age as an explanatory variable in their model. Experience could be valuable, however, but when the age (in years) was included as a quadratic in the model, neither term was statistically significant, so the simple specification with just age (in years) is shown below. The original study of Black head coaches in the NFL by Madden (2004) and the follow-up paper by Madden and Ruther (2011) both included experience in their models.

A dummy variable was included for managers who are in their first year with the organization. Managers are generally given some time to put their strategies into place for an organization, which means that first-year managers are seldom dismissed. This variable is expected to have a negative impact in the model. Years with Organization was included as a quadratic in the model. The number of years in an organization is likely to have a non-linear effect over time with respect to dismissal or retention. In a comparable manner, years of managerial experience outside of the current organization is also included in the model, along with the square of this variable.

Win percentage of the most recent completed season was included to account for team performance, with common expectations being that higher win percentages should lead to fewer dismissals. However, as noted previously, performance compared to expectations are likely to be more important in this context than just the overall win percentage. Therefore, the actual win percentage minus the expected win percentage is included to account for expectations in the model. It is expected that if a manager outperforms expectations, higher win percentage than expected, he is less likely to be dismissed. However, if the manager disappoints compared to expectations, he is more likely to be dismissed.

Dummy variables for minority managers are included in the model in the form of dummy variables. A simple binary variable that takes the value of one when the manager is a minority was included as one specification, while dummy variables for individual minority groups (Black, Hispanic, Asian) were also included in an alternative model. A dummy variable also was included in the model for managers who were former MLB players, with the dummy taking a value of one if the manager was a former MLB player.

Summary statistics for the non-binary variables are shown in Table 1 below, while a frequency table of the binary variables, in terms of manager-years for former player and minority statuses, is shown in Table 2.

In relation to minority and former player status, model specification I includes the dummy for minority status and the dummy for former player status. The second model specification includes the dummy for former player status and the individual minority group dummy variables.

Table 1. Summary Statistics of Non-Binary Variables – Number of Observations:910

Variable	Mean	Median	Standard Deviation
Season Win Percentage	0.4985	0.5000	0.0952
Expected Win Percentage (Based on Season Total)	0.4916	0.5062	0.0768
Actual Minus Expected Win Percentage	0.0069	0.0000	0.1033
Years with Organization	11.4824	11.0000	6.8618
Other Managerial Experience (Years)	3.4109	0.0000	4.7343
Age (Years)	52.0791	51.0000	7.5374

Table 2. Frequency Table of Binary Variables – Number of Observations: 910

Variable	Frequency
Former MLB Player (Seasons)	759
Black Manager (Seasons)	91
Hispanic Manager (Seasons)	42
Asian Manager (Seasons)	8
Dismissed	207

Table 3. Logit Model Results: Dependent Variable – Manager Dismissed

Variable	I.	II.
Intercept	0.2322	0.2396
	(0.2194)	(0.2253)
Age (Years)	0.0258 *	0.0266 *
	(1.7345)	(1.7202)
First Year Manager Dummy	-1.2889 **	-1.2922 ***
	(-4.6776)	(-4.6886)
Years with Organization	-0.0739	-0.0754
	(-0.9608)	(-0.9796)
Years with Organization^2	0.0041	0.0042
	(0.8169)	(0.8309)
Other Managerial Experience in Years	-0.0263	-0.0277
	(-0.5340)	(-0.5612)
Other Managerial Experience in Years^2	-0.0002	-0.0001
	(-0.0482)	(-0.0359)
Win Percentage	-4.3505 ***	-4.3206 ***
	(-2.8694)	(2.8268)
Difference Between Actual and Expected Win Percentage	-5.1904 ***	-5.2217 ***
	(-3.2253)	(-3.2224)
Minority	0.2790	-
	(1.1765)	-
Black Manager	-	0.3094
	-	(1.0917)
Hispanic Manager	-	0.1811
	-	(0.4522)
Asian Manager	-	-0.4262
	-	(-0.3622)
Former MLB Player	-0.3299	-0.3325
	(-1.4764)	(-1.4776)
McFadden R-Squared	0.1092	0.0192
Likelihood Ratio Statistic	106.5248	106.5241
LR Statistic Probability Value	0.0000	0.0000

Statistical significance is noted by *-notation. * is significant at 10% level, ** at the 5% level, and *** at the 1% level.

Each model specification had similar findings. In each case, the dummy for the manager's first year with the organization was found to have a negative and statistically significant effect on if a manager was dismissed. This illustrates that managers are given time to implement their strategies and structure, making a first-year dismissal highly unlikely. The years with organization and other managerial experience variables, and their squares, were not found to be statistically significant.

Experience with the organization and prior managerial experience do not appear to play a key role in whether the manager gets dismissed or retained. Win percentage was found to have a negative and statistically significant relationship to if a manager was dismissed. All else equal, winning a higher percentage of games leads to a lower likelihood of being dismissed.

The role of expectations compared to actual performance was shown to have a statistically significant effect across both model specifications. This variable was found to be negative and was statistically significant at the 1% level. The greater the difference between actual win percentage of the team and its expected win percentage, the less likely the manager was to be dismissed. Managers who exceeded expectations were more likely to keep their positions, while those who disappointed, compared to expectations, were more likely to be replaced. Converting the logit coefficient on the difference between actual and expected win percentage to probability, underperforming your expected win percentage as a Major League Baseball manager by 0.1 will lead to a 9.9% increase in the likelihood of being dismissed from your position.

The dummy variables related to race were not found to be statistically significant in any specification. Whether included as a straightforward binary variable for a minority manager or broken down into specific groups in the second specification, none of the race-related variables were found to be statistically significant. Also, being a former MLB player was also not found to have a significant effect on if a manager was retained or dismissed. Overall, race and former player status were not found to play a significant role in decisions to retain or dismiss managers in Major League Baseball.

Discussion and Conclusions

When evaluating manager performance in any field, it is important to consider the role of expectations compared to actual performance. Actual performance, in terms of revenues or profits in the business world, or in terms of wins as it relates to directly to Major League Baseball managers, as observed in this study, may convey information, but mostly only when coupled with expected performance. The role of actual versus expected performance is studied as it relates to baseball managers using prices in a simple financial market, the season win totals posted in the betting market.

In the season win totals market, an over/under wagering price is posted where a bettor can wager if the team will exceed or fall short of this number of wins. Converting these win totals into win percentages and comparing them with actual win percentage performance during the season allow for a simple computation of performance compared to expectations. If a team wins more than their season win total, the team and its manager are assumed to have outperformed compared to expectations. On the other hand, if the team wins fewer games than expected, the team and manager are assumed to have disappointed compared to expectations. Logically, this performance compared to expectations has major ramifications for job retention as outperforming expectations is likely to lead to retention, while disappointment is more likely to lead to dismissal.

Using data from 1990-2021, the difference between actual and expected win percentage for Major League Baseball managers is computed and included this in a logit model of whether a manager was dismissed or retained. This variable was shown to be highly statistically significant as when actual performance was worse than expected performance, the manager was found to be more likely to be dismissed. In addition to the role of expectations compared to actual performance, manager years in organization, its square, and win percentage were also found to be statistically significant.

The role of discrimination was then evaluated, using dummy variables based on race of the manager and former player status. If race played a role in dismissal versus retention, once accounting for performance compared to expectations, dummy variables for minority status (or individual race dummy variables) or if a manager was a former player should account for any discrimination in this marketplace. The logit results revealed no statistically significant results as it related to any of the race-related dummy variables, nor the former player status variable. In short, no evidence of discrimination was found through these tests and the dismissal or retention of managers was found to be driven by their performance compared to expectations.

These findings are important for two key reasons. First, without measurable expectations, any analysis involving retention and dismissal of any employee is likely missing key information. Absolute performance only goes so far in explaining personnel decisions in the workplace as expectations as a benchmark is an imperative factor. The second reason is that without expectations in the model, hot-button societal issues, such as the role of race as studied in this paper (which could be easily extended to studies of gender, age, etc.), need to have results interpreted with caution, as what may appear to be bias could simply be a function of performance compared to management or owner expectations.

While no evidence of discrimination was found in this study in terms of race or former player status, it should be noted only direct testing for treatment discrimination was conducted, not access discrimination. In other words, the data used for this study were only able to address potential discrimination that could have occurred after the manager had already been hired. Specifically, the tests were for if managers were treated differently in terms of retention or dismissal after they already had the job. The data and analysis for this study cannot say anything about access discrimination, specifically if there are any barriers to becoming a Major League Baseball manager due to factors such as race. With that in mind, however, this study offers evidence that whether a manager keeps or loses his job in Major League Baseball appears to be driven by merit, rather than other factors.

References

Barros C, Frick B and Passos J (2009) Coaching for survival: The hazards of head coach careers in the German Bundesliga. Applied Economics 41:3303–11

Brickley JA (2003) Empirical research on CEO turnover and firm-performance: a discussion. Journal of Accounting and Economics 36(1-3):227-233

Brown A, Yang F (2015) Does society underestimate women? Evidence from the performance of female jockeys in horse racing. Journal of Economic Behavior Organization 111:106-118

Buraimo B, Bryson A, Simmons R (2017) Time to go? Head coach quits and dismissals in professional football. Working Paper No. 10600, Institute for the Study of Labor (IZA)

Cunningham G (2019) Diversity and Inclusion in Sport Organizations: A Multilevel Perspective (4th ed). Routledge

del Corral J, Maroto A, Gallardo A (2017) Are former professional athletes and native better coaches? Evidence from Spanish basketball. Journal of Sports Economics 18(7):698-719

Elaad G, Jelnov A, Kantor J (2018) You do not have to succeed, just do not fail: When do soccer coaches get fired? Managerial and Decision Economics 39(3):269-274

Farrell KA, Whidbee DA (2003) Impact of firm performance expectations on CEO turnover and replacement decision. Journal of Accounting and Economics 36(1-3):165-196

Foreman JJ, Turick RM (2021) The effect of race on lateral moves to coach central positions. Sport Mgmt Rev 24(1):69-91

Foreman JJ, Soebbing BP (2015) The role of candidate availability in CEO dismissals: an examination of the National Football League. Journal of Management Policy and Practice 16(2):11-25

Foreman JJ, Soebbing BP, Seifried CS (2021) Executive deviance as a sociopolitical force in dismissals. Deviant Behavior 42(2):193-209

Franck E, Nüesch S, Pieper J (2010) How expectations affect managerial change. Working Papers 0139, University of Zurich, Institute for Strategy and Business Economics (ISU)

Frederickson JW, Hambrick DC, Baumrin S (1988) A model of CEO dismissal. Academy of Management Review 13:255-270

Frick B, Barros C, Prinz J (2010) Analyzing head coach dismissals in the German "Bundesliga" with a mixed logit approach. European Journal of Operational Research 200(1):151-159

Gomez-Gonzalez C, del Corral J, Maroto A, Simmons R (2018) Racial differences in the labor market: The case of expected performance and dismissals of head coaches in NBA. Working Paper 2018-2. DAEF, UCLM

Heilman M, Caleo S (2018) Combatting gender discrimination: A lack of fit framework. Group Processes & Intergroup Relations 21(5):725-744

Holmes P (2010) Win or go home: Why college football coaches get fired. Journal of Sports Economics 11(1):1-22

Humphreys B, Paul R, Weinbach A (2016) Performance expectations and the tenure of head coaches: Evidence from NCAA football. Research in Economics 70(3):482-492

Lefgren LJ, Platt B, Price J, Higbee S (2019) Outcome based accountability: Theory and evidence. Journal of Economic Behavior & Organization 160:121-137

Madden J (2004) Differences in the success of NFL coaches by race, 1990-2002: Evidence of last hire, first fire. Journal of Sport Economics 5(1):6-19

Madden J, Ruther M (2011) Has the NFL's Rooney Rule efforts "leveled the field" for African American head coach candidates? Journal of Sport Economics 12(2):127-142

Paul RJ, Weinbach AP, Wilson M (2022) Betting market expectations and NFL coach retention and firing. Journal of Gambling Business and Economics (forthcoming).

Pieper J, Nüesch S, Franck E (2014) How performance expectations affect managerial change. Schmalenbach Bus Rev 66:5-23

Roach MA (2020) Team performance relative to expectations and its impact on player personnel decisions in major league baseball. Available at SSRN 3701851

Salaga S, Juravich M (2020) National Football League head coach race, performance, retention, and dismissal. Sport Management Review 23(5):978-991

van Ours JC, van Tuijl MA (2016) In-season head-coach dismissals and the performance of professional football teams. Economic Inquiry 54(1):591-604

Wangrow DB, Schepker DJ, Barker III, VL (2018) Power, performance, and expectations in the dismissal of NBA coaches: A survival analysis study. Sport Management Review 21(4):333-346

Wicker P, Breuer C, Dallmeyer S (2021) The gender earnings gap among elite athletes in semi-professional sports. Managing Sport and Leisure, DOI: 10.1080/2375472.2021.1944819

Optimizing NBA Roster Construction

Nick Riccardi, Syracuse University

Abstract

This study aims to quantify the effect that complementary player types have on team success in the National Basketball Association. Using cluster analysis, player-seasons are redefined from their traditional basketball positions to better encompass the roles that players play. For the 10 seasons of data, the best player for each of the 30 teams in the league is determined and teams are grouped based on the cluster of their best player. Ordinary Least Squares regressions are performed to test what player types fit together best. The results of this study show the importance of complementary workers to a firm's success.

JEL Codes: C1, J0
Keywords: team performance, NBA, clustering

Introduction

The success of a professional sports team is dependent on, among other factors, the level of talent that the team possesses and how the players on the team fit together. The National Basketball Association (NBA) is no exception to this and perhaps the most reliant on these two factors given that only five players for a team are on the court at once and a team's best players play most of the game. Therefore, it is vitally important for NBA general managers to acquire players who are good enough to bring the team success while also fitting together well on the court. Traditional basketball mindsets can make this task more daunting than it already is. With how the game is currently played in the era that emphasizes three-pointers[1] and the use of analytics, players are playing in less position-specific roles than they have before. This means that general managers need to assess the specific style in which a player plays when they evaluate them.

Not only do general managers need to be able to determine the playstyle of a player, but they also need to be able to figure out which other playstyles fit well with that style. Building an NBA team begins with the process of putting complementary players around the team's best player. In this study, player positions will be redefined based on the styles of players across the league. Using cluster analysis, player-seasons will be grouped together based on similar statistics. Linear regression models will be conducted with the purpose of figuring out which playstyles work well with others. The goal is to determine the types of players that complement specific styles of play in order to figure out how to build around the style of a team's best player.

Literature Review

Positions in basketball have always followed the traditional specification of having five players on the court, usually a point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C). McMahan (2018) states that this declaration of categorizing players into these five positions is the result of overall NBA strategy. Definitions of the roles for each of these positions are provided on the NBA's website ("Basketball Positions," nd). A point guard is a player who "runs the offense and usually is the team's best dribbler and passer. The point guard defends the opponent's point guard and tries to steal the ball." Shooting guards are described as "usually the team's best shooter. The shooting guard can make shots from long distance and also is a good dribbler." A small forward is perhaps the most versatile of the five positions, given that he "plays against small and large players. They roam all over on the court. Small forwards can score from long shots and close ones." The next position, a power forward, "does many of the things a center does, playing near the basket while rebounding and defending taller players. But power forwards also take longer shots than centers." The last of the five traditional positions is the center, which "is the tallest player on each team, playing near the basket. On offense, the center tries to score on close shots and rebound. But on defense, the center tries to block opponents' shots and rebound their misses."

The increase in overall athleticism and skill of players coupled with the use of analytics in today's modern NBA has changed the way these traditional positions are viewed. Now, with the increase in the importance of three-point shooting and being able to switch defensive assignments, basketball has become a positionless game. The current NBA commissioner, Adam Silver, shared this sentiment before Game 1 of the 2022 NBA Finals, saying, "We're a league that has moved increasingly to positionless basketball" (Aschburner, 2022). This is in part due to players like Dirk Nowitzki, a 7'0" power forward who could shoot from long distances with the same skill as guards. In today's NBA, power forwards and centers, also known in combination as "bigs," are not only encouraged to shoot three-pointers, but without the skill, can become obsolete. Over the past two decades in the NBA, teams have been taking more three-pointers each year, with a major increase in threes attempted per game starting in 2013-2014 (Shea, nd). For this reason, the center position, generally assigned to the tallest players in basketball, is dwindling. Ziller (2017) notes that this change in playstyle has led to the typical duties of a center no longer being needed.

The NBA is in need of redefining positions by classifying players by their player type, not the traditional position they play. Many researchers have attempted to solve this problem through machine learning techniques, such as cluster analysis. Cluster analysis can be used to group together players based on similarities of statistics. Having the goal of finding complimentary player types in mind, only research considering team success in relation to clustering NBA players will be discussed. One such example comes from Kalman and Bosch (2020), who modeled lineup efficiency after clustering 3,608 NBA players from 2009-2018. The statistics chosen for clustering in this paper were based on skills, habits, and opportunity. Their clustering algorithm resulted in nine different clusters of players. From there, the authors modeled five-man combinations of clusters to predict Net Rating, which is the scoring differential per 100 possessions. A linear regression model was conducted to determine the effect that the number of players from each of the nine clusters within the combination of five players has on Net Rating. Lastly, a Random Forest Model was used to predict the Net Rating of all possible five-man combinations of clusters.

Zhang et al. (2018) clustered 354 players from the 2015-2016 season based on experience, weight, and height which resulted in five clusters described by these three factors. The authors then analyzed the distribution of clusters across different levels (based on performance) of teams. Patel (2017) also used one season of data to cluster 486 players from the 2016-2017 season based on per-100-possession stats. The clustering resulted in four clusters of players, which the author described as "The Paint Protectors," "The Supporters," "The Shooters," and "The Insiders." The author did not find significant relationships between team success and the cluster membership of a team's players. However, a significant result was found regarding the distance of players from their cluster centroid and team success.

Duman et al. (2021) clustered players within their traditional basketball position. That is, within each of the five traditional basketball positions, clustering was performed to distinguish the player types within a position. Four different clusters were created amongst players who were identified as point guards, shooting guards, and small forwards, respectively, while five clusters were created for power forwards and six for centers. The authors found the clusters of each of the five positions that were part of the most successful teams as well as pairs of clusters across two traditional positions. Osken and Onay (2022) used the clustering of NBA players to predict the outcome of NBA games. In this paper, players from the 2012-2013 to the 2017-2018 seasons were clustered using box score, advanced efficiency, and shot selection data. The authors predict the winners of NBA games using an artificial neural network that takes into account the minutes played by each cluster for each team as well as factors such as win percentages of teams, month of season, and days of rest for teams. The prediction accuracy was greater than 75%.

Furthermore, Tsai (2017) clustered all NBA players who averaged one shot per game or more from 2010 to 2016. First, the author clustered players using cumulative shot chart data. The shot charts were converted into heat maps which were then converted into a data matrix of players' shots. K-means clustering was then used on this matrix, and it was determined that seven clusters were the optimal amount. A second cluster analysis was based on player performance statistics, 16 in total, that were taken from the NBA's website. This clustering led to eight different groups. The author then took each of the matrices created for the two k-means clustering analyses and combined them into one. This matrix revealed an R-squared value of .71 in terms of its correlation to predicted winning percentages for teams throughout the league.

This study adds to the existing literature regarding clustering NBA players and team success by evaluating team success under the condition of the player type of the team's best player. This context is important because teams do not always have access to the types of players they desire and need to optimize their roster under the constraints of available players and assets. Additionally, the same makeup of players in a lineup will inherently be different depending on the role that the team's best player plays. For example, if a team's best player is a pass-first player, then team success will be highly dependent on the ability of the players around the best player to score off their passes. On the other hand, if the best player on a team is a score-first, high-shot attempt player, then the players around him might need to be good at setting screens and providing space on the court for the best player to get shots off. These two examples might have the same makeup of player styles on the team, but without the context of the playstyle of the best player, the distinction can't be made on whether the playstyles are a good fit. Furthermore, this study includes the 10 most recently completed regular seasons of the NBA. These 10 seasons begin with the start of the current three-point era that the NBA is in. Therefore, this study is a novel approach to finding the relationship between player types and team success in the NBA by considering the differences in complementary player types based on the player type of a team's best player as well as including the most recent data available.

Data and Methodology

The data gathered for this study are in the form of player-seasons gathered from basketballreference.com. That is, each row of data represents a given NBA player during a given season. This includes per-game, shooting, advanced, totals, and play-by-play statistics for every player from the 10 seasons between 2013-2014 and 2022-2023. The per-game statistics represent typical box score data such as points, assists, rebounds, blocks, steals, field goal percentage, and more. The totals statistics encompass the same statistics as per-game but are sums for each player for the entire season. To get an understanding of where and how players are taking their shots, the shooting statistics shine a light by providing the percentage of shots each player takes from five distance ranges, their field goal percentage from each range, the percentage of their shots that are assisted, and more. The distance ranges include zero to three feet away from the basket, three to 10 feet, 10 to 16 feet, 16 feet to the three-point line, and three-pointers. Advanced statistics include rate statistics, such as rebounding percentage, assist percentage, and steal percentage, as well as metrics derived to encompass a player's value on offense, defense, and in total. The author invites the reader to access the glossary provided by Basketball Reference which includes definitions for many of the statistics gathered ("Glossary," nd). An example of a definition for a rate statistic is that of TRB% (Total Rebound Percentage), which is "an estimate of the percentage of available rebounds a player grabbed while he was on the floor." Lastly, play-by-play statistics were recorded to get an estimate of where each player plays a majority of their minutes based on traditional player positions. Based on these estimates, each player will be assigned the position that they most frequently play.

Limitations were imposed on which player-seasons would be included in the analysis. Any player who did not play at least half of the season (41 games in a normal NBA season[2] as well as at least 10 minutes per game were taken out of the dataset. The number of games played minimum requirement was implemented because if a player has a small sample size, their effect on a team's winning percentage could be disproportionate. As far as the minutes per game minimum, the goal was for players in the analysis to have an established role on their teams. Playing at least 10 minutes per game warrants being part of a consistent rotation.

These restrictions limited the dataset to 3,145 players over 10 seasons. With these player-seasons, a k-means clustering algorithm was performed with the goal of redefining the traditional player positions into functional player types. K-means clustering "aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster". The algorithm works by minimizing within-cluster variances or squared Euclidean distances. To start, centroids are randomly selected to serve as beginning points for clusters and then iterative calculations are performed for centroid positioning optimization (Medium, 2018). Since k-means clustering is an unsupervised algorithm, the results need to be interpreted by examining what similarities observations have within each cluster as well as what distinguishes the observations in each cluster from the other clusters.

Out of the performance statistics gathered for the player-seasons, the statistics used for analysis were total rebounding percentage (TRB%), assist percentage (AST%), steal percentage (STL%), block percentage (BLK%), turnover percentage (TOV%), standard deviation of the proportion of shots taken from each distance range (SDSH), shot percentage from each distance range, percentage of two-point makes that were assisted, percentage of three-point makes that were assisted, free throw rate, and personal fouls per game. Rate statistics were used rather than their per-game or totals counterparts to eliminate the opportunity that players get. More important than the counts that players rack up for these statistics is how often they occur when the player is on the floor.

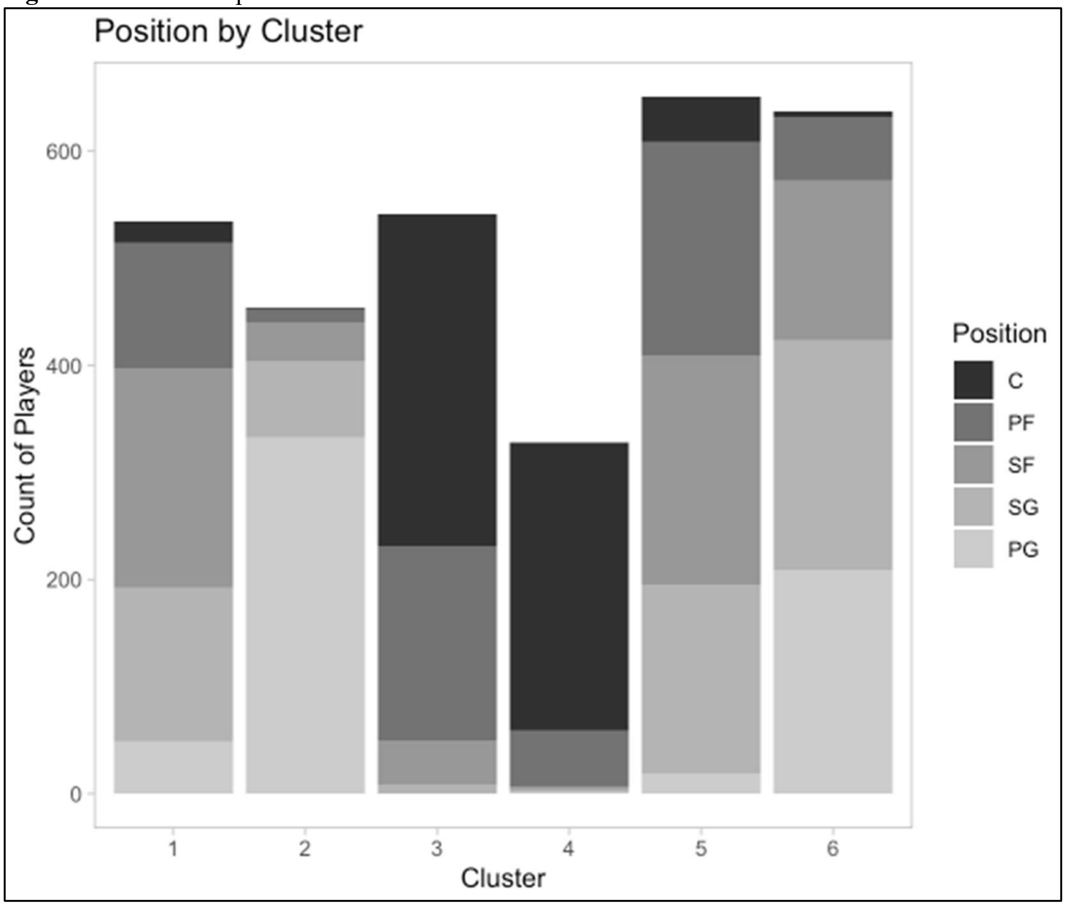
Before the cluster analysis was conducted, each variable was standardized to not disproportionately influence the separation of observations based on differences in scales of variables. Given the large number of variables used for clustering, Principal Component Analysis (PCA) was used to reduce the number of dimensions. PCA is a technique for feature extraction, which creates new variables (the number of which is the same as the number of the original variables) that are combinations of the variables supplied (Brems, 2017). The created variables, called components, are ordered by the proportion of variance they explain between the observations. The first component explains the most amount of variance while the last explains the least amount. Dimensionality reduction is achieved by taking the number of dimensions that account for a certain threshold of cumulative variance explained. In this case, the threshold was set at 75% of the variance explained which was accomplished by the first seven principal components. With these seven components, k-means clustering was performed. The optimal number of clusters, six, was determined by the balance of the count of observations in each cluster and the between sum of squares accounting for 52.1% of the total sum of squares.

After the cluster analysis was conducted and each player-season was assigned to one of the clusters, Ordinary Least Squares regression models were run to figure out how each cluster impacted a team's success depending on which cluster the team's best player came from. Team success was determined by Pythagorean win percentage, which was the dependent variable in the models. Pythagorean win percentage is an estimate of a team's strength based on the number of points they score and allow over a season. A team's best player was based on which player had the greatest Real Plus-Minus (RPM). RPM is an advanced statistic created by ESPN that provides a "Player's estimated on-court impact on team performance, measured in net point differential per 100 offensive and defensive possessions" (ESPN, nd). Limitations were put into place for which players could be declared a team's best player. The best player for each team was selected from the players on that team who played at least 25 minutes per game. This was implemented to prevent role players from being selected as the best player on a team. By playing limited minutes, a player's RPM may not realistically account for their contribution to their team's success. Teams were then grouped together based on which cluster their best player resided in. The independent variables in each of the models were the minutes played by a team's best player and the sum of the minutes played by each cluster for the rest of the players for a team.

Cluster Analysis

The first step in analyzing the results is getting an understanding of what separates the observations into their respective clusters. Figure 1 provides the first glance at the clusters by showing the breakdown of traditional positions by cluster. While the goal of the clustering was to move from traditional positions to player types, it is still informative to analyze the positional breakdown.

Figure 1. Traditional positions breakdown for each cluster



Clusters one and five have similar compositions, being mostly occupied by shooting guards, small forwards, and power forwards. Almost three-fourths of the second cluster is comprised of point guards, while only 14 of these 454 players are bigs (13 power forwards, one center). Clusters three and four, on the other hand, are made up of mostly bigs, with cluster three having 90.8% of players being bigs with zero point guards and cluster four having 97.7% bigs (82% centers). Cluster six has a distribution that's not as refined as clusters two, three, and four while not being as spread out as clusters one and five. Roughly 90% of this cluster is made up of point guards, shooting guards, and small forwards, with guards accounting for about two-thirds of the players within the cluster and almost split perfectly between the two guard positions.

Next, visualizing the shot profile of the clusters helps to understand a major aspect of the offensive side of the game. Figure 2 below shows the mean proportion of shots taken from each distance range for the six clusters. Like the positional breakdown, clusters one and five also have similar shot distributions. These two clusters take the highest proportion of shots from beyond the three-point line (0.466 for cluster one, 0.505 for cluster five), and the smallest proportion of shots between three feet and the three-point line (0.264 for cluster one, 0.296 for cluster five). Clusters two and six also have similar shot distributions, sporting the two lowest SDSH, showing that they take shots from all over the court. The greatest difference in proportion between these two clusters across any distance range is just 0.046 in the zero to three feet range. Lastly, clusters three and four take the shortest distance shots across the six clusters. While cluster three is relatively balanced (third lowest SDSH), over 60% of shots from these players come from within ten feet of the basket. Still, this number pales in comparison to the 83.7% of shots that come from within 10 feet for cluster four. Additionally, while cluster three still takes over one-fifth of their shots from beyond the three-point line, cluster four players only have a proportion of 0.006 in that range.

Figure 2. Proportion of shots taken from each distance range by cluster

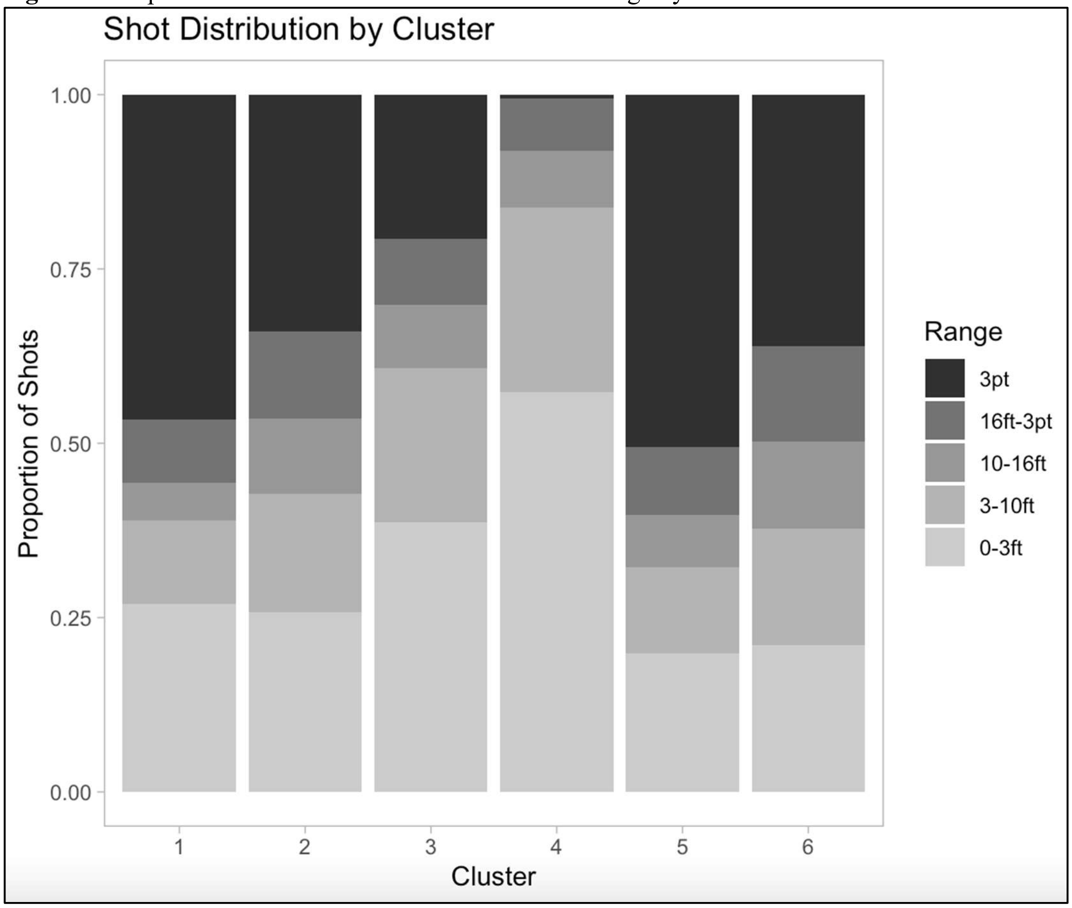


Table 1 provides the means of variables by each cluster. Bolded variables are those that were explicitly used in the dimensionality reduction and subsequent cluster analysis. The appendix provides definitions for each of the variables listed. Figure 1, Figure 2, and Table 1 provide the information necessary to describe the players in each cluster and give them labels based on their functional roles. For cluster one players, it is easier to focus on the negative aspects of their averages rather than the positives, considering how few there are. Cluster one players are the worst offensively out of the six clusters. These players rank last in overall field goal percentage, having the lowest field goal percentages in the three ranges between zero and 16 feet from the basket and second to last between 16 feet and the three-point line.

Their AST% is marginally better than the bottom two clusters in the statistic and their USG% ranks last, showing that teams generally do not turn to them on the offensive side of the game. However, cluster one players provide value on defense and from beyond the three-point line. The cluster averages the second highest STL%, third highest BLK% (highest among non-big-dominated clusters), and third highest three-point percentage. Players in cluster one can be considered defensive specialists. Common players in this cluster include Danny Green (6x), Kentavious Caldwell-Pope (5x), and P.J. Tucker (6x).

The most important statistic to describe cluster two players is AST% in which they average the highest percentage by a wide margin. Additionally, cluster two players average the highest STL%. Both findings are not surprising given that the cluster is dominated by point guards. These players average the lowest percentage of their field goals being assisted, the second highest USG%, the most minutes played per game. A couple downfalls to players in this cluster are that while they average the highest AST%, they also average the highest TOV%, and they are not particularly great at shooting from any specific distance range. Given their highest AST% and low assisted percentage, players in cluster two are playmakers, creating shots for their teammates and themselves. Some notable players from cluster two are Chris Paul (10x), James Harden (10x), and LeBron James (8x).

Cluster three players are skilled across many statistics. They average the second highest TRB%, third highest AST%, second highest BLK%, highest field goal percentages within 10 feet, and the second highest free throw rate. While they take almost 40% of their shots beyond 10 feet, they are not as effective as close to the basket, generally shooting about average from farther distances. Still, these players can stretch the floor with outside shooting while being highly effective close to the basket, are strong rebounders, can move the ball well, and have a strong defensive presence. Players in cluster three can generally be thought of as versatile bigs. Examples of players from this cluster include Anthony Davis (9x), Giannis Antetokounmpo (9x), and Nikola Jokic (7x).

Players in cluster four can also generally be thought of as bigs, but in the traditional sense. These players fit closely to the definition of a center, highlighted by ranking first in TRB% and BLK% and rarely taking shots outside of 10 feet from the basket, hence the highest SDSH belonging to the cluster. Their average shot distance is less than five feet, they are second lowest in USG%, highest in field goal percentage, lowest in AST%, second lowest in STL%, and likely only take threes when they must heave the ball up at the end of the shot clock or quarter, evident by their extremely low three-point rate and three-point percentage. Once again, this cluster consists of traditional bigs. A few players that belong to this cluster often are Andre Drummond (9x), DeAndre Jordan (8x), and Rudy Gobert (9x).

Table 1. Variable means by cluster

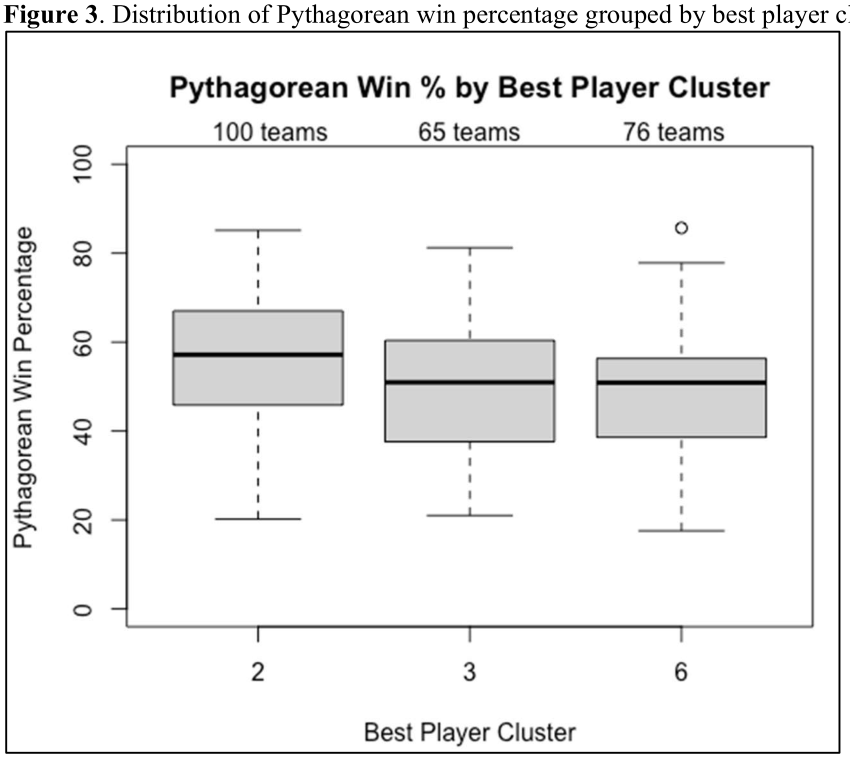
Variable	Cluster One	Cluster Two	Cluster Three	Cluster Four	Cluster Five	Cluster Six
TRB%	8.54	7.48	14.8	17.2	8.33	7.23
AST%	9.92	28.9	11.6	8.22	8.57	18.4
STL%	1.73	2.17	1.39	1.32	1.27	1.46
BLK%	1.37	0.949	3.25	3.70	1.17	0.853
TOV%	12.2	16.2	12.5	14.5	9.39	11.8
SDSH	0.200	0.136	0.166	0.246	0.201	0.123
FG% 0-3ft	59.8	61	70.1	67.5	67.2	62.7
FG% 3-10ft	29.7	38	43.4	40.5	41.5	41.6
FG% 10-16ft	29.9	39.9	39.5	36.2	41.5	43.1
FG% 16ft-3pt	31.6	38.3	37.4	26.5	38.3	40.9
3P%	34.5	33.3	32.4	0.96	36.5	35.5
% 2P Ast'd	57.2	27.9	65.5	68.7	66.1	37.2
% 3P Ast'd	92.4	69.1	95	2.5	94.9	78.9
FT Rate	0.219	0.276	0.316	0.404	0.176	0.237
PF/G	1.84	2.07	2.51	2.33	1.65	1.89
Min/G	21.8	27.7	25.3	21.7	22.2	27.4
PTS/G	7.71	13.6	12.5	8.29	9.03	14.1
TRB/G	3.37	3.85	6.79	6.91	3.31	3.65
AST/G	1.51	5.28	1.97	1.19	1.33	3.24
STL/G	0.772	1.21	0.719	0.578	0.582	0.81
BLK/G	0.346	0.317	0.945	0.96	0.3	0.28
FGA/G	6.63	11	9.43	6.06	7.39	11.5
FG%	41.5	43.5	51.7	56	44.5	44.4
3PA/G	3.12	3.71	1.97	0.025	3.66	4.09
FTA/G	1.45	3.26	3.01	2.4	1.33	2.84
USG%	16.3	22.2	20.2	16.5	16.9	22.3
Avg. Dist	15.1	13.8	9.91	4.68	16.5	14.7
% Shots 0-3ft	0.269	0.257	0.387	0.573	0.198	0.211
% Shots 3-10ft	0.12	0.17	0.221	0.264	0.124	0.167
% Shots 10-16ft	0.054	0.109	0.091	0.082	0.074	0.124
% Shots 16ft-3pt	0.090	0.124	0.095	0.074	0.098	0.137
% Shots 3P	0.466	0.34	0.207	0.006	0.505	0.361
% FG Ast'd	73.5	41.9	71.6	68.2	80.6	52.2

There are a few defining characteristics of players in cluster five. Most evident of these are statistics relating to three-point shooting. These players take the greatest proportion of three-point shots while averaging the highest percentage from three as well. Players from this cluster are also effective between 10 feet and the three-point line and have the highest percentage of their field goals assisted. Other than their ability to catch and shoot the ball, they do not provide much value elsewhere. Cluster five players are low in AST% and TRB% and average the lowest STL% and third lowest BLK%. While their TOV% is the lowest of the six clusters, this is a microcosm of not having the ball in their hands for long, evident by their low USG%, low free throw rate, and high assisted percentage. Players from this cluster can be described as perimeter shot takers. Prime examples of cluster five players include JJ Redick (6x), Terrence Ross (8x), and Klay Thompson (7x).

Lastly, cluster six players have an affinity for scoring the ball. Their lowest SDSH amongst the clusters show that they are willing to take shots from anywhere on the court and they tend to be effective everywhere outside of three feet from the basket, ranking first or second in field goal percentage across the remaining distance ranges. Not surprisingly, these players take the most shots per game, average the most points, and have the highest USG%. Although they tend to take many shots, they also create for their teammates, evident by ranking second in AST%. Cluster six players tend to not be as effective on the defensive end, averaging the lowest TRB% and BLK%. Players in cluster six are well-rounded scorers. Some players commonly found in cluster six are Bradley Beal (8x), DeMar DeRozan (10x), and Jordan Clarkson (9x).

The next step in the analysis is to group teams together based on the cluster of the team's best player. As it was mentioned before, the best player on a team was determined by RPM. Given the playstyles of the six clusters, teams are inherently more likely to have their best player belong to a particular cluster. Therefore, it is unsurprising to see that 20 teams had their best player come from cluster one, 100 teams from cluster two, 65 teams from cluster three, 18 teams from cluster four, 21 teams from cluster five, and 76 teams from cluster six. Given these sample sizes, it is only appropriate to analyze teams whose best player is from cluster two, three, or six. These three clusters account for over 80% of all teams. Figure 3 presents boxplots of teams' Pythagorean win percentage grouped by best player cluster.

Figure 3. Distribution of Pythagorean win percentage grouped by best player cluster



Teams with a playmaker (cluster two) as their best player tend to fair the best out of the three groupings. Welch two-sample t-tests confirm this result. There was a significant difference in mean Pythagorean win percentage between cluster two teams (M = 55.7, SD = 14.6) and cluster three teams (M = 49.1, SD = 15.6) at the 1% level (p = 0.008) and a significant difference between cluster two teams and cluster six teams (M = 48.4, SD = 15.7) at the 1% level (p = 0.002). A Welch two-same t-test of mean Pythagorean win percentage between cluster three and cluster six teams did not yield a significant result.

Empirical Models

With an established understanding of the six clusters and teams grouped by the cluster of their best player, the last step in the analysis is to create linear regression models. Three models were specified, one each for cluster two teams (Model I), cluster three teams (Model II), and cluster six teams (Model III). The dependent variable for each model is Pythagorean win percentage and the independent variables are the total minutes played for the season by a team's best player (BPM) and the sum of minutes played for the season by each of the six clusters (denoted as C1M for cluster one minutes, C2M for cluster two, and so on).

Although the dependent variable has natural lower (0) and upper (100) bounds, predictions supplied by the regressions were within the possible range of values[3], so censoring was avoided. The following tables display summary statistics for the independent variables for the three subsections of teams. All variables are scaled in thousands of minutes.

Table 2. Cluster two teams variable summaries (in thousands of minutes)

Variable	Minimum	Median	Mean	Maximum	Std. Deviation
BPM	1.173	2.439	2.424	3.125	0.389
C1M	0	2.402	2.529	8.268	1.798
C2M	0	1.159	1.292	5.967	1.390
C3M	0	2.903	2.887	6.885	1.777
C4M	0	1.371	1.697	5.548	1.524
C5M	0	3.333	3.622	9.536	2.175
C6M	0	2.843	2.909	7.807	1.808

Table 3. Cluster three teams variable summaries (in thousands of minutes)

Variable	Minimum	Median	Mean	Maximum	Std. Deviation
BPM	1.510	2.284	2.253	3.030	0.360
C1M	0	2.396	2.502	7.757	1.773
C2M	0	2.266	2.317	5.215	1.242
C3M	0	2.041	2.028	5.946	1.495
C4M	0	0.549	0.989	4.203	1.187
C5M	0	2.780	3.022	7.340	1.740
C6M	0	4.022	3.973	10.271	2.316

Table 4. Cluster six teams variable summaries (in thousands of minutes)

Variable	Minimum	Median	Mean	Maximum	Std. Deviation
BPM	1.267	2.360	2.344	3.122	0.426
C1M	0	1.626	2.059	5.704	1.682
C2M	0	2.055	2.047	6.499	1.472
C3M	0	2.716	2.839	6.977	1.780
C4M	0	1.700	1.610	4.381	1.329
C5M	0	3.047	3.109	9.335	2.079
C6M	0	3.418	3.179	7.395	1.942

Ordinary Least Squares regressions were conducted by the specificities offered above. A Breusch-Pagan test revealed heteroskedasticity in Model III, so weighted least squares were applied to the equation as done by Yobero (2016). Table 5 presents the results of the three models.

The results of the models can best be analyzed by comparing the coefficients of significant variables. For example, in Model I, all six clusters have significant and positive coefficients, suggesting that all player types are beneficial to be put around playmakers. Individual coefficients represent the predicted percentage increase in Pythagorean win percentage by increasing minutes played by a particular cluster by 1000 minutes. However, what's important to note in terms of making predictions is the significant negative constant. While the constant is meaningless in this context (a team cannot have a negative win percentage), it is important in making predictions.

As an example, take C6M in Model I, as this is the lowest coefficient across all six clusters. In a normal NBA season, if a team does not play any overtime games, over 82 games they will total 19,680 minutes played[4]. Using this number as well as the average minutes played by the best player on cluster two teams, if the remaining minutes on a cluster two team were given to only cluster six players, the predicted Pythagorean win percentage is just 29.5%. For context, the lowest prediction for cluster two teams was 36.2%. Contrast this with giving all the remaining minutes to cluster five players, which has the greatest coefficient of the six clusters, and the number jumps to 89.5%. While both cases are unrealistic as lineups need a balance of players, and there are natural upper limits of playing time by cluster outlined by the maximum number of minutes that each cluster received in their relative grouping, they show the predicted difference of the impact of different clusters on team success. Therefore, the results highlight the opportunity costs of giving certain clusters minutes over others. If a cluster two team were to take 1000 minutes that were played by cluster six players and gave them to cluster five players instead, their predicted Pythagorean win percentage would increase by about 3.5%, all else equal. Table 6 below shows the relative rank of clusters for each model and subsection of teams.

Table 5. Regression results

	Model I (Cluster two teams)		Model II (Cluster three teams)		Model III (Cluster six teams)	
Variable	Coefficient	Prob.	Coefficient	Prob.	Coefficient	Prob.
Constant	-43.176	0.029**	-20.060	0.173	-35.781	0.026**
BPM	10.939	0.002***	16.676	0.002***	14.268	0.000***
C1M	5.011	0.000***	2.079	0.108	2.771	0.044**
C2M	4.015	0.004***	-0.739	0.602	1.911	0.159
C3M	5.776	0.000***	1.863	0.258	3.837	0.002***
C4M	4.549	0.002***	1.655	0.392	5.095	0.001***
C5M	6.154	0.000***	4.955	0.000***	3.973	0.000***
C6M	2.676	0.044**	1.940	0.042**	3.042	0.013**

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 6. Supporting clusters ranks by subsection of teams

	Cluster two teams		Cluster three teams		Cluster six teams	
Cluster	Coefficient	Rank	Coefficient	Rank	Coefficient	Rank
1	5.011	3	Insignificant	T-4	2.771	5
2	4.015	5	Insignificant	T-4	Insignificant	6
3	5.776	2	Insignificant	T-4	3.837	3
4	4.549	4	Insignificant	T-4	5.095	1
5	6.154	1	4.955	1	3.973	2
6	2.676	6	1.940	2	3.042	4

Discussion

Several takeaways related to complementary player types and predicted Pythagorean win percentage are provided by the regression models:

1. Unsurprisingly, all models suggest that the biggest impact comes from a team's best player. In all three models, BPM is significant with the largest positive coefficient.
2. The models suggest that cluster five players are universally the most impactful, ranking as the greatest coefficient of the six clusters for cluster two and three teams and the second highest for cluster six teams. This reflects the current state of the NBA and the need for role players to be able shoot three-pointers. Their catch and shoot ability and low USG% pairs nicely with any playstyle.
3. The relationship between cluster two and six players suggests that these two clusters are incompatible. Playing cluster six players on teams with a playmaker as the best player provides little value and cluster two players were the only ones not significant in Model III. This relationship makes sense as both types of players want the ball in their hands and having both on the court at once would likely lead to chemistry issues.
4. While cluster two and six players do not appear to be helpful to one another, cluster three players rank second for cluster two teams and third for cluster six teams, suggesting they are the most compatible as supporting players out of these three clusters.
5. Although cluster two teams tend to have greater success than cluster three and six teams, as supporting players, cluster two players do not seem to be helpful. Minutes played by this cluster are insignificant in Models II and III and rank second to last in Model I.
6. Model II suggests that cluster three teams are the least dependent on the makeup of the roster. Instead, a recipe for success for these teams is availability of their best player, given by the large, positive coefficient for BPM, and having players that can shoot the ball well, evident by clusters five and six being the only significant variables.
7. The cluster with the most varied impact across the three models is cluster four. The cluster ranks first for cluster six teams, fourth for cluster two teams, and is insignificant for cluster three teams. The insignificance in Model II is not surprising, as playing a traditional big with a versatile big would likely clog up the area near the basket for both players and not provide much defensive versatility. Cluster six teams, however, welcome the defensive prowess of cluster four players, while cluster two teams seem to benefit more by the defensive abilities of cluster three and one players rather than those from cluster four.
8. Cluster one players only seem to provide value to cluster two teams, ranking second to last for cluster six teams and not being significant in Model II.

Conclusion

This study aimed to quantify the effects that different playstyles of NBA players have on a team's success. By first clustering NBA players based on a variety of statistics that describe the way they play, traditional player positions were redefined to more accurately account for what players do when they are on the court. All 300 teams over the 10 years of data were put into groups based on the playstyle of their best player. Ordinary Least Squares regressions were conducted to show what playstyles are best to put around different styles of players. The models suggest that depending on the playstyle of a team's best player, different playstyles are impactful on team success. Teams whose best player can be described as a playmaker have higher predicted Pythagorean win percentages when surrounded by perimeter shot takers, versatile and traditional bigs, and defensive specialists than when minutes are given to other playmakers and score-first players. When the best player on a team is a versatile big, the model suggests that only perimeter shot takers and well-rounded scorers impact success compared to other playstyles. Lastly, teams that have a well-rounded scorer as their best player are predicted to benefit most from traditional bigs, perimeter shot takers, and versatile bigs.

Notes

1. The author invites the reader to look at Coach and A.D. (nd) or other sources for a glossary of basketball terminology.
2. The 2019-2020 NBA season was shortened due to the COVID-19 pandemic. Games played by teams during this season ranged from 63 to 75. The games played minimum requirement was based on the number of games each individual team played. The 2020-2021 NBA season was shortened to 72 games. The games played minimum requirement was set at 36 for this season.
3. Predictions for Model I ranged from 36.2%-76.5%, Model II ranged from 19.1%-70.2%, Model III ranged from 20.9%-65.8%.
4. The 2022-2023 average minutes played per team was 19,828 due to a small number of games going to overtime.

References

Aschburner S (2022) Commissioner Adam Silver discusses league's 'positionless basketball' at Finals press conference. NBA https://www.nba.com/news/nba-commissioner-adam-silver-discusses-leagues-positionless-basketball-at-annual-finals-press-conference

Brems M (2017) A one-stop shop for Principal Component Analysis. Towards Data Science https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c

Coach and A.D. (nd) A glossary of basketball terms. https://coachad.com/glossary-of-basketball-terms/

Duman EA, Sennaroğlu B, Tuzkaya G (2021) A cluster analysis of basketball players for each of the five traditionally defined positions. Journal of Sports Engineering and Technology https://doi.org/10.1177/17543371211062064

ESPN (nd) NBA Real Plus-Minus - 2023-24. https://www.espn.com/nba/statistics/rpm/_/sort/DRPM

Kalman S, Bosch J (2020) NBA lineup analysis on clustered player tendencies: a new approach to the positions of basketball and modeling lineup efficiency of soft lineup aggregates. In 14th annual MIT Sloan Sports Analytics Conference

McMahan I (2018) How (and why) position-less lineups have taken over the NBA playoffs. https://www.theguardian.com/sport/blog/2018/may/01/how-and-why-position-less-lineups-have-taken-over-the-nba-playoffs

Medium (2018) Understanding k-means clustering in machine learning. Towards Data Science https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

NBA (nd) Basketball positions. Jr NBA https://jr.nba.com/basketball-positions/

Osken C, Onay C (2022) Predicting the winning team in basketball: a novel approach. Heliyon https://www.cell.com/heliyon/pdf/S2405-8440(22)03477-6.pdf

Patel R (2017) Clustering professional basketball players by performance (Master's thesis, University of California Los Angeles) https://www.proquest.com/docview/1989767866?fromopenview=true&pq-origsite=gscholar

Shea S (nd) The 3-point revolution. https://shottracker.com/articles/the-3-point-revolution

Sports Reference Glossary (nd) Basketball Reference https://www.basketball-reference.com/about/glossary.html

Tsai P (2017) A metallurgical scientist's approach to predicting NBA team success. https://towardsdatascience.com/a-metallurgical-scientists-approach-to-predicting-nba-team-success-4bfa7b2bd6a7

Yobero C (2016) Methods for detecting and resolving heteroskedasticity. RPubs https://rpubs.com/cyobero/187387

Zhang S, Lorenzo A, Gómez MA, Mateus N, Gonçalves B, Sampaio J (2018) Clustering performances in the NBA according to players' anthropometric attributes and playing experience. Journal of Sports Sciences 36(22):2511-2520 https://doi.org/10.1080/02640414.2018.1466493

Ziller T (2017) The center position in the NBA is dying. https://www.sbnation.com/2017/7/17/15981266/nba-center-position-bad-free-agents

Appendix

Appendix: Variable descriptions, some quoted from basketballreference.com

Variable	
TRB%	An estimate of the percentage of available rebounds a player grabbed while they were on the floor
AST%	An estimate of the percentage of teammate field goals a player assisted while they were on the floor
STL%	An estimate of the percentage of opponent possessions that end with a steal by the player while they were on the floor
BLK%	An estimate of the percentage of opponent two-point field goal attempts blocked by the player while they were on the floor
TOV%	An estimate of turnovers committed per 100 plays.
SDSH	standard deviation of the proportion of shots taken from each distance range (derived)
FG% 0-3ft	Field goal percentage on shots between zero and three feet from the basket
FG% 3-10ft	Field goal percentage on shots between three and 10 feet from the basket
FG% 10-16ft	Field goal percentage on shots between 10 and 16 feet from the basket
FG% 16ft-3pt	Field goal percentage on shots between 16 feet from the basket and the three-point line
3P%	Three-point percentage
% 2P Ast'd	Percentage of two-point shots made that were assisted
% 3P Ast'd	Percentage of three-point shots made that were assisted
FT Rate	Number of free throw attempts per field goal attempt
PF/G	Personal fouls per game
Min/G	Minutes played per game
PTS/G	Points scored per game
TRB/G	Total rebounds per game
AST/G	Assists per game
STL/G	Steals per game
BLK/G	Blocks per game
FGA/G	Field goal attempts per game
FG%	Field goal percentage
3PA/G	Three-point attempts per game
FTA/G	Free throw attempts per game
USG%	An estimate of the percentage of team plays used by a player while they were on the floor.
Avg. Dist	Average shot distance (in feet)
% Shots 0-3ft	Proportion of shots taken between zero and three feet from the basket
% Shots 3-10ft	Proportion of shots taken between three and 10 feet from the basket
% Shots 10-16ft	Proportion of shots taken between 10 and 16 feet from the basket
% Shots 16ft-3pt	Proportion of shots taken between 16 feet from the basket and the three-point line
% Shots 3P	Proportion of shots taken from beyond the three-point line
% FG Ast'd	Percentage of shots made that were assisted (derived)

The Effects of COVID-19 and the Increasing Relationship between Individual Investor Sentiment, Cryptocurrencies, and the U.S. Market

Juan A. Rodriguez-Nieto and Kseniia Eremina, Drury University

Abstract

This research analyzes the effect of the COVID-19 pandemic on the relationship between individual (retail) investor sentiment, cryptocurrencies, and the U.S. stock market. Using the DCC-GARCH model, the dynamic conditional correlations between Bitcoin, Ethereum, Individual Investor Sentiment, VIX, and the U.S. stock market before and during the pandemic are obtained. The DCC-GARCH model is used to identify the increasing relationship between individual investor sentiment and the U.S. Stock market performance. This paper sheds light on the growing significance and influence of individual (retail) investors in the U.S. stock and cryptocurrency markets, with the COVID-19 pandemic catalyzing this accelerated influence.

JEL Codes: G01 • F65
Keywords: Cryptocurrencies, COVID-19, Stock Market, Retail Investor

Introduction

In 2008, the first cryptocurrency, Bitcoin (BTC), was introduced to the world by Satoshi Nakamoto (2008). Bitcoin is a purely peer-to-peer version of electronic cash that allows online payments to be sent directly from one party to another without needing a trusted third party (financial institution); thus, the proposed system of electronic transactions relies on something other than trust. Bitcoin's price is very volatile, creating significant financial risk. In 2010, the market value of Bitcoin was less than 5 cents, while in 2021, it reached its all-time high of $68,789 before closing at $64,305.94 (Marr, 2022).

Ethereum (ETH) was introduced in 2015 by Vitalik Buterin. It is another platform using blockchain technology, but unlike Bitcoin, Ethereum offers different methods of exchange, including cryptocurrency, smart contacts, and the Ethereum Virtual Machine (EVM). In July 2015, the Ethereum token launched at $0.43; in 2021, it reached its all-time high above $4,500 during the bull cycle (Marr, 2018).

However, the new coronavirus disease (COVID-19) started in December 2019 in Wuhan, China. The virus quickly spread around the world, causing the declaration of a pandemic. As of 16 March 2023, 760,360,956 cases have been confirmed, with 6,873,477 deaths reported to The World Health Organization (WHO, nd). The COVID-19 pandemic has affected markets around the world in a short time. Stocks, funds, and other market instruments have dropped prices sharply, which has hurt the economy.

During COVID-19, a new phenomenon led by individual (retail) investors attracted the media's and scholars' attention. While professional and institutional investors are considered to behave as rational investors, retail investors act by impulse and seem to be coordinating strategies by sharing information on social media platforms such as Reddit. The introduction and adoption by retail investors of affordable and easy-to-use fintech trading platforms during this period, paired with coordinated efforts using social media, resulted in the "meme stock" phenomenon in early 2021 with the short squeeze of GameStop, AMC, and the Dogecoin bubble of 2021.

This paper contributes to the literature by assessing the impact of COVID-19 on the strengthening relationship between retail investor sentiment, Bitcoin, Ethereum, and the U.S. stock market. The sample is divided into two periods: Pre-COVID-19 and COVID-19, and then proceed into a more profound analysis by breaking the data sample into three periods: Pre-COVID-19, the height of the pandemic and after the height of the pandemic.

The remainder of this paper has the following structure: The following section gives an overview of the existing literature. Section three describes the data used and provides a summary of descriptive statistics. In the fourth section, theoretical and empirical econometric models are explained. Section five presents the results of the different econometric models. Finally, section six concludes.

Literature Review

To better understand the topic and identify how cryptocurrencies and the U.S. market were affected by the COVID-19 disease, other research investigating the effects of investor sentiment on cryptocurrencies and the stock market before, during, and after the pandemic is reviewed. Knowing how global financial and crypto markets reacted to COVID-19 is crucial to investors and beneficiaries as traders need to know what changes should be made in their portfolios to continue generating profits even in times of uncertainty.

Several studies focus on Bitcoin's performance before and during the COVID-19 pandemic and whether cryptocurrency can be used as a hedge against the stock market. Xu (2022) attempts to examine the dynamic conditional volatility correlation between Bitcoin and stock markets before and after the COVID-19 outbreak. The CoinDesk website is used to obtain Bitcoin's price data series with a daily frequency from March 11, 2019, to March 12, 2021, while data of stock closing price covering six countries (developed countries: the United States, the United Kingdom; developing countries: Japan, China, India, and Brazil) with a daily frequency is taken from Investing. The researcher finds that after COVID-19, the volatility relation is strengthened and keeps a positive relation most of the time. In addition, the developed countries' stock markets have a relatively weak relationship after COVID, while the stock markets of developing countries have a relatively weak relationship except for Brazil. Overall, the results underscore that Bitcoin is not a good hedge against the stock market as the optimal allocation of Bitcoin is not high.

Athari and Hung (2022) explore the time-frequency return connectedness of the four most relevant asset classes, such as equity (novel proxy of the S&P500), digital asset (novel proxy of the S&P Cryptocurrency MegaCAP Index), commodity (novel proxy of the S&P Goldman Sachs Commodity Index), and fixed income (novel proxy of the S&P Global Developed Sovereign Bond Index). The daily data is obtained from the DataStream for the period from 02/28/2017 to 09/30/2021, which is divided into two subperiods: the pre-COVID-19 pandemic period from 02/28/2017 to 03/10/2020 and the COVID-19 pandemic period from 03/11/2020 to 09/30/2021. First, the authors find strong relationships at different frequencies during the COVID-19 pandemic, which differs from the pre-COVID-19 period. Second, they also identify a significant causal link between the variables during the COVID compared to pre-COVID, indicating the lack of hedging opportunities.

The above papers find strong volatility relationships during and after the COVID-19 pandemic, which differs from the pre-COVID-19 period. In addition, they all state that Bitcoin is not a good hedge. In this research, a positive relationship strengthened during the COVID-19 outbreak and the height of the pandemic.

Marobhe (2021) examines the sensitivity of cryptocurrency returns and several stock market indexes to the coronavirus disease 2019 (COVID-19). Marobhe applies the Bayesian structural vector autoregression to explore the phenomenon in Bitcoin, Ethereum, and Litecoin during the pandemic period from 2nd January 2020 to 30th June 2021. The findings indicate that all three cryptocurrencies experienced major adverse return shocks during the first wave of COVID-19. However, they recovered by April 2020 and remained resistant to further COVID-19 panic shocks. As for major stock indices, S&P 500, FTSE 100, and SSE Composite, they were susceptible throughout all the two waves of COVID-19. The results provide evidence to support the hypothesis that cryptocurrency is a safe haven during the coronavirus pandemic.

The time-varying correlations between six cryptocurrencies and S&P 500 index markets using a copula-ADCC-EGARCH model are examined by Tiwari et al. (2019) to investigate the hedging role of cryptocurrency against the risk of stock returns. The cryptocurrency data (Ripple, Dash, Stellar, Litecoin, Ethereum, and Bitcoin) are extracted from the Coindesk Price Index, while daily S&P 500 index prices are taken from DataStream. Time spans from August 7, 2015, to June 15, 2018. The authors find very low time-varying correlations, which are close to zero, that indicate that cryptocurrency serves as a hedge asset against the risk of the S&P 500 stock market. They also investigate that Litecoin is the most effective hedge investment against the S&P 500 stock market risk.

In contrast to Tiwari et al. (2019), herein is identified increased conditional correlations between Bitcoin and Ethereum due to the pandemic, reaching 0.8084 after the height of the pandemic. The following four publications observe the impact of different investor sentiments on the cryptocurrency and stock market before COVID-19, during COVID-19, and during the liquidity crisis of 2008-2009.

Güler (2021) studies the impact of investor sentiment on Bitcoin returns as well as conditional volatility during the COVID-19 outbreak by using three investor-sentiment proxies, such as the Bitcoin trading volume, the Crypto Fear & Greed Index, and the America Association of Individual Investor Index. The author concludes that both rational and irrational investor sentiments impact Bitcoin return, meaning that the Bitcoin market is affected and driven by rational investors, emotions, and noise traders. They identify that excellent news has more impact on Bitcoin prices than bad news, which can be referred to as a fear of missing out (FOMO) behavior of speculative and irrational investors.

The effect of several measures of Twitter-based sentiment on cryptocurrencies during the COVID-19 pandemic is studied by Kyriazis et al. (2022). The period since the beginning of the COVID-19 pandemic, starting from 1st January 2020 and ending on 25th July 2021, is observed. To detect causality, the authors use eight Twitter-derived uncertainty measures. In addition, daily data based on the ten largest cryptocurrencies (Bitcoin, Ethereum, Binance Coin, Cardano, Ripple, Dogecoin, Bitcoin Cash, Litecoin, Ethereum Classic, and Stellar) by market capitalization during the examined periods are applied for the empirical examination. Researchers find that Twitter-derived sentiment measures cannot explain the identified volatilities in low nominally priced cryptocurrencies, even in intensely distressed periods such as the COVID-19 pandemic, considered a major international financial crisis. They also conclude that investors willing to hedge their portfolios from the effects of the pandemic would benefit by investing in low nominally priced yet highly capitalized cryptocurrencies.

Koutmos (2022) attempts to quantify a robust sentiment-return relation using a bootstrapped quantile regression procedure, which he strives to achieve empirically and theoretically. A proxy for daily investor sentiment is constructed using a unique data set of intraday *buy* and *sell* orders from 2015 to 2020. Koutmos finds that rising sentiment is linked to price increases, while declining sentiment is related to price decreases. In this research, the author also reconciles the observations in models seeking to explain Bitcoin prices. This may arise because Bitcoin prices undergo regime shifts, or conventional regression models tend to focus on the mean of the distribution of Bitcoin price.

Huerta et al. (2016) examine the relationship between investor sentiment and REIT returns and volatility from December 2001 to February 2013 with a focus on the REIT liquidity crisis of 2008-2009. The REIT index and investor sentiment data are extracted from Thomson Reuters' Datastream. In addition to proxy investor sentiment, survey-based weekly sentiment measures are constituted from the American Association of Individual Investors (AAII) and Investor's Intelligence (II). Huerta et al. observe that investor sentiment significantly impacts REIT returns – institutional investor sentiment is realized to have a more considerable impact on REIT returns and volatility than individual investor sentiment. The authors find that institutions and individual investor sentiment have a positive and statistically significant effect on volatility. Researchers also point out that during the REIT liquidity crisis, institutional investor sentiment was a vital factor affecting excess returns during the crisis, while individual investor sentiment was not essential.

Ustalar et al. (2022) analyze the volatility spread between the cryptocurrency and global stock markets considering the COVID-19 pandemic. The analysis of volatility transmissions between cryptocurrency and the stock market was performed by applying the Constant Conditional Correlation Multivariate GARCH (CCC-GARCH) model. The time spans from 1st December 2019 to 1st July 2022, and daily closing prices are used for the analysis. The authors use Bitcoin to represent the cryptocurrency market and S&P 500, FTSE 100, SSEC, and NIKKEI indices for the global stock markets. The authors find a positive correlation between Bitcoin and the stock market indices. The estimated conditional correlation parameters are evaluated higher for the S&P 500 index and lower for the SSEC index than other stock market indices. Ustalar et al. conclude that risk and information transmission exists between cryptocurrency and global stock markets.

Data and Summary Statistics

The data used in this research are from the Federal Reserve Economic Data (includes CDOE Volatility Index [VIX], S&P 500 Index, Coinbase Bitcoin, and Coinbase Ethereum). Sentiment data is obtained from the American Association of Individual Investors (AAII, nd), which since 1987 provides insights into the options of individual investors by asking them their thoughts on where the market is heading in the next six months (AAII, nd). Weekly data is extracted starting from 06/02/2016 to 01/19/2023, capturing a total of 347 weeks. Data are reported in returns (S&P 500, Bitcoin, and Ethereum) and first differences (Sentiment and VIX). STATA and Excel are used to obtain dynamic conditional correlations between all variables and determine any significant changes during the defined periods.

The World Health Organization declared on 11 March 2020 that COVID-19 was a 'global pandemic'. Effective on June 12, 2022, the CDC rescinded the order requiring people to show a negative COVID-19 test result or documentation of recovery from COVID-19 before boarding a flight to the United States (CDC, 2022). Those dates are used to determine the three periods: pre-COVID-19, the height of the pandemic, and after the height of the pandemic.

Figures 1 to 3 represent Bitcoin, Ethereum, and S&P500 stock prices from 06/02/2016 to 01/19/2023. All exhibit that after the WHO declaration of the novel coronavirus (COVID-19) outbreak as a global pandemic (11 March 2020), the prices of both cryptocurrencies as well as S&P500, which is taken as a proxy for the U.S. stock market, dropped significantly. Bitcoin and Ethereum prices decreased by almost 11%, while S&P500 fell by 9%, with a continuous decline lasting more than a month.

However, cryptocurrency prices started fluctuating more as uncertainty about inflation and the emergence of a new COVID-19 variant, Omicron, continued to spook investors. By the end of 2022, the 'crypto winter' began, and Bitcoin dropped below $20,000; Ethereum price plummeted to $944 in March 2022 due to the crypto crush flowing the Russian-Ukrainian conflict. As for S&P500, the COVID-19 pandemic and the subsequent recessions caused it to plummet nearly 20% to $3,970.63. According to Senior Investment Strategy Director at the U.S. Bank Wealth Management, Rob Haworth, the 2022 market downturn could be caused by the rising level of uncertainly for investors, which incurred because of high inflation and the economic fallout from Russia's invasion of Ukraine (U.S. Bank, 2021).

Figure 1. Bitcoin Prices from 6/2/2016 to 1/19/2023

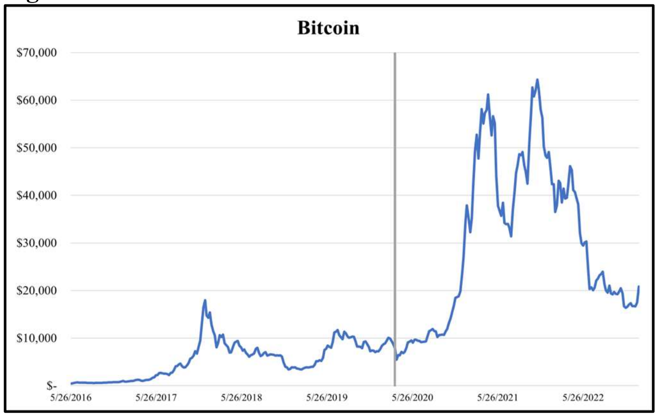


Figure 2. Ethereum Prices from 6/2/2016 to 1/19/2023

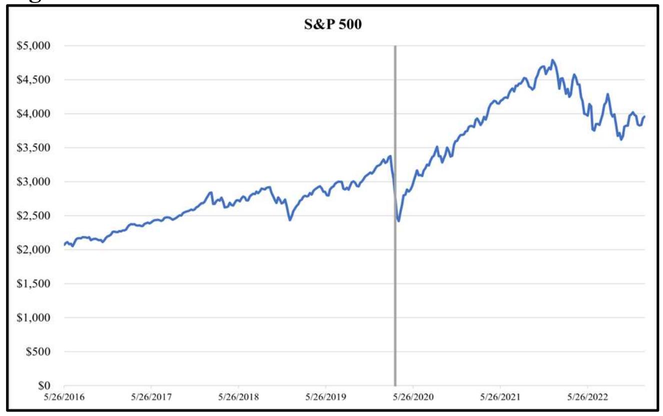
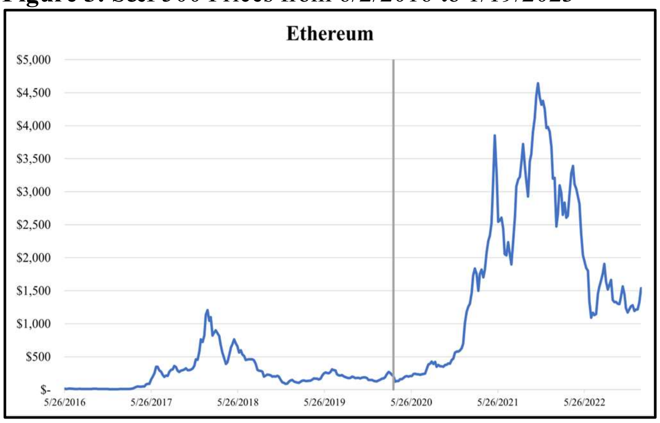


Figure 3. S&P500 Prices from 6/2/2016 to 1/19/2023



Note: Vertical line notes the beginning of the COVID-19 pandemic according to The World Health Organization (WHO)

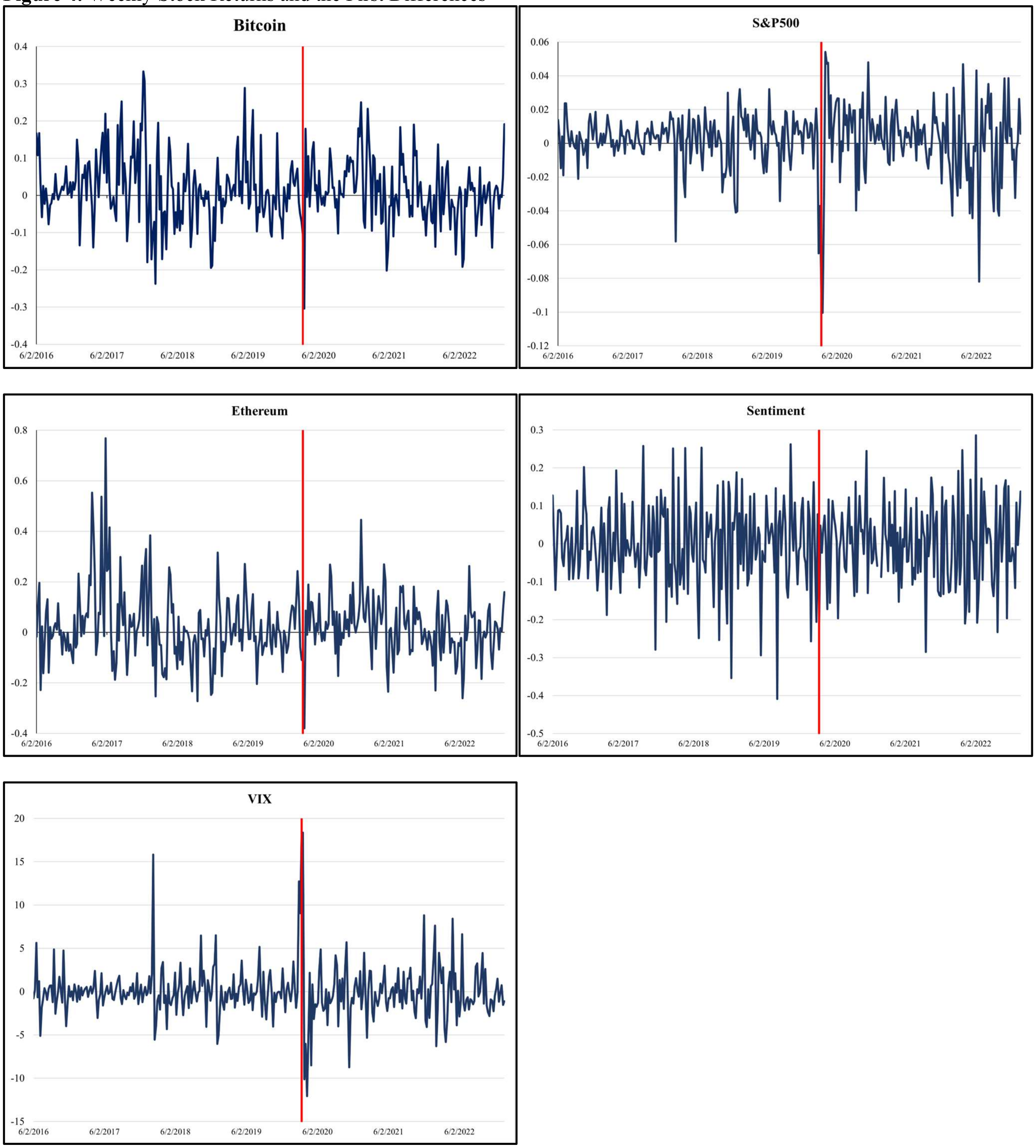
Table 1 summarizes data from 347 weeks of Bitcoin, Ethereum, and S&P500 prices. The statistics table shows that the minimum Bitcoin price was $447.65, while it soared to $64,305.94. As for Ethereum, it was as low as $7.41 and peaked at $4,643.50. S&P500 from June 2016 to January 2023 looked more stable, with the lowest being $2,048.73 and the highest hitting $4,787.33. The standard deviation is considered to better understand the price volatility. Bitcoin shows higher price volatility than Ethereum, being $16,643.20 and $1,129.97, respectively. The price of the S&P500 was less volatile, with a standard deviation of $757.32.

Table 1. Summary Statistics Table

	Bitcoin	Ethereum	S&P500
Mean	$16,024.26	$902.97	$3,181.21
Median	$9,132.74	$309.51	$2,923.77
Minimum	$447.65	$7.41	$2,048.73
Maximum	$64,305.94	$4,643.50	$4,787.33
Standard Deviation	$16,643.20	$1,129.97	$757.32
Observations (in weeks)	347	347	347

Figure 4 presents the weekly returns for Bitcoin, Ethereum, and the S&P500 index and the first differences for Individual Investor Sentiment and VIX during the sample period. Volatility for all variables increased dramatically around the pandemic declaration on March 11, 2020. After the COVID-19 pandemic, both cryptocurrencies, Bitcoin and Ethereum, aligned with traditional markets, hinting at contagion, since the coronavirus disease created a high degree of uncertainty and volatility across the global financial markets that prompted investors to reduce their exposure to risky assets, including cryptocurrencies.

Figure 4. Weekly Stock Returns and the First Differences



Note: Vertical line notes the beginning of the COVID-19 pandemic according to The World Health Organization (WHO)

Theoretical Model

The GARCH model is an autoregressive moving average model based on the weighted average of past squared residuals first introduced by Tim Bollerslev (1986). In 2002, the DCC-GARCH model was proposed by Robert Engel (2002), and the coefficients of the standardized residuals can be estimated using this model.

The DCC-GARCH(1,1) model is defined by:

$$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1}^2) + \beta(\sigma_{t-1}^2) + \lambda_2 S_{t-1} \quad (1)$$

Where α and β designate the GARCH coefficient, while λ_2 denotes the effect of investor sentiment on the conditional volatility.

In this research, the DCC-GARCH model is used to determine contagion by identifying whether increased dynamic conditional correlations between Bitcoin, Ethereum, Individual Investor Sentiment, the U.S. market volatility (VIX), and the U.S. stock market occurred during the COVID-19 outbreak.

The DCC-GARCH model measures the pairwise dynamic correlations between Bitcoin, Ethereum, S&P500, VIX, and Individual Investor Sentiment. The model is used as follows:

$$r_{Bit=} \gamma_0 + \gamma_1 r_{t-1}^{Bit} + \gamma_2 r_{t-1}^{Ethereum} + \gamma_3 r_{t-1}^{S\&P500} + \gamma_3 r_{t-1}^{VIX} + \gamma_3 r_{t-1}^{AAII} + \varepsilon_t \quad (2)$$

Two empirical regression models are built to analyze the conditional correlation dynamics before and during COVID-19 and then break the COVID-19 period into two, namely: during and after the height of the pandemic.
The models are the following:

$$\hat{\rho}_{ij,t} = \lambda_0 + \lambda_1 DV1_t + \epsilon_t, \qquad \text{for } i \neq j \quad (3)$$

$$\hat{\rho}_{ij,t} = \lambda_0 + \lambda_1 DV2_t + \lambda_2 DV3_t + \epsilon_t, \qquad \text{for } i \neq j \quad (4)$$

The Pre-COVID-19 period starts on 06/02/2016 and ends on 03/12/2020. For the first model, the COVID-19 period spans from 03/12/2022 to 01/19/2023. For the second model, the height of the pandemic period includes data from 03/12/2020 to 06/12/2022 and after the height of the pandemic period starting on 06/16/2022 and ending on 01/19/2023.

The dependent variable $\hat{\rho}_{ij,t}$ represents the forecast conditional correlation by the DCC-GARCH and indicates the strength of the relationship and its direction between pairs *i* and *j* at time *t*. It is measured on a scale from -1 to +1.

Equation 3 includes the dummy variable $DV1_t$, which represents the COVID-19 period starting on March 11, 2020. In equation 4, $DV2_t$ is the dummy variable for the height of the pandemic that spans from March 11, 2020, to June 11, 2022, and $DV3_t$ represents the period after the height of the pandemic starting June 12, 20202 until January 19, 2023. These dummy variables take the value of one (1) if the observation is included in the period and zero (0) if it is not. Using equations 3 and 4, the dynamic conditional correlation coefficients are regressed on the dummy variables and capture the effect of each COVID-19 period relative to the pre-COVID-19 period.

Following Rodriguez and Mollick (2021), it is acknowledged that incorporating lagged dependent and independent variables in the DDC–GARCH model implicitly accounts for all omitted variables and effectively minimizes the potential omitted variable bias. Additionally, the DCC-GARCH model without asymmetric extensions is selected, as the primary objective is to uncover the conditional correlation dynamics between the variables for subsequent analysis.

Results

This paper analyzes the effect of COVID-19 and individual investor sentiment on the relationship between cryptocurrencies and the U.S. stock market. Tables 2 to 5 show the results of the research analysis. The relationships between dependent and independent variables are statistically significant at 1%. Results in Table 2 represent regression analysis of correlation coefficients for the period observed (06/02/2016 – 01/19/2023). There is a strong correlation between Bitcoin and Ethereum and between S&P500 and Individual Investor Sentiment 0.6306 0.5039, respectively. As for the relationship between Bitcoin and S&P500, as well as Bitcoin and AAII, a positive correlation between the variables of 0.2885 and 0.1595 is observed, respectively. The U.S. market volatility index (VIX), as expected, is negatively correlated with all other variables, such as Bitcoin (-0.2804), Ethereum (-0.2391), S&P500 (-0.7823), and Individual Investor Sentiment (-0.4394).

Tables 3 to 5 present the results for analyzing conditional correlation coefficients for Bitcoin, considering different COVID-19 outbreaks. A positive relationship between Bitcoin and S&P500 is found before the start of the coronavirus disease that equals 0.1879, and during COVID-19, it increased to 0.3097. The correlations between Ethereum and S&P500, and Investor Sentiment and S&P500 were positive before the COVID-19 outbreak and rose as COVID-19 began, being 0.3210 and 0.6088, respectively.

As for the height of the pandemic and after the height of the pandemic periods, a strengthening correlation between Bitcoin and Ethereum is noted, which has a long-lasting effect that could be associated with the geopolitical situation, being 0.6890 during the height of the pandemic and 0.8084 after the height of the pandemic. A stronger relationship after the height of the pandemic between Ethereum and VIX and Bitcoin and VIX is found, which are 0.3874 and 0.4330, respectively. This indicates that investors likely consider Bitcoin and Ethereum similar investments, which is why their relationship has strengthened.

Table 2. Regression Analysis of Correlation Coefficients for the Period from 06/02/2016 to 01/19/2023

	Returns Bitcoin	Returns Ethereum	Returns S&P500	First Difference Sentiment	First Difference VIX
Returns_Bitcoin	1.00				
Returns_Ethereum	0.63***	1.00			
Returns_S&P500	0.29***	0.30***	1.00		
First Difference_Sentiment	0.16***	0.16***	0.50***	1.00	
First Difference_VIX	-0.28***	-0.24***	-0.78***	-0.44	1.00

* significant at the 1% level, ** significant at the 5% level, *** significant at the 10% level

Table 3. Regression Analysis of Conditional Correlation Coefficients – Bitcoin

	Bitcoin S&P500	Bitcoin Ethereum	Bitcoin Sentiment	Bitcoin VIX
Intercept	0.1879***		0.1239***	-0.2012***
COVID-19	0.1218***		0.1109***	-0.0940***
Adjusted R^2	0.2806		0.2365	0.2028
Intercept		0.5925***	0.1239***	-0.2012***
Pandemic Height		0.0955***	0.0894***	-0.0778***
After Pandemic Height		0.1194***	0.1901***	-0.1540***
Adjusted R^2		0.1762	0.2927	0.2405
Observations	347	347	347	347

* significant at the 1% level, ** significant at the 5% level, *** significant at the 10% level

Table 4. Regression Analysis of Conditional Correlation Coefficients – Ethereum

	Ethereum S&P500	Ethereum Sentiment	Ethereum VIX
Intercept	0.2141***	0.1606***	-0.1921***
COVID-19	0.1069***	0.0837***	-0.0704***
Adjusted R^2	0.2435	0.1794	0.1161
Intercept			-0.1921***
Pandemic Height			-0.0502***
After Pandemic Height			-0.1451***
Adjusted R^2			0.1777
Observations	347	347	347

* significant at the 1% level, ** significant at the 5% level, *** significant at the 10% level

Table 5. Regression Analysis of Conditional Correlation Coefficients – Sentiment

	Sentiment S&P500	Sentiment VIX
Intercept	0.5460***	-0.4503***
COVID-19	0.0628***	-0.0473***
Adjusted R^2	0.1188	0.0873
Intercept	0.5460***	-0.4503***
Pandemic Height	0.0576***	-0.0451***
After Pandemic Height	0.0821***	-0.0553***
Adjusted R^2	0.1217	0.0859
Observations	347	347

* significant at the 1% level, ** significant at the 5% level, *** significant at the 10% level

Conclusion

This paper analyzes the effect of COVID-19 on the relationship between individual/retail investor sentiment, Bitcoin, Ethereum, and the U.S. stock market. The Dynamic Conditional Correlation GARCH (DCC-GARCH) model obtains the pairwise dynamic correlations between cryptocurrencies (Bitcoin and Ethereum), S&P500, VIX, and AAII. The model provides the analysis of the data from 06/02/2016 to 01/19/2023.

First, increased correlations are identified during periods of market instability, making it hard for investors to diversify their portfolios. This research shows positive and increasing correlations between individual investors and all asset classes during COVID-19. These correlations continue to grow after the height of the pandemic.

Second, increasing easy access to social media and online investment training platforms like Reddit allows investors to share their knowledge, experience, and strategies. Due to this way of communication, individual investor herding is rising. The correlation between retail investor sentiment and these asset classes is rising with retail investors investing in similar asset classes.

Third, the relationship between Bitcoin and Ethereum has strengthened after the height of the COVID-19 pandemic, meaning that investors will likely treat these cryptocurrencies as similar investments. Furthermore, increased demand among retail investors for Bitcoin and Ethereum during COVID-19 was observed. This is because the pandemic created significant volatility in the stock market, forcing investors to choose other asset classes that perform better during time of uncertainty.

The results show that the relationship between Bitcoin and Ethereum, Bitcoin and S&P500, and Bitcoin and Individual Investor Sentiment increased during the pandemic. Furthermore, the correlation between mentioned pairs continues to grow even after the height of the pandemic. This can have a long-lasting effect associated with the Russian-Ukrainian conflict.

Acknowledgments

The anonymous reviewers and the editor of the Journal of the Academy of Economics and Finance are thanked for their feedback as well as the Academy of Economics and Finance, the Breech College of Business at Drury University, and the Ph.D. Project for their support.

References

American Association of Individual Investors (nd) The AAII Investor Sentiment Survey. Retrieved April 16, 2023 from https://www.aaii.com/sentimentsurvey

Athari SA, Hung NT (2022) Time–Frequency return co-movement among asset classes around the covid-19 outbreak: portfolio implications. Journal of Economics and Finance 46(4):736–756. https://doi.org/10.1007/s12197-022-09594-8

Atıcı Ustalar S, Ayar E, Şanlısoy S (2022) The volatility transmission between cryptocurrency and global stock market indices: case of Covid-19 period. İzmir İktisat Dergisi 37(2):443-459 DOI: 10.24988/ije.1034580

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31(3): 307–327

Centers for Disease Control and Prevention (2022, June 10) CDC Rescinds Order Requiring Negative Pre-Departure COVID-19 Test Before Flight to the US. Centers for Disease Control and Prevention. Retrieved April 16, 2023, from https://www.cdc.gov/media/releases/2022/s0610-COVID-19-test.html

Engle R (2002) Dynamic conditional correlation. Journal of Business & Economic Statistics 20(3):339–350

Güler D (2021) The impact of investor sentiment on bitcoin returns and conditional volatilities during the era of Covid-19. Journal of Behavioral Finance 1–14 https://doi.org/10.1080/15427560.2021.1975285

Huerta D, Egly PV, Escobari D (2016) The liquidity crisis, investor sentiment, and reit returns and volatility. Journal of Real Estate Portfolio Management 22(1):47–62

Koutmos D (2022) Investor sentiment and bitcoin prices. Review of Quantitative Finance and Accounting 60(1):1–29.

Kyriazis N, Papadamou S, Tzeremes P, Corbet S (2022) The differential influence of social media sentiment on cryptocurrency returns and volatility during COVID-19. The Quarterly Review of Economics and Finance 89: 307-317

Marohbe MI (2022) Cryptocurrency as a safe haven for investment portfolios amid Covid-19 panic cases of Bitcoin, Ethereum and Litecoin. China Finance Review International 12(1):51–68. https://doi.org/10.1108/cfri-09-2021-0187

Marr B (2018, March 20) Blockchain: A very short history of ethereum everyone should read. Forbes. Retrieved April 16, 2023, from https://www.forbes.com/sites/bernardmarr/2018/02/02/blockchain-a-very-short-history-of-ethereum-everyone-should-read/

Marr B (2022, October 12) A short history of bitcoin and crypto currency everyone should read. Forbes. Retrieved April 16, 2023, from https://www.forbes.com/sites/bernardmarr/2017/12/06/a-short-history-of-bitcoin-and-crypto-currency-everyone-should-read/

Nakamoto S (2008) Bitcoin: A Peer-to-Peer Electronic Cash System
Rodriguez-Nieto JA, Mollick AV (2021) The US financial crisis, market volatility, credit risk and stock returns in the Americas. Financial Markets and Portfolio Management 35:225–254
Tiwari AK, Raheem ID, Kang SH (2019) Time-Varying dynamic conditional correlation between stock and cryptocurrency markets using the copula-ADCC-EGARCH Model. Physica A: Statistical Mechanics and Its Applications: 535, 122295. https://doi.org/10.1016/j.physa.2019.122295
US Bank (2021, December 20) How far will the market correction go? Retrieved April 16, 2023, from https://www.usbank.com/investing/financial-perspectives/market-news/is-a-market-correction-coming.html
World Health Organization (nd) Coronavirus Disease (covid-19). World Health Organization. Retrieved April 14, 2023 from https://www.who.int/emergencies/diseases/novel-coronavirus-2019
Xu C (2022) The dynamic relation between bitcoin volatility and stock volatility: DCC-GARCH approach. Proceedings of the International Conference on Information Economy, Data Modeling and Cloud Computing, ICIDC

Financial Benchmarking: Financial Metrics across U.S. Industries

Carlos Trejo-Pech, University of Tennessee at Knoxville

Abstract

This study analyzes the financial performance of non-financial publicly traded firms across 16 industries over two decades. The analysis covers 25 commonly used financial metrics grouped into four categories: accruals-based ratios, risk and risk-adjusted profitability metrics, cash flow vs. accruals earnings, and market prices. The results can be used for benchmarking purposes in research and teaching. In addition, the analysis identifies sectors ranked as high and low financial performers by alternative financial metrics. Overall, the study finds a close relationship between stock returns and free cash flow and, to a less extent, between stock returns and profits and risk-adjusted profits.

JEL Codes: G10, G30
Keywords: Financial performance, financial benchmarks, financial metrics, ROI decomposition

Introduction

Financial analysis, conducted by evaluating an array of financial ratios and economic metrics (referred to as financial metrics in this study), helps support informed recommendations and decisions by managers, creditors, security analysts, and investors. In addition, financial analysts support their recommendations and decisions with benchmarking (Eklund et al. 2003). Thus, research analyzing financial metrics, for benchmarking, in specific industries is expected. For instance, Vogel and Graham examined the financial performance of a group of airports. Singh and Schmidgall (2002) investigated the most commonly used financial metrics in the lodging industry. Bouras et al. (2014) evaluated the effects of diversification on the financial performance of grocery stores. Schaufele and Sparling (2011), Katchova and Enlow (2013), Jackson and Singh (2015), and Trejo-Pech et al. (2023) investigated different research questions employing financial ratios mainly in the food and beverage industry.

With a few exceptions, the abovementioned studies employ a small set of financial metrics to characterize industries regarding financial performance, thus providing valuable information for intra-industry benchmarking. However, those studies do not benchmark the financial metrics of the industry they focus on with other sectors. This study analyzes the financial performance of U.S. publicly traded firms across 16 industries over two decades. The evaluation covers 25 financial metrics, including four broad categories: accruals-based financial ratios, risk and risk-adjusted profitability metrics, cash flow vs. accruals earnings, and market prices. The authors believe this is the first study to comprehensively evaluate the U.S. market by industry. There are at least two sources of financial metrics by industry, Damodaran Online (Damodaran, 2023) and the Financial Ratios Suite by Wharton Research Data Services (WRDS, 2022). While these sources provide a long list of readily available financial metrics across industries, they are databases, and Financial Ratios Suite is only available for subscribers. In addition, these sources do not compute risk and risk-adjusted profitability metrics, as done in this study, which is relevant from an economic perspective. This article uses financial statements and market data from two WRDS databases to compute and analyze 25 financial metrics across 16 industries using Fama and French's industries classification (Fama and French, 2022).

This study contributes to the financial management literature by computing and analyzing benchmark metrics from the last two decades and by providing a framework to evaluate the financial performance of a group of firms or industries. Notably, the study analyzes profitability, which has been reported among the most important financial ratios categories (Gibson, 1987; Singh and Schmidgall, 2002) by (a) decomposing the return on investment (ROI) into a margin, asset efficiency, and leverage component according to the DuPont decomposition, (b) evaluating other popular financial ratios within the margin, asset efficiency, and leverage categories, (c) evaluating risk-adjusted profitability and risk metrics, and (d) relating profitability and risk-adjusted profitability metrics with cash flow-based metrics and stock returns. Lastly, this article identifies industries that consistently rank at the top and the bottom of the U.S. market according to a few relevant metrics. The results of this study can also be helpful for teaching purposes. For example, some financial management textbooks (Brigham and Houston, 2019) use hypothetical benchmarks when covering financial analysis. Using actual financial measures will arguably improve the learning process because learning financial analysis according to the actual figures of the U.S. market is more interesting for students and will likely provide them with a reference for subsequent analysis of specific firms.

One challenge and limitation of this study is selecting a manageable yet informative and helpful set of financial metrics to study. This is because many financial ratios are used in practice; no consensus exists regarding which financial metrics are the most important, and financial ratios contain overlapping information (Chen and Shimerda, 1981). Therefore, this article selects financial metrics guided by prior research (Matsumoto, Shivaswamy, and Hoban, 1998; Gibson, 1987; Singh and Schmidgall, 2002; Chen and Shimerda, 1981; Lewellen, 2004; Delen, Kuzey, and Uyar, 2013; Trejo-Pech, Noguera, and White, 2015). Those studies have identified financial ratios with predictive power to forecast stock returns, financial ratios that are more important to surveyed financial analysts and managers, and financial ratios most employed by equity analysts for their stock recommendations. An inspection of financial ratios currently used by S&P's Capital IQ NetAdvantage (S&P Net Advantage, 2023) also guided the selection of metrics. A second limitation of this study is deciding how narrow or broad-defined industries are for this analysis. Fama and French classify firms into 5, 10, 12, 17, 30, 38, 48, and 49 sectors. The 17 industries classification was arbitrarily chosen in the middle of the spectrum, recognizing that this presents a limitation.

Data and Methods

Databases and Industries

Data at the firm level are from two databases maintained by Wharton Research Data Services (WRDS, 2022): COMPUSTAT | North America | Fundamental Annual and Beta Suite by WRDS. The data were downloaded and processed in April 2022. Based on firms' Standard Industry Classification (SIC) codes, companies were grouped into industries according to Fama and French's (F&F) 17 industries classification (Fama and French, 2022). Given that the financial ratios of nonfinancial firms are not comparable to those of financial firms (i.e., banks, insurance companies, and other financials), the financial industry was removed from the sample. In addition, firms with COMPUSTAT ISO Country Code other than "USA" were removed from the database to focus on U.S. firms. A secondary source used in this study is the U.S. Department of Treasury website, where risk-free rates were obtained. Thus, this analysis covers financial metrics of nonfinancial American-based publicly traded firms from 2000 to 2021.

Categories of Financial Metrics

This paper covers four broad categories of financial metrics: (1) accruals-based financial ratios, (2) risk and risk-adjusted profitability, (3) cash flow and accrual earnings per share, and (4) market prices. This section describes what ratios are analyzed, the corresponding formulas, and why particular ratios were selected.

Except for one financial ratio (cash flow from operations to interest), the first category has accruals-based financial ratios. The analysis starts with profitability ratios return on investment (ROI), return on equity (ROE), and return on assets (ROA). ROI is calculated by,

$$ROI = \frac{NOPAT}{Capital} = \frac{EBIT \times (1 - tax)}{Avg.(D + E)}, \tag{1}$$

where NOPAT is net operating profit after taxes, EBIT is COMPUSTAT "earnings before interest and taxes" or operating income, tax is calculated by dividing "income taxes" by "pretax income," total debt (D) is "total debt including current," and equity (E) is "stockholders equity." (Quotation marks are used in this section to indicate variables from COMPUSTAT. However, quotation marks are not used below.) Average (Avg.) is the simple average of the value in the current and previous year. Average values of balance sheet items are used to calculate profitability ratios following WRDS Industry Financial Ratio and S&P Capital IQ Net Advantage (WRDS Research Team, 2016; S&P Net Advantage, 2023).

ROE is calculated by dividing net income (loss) by average equity, and ROA is net income (loss) by average assets. While ROI, ROE, and ROA are proxies of profitability used by equity analysts (Trejo-Pech, Noguera, and White, 2015), ROI is preferred in this study because capital (debt plus equity) captures the two sources of financing with cost, arguably making it a better proxy from the perspective of investors (Schill, 2017). Thus, after analyzing ROI, ROA, and ROE, ROI is further decomposed into three of its drivers following the DuPont decomposition approach. The spirit of the DuPont model, which has been modified and adjusted many times to capture better investment returns (Gupta, Synn, and Upton, 2019), is to breakdown ROE into a margin (i.e., net income to sales), operational efficiency (sales to assets), and leverage (assets to equity) component. Following the same framework, this study decomposes ROI into NOPAT margin, assets turnover, and a non-interest-bearing leverage component (equation (2)). Since assets to capital (Asset_Cap), the last term in equation (2), measures how much a firm uses non-interest-bearing liabilities or free-financing such as suppliers trade credit, this article refers to Asset_Cap as non-interest-bearing leverage. In passing, notice from equation (2) that while Asset_Cap is not a commonly used leverage ratio, it is the product of two typical leverage ratios.

$$ROI = \frac{NOPAT}{Capital} = \frac{NOPAT}{Sales} \times \frac{Sales}{Assets} \times \frac{Assets}{Equity} \times \frac{Equity}{Capital} = \frac{NOPAT}{Sales} \times \frac{Sales}{Assets} \times \frac{Assets}{Capital} \quad (2)$$

Despite the ROI decomposition being a parsimonious model providing insights on performance related to margins, asset efficiency, and leverage, other financial ratios within each category are widespread, according to research surveying financial managers, equity analysts, recommendation reports, or financial management textbooks (Matsumoto, Shivaswamy, and Hoban, 1998; Gibson, 1987; Singh and Schmidgall, 2002; Chen and Shimerda, 1981; Trejo-Pech, Noguera, and White, 2015). Therefore, this study also analyzes popular financial ratios within the margin, asset efficiency, and leverage/solvency groups.

Regarding margins, the EBIT margin (EBIT%), defined as EBIT to revenue, the EBITDA margin (EBITDA%)—EBITDA to revenue—, and the net income margin (NI%)—net income (loss) to revenue—are included. On asset efficiency, the property, plant, and equipment turnover (PPE_TO), accounts receivable turnover (AR_TO), and inventory turnover (INV_TO) are studied. PPE_TO, AR_TO, and INV_TO are calculated as revenue divided by year-to-year average PPE, revenue divided by average year-to-year accounts receivable, and cost of goods sold divided by average year-to-year inventory.

Lastly, within the leverage and solvency category, the study includes the following widely used financial ratios (S&P Net Advantage 2023): debt to capital (Debt_Cap), debt to assets (Debt_Asset), debt to EBITDA (D_EBITDA), EBIT to interest expenses (EBIT_Int), and cash flow from operations to interest expenses (CFO_Int). Cash flow from operation (CFO) was calculated according to equation (3), with ΔWC representing working capital in *t* minus working capital in *t-1*. Working capital is calculated as current assets minus current liabilities. *D&A* is depreciation and amortization.

$$CFO = Net\ Income\ (loss) + D\&A - \Delta WC \quad (3)$$

Table 1 summarizes the accruals-based financial ratios included in the study and the sequence in which they are evaluated across industries. Table 1 also contains the other categories of ratios, discussed next.

Table 1. Financial metrics by categories

1. Accruals-based financial ratios
1.1. Profitability: ROI, ROE, and ROA
1.2. Drivers of ROI: NOPAT% (margin), ATO (efficiency), and Asset_Cap (non-interest-bearing leverage)
1.2.1. Margins: EBIT%, EBITDA%, and NI%
1.2.2. Efficiency: ATO, PPE_TO, AR_TO, and INV_TO
1.2.3. Leverage and Solvency: Debt_Cap, Debt_Asset, D_EBITDA, EBIT_Int, and CFO_Int
2. Risk and risk-adjusted profitability
2.1. EVA = ROI-WACC
2.2. WACC and Beta
3. Cash flow vs accrual earnings
3.1. FCFPS and EPS
4. Market prices
4.1. Stock returns
4.2. Market multiples: PE and FV to EBITDA

The second broad category of metrics is risk and risk-adjusted profitability, which includes a firm's systematic risk (beta), the weighted average cost of capital (WACC), and economic value added margin (EVA). Beta is estimated using Beta Suite by WRDS, as in Trejo-Pech et al. (2021). Specifically, Beta Suite was specified to estimate 60-month rolling regressions whenever available in this database. For firms with less than 60 monthly returns, the software was restricted to estimating betas only if the firms had stock returns for at least 36 months. This 3-5 years length window is commonly used in practice (Brotherson et al. 2013).

Betas are analyzed separately and used to estimate the expected cost of equity (*e*) by applying the Capital Asset Price Model (CAPM) (Sharpe, 1964). CAPM, equation (4), is practitioners' most popular capital asset pricing model (Graham and Harvey, 2018).

$$e_{i,t} = rf_t + \beta_{i,t} \times rp \quad (4)$$

The firm's cost of equity is estimated every year by proxying the risk-free rate (*rf*) with the simple average of the daily annualized rates for the long-term composite bond (US Department of the Treasury, 2022) and using a 6.5% market risk premium (*rp*) (Brotherson et al., 2013; C. Trejo-Pech, DeLong, and Johansson, 2023). Estimated *e's* are inputs to calculate the WACC, according to Equation (5):

$$WACC = \left\{\frac{D}{D+MCap} \times d \times (1 - tax)\right\} + \left\{\frac{MCap}{D+MCap} \times e\right\}, \tag{5}$$

where total debt (*D*) has been defined above, market capitalization (*MCap*) is calculated as the closing price at the end of the year times the number of shares outstanding, and the cost of debt (*d*) is calculated by dividing interest and related expenses by total debt, *tax* is the income tax rate as defined before, and *e* is from equation (4).

Lastly in this category, EVA margin (EVA) is the difference between ROI and WACC. EVA is a risk-adjusted profitability metric because it subtracts the firm's opportunity cost of capital rate from the after-tax (but before interest) profit rate. EVA is a proxy of a firm's *economic* profit rate or residual income, which economic theory predicts to be zero (i.e., not abnormal returns) in the long term.

This study's third category of financial metrics compares cash flow and earnings accrual metrics free cash flow per share (FCFPS) and earnings per share (EPS). FCFPS is calculated by,

$$FCFPS = \frac{NI + D\&A - CAPEX - \Delta WC}{Shares}, \tag{6}$$

where *CAPEX* is capital expenditures and *Shares* is COMPUSTAT "shares used to calculate earnings per share." EPS is COMPUSTAT "earnings per share diluted no extraordinary."

The final category, market prices, includes stock returns and market multiples. Stock returns are calculated yearly as the natural logarithm of the quotient of the closing price in *t* and the closing price in *t-1*. Two market multiples are analyzed, price to earnings (PE) and firm value to EBITDA (FV to EBITDA). PE is obtained by dividing the closing price by EPS. FV to EBITDA is calculated with equation (7), with all variables already defined, except *Cash* which is balance sheet item cash plus cash equivalents.

$$FV\ to\ EBITDA = \frac{MCap + D - Cash}{EBITDA}, \tag{7}$$

All observations with zero or negative total assets, stockholders' equity, capital, or revenue were removed from the database (WRDS Research Team, 2016; Cocco and Volpin, 2013). In addition, all variables in the study were winsorized every year at the 1% and 99% levels to remove outliers (WRDS Research Team, 2016). Table 1 contains all financial metrics presented in this section.

Analysis Approach

Median values are more appropriate than means or other statistics for financial ratio analysis by industry (WRDS Research Team 2016). In addition, with a few exceptions (beta, WACC, PE, FV to EBITDA, and FCFPS), the financial metrics analyzed in this article have skewed distributions (i.e., skewness absolute value higher than 1.0). Therefore, this study used median values to characterize financial performance by industry. The analysis is done by computing and comparing the financial metrics across the 16 F&F nonfinancial industries over the 22 years from 2000 to 2021.

Industry comparisons are made by ranking industries according to performance and focusing on high and low performers. High-performing industries are defined as those ranked in the top five of the 16 industries. Similarly, low-performing industries are those ranked in the bottom five. Industries ranking in the middle are referred to as moderate performers. In addition to rankings, statistical tests are conducted to compare the performance of an industry relative to the U.S. market performance for selected financial metrics. Specifically, for industry *i,* the performance of the U.S. market *i* is the median value of the financial metric pooling all industries except industry *i*. Thus, the market's financial performance to which each industry is compared varies.

The difference between the financial performance of industry *i* (Ind_i) and the market *i* (Mkt_i), Ind_i-Mkt_i, is statistically tested using median equalities tests with quantile regression (Conroy, 2012; C. Trejo-Pech, DeLong, and Johansson, 2023). The null hypothesis of median equalities between two groups (Equation 8) is tested with the STATA procedure qreg with the option vce(robust) for robust standard errors.

$$H_0{:}\, Ind_{i,median\ financial\ metric\ j} = Mkt_{i,median\ financial\ metric\ j} \tag{8}$$

Results and Discussion

Profitability

Table 2 shows profitability ratios ROI, ROE, and ROA by industry. Each financial ratio in Table 2 has three columns. The first column shows the rankings by profitability ratios—from largest to smallest—listed in the second column. The third column, Ind_i-Mkt_i, has the difference between the profitability ratio of industry *i* and market *i*. (The profitability ratio of market *i* is calculated for each industry (*i*) as the median of the profitability of all industries except industry *i*. Thus, the market's profitability to which each industry is compared varies across industries.) Using quantile regression, the Ind_i-Mkt_i column also indicates whether the difference between medians is statistically significant (Equation 8). In this paper, Ind_i-Mkt_i fis calculated or selected metrics only (i.e., profitability, risk-adjusted profitability, and stock returns). Finally, the bottom of Table 2 shows the median of the profitability ratio pooling all industries (i.e., the U.S. market hereafter). As explained in the Analysis Approach section, industries ranked in the top five are defined as high performers, industries in the bottom five are low performers, and industries in the middle are moderate performers.

Table 2. ROA, ROE, and ROI by industry (median values, 2000 to 2021)

	ROI			ROE			ROA		
	Rank	Ratio	Ind_i-Mkt_i	Rank	Ratio	Ind_i-Mkt_i	Rank	Ratio	Ind_i-Mkt_i
CLTHS	1	0.101	0.037***	4	0.114	0.037***	1	0.052	0.028***
FABPR	2	0.096	0.032***	2	0.117	0.039***	6	0.039	0.015***
CARS	3	0.095	0.031***	1	0.130	0.006*	8	0.030	0.006*
RTAIL	4	0.091	0.028***	5	0.112	0.036***	3	0.041	0.018***
CHEMS	5	0.087	0.023***	6	0.108	0.031***	7	0.034	0.01***
DURBL	6	0.084	0.02***	9	0.094	0.016***	10	0.027	0.002
TRANS	7	0.083	0.02***	3	0.115	0.039***	2	0.043	0.019***
FOOD	8	0.078	0.014***	7	0.104	0.027***	5	0.040	0.016***
CNSTR	9	0.076	0.012***	8	0.094	0.017***	4	0.040	0.017***
STEEL	10	0.073	0.008**	13	0.065	-0.012**	12	0.023	-0.001
CNSUM	11	0.065	0	11	0.074	-0.003	16	-0.006	-0.03***
UTILS	12	0.064	0	10	0.092	0.018***	9	0.029	0.008***
MACHN	13	0.063	-0.001	12	0.069	-0.01***	11	0.025	0.001
MINES	14	0.054	-0.01**	14	0.056	-0.022***	14	0.019	-0.004**
OTHER	15	0.051	-0.021***	15	0.049	-0.04***	15	0.005	-0.025***
OIL	16	0.039	-0.027***	16	0.041	-0.037***	13	0.020	-0.003**
US Market		0.064			0.077			0.024	

Industries: Textiles, apparel & footwear (CLTHS), fabricated products (FABPR), automobiles (CARS), retail stores (RTAIL), chemicals (CHEMS), consumer durables (DURBL), transportation (TRANS), food (FOOD), construction (CNSTR), steel works (STEEL), Drugs, Soap, Perfumes, Tobacco (CNSUM), utilities (UTILS), machinery and business equipment (MACHN), mining and minerals (MINES), others (OTHER), and oil and petroleum products (OIL).
Industries definitions by SIC codes available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

The U.S. market median ROI, ROE, and ROA are 6.4% (varying from 3.9% to 10.1%), 7.7% (4.1% to 13%), and 2.4% (-0.6% to 5.2%). As expected, the three proxies of profitability have high correlation coefficients, ranging between 0.65 and 0.81 in the pooled data. Due to the high correlations, the rankings by industry across profitability ratios in Table 2 are generally consistent. High (low) performing industries in terms of ROI are, with a few exceptions, also high (low) performers in terms of the two other profitability ratios. The CLTHS, FABPR, CARS, RTAIL, and CHEMS industries are high ROI performers (complete industry definition provided in Table 2). For example, the CARS industry yields a 9.5% ROI and 3.1 percentual points above the market's ROI. According to quantile regression, the difference between this industry's ROI and the market's is statistically significant at 1%. In contrast, OIL, OTHER, MINES, MACHN, and UTILS are low ROI sectors. Furthermore, MINES, OTHER, and OIL yielded ROI, ROA, and ROE that are statistically significant below the market (i.e., negative values in the 'Ind_i-Mkt_i' column). The following section further evaluates ROI.

Drivers of Return on Investment

Table 3 decomposes ROI into NOPAT margin (NOPAT%), assets turnover (ATO), and Assets to Capital (Asset_Cap) following the DuPont decomposition framework (instead of decomposing ROE, ROI is broken down). The ROI decomposition identifies three drivers of profitability: profit margin, asset efficiency, and non-interest-bearing leverage. The industries in Table 3 are ranked in terms of ROI, as in Table 2, to facilitate continuity in the discussion. As expected, the correlations amongst the ROI components are relatively low because they belong to different categories of ratios. For example, the correlation coefficient of the pooled data between NOPAT% and Asset_Cap is 0.048. The correlation coefficient between NOPAT% and ATO is 0.179, and between Asset_Cap and ATO is 0.33. Those low correlations make the rankings by one ROI driver differ from rankings by another component.

Table 3. ROI decomposed by industry (median values from 2000 to 2021)

	NOPAT%		**ATO**		**Asset_Cap**	
	Rank	Ratio	Rank	Ratio	Rank	Ratio
CLTHS	5	0.059	2	1.37x	10	1.34x
FABPR	2	0.071	8	1.02x	4	1.42x
CARS	6	0.050	4	1.25x	1	1.52x
RTAIL	16	0.037	1	1.78x	2	1.45x
CHEMS	3	0.065	10	0.87x	8	1.37x
DURBL	7	0.049	7	1.18x	6	1.39x
TRANS	4	0.061	9	0.91x	5	1.39x
FOOD	9	0.044	3	1.26x	7	1.37x
CNSTR	12	0.043	6	1.19x	11	1.34x
STEEL	15	0.038	5	1.20x	9	1.37x
CNSUM	10	0.044	13	0.71x	13	1.32x
UTILS	1	0.162	16	0.25x	3	1.42x
MACHN	8	0.048	11	0.86x	14	1.31x
MINES	11	0.043	14	0.60x	15	1.30x
OTHER	14	0.040	12	0.76x	12	1.33x
OIL	13	0.042	15	0.36x	16	1.26x
US Market		0.045		0.88x		1.34x

Refer to Table 2 for industries definitions.

The high ROI in CLTHS is primarily driven by high efficiency (ranked 2nd) and high margin (5th) (non-interest-bearing leverage is moderate in this industry). ROI's leading drivers in FABPR are high margin (2nd) and high non-interest-bearing leverage (4th) industry. ROI in CARS is primarily driven by high non-interest-bearing leverage (1st), high assets efficiency (4th), and moderate-to-high margin (6th). RTAIL's high ROI is driven by asset efficiency (1st) and non-interest-bearing leverage (2nd). Notably, RTAIL, a high ROI industry, is ranked at the bottom (16th) regarding margin. Lastly, high ROI in CHEMS is mainly driven by high margins (3rd).

Among the high ROI industries, it is worth emphasizing stylized facts of CARS and RTAIL. First, the automobile industry is the only one consistently ranked high in terms of the three drivers of profitability, profit margin, asset efficiency, and non-interest-bearing leverage. This result is remarkable for an industry and is consistent with research that considers the automobile industry exerts high market power (Kwoka, 1984; Grieco, Murry, and Yurukoglu, 2021). Second, the retail industry is usually stylized as a low-margin industry, which the results herein confirm, given that the profit margin in this industry is ranked at the bottom among the 16 sectors. However, emphasizing the low margins in this industry, as it is sometimes done in research reports, the business press, and textbooks (Diment, 2023; Kang, 2023; Saghir and Jönson, 2001), might be misleading because RTAIL is a highly efficient industry in terms of asset turnover and non-interest-bearing leverage. This combination puts retailers among the top performers regarding ROI, ROE, and ROA (Table 2). Kroger and Walmart are prominent examples in this industry.

Among the low ROI industries, MINES, OTHER, and OIL are consistently ranked in the bottom five industries across all ROI drivers (except NOPAT%, ranked 11th, as moderate-to-low). This is consistent with the results in Table 2 that identified these sectors among the least profitable. Table 3 also shows that the utility industry is the least asset-efficient industry (ranked 16th) but, at the same time, has the highest profit margin and high non-interest-bearing leverage (3rd). This result highlights the characteristics expected in natural monopoly industries such as UTILS (Primeaux, 1979).

Other Margin, Leverage & Solvency, and Assets Efficiency Ratios

While Table 3 gives key drivers of profitability, other financial ratios within these categories are commonly used as proxies for margin, leverage, and asset efficiency performance. Mainly, the EBIT margin (EBIT%), EBITDA margin (EBITDA%), and net margin (NI%) are margin ratios commonly used in practice (S&P Net Advantage 2023; Trejo-Pech, Noguera, and White 2015). Regarding leverage and solvency ratios, the Debt to Capital (Debt_Cap), Debt to Assets (Debt_Asset), Debt to EBITDA (D_EBITDA), EBIT to Interest (EBIT_Int), and Cash flow to Interest (CFO_Int) are selected proxies of leverage and solvency. Lastly, in addition to the asset turnover ratio—discussed in the previous section—other turnover ratios include the property, plant, and equipment to sales (PPE turnover), the account receivables to sales (AR turnover), and the inventory to sales (INV turnover) ratios. These ratios are calculated for benchmarking and relevant results are discussed in this section.

Table 4. Selected margin proxies by industry (median values from 2000 to 2021)

	EBIT%		**EBITDA%**		**NI%**	
	Rank	Ratio	Rank	Ratio	Rank	Ratio
UTILS	1	0.168	1	0.274	1	0.081
FABPR	2	0.085	3	0.125	4	0.037
TRANS	3	0.074	4	0.123	2	0.040
CHEMS	4	0.073	6	0.118	5	0.032
CLTHS	5	0.073	7	0.098	3	0.037
OIL	6	0.059	2	0.214	8	0.023
DURBL	7	0.058	8	0.094	12	0.020
CNSTR	8	0.058	12	0.084	6	0.029
FOOD	9	0.056	9	0.091	7	0.028
CARS	10	0.055	10	0.089	11	0.021
STEEL	11	0.050	13	0.084	14	0.018
MACHN	12	0.048	11	0.086	9	0.022
RTAIL	13	0.045	15	0.077	10	0.021
MINES	14	0.044	5	0.121	13	0.018
OTHER	15	0.034	14	0.078	15	0.003
CNSUM	16	0.020	16	0.049	16	-0.011
US Market		0.054		0.097		0.022

Refer to Table 2 for industries definitions.

Table 4 provides alternative margin ratios to NOPAT%—discussed in the previous section—. As expected, the correlations among margin ratios are very high, particularly amongst NOPAT%, NI%, EBITDA%, and EBIT%, with correlation coefficients above 0.96. The correlation coefficient between GM% and the other margins is also high, between 0.78 and 0.79. Thus, except for GM%, these high correlations produce similar rankings regardless of the margin proxy chosen. This highlights that financial ratios within the same category provide redundant information due to overlapping, as has been reported (Chen and Shimerda, 1981). Reinforcing the result on NOPAT margin in the previous section, Table 4 shows that UTILS ranks first in profit margin in terms of NI%, EBITDA%, and EBIT%. Other high-profit margin industries are TRANS, CLTHS, and FABPR.

Table 5 gives proxies of leverage and coverage ratios. When leverage is measured as Debt to Capital, the aggregated U.S. market has 33 cents of debt for each dollar of capital (i.e., one-third of debt and two-thirds of equity). U.S. leverage, measured as debt to assets, is 17%. The Debt to EBITDA ratio, which adjusts leverage by a cash proxy, indicates the U.S. market owes 73 cents of total debt for every EBITDA dollar generated. UTILS, CHEMS, and TRANS are high levered industries across the three leverage ratios. Similarly, alternative leverage ratios consistently rank MACHN, CLTHS, DURBL, and OTHER as low-levered sectors.

Table 5 also gives coverage ratios. A high (strong) coverage ratio and high leverage suggest payment capacity to support such high leverage. This is the case of the transportation (TRANS) industry, ranked fourth by EBIT_Int and second by CFO_Int and highly levered. The rest of the high levered industries in Table 5 have moderate payment capacity. In the other extreme of leverage, the combination of low leverage and low coverage suggests that the low payment capacity drives the low leverage. This is the case of OTHER and DURBL. More interesting, CLTHS combines low leverage with high coverage ratio. (Ranked 15^{th}, 14^{th}, and 13^{th} in terms of Debt_Cap, Debt_Asset, and D_EBITDA, but ranked1^{st} and 3^{rd} in terms of EBIT_Int and CFO_Int.) This combination suggests a cushion to growth debt if investment opportunities arise in the textiles, apparel & footwear industry.

Regarding proxies of asset efficiency, Table 6 provides alternative turnover ratios. As expected, manufacturing industries such as CLTHS, MACHN, and CNSTR manage long-term assets more efficiently. In contrast, service-oriented industries like TRANS and UTILS are ranked at the bottom regarding PPE_TO. Regarding account receivables turnover, the first and second-ranked industries are RTAIL and FOOD. This may be explained by firms in these industries (e.g., grocery stores and restaurants) collecting sales almost immediately.

Table 5. Selected leverage and solvency proxies by industry (median values from 2000 to 2021)

	Debt_Cap		**Debt_Asset**		**D_EBITDA**		**EBIT_Int**		**CFO_Int**	
	Rank	Ratio	Rank	Ratio	Rank	Ratio	Rank	Ratio	Rank	Ratio
UTILS	1	51%	1	33%	1	3.76x	11	3.11x	8	3.32x
CHEMS	2	45%	4	26%	5	1.73x	8	3.43x	11	3.00x
CARS	3	43%	8	22%	9	1.30x	10	3.19x	9	3.32x
TRANS	4	42%	3	26%	2	1.88x	4	4.07x	2	4.66x
CNSTR	5	40%	7	24%	7	1.62x	7	3.87x	7	3.48x
OIL	6	39%	2	27%	8	1.39x	15	1.83x	4	4.00x
FABPR	7	37%	5	24%	3	1.87x	2	4.45x	6	3.64x
STEEL	8	36%	6	24%	4	1.85x	9	3.20x	13	2.54x
FOOD	9	35%	9	22%	6	1.66x	5	4.04x	5	3.74x
CNSUM	10	35%	15	13%	16	0.00x	14	1.95x	16	0.85x
RTAIL	11	34%	11	19%	12	1.03x	3	4.18x	1	5.05x
MINES	12	34%	10	22%	10	1.24x	13	2.05x	12	2.77x
OTHER	13	31%	13	14%	14	0.20x	16	1.74x	15	1.68x
DURBL	14	28%	12	18%	11	1.11x	12	2.92x	14	2.20x
CLTHS	15	21%	14	13%	13	0.72x	1	6.33x	3	4.01x
MACHN	16	21%	16	9%	15	0.20x	6	3.97x	10	3.20x
US Market		33%		17%		0.73x		2.83x		2.87x

Refer to Table 2 for industries definitions.

Risk and Risk-Adjusted Profitability

Table 7 provides a risk-adjusted profitability metric, the economic value added margin, EVA, which considers profitability and risk (i.e., EVA=ROI-WACC). Table 7 also breakdowns EVA into risk metrics WACC and beta. Like with profitability ratios (Table 2), differences between each industry EVA and market EVA (Ind_i-Mkt_i) and tests for statistical significance using quantile regression are tabulated.

Table 6. Selected asset efficiency proxies by industry (median values from 2000 to 2021)

	PPE_TO		**AR_TO**		**INV_TO**	
	Rank	Ratio	Rank	Ratio	Rank	Ratio
CLTHS	1	10.01x	7	7.66x	15	3.58x
DURBL	2	7.13x	11	6.80x	13	4.69x
OTHER	3	7.12x	14	6.34x	4	6.96x
MACHN	4	7.08x	16	6.10x	14	3.84x
CNSTR	5	7.01x	6	7.77x	9	5.69x
RTAIL	6	5.81x	1	55.77x	6	6.71x
CNSUM	7	5.53x	10	7.11x	16	3.06x
CARS	8	5.11x	9	7.19x	7	6.05x
FOOD	9	4.83x	2	11.26x	5	6.84x
FABPR	10	4.63x	12	6.78x	12	4.71x
STEEL	11	3.43x	5	8.02x	10	5.36x
CHEMS	12	3.05x	13	6.75x	11	4.90x
TRANS	13	2.85x	4	8.65x	2	15.61x
MINES	14	1.06x	3	9.68x	8	5.80x
OIL	15	0.52x	15	6.19x	1	19.68x
UTILS	16	0.52x	8	7.48x	3	13.41x
US Market		5.29x		6.97x		5.87x

Refer to Table 2 for industries definitions.

The EVA median value for the U.S. market during the 22-year evaluation is practically zero (-0.002). This result is aligned with microeconomics theory predicting that firms yield zero residual (economic) profits in the long term as they enter a steady-state equilibrium. In other words, at zero economic gains, firms in the market generate just enough profits to pay debt and equity holders the profits they expect according to their risk expectations captured by the CAPM and WACC. As expected, there is variation in EVA across industries. The high EVA industries are CNSUM, FOOD, CARS, RTAIL, and FABPR. For example, FOOD, with 2.2.% median EVA, generated 2.5 percentual points above the EVA median of the other 15 industries according to quantile regression. This difference is statistically significant at 1%. Moreover, three of these high EVA industries, RTAIL, CARS, and FABPR, are high ROI sectors (Table 2), meaning these industries also have a competitive WACC that keeps them as high-risk-adjusted profitability performers. In contrast, the low EVA industries are MACHN, OIL, STEEL, MINES, and OTHER. Except for STEEL, all these industries were also ranked low ROI performers in Table 2.

Table 7 also gives WACC and beta median values. Unlike EVA (and the previous metrics), WACC and beta are ranked from low to high because these metrics measure the cost of capital and risk, meaning that industries ranked at the top (bottom) are more (less) competitive. The UTILS industry has the lowest WACC with a 5.5% median value, followed by FOOD (6.9%) and MINES (8.3%). Except for the low WACCs in UTILS and FOOD and the highest WACC (10.8%) in MACHN, there is little variability of WACC across the rest of the industries, with most WACC values primarily concentrated between 8.3% (MINES) and 9.6% (STEEL). The overall WACC for the U.S. market is 9.1%, close to the stylized 10% WACC typically used in corporate finance textbooks.

Lastly, Table 7 ranks beta values across industries. As predicted by the portfolio and asset pricing theory, the beta for the U.S. market is about 1.0 (1.019 in Table 7). According to theory, the risk of a diversified portfolio containing all assets in the economy represents the benchmark to compare the risk of specific firms or groups of firms. By construction, this benchmark is 1.0. Table 7 also illustrates a close relation between beta and WACC rankings, emphasizing the beta's importance as a WACC driver. For instance, a low 0.339 beta in UTILS drives this industry's lowest WACC value of 5.5%. Lastly, the low betas and WACCs in FOOD and CNSUM position these industries as high-risk-adjusted profitability sectors.

To recap, Table 2 through Table 7 rank industries by accrual-based financial ratios primarily and stock-market (beta and WACC) metrics. The EVA results in Table 7 are notable because EVA adjusts profitability (ROI) for risk (WACC). The analysis identifies high and low-performing industries and discusses connections between several financial ratios commonly used in practice. The relationships between several financial metrics are more apparent for some sectors with extreme performance. The following section analyzes mainly cash flow vs. accrual profit and market multiples and stock returns.

Table 7. EVA, WACC, and beta by industry (median values from 2000 to 2021)

	EVA			**WACC**		**Beta**	
	Rank	Value	Ind_i-Mkt_i	Rank	Value	Rank	Value
CNSUM	1	0.049	0.052***	4	0.084	3	0.782
FOOD	2	0.022	0.025***	2	0.069	2	0.616
RTAIL	3	0.021	0.024***	8	0.086	6	0.957
CARS	4	0.019	0.022***	10	0.088	13	1.056
FABPR	5	0.019	0.022***	6	0.084	4	0.921
UTILS	6	0.014	0.016***	1	0.055	1	0.339
TRANS	7	0.013	0.016***	5	0.084	9	0.987
CHEMS	8	0.013	0.015***	7	0.086	8	0.969
CLTHS	9	0.012	0.014***	13	0.092	11	1.033
DURBL	10	0.006	0.008***	12	0.090	10	0.989
CNSTR	11	-0.002	0	11	0.088	12	1.041
OTHER	12	-0.007	-0.007***	14	0.092	14	1.071
MINES	13	-0.010	-0.008*	3	0.083	5	0.936
STEEL	14	-0.020	-0.018***	15	0.096	15	1.235
OIL	15	-0.022	-0.02***	9	0.086	7	0.967
MACHN	16	-0.028	-0.031***	16	0.108	16	1.333
US Market		-0.002			0.091		1.019

Refer to Table 2 for industries definitions.

Earnings and Free Cash Flow Per Share

Earnings per share (EPS) and free cash flow per share (FCFPS) are relevant metrics equity analysts analyze for their recommendations and used by financial managers (Matsumoto, Shivaswamy, and Hoban, 1998; Lewellen, 2004; Trejo-Pech, Noguera, and White, 2015). Furthermore, accruals-based EPS and cash flow-based FCFPS contain valuable information for investors, and both are among the most critical measures firms report to outsiders (Graham, Harvey, and Rajgopal 2006).

EPS and FCFPS are tabulated by industry in Table 8. While the match is not perfect between high FCFPS and EPS industries, there are coincidences among the rankings. FABPR, CLTHS, and CNSTR are high FCPFS and EPS industries. In addition, FOOD is a high FCFPS industry and a moderate-to-high EPS sector. Low performers coincide regardless of ranking.

OIL, UTILS, CNSUM, OTHER, and MINES are low FCFPS and EPS industries. In addition, three industries ranked high FCFPS are also ranked high in terms of profits (Table 2), and all low FCFPS industries are also low profitable.

In general, the results in Table 8 suggest a consistency between the performance of industries in terms of accrual profits and free cash flow. This is relevant because the ‘cash is king’ commonplace suggests a discrepancy between accrual profits and cash. However, Graham et al. (2006) state that while cash may be more important for private, new, and fast-growing companies, both metrics may be equally crucial for publicly traded and mature firms. The results herein seem to corroborate the coincidence between profits and cash-based metrics.

Stock Returns and Market Multiples

This final section focuses on stock returns and market multiples (Table 9). FABPR, CNSTR, CLTHS, STEEL, and FOOD yield high returns, while OIL, MINES, OTHER, CNSUM, and DURBL give low stock returns. Establishing ties between stock returns and financial ratios is challenging because stock returns capture investors’ interpretations of current and expected/future performance. However, results in this study provide clues about relationships between stock returns and financial ratios, prominently with free cash flow and, to less extent, with EPS and EVA. For instance, high stock return industries FABPR, CNSTR, CLTHS, and FOOD are also ranked as high FCFPS industries. Furthermore, the top three (FABPR, CNSTR, CLTHS) in terms of stock returns are also high EPS industries. FOOD is high in terms of stock returns, FCFPS, and EVA. The relationship between stock returns and these financial metrics is even more apparent for low-performing industries. OIL, MINES, and OTHER are low industries in terms of stock returns, FCFPS, EPS, and EVA. This study finds a closer relationship between stock returns and free cash flow across industries.

Table 8. Free cash flow per share and earnings per share by industry (median values from 2000 to 2021)

	FCFPS		**EPS**	
	Rank	Ratio	Rank	Ratio
FABPR	1	0.471	4	0.585
CLTHS	2	0.290	2	0.815
CHEMS	3	0.232	8	0.405
FOOD	4	0.196	6	0.465
CNSTR	5	0.182	3	0.660
TRANS	6	0.145	1	0.830
CARS	7	0.143	7	0.455
RTAIL	8	0.124	5	0.490
DURBL	9	0.096	11	0.120
STEEL	10	0.073	9	0.295
MACHN	11	0.001	10	0.130
MINES	12	-0.013	13	0.000
OTHER	13	-0.013	14	0.000
CNSUM	14	-0.041	15	0.000
UTILS	15	-0.422	16	0.000
OIL	16	-0.437	12	0.020
US Market		-0.003		0.060

Refer to Table 2 for industries definitions.

Table 9 (on the following page) also provides PE and FV to EBITDA. High-stock return industries are expected to be priced high and vice versa. Otherwise, buying and selling opportunities arise. Indeed, all low stock performers in Table 9 are or tend to be ranked as low or cheap industries. Regarding high stock return industries, FABPR, CNSTR, and FOOD are, simultaneously, high priced industries according to either of the market multiples. This is not the case for CLTHS, which has moderate-to-low market multiples, suggesting a buying opportunity.

Closing

This study analyzed the financial performance of non-financial publicly traded firms across 16 industries over two decades. The analysis covered 25 commonly used in-practice financial metrics grouped into four categories: accruals-based financial ratios, risk and risk-adjusted profitability metrics, cash flow vs. accruals earnings, and market prices. Results from this study can be used for benchmarking purposes in research and teaching. This article also contributes to the financial management literature by providing a framework to evaluate a firm financial performance employing a relatively large set of financial metrics and industries. Notably, the article analyzed profitability by (a) decomposing the return on investment into its key drivers, (b) evaluating an array of popular financial ratios, (c) evaluating risk-adjusted profitability and risk metrics, and (d) relating profitability and risk-adjusted profitability metrics with cash flow-based metrics and stock returns. In addition, the analysis identified industries with financial performance consistently ranking at the top and the bottom of the U.S. market. Overall, the study found a close relationship between stock returns and free cash flow and, to a less extent, between stock returns and profits and risk-adjusted profits.

Table 9. Stock returns, Price to Earning and Firm Value to EBITDA market multiples by industries (median values from 2000 to 2021)

	Stock Return			**PE**		**FV to EBITDA**	
	Rank	Return	Ind_i-Mkt_i	Rank	Ratio	Rank	Ratio
FABPR	1	0.069	0.036*	3	15.38x	3	11.03x
CNSTR	2	0.068	0.047***	7	13.15x	5	10.75x
CLTHS	3	0.059	0.027*	6	13.84x	13	9.44x
STEEL	4	0.058	0.017	12	9.65x	9	9.91x
FOOD	5	0.054	0.038***	2	16.80x	2	11.80x
TRANS	6	0.051	0.036***	5	14.89x	6	10.70x
UTILS	7	0.049	0.048***	1	17.24x	1	13.41x
CHEMS	8	0.047	0.026**	11	12.17x	7	10.35x
CARS	9	0.038	0.006	8	12.75x	8	10.23x
RTAIL	10	0.031	0.005	4	14.90x	11	9.67x
MACHN	11	0.027	-0.003	9	12.53x	4	10.89x
DURBL	12	0.026	-0.012	10	12.30x	10	9.70x
CNSUM	13	0.024	0.006	16	3.12x	16	7.49x
OTHER	14	0.011	-0.027***	15	5.79x	12	9.66x
MINES	15	0.008	-0.012	13	7.04x	14	8.80x
OIL	16	-0.017	-0.012	14	6.36x	15	7.98x
US Market		0.026			11.52x		10.15x

Refer to Table 2 for industries definitions.

Acknowledgements

Carlos Trejo-Pech acknowledges that this work was partially supported by the USDA National Institute of Food and Agriculture, Hatch Multi-State project 1020537.

References

Bouras A, Bouras D, Ajuzie E (2014) An empirical analysis of the effect of diversification on financial performance: The case of the US grocery store industry. Review of Applied Economics 10:69–81

Brigham EF, and Houston JF (2019) Fundamentals of Financial Management. 10th ed. Boston, MA: Cengage Learning Inc

Brotherson T, Eades K, Harris R, Higgins R (2013) Best practices in estimating the cost of capital: An update. Journal of Applied Finance 23(1):1–19

Chen K, Shimerda T (1981) An empirical analysis of useful financial ratios. Financial Management 10(1):51–60

Cocco JF, Volpin PF (2013) Corporate pension plans as takeover deterrents. Journal of Financial and Quantitative Analysis 48(4):1119–44. https://doi.org/10.1017/S0022109013000355

Conroy R (2012) What hypotheses do 'nonparametric' two-group tests actually test? The Stata Journal: Promoting Communications on Statistics and Stata 12(2):182–90. https://doi.org/10.1177/1536867X1201200202

Damodaran A (2023) Damodaran Online. NYU Stern School of Business, January 2023. https://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html#returns

Delen D, Kuzey C, Uyar A (2013) Measuring firm performance using financial ratios: A decision tree approach. Expert Systems with Applications 40(10):3970–83. https://doi.org/10.1016/j.eswa.2013.01.012

Diment D (2023) Supermarkets and Grocery Stores in the U.S. Industry Report 44511. IBISWorld Industry Report

Eklund T, Back B, Vanharanta H, Visa A (2003) Using the self-organizing map as a visualization tool in financial benchmarking. Information Visualization 2(3):171–81. https://doi.org/10.1057/palgrave.ivs.9500048

Fama E, French K (2022) Data Library Current Research Returns. Kenneth French's Website http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Gibson C (1987) How chartered financial analysts view financial ratios. Financial Analysts Journal 43(3):74–76

Graham J, Harvey C (2018) The equity risk premium in 2018. SSRN Paper ID 3151162. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3151162

Graham J, Harvey C, Rajgopal S (2006) Value destruction and financial reporting decisions. Fin Analysts Journal 62:27–39

Grieco P, Murry C, Yurukoglu A (2021) The evolution of market power in the US auto industry. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w29013

Gupta N, Synn W, Upton K (2019) Updating the 100-year-old DuPont analysis: Using ROME vs. ROE. Advances in Financial Education 45:1–14

Jackson L, Singh D (2015) Environmental rankings and financial performance: An analysis of firms in the US food and beverage supply chain. Tourism Management Perspectives 14:25–33. https://doi.org/10.1016/j.tmp.2015.02.002

Kang J (2023) A supermarket megamerger will redefine what you buy at the grocery store. The Wall Street Journal, March 11, 2023, sec. Business

Katchova AL, Enlow SL (2013) Financial performance of publicly-traded agribusinesses. Agricultural Finance Review 73(1):58–73. https://doi.org/10.1108/00021461311321311

Kwoka JE (1984) Market power and market change in the US automobile industry. The Journal of Industrial Economics 32(4):509–22. https://doi.org/10.2307/2098233

Lewellen J (2004) Predicting returns with financial ratios. Journal of Financial Economics 74(2):209–35

Matsumoto K, Shivaswamy M, Hoban J (1998) Security analysts' views of the financial ratios of manufacturers and retailers. Financial Practice and Education (July):44–55

Primeaux W (1979) Some Problems with Natural Monopolies. The Antitrust Bulletin (Spring):63–85

Saghir M, Jönson G (2001) Packaging handling evaluation methods in the grocery retail industry. Packaging Technology and Science 14(1):21–29 https://doi.org/10.1002/pts.523

Schaufele B, Sparling D (2011) Regulation and the financial performance of Canadian agribusinesses. Agricultural Finance Review 71(2):201–17 https://doi.org/10.1108/00021461111152573

Schill MJ (2017) Business valuation: Standard approaches and applications. Darden Business Publishing, University of Virginia UV6586 (Rev. Nov. 20, 2017):1–14

Sharpe W (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance 19(3):425–42. https://doi.org/10.2307/2977928

Singh A, Schmidgall R (2002) Analysis of financial ratios commonly used by lodging financial executives. Journal of Leisure Property 2(3):201–13

S&P Net Advantage (2023) Standard and Poor's Capital IQ. S&P Global Market Intelligence. May 2023. www.capitaliq.com.

Trejo-Pech C, DeLong K, Johansson R (2023) How Does the financial performance of sugar-using firms compare to other agribusinesses? An accounting and economic profit rates analysis. Agricultural Finance Review 83(3):1–25

Trejo-Pech C, Larson JA, English BC, Yu TE (2021) Discount rates and stochastic techno-economic analysis for a prospective pennycress (Thlaspi Arvense L.) Sustainable Aviation Fuel Supply Chain. Working Paper, Univ of Tennessee

Trejo-Pech C, Noguera M, White S (2015) Financial ratios used by equity analysts in Mexico and stock returns. Contaduría y Administración 60(3):578–92. https://doi.org/10.1016/j.cya.2015.02.001

US Department of the Treasury (2022) Daily Treasury Long-Term Rates. https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_long_term_rate&field_tdr_date_value=2000.

(2022) The Global Standard for Business Research. May 2022. https://wrds-www.wharton.upenn.edu/

Wharton Research Data Services (WRDS) Research Team (2016) WRDS Industry Financial Ratio, https://wrds-www.wharton.upenn.edu/pages/grid-items/financial-ratio-industry-level/

PARKER
COLLEGE OF BUSINESS
GEORGIA SOUTHERN UNIVERSITY

Department of Economics
Department of Finance
11935 Abercorn Street
Savannah, GA 31419