1-1-2017

The S&P 500© Is Not a Leading Indicator for US GDP Over Policy-Relevant Time Frames

William D. Campbell Sr.

Eastern Illinois University

This research is a product of the graduate program in Economics at Eastern Illinois University. Find out more about the program.

Recommended Citation
http://thekeep.eiu.edu/theses/2746

This Thesis is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.
Preserving, reproducing, and distributing thesis research is an important part of Booth Library’s responsibility to provide access to scholarship. In order to further this goal, Booth Library makes all graduate theses completed as part of a degree program at Eastern Illinois University available for personal study, research, and other not-for-profit educational purposes. Under 17 U.S.C. § 108, the library may reproduce and distribute a copy without infringing on copyright; however, professional courtesy dictates that permission be requested from the author before doing so.

Your signatures affirm the following:

- The graduate candidate is the author of this thesis.
- The graduate candidate retains the copyright and intellectual property rights associated with the original research, creative activity, and intellectual or artistic content of the thesis.
- The graduate candidate certifies her/his compliance with federal copyright law (Title 17 of the U. S. Code) and her/his right to authorize reproduction and distribution of all copyrighted materials included in this thesis.
- The graduate candidate in consultation with the faculty advisor grants Booth Library the non-exclusive, perpetual right to make copies of the thesis freely and publicly available without restriction, by means of any current or successive technology, including by not limited to photocopying, microfilm, digitization, or internet.
- The graduate candidate acknowledges that by depositing her/his thesis with Booth Library, her/his work is available for viewing by the public and may be borrowed through the library’s circulation and interlibrary loan departments, or accessed electronically.
- The graduate candidate waives the confidentiality provisions of the Family Educational Rights and Privacy Act (FERPA) (20 U. S. C. § 1232g; 34 CFR Part 99) with respect to the contents of the thesis and with respect to information concerning authorship of the thesis, including name and status as a student at Eastern Illinois University.

I have conferred with my graduate faculty advisor. My signature below indicates that I have read and agree with the above statements, and hereby give my permission to allow Booth Library to reproduce and distribute my thesis. My adviser’s signature indicates concurrence to reproduce and distribute the thesis.

Graduate Candidate Signature

Printed Name

M. A. - Economics

Graduate Degree Program

Faculty Advisor Signature

Printed Name

Date 6/1/2017

Please submit in duplicate.
THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

YEAR

I HEREBY RECOMMEND THAT THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

6/15/17
THESIS COMMITTEE CHAIR

6/15/17
DEPARTMENT/SCHOOL CHAIR
OR CHAIR'S DESIGNEE

THESIS COMMITTEE MEMBER

THESIS COMMITTEE MEMBER

THESIS COMMITTEE MEMBER

6-5-17
DATE

DATE

DATE

DATE
The S&P 500© Is Not a Leading Indicator for US GDP

Over Policy-Relevant Time Frames.

William D. Campbell, Sr.

Thesis: Master of Arts in Economics

Eastern Illinois University

May 4, 2017

Abstract:

Simply, we find S&P 500© returns are no longer a statistically meaningful leading indicator for growth in “real” per capita GDP. Technically, we find the SP500 fails to Granger cause real per capita GDP over policy relevant time frames while confirming nominal (non-inflation adjusted) SP500 quarterly returns continue to Granger cause one-period ahead quarterly growth in inflation adjusted (real) US Per-Capita GDP over very long (and perhaps less meaningful from fiscal and monetary perspectives) time frames. In addition, we identify a likely transition period when the SP500 switched from a leading indicator to a lagging indicator. Therefore, in keeping with the principal finding of this paper, we suggest great restraint is warranted when using equity market returns as a basis for economic policy.
1 Introduction

This paper first investigates the long-term relationship between US GDP growth and S&P 500 returns, including: (a) the determination of any sense of causality between US GDP and equity markets over very long time frames and (b) a critique of the common comparison of growth in inflation adjusted ("real") GDP in the USA to nominal equity index returns.

The second objective is our primary objective. We now correct for the major failures of other papers and re-examine the GDP / SP500 relationship over policy relevant (10-20 year) time frames. This re-examination also discusses the SP500’s transition from a leading to a lagging indicator.

The paper concludes by suggesting possible additional avenues for research regarding equity market returns and US GDP growth.

2 Background & Theory

The primary objective of this paper is to determine if there is statistical evidence for the notion that equity market returns lead GDP growth in the United States over policy relevant time frames. To phrase this question in somewhat more practical language, we ask if there is statistical evidence to support S&P 500’s status as an official "leading indicator" (Levanon 2011).
Our core assessment tools are: 1. Ensuring the two indices (GDP growth and S&P 500 returns) are on an equivalent return basis and 2. a technique known as Granger Causality. We apply the Granger Causality test to determine which way causality runs (if any).

Two established theories support the notion that a reasonably strong relationship should exist between equity market returns and GDP growth. Although we will not delve into the details or histories of the theories, it is relevant to note them here.

The first, the Efficient Market Hypothesis, essentially asserts “all known and expected information is embedded in the last price” (Fama, 1970). This theory is interesting because it could support statistical causality in either direction, although it is commonly used to support rGDP1 Granger causing stock market returns.

We also give a nod to the potential relevancy of the Rational Expectations Theory (Sargent, 2008), particularly in those situations where SP500 is found to Granger cause rGDP.

We ignore leverage in this paper. Admittedly, changes in credit availability to non-financial businesses can impact corporate profits which in turn impact GDP. We note, much has been written about how changes in available financial market leverage can have material impact on short-run SP500 returns and this aspect may be relevant to future

\(^1\) rGDP is this paper’s notation for “real GDP” as defined in the Common Acronyms section
research. We also recognize credit availability to consumers varies significantly over time. However, investigating the impact of the changes in available credit (to consumers, to financial assets, or jointly) is outside the scope of this paper.

In addition, we purposefully reserve most other macroeconomic (e.g. Proprietors’ Income) and financial criteria to later studies which might investigate whether changes in government policies with critical reliance upon equity market returns are warranted.

Finally, while embedded in our null hypothesis, we ignore a critical SP500 definition. Namely, a company can have far fewer than 50% of its revenues, workforce, and/or assets in the USA and still qualify for inclusion in the SP500. In short, a company qualified as “American” for SP500 purposes can have very little relationship to or impact on US GDP.
3 **Common Acronyms and Abbreviations**

GDP: Nominal Gross Domestic Product as calculated by the BEA

GDPDEF: GDP Deflator as calculated by the BEA

rGDP: Real GDP (GDP adjusted for inflation) as calculated by the BEA, if a suffix is used then a deflator other than the GDPDEF was employed

GDPpc: GDP per capita, if the “r” prefix is used then “real” GDP per capita

GDP%: the percent change in GDP over a given time period, “r” prefix rule

(not annualized unless specifically stated)

PCEDEF: the deflator calculated by the BEA for Personal Consumption Expenditures

PPI: Producer Price Index calculated by the BLS

CPI: Consumer Price Index for Urban Consumers (CPI-u) calculated by the BLS

SP500: the Standard and Poor’s 500 Stock Index (of large cap stocks) as calculated by S&P Dow Jones Indices

SP500%: the percent change in the SP500 over a given time period

(not annualized unless specifically stated)

SPTR: the Total Return Index derived by S&P Dow Jones Indices for the SP500

(this index specifically includes dividends where the SP500 does not)
SPTR%: the percent change in the SPTR over a given time period

(not annualized unless specifically stated)

DJIA: the Dow Jones Industrial Average calculated by S&P Down Jones Indices

NDX: the NASDAQ-100 stock index as calculated by NASDAQ, this Index

is not restricted to US domestic companies but excludes financial companies

FED: any of the Federal Reserve entities inclusive of the Board of Governors

and the Regional Federal Reserve Banks and the FOMC

FOMC: The Federal Reserve Open Market Committee

BEA: Bureau of Economic Analysis, within the US Department of Commerce

BLS: the Bureau of Labor Statistics, within the US Department of Labor

Vintage: Economic statistics, particularly those released by the BEA and BLS are

frequently revised months or even years after the fact. The BEA and BLS preserve the

original releases as date-specific vintages.

4 A Review of Relevant Literature

There is a wide and generally rich body of research investigating the relationship

between GDP and other macroeconomic statistics and the equity markets. Contributors
to this field include Nobel Laureates such as Tobin, Modigliani, Markowitz, Miller, Sharpe, Kahneman, and Fama. ²

Before we begin our investigation into Granger causality over long time frames between returns on US equities and US Real GDP (rGDP), we acknowledge events such as the oil price shocks in the 1970s, the 1987 stock market crash, the 1998 credit contraction and the 2002-06 mortgage bubble could create “breaks”³ in the relationship between the two series and possibly impact causality in policy relevant (shorter) time frames. The credit contraction of 2008-09 resulted in a significant change in the focus of research in this area, therefore, we break the Literature Review into two portions, before and after the credit contraction of 2008-09.

4.1 Pre-2008 Literature

Saunders and Ghosh (2006), using data from an intermediate time frame (1971-2005) assert: (1) S&P500 returns are not cointegrated with rGDP although the NDX and DJIA were and importantly for this paper (2) rGDPc "Granger Causes" changes in returns from the S&P500. Interestingly, they also note the data the mortgage bubble of

² Award years: Tobin (1981), Modigliani (1985), Markowitz (1990), Miller (1990), Sharpe (1990), Kahneman (2002), and Fama (2013).

³ Chow tests can be used to test for the existence of breaks.
2000-2005 coincided with a significant increase in the connection between rGDP and equity market returns.

Ulrich (2002), using overlapping but somewhat earlier data from 1959-2002 also finds a strong relationship between equity market returns and rGDP. Interestingly, Ulrich determined: (a) there is collinearity unless the natural logs of the two series are used and (b) adding other macroeconomic variables such as Consumer Confidence increased the robustness of his model.

One of the most frequently cited pre-2008 studies (Cominicioli, 1996) used data from 1970-1994 and also found a strong relationship between rGDP and the S&P500. However, in contrast Saunders & Ghosh (2006) yet in line with the SP500’s status as an official Leading Indicator (Levanon 2011), Cominicioli found Granger causality running from equities to rGDP.

4.2 Post 2008 Literature

Given the focus of the pre-2008 research, it is interesting to note that much of the post-2008 literature seeks to describe, explain, or justify the apparent divergence between US equity returns and the underlying economy.

For example, Cooper and Dynan (2013) investigate the heterogeneity of the post 2008 financial asset recovery as a reason for the unusually long lag time between the recovery of financial assets and increased consumption. In particular, they focus on the demographic and geographic distribution of asset holders relative to the increased asset
values and how this relationship has lengthened the lag between a recovery in financial assets (primarily stocks) and rGDP.

On the other hand, Duca, Murphy, and Organ (April 2016) assert the recovery in asset prices, lower household debt, and increased access to consumer credit "bolstered U.S. consumer spending" starting in the middle of 2015. Interestingly, they too comment on the heterogeneity of assets, asset holders, and relative marginal propensities to consume (mpc). In particular, they note a dramatic difference between the mpc from "liquid assets" (nearly 9%) and illiquid assets like homes and pensions (roughly 1.5%). This, when juxtaposed against the holders of those assets, helps explain part of the apparent divergence between equity returns and increased consumption and, in turn, rGDP.

A slightly different tack is taken by Curdia and Ferrero (2013). These two senior economists from the Federal Reserve assert the Fed's Large Scale Asset Purchases ("QE") programs largely failed with respect to the stated goal of stimulating the economy. Unfortunately, for the purpose of this paper, they don't make a direct connection to the performance of the S&P 500 during their very short study period (2009-2013).

In stark contrast to all of the above, two officers of the Federal Reserve Bank of Boston, Fuhrer and Olivei (2011), assert the "QE" programs should add, on average 2.25% to rGDP for every 100 basis point decline in the yield of the US Treasury 10-year note. They also assert the lag should be only 2 years from the onset of the program. Given the yield on the UST 10yr subsequently declined by roughly 75 basis points from
the initiation of the QE programs to the cessation of new active purchases, this should have translated into just over a 1.5% increase in rGDP, which should be reflected in the SP500.

It is important to note the authors of the two 2013 papers and the 2016 paper (above) had the benefit of observing history, while Boston Fed paper was a forward-looking paper written during a period of intense debate about the appropriateness of Large Scale Asset Purchases.

4.3 **Causality: Equity Returns ⇔ rGDP**

There are many theories about the relationship of equity returns to economic growth. A discussion of the merits of the various theories is outside the scope of this paper. However, one hypothesis, the Efficient Market Hypothesis (Fama, 1970), is worth mentioning. Holmstrom (2015) provides a succinct summary:

"*The Efficient Market Hypothesis posits that information will be reflected rapidly in share prices and as a first approximation this seems to be empirically true.*"

Perhaps oversimplifying, we restate the above as: all public and private information, **whether actually known or merely expected**, is embodied in the current price of any financial asset. At the risk of going too far afield, we note some critics like Krugman (referenced in Malkiel 2011) assert the presence of asset bubbles/bursts as proof EMH fails. However, Malkiel (2011) and others point out EMH requires neither globally accurate expectations or perfectly accurate current prices.
We dwell on EMH because, in some ways, it underpins both dominant camps of "rGDP $\Leftrightarrow$ SP500 return" thinking. For example, if changes in rGDP Granger cause SP500 returns then, as more becomes known about rGDP, the body of known information increases and the SP500 efficiently adjusts to the new total knowledge. On the other hand, if the SP500 Granger causes rGDP, then it can be asserted SP500 investors expect a change in rGDP, and in turn, a change in the value of large companies.

The second camp finds important support in papers such as the often referenced Cominicioli (1996) and slightly modified in Croux & Reusens (2011). In addition, The Conference Board, provides implicit support to the second camp by continuing to include the SP500 (despite rather weak results) in its construction of the Leading Economic Indicators (Levanon, et al. 2011).

However, others such as Hymans (in Bosworth, 1975), Ulrich (2002) and Saunders & Ghosh (2006) find solid statistical evidence that changes in rGDP "cause" changes in SP500 returns.

4.4 Causality: Granger causality

This paper uses the common linear one-period ahead form of Granger causality (Granger 1969) when testing whether rGDP "causes" SP500 or vice-versa.

A practical language definition of the Granger test is: X Granger-causes Y if a regression using lagged values of X -and- lagged values of Y, provide statistically better
predictions about subsequent values of Y than a regression consisting only of lagged values of Y.

It is important to note the Granger test does not directly answer the question, "does A cause B". Rather, the Granger test essentially (in the context of this paper) tests to see if A precedes B on a statistically consistent basis (or vice versa, or bidirectional). A more technical definition can be found in Granger (1969).

However, while it is probably more correct to alter the term Granger causality to "Granger precedence", the field refers to this test as "Granger causality". In addition, if an economist desires to use A to forecast B, and if A is found to consistently precede (or forecast) B in a statistically significant manner... does the economist truly care if actual causality exists? *N.B. – briefly diverging from economics to finance, it is likely a trader or business unit manager values accuracy of forecasts over actual causality.*

Finally, other tests for "causality", such as the Structural Causal Model (Pearl 2009) exist but are generally more complex and, for the purposes of this paper are unlikely to materially improve on the Granger tests.

### 4.5 Returns, Inflation, and "Real" Indices

We narrowly and precisely define the return types, inflation measures, "real" measures (whether macroeconomic statistics or returns), and "re-indexed" measures in the Methodology section of this paper.
However, it is important to realize there is a great deal of debate about which measures are appropriate for the task at hand. For example, rGDPpc is seasonally adjusted (smoothed), adjusted for inflation, and adjusted for population growth (NIPA, 2016). In stark contrast, the most popular measures for the US equity markets (including the SP500) are not seasonally, not inflation adjusted, and not adjusted for population growth. (SP Indices Manual, 2016).

4.6 Four other common issues potentially degrade many studies.

First, the managers of the SP500 index do not include common dividends in the index (SP Indices Manual, 2016), this important aspect of the return from equities is omitted from most studies. Interestingly, as of this writing, the St. Louis Federal Reserve's "FRED©" Database does not publish the index which includes dividends, the SP500 Total Return index.

Second, many studies "deflate" the SP500 with the CPI-u while simultaneously using the GDPDEF to (implicitly) deflate rGDP. In addition to the obvious potential for discrepancy, the use of different deflators provokes the question, "which inflation index is the right one to use?" (Case & Wachter, 2011) (Hadbury et al, 2013)

Thirdly, given the rGDP is seasonally adjusted (X13 per NIPA) whereas the SP500 returns generally are not smoothed, a potential source of uncertainty based on the smoothing routines is introduced.

Lastly, the historical data for both the rGDP and the SP500 are subject to explicit biases. The rGDP is subject to quarterly, annual, pentennial, and ad-hoc revisions
resulting in a distorted data series that bears only slight resemblance to the information available to decision- and policy- makers at any historical point in time (Chang & Li, 2015) (Moulton, 1999). On the other hand, given the relatively frequent replacement of companies included in the SP500, the equity index suffers from both an inclusion and survivor bias (Denis et al, 2015).

4.7 Market Value of the SP500 (ex/with Survivor Bias) vs Price Index

Tobin, in his 1976 paper with Brainard (Tobin & Brainard, 1976) introducing what came to be known as "Tobin's q", attributes the following quote to Keynes (General Theory, p 151):

"[The] daily revaluations of the Stock Exchange, though they are primarily made to facilitate transfers of old investments between one individual and another, inevitably exert a decisive influence on the rate of current investment. For there is not sense in building up a new enterprise at a cost greater than that at which a similar existing enterprise can be purchased; whilst there is an inducement to spend on a new project what may seem an extravagant sum, if it can be floated off on the Stock Exchange at an immediate profit."

This quote is very relevant to the current paper because it makes no reference to the per-share price, it only references the total value of the enterprise. Tobin and Brainard shared that focus. So did the inventors of the SP500. This paper recognizes the SP500 is a “market capitalization” index and implicitly adopts the enterprise value perspective.
However, although the SP500 remains a "market capitalization" index, it doesn't truly reflect the current market value of the component companies. The index managers employ a variety of smoothing routines (i.e. "index math") accommodating events such as corporate actions, additions to the index, and deletions from the index. Additionally, frequent additions of companies to / deletion of companies from the SP500 index exacerbate the smoothing adjustments.

4.7.1 Importantly, to properly follow the market value based work by Tobin (1976), Shiller ("CAPE" Shiller, 1988), and recently Damodaran (2013), one must effectively recalculate the entire current and past market values for the entire index. This task is very difficult, accordingly most studies are performed on the raw index or first differences of the raw index. Even studies referenced in this paper such as Cominicioli (1996), Levanon, et al. (2011), and Ulrich (2002) employ the raw or differenced raw index. While we acknowledge the potential for error, for comparison purposes, we perform some calculations based on the raw (or first differenced raw) SP500 Index. We then employ the raw or first differenced SP500 Total Return Index, and the Total Market Value of the SP500. In addition, as per Damodaran (2013) we also use a modified Total Market Value which includes the market value for previously removed companies (an effort to remove the survivor bias).
Interestingly, using the "right" index may not result in statistically different results. Damodaran (2013) provides several examples of qualitatively similar results when using the SP500 Total Return index in comparison to a constructed total market value.

4.8 1977/78 "Break" in rGDP and GDPDEF series

The final section of the review of existing literature focuses on an apparent break in the GDP and GDPDEF data series during 1977-78. Kitov (2010) illustrates this break between the CPI-u and the GDPDEF on his blog and refers to a methodology change by the BEA around that time. Interestingly, the impact noted by Kitov is actually a combination of several changes implemented during the 1999 major revisions.

According to Moulton (1999), one much later yet significant change in methodology (to chain-weights) caused the BEA to revise inflation data back to 1978. However, and importantly, no deeper historical data (pre-1978) were revised:

"Incorporation of the geometric mean type consumer price indexes (CPI's) that are currently used to deflate consumer expenditures beginning with 1995 to deflate consumer expenditures back to 1978, increasing the consistency and accuracy of the time series for real PCE and real GDP."

In addition, they note "[b]eginning with 1977, an annual chain-type price index will be calculated… for each nonfinancial industry." (Moulton, 1999, page 8).
4.9 *Outside this paper's scope yet worth mentioning*

Four more items are worth noting with respect to the potential sources for the 1977/78 break. A thorough discussion of these items is outside the scope of this paper, therefore we simply note them for reference and possible future research.

a) During the mid to late 1970s, junk bonds and Leveraged Buy Outs ("LBOs") became a popular method to exploit the interest expense deduction (Holmstrom, 2015 and Warren, 1974) this financial innovation continued to expand rapidly until the demise of Drexel Burnham Lambert in 1989 and had an associated effect of boosting selected stock prices,

b) Government efforts to mitigate 1970s price volatility artificially impacted reported prices. Interestingly, some price controls that lasted until Reagan removed them in 1981 (Griswold, 2011),

c) Negotiable Orders of Withdrawal ("NOW Accounts") (Kaplan, 1973) had a surprising impact on consumer behavior and on the US Money Supply, and

d) Short term US interest rates\(^4\) saw unprecedented volatility during the middle to late 1970s creating problems for the economy, policy makers, and statisticians.

\(^4\) The Federal Funds rate spiked to nearly 13% in July 1974, fell back to 5% by 1977 and spiked again to over 17% by July 1980. Board of Governors of the Federal Reserve System (US), Effective Federal Funds Rate [FEDFUNDS], retrieved from FRED, Federal Reserve Bank of St. Louis.
5 Methodology & Detailed Definitions

Common knowledge is frequently wrong and/or oversimplified. Given the focus of this paper, it is imperative to ensure critical terms are carefully defined. This section provides detailed descriptions for several of those terms and contrasts those terms to similar terms.

5.1 Equity Market Indices

We first address key features of four popular equity market indices: the SP500, DJIA, NDX, and Wilshire 5000. Three of these four are “modified market capitalization weighted” indices. The fourth, the DJIA, is a price weighted index.

The primary difference, for our purposes, is the capitalization weighted indices are more likely to reflect population-wide investment patterns and returns. The two relevant principal drawbacks to the market capitalization indices are: (1) volatility of returns among the very large companies far outweighs contributions from smaller companies and (2) market capitalization indices are considerably more difficult to construct.

A second difference is the breadth and sectorization of coverage. The Wilshire5000 includes “all actively traded” stocks in the US regardless of domicile (as of this writing roughly 3500 companies (Wilshire, 2017)), the SP500 includes roughly 500 of the largest US domiciled companies (S&P Dow Jones Indices, 2016), the NDX includes 100 large non-financial companies listed on the NASDAQ (regardless of domicile per NASDAQ, 2017), and the DJIA includes only 30 of the largest US stocks.
(S&P Dow Jones Indices, 2016). In addition, with the exception of the Wilshire5000, the indices make attempts to ensure various sectors of the economy are proportionately represented.

Interestingly, despite the different methods, the above listed indices are highly correlated (both in level and returns) over both intermediate and long terms (CMEgroup / Klein, 2008), especially when US equity markets are near extremes (Michel, et al 2015). The correlation, data availability, and breadth all contributed to my decision to select the SP500 as the equity index of interest for this paper.

Unfortunately, all of the above indices share two common problems, they omit returns to investors resulting from common dividends and they all have “survivor and/or inclusion bias”. Except for the very long term benchmark analysis (where we used data methodology consistent with prior studies), and largely because an appropriate index is not available in the public domain, we created our own SP500 based total value index. Essentially, we used the SPTR as a base, enhanced that index with dividends, and adjusted the result for inflation. The details of the transformation process are covered below in the Data section.
5.2 Inflation Definitions & Comparisons

"Nobody's right if everybody's wrong"5

A common question in macroeconomics as well as financial economics is, “What is the right inflation gauge” to use? Clearly, given the plethora of inflation measures, the answer largely depends on the user’s objectives. The GDPDEF at first glance, seems to be the preferred choice for this paper. And, in fact, this is the index we use to deflate GDP. However, given the construction of the GDPDEF, unwanted bias and/or errors may be introduced when comparing time series of rGDPpc against financial market data. This problem is especially true for time frames spanning or following an apparent 1977-1978 break in either the GDP or the GDPDEF. Thus, in the interest of completeness, we discuss GDPDEF and briefly note three other principal measures of inflation.

This section introduces and discusses some of the relevant strengths and weaknesses of four important inflation indexes: GDPDEF, the PCEDEF, the PPI, and the CPI-u. We do not discuss most of the sub-indices (eg. “ex-food and energy”) as they are not relevant to this paper. In addition, we point out most economic indices are “seasonally adjusted (BEA blog, 2015)” or otherwise smoothed while most financial

5 “For What It’s Worth”, from the album “Buffalo Springfield” by Buffalo Springfield, Gold Star Studios, December 5, 1966 (written by Stephen Stills)
market returns are not (Damodaran, 2017). This of course introduces the possibility of a variety of errors when comparing rGDPpc to equity market returns; however, we leave that discussion to another paper.

5.2.1 GDP Deflator “GDPDEF”

The official definition and methodology for deriving the GDPDEF can be found in Chapter 4 of the current “Concepts and Methods of the U.S. National Income and Product Accounts” (NIPA, October 2016). A skim of this section reveals several interesting aspects of the GDPDEF and at least two errors in the “common knowledge” about this measure. One aspect particularly relevant to this paper is the GDPDEF incorporates portions of the CPI-u, the PPI, and has a circular relationship with the PCEDEF. Therefore, over long periods of time, the GDPDEF should closely resemble the other three indices. In fact, given the relatively high percentage contribution by PCE to GDP, we should expect the PCEDEF and the CPI-u to be very closely related to the GDPDEF. Studies have shown these close relationships exist, particularly before 1978 (Kitov, 2010).

The first “common knowledge” error, on the other hand, probably arises from the structures of popular economic databases (Stierholz, 2012) and/or from oversimplification in the classroom. The GDPDEF does not arise from direct observation of price changes (even adjusted for quality differences). We quote from NIPA (2016):
“The chain-dollar estimates are used in the calculation of another price index, the implicit price deflator (IPD). The for period \( t \) is calculated as the ratio of the current-dollar value to the corresponding chained-dollar value, multiplied by 100, as follows:

\[
IPD^p_t = \frac{\sum p_t^* q_t}{CD_t^p} \times 100.
\]

For all aggregates and components and for all time periods, the value of the IPD is very close to the value of the corresponding chain-type price index. Note that this definition of the IPD differs from that used before the introduction of chain-type measures in 1996, when the IPD was defined as the ratio of the current-dollar value to the corresponding constant-dollar value.”

I simplify and emphasize the common misconception: the GDPDEF is not a stand-alone metric, it is derived from comparing the chain-weight rGDP (Cahill, 2002) to the nominal GDP. In other words, the “Real GDP” is not the derived value, the deflator is the derived value.

This concept is critical to understanding how the GDPDEF can change (sometimes dramatically) from one vintage to the next and why using only the most recent revision often introduces errors in time-series analysis.

Again, we emphasize decision making is predicated on the information available at the time of the decision (and expectations about future information). The official data series for GDP, rGDP, GDPDEF, and most of the remaining indices reflect only the most
recent revision. Importantly, the most recent revision is sometimes decades after the policymaking and/or business decision period.

Therefore, if one is investigating causality, it is critical to use the then-current vintage in an effort to identify then-current information.

The second error arises from the chain method itself. The chain-weighting methodology used by the BEA is subject to a phenomena known as “chain drift”. Essentially, the farther away (in time) from the chained reference data one moves, the larger the residual error becomes. This drift, even without major changes in product quality or technology, would be sufficient reason for regular “benchmark revisions” (NIPA, 2016).

Lastly, we address the paired concepts of “imputations” and “new additions” to GDP, their impact on historical measures of rGDP and GDPDEF.

Imputations are (generally) modeled estimates of portions of the economy the BEA desires to include in GDP or GDPDEF which are either too difficult or too costly to obtain in a timely manner. An example might be the “rent equivalent” contribution to GDP made by the physical house in which a homeowner resides\(^6\) (NIPA, 2016). A new addition arises from the advent of a new technology or product sector (e.g. email) or the

\(^6\) WDC comment: yes, as strange as this sounds, the physical house is adding to GDP simply if someone lives in the structure.
inclusion of a previously excluded sector (e.g. as of 2013 “Movies and Films” are included as “Private Fixed Investments”) (NIPA, 2016).

Imputations, by their nature, are guesses but must be done. These introduce smoothed projections of non-smooth production and therefore introduce errors to both GDP and GDPDEF (BEA FAQ, 2017).

New Additions present a special problem with respect to historic revisions. It is not possible to estimate, given previously existing societal mores and consumption patterns, changes in historical consumption patterns (e.g. substitution effect) if the novel product had been available. Accordingly, the BEA simply adds the new amount to the current year and adjust prior periods in a deflated pro-rata manner.

It is beyond the scope of this paper to suggest a better solution to the drift, imputation, and addition problems.

Importantly, for the purposes of this paper, it seems reasonable to chain then-current percent changes in historic vintages of nominal GDP, real GDP, and the GDPDEF and re-index the entire series inclusive of recent additions on a compounded basis.

5.2.2 PCE Price Index (and its many relatives)

The headline form of the PCE Price Index “measures the prices paid for the goods and services purchased by ’persons’” (NIPA / BEA, 2016). Although not technically correct we will, without harming this paper, refer to the headline PCE Price Index as the “PCE deflator” and abbreviate it as PCEDEF. The PCEDEF includes household
expenditures on items ranging from food and fuel to pension plan expenses. Given the inclusion of substantial portions of the CPI-u in the PCEDEF, it is reasonable to expect close similarities between these two measures. A short yet complete comparison can be found in “A Reconciliation between the Consumer Price Index and the Personal Consumption Expenditures Price Index” (McCully et al, 2007). The oversimplified summary of the “Reconciliation” paper determines three primary sources of difference: (a) the PCEDEF is designed to capture more consumers (non-urban as well as urban) therefore the weights given to identical consumption items vary between the two indices, (b) differences in the scope of expenses, and (c) differences in seasonal adjustment methods.

Finally, given the PCEDEF is a sub-index of the GDPDEF and the very tight relationship to the GDPDEF over both short and long periods, it is reasonable to question why it is included in this comparison. Quite simply, the PCEDEF (at this writing) is one of the FOMC’s preferred gauges of inflation measurement (BOG Federal Reserve, 2017). Therefore, even though we will not use it in the following analysis, the PCEDEF deserves an honorable mention.

5.2.3 PPI

The PPI is referred to as a “business index” and, if a core contention of Real Business Cycle theory is correct (businesses lead recessions, not consumers) (see Mankiw, 1989 for critique), the PPI is more relevant to business planning than GDPDEF or CPI-u. By extension, if the PPI is more relevant to business planning, then it is
probably more relevant when comparing equity market returns to rGDP. A measure of rGDP obtained by deflating nominal GDP with the PPI could be denoted as rGDPppi.

Technically, the PPI measures average changes in prices received by domestic producers for their output. This is markedly different than the CPI-u which measures prices paid by urban consumers. While the PPI also uses imputed values, a very high percent of the PPI’s physical good inputs is directly sampled by the BLS. The PPI coverage of the service sector of the economy, at just over 70%, has markedly improved over the last decade but remains incomplete (BLS PPI, 2014).

A major problem of the PPI is its relatively short history in its current form. The PPI underwent substantial methodology changes over the last 20 years with a dramatic expansion of the “service industry” information.

5.2.4 CPI-u

The CPI-u, “represents the buying habits of the residents of urban or metropolitan areas in the United States” (BLS CPI, 2015)

The preceding quote oversimplifies a very complex calculation, one which differs in many critical respects from the PCEDEF, GDPDEF, and the PPI. For example, the CPI-u excludes home purchases as a consumption item (these are considered investments) but includes an estimate of urban rent changes (the GDPDEF and PCE have derived rent equivalency imputations that occasionally differ quite markedly from the urban rent component of the CPI-u). Sales taxes, included in the CPI-u, are another example of a difference between the CPI-u and the PCEDEF. A third and final
significant difference between the CPI-u and the previously discussed inflation indices is
the relatively low rate of revision to historical CPI-u. This low revision rate results from
a much more extensive data collection method that results in fewer imputations and
therefore reduced revisions.

For the purposes of many papers, the lower revision rate of the CPI-u elevates its
status as a contemporaneous input for decision makers, particularly when compared to
the GDPDEF. Therefore, it may be reasonable to “deflate” vintages of nominal GDP by
the CPI-u (resulting in rGDPcpi) when investigating the relationship between SP500 and
a measure of “real” GDP.

5.3 Common Measures of Investment Returns

Return on investment (“ROI” or simply “return”) is a simple sounding relatively
straightforward concept. However, like inflation, there are a variety of ways to calculate
returns on financial assets. One could simply measure the total dollars on hand at the end
of a period versus the total dollars invested at the start, or one could pursue any of the
complex spread valuation methods to account for different cash flows (in /out) at
different points of time assessed at variant spreads to riskless returns.

Given the goal of this paper, our primary concern with respect to ROI is to use a
method reasonably comparable to the method used in determining changes in rGDP.
Accordingly, we define several simple measures of return and comment on possible
variants.
5.3.1 Single Period Returns

We define single period returns as the value obtained by subtracting 1 from the quotient of the amount returned divided by the amount invested. We also refer to this result as a “periodic” return. It is, in most cases, analogous to simple interest.

The single period return is useful as a component of “re-indexing” economic and financial series to common bases. However, one has to take care to ensure the return is quoted as a periodic return rather than the annualized rate the periodic return represents (whether compounded or linearly extended).

5.3.2 Multiple Period Returns

Most implementations of multiple period return transform the single period return, measured over several periods, into an annualized return. Commonly this is done by compounding / decompounding the simple return as needed. Occasionally; however, papers will simply average a string of periodic return and report the result as an annualized return. In addition, some commonly accepted implementations use a blend of the two methods, an obvious example is the standard “yield-to-maturity” calculation for US Treasury bonds and notes. Clearly, as in the single period returns, care must be taken to determine the annualization method used by any given author.

5.3.3 Inflation Adjusted or “Real” Returns

Most modern economic papers convert nominal returns from stocks, bonds, and commodities to real returns by adjusting the periodic or annualized returns by some measure of inflation. The often cited study by Ibbottson (1976) is a good example of an
early post-Big Bang work relying on inflation adjusted returns. A similar recent article by Ramraika (2014) illustrates the continuing and expanding use of inflation adjusted or real returns. Unfortunately, the financial press and the referenced papers generally compare nominal equity market returns with rGDP, for consistency we will also use nominal equity market returns in our long-term benchmark study and correct for this in our policy-relevant studies.

5.3.4 Index, Mutual Fund, Stock, and ETF Returns

A failing of many common measures of return from stocks, mutual funds, ETFs, and equity market indices is the omission of common dividends. Interestingly, the same professionals and pundits who automatically include interest payments on bonds in the yield-to-maturity conveniently ignore that portion of the return on stocks attributable to dividends (and reinvested dividends).

This omission may arise from the relatively uncertain path of common dividends and the accompanying difficulty of tracking them or it may be a byproduct of long standing industry habits. For example, the very good economic database maintained by the St. Louis Fed (“FRED®”) does not contain the SPTR or any other total return index on mainstream US equity indices. Interestingly, even S&P Dow Jones Indices, the producer of the SP500 and its total return companion the SPTR, subordinates the SPTR on its website.

Unfortunately, due to relatively active management of the component stocks in the SP500 and the amazingly sparse publicly available dividend data for former
component companies, we are unable to construct a true historical total return index for
the SP500. However, working under the assumption that many stocks dropped for
performance reasons from the SP500 were paying little to no dividend, we were able to
recreate an implied estimated dividend series for the SP500 ex-survivor bias. This series,
along with the regular SPTR, was used in our analysis.

5.3.5 Other Important Measures of Equity Market Returns

We do not use the following measures in this paper; however, we would be remiss
if we did not mention risk-adjusted returns in this section. Measures such as the Sharpe
Ratio (named for Nobel Laureate William Sharpe), RAROC, VaR, Sortino, and ERP
(Equity Risk Premium) have been used by the financial community for at least a half
century. Recently, they gained considerable traction among economists and have drawn
the interest of behavioral economists such as Kahneman. While these measures are not
immediately germane to this paper, their concepts suggest interesting future research.

5.3.6 Growth Rate for GDP and other NIPA Measures

Finally, since Investment Returns represent the “growth of an investment”, we
include the definition of GDP Growth (BEA GDP Math, 2008). The BEA notes the
formula it uses to calculate the average annual growth is a variant of the compound
interest formula:

\[ r = \{ \left[ \frac{(GDP_t / GDP_0)^{(m/n)}}{m/n} \right] - 1 \} \times 100 \]
Where:

GDPt is the level of activity in the later period;

GDP0 is the level of activity in the earlier period;

m is the periodicity of the data (for example, 1 for annual data, 4 for quarterly data, or 12 for monthly data); and

n is the number of periods between the earlier period and the later period (that is t-0).

5.4 Additional Methods

Data Conversion

Data acquisition, cleaning, and conversion proved to be the most difficult aspect of this paper. Specifically, unexpected challenges arose with respect to macroeconomic data (i.e. deflating vintages).

5.4.1 Macroeconomic Data

As noted above, several economic data series are subject to frequent revision. In addition, some economic data series ceased to exist, some are relatively recent, and at least one changed in very material ways (PPI). The challenge was to knit together sub-indices and/or vintages of nominal and real data to produce a consistent set of data.

Importantly, given most resources neither mention or explicitly correct for this problem, we can imply most studies rely on poor or even inappropriate data. For
example, if one attempts to investigate decision making processes over time, it is inappropriate to use the most recent revisions of rGDP since those revisions do not reflect what was known (facts or expectations) at the time of the decision.

For the purposes of this paper, we’ve only adjusted nominal GDP with matched vintages of GDPDEF to obtain then-current rGDP; however, to mimic decision making information at then-current moments in time, we think it would be interesting to deflate vintage GDP with date matched PPI and CPI-u.

5.4.2 Equity Market Data

We expected and experienced significant difficulty in obtaining accurate equity market data. First, primarily for licensing reasons, it is very difficult to retrieve historical equity index data for time periods dating back more than 10 years. Fortunately, we had access to a sufficiently robust database constructed and maintained by a private individual dating to before 1960. In addition, by combining data from three other databases with the aforementioned private data, we were able to construct a time series for an estimate the Total Market Value ("TMV") for the Survivor-Bias Free SP500 ("SP500f"). Essentially, this process involved identifying companies dropped from the SP500 for performance reasons then determining the post-removal market values for those companies, then adding the "missing" market value back to the total market value for the current SP500 at any point in time. We emphasize that we focused on companies removed from the index for performance reasons. Mergers, acquisitions, name changes,
and other events can cause a symbol change or removal of an existing company; however, those events do not create survivor bias.

Once we had the TMV for the SP500f then we had to arrive at an estimate of the dividends paid by the companies. This proved to be a significant hurdle. In the end, we used a combination of: (a) the implied dividend rate from the SPTR compared to the SP500 and (b) the published dividend rate for the SP00. Interestingly, while very similar, the two rates were not identical. We then applied that rate to the SP500f to arrive at a cash value for the dividend and added that cash value to the TMV to arrive at a TMV+Dividend or an SP500f Total Return (“SPTRf”).

Finally, to ensure accurate and reasonable comparisons of this new SPTRf to macroeconomic data, we adjusted the SPTRf for inflation.

5.4.3 Re-indexing, First Differences, and Stationarity

Graphical presentations are an important part of any paper, including the present paper. However, when presenting rGDP and SP500 on the same graph, the difference in scales generally requires a right and left vertical axis. In addition, since the baseline or reference date for macroeconomic data series varies substantially (even within an index family), straightforward comparisons are often difficult. For example, if an index moves from 200 to 200.5 in a given time period and another moves from 27 to 27.0675, this is the same percent move but not intuitively obvious.
Therefore, to facilitate inspection and presentation, we re-indexed many of the economic and return series to a common date (January 1, 1991). Interestingly, this somewhat aesthetically based process greatly assisted error and outlier checking.

Somewhat related to re-indexing is the process of “first differencing”. Most of the data series (financial and economic) we worked with are known to be non-stationary. Accordingly, for the Granger causality tests, we transformed the relevant data series to periodic percent changes (non-annualized). Unless otherwise noted, the reader should assume all first differenced series were tested (see 7.1 below) and found stationary and first differenced series were used in the analysis.

5.4.4 Time Frames

Intuition suggests ours and others’ results could be impacted by the selected time frame. For example, a randomly selected 40 consecutive quarter period might possibly land on a period of relatively low volatility and stable growth patterns or it could straddle a period of wrenching change and high volatility. The same tests conducted on those two periods could easily result in opposite views of whether rGDPpc Granger causes SP500 or vice versa.

In addition, the long term Granger test (the benchmark test in this paper) might show causality running in one direction while a shorter time frame might show the opposing direction. Thus, (with a nod to Keynes) given very-long-run results are known with certainty, we acknowledge policymakers have more interest in short or intermediate time frames. Finally, in keeping with our stated goal (does the SP500 Granger cause rGDP
over policy relevant time frames), it is necessary to test a variety of rolling short and intermediate term tests.

Therefore, we conducted repeated trials of the Granger causality test. We tested the entire term of available data, 40 quarter periods starting on 5 year cycles, and 76 quarter long periods starting on 19 year cycles. We note the 40 quarter periods, once adjusted for the larger lags, have relatively low remaining degrees of freedom. Thus, while they might be of interest if one uses higher frequency data, they are omitted from this paper. In addition, given the data was easily split into consecutive thirds of 19 years each, we used 19 year (instead of 20 year) periods for the policy relevant tests.

We also conducted Chow tests to identify statistically significant breaks in the data.

5.4.5 "Street Version": rGDP v Nominal SP500

While seemingly misaligned with the goal of this paper and the focus on properly matching data series, we recognize a need to show a commonly accepted benchmark. Therefore, despite the previously noted critical problems with comparing growth rates of rGDP to nominal SP500 returns, we ran this test over the very long term (the entirety of the available data) and include those results.
5.5 **Chow tests**

The method commonly used when testing for a break in the data is the Chow Test. According to the User Manual for the statistical software “Shazam”, the Chow test formula is:

$$\text{CHOW} = \frac{(\text{SSE} - \text{SSE}_1 - \text{SSE}_2)/K}{(\text{SSE}_1 + \text{SSE}_2) / (N_1 + N_2 - 2K)}$$

Where:

- SSE1 and SSE2 are the sum of the squared errors from the first and second portions of the data series (the portions on either side of the split or break)
- N1 and N2 are the number of observations on either side of the split
- K is the number of estimated parameters. (SHAZAM, 2011)

Thus if the result is less than the critical value from an F-distribution with parameters (K, N1+N2-2K ) distribution “then there is no evidence for a structural break” (SHAZAM, 2011).

However, if the result is greater than the critical value, there is statistical evidence of a structural break in the data.

Importantly, the identified break point is not necessarily the exact location of a change in the data but rather the point at which the change in the regression statistics become significant.
6 Graphical Summary of Data

Below we provide a quick graphical summary of the re-indexed (basis 1991) data for the real per-capita GDP, the nominal SP500, the deflated SP500 (using the GDPDEF) and the quarter to quarter percent changes in the three series.

It is reasonably easy to see strong tendencies in the raw data to trend while the first differenced (percent change) data shows very little trend.

7 Empirical Results Part 1: Long Term Benchmark Results

7.1 Adjusted Dickey-Fuller test for stationarity – even though we assumed non-stationarity for raw data and stationarity for first differenced data per common use
cases, we tested both sets of series. The raw data proved to be non-stationary and the first differenced data proved to be stationary.

7.2 Test for autocorrelation (Durbin-h vs Durbin-Watson)

This stat is shown in APPENDIX 1 with each set of regressions (lag 8, 4, and 2 restricted and unrestricted). The reader will find all Durbin-Watson statistics are extremely close to 2.0 thus, “apparently”, indicating no material autocorrelation for any of the restricted or unrestricted regressions.

However, and very importantly, the standard DW test is not the correct test to use. All data series considered in this paper include lagged values of the primary data, therefore the proper test is the Durbin-h test.

\[ h = \hat{\rho} \sqrt{\frac{N}{1 - N \hat{\sigma}^2_\alpha}} \]

where \[ \hat{\rho} = \frac{\sum_{t=2}^{N} e_t e_{t-1}}{\sum_{t=2}^{N} e_t^2} \]

Unfortunately, as is clear from the above equation, the Durbin-h statistic cannot be calculated by most statistical applications if the term under the radical is negative (compels imaginary space). A careful inspection of the above equation reveals, due to the reliance on “N” in the denominator’s subtracted term, large samples are prone to negative terms under the radical. Indeed, this was the case with our data. Fortunately,
for the policy-relevant periods, the combination of smaller N and/or smaller variances permitted calculation of the Durbin-h statistics. Those results are reported in APPENDIX 3.

7.3  \textit{Granger Causality test (linear) based on the \textquotedblleft F-test\textquotedblright} 

Granger Causality tests require us to test if the rGDPpc “causes” SP500%. If we find the unrestricted model (i.e. the model including both lagged rGDPpc as well as lagged SP500%) has a better fit than the restricted model (i.e. lagged values of SP500% forecasting itself) then we can say the rGDPpc “Granger causes” the SP500%.

To do this, we create a “Restricted model” where we regress SP500% on past values of itself, lagged back up to and including 8 quarters (where \(b_i\) = the coefficient for each lagged term):

\[
SP500\% = \alpha + b_1(SP500\% - 1) + b_2(SP500\% - 2) + b_3(SP500\% - 3) + b_4(SP500\% - 4) + b_5(SP500\% - 5) + b_6(SP500\% - 6) + b_7(SP500\% - 7) + b_8(SP500\% - 8) + u
\]

Then we repeat the regression, this time creating an “Unrestricted model” by adding as new independent variables 8 lags of rGDPpc (for a total of 17 variables, (where \(b_i\) = the coefficient for each lagged SP500% term and \(g_i\) = the coefficient for each lagged GDP term).
\[ SP500\% = \alpha + b_1(\text{SP500}\% - 1) + b_2(\text{SP500}\% - 2) + b_3(\text{SP500}\% - 3) + b_4(\text{SP500}\% - 4) + b_5(\text{SP500}\% - 5) + b_6(\text{SP500}\% - 6) + b_7(\text{SP500}\% - 7) + b_8(\text{SP500}\% - 8) + g_1(\text{rGDPpc} - 1) \\
+ g_2(\text{rGDPpc} - 2) + g_3(\text{rGDPpc} - 3) + g_4(\text{rGDPpc} - 4) + g_5(\text{rGDPpc} - 5) + g_6(\text{rGDPpc} - 6) + g_7(\text{rGDPpc} - 7) + g_8(\text{rGDPpc} - 8) + \epsilon \]

Then we perform an “F-test” comparing the Restricted and Unrestricted models as:

\[ F_{\text{test}} = \frac{(SSE_r - SSE_u)}{m} / \frac{(SSE_u / (n-k-1))}{1} \]

Where: 
- \( SSE_r \) = sum of the squared errors from the restricted model
- \( SSE_u \) = sum of the squared errors from the unrestricted model
- \( m \) = number of omitted lags from the "suspected causal" variable
- \( n \) = number of observations
- \( k \) = total number of parameters in the UNrestricted model

Then, we compare the F-stat to a critical value to determine its significance. We used generally accepted significance levels of 10\%, 5\%, and 1% in our tests. If we found we could reject the “Null” hypothesis (H0) of “the additional variable did not improve the regression model, then we could assert the additional variable “Granger caused” the dependent variable. In the first case above, the Dependent variable was the SP500\% and the additional variable was rGDP.
Then, we repeated the above series of steps and tests for lags of only 4 quarters and only 2 quarters. We note, it is somewhat uncommon to use lags of only 2 quarters when working with GDP data; however, based on the AIC stats for the 8, 4, and 2 lag models we tested this short lag model as well.

Finally, to test for causality running from SPTR% to rGDP%, we reversed the dependent and independent variables (series) and repeated the above tests.

The intermediate and final results are presented in the table below. The last section of the table illustrates that all three combinations of lags (8, 6, and 2 quarters) reject the hypothesis that rGDPpc Granger causes SP500% and support the hypothesis that SP500% Granger causes rGDPpc. (note: because of Shazam labeling restrictions, SP500% is denoted as NOM_SP500 and rGDPpc is denoted as rGDPpc09):

<table>
<thead>
<tr>
<th>Lag</th>
<th>F-test</th>
<th>ConLevel</th>
<th>Lag</th>
<th>F-test</th>
<th>ConLevel</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>138.8</td>
<td>0.000023</td>
<td>4</td>
<td>130.65</td>
<td>0.000207</td>
</tr>
<tr>
<td>4</td>
<td>141.54</td>
<td>0.000017</td>
<td>9</td>
<td>134.17</td>
<td>0.000002</td>
</tr>
<tr>
<td>2</td>
<td>141.63</td>
<td>0.000012</td>
<td>5</td>
<td>137.20</td>
<td>0.000002</td>
</tr>
</tbody>
</table>

\[
F_{\text{test}} = \frac{\text{SSE}_r \cdot \text{m}}{\text{SSE}_u / (n-k)}
\]

where

- \( \text{SSE}_r \) = sum of the squared errors from the restricted model
- \( \text{SSE}_u \) = sum of the squared errors from the unrestricted model
- \( m \) = number of omitted lags from the "suspected causal" variable
- \( n \) = number of observations
- \( k \) = total number of parameters in the Unrestricted model

All Models: \( n = 229 \) Observations

<table>
<thead>
<tr>
<th>Lag</th>
<th>F-stat</th>
<th>ConLevel</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2.968035969</td>
<td>0.0000023</td>
</tr>
<tr>
<td>4</td>
<td>4.584385763</td>
<td>0.000017</td>
</tr>
<tr>
<td>2</td>
<td>8.811576942</td>
<td>0.0000207</td>
</tr>
</tbody>
</table>

In all three of the above, we fail to reject the null hypothesis that rGDPpc Granger causes SP500%.
8 Empirical Results Part 2: Policy Relevant Data & Terms

8.1 Adjusted Dickey-Fuller test for stationarity

We repeated the ADF tests for the adjusted and reindexed data (e.g. the SPTR). The reindexed data proved (as for the raw data) to be non-stationary and the first differenced data proved to be stationary.

8.2 Test for autocorrelation (Durbin-h vs DW)

While we repeated the DW tests for the adjusted and reindexed data, for each set of regressions (lag 8, 4, and 2 restricted and unrestricted), we remind the reader of our previous comments regarding the inapplicability of the standard DW test and the relevancy of the Durbin-h test. Again, standard DW tests are ignored and the Durbin-h tests, if they were able to be calculated, are reported in APPENDIX 3.

8.3 Granger Causality test (linear) based on the “F-test”

Granger Causality tests require us to test if the rGDPpc “causes” real SPTR%. If we find the unrestricted model (i.e. the model including both lagged rGDPpc as well as lagged real SPTR%) has a better fit than the restricted model (i.e. lagged values of real SPTR% forecasting itself) then we can say the rGDPpc “Granger causes” the real SPTR%.
To do this, we create a “Restricted model” where we regress real SPTR% on past values of itself, lagged back up to and including 8 quarters (where \( b_i \) = the coefficient for each lagged term):

\[
\begin{align*}
\text{pcRSP} &= \alpha + b_1(\text{pcRSP}-1) + b_2(\text{pcRSP}-2) + b_3(\text{pcRSP}-3) + b_4(\text{pcRSP}-4) + b_5(\text{pcRSP}-5) + b_6(\text{pcRSP}-6) + b_7(\text{pcRSP}-7) + b_8(\text{pcRSP}-8) + u
\end{align*}
\]

Then we repeat the regression, this time creating an “Unrestricted model” by adding as new independent variables 8 lags of rGDPpc (for a total of 17 variables, (where \( b_i \) = the coefficient for each lagged pcRSP % term and \( g_i \) = the coefficient for each lagged GDP term).

\[
\begin{align*}
\text{pcRSP} &= \alpha + b_1(\text{pcRSP}-1) + b_2(\text{pcRSP}-2) + b_3(\text{pcRSP}-3) + b_4(\text{pcRSP}-4) + b_5(\text{pcRSP}-5) + b_6(\text{pcRSP}-6) + b_7(\text{pcRSP}-7) + b_8(\text{pcRSP}-8) + g_1(\text{rGDPpc-1}) + g_2(\text{rGDPpc -2}) \\
&+ g_3(\text{rGDPpc -3}) + g_4(\text{rGDPpc -4}) + g_5(\text{rGDPpc -5}) + g_6(\text{rGDPpc -6}) + g_7(\text{rGDPpc -7}) + g_8(\text{rGDPpc -8}) + u
\end{align*}
\]

Then we perform an “F-test” comparing the Restricted and Unrestricted models as:

\[
F\text{ test} = ( (\text{SSE}_r - \text{SSE}_u) / m ) / (\text{SSE}_u / (n-k-1))
\]

43
Where: \( \text{SSE}_r \) = sum of the squared errors from the restricted model

\( \text{SSE}_u \) = sum of the squared errors from the unrestricted model

\( m \) = number of omitted lags from the "suspected causal" variable

\( n \) = number of observations

\( k \) = total number of parameters in the UNrestricted model

Then, we compare the F-stat to a critical value to determine its significance. We used generally accepted significance levels of 10\%, 5\%, and 1\% in our tests. If we found we could reject the “Null” hypothesis (H0) of “the additional variable did not improve the regression model, then we could assert the additional variable “Granger caused” the dependent variable. In the first case above, the Dependent variable was the SP500\% and the additional variable was rGDP.

Then, we repeated the above series of steps and tests for lags of only 4 quarters and only 2 quarters. We note, it is somewhat uncommon to use lags of only 2 quarters when working with GDP data; however, based on the AIC stats for the 8, 4, and 2 lag models we tested this short lag model as well.

Finally, to test for causality running from SPTR\% to rGDP\%, we reversed the dependent and independent variables (series) and repeated the above tests.

The intermediate and final results are presented in the table below. The last section of the table illustrates an important change in the direction of causality. We find
for the earliest time frames (i.e. before the mid 1980s) all three combinations of lags (8, 6, and 2 quarters) reject the hypothesis that rGDPpc Granger causes real SPTR% and fail to reject the hypothesis that real SPTR% Granger causes rGDPpc. However, we find for the most recent time frames (i.e. after the mid 1980s) all three combinations of lags (8, 6, and 2 quarters) fail to reject the hypothesis that rGDPpc Granger causes real SPTR% and reject the hypothesis that real SPTR% Granger causes rGDPpc. (note: because of Shazam labeling restrictions, real SPTR% is denoted as pcRSP09 and rGDPpc is denoted as rGDPpc09).

Restating the above in simplified and plain terms:

Prior to the mid 1980s we see the SP500 is a leading indicator for the economy (i.e. the SP500 Granger causes rGDP). However, and in stark contrast, after the mid 1990s the SP500 ceases to be a leading indicator (rGDP Granger causes SP500%).
First 19 Years: “Accept” SP500 Granger causes rGDP

<table>
<thead>
<tr>
<th>RESTRICTED MODELS with the Dependent Variable as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCAP/rGDPpc09</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Lag8 69.392</td>
</tr>
<tr>
<td>Lag4 72.792</td>
</tr>
<tr>
<td>Lag2 72.986</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UN-RESTRICTED MODELS with the Dependent Variable as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCAP/rGDPpc09</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Lag8 59.764</td>
</tr>
<tr>
<td>Lag4 63.593</td>
</tr>
<tr>
<td>Lag2 63.357</td>
</tr>
</tbody>
</table>

- F-stat | ConLevel | F-stat | ConLevel |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag8 1.20825246</td>
<td>0.3097871</td>
<td>0.86418619</td>
<td>0.5513506</td>
</tr>
<tr>
<td>Lag4 2.45922859</td>
<td>0.0013694</td>
<td>1.54214831</td>
<td>0.1998832</td>
</tr>
<tr>
<td>Lag2 5.47121874</td>
<td>0.0013694</td>
<td>0.7531394</td>
<td>0.4828836</td>
</tr>
</tbody>
</table>

In all of the three of the above, we "REJECT THE NULL" that pSP90 < 0.05. 
If we "FAIL to REJECT" then we "accept" SP DOES NOT CAUSE rGDP.

Second 19 Years: No Granger Causality in Either Direction

<table>
<thead>
<tr>
<th>RESTRICTED MODELS with the Dependent Variable as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCAP/rGDPpc09</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Lag8 49.727</td>
</tr>
<tr>
<td>Lag4 51.532</td>
</tr>
<tr>
<td>Lag2 51.675</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UN-RESTRICTED MODELS with the Dependent Variable as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCAP/rGDPpc09</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Lag8 47.951</td>
</tr>
<tr>
<td>Lag4 47.446</td>
</tr>
<tr>
<td>Lag2 47.982</td>
</tr>
</tbody>
</table>

- F-stat | ConLevel | F-stat | ConLevel |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag8 1.175129678</td>
<td>0.3290585</td>
<td>1.328384002</td>
<td>0.322724</td>
</tr>
<tr>
<td>Lag4 1.646022257</td>
<td>0.3227892</td>
<td>2.138604668</td>
<td>0.187528</td>
</tr>
<tr>
<td>Lag2 2.77078046</td>
<td>0.0092984</td>
<td>0.771136595</td>
<td>0.494001</td>
</tr>
</tbody>
</table>

In all of the above, we "FAIL to REJECT" the null that pSP < 0.05.
If we "FAIL to REJECT" then we "accept" GDP DOES NOT CAUSE SP.
9 Summary of Results

One objective of this paper was to determine if there was any sense of causality between GDP and the equity markets in the USA over long time frames. This benchmark test relied on the same data errors made in many commonly cited papers. Accordingly, we were able, for the 57 year period studied, to support the hypothesis that nominal quarterly percent returns from the SP500 Granger cause quarterly growth in real GDP. In addition, we can reject the hypothesis that quarterly growth in real GDP Granger causes quarterly percent returns in the SP500. We are compelled to reinforce two aspects of this benchmark test. First, for consistency with other papers, we repeated the same data errors including using “nominal” returns from the SP500 in combination with inflation adjusted (real) growth in GDP. The second aspect is the minimal relevance a nearly 60 year period has with respect to business decisions, consumer decisions, and especially fiscal / monetary policy formulation.
9.1 Three Questions

The primary objective was to correct for the major errors found in many other papers and then determine if the series of “real” returns from the SPTR (our proxy for the total value of the equity market per Keynes / Tobin above) were indeed a leading indicator for real per capita growth in GDP over policy relevant time frames. Unfortunately for many in the finance and financial media business, the SP500 is no longer a leading indicator. In technical terms, the real returns from the SP500 no longer Granger cause growth in real per capita GDP over policy relevant periods.

Our findings beg three important questions. First, why is an obviously important variable, inflation commonly omitted from equity market returns while included in GDP growth, and second a related question, “which inflation measure is appropriate”? Our findings, when combined with the tendency of financial markets to “overshoot”, as extended from Dornbusch (1976), give rise to the third obvious question. Given the importance of the NIPA category of “Proprietors Income” and the existence of overshooting, if the goal of a model is to provide a reasonable basis for policymaking why is “Proprietors Income” commonly omitted.

Finally, we propose a possible solution for the “quarterly data” problem frequently noted in GDP related research. We previously discussed that official GDP data is quarterly and subject to sometimes significant revisions. However, we relied on vintages in the production of this paper. We now note the vintages are produced monthly and we note decisions (business, personal, and policy) are made or revised concurrent with those monthly vintages. This reality, in effect, reduces the commonly referred to GDP data
series to irrelevance as far as policy is concerned. Why? Because the data in the official series did not exist at the time of the decision(s). In addition, the monthly frequency of the vintages has the effect of increasing from 4 to 12 the number of observations per year (if we view each new vintage release as a new piece of information). This could, subject to additional research, improve forecasts as the nexus to decision making improves. In turn, this could improve economic policy decisions.

9.2 Chow Tests for possible breaks in the data

The benchmark test (the very long term test including uncorrected and common data errors) indicated the changes in the SP500 returns Granger cause rGDPpc growth. Interestingly, if we use the proper data (i.e. real SP500 returns) for the shorter (19 year) time frames, we find the real SP500 returns Granger cause rGDPpc growth until the mid 1980s. Subsequent to the mid 1980s, the causality first become neutral and then flips so the rGDPpc Granger causes rSP500 returns.

Given this steady progression of change in the direction of causality, we were compelled to see if the data exhibited statistically relevant “breaks”. A break, in common language, would be a material change in the relationship between the two data series. In effect, we could say that one relationship existed before a break and a second relationship existed after a break.

We ran the Chow test for rolling split dates from the beginning of the 57 year period until the end. The series of 188 unique tests revealed a lone cluster of “breaks” in the mid 1980s. Interestingly, this cluster of breaks coincides with the switching of
causality (from SP500 Granger causing rGDP to rGDP Granger causing SP500). The table below shows the results of the Chow tests from just before the cluster starts until just after it terminates.

There were 238 total observations (arranged in rows in Excel) with each test commencing at the 11th row (10th observation) due to the required lags at the start of the data set. The “Break Row” is the row (observation+1) identified by the Chow test as a structural break in the data. Therefore, “Row 107” (the first break discovered) corresponds to Q2_1983 and “Row 118” (the last break discovered) corresponds to Q1_1986.

<table>
<thead>
<tr>
<th>TestNum</th>
<th>BeginRow</th>
<th>BreakRow</th>
<th>EndRow</th>
<th>SSSE_c</th>
<th>SSSE_1</th>
<th>SSSE_2</th>
<th>ChowStat</th>
<th>Fdist pval</th>
<th>Result @ 10% Result @ 5%</th>
<th>BreakDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>11</td>
<td>100</td>
<td>239</td>
<td>128.7782853</td>
<td>78.14749031</td>
<td>42.94006166</td>
<td>1.489046295</td>
<td>0.15348133</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>51</td>
<td>11</td>
<td>101</td>
<td>239</td>
<td>128.7782853</td>
<td>81.28836775</td>
<td>40.62717715</td>
<td>1.317907777</td>
<td>0.22806341</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>52</td>
<td>11</td>
<td>102</td>
<td>239</td>
<td>128.7782853</td>
<td>84.70493105</td>
<td>38.42063878</td>
<td>1.076338378</td>
<td>0.38150744</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>53</td>
<td>11</td>
<td>103</td>
<td>239</td>
<td>128.7782853</td>
<td>84.76328721</td>
<td>37.30567592</td>
<td>1.288565791</td>
<td>0.24452306</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>54</td>
<td>11</td>
<td>104</td>
<td>239</td>
<td>128.7782853</td>
<td>84.99289829</td>
<td>36.96706865</td>
<td>1.315673344</td>
<td>0.23726313</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>55</td>
<td>11</td>
<td>105</td>
<td>239</td>
<td>128.7782853</td>
<td>85.00276575</td>
<td>36.94285089</td>
<td>1.313602945</td>
<td>0.23122176</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>56</td>
<td>11</td>
<td>106</td>
<td>239</td>
<td>128.7782853</td>
<td>85.14346663</td>
<td>35.52648934</td>
<td>1.575300478</td>
<td>0.12421001</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>57</td>
<td>11</td>
<td>107</td>
<td>239</td>
<td>128.7782853</td>
<td>85.73310997</td>
<td>34.06592626</td>
<td>1.748665696</td>
<td>0.07983142</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>58</td>
<td>11</td>
<td>108</td>
<td>239</td>
<td>128.7782853</td>
<td>85.91773887</td>
<td>33.85077664</td>
<td>1.763641999</td>
<td>0.07676777</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>59</td>
<td>11</td>
<td>109</td>
<td>239</td>
<td>128.7782853</td>
<td>86.00690602</td>
<td>33.40501197</td>
<td>1.839301762</td>
<td>0.06290567</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>60</td>
<td>11</td>
<td>110</td>
<td>239</td>
<td>128.7782853</td>
<td>86.80586811</td>
<td>32.77127434</td>
<td>1.803987607</td>
<td>0.06903635</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>61</td>
<td>11</td>
<td>111</td>
<td>239</td>
<td>128.7782853</td>
<td>87.25092928</td>
<td>32.47043834</td>
<td>1.772813208</td>
<td>0.07494473</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>62</td>
<td>11</td>
<td>112</td>
<td>239</td>
<td>128.7782853</td>
<td>87.28280788</td>
<td>32.4658478</td>
<td>1.767825028</td>
<td>0.07593145</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>63</td>
<td>11</td>
<td>113</td>
<td>239</td>
<td>128.7782853</td>
<td>87.32613326</td>
<td>32.46194941</td>
<td>1.799022156</td>
<td>0.07758021</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>64</td>
<td>11</td>
<td>114</td>
<td>239</td>
<td>128.7782853</td>
<td>87.39268329</td>
<td>32.38743843</td>
<td>1.774592384</td>
<td>0.07459529</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>65</td>
<td>11</td>
<td>115</td>
<td>239</td>
<td>128.7782853</td>
<td>87.33799189</td>
<td>32.38526568</td>
<td>1.773411309</td>
<td>0.07482625</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>66</td>
<td>11</td>
<td>116</td>
<td>239</td>
<td>128.7782853</td>
<td>87.38176555</td>
<td>31.94638474</td>
<td>1.856671426</td>
<td>0.06001223</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>67</td>
<td>11</td>
<td>117</td>
<td>239</td>
<td>128.7782853</td>
<td>87.70718507</td>
<td>31.89019839</td>
<td>1.799714487</td>
<td>0.06982024</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>68</td>
<td>11</td>
<td>118</td>
<td>239</td>
<td>128.7782853</td>
<td>87.74962146</td>
<td>31.8630199</td>
<td>1.797064091</td>
<td>0.07031059</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>69</td>
<td>11</td>
<td>119</td>
<td>239</td>
<td>128.7782853</td>
<td>89.01924561</td>
<td>31.80288585</td>
<td>1.54834823</td>
<td>0.13427591</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>70</td>
<td>11</td>
<td>120</td>
<td>239</td>
<td>128.7782853</td>
<td>89.21444827</td>
<td>31.7969467</td>
<td>1.504737897</td>
<td>0.14775489</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>71</td>
<td>11</td>
<td>121</td>
<td>239</td>
<td>128.7782853</td>
<td>89.8856583</td>
<td>31.68379583</td>
<td>1.389474365</td>
<td>0.19437266</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>72</td>
<td>11</td>
<td>122</td>
<td>239</td>
<td>128.7782853</td>
<td>89.89278772</td>
<td>31.67340944</td>
<td>1.390875159</td>
<td>0.19374024</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>73</td>
<td>11</td>
<td>123</td>
<td>239</td>
<td>128.7782853</td>
<td>89.91755802</td>
<td>31.2219632</td>
<td>1.478349746</td>
<td>0.15749287</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>74</td>
<td>11</td>
<td>124</td>
<td>239</td>
<td>128.7782853</td>
<td>90.65190551</td>
<td>31.2087581</td>
<td>1.338082587</td>
<td>0.22338306</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>75</td>
<td>11</td>
<td>125</td>
<td>239</td>
<td>128.7782853</td>
<td>90.97664399</td>
<td>30.78140874</td>
<td>1.355776284</td>
<td>0.21203608</td>
<td>Fail to Reject</td>
<td>Fail to Reject</td>
</tr>
</tbody>
</table>
We reserve investigation into possible causes of the mid-1980s break for later research. However, we suggest the catalyst for the break likely falls within one or both of a pair of secular changes with roots in the early 1980s. Importantly, these changes continued to evolve and accelerate through the 1990s and only momentarily slowed down during the deleveraging of 2007-10.

The first broad category focuses on changes in the SP500 construction. Prior to the 1980s, the population of stocks included in the index was relatively stable; however, beginning in the 1980s the index managers became much more active with respect to managing the component stocks. For example, we note two items: (1) there were major changes to the rules in the late 1990s to allow the managers to include high profile technology stocks, (2) the annual replacement rate increased dramatically in the 1990s. The second major index change essentially relaxed the definition of an “American” stock which, in turn, allowed for the inclusion of still more high profile (and volatile) stocks.

The second broad category addresses changes in financing techniques and regulations. Again, within this category are two major sub-categories: regulation and financing techniques. We note substantial deregulation of financial markets and entities during the late 1980s and 1990s contributed to the rise (and fall) of many new types of securities and derivatives. The net effect of much of the associated innovation was a substantial increase in securities-related leverage. Interestingly, while the putative goal of the innovation was to increase Main Street’s access to attractive financing, data appear to suggest Wall Street’s increase in leverage far outpaced Main Street’s. We speculate, if true, this relative change negatively impacted the ability of the SP500 to lead changes in
real GDP. Finally, we note the Federal Reserve (from 1987-2011) embarked on a long period of easy money. The vast body of research surrounding the unprecedented period of low real rates is conflicted about the net impact of these policies. Certainly, the low real rates provided some benefit to the real economy. However, recent research (including some from the regional Federal Reserve banks) have begun to question the efficacy of the “QE” programs with a particular focus on the relative benefits accruing to Main Street vs Wall Street.

Clearly, there is much to discuss regarding the reasons for the 1980s break; however, as note above – we reserve that topic for later research.

10 Possible paths for further research

We conclude with a few thoughts on potential avenues of research.

Data:

It seems reasonable to investigate a derivative of the “total market value” approach as noted in the Tobin/Keynes comment. An attempt to model the “total market value” by focusing on non-tangible book value. The “real tangible book value” approach has the added advantage of attempting to accommodate changes in leverage of the economy and financial assets.
It might also be interesting to look at the equity markets on a “real per capita” basis in the same manner as we consider real GDP. This could eliminate growth in the stock market resulting solely from population growth.

Of course, as noted in the previous section, if one’s goal is to forecast rGDP growth, then it seems relevant to include real Proprietors Profits.

**Time frames:**

If we acknowledge policies should be grounded in shorter term time frames (5 – 15 years) to encourage innovation, mitigate excessive shortages / surpluses, we should also recognize the relevance of 50+ year horizons to promote sustainable programs and enterprises. Accordingly, policy makers should consider short time frames as well as very long time frames. Given the “quarterly” nature of GDP data and common statistical guidelines, 10 years appears to be the lower limit for any technique requiring lagged data; however, as noted above, an interesting caveat to this 10 year limit arises from the monthly availability of GDP estimates (generally three consecutive months of a very first preliminary estimate, a second estimate eliminating most modeled inputs, and a final estimate). Acceptance of this idea could shorten enable statistically sound testing periods and, by extension, policy making cycles to 3 to 5 years.

**Methods:**

Granger causality, our principal tool in this paper, seems a bit restrictive and perhaps even a bit primitive for this area of study. Modifications have been made to the basic linear methodology used herein (one period ahead, linear model). Some of these
extensions allow for non-linear base models and multiple period ahead (either consecutive or “skipped” periods) forecasting.

In addition, Granger causality (and many other common methods) suffer from a bias toward significance (Lin et al, 2013) when the number of observations grows very large. Finding and applying a consistent method to distinguish between “statistical significance” and “practical significance” to re-test existing data would be very interesting.

Finally, other methods attempt to model the “memory” a market or economy might have regarding a specific event. Still other methods attempt to incorporate the divergence between forecasts of an event (e.g. an unemployment report) and the actual report. All of these other methods offer potentially interesting paths for additional research. This last grouping of models, when combined with policy relevant time frames and tangible book value + Proprietors Income data seems to have great potential with respect to forecasting changes in real GDP.
11 References


https://www.bea.gov/faq/?faq_id=488


https://blog.bea.gov/2015/06/10/a-snapshot-of-the-seasonal-adjustment-process-for-gdp-2/


https://www.federalreserve.gov/faqs/economy_14400.htm


Damodaran, A. (2017, March 15). *Variables used in Data Set*. Retrieved from Damodaran Online:

http://people.stern.nyu.edu/adamodar/New_Home_Page/datafile/variable.htm


Economic Index(c) for the United States. New York City, NY: The Conference Board.


NASDAQ. (2017). NASDAQ-100: Index Methodology. NASDAQ, INC.


APPENDIX 1

** THIS SUMMARY DATA PRINTOUT IS FOR **

** NOMINAL **
SP500 Quarterly Percent RETURNS

vs

Quarterly Changes in Per Capital Real GDP

(with associated lags for entire 1957-2016 term)

I. This first section summarizes tests run on Shazam with
quarterly growth in PerCapita Real GDP as the
Dependent Variable. "Restricted" models exclude
quarterly returns for Nominal SP500 as an independent
variable.

** GDP is DEPENDENT, RESTRICTED MODEL, LAG 8 **

OLS ESTIMATION

229 OBSERVATIONS      DEPENDENT VARIABLE = PCRGDP09

R-SQUARE = 0.1467      R-SQUARE ADJUSTED = 0.1157

VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.63092
** GDP is DEPENDENT, RESTRICTED MODEL, LAG 4

OLS ESTIMATION

229 OBSERVATIONS  
DEPENDENT VARIABLE: PCRGDP09

R-SQUARE = 0.1299  
R-SQUARE ADJUSTED = 0.1144

VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.63187

SUM OF SQUARED ERRORS-SSE= 141.54

AKAIKE (1974) INFORMATION CRITERION - AIC = 0.64566

DURBIN-WATSON = 1.9933

** GDP is DEPENDENT, RESTRICTED MODEL, LAG 2

OLS ESTIMATION

229 OBSERVATIONS  
DEPENDENT VARIABLE: PCRGDP09

R-SQUARE = 0.1293  
R-SQUARE ADJUSTED = 0.1216

VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.62669
** GDP is DEPENDENT, UN-RESTRICTED MODEL & NOMINAL SP500, **

LAG 8

OLS ESTIMATION

229 OBSERVATIONS DEPENDENT VARIABLE= PCRGDP09

R-SQUARE = 0.2327 R-SQUARE ADJUSTED = 0.1748

VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.58877

SUM OF SQUARES ERRORS-SSE= 124.82

AKAIKE (1974) INFORMATION CRITERION - AIC = 0.63231

DURBIN-WATSON = 1.9943

** GDP is DEPENDENT, UN-RESTRICTED MODEL & NOMINAL SP500, **

LAG 4
OLS ESTIMATION

229 OBSERVATIONS    DEPENDENT VARIABLE= PCRGDP09

R-SQUARE = 0.1968    R-SQUARE ADJUSTED = 0.1676

VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.59387

SUM OF SQUARED ERRORS-SSE= 130.65

AKAIKE (1974) INFORMATION CRITERION - AIC = 0.61718

DURBIN-WATSON = 1.9915

** GDP is DEPENDENT, UN-RESTRICTED MODEL & NOMINAL SP500, LAG 2

OLS ESTIMATION

229 OBSERVATIONS    DEPENDENT VARIABLE= PCRGDP09

R-SQUARE = 0.1928    R-SQUARE ADJUSTED = 0.1784

VARIANCE OF THE ESTIMATE-SIGMA**2 = 0.58617

SUM OF SQUARED ERRORS-SSE= 131.36

AKAIKE (1974) INFORMATION CRITERION - AIC = 0.59896

DURBIN-WATSON = 2.0057
II. This second section summarizes tests run on Shazam with quarterly returns from the Nominal SP500 as the Dependent Variable. "Restricted" models exclude quarterly growth in PerCapita Real GDP as an independent variable.

** SP500 NOM is DEPENDENT, RESTRICTED MODEL & rGDPpc, LAG 8

OLS ESTIMATION

229 OBSERVATIONS

DEPENDENT VARIABLE= PCNOMSP

R-SQUARE = 0.0274
R-SQUARE ADJUSTED = -0.0080

VARIANCE OF THE ESTIMATE-SIGMA**2 = 62.143

SUM OF Squared ERRORS-SSE= 13671.

AKAIKE (1974) INFORMATION CRITERION - AIC = 64.583

DURBIN-WATSON = 1.9985

** SP500 NOM is DEPENDENT, RESTRICTED MODEL & rGDPpc, LAG 4

OLS ESTIMATION

229 OBSERVATIONS

DEPENDENT VARIABLE= PCNOMSP

66
R-SQUARE = 0.0117  R-SQUARE ADJUSTED = -0.0059

VARIANCE OF THE ESTIMATE-SIGMA**2 = 62.016

SUM OF SQUARED ERRORS-SSE= 13892.

AKAIKE (1974) INFORMATION CRITERION - AIC = 63.369

DURBIN-WATSON = 1.9986

** SP500_NOM is DEPENDENT, RESTRICTED MODEL & rGDPpc, LAG 2

OLS ESTIMATION

229 OBSERVATIONS  DEPENDENT VARIABLE= PCNOMSP

R-SQUARE = 0.0113  R-SQUARE ADJUSTED = 0.0025

VARIANCE OF THE ESTIMATE-SIGMA**2 = 61.493

SUM OF SQUARED ERRORS-SSE= 13897.

AKAIKE (1974) INFORMATION CRITERION - AIC = 62.299

DURBIN-WATSON = 2.0002

** SP500_NOM is DEPENDENT, UN-RESTRICTED MODEL & rGDPpc, LAG 8

OLS ESTIMATION
** SP500_NOM is DEPENDENT, UN-RESTRICTED MODEL & rGDPpc, LAG 4 **

** OLS ESTIMATION **

229 OBSERVATIONS DEPENDENT VARIABLE= PCNOMSP

R-SQUARE = 0.0710 R-SQUARE ADJUSTED = 0.0008

VARIANCE OF THE ESTIMATE-SIGMA**2 = 61.598

SUM OF SQUARED ERRORS-SSE= 13059.


DURBIN-WATSON = 1.9996

229 OBSERVATIONS DEPENDENT VARIABLE= PCNOMSP

R-SQUARE = 0.0455 R-SQUARE ADJUSTED = 0.0108

VARIANCE OF THE ESTIMATE-SIGMA**2 = 60.984

SUM OF SQUARED ERRORS-SSE= 13417.

AKAIKE (1974) INFORMATION CRITERION - AIC = 63.379

DURBIN-WATSON = 2.0014
** SP500 NOM is DEPENDENT, UN-RESTRICTED MODEL & rGDPpc, LAG 2 **

OLS ESTIMATION

229 OBSERVATIONS DEPENDENT VARIABLE= PCNOMSP

R-SQUARE = 0.0239 R-SQUARE ADJUSTED = 0.0065

VARIANCE OF THE ESTIMATE-SIGMA**2 = 61.250

SUM OF SquARED ERRORS-SSE= 13720.

AKAIKE (1974) INFORMATION CRITERION - AIC = 62.587

DURBIN-WATSON = 2.0102