unpublished

An Empirical Examination of Interindustry Cross-Correlation

Aaron Pollack

Abstract
This report presents an examination of the relationship between sectors of the US economy by examining cross-correlation between the performance of a selection of sector-focused ETF's (exchange traded funds) from May 2006 through August 2018. We examine the ways in which the performance of a given industry is related to that of other industries and to the performance of the American economy as a whole. In doing so, we explore empirically, various economic relationships that exist between and among sectors and industries within the US economy.

Keywords
Time Series Analysis — Economics — Finance

Contents

	Introduction	**1**
1	**Data**	**1**
1.1	Data Preparation	2
2	**Methods**	**2**
3	**Results**	**3**
3.1	General Cross-correlations	3
3.2	Cross-correlations With Lag	3
4	**Discussion**	**4**
4.1	Cross-correlation With the Market	4
4.2	Interindustry Cross-correlation	5
4.3	Cross-correlation With Lags	6
5	**Sources of Error**	**7**
6	**Conclusion**	**7**
	References	**7**

Introduction

Interactions between sectors of the economy are of great interest to economists, financiers, and economic onlookers. An intuitive and empirical understanding of correlations between the behavior and performance of the various industries within the economy could prove both enlightening and financially profitable.

Intuitively, one could imagine that the performance of complimentary or adjacent industries would be closely related, or that one industry's performance might lead or lag that of the other. Indeed, as an example, personal expenditure on durable goods was seen to lead the performance of the economy for many decades. [1] Although in the current age of automation and globalization this effect is much less pronounced, for a time this indicator could credibly predict economic downturns.[2] The presence of such cross-industry economic effects suggests the presence of empirically visible cross-industry or cross-sector correlations.

We wish to explore, empirically to what degree – if at all – these relationships can be observed. To do this, we examine cross-correlation in the performance of ETF's (exchange traded funds) meant to track the performance of a specific industry or sector of the economy. Additionally, we will examine the lead and lag effects present in the correlations between the ETFs' performance and that of the general business cycle. The presence of such lead or lag effects might indicate that a given sector's performance could predict the performance of the general economy or that one sector's performance could predict that of another. Note that throughout our analysis, unless otherwise stated, the terms "sector" and "industry" are used interchangeably to refer to a sub-section of the economy related to the production of a specific, closely related set of goods or services.

1. Data

The data used for our analysis comes from the historic prices of various exchange traded funds (ETF's). ETF's are securities which, like mutual funds and index funds, are designed to track the performance of a specific index, commodity, bond, equity market, industry, or any other well-defined basket of securities. However, unlike mutual funds and index funds, ETF's are freely traded on exchanges and so have comparatively higher liquidity while investors enjoy lower berries to entry and exit. In theory, the performance of an ETF closely tracks the performance of it's underlying assets.

This study makes user of a collection of ETF's offered by Black Rock as part of their iShares line of investment products. These specific funds were chosen for their offering broad, well diversified exposure to various sectors throughout the economy. As this study seeks to examine correlations specifically within the American economy, ETFs's with a strong or exclusive focus on exposure to the American economy were chosen where available.

Note that most, though not all, of the ETF's chosen represent exclusive exposure to the United States market. Three funds include exposure to other world markets: iShares Transportation, iShares Semiconductor, and iShares Biotechnology. Note, however, that in all three cases these funds have a primary focus on US markets therefore their performance should be instructive of the performance of their respective sector within the American economy.

Table 1. Table of Sector-focused Securities

Fund Name	Ticker Symbol
iShares U.S. Telecommunications ETF	IYZ
iShares U.S. Pharmaceuticals ETF	IHE
iShares U.S. Consumer Goods ETF	IYK
iShares U.S. Basic Materials ETF	IYM
iShares Transportation Average ETF	IYT
iShares U.S. Industrials ETF	IYJ
iShares U.S. Consumer Services ETF	IYC
iShares U.S. Energy ETF	IYE
iShares U.S. Financials ETF	IYF
iShares PHLX Semiconductor ETF	SOXX
iShares U.S. Healthcare ETF	IYH
iShares U.S. Technology ETF	IYW
iShares U.S. Real Estate ETF	IYR
iShares Nasdaq Biotechnology ETF	IBB
iShares U.S. Aerospace & Defense ETF	ITA
iShares U.S. Utilities ETF	IDU
iShares U.S. Insurance ETF	IAK

To capture a measure of the performance of the baseline business, a broadly diversified measure of the entire US stock market was needed. Luckily, many investment firms offer investment vehicles meant to track the performance of the stock market in general. This study makes use of iShares Core S&P Total U.S. Stock Market ETF (ITOT) a fund which includes over 3000 underlying equities and offers broad exposure to the entire US equity market.

In all cases the metric used to measure the "value" of the fund was the *Net Asset Value* which measures the net value of the fund (tabulating held securities, cash assets, liabilities, etc) per outstanding share of the fund. Being a more direct measure of the value of the underlying securities in the fund, this metric was chosen over simple market price for it's being more readily comparable between ETF's.

Data was retrieved for all securities between the dates of May 1, 2006 and August 13, 2018 – 3094 trading days in total. We are left with a collection of time series which, in theory, closely track the performance of various industries within the US economy as well as a metric which measures the performance of the US economy as a whole.

1.1 Data Preparation

Assessments of cross-correlations in time series data frequently take great care to mitigate the effects of spurious correlations arising from underlying trends shared between the time series. [3], [4] Data is typically prepared by attempting to remove a trend from the data. [5] However, this study explicitly wishes to asses cross-correlations in underlying trends shared between our data sets. For this reason, we will asses the presence of cross-correlations between time series both with their trends removed and with their trends isolated.

Whether we wish to use the trend-isolated data or the detrended data, we shall need a method to isolate the stochastic component from the underlying trend of our time series. Were we able to make an assumption of an underlying trend within our data, this would be relatively easy and statistically robust. Unfortunately financial data does not offer the luxury of such assumptions. No polynomial model, linear or otherwise, is expected to fit our data. For this reason, we shall make use of a *100 day moving average* to isolate the underlying trend from the stochastic elements of our data. Such a model allows us to relax any assumptions concerning the degree of the underlying trend present in our time series. The 100 day moving average effectively isolates the underlying trend of the time series from the stochastic residual component of the time series.

$$observation = trend - residual$$

2. Methods

There exist several widely-used measures of correlation between variables. This study makes use of the Pearson, Spearman, and Kendall correlation coefficients. While all three metrics measure to what degree two data sets co-very from their respective means, each offer slightly different properties and comparing results across all three offers deeper insight into the underlying properties of our data.

Pearson's Correlation Coefficient The first, most basic, coefficient metric is the Pearson correlation coefficient which measures to what degree there exists a *linear* relationship between the two variables, one wherein the two variables change at fixed rate. In practice, we expect any cross-correlations in our data to be more complicated that a simple linear relationship, but Pearson's metric offers a solid starting point.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Spearman's Rho The second correlation metric is Spearman's rank correlation coefficient which describes correlations with somewhat more nuance. While Pearson's coefficient describes only linear relationships, Spearman's Rho captures any *monotonic* relationship, regardless of the degree of the underlying model, by assessing correlation in the rank-ordering of the two variables. The assumption of a linear relationship between the variables is relaxed; Spearman's rho captures any relationship in which the two variables co-vary, not necessarily at a fixed rate.

$$r_{X,Y} = \rho_{rk(X),rk(Y)} = \frac{cov(rk_X, rk_Y)}{\sigma_{rk_X} \sigma rk_y}$$

Kendall's Tau Finally, we make use of Kendall rank correlation coefficient. Like Spearman's Rho, this measure of correlation really measures the relationship in the ranked list of values for the two variables. Kendall's tau measures the number of pairs observations of X and Y who's ranks match (said to be *concordant* pairs) as a fraction of the total number of pairs. Like Spearman's rho, Kendall's tau captures any monotonic relationship between the variables; the assumption of linearity is similarly relaxed.

$$\tau = \frac{(concordant - pairs) - (discordant - pairs)}{n(n-1)/2}$$

3. Results

Our analysis is divided into two sections. Section one includes analysis of cross-correlation between the performance of each sector ETF with that of the business cycle (represented as the total stock market ETF). Section two concerns analysis of leading and lagging effects present in the cross-correlations. That is to say, to what degree if at all does one sector's performance tracks that of another sector – or the market – but at a lag.

3.1 General Cross-correlations

Table 2. Cross-correlations: 100 Day Moving Average

	Correlation Metrics		
ETF Ticker	Pearson	Spearman	Kendall
IYZ	0.77	0.85	0.64
IHE	0.91	0.88	0.71
IYK	0.96	0.94	0.81
IYM	0.83	0.79	0.61
IYT	0.97	0.93	0.80
IYJ	0.98	0.97	0.86
IYC	0.97	0.96	0.85
IYE	0.34	0.34	0.25
IYF	0.70	0.73	0.59
SOXX	0.94	0.96	0.84
IYH	0.96	0.95	0.82
IYW	0.97	0.92	0.80
IYR	0.73	0.92	0.65
IBB	0.91	0.88	0.70
ITA	0.97	0.94	0.83
IDU	0.94	0.93	0.80
IAK	0.85	0.84	0.67

Tabulated in table 2 are the correlations between each respective ETF's underlying trend and that of the total stock market, represented by the iShares total stock market ETF described above. Three different types of correlation metrics are reported. Note that, in the table above, the time series representing each ETF is the 100 day moving average of the daily net asset value of each security. Therefore, the correlation metrics should be interpreted as an indicator of how closely the *underlying trend of the ETF* mirrors that of the market.

Table 3. Cross-correlations: Detrended

	Correlation Metrics		
ETF Ticker	Pearson	Spearman	Kendall
IYZ	0.024	-0.035	-0.022
IHE	0.072	0.145	0.115
IYK	0.102	0.110	0.077
IYM	0.160	0.047	0.035
IYT	0.267	0.189	0.132
IYJ	0.311	0.267	0.193
IYC	0.384	0.310	0.218
IYE	0.063	0.040	0.030
IYF	0.343	0.269	0.187
SOXX	0.481	0.448	0.328
IYH	0.301	0.365	0.260
IYW	0.481	0.415	0.301
IYR	0.125	-0.012	-0.006
IBB	0.159	0.267	0.199
ITA	0.475	0.463	0.339
IDU	0.148	0.141	0.098
IAK	0.272	0.213	0.150

Tabulated in table 3 are the cross-correlation metrics of each time series after they've been detrended by subtracting the 100 day moving average leaving only the stochastic residuals. These values can be interpreted as measure of to what degree, if at all, the two time series are correlated in absence of any effects arising from shared underlying trends. As we can see from the table, the answer of "to what degree" is, frequently, "not at all". The cross-correlations across the detrended data are observed to be much weaker than than those in the 100 day moving averages.

3.2 Cross-correlations With Lag

Next, we consider lag effects. Lag effects describe a relationship between two time series wherein one of the time series has been shifted backwards or forwards by a set number of periods. In this case, we wish to consider the presence of lag effects in the cross-correlation of our ETF's. Table 4 tabulates, for for each sector focused ETF, the maximum observed cross-correlation with the ITOT Total Market ETF, and at what lag this correlation was observed.

Note that a negative lag value indicates that the sector ETF time series was shifted *backwards* by n days while a positive lag value indicates the opposite. In all cases, cross-correlations were computed using the moving averages of the

Table 4. Maximum Cross-correlations With Lags

ETF Ticker	Lag (Days)	Correlation (Spearman)
IYZ	-44	0.831
IHE	284	0.920
IYK	-27	0.950
IYM	-68	0.829
IYT	-39	0.950
IYJ	-53	0.986
IYC	-8	0.962
IYE	-50	0.374
IYF	-255	0.834
SOXX	-41	0.972
IYH	-44	0.965
IYW	-37	0.932
IYR	-242	0.863
IBB	-299	0.900
ITA	-61	0.970
IDU	-82	0.964
IAK	-109	0.888

net asset value for each security and the Spearman correlation metric.

4. Discussion

This study is intuitively motivated by the notion that certain sectors of the economy would be more closely linked to the performance of the business cycle than others, that certain pairs of sectors might be closely linked, and that certain sectors and industries might closely follow the performance of the business cycle but at a lag. To varying degrees, these intuitive assumptions are demonstrated empirically by our analysis.

4.1 Cross-correlation With the Market

The first component of our inquiry concerned how the performance of various industries, represented by the performance of their respective ETF's, is related to the performance of the economy in general, represented by the iShares Total Stock Market ETF.

Comparing Linear and Monotonic Correlations

We made use of three different correlation metrics, each of which offers a slightly different representation of the relationship between the two time series. Recall that Pearson's correlation coefficient characterizes the degree of *linear relationship* (moving up of down together at a fixed rate) between the two series. The Spearman and Kendall coefficients, by contrast, characterize the degree of *monotonic relationship* – moving up or down together but not necessarily at the same rate. In short, if the Pearson statistic is higher than the Kendall and Spearman statistics, than the relationship is more strongly linear and is characterized by a fixed rate of change between the two series.

Given the complexity of financial data, we would expect any relationship to resist the simplicity of linearity. In practice, however, we observe that many of our time series exhibit more strongly linear correlations than monotonic correlations.This suggests that, for the most part, cross-correlations between sectors and the market – particularly with respect to underlying trends – are characterized by linearity more readily than previously thought.

Trend-isolated Correlations

We conducted our correlation analyses on both the underlying trend of each time series (represented by a 100 day moving average) and on the isolated stochastic residuals of each time series (represented as the detrended time series, after removing our 100 day average).

The sectors' correlations with the market were strikingly high for the 100 day moving averages. This makes some intuitive sense. We expect each individual industry of the economy to respond to the same broad economic trends, therefore we would expect a significant amount of co-relation in the underlying trends of each time series. The industries which displayed the weakest correlation with the market were **Financial Services** (represented by IYF) and **Energy** (represented by IYE).

In the case of the financial services industry, the relatively week correlation observed with the wider market could be related to the financial crisis of 2008. As our data spans 2006 to 2018, the effect of the financial crisis looms large over the period surveyed. Given this, the relatively weak correlation between the general market and the performance of the financial industry could reflect a divergence in performance between the wider market and the Financial Services industry brought on by the Great Recession – which effected the financial services industry most deeply. Alternatively, it could reflect a general, broad detachment between the performance of the market and the financial industry specifically.

The relationship between the performance of the market and that of the Energy industry is also strikingly weak. The low correlation could reflect the industry's heightened sensitivity to exogenous demand and supply shocks. As an illustrating example, consider how the price of oil is, at best, weakly related to the performance of the American or world economy. The energy industry – being especially sensitive to the price of oil and other commodities, especially during the time frame surveyed – would therefore be less closely correlated with the performance of the American economy.

Detrended Correlations

Our analysis of the detrended data provides further insight into each industry's correlation with the general market. Here, with the underlying trend removed, the correlation values characterize the degree to which the two time series are are still co-related. In theory, the observation of a high degree of correlation in the residuals would indicate an especially high sensitivity to the underlying market or an especially strong relationship between the given sector and the wider economy.

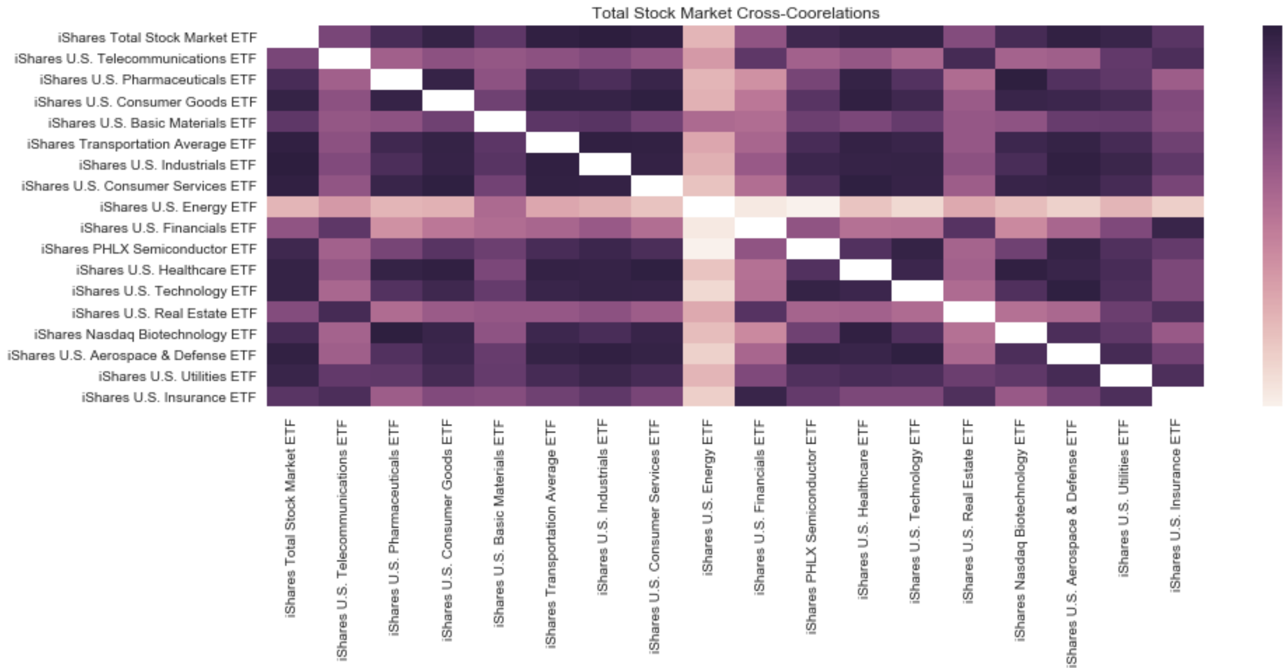

Figure 1. A heatmap depicting cross-correlation between the moving-averages of each pair of ETF's. Note that the white diagonal corresponds to the intentionally omitted values measuring each time series' correlation with itself.

As expected, the correlation values are much lower than those observed for the moving average time series. Here, we see the most highly correlated industries are **Technology** (represented by IYW), **Semiconductors** (represented by SOXX) and **Consumer Goods** (represented by IYC).

Intuitively, one could imagine that a heavily service-focused industry like Information Technology would be especially sensitive to the performance of the general market. Alternatively, it may be that Information Technology – being of eminent importance in the growth of the American economy – closely follows economic growth simply because it's responsible for so much of that growth. Regardless of the reason for the industry's appearing so closely tied to the performance of the market, it makes intuitive sense that the fate of an industry like Semiconductors – which produces the components used in the computers that make information Technology possible – would be closely tied to that of the Technology sector in general.

The high degree of correlation between the general market and the market for consumer goods likely reflects the sensitivity of consumers to economic boom and bust. The buying patterns of everyday consumers are closely tied to the performance of the wider economy and therefore (although the two are, of course, not the same thing) the total stock market. A good rising stock market suggests that consumers, in theory, have more money to spend and feel more confident doing so. A market downturn downturn, by contrast, would stifle consumer spending and dampen the performance of the consumer goods industry. This duality would lead to an exceptionally close relationship between the performance of the consumer goods industry and that of the wider stock market.

The observations described above are only some of the intuitive motivated, economic phenomena represented empirically by our analysis.

4.2 Interindustry Cross-correlation

Although the primary purpose of our investigation is to determine the presence of cross-correlation between industries and the economy, our data set allows us to asses the presence of cross-correlations between the industries themselves. Figure 1 depicts the cross-correlations between each industry and the total stock market as well as between each industry pair. Recall that the underlying data used to generate the plot is the detrended, 100 day moving averages discussed in section 1.1.

We observe interesting corollaries to the empirical effects previously discussed in section 4.1. We noted earlier that the **Energy Sector** seems to be especially weekly correlated the total stock market. We observe that this trend is mirrored when assessing the performance of the Energy industry vis-a-vis that of the other industries. Recalling our economic intuition described earlier, this effect could be due to the relative sensitivity of the energy industry to exogenous supply shocks and rapid changes in the prices of key inputs (like oil), making the industry's performance more closely tied to forces outside the scope of our investigation than to economic forces represented by the other industries.

A few industry pairs seem to exhibit especially interesting behavior. Specifically, We observe that the **Real Estate** industry, and the **Telecommunications** industry, which exhibit relatively weak correlation over all, appear relatively highly correlated with each other.

Figure 2. A heatmap depicting cross-correlation, including lags measured in days, of the moving-average for each ETF with that of the total stock market. Note that a positive lag indicates a shift *forward* of the sector-focused ETF while a negative lag indicates a shift backwards of the same

Similarly, the **Financial Services** industry appears most highly correlated with the **Real Estate** industry and with the *Insurance* industry. In the case of Financial Services and Real Estate, the industries appear weekly correlated with their peers yet comparatively strongly correlated with each other. One imagines that this reflects the effects of the 2008 financial crisis wherein these three industries were most deeply effected. As such, these two industries appear are more closely linked to each other than with with any other industry under consideration. Whether this apparent correlation is a simple fluke due to the limited pre-crisis data included in our or, or representative of a deeper economic relationship, is a question that requires further study to answer.

4.3 Cross-correlation With Lags

Finally we consider the presence of lag effects in our data. Specifically, we wish to examine to what degree the performance of a given industry leads or lags that of the total stock market. Table 4 lists, for each ETF, the lag (in days) at which the highest correlation was observed. In theory, this gives us some idea as to the size and direction of the lag effect between the two securities. Figure 2 depicts a more comprehensive picture of the presence of leading and lagging effects by plotting each industry's correlation with the stock market at each of various lags. Recall that, for the purposes of our analysis, a positive lag value represents a shift in the *Sector ETF* by n days forward while a negative lag value indicates a shift backwards n days.

We note first that most of the lag effects appear to be negative. That is to say, in many cases, the performance of the given ETF is most closely correlated with *past* market performance. In the case of **Industrials**, **Semiconductors**, and **Aerospace & Defense** we observe a lag of roughly 90 days. This indicates that the performance of these industries is most closely correlated with the performance of the market 90 days ago. Or, put another way, what the market does today these industries will do in roughly 90 days. The clearest, intuitive reason for this empirical observation is that these industries – heavy industries, all – make plans on a very long time horizon and are relatively less-able to react to sudden market shifts. On the supply side, in the case of all three industries, making significant shifts in production capacity and developing new products takes months of years. If Boeing, a major Aerospace firm, wants to deploy a new factory to react to increased market demand for its planes, doing so will take years. If, during an economic slowdown, demand for planes decreases and so Boeing wishes to scale down, this too will take months at least. On the demand side, many consumers of the products of heavy industry are, themselves, operating on long time horizons and so will not or cannot change their consumption patterns quickly in response to economic downturn. As an illustrating example, if the economy suddenly declines, causing few people to buy airline tickets, Delta Airlines – a major airline and consumer of Boeing jets – cannot quickly adjust their purchases of airplanes from Boeing. In short, the performance of heavy industries, which have large, sunk costs in physical capital, long production time lines, and who's clients are unable to react quickly to market forces due to the nature of their product are seen to lag the performance of the general stock market significantly.

Relatively few industries appear to lead the market. The most striking exceptions are the **Consumer Goods** industry and the **Consumer Services** industry. Although, in the case of both industries, the *maximum* lag effect was observed for a negative lag, a more consistent, higher degree of correlation is observed for positive lags than for no lag. This observation is more readily illustrated in figure 2. Both industries appear to lead the market by between 30 and 60 days, although the effect is much more pronounced for the consumer goods industry. The presence of a leading effect – a "positive lag" effect for which $n > 0$ – would indicate that the performance of a given industry is most closely related with what the market *will do* in the future. Put a different way, a positive lag effect suggests that the performance of the industry today is correlated with the the performance of the total stock market in the future.

In the case of the two industries listed above, one intuitive economic explanation for the observed behavior centers around the effect that consumer spending has on Gross Domestic Product. Although consumer spending represents only a small portion of our analysis here, it represents a significant fraction of the US economy. Given that consumer spending contributes – directly and indirectly – such a significant portion to the American economy, one could imagine that a decline in consumer focused industries would signal an upcoming downturn in the general market as the effects of an industry-isolated downturn are transmitted throughout the economy.

5. Sources of Error

The empirical methods used in this investigation are quite simple. Despite this – or perhaps because of this – various sources of error exist in our approach which must be highlighted here.

The first, and most obvious, source of error concerns a foundational assumption made by this study – that the Net Asset Value of an exchange traded fund meant to track the performance of a given sector accurately measures the performance of that sector in the context of the greater economy. Although the financial instruments chosen for use in our analysis were chosen specifically because they are designed to closely track their respective industry's performance through broad exposure and diversified asset holdings, it is certainly the possible the performance of a sector might not be well reflected in the performance of a given sector-focused ETF.

Secondly, it is possible that the Total Stock Market ETF chosen to represent the performance of the economy as a whole does not, in fact, do so. The ITOT security was chosen for its offering a tremendous amount of diversified exposure to the American stock market. However, as is frequently pointed out to economic and political pundits, the stock market is not the same thing as the economy. In interpreting our results, one should be mindful of the implicit assumption being made that the performance of the total stock market is an acceptable stand-in for the performance of the total economy and of the business cycle; an assumption which, though convenient to our analysis and potentially empirically valid means that large swaths of the American economy – small businesses, public sector firms, privately held firms, etc. – are not included in our study.

The third source of error concerns the manner in which we prepared our data before running correlations between the time series. Our use of trend-isolated moving-average time series may have caused our empirical analysis to fall pray to the perils of spurious correlations. Our intention for using moving averages was to explicitly analyze the cross-correlation effects between the underlying trends shared across the economy. However, these moving average correlations, specifically, may not be interpretable as we expect.

Finally, a word must be said about the intuitive economic explanations discussed in section 4. Our intention for offering possible economic explanations for the observed data should not be confused with an assertion of causality. This study does not attempt to characterize any causal relationships between correlations in the performance of the economic sectors nor to assert any causal roots to the empirical phenomena observed.

6. Conclusion

In this report we have outlined our approach for assessing the cross-correlation effects between various sectors of the American economy, as represented by the Net Asset Value of a selection of sector-focused ETF's designed to offer an investor broad exposure to a specific sector or industry within the market. We correlated the performance of these ETF's with that of a total stock market ETF meant to represent the performance of the entire economy. We explored correlations on trend-isolated and detrended time series data sets. Finally, we assessed the presence of lag effects in our correlations – exploring whether a given sectors performance today mirrors the performance of the market in the past, or might suggest the market's performance in the future.

Although every attempt has been taken to ensure that the empirical observations used in our analysis are meaningful and interpretable, nothing in this report should be considered sound investment advice. Our attempt was not to offer such advice, nor do we feel that we have done so. Our analysis gave rise to various empirical observations which could, in theory, be explained intuitively by underlying economic forces. Taken together our analysis represents an attempt to characterize the historic dynamics between and among various sectors and the American economy.

References

[1] U.S. Bureau of Economic Analysis. Personal consumption expenditures: Durable goods. ret, August 18, 2018.

[2] Victor Zarnowitz. *Business Cycles: Theory, History, Indicators, and Forecasting*, chapter Composite Indexes of Leading, Coincident, and Lagging Indicators, pages 316 – 356. 1992.

[3] G Udny Yule. Why do we sometimes get nonsense-correlations between time series? a study in sampling and the nature of time series. *Journal of the Royal Statistical Society*, 89:1–63, 1929.

[4] Søren Johansen. Correlation, regression, and cointegration of nonstationary economic time series. 2007.

[5] Jason Brownlee. How to use and remove trend information from time series data in python. 2016.