Abstract
This paper sheds light on the relationship between the Nasdaq Composite Index and a newly proposed Energy Futures Conditions Index (EFCI). While various financial conditions indices provide information about the financial stability of a country, the existence of an energy condition index, using futures markets, is scarce. Using weekly data over the period 1992–2017, this paper introduces an energy futures index using principal component analysis and test its predictability over the Nasdaq Composite Index. The EFCI captures 95% of the variability inherent in crude oil, heating oil and natural gas futures’ total reportable positions. Stability in forecast errors over different lags suggests a one week lag is sufficient to forecast weekly Nasdaq Composite Index. 95% prediction levels support that the estimated model captures actual equity market index values, except for the 2000 technology bubble. Distributions of level data were non-normal, not serially correlated and homoscedastic under the whole sample period, with diagnostics on pre and post technology bubble crisis showing mixed results. While differencing ensured homoscedastic errors in the forecasting model, Granger causality supported non-causality from both energy futures and equity markets, suggesting no evidence of cross market information flows.

Keywords
Nasdaq Composite Index, energy futures markets, equity market

JEL Classification
G15, G18, Q47

INTRODUCTION
The role of speculators in globalized markets can arguably be traced back to early studies like Kaldor (1939), Working (1953), Nurske (1944) and Friedman (1953). While the first two authors postulated that speculators are market destabilizers by allowing for speculative decisions based on other players’ behavior, the latter two authors support that speculators can help in providing liquidity thereby decreasing volatility in markets. Despite some authors like Meese and Rogoff (1983) find it hard to explain movements in some currency futures markets, Houthakker (1957) and Yoo and Maddala (1991) found speculators in commodities markets to be more profitable. In contrast, Hartzmark (1987) and Khoury and PERRAKIS (1998) found hedgers to pick the future direction of prices better than speculators. While studies like FIGLEWSKI (1981) and Santoni (1987) looked at the effect of futures prices on spot prices, the first study found higher volatility in post futures periods compared to lower volatility in the second study. Following major events in the 1990s, Gurrib (2009) reported that the volatility in the biggest futures markets players’ positions gradually waned off, suggesting that they react well to news. More recently, Gurrib (2018) analyzed the relationship between leading currency futures and major financial conditions indexes and reported that only Chicago’s National...
Financial Condition Index (NFCI) was able to forecast the weekly Japanese Yen net positions. Aggarwal (1988) found an increase in volatility following the introduction of futures markets and also volatility increases, over time, where no futures markets were introduced. This suggests that futures markets are not necessarily linked to volatility in other markets, and that financial conditions can drive volatility in markets.

Support of the relationship between financial markets regulation and financial conditions can be found in studies like IMF (2017), Adrian, Boyarchenko, and Giannone (2016), Koop and Korobilis (2014), Dudley (2010), Aramonte, Jahan-Pavar, Rosen, and Schindler (2017), where the latter propel that FCIs are generally correlated and can yet yield different values on financial conditions. While FCIs might be constructed differently, they largely consist of financial products ranging from equity returns to risk on debt instruments. Kliesen, Owyang, and Vermann (2012) summarized the major variables in US FCIs. Despite studies that focus on the relationship between energy futures conditions and equity markets movements is scarce, some studies looked at the relationships between net positions of key market players in specific futures markets and risk and return. For instance, Gurrib (2009) adopted ARCH models to predict the largest speculators’ and hedgers’ positions and reported these models to be weak in forecasting one month return. Gurrib (2008) looked at the impact of major macroeconomic events on futures markets key players’ net positions and found short lived structural breaks. Gurrib (2018) found that current and lagged FCIs values and the most actively traded USD based currency pairs share low correlations.

The focus of this study on energy futures markets is backed by IBRD (2017), which reported that US was the biggest consumer and producer of crude oil and natural gas in 2016. Despite China has surpassed the US in terms of crude oil imports, the US crude oil market, natural gas and heating oil markets remain among the most actively futures markets, with the New York Mercantile Exchange (NYMEX) leading other exchanges such as the Intercontinental Exchange (ICE). The study builds onto the current literature on various grounds. First, although research on global financial conditions and equity markets exists, the relationship between energy futures markets, equity indices and financial conditions is rare or non-existent. It is the first paper to analyze if the largest hedgers and speculators’ total reportable positions, embedded through a proposed Energy Futures Conditions Index, can affect major US stock market indices such as the Nasdaq Composite Index. Findings from this paper are critical, since it helps to shed light whether the biggest players’ transactions, through reportable positions in the energy futures markets, can potentially affect stock market index movements. This allows regulatory bodies such as the Commodity Futures Trading Commission (CFTC) and the Securities Exchange Commission (SEC) to be more informative in their mandate of promoting greater price stability in financial markets. With the US being among the top two consumers and producers of crude oil and natural gas, the findings help to ascertain whether significant information flows between the US energy futures markets and the equity market. The remaining part of the paper provides some review of existing literature, which is followed by the methodological approach adopted and data used. The analysis section then follows with descriptive statistics, forecasting results, including diagnostics, before ending with a conclusion.

1. LITERATURE REVIEW

The International Monetary Fund (IMF), as an international body fostering financial stability, suggested in IMF (2017) that globalization can potentially obfuscate the administration of domestically based financial conditions, particularly in instances where local economies have gradually transformed into global economies, such that external factors need to be considered when pursuing domestic goals. Although the Organization for Economic Cooperation and Development (OECD) and IMF construct FCIs based on nationally localized data, global financial conditions are led by the US, which remains the leading sage in the current global monetary system. For instance, Rey (2013) showed that US FCIs and the Chicago Board of Exchange Volatility Index (CBOE VIX)
are strongly positively correlated. In the same manner, IMF (2014) supports that the US Dollar is vested as an international currency with critical functions in the finance arena.

More importantly, the essence of a stable financial system compared to one witnessing financial stress needs to be documented. Hakkio and Keeton (2009) succinctly reported the features surrounding financial stress, which is defined as noise to the usual functioning of financial markets. Although every period of financial disruption can be dissimilar, they noted some common traits such as an enhanced uncertainty regarding fundamental asset values, more uncertainty about how one believes other investors behave, an increase in information asymmetry, and a shift towards less risky yet more liquid investments. The reliance on subjective guesses on how other investors’ make decision as opposed to using fundamentals as key metrics result to a surge in price fluctuations. While jumps in asymmetric information can be validated with investors dropping their self-confidence on the quality of lenders’ credit ratings and lenders having more challenges to decide on the true quality of borrowers, the flight to quality during a period of instability, lead to an expectation of lower return for safer assets holders. Caballero and Kurlat (2008) found such times are typically followed by a rise in borrowing costs for riskier borrowers. Similarly, illiquid assets’ lenders mostly endure higher borrowing costs during financial disruptions to reward risk takers for the heightened risk of continuing to hold their current investments.

With the importance of financial stability justified, it is beneficial to capture that FCIs have been constructed using different methods like macroeconomic models (see Beaton, Lalonde, & Luu, 2009), Vector Auto Regressive models (VAR) (see Swiston, 2008), and Principal Component Analysis (PCA). As the paper uses PCA initially, laying down previous studies covering FCI using PCA is warranted. Hatzius (2010) adopted PCA to choose the first principal component and used it as the FCI variable to predict economic growth. Similarly, Gomez (2011) extracted the main ingredient from indicators such as exchange rates, asset prices and interest rates, and constructed an FCI for Colombia using variance probability of the principal components as weights. The use of PCA in this study is moved by the capacity of the technique to capture most variability in major energy futures contracts under uncorrelated components dubbed as principal components. The principal component(s) can then be used to test the predictability of the stock market index.

The study is motivated by previous research work, which looked at spill-over effects in equity and energy markets. While Lin and Tamvakis (2001) found substantial spill-over effects among crude oil markets, Hamao, Masulis, and Nag (1990), King and Wadhwa (1990) supported the same but among stock markets. Panagiotidis and Rutledge (2007) provided evidence that UK crude oil and natural gas shared long-term pricing relationships. Gurrib (2018) modelled a financial condition index by combining various major US FCIs and tested its forecasting ability over the most active foreign currencies and reported major financial condition indices to be weak in predicting foreign currency spot values. Bessembinder and Chan (1992) modelled the use of economic variables like equity dividend yields and rejected the hypothesis that futures and equity markets contain different risk premia. Our study closes the gap in that it is the first to introduce an energy index based on the energy futures markets’ largest players and assess if it can be used to predict 1 week ahead stock market index values.

2. RESEARCH METHODOLOGY

Basically, PCA is a method, which targets a reduction in the dimensions that connect variables, whilst retaining most of the variability among the variables. Alternatively stated, it is a mathematically based process, which transforms correlated variables into a number of uncorrelated ones dubbed as principal components. The first principal component captures the highest variation in the data, followed by the second principal component and so on. The PCA model is centered on eigenvalues and eigenvectors, where the former represents the variance of all variables accounted by a factor and the latter accounts for a scaled direction of a non-zero vector as follows:

$$|A - \gamma I| = 0,$$  (1)
\( (A - \gamma I) \varphi = 0, \) \hspace{1cm} (2)

where \( A \) is a square matrix in the form of
\[
\begin{pmatrix}
cov_{1,1} & cov_{1,2} \\
cov_{1,2} & cov_{2,2}
\end{pmatrix},
\]

\( \varphi \) is a vector, \( \gamma \) is a scalar that satisfy equation 2, and \( I \) is an identity matrix. The eigenvalues of \( A \) are calculated from the determinant of equation 1, followed by eigenvectors \( \varphi \) for each eigenvalue, by using reduced matrix to row echelon form
\[
\begin{pmatrix}
a & \cdots & b \\
0 & \ddots & \vdots \\
0 & 0 & c
\end{pmatrix}
\]

and reduced matrix to reduced row echelon form
\[
\begin{pmatrix}
1 & \cdots & b \\
0 & \ddots & \vdots \\
0 & 0 & 1
\end{pmatrix}
\]

\( cov_{1,1} \) and \( cov_{2,2} \) represent the variance of specific FCIs, while \( cov_{1,2} \) represents the covariance between any two FCIs. To identify periods, which have witnessed large fluctuations, the FCIs are scaled by their respective standard deviations, after having been demeaned. For instance, an index value of –3 (3) is associated with financial conditions that are tighter (looser) than on average by three standard deviations. This common approach of standardization can also be found in Cardarelli, Elekdag, and Lall (2011), Nelson and Perli (2007). The uncorrelated linear combinations of standardized variables form the principal components as follows:
\[
\sigma_{PC_1} > \sigma_{PC_2} > \sigma_{PC_3} \ldots > \sigma_{PC_n},
\]

where \( \sum_{i=1}^{n} \) is number of FCIs and \( \sigma_{PC_1,2} \) represents the variance of the principal component 1, principal component 2, etc. Alternatively stated, the eigenvalues drop as we move from first principal component to the next one. The first principal component (PC1), which captures most of the variability in the FCIs, is essentially the Energy Futures Conditions Index (EFCI) model, where the second and subsequent principal components are uncorrelated with each other.

3. DATA

Data on the light sweet crude oil, natural gas and heating oil used is captured by the New York Mercantile Exchange (NYMEX) and provided by Commodity Futures Trading Commission (CFTC). Weekly data frequency is adopted, since the Commitment of Traders (COT) data is weekly based. CFTC categorizes traders who manage their business risks by hedging in futures as commercials, and the rest as non-commercials (CFTC, 2018). For the purpose of this study, net positions of speculators (hedgers) are calculated by taking the difference between non-commercial (commercial) long and short positions. Table 1 provides a summary of the heating oil, crude oil and natural gas futures markets and the correlation coefficients of the largest hedgers and speculators’ net positions. It can be deducted that reportable positions represent a significant portion of the total open interest, with a range of 0.39-0.96 in the natural gas market. The largest speculators and hedgers share strongly negative correlations across with correlation coefficients approaching –1. This is in line with Keynes (1930) and Gurrib (2009) who found that hedgers are mostly net short due to their requirements, to protect their investments from falling prices. Crude oil hedgers and speculators share the highest negative correlation of –0.996 suggesting that the largest speculators and hedgers in this markets take opposite positions.

Although crude oil and heating oil data are available since January 1986, data for the natural gas were available from April 1990. For consistency, weekly data are gathered across the three markets over October 2, 1992 – December 29, 2017. Nasdaq Composite Index data are collected from St Louis Federal Reserve Database (FRED). Other major market indices such as S&P 500 and the Dow Jones Industrial Average (DJIA) are not used, since the data availability dated back to 2008 only and we want to ensure consistency with the energy futures data. Table 2 reports the correlations among the Nasdaq Composite Index, net positions of large hedgers and speculators in the heating oil, crude oil and natural gas markets, and the total reportable positions (long and short) under each of the energy markets. Only speculators (hedgers) crude oil net positions were found to be strongly positively
(negatively) correlated with the three market indices. The net positions held in the three energy markets, for either players, are not strongly correlated with each other, suggesting initially that hedgers and/or speculators within each energy futures market are not affected by other energy market players’ net positions. However, on a broader basis, total long and short reportable positions in heating oil, crude oil and natural gas futures markets shared strongly correlations with each other across markets, suggesting markets reportable positions (long or short) are related to each other at a broad level, but not net positions of specific players among markets. For instance, the heating oil total reportable long positions are strongly positively correlated with the crude oil and natural gas with 94% and 88% correlation values. This is in line with EIA (2017) who reported correlations between daily futures price changes of crude oil with other commodity markets mostly rose during the period 2011–2017. Further, correlation values increase significantly when total reporting positions (long or short) are assessed against the equity market index. This is also in line with EIA (2017), which found stronger positive correlations between crude oil energy futures and financials such as S&P 500. As per Bloomberg (2018), crude oil and natural gas retain a significant 15% and 8% target weight in the Bloomberg Commodity Index. Figure 1 shows the net positions of hedgers and speculators in the heating oil, crude oil and natural gas markets. Key market players in energy futures markets reduced their net positions during the September 2008 financial crisis and 2000–2002 technology bubble. This was confirmed by the

Table 1. Contract specifications

<table>
<thead>
<tr>
<th>Futures market</th>
<th>Exchange</th>
<th>Contract size</th>
<th>Long reportable positions/OI</th>
<th>Short reportable positions/OI</th>
<th>Hedgers and speculators’ net positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heating oil</td>
<td>NYMEX</td>
<td>42,000 US gallons</td>
<td>0.51-0.91</td>
<td>0.61-0.94</td>
<td>–0.941</td>
</tr>
<tr>
<td>Crude oil, light sweet</td>
<td>NYMEX</td>
<td>1000 barrels</td>
<td>0.66-0.96</td>
<td>0.65-0.97</td>
<td>–0.996</td>
</tr>
<tr>
<td>Natural gas</td>
<td>NYMEX</td>
<td>10,000 MMBTU</td>
<td>0.39-0.96</td>
<td>0.41-0.98</td>
<td>–0.990</td>
</tr>
</tbody>
</table>

Note: NYMEX is the New York Mercantile Exchange, MMBTU is equal to 1 million British Thermal Units, OI is open interest.

Table 2. Correlation among heating oil, natural gas and crude oil net positions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Nasdaq Composite Index</th>
<th>HO NP-H</th>
<th>HO NP-S</th>
<th>CROIL NP-H</th>
<th>CROIL NP-S</th>
<th>NGAS NP-H</th>
<th>NGAS NP-S</th>
<th>HO TRP L</th>
<th>HO TRP S</th>
<th>COIL TRP L</th>
<th>COIL TRP S</th>
<th>NGAS TRP L</th>
<th>NGAS TRP S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq Composite Index</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HO NP-H</td>
<td>–0.11</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HO NP-S</td>
<td>0.15</td>
<td>–0.94</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CROIL NP-H</td>
<td>–0.82</td>
<td>0.28</td>
<td>–0.26</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CROIL NP-S</td>
<td>0.84</td>
<td>–0.24</td>
<td>0.24</td>
<td>–1.00</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NGAS NP-H</td>
<td>0.41</td>
<td>–0.05</td>
<td>0.09</td>
<td>–0.49</td>
<td>0.48</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NGAS NP-S</td>
<td>–0.45</td>
<td>0.07</td>
<td>–0.13</td>
<td>0.52</td>
<td>–0.51</td>
<td>–0.99</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HO TRP L</td>
<td>0.80</td>
<td>–0.18</td>
<td>0.27</td>
<td>–0.81</td>
<td>0.82</td>
<td>0.70</td>
<td>–0.74</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>HO TRP S</td>
<td>0.80</td>
<td>–0.24</td>
<td>0.31</td>
<td>–0.83</td>
<td>0.84</td>
<td>0.69</td>
<td>–0.74</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>COIL TRP L</td>
<td>0.81</td>
<td>–0.20</td>
<td>0.31</td>
<td>–0.84</td>
<td>0.85</td>
<td>0.62</td>
<td>–0.67</td>
<td>0.94</td>
<td>0.93</td>
<td>1.00</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>COIL TRP S</td>
<td>0.81</td>
<td>–0.21</td>
<td>0.31</td>
<td>–0.84</td>
<td>0.85</td>
<td>0.62</td>
<td>–0.68</td>
<td>0.94</td>
<td>0.93</td>
<td>1.00</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NGAS TRP L</td>
<td>0.77</td>
<td>–0.13</td>
<td>0.23</td>
<td>–0.80</td>
<td>0.81</td>
<td>0.57</td>
<td>–0.63</td>
<td>0.88</td>
<td>0.87</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
<td>–</td>
</tr>
<tr>
<td>NGAS TRP S</td>
<td>0.76</td>
<td>–0.13</td>
<td>0.23</td>
<td>–0.79</td>
<td>0.81</td>
<td>0.57</td>
<td>–0.63</td>
<td>0.88</td>
<td>0.87</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: HO, CROIL and NGAS represent the heating oil futures, crude oil futures and natural gas. NP-H and NP-S represent the net positions of large hedgers and large speculators respectively, and are calculated by taking the difference between long and short positions. TRP-L and TRP-S represent the total reportable positions, which are long and short, respectively. Bold figures represent correlations with r-squared values greater than 0.36.
decrease in the total reportable positions during those periods, and the stock market index, which retracted twice during these correction waves. Both total reportable positions and the stock market index resumed their long-run upward following crisis periods. For the later part of this study, only total reportable positions, both long and short, are included, since net positions across futures markets were weakly correlated.

4. RESEARCH FINDINGS

The results of the PCA are decomposed in Figure 2. As observed in the scree plot, the first principal component (PC1), which has an eigenvalue of 2.535, explains nearly 95% of all variations, which exist among all the total reportable positions in the three energy futures markets. The cumulative variability increases only slightly.
after including the second principal component (PC2), suggesting that the first principal component is sufficient to account for major variations among heating oil, crude oil and natural gas. The correlation circle supports that the second principal component only contribute to another 4.13% of the total variation in energy markets reportable positions. This is in line with relatively higher squared cosines values of EFCI compared to PC2 and PC3. Although not reported here, the eigenvalues for the second and third principal components drop significantly to 0.248 and 0.058, respectively. The eigenvectors for the first principal component of heating oil, crude oil and natural gas long (short) reportable positions are 0.404 (0.402), 0.415 (0.415), and 0.407 (0.407), with correlations between the EFCI and HO, CROIL and NGAS at 0.963 (0.960), 0.990 (0.990) and 0.970 (0.971), respectively.

Figure 3 displays the EFCI and total reportable long positions over the period 1992–2017. Although not displayed here, EFCI and total reportable short positions shared similar relationships. As observed in the three graphs, EFCI tracked closely the performance of the three energy markets, including the 2007–2008 global financial crisis and to a less extent the 2000–2001 technology bubble. It is also important to note that the natural gas market experienced more than usual fluctuations in its total reportable long positions during the period 2012–2014, strong inventories, production growth and warmer than normal winter seasons like the El Nino phenomena (EIA, 2016). The total reportable short positions in the natural gas market also observed similar abrupt volatility change not captured by the EFCI, suggesting specific rather than broad energy market factors affected markets like natural gas.

### Table 1: Squared cosines of the variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>EFCI</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
</tr>
</thead>
<tbody>
<tr>
<td>HO TRP L</td>
<td>0.928</td>
<td>0.068</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>HO TRP S</td>
<td>0.925</td>
<td>0.075</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>CROIL TRP L</td>
<td>0.979</td>
<td>0.002</td>
<td>0.019</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>CROIL TRP S</td>
<td>0.980</td>
<td>0.001</td>
<td>0.018</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>NGAS TRP L</td>
<td>0.941</td>
<td>0.051</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>NGAS TRP S</td>
<td>0.942</td>
<td>0.051</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Note: HO, CROIL and NGAS represent the heating oil futures, crude oil futures and natural gas. TRP-L and TRP-S represent the total reportable positions, which are long and short, respectively. Values in bold correspond for each variable to the factor for which the squared cosine is the largest.*

**Figure 2.** Principal component analysis

**Figure 3.** Scree plot

[Diagram of variables (axes EFCI and F2: 98.97%), squared cosines of the variables, and scree plot showing cumulative variability (%).]
In line with Gurrib (2018) who looked at the relationship between major currency futures and major financial conditions indexes, IMF (2017) and Stock and Watson (2002) who used PCA to predict excess stocks returns and macroeconomic variables over different time periods, and EIA (2017) who postulated energy markets like crude oil share similar risk and return relationships with stocks in the last decade, this study extends the application of PCA by analyzing the effect of the proposed EFCI onto the Nasdaq Composite Index, which represents a market capitalization weighted index of roughly 3000 companies on the Nasdaq Stock Exchange. While the Root Mean Squared Error (RMSE) is widely used in literature, the Normalized Root Mean Squared Error values (NRMSE) are also adopted here to allow for the difference in the units, when sampled over different periods. The following model linking the equity market index with the energy futures conditions index is proposed:

**Figure 3. Energy Futures Conditions Index model (EFCI)**

Note: The Energy Futures Conditions Index (EFCI) is displayed on the right hand side vertical axis. HO, CROIL and NGAS represent the heating oil futures, crude oil futures and natural gas. NP-H and NP-S represent the net positions of large hedgers and large speculators respectively, and are calculated by taking the difference between long and short positions. TRP-L represent the total reportable positions, which are long in their futures positions.
where \( i \) represents the Nasdaq Composite Index, \( EFCI_{t-n} \) is the Energy Futures Conditions Index, where \( n \) ranges from 1, 2, 3, 10, and is used to estimate \( M_{t} \), which represents the current market indices values. Current EFCI values are also regressed against the current market index for comparison purposes. The NRMSE is the RMSE adjusted to the difference between the minimum and maximum observed value of the EFCI. Table 3 reports the forecast errors based on the NRMSE and RMSE of the model in equation 4, using current values and 1, 2, 3 and 10 weekly lags in the \( EFCI \) data. The RMSE for the market index increased slightly, as the number of lags increased. Due to the non-sensitivity of forecast errors as number of lags is increased, a 1 week lag in the \( EFCI \) is retained as a factor for predicting market index values.

Further, this study captures how the estimated model helps to explain actual market index values. Figure 4 displays the actual and estimated market index values over the period 1992–2017, including a lower and upper boundary level set at 2 standard deviations. As observed, the estimated values of the market index tracked closely the actual values. The only noticeable exception was the heightened volatility observed in the 2000 period, which was caused by the technology bubble. Our model, which is based on the energy futures index, failed to capture this event, as observed earlier in Figure 3, where energy futures did not witness similar impacts during the same period as those experienced by the equity market index. Table 4 reports the \( r^{2} \)-squared values, \( p \)-values of the \( EFCI_{t-1} \) coefficient and \( F \)-statistics. Coefficient of determination values fluctuating between 0.66 to 0.81, and \( p \)-values of both the independent variable and \( F \)-statistics at zero, suggest that the energy futures conditions index is significant in explaining next week’s equity market index value.

### Table 3. Forecast errors

<table>
<thead>
<tr>
<th>Lags</th>
<th>Root Mean Squared Error (RMSE)</th>
<th>Normalized RMSE (NRMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t )</td>
<td>( t-1 )</td>
</tr>
<tr>
<td>Nasdaq Composite Index</td>
<td>820</td>
<td>821</td>
</tr>
</tbody>
</table>

Note: Nascompf represents the estimated values of the Nasdaq Composite Index. Two standard errors are used to calculate the lower and upper bounds.

Further, this study captures how the estimated model helps to explain actual market index values. Figure 4 displays the actual and estimated market index values over the period 1992–2017, including a lower and upper boundary level set at 2 standard deviations. As observed, the estimated values of the market index tracked closely the actual values. The only noticeable exception was the heightened volatility observed in the 2000 period, which was caused by the technology bubble. Our model, which is based on the energy futures index, failed to capture this event, as observed earlier in Figure 3, where energy futures did not witness similar impacts during the same period as those experienced by the equity market index. Table 4 reports the \( r^{2} \)-squared values, \( p \)-values of the \( EFCI_{t-1} \) coefficient and \( F \)-statistics. Coefficient of determination values fluctuating between 0.66 to 0.81, and \( p \)-values of both the independent variable and \( F \)-statistics at zero, suggest that the energy futures conditions index is significant in explaining next week’s equity market index value.

Source: Author.

![Figure 4. Actual and estimated market index (1992–2017)](http://dx.doi.org/10.21511/imfi.15(4).2018.01)
Table 4. Regression statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>R-squared</th>
<th>$EFCI_{t-1}$</th>
<th>F-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq Composite Index</td>
<td>0.660</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Diagnostic tests – while the $r$-squared values and $p$-values of EFCI coefficients and $F$-statistics point to a reliable forecast model initially, in order to validate the use of the model based on equation 4, it is important to carry out some diagnostic tests on the model. While not reported here, the model from equation 4 suffers from non-normal distribution, autocorrelation and heteroscedasticity. While non-normal distribution was expected due to positively skewed and kurtosis values reported earlier, autocorrelation and heteroscedasticity needs to be addressed. The non-normality in our model is consistent with Hilliard and Reis (1999) who found non-normality in most futures markets. While we are using total reportable positions, our study is also consistent with Blattberg and Gonedes (1974) who found leptokurtic distributions for hedgers and speculators. To potentially eliminate autocorrelation in the model, a 1 week lag of the market index is included as an independent variable. To make the model lean more towards homoscedasticity, both equity and energy futures indexes are transformed in logarithmic variables. A positive constant is imposed on EFCI values to avoid negative logarithmic calculations. The updated model is represented as follows:

$$M^i_t = \alpha + \beta EFCI_{t-1} + \alpha' M^i_{t-1} + \epsilon_t, \quad (5)$$

where $i$ represents the Nasdaq Composite Index, $EFCI_{t-1}$ is the Energy Futures Conditions Index and is used to estimate $M^i_{t-1}$, which represents the current equity market index values, $M^i_{t-1}$ is the one-week lag variable of the market index.

As observed in Table 5, the high $r$-squared values are attributed to the independent lagged market index value. The lagged EFCI coefficient was significant only at 10% significance level. Due to the logarithmic transformations, the Jarque-Bera normality test rejected the hypothesis of normally distributed residual data. More importantly, the Breusch-Godfrey autocorrelation test reported $p$-values greater than 5% for the Nasdaq Composite Index, suggesting the removal of autocorrelation in the model. Similarly, the Breusch-Pagan-Godfrey heteroscedasticity test reported $p$-values greater than 5%, suggesting a homoscedastic model.

Due to the inability of the EFCI model to capture the 2000–2001 technological bubble global event, as observed in Figure 4, it is important to ascertain if indeed there is a break during this time period. The whole sample period is tested for any significant breakpoint in line with Bai-Perron (2003), and findings shows a structural break around March 2000. While not reported here, the residual plot showed a spike around that period. Table 6 reports the pre and post technological bubble crisis robustness test results. While the Jarque-Bera test continues to point towards non-normality, the results are mixed in terms of the removal of autocorrelation and heteroscedasticity in the model. The Breusch-Pagan-Godfrey heteroscedasticity test suggest that the error variances are not all equal in both the pre and post financial crisis model. This contradicts earlier findings when the model was found to be homoscedastic, when applied over the full sample size. The EFCI lagged coefficient was found significant in forecasting Nasdaq Composite Index values following the crisis. The $p$-values of the $F$-statistics remained at ze-

Table 5. Regression statistics and diagnostic tests

<table>
<thead>
<tr>
<th>Regression statistics</th>
<th>R-squared</th>
<th>Log($EFCI_{t-1}$)</th>
<th>F-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq Composite Index</td>
<td>0.997</td>
<td>0.081</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Diagnostic Tests

<table>
<thead>
<tr>
<th>Diagnostic Tests</th>
<th>Jarque-Bera normality test</th>
<th>Breusch-Godfrey autocorrelation LM test</th>
<th>Breusch-Pagan-Godfrey heteroscedasticity test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample period</td>
<td>10/16/1992-12/29/2017</td>
<td>0.380</td>
<td>0.232</td>
</tr>
<tr>
<td>Nasdaq Composite Index</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For the regression statistics, only the $p$-values of the log ($EFCI_{t-1}$) and $F$-statistics are displayed. For the Jarque-Bera normality test, only the $p$-value of the Jarque-Bera test statistic is reported. For the Breusch-Godfrey serial correlation LM test, the $p$-value of observation$r$-squared value is shown. Two residual lags are used. For the heteroscedasticity test, only the $p$-value of the observation$r$-squared value is displayed. The Breusch-Pagan-Godfrey test is used.
Investment Management and Financial Innovations, Volume 15, Issue 4, 2018

Table 6. Pre and post crisis robustness test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Period</th>
<th>Jarque-Bera normality test</th>
<th>Breusch-Godfrey autocorrelation LM test</th>
<th>Breusch-Pagan-Godfrey heteroskedasticity test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre bubble crisis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasdaq Composite Index</td>
<td>10/16/1992–4/07/2000</td>
<td>0.000</td>
<td>0.528</td>
<td>0.000</td>
</tr>
<tr>
<td>Post bubble crisis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasdaq Composite Index</td>
<td>4/14/2000–12/29/2017</td>
<td>0.000</td>
<td>0.728</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Note: For the regression statistics, only the $p$-values of the log ($EFCI_{t-1}$) and $F$-statistics are displayed. For the Jarque-Bera normality test, only the $p$-value of the Jarque-Bera test statistic is reported. For the Breusch-Godfrey serial correlation LM test, the $p$-value of number of observations$r$-squared value is shown. Two lags are used as standard number lags in the residuals. For the heteroskedasticity test, only the $p$-value of the number of observations$r$-squared value is displayed. The Breusch-Pagan-Godfrey test is used. The whole sample is tested for any significant breakpoint using the Bai-Perron (2003) breakpoint test.

ro, with however, heteroskedasticity presence in both post and pre crisis periods. Findings suggest that the proposed model in equation 5 is influenced by the sample size, and not consistently reliable. The proposed Energy Futures Conditions Index, despite capturing 95% of variability in the three energy futures under analysis, and despite explaining most of the movements in the equity market index, failed in the diagnostic parts, where it revealed non-homoscedastic presence, when the sample is broken down into pre and post crisis periods.

Due to the mixed results obtained above when carrying out the diagnostic tests over different sample period, to avoid the possibility of spurious regressions, stationarity test is carried out using the Augmented Dickey-Fuller (ADF) test at 1%, 5%, and 10% level. The ADF test is conducted using equation variations, namely, with an intercept, trend and intercept, and no intercept and trend. The number of lags is selected by minimizing the Schwarz Information Criteria (SIC). The $t$-statistics result are reported in Table 7. To account for ADF sensitivity in the presence of potential breakpoints, the test is conducted over full, post and pre breakpoints periods. The full sample is broken down into pre and post 2000 crisis periods. The breakpoint date is based on the breakpoint unit root test with trend and intercept included in the specification, with a breakpoint objective of minimizing the intercept break $t$-statistic as proposed in Qu and Perron (2007). The first graph in Figure 5 displays the Dickey-Fuller autoregressive coefficients supporting the break date at 03/10/2000 for the Nasdaq Composite Index. Compared to the Nasdaq Composite Index, which displayed a break around the 2000 technology bubble crisis, the energy futures conditions index exhibited a break around the latest global financial crisis, more specifically at 07/20/2007 as per the second graph in Figure 5. This suggests that at this stage, both series, while being stationary, after differencing, might not be sharing strong relationship in forecasting each other. Under the full 1992–2017 sample data, while the Nasdaq Composite Index and EFCI were non-stationary at levels, they were all found to be stationary after 1$^{st}$ order differencing, where the ADF $t$-statistics were greater than the critical values at 1%, 5% and 10% levels. Under the pre 2000 bubble period, the stock market index was stationary after 1$^{st}$ order differencing. Post 2000 crisis period, the ADF tests suggest stationarity after 1$^{st}$ order differencing, except in one instance, where the Nasdaq Composite Index level data were found to be stationary at 10% level, when an intercept and trend is included in the ADF model. After 1$^{st}$ order differencing, all post 2000 crisis data were stationary at 1%, 5% and 10% significance levels. Similarly, the EFCI, in both the pre and post 2007 breakpoint, was found to be stationary after 1$^{st}$ order differencing. Differenced values are used for both EFCI and Nasdaq Composite Index in further sections of the study.

As observed in Table 8, the relationship between the Nasdaq Composite Index and the EFCI is tested for heteroskedasticity using the Breusch-Pagan-Godfrey test. Based on earlier breakpoint dates noticed under both the Nasdaq Composite Index and EFCI earlier, the heteroskedasticity test is carried out over different sample periods, namely, a full period 1992–2017, pre and post 2000 crisis.
period, and pre and post 2007 crisis period. While there is no discernible relationship or prior studies relating the EFCI and stock market index, both EFCI and the Nasdaq Composite index are used as dependent and independent variables and tested for heteroskedasticity. Findings from Table 8 support that the model is homoscedastic in the error terms in both the 2000 and 2007 pre and post crisis periods, except for the pre 2000 crisis period, where heteroskedasticity presence was noted at 5% level, when both the EFCI and Nasdaq Composite Index were used as independent and dependent variables. This result questions the presence of homoscedasticity when tested over the full 1992–2017 sample period. The presence of heteroskedasticity in the pre 2000 crisis can be explained

### Table 7. Stationarity test

<table>
<thead>
<tr>
<th>Variable</th>
<th>1992–2017</th>
<th>Pre structural break</th>
<th>Post structural break</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq Composite Index</td>
<td>0.687 –0.7 2.136</td>
<td>4.713 4.890 4.466</td>
<td>0.520 –3.326** 0.867</td>
</tr>
<tr>
<td>Energy Futures Conditions Index</td>
<td>1.031 –1.939 0.931</td>
<td>1.403 –0.452 –1.262</td>
<td>0.235 –2.697 1.382</td>
</tr>
<tr>
<td></td>
<td>–11.35 –11.07 –11.22</td>
<td>–9.06 –9.27 –8.79</td>
<td>–6.92 –7.06 –6.77</td>
</tr>
</tbody>
</table>

Note: The stationarity test is based on the Augmented Dickey-Fuller (ADF) test statistic at 1%, 5% and 10% significance level. The test is conducted over the period 1992–2017 and pre and post 2000 bubble crisis periods. The first row under the Nasdaq Composite Index and EFCI series report the t-statistic using level data. The second row reports t-statistic for 1st differenced data. Italic values are stationary after 1st order differencing at 1%, 5% and 10% level. ** values were found to be stationary at level, since the ADF test statistic was greater than the test critical value at 10% significance level. The breakpoint date is based on the breakpoint unit root test, with trend and intercept included in the trend specification. EFCI is the Energy Futures Conditions Index.

![Figure 5. Dickey-Fuller autoregressive coefficients](http://dx.doi.org/10.21511/imfi.15(4).2018.01)
by the fact that the energy futures index was not affected by the 2000 crisis, relative to the Nasdaq Composite Index, which manifested increased volatility around the 2000 crisis period. The regression model of EFCI as a dependent variable, with the stock market index as an independent variable, produced variance in the errors terms, which were non constant. To ensure robustness in further tests and analysis, only pre and post 2007 crisis are used to assess the relationship between the stock market and energy futures indexes.

Table 9 provides a summary of the forecast evaluation results upon regressing the Nasdaq Composite Index onto the EFCI and vice versa. Root Mean Squared Errors (RMSE) for forecasting 2007–2017 EFCI values is provided using pre and post 2007 crisis period data. As observed in Table 9, the use of the pre 2007 crisis period Nasdaq Composite Index to predict EFCI values, and the use of EFCI to predict Nasdaq Composite Index, produced relatively higher forecast errors compared to the use of only post 2007 data to predict 2007–2017 EFCI and Nasdaq Composite Index values. This suggests that the use of post 2007 crisis data provided less deviation in the forecasted values compared to the actual values. As displayed in Panel A of Figure 6, the use of 2000–2007 EFCI values captured forecasted values of the stock market index, within ±2 standard errors, except in late 2008. This can be explained by events like Lehman Brothers collapse in late 2008, which had a more substantial effect on the stock market index compared to energy futures markets, which is represented here by the EFCI. Compared to Panel A, where forecasted values of the stock market index were lower than actual values, Panel B displays actual values of the stock market index being lower than forecasted values. Both actual and forecasted values of the Nasdaq Composite Index are trending upwards for the period 2007–2017. Lastly, but not least, while the use of 2000–2007 stock market values was useful in predicting 2007–2017 EFCI values, the second graph under Panel B, supports that using post 2007 stock market values help predict EFCI values better, with more instances where the actual and forecasting EFCI values hardly deviate hardly from each other. Both EFCI actual and forecasted values trended upwards, with forecasted values being closer to the actual values. Although not reported here, using the Breusch-Godfrey serial correlation LM test and Breusch-Pagan-Godfrey heteroskedasticity test, the post 2007 model was still robust with no serial correlation and homoscedastic errors. The gap between the lower and upper forecasting bounds in all four graphs can be attributed to relatively high bias proportions, witnessed by the increasing squared deviations between forecasted and actual mean values of the dependent variables.

Table 9. Forecast evaluation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq Composite Index</td>
<td>0.9261</td>
<td>0.0293</td>
<td>0.9568</td>
<td>0.8831</td>
<td>0.4368</td>
</tr>
<tr>
<td>EFCI</td>
<td>0.1166</td>
<td>0.0000</td>
<td>0.2115</td>
<td>0.8824</td>
<td>0.1416</td>
</tr>
</tbody>
</table>

Table 8. Heteroskedasticity test

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq Composite Index</td>
<td>0.9261</td>
<td>0.0293</td>
<td>0.9568</td>
<td>0.8831</td>
<td>0.4368</td>
</tr>
<tr>
<td>EFCI</td>
<td>0.1166</td>
<td>0.0000</td>
<td>0.2115</td>
<td>0.8824</td>
<td>0.1416</td>
</tr>
</tbody>
</table>

Note: The heteroskedasticity test is based on Breusch-Pagan Godfrey test. The Nasdaq Composite Index is regressed against EFCI, followed by the EFCI regressed onto the Nasdaq Composite Index. 2000 crisis is based on the unit root breakpoint test, where 03/10/2000 was found to be the breakpoint date. 2007 crisis is based on the unit root breakpoint test, where 07/20/2007 was found as a break date. Pre (post) 2000 crisis is set from 10/09/1992 till 03/10/2000 (03/10/2000–12/29/2017) and Pre (Post) 2007 crisis is set from 10/09/1992 till 07/20/2007 (07/20/2007–12/29/2017). The italic values are the significant p-values of observation*r-squared values, which support the presence of homoscedasticity.
While the above analysis demonstrates some forecasting abilities between the Energy Futures Conditions Index (EFCI) and the stock market index, it is worthwhile to assess whether the energy market driven data or the equity market one is leading the other. To this end, the Granger causality test as proposed by Granger (1980) is carried out and reported the two markets. Grosche (2012) provides a good overview of the use of Granger causality test in commodity futures markets. While not reported here, both differenced EFCI and Nasdaq Composite Index were found not to Granger cause each other, where the $p$-values were 0.54 and 0.38, leading to a non-rejection of the null hypothesis that the differenced EFCI values do not Granger cause the differenced Nasdaq Composite Index values and vice versa. This suggests that while the Energy Futures Conditions Index can predict the Nasdaq Composite Index and vice versa, in a not serially correlated and homoscedastic model, the findings suggest that information does not flow across the two markets. This is also supported by earlier findings that structural breaks in the two different markets occurred at different instances during the period 1992–2017.

CONCLUSION

This paper introduces an energy futures index based on the most actively traded energy futures contracts in the US. The use of principal component analysis allows the energy futures index to capture nearly 95% of the variability existing in the crude oil, heating oil and natural gas futures markets, where US and China are the leading producers and consumers. Initially, the proposed energy futures index model produced stable forecast errors over different lags imposed, and explained most of the actual market index values of the Nasdaq Composite index. However, diagnostic tests revealed non normal,
auto-correlated and heteroskedasticity. Logarithmically transformed and calibrated EFCI and stock market index data, and the inclusion of a 1 week lagged stock market index independent variable, resulted in a non-normal, non-auto correlated and homoscedastic model, when tested over the full 1992–2017 sample. To account for potential breakpoints in the model, the sample was broken down into pre and post technology bubble crisis periods and produced mixed robustness results. Differenced stock market and energy futures conditions index data led to stability in the model, where the model was resilient to serial correlation and heteroskedasticity. While the stock market index witnessed a structural break in 2007, energy futures experienced a similar break in 2000. The use of 2000–2007 stock market index to predict 2007–2017, values within 2 standard errors, captured most movements in the energy futures index, and vice versa. Although using post 2007 crisis data produced more accurate forecasting results, findings were subject to relatively large bias proportions. This was further supported, since neither the energy futures index nor stock market index granger caused each other. This suggests that the two markets, while witnessing homoscedastic errors when brought together in a relationship model, show weak sign of predictability and cross market information flows. Further research can tap into expanding the application of energy futures conditions index model onto other commodity futures and spot markets.

REFERENCES


